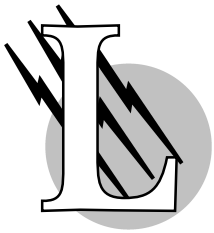


Laser–Materials Interactions

Michael Bass
Aravinda Kar
University of Central Florida

I. Deposition of Light Energy in Matter
II. Laser Heating of Materials
III. Responses of Materials

GLOSSARY

Complex propagation constant Quantity $\gamma(=\alpha + j\beta)$ used to describe the propagation of light in a general medium; α gives the attenuation and β the wave vector magnitude, both in reciprocal centimeters.

Laser-driven plasma Optically thick plasma found near a surface heated to the point of vaporization by an intense laser beam; the beam ionizes some of the vapor and through the process of avalanche breakdown creates the plasma.

Nonlinear absorption Process in which the attenuation coefficient itself is a function of the light intensity.

Photoablation Process by which short-wave-length laser light breaks molecular bonds in matter and causes material to be ejected from the surface with very little heating of the sample.

Surface electromagnetic wave Wave formed on an irradiated surface when an irregularity on the surface scatters the light energy; the amount of energy coupled can be significant.

THE INTERACTION of laser light with matter is in many ways like the interaction of ordinary light with matter. The difference between them and the interest in laser light arises from the fact that laser light can be much more intense than light from other sources. This property of laser light has opened new research areas in nonlinear optics, laser plasma generation, and new types of materials processing. It has also led to applications such as laser cutting, drilling, welding, and heat treating. The high fluxes possible allow consideration of such futuristic applications as power generation through inertial confinement fusion, X-ray generation for microelectronics, and laser weaponry.

I. DEPOSITION OF LIGHT ENERGY IN MATTER

A. Absorption

The interaction between light, a form of electromagnetic energy, and materials takes place between the optical electric field and the most nearly free electrons in the medium. In quantum mechanical terms, one speaks of interactions between photons and electrons in the presence of the lattice of nuclei. The lattice must be present to ensure momentum conservation. For the purposes of this article, the classical viewpoint of an electric field causing electrons to move in the presence of damping or frictional forces is sufficient to describe the process of optical absorption.

Conduction band electrons in metals are able to interact with an optical electric field. Their motion is damped by collisions with the vibrating lattice and so some of the light energy is transferred to the lattice. In this manner the material is heated. In semiconductors the motions of both electrons in the conduction band and holes in the valence band must be considered. In dielectrics the electrons are effectively bound to the atoms or molecules that compose the material. The applied optical field induces a polarization in the material. Upon relaxation some of the energy in the polarization is coupled to the lattice and the material is heated. These processes of absorption of energy from an optical field can be treated by classical electromagnetics. That is, Maxwell's equations, the constitutive equations of matter, and the boundary conditions for each material can be solved in each case.

1. Absorption from an Electromagnetic Point of View

a. Propagation in a general dielectric material. Let the propagation of the optical electromagnetic wave in a medium be given by the complex propagation constant γ where

$$\gamma = \alpha + j\beta, \tag{1}$$

so that a plane wave electric field is described by

$$E(z,t) = \mathrm{Re}\{E(0)\exp(-\gamma z)\exp[j(\omega t - \beta z)]\}. \tag{2}$$

Here ω is the radian frequency of the field, α the attenuation coefficient, and β the propagation coefficient.

A dielectric is described by the complex dielectric constant

$$\varepsilon = \varepsilon_1 + j\varepsilon_2, \tag{3}$$

and a magnetic susceptibility μ. This enables us to define a complex index of refraction

$$\eta = n + jk, \tag{4}$$

in which n determines the propagation vector and k determines the attenuation of the wave in the medium. It is easy to show that

$$n = \left\{\left[\left(\varepsilon_1^2 + \varepsilon_2^2\right)^{1/2} + \varepsilon_1\right]/2\right\}^{1/2} \tag{5}$$

and

$$k = \left\{\left[\left(\varepsilon_1^2 + \varepsilon_2^2\right)^{1/2} + \varepsilon_1\right]/2\right\}^{1/2}. \tag{6}$$

By applying Maxwell's equations and the dielectric boundary conditions as to the case of a lossy dielectric we find that

$$\gamma = j\omega\{(\mu\varepsilon_1)[1 - j(\varepsilon_2/\varepsilon_1)]\}^{1/2}. \tag{7}$$

Note that if we had treated the lossy dielectric as a medium having a finite conductivity σ, we would have obtained

$$\gamma = j\omega\{(\mu\varepsilon)[1 - j(\sigma/\omega\varepsilon)]\}^{1/2}, \tag{8}$$

where ε is the dielectric permittivity. It is clear that Eqs. (7) and (8) are identical with ε replaced by ε_1 and σ/ω by $-\varepsilon_2$.

Lossless dielectrics are media having σ identically zero. The complex propagation constant is then purely imaginary and there is no absorption in the medium.

b. Perfect metals. A perfect metal is a material having an infinite conductivity or a material in which

$$\alpha/\omega\varepsilon \gg 1.$$

In this case we find

$$\alpha = \beta = (\omega\mu\sigma/2)^{1/2}, \tag{9}$$

which means that a field propagating in a metal will be attenuated by a factor of $1/e$ when it has traveled a distance

$$\delta = (2/\omega\mu\sigma)^{1/2}. \tag{10}$$

The quantity δ is called the skin depth and at optical frequencies for most metals it is ~50 nm or about one-tenth of the wavelength of green light. After a light beam has propagated one skin depth into a metal, its intensity is reduced to $1/e^2$ or 0.135 of its value on the surface.

The energy that is no longer in the electromagnetic wave when light is attenuated in a metal drives conduction currents in the so-called skin layer. These in turn produce heating through ohmic losses and the light energy thereby appears in the metal as heat. Heating of the metal at depths below the skin layer takes place by means of thermal conduction.

c. Lossy, nonconductive dielectrics. As mentioned above, the results derived for a general dielectric are equivalent to those for a metal when the proper substitutions are made for ε_1 and ε_2. This allows us to consider the case where

$$\varepsilon_2/\varepsilon_1 \gg 1$$

and write the attenuation coefficient for the lossy dielectric as

$$\begin{aligned}\alpha &= \omega(\mu\varepsilon_2/2)^{1/2} \\ &= (2\pi/\lambda_0)(\mu_R\varepsilon_{2R}/2)^{1/2},\end{aligned} \tag{11}$$

where $(\mu_0\varepsilon_0)^{1/2}$ is recognized as the speed of light in free space, $\omega = 2\pi c/\lambda_0$, and the subscript "R" indicates the relative permittivity and susceptibility. For electromagnetic radiation with a free space wavelength of 1000 nm propagating in such a dielectric we see that

$$\alpha > 10^4\ \mathrm{cm}^{-1}.$$

This means that the electromagnetic energy will be reduced to $1/e^2$ of its value within 10^{-4} cm.

The nonconductive dielectric is a medium with very few free or conduction band electrons at room temperature and so we cannot consider ohmic losses as the mechanism for heating. Its interactions with electromagnetic radiation involve valence band electrons. When such electrons in this type of material absorb light energy they can be raised either to the conduction band or to some impurity state lying within the band gap. They then relax back to the valence band. The process of relaxation can be radiative (fluorescence and phosphorescence) or nonradiative. In either case phonons (lattice vibrations) are generated and the material is heated. The heat generation is localized to the region in which the absorption occurs and, as in metals, heating beyond this region takes place by thermal heat conduction.

2. Lorentz and Drude Models

The classical theory of absorption in dielectric materials is due to H. A. Lorentz and in metals it is the result of the work of P. K. L. Drude. Both models treat the optically active electrons in a material as classical oscillators. In the Lorentz model the electron is considered to be bound to the nucleus by a harmonic restoring force. In this manner, Lorentz's picture is that of the nonconductive dielectric. Drude considered the electrons to be free and set the restoring force in the Lorentz model equal to zero. Both models include a damping term in the electron's equation of motion which in more modern terms is recognized as a result of electron–phonon collisions.

These models solve for the electron's motion in the presence of the electromagnetic field as a driving force. From this, it is possible to write an expression for the polarization induced in the medium and from that to derive the dielectric constant. The Lorentz model for dielectrics gives the relative real and imaginary parts of the dielectric constant as

$$\varepsilon_{1R} = 1 + \left(Ne^2/\varepsilon_0 m\right)\frac{\omega_0^2 - \omega^2}{\left(\omega_0^2 - \omega^2\right)^2 + \Gamma^2\omega^2} \tag{12}$$

and

$$\varepsilon_{2R} = \left(Ne^2/\varepsilon_0 m\right)\frac{\Gamma\omega}{\left(\omega_0^2 - \omega^2\right)^2 + \Gamma^2\omega^2}. \tag{13}$$

In these expressions N is the number of dipoles per unit volume, e the electron charge, m the electron mass, Γ the damping constant, ω_0 the resonance radian frequency of the harmonically bound electron, ω the radian frequency of the field, and ε_0 the permittivity of free space. Equations (12) and (13) are sketched in Fig. 1. The range of frequencies where ε_1 increases with frequency is referred

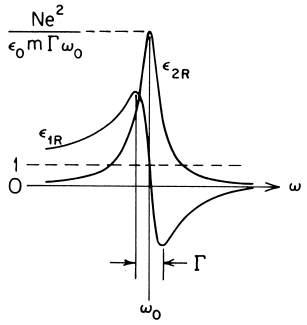


FIGURE 1 Frequency dependences of ϵ_{1R} and ϵ_{2R}.

to as the range of normal dispersion, and the region near $\omega = \omega_0$ where it decreases with frequency is called the range of anomalous dispersion.

In the preceding discussion the contributions of the electronic polarizability to the dielectric constant were considered. Other contributions occur such as stimulation of vibrations of ions in ionic crystals. This type of contribution is very small at optical frequencies because of the large mass of the ions compared to that of the electrons. In other words, the ionic polarizability is much smaller than the electronic polarizability at optical frequencies. We can therefore consider only the electronic terms when evaluating optical absorption. By inserting Eq. (13) into Eq. (11) we obtain the absorption coefficient of the lossy, nonconductive dielectric in terms of the properties of the electron's kinematics assumed by Lorentz. That is,

$$\alpha = (2\pi/\lambda)(\mu_R/2)\left(Ne^2/\varepsilon_0 m\right)\left(\frac{\Gamma\omega}{\left(\omega_0^2 - \omega^2\right)^2 + \Gamma^2\omega^2}\right)^{1/2}. \tag{14}$$

The Drude model for metals assumes that the electrons are free to move. This means that it is identical to the Lorentz model except that ω_0 is set equal to zero. The real and imaginary parts of the metal's dielectric constant are then given by

$$\varepsilon_{1R} = 1 - \left(Ne^2\varepsilon_0 m\right)\frac{1}{\omega^2 + \Gamma^2} \tag{15}$$

$$\varepsilon_{2R} = \left(Ne^2\varepsilon_0 m\right)\frac{\Gamma}{\omega(\omega^2 + \Gamma^2)}. \tag{16}$$

The quantity Γ is related to the mean time between electron collisions with lattice vibrations. T (i.e., to the problem of electron–phonon scattering). By considering the motion of electrons able to make collisions with lattice vibrations in an electric field E having radian frequency ω, it is straightforward to show that the average velocity is

$$v = -\frac{eE}{m}\frac{\mathrm{T}}{1 - j\omega\mathrm{T}}. \tag{17}$$

The conductivity at this frequency is then

$$\sigma = \frac{Ne^2\mathrm{T}}{m}\frac{1}{1-j\omega\mathrm{T}}, \tag{18}$$

where the dc conductivity is given by

$$\sigma_{\mathrm{dc}} = Ne^2\mathrm{T}/m. \tag{19}$$

From Eqs. (15) and (16) we see that if we allow $\Gamma = 1/\mathrm{T}$, then

$$\varepsilon_1 = 1 - \left(\frac{\sigma_{\mathrm{dc}}}{\varepsilon_0}\right)\mathrm{T}\frac{1}{\omega^2\mathrm{T}^2+1} \tag{20a}$$

and

$$\varepsilon_2 = \left(\frac{\sigma_{\mathrm{dc}}}{\varepsilon_0}\right)\left(\frac{1}{\omega}\right)\frac{1}{\omega^2\mathrm{T}^2+1}. \tag{20b}$$

At electromagnetic field frequencies that are low, that is, when $\omega\mathrm{T} \ll 1$, we have

$$\varepsilon_1 = 1 - \sigma_{\mathrm{dc}}\mathrm{T}/\varepsilon_0 \tag{21}$$

and

$$\varepsilon_2 = \sigma_{\mathrm{dc}}/\varepsilon_0\omega \tag{22}$$

At such frequencies $\varepsilon_2 \gg \varepsilon_1$, and since $\Gamma = 1/\mathrm{T}$ Eq. (14) gives

$$\alpha = (\omega\mu\sigma/2)^{1/2}, \tag{23}$$

which is exactly the result we obtained earlier when treating absorption from an electromagnetic point of view. In other words, the optical properties and the dc conductivity of a perfect metal are related through the fact that each is determined by the motion of free electrons. At high frequencies transitions involving bound or valence band electrons are possible and there will be a noticeable deviation from this simple result of the Drude model. However, the experimental data reported for most metals are in good agreement with the Drude prediction at wavelengths as short as 1 μm.

3. Temperature Dependence

Another aspect of the absorption of light energy by metals that should be noted is the fact that it increases with temperature. This is important because during laser irradiation the temperature of a metal will increase and so will the absorption. The coupling of energy into the metal is therefore dependent on the temperature dependence of the absorption. This property is easy to understand if we remember that all the light that gets into a metal is absorbed in it. The question that must be addressed is how the amount of incident optical energy that is not reflected from the metal's surface depends on temperature. Recalling the Fresnel expression for electric field reflectance and applying to it the real and imaginary parts of the complex index of refraction for a metal–air interface, we can write the field reflectivity. Multiplying this by its complex conjugate, we find the intensity reflection coefficient for a metal

$$R_{\mathrm{I}} = 1 - 2\mu\varepsilon_0\omega/\sigma \tag{24}$$

Since the conductivity σ decreases with increasing temperature, R_1 decreases with increasing temperature. As a result, more incident energy actually gets into the metal and is absorbed when the temperature is raised. This is true even though the absorption at high temperatures takes place in a deeper skin depth.

B. Nonlinear Absorption

Since the advent of lasers made possible highintensity optical fields it has been possible to explore interactions of light with matter in which the response of the material is not linear with the optical electric field. This subject, known as nonlinear optics, has become a major field of research and has led to such useful optical devices as frequency converters, tunable parametric devices, and optically bistable elements. This section presents a semiclassical discussion of nonlinear optics and how it gives rise to nonlinear absorptions, in particular two-photon absorption (TPA).

The polarization of a medium and the optical electric field applied to it are linked by the material's susceptibility X, a tensor quantity. In the previous section we considered the limit of small optical fields, where the susceptibility is a function of the dielectric constants only and is independent of the field. In this case the polarization vector **P** is related to the optical electric field **E** by the expression

$$\mathbf{P} = \mathbf{X} * \mathbf{E} \tag{25}$$

Equation (25) is the relationship on which optics was built prior to 1961. As a result of lasers and the high fields they produce it is now necessary to allow that X can be a function of the optical field. This is accomplished by writing the field-dependent susceptibility as a Taylor series expansion in powers of the optical field and thus expressing the polarization as

$$\mathbf{P} = \mathbf{X_1} * \mathbf{E} + \mathbf{X_2} * \mathbf{E} * \mathbf{E} + \mathbf{X_3} * \mathbf{E} * \mathbf{E} * \mathbf{E} + \cdots \tag{26}$$

Maxwell considered this form of relation in his classical treatise on electricity and magnetism but, for simplicity, retained only the first-order term.

The first term on the right of Eq. (26) gives rise to the "linear" optics discussed previously. The second term gives rise to optical second-harmonic generation or frequency doubling and optical rectification. The third term results in third-harmonic generation and self-focusing.

Absorption processes involving one or more photons are also described by the susceptibilities X_n employed in Eq. (26). By considering the continuity equation for the flow of energy and the fact that the susceptibilities are complex quantities.

$$X_n = X_n' + jX_n'',$$

the imaginary part of the third-order term can be shown to give rise to an altered form of Beer's law. That is,

$$I(z) = \frac{I(0)\exp(-\alpha_0 z)}{1 + (\alpha_{TPA}/\alpha_0)I(0)[1 - \exp(-\alpha_0 z)]} \qquad (27)$$

where $\alpha_0 = 4\pi k X_1$ is the conventional linear attenuation coefficient, k the propagation vector in the medium, and α_{TPA} the nonlinear attenuation coefficient, given by

$$\alpha_{TPA} = \frac{32\pi^2}{c^2}\omega X_3'' I \qquad (28)$$

This is the part of the absorption that is linearly dependent on the intensity. From a quantum mechanical point of view, such a property is the result of processes in which two photons are absorbed simultaneously. In other words, a material with energy levels separated by U must be considered able to absorb simultaneously two photons, each having energy $U/2$. As a result, when studying the absorption of light in materials it is no longer sufficient to measure it at one low intensity. Contributions from such higher order processes as TPA can occur and must be measured. For example, TPA is sufficient in some semiconductors to enable their excitation to lasing inversions. It is also a powerful tool for studying the presence of deep-level dopants in semiconductors, since they contribute energy levels that enhance TPA.

Higher-order multiphoton absorption processes are possible. There is some evidence that three-photon processes have been detected. However, nonlinear processes higher than TPA require optical fields that are very high. These fields may result in TPA, and the electrons that are thereby freed may be accelerated by the optical field to form a catastrophic electron avalanche breakdown. When this happens the material's effective absorptivity becomes 100%, too much energy is deposited in the irradiated volume, and severe mechanical damage known as intrinsic laser-induced damage follows. This sequence of events is responsible for setting intensity limits for materials and, consequently, for fixing the minimum size of optical components used in high-power lasing systems.

C. Laser-Driven Plasma Coupling and Decoupling

One of the most spectacular features of laser–materials interactions is the formation of a laser-driven plasma in the irradiated region (see Fig. 2). This phenomenon occurs in almost all types of intense laser irradiation, continuous wave (cw) or pulsed, on surfaces, and in the bulk of solids, liquids, and gases. It can prevent the laser light from further coupling to the sample or it can enhance the coupling. Its most important applications are in driving the target compression used to produce fusion energy by inertial confinement and in creating point sources of soft X-rays for X-ray lithography in microelectronics.

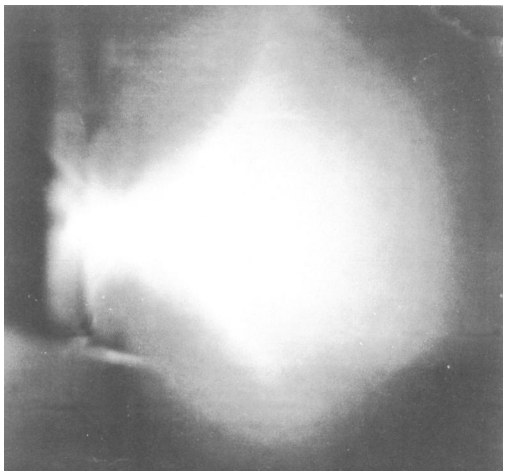


FIGURE 2 Plasma generated by a pulsed CO_2 transverse electrical discharge at atmospheric pressure (TEA) laser incident from the right on a copper target in vacuum. The outer region contains hot copper atoms. The central region is a very bright blue-white color, indicating highly ionized, very hot copper vapor.

When an intense laser beam is focused inside a transparent medium a laser-induced break-down and plasma can be formed in as short a time as 6 psec. This very fast formation is the result of the very highly nonlinear process of plasma formation. The process is so complex that it remains a major subject of current research. As a result, only a qualitative discussion is presented in this article.

If the target reaches a high enough temperature to induce mass loss, it is essential to account for the energy carried away by the removed material in order to determine the energy absorbed. Mass can be removed by such processes as vaporization, melt removal, and pyrolysis or burning. When vaporization occurs some of the ejected material can be ionized. Then, through the process of inverse bremsstrahlung, the vapor can be broken down by the optical field to produce an optically thick plasma. This plasma is a very hot blackbody radiator and emits an intense blue-white light called a laser-induced spark.

As most irradiations take place in air, we consider that case. At laser intensities slightly greater than the plasma formation threshold intensity, a laser-supported combustion (LSC) wave can be ignited. These waves occur at

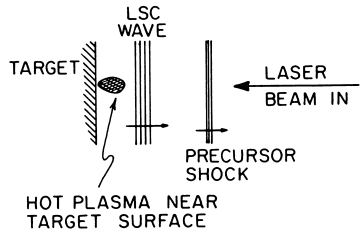


FIGURE 3 Sketch of a one-dimensional laser-supported combustion (LSC) wave. The hot plasma near the target can radiate ultraviolet light, which is strongly absorbed by the target material. The LSC wave is of low enough density to allow the laser light to reach and heat the plasma.

intensities from 2×10^4 to 10^7 W/cm^2 with both pulsed and cw lasers. The ignition of an LSC wave takes place in the ejected target vapor and the heated vapor transfers energy to the surrounding air. The LSC wave thus generated propagates away from the target surface along the beam path and drives a shock wave ahead of itself in the air (see Fig. 3).

The nature of the coupling of the light to the material when a plasma is formed depends on the beam parameters of energy, spot size, and pulse duration as well as air pressure and target material. For long pulse times and low laser intensities the formation of an LSC wave usually results in decreased coupling as the wave propagates away from the surface. This is the case in most laser materials processing applications requiring melting or material removal (i.e., welding, drilling, and cutting).

At high intensities a laser-supported detonation (LSD) wave is ignited. In this case absorption takes place in a thin zone of hot, high-pressure air behind the detonation wave (see Fig. 4). Since this wave takes air away from the surface, expansion fans form to satisfy the boundary conditions at the target surface. The plasma remains nearly one-dimensional until the expansion fans from the edge reach the center. This time is given approximately by the beam radius divided by the speed of sound in the plasma. In the vicinity of the surface there is no laser absorption

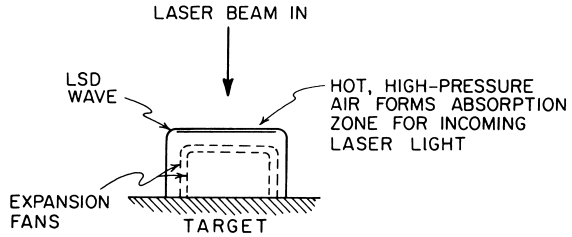


FIGURE 4 Sketch of a laser-supported detonation (LSD) wave. The incoming laser light is absorbed in the LSD wave front as is propagates away from the target. This spreads any reradiated energy in the ultraviolet over a large area and reduces the local energy deposition.

and the plasma properties are determined by its isentropic expansion. The LSD wave plasma expansion away from the surface is very rapid. As a result, though it intercepts and absorbs incoming laser energy, the LSD wave plasma does not reradiate this energy as strongly absorbed ultraviolet light into the area originally irradiated.

Computations for two-dimensional LSD wave plasmas show that for short laser pulses (i.e., shorter than 1 μsec) the overall coupling of light to metal targets can be as large as 25%. Since most metals have absorptions near 1–2% in the infrared, this represents a substantial enhancement of the coupling. However, the coupling remains the same for increasing intensities but is spread out over increasing areas compared to that of the irradiated spot. Therefore, for short pulses, while the total coupling coefficient may be large, the energy deposited in the irradiated area may be smaller than if the LSD wave plasma had never been formed.

When an LSC wave plasma is formed, it is possible to find enhanced coupling in the irradiated region. This results from energy transfer by radiation from the hot, high-pressure plasma adjacent to the surface. The LSC wave propagates into the air at a low speed, and a large fraction of the incoming laser energy is used to heat the plasma to a temperature in excess of 20,000 K. This plasma radiates very efficiently in the ultraviolet part of the spectrum, where most materials, in particular, metals, absorb more strongly than in the visible and infrared. As long as the LSC wave expansion remains nearly one-dimensional the plasma-radiated ultraviolet light remains localized in the originally irradiated area and the local heating is enhanced.

Efficient local coupling of incident energy to the sample requires that the plasma be ignited near the surface, an LSC wave be generated, and the plasma expansion remain one-dimensional during the irradiation. These requirements define a range of intensities over which enhanced coupling will be observed. The minimum intensity is clearly that which ignites a plasma at the surface in the vaporous ejecta. The maximum intensity is that which generates an LSD wave instead of an LSC wave. Pulse durations or dwell times of cw sources must be less than that required for radial expansion of the plasma to set in over the dimension of the irradiated spot. Thus, enhanced coupling will occur for pulses shorter than the spot radius divided by the speed of sound in the plasma ($\sim 5 \times 10^5$ cm/sec). Once radial expansion begins, the pressure and the plasma temperature drop rapidly. Concomitantly, so does the efficiency of the plasma as an ultraviolet light radiator.

As mentioned above, the irradiation conditions used in many laser materials processing applications correspond to the two-dimensional LSC wave plasma case. Therefore, plasma formation reduces the coupling of laser energy into

the irradiated region, and further increases in the laser intensity, either by increasing the laser power or decreasing the spot size, are futile. To deal with this problem. He cover gas is often used to suppress plasma formation. In some cases strong gas jets are used to blow the plasma away and allow the laser light to enter the irradiated material.

Another aspect of laser-driven vaporization and plasma formation that should be considered is the momentum transferred to the surface. In leaving the surface at high speed the ejected material carries away a substantial amount of momentum. This appears as a recoil momentum of the surface and, since it occurs during the short duration of the laser irradiation, it results in a substantial impulse to the surface. This process has been used to shock-harden certain materials. An ablative coating is placed on the area to be shock-hardened and it is struck by a pulsed laser beam. The impulse due to the removal of the ablative coating produces the desired hardening. The transfer of an impulse by laser irradiation is another subject of major research interest, in part due to its potential value as a laser weapons effect.

1. Analysis of Laser–Plasma Interaction

A vapor–plasma plume consisting of ions, electrons, and vapor particles is generally formed above the workpiece during various types of laser materials processing. The laser beam propagates through this plume before reaching the substrate surface. It loses a considerable amount of its energy in the plume above a certain critical power density of the incident beam for a given process. Such partitioning of laser energy between the plume and substrate is important because it determines the amount of laser energy that is actually utilized for materials processing.

Several physical phenomena need to be considered to analyze the partitioning of laser energy during materials processing. These include laser heating, melting, and vaporization, conservation of mass, momentum and energy, absorption and scattering of the laser radiation in the plume, gas dynamics, plasma dynamics, and equation of state. The following analysis provides a simplified approach to understand this complex phenomenon of laser energy partitioning under quasi-steady-state conditions for a pulsed laser beam, which is stationary relative to the substrate.

Mass balance:

$$\pi \mathrm{r}^2 \mathrm{h_p} \rho_\mathrm{v} - \pi \mathrm{r_0^2} \rho_\mathrm{l} \mathrm{S_v} \tau = 0,$$

where r, $\mathrm{h_p}$, and ρ_v are the radius, height, and density of the vapor-plasma plume, respectively. $\mathrm{r_o}$ is the radius of the incoming laser beam, ρ_1 is the density of workpiece at melting temperature, $\mathrm{S_v}$ is the liquid-vapor interface velocity, and t is the pulse duration. The diameter of the plume is considered to be linearly proportional to its height, that is, $D = \mathrm{sh_p}$, where s is the slenderness ratio. In this equation, the amount of material contained in the plume is equated to the amount of material vaporizing at the liquid–vapor interface.

The energy balance for the plume temperature is given by the following equation.

Plume energy balance:

$$\mathrm{P}(1 - \mathrm{e}^{-\mu \mathrm{h_p}}) - 2\pi \mathrm{r h_p h_{eff}}(\mathrm{T} - \mathrm{T}_\infty) = 0,$$

where P, m, $\mathrm{h_{eff}}$, T, and T_∞ are the laser power, absorption coefficient of the plume for the incident laser, effective heat transfer coefficient, absolute temperatures of the plume, and surrounding, respectively. The first term on the left-hand side of this equation represents the energy absorption by the plume. This term is obtained from the Beer–Lambert law by considering the plume as the attenuating medium. The second term corresponds to the energy loss from the plasma to the surrounding due to the radiation and convection effects that are combined together by defining an effective heat transfer coefficient $\mathrm{h_{eff}}$ as follows:

$$\mathrm{h_{eff}} = \mathrm{h_c} + \varepsilon \sigma \left(\mathrm{T_v^2} + \mathrm{T_\infty^2}\right)(\mathrm{T_v} + \mathrm{T}_\infty),$$

where $\mathrm{h_c}$ is the convection heat transfer coefficient, ε is the emissivity of the plume, s is the Stefan-Boltzmann constant, and $\mathrm{T_v}$ is the boiling temperature of the workpiece.

Equation of state for the plume:

$$pV = RT_s,$$

where p is the pressure exerted on the melt surface by the plume, R is the universal gas constant, T_s is the melt surface temperature, and V is the plume volume per unit mole.

Clasius–Clapeyron equation for vaporization:

$$\ln(p/p_c) = 4.693(1 - T_c/T_s),$$

where P_c and T_c are the critical pressure and temperature, respectively, which are given as $T_c = 9340$ K and $P_c = 10291.8$ atm for iron.

Plasma ionization (Saha-Eggert equation):

$$\frac{n_\mathrm{I} n_e}{n_{neu}} = \frac{2\mathrm{U_1}(2\pi m_e k \mathrm{T})^{3/2} \exp[-\mathrm{I/KT}]}{\mathrm{U_0} h^3},$$

where n_I, n_e, and n_{neu} are the ion, electron, and neutral densities, respectively. $\mathrm{U_1}$ and $\mathrm{U_0}$ are the statistical weights of the singly ionized and neutral states respectively. E is the first ionization potential, h is the Planck constant, k is the Boltzmann constant, and m_e is the mass of an electron. At high vapor temperatures, there is sufficient number of charged particles in the plume and

the electron-ion inverse bremsstrahlung process becomes dominant. The absorption coefficient for this process is given by the Bremsstrahlung absorption coefficient:

$$\eta = \frac{n_i n_e Z^2 e^6 \ln[2.25\mathrm{KT}/\mathrm{h}\upsilon]}{24\pi^3 \varepsilon_0^3 \mathrm{c}^3 \mathrm{m_e nKT}\sqrt{2\pi m_e \mathrm{KT}}}\lambda^2,$$

where Z is the charge number, e is the electronic charge, c is the velocity of light in vacuum, and e_o is the permittivity of free space. n and 1 are the frequency and wavelength of the laser light.

Stefan condition for vaporization:

$$AI + k_l\left[\frac{T_s - T_m}{(S_m - S_v)\tau}\right] = \rho_l S_v L_v$$

Stefan condition for melting:

$$\mathrm{k_l}\left(\frac{\mathrm{T_s - T_m}}{(\mathrm{S_m - S_v})\tau}\right) - \mathrm{k_s}\left(\frac{\mathrm{T_m - T_\infty}}{2\sqrt{\alpha\tau}}\right) = \rho_1 \mathrm{L_m S_m},$$

where A is the absorptivity of the melt surface, and I is the laser irradiance at the melt surface. k_l and ρ_l are the thermal conductivity and density of the substrate at the melting temperature, respectively. L_v and L_m are the latent heats of boiling and melting, respectively. a is the thermal diffusivity, k_s is the thermal conductivity, and S_m is the solid–liquid interface velocity. The irradiance (I) reaching the substrate surface and the substrate surface temperature are estimated by the following expressions respectively.

Beer–Lambert law:

$$\mathrm{I} = \mathrm{I_0 e^{-\mu h_p}}.$$

Substrate surface temperature:

$$\mathrm{T_s} = \mathrm{AI}\sqrt{\frac{4\tau}{\pi\rho_1 \mathrm{c_{eff}} \mathrm{k_l}}},$$

where $\mathrm{I_0}$ is the irradiance of the incident laser beam at the top surface of the plume, $\mathrm{c_{eff}}$ is the effective heat capacity that accounts for the latent heat of melting and specific heat capacity of the workpiece $\mathrm{c_{ps}}$, which is taken as $\mathrm{c_{eff}} = \mathrm{c_{ps}} + \mathrm{L_m/T_m}$. The vapor–liquid interface velocity attains a constant value at the quasi-steady state depending on the plasma volume and temperature. The mass of material vaporized per unit area per unit time (J) is given by Vapor flux:

$$J = S_v \rho_l.$$

These equations describe various aspects of the process physics and represented a set of coupled nonlinear algebraic equations that need to be solved numerically. Typical results for the partitioning of laser energy is presented in Figure 5 for stainless steel (SS 316) substrate irradiated with a Nd:YAG laser of pulse duration 160 ns. Pinhole experiments were conducted to obtain the experimental results of this figure by actually measuring the laser energy that reaches the substrate surface after the beam propagates through the plume. The theory overestimates the results compared to the experimental results because the theory is based on the Bremsstrahlung absorption of the laser radiation, whereas the beam may also be attenuated due to the scattering of the beam in the plume.

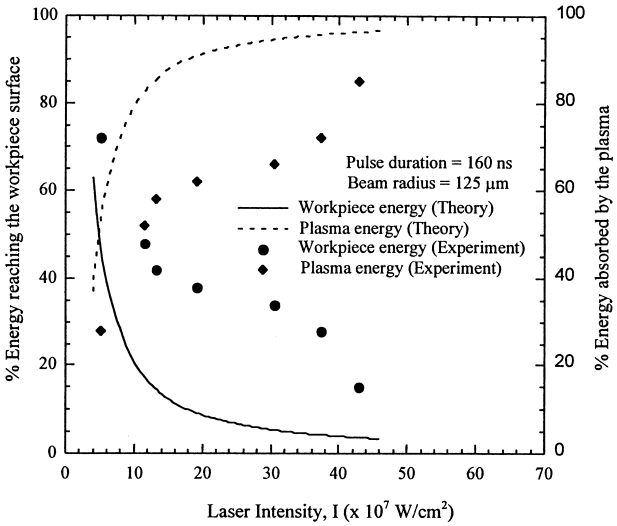


FIGURE 5 Partitioning of laser energy between the plasma and workpiece.

D. Surface Electromagnetic Waves

Researchers studying such diverse topics as laser processing of semiconductors, laser-induced damage, laser materials processing, and laser-driven deposition processes have observed the formation of "ripples" in the irradiated region (see Fig. 6). Such structures on the surface of metals, dielectrics, and semiconductors can be divided into two categories: (1) resonant periodic structures (RPS) and (2) nonresonant periodic structures (NRPS). RPS are formed due to the interaction of the incident laser beam with the surface electromagnetic waves (electromagnetic waves that propagate across the surface), and their periods are influenced by the wavelength, polarization, and incidence angle of the beam. The RPS formation theory is generally based on (1) electrodynamic analysis, in which the modulation of the energy absorbed at the substrate surface is studied by considering the interference between the incident and surface electromagnetic waves, and (2) thermophysical analysis, in which the nonuniform dissipation of energy at the substrate surface is studied by considering microscopic changes in the crystal structure during laser heating. The formation of NRPS is generally attributed to the melting of metal surface in pulsed laser heating. It involves nonuniform vaporization of the metal, and capillary oscillations of the molten surface layer due to instabilities at the plasma–substrate interface.

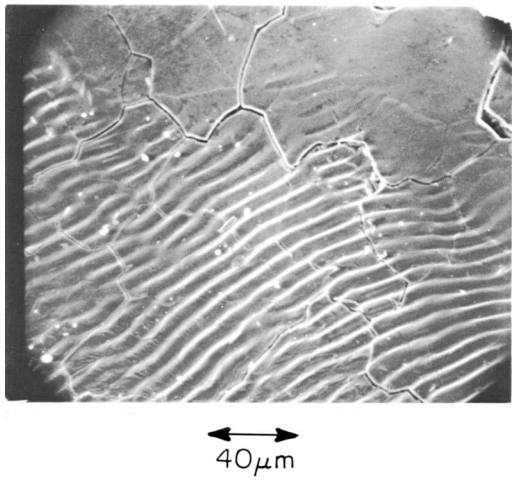


FIGURE 6 Scanning electron micrograph of ripples spaced by ~10 μm produced in 304 stainless steel exposed to a cw CO_2 laser able to melt metal. The sample was stationary under a normally incident 10.6-μm beam polarized perpendicular to the ripples. The residual ripples are found near the edges of the melt pool, where freezing is sufficiently rapid to preserve the ripples. In the central part of the melt pool convective motion in the melt obscures the ripples. The ripples are the result of surface electromagnetic waves (SEW) launched by scattering the incident light and subsequent interference between two electromagnetic waves.

All of these observations had in common the facts that linearly polarized lasers were used and that the material in the irradiated region had reached its melting point. The ripples were spaced by the light wavelength when the light was at normal incidence. These facts and additional data demonstrated that the ripples were the result of the following sequence of events:

1. The incident light was scattered by some surface irregularity and a surface electromagnetic wave was launched.
2. In the region irradiated this wave was able to interfere with the incident light.
3. The interference pattern produced localized differences in the surface temperature, and when the material became molten it flowed due to surface tension effects or differential vapor pressures to conform with the varying light intensity.
4. The molten material froze quickly enough following irradiation that the ripples were detectable afterward.

The key element in the formation of these ripples is the fact that surface electromagnetic waves (SEW) were generated. This is another means by which light energy can be coupled into a material and one that was not obvious from prelaser knowledge. The fact that a gratinglike interference pattern is set up allows one to apply the electromagnetic theory of gratings to understand the process. From this theory, it is clear that for a particular line spacing, there is a particular grating height that can dramatically enhance the absorption of a material over that of a smooth surface. The generation of SEW can therefore result in greatly enhanced coupling of laser light to materials.

On the other hand, if the surface waves propagate, they may carry energy away from the irradiated area. If the dimension of the irradiated area is small compared to the SEW attenuation length, this may be a major factor in redistributing the absorbed energy. The irradiated region will not become as hot as it might if no SEW were generated. It should be noted that this effect would give rise to a beam spot size dependence of laser heating when the beam diameters used are in the range of one SEW attenuation length.

E. Photoablation

A new and exciting light–matter interaction called photoablation has been demonstrated with the advent of high peak intensity ultraviolet lasers. This occurs with the excimer lasers (i.e., ArF, KrF, XeF, or XeCl) operating at wavelengths from ~250 to 350 nm. The lasers provide high energies (~1 J or more per pulse) in pulses of ~30 nsec duration. Most organic materials absorb very strongly at these wavelengths. The energies of photons in the ultraviolet are sufficient to efficiently break bonds in such materials. When this happens the fragments are subjected to very high pressures and are ejected from the surface (see Fig. 7). The process takes place entirely within one absorption depth (~20 nm) and so provides a very controllable means of material removal. In addition, the

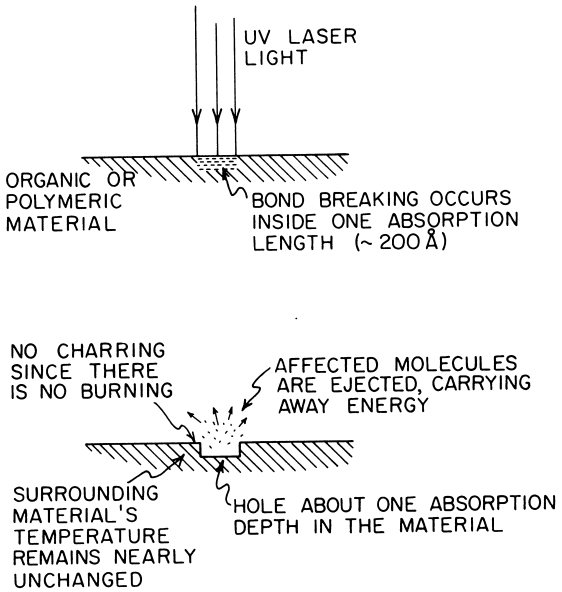


FIGURE 7 Sketch of the process of ultraviolet laser irradiation-induced photoablation.

energy deposited in the material is removed as kinetic energy of the ejecta. As a result, there is very little thermal heating of the sample and the process is confined to the area irradiated. With lasers operating at reasonably high repetition rates, this process can be used to remove significant amounts of material.

The use of excimer lasers and photoablation is still under consideration in research and development laboratories. However, it is clearly suitable for removing protective coatings or insulation and for certain medical applications where great delicacy is required.

II. LASER HEATING OF MATERIALS

A. Thermal Diffusion Problem and Thermal Diffusivity

We are concerned with the heating of materials that results from several of the laser light absorption processes discussed in the previous sections. This can be approached in a number of nearly equivalent ways, all of which require a knowledge of (1) the target's optical and thermal properties during the irradiation, (2) the laser beam distribution, (3) the dynamics of the irradiation process, and (4) the processes of phase change in the target material. As would be expected, it is extremely difficult to solve the heat flow problem exactly in the general case and so reasonable approximations are used. In addition, only problems that are easy to solve are attempted. These are then useful as guides to the solution of other problems.

Consider the equation of heat conduction in a solid with the laser energy absorbed on the irradiated surface as a heat source. By judiciously selecting the beam geometry, dwell time, and sample configuration, the problem may be reduced to solvable one- and two-dimensional heat flow analyses. Phase transitions can be included and the temperature distributions that are produced can be calculated. Examples will be selected to provide specific guidance in the choice of lasers and materials. The result of all this will be an idea of the effects that one may produce by laser heating of solids.

The equation for heat flow in a three-dimensional solid is

$$\rho C\frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(K\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(K\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(K\frac{\partial T}{\partial z}\right) + A(x, y, z, t), \qquad (29)$$

where ρ is the material density in grams per cubic centimeter, K the thermal conductivity in watts per centimeter per degree Celsius, and C the heat capacity in joules per centimeter per degree Celsius; these are material properties that will depend on temperature and position. The quantity $A(x, y, z, t)$ is the rate at which heat is supplied to the solid per unit time per unit volume in joules per second per cubic centimeter and $T = T(x, y, z, t)$ is the resulting temperature distribution in the material. Figure 8 defines the coordinate system used.

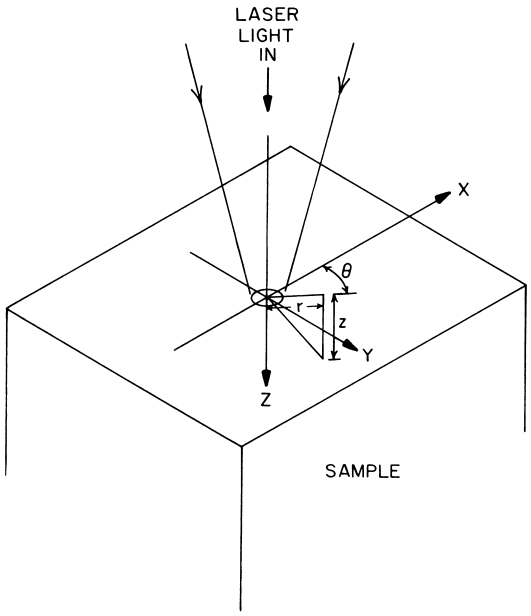


FIGURE 8 Sketch of coordinate system used in evaluating responses of materials to laser irradiation.

The temperature dependence of the properties results in a nonlinear equation that is very difficult to solve exactly. Where the functional dependence of these quantities on temperature is known, it is sometimes possible to use numerical integration techniques to obtain a solution. A further complication arises from the temperature dependence of $A(x, y, z, t)$ through that of the material's absorptivity. When phase transitions occur one can attempt a solution of the problem by solving for each phase separately and including the heat required for the transition where appropriate.

For most materials the thermal properties do not vary greatly with temperature and can be assigned an average value for the temperature range to be studied. In this case it is possible to solve the heat flow problem. A further simplification is obtained by assuming that the material is homogeneous and isotropic. Under these conditions Eq. (29) reduces to

$$\nabla^2 T - \frac{1}{\kappa}\frac{\partial T}{\partial t} = -\frac{A(x, y, z, t)}{K}, \qquad (30)$$

where $\kappa = K/\rho C$ is the thermal diffusivity. In the steady state $\partial T/\partial t = 0$, resulting in

$$\nabla^2 T = -\frac{A(x, y, z, t)}{K}. \qquad (31)$$

If there is no heat source, as in the case of cooling of heated material, the temperature distribution will be given by

$$\nabla^2 T = \frac{1}{\kappa}\frac{\partial T}{\partial t} \tag{32}$$

in the time-dependent case and

$$\nabla^2 T = 0 \tag{33}$$

in the steady state.

A physical insight into the meaning of the quantity κ is useful. A simple analysis of the units show that

$$(\kappa t)^{1/2} = \text{distance}.$$

This distance or a multiple thereof is generally known as the thermal diffusion distance for the particular problem and is a handy quantity to use when scaling the effects of laser heating. It is often useful to know how the optical absorption depth compares with the thermal diffusion distance during a laser irradiation. To see the meaning of the thermal diffusion depth more clearly, consider the following particular solutions of Eq. (32):

$$T = T_0 e^{t/\mathrm{T} \pm z/z_\mathrm{D}}$$

and so

$$\frac{\partial T}{\partial t} = \frac{1}{\mathrm{T}} T$$

and

$$\frac{\partial^2 T}{\partial z^2} = \left(\frac{1}{z_\mathrm{D}}\right)^2 T$$

Thus

$$\left[\left(\frac{1}{z_\mathrm{D}}\right)^2 - \frac{1}{\kappa \mathrm{T}}\right] T = 0,$$

and we have a solution if

$$z_\mathrm{D} = (\kappa \mathrm{T})^{1/2}.$$

Therefore a characteristic distance is related to a characteristic time by $(\kappa \mathrm{T})^{1/2}$.

For a more realistic case consider

$$T = T_0 t^{-1/2} e^{-z^2/4\kappa t} \quad \text{for } t > 0$$

Then we have

$$\frac{\partial T}{\partial t} = T_0 \left(-\frac{1}{2t^{3/2}} + \frac{z^2}{4\kappa t^{5/2}}\right) e^{-z^2/4\kappa t}$$

and

$$\frac{\partial^2 T}{\partial z^2} = T_0 \left(-\frac{1}{2\kappa t^{3/2}} + \frac{z^2}{4\kappa t^{5/2}}\right) e^{-z^2/4\kappa t}$$

which satisfy

$$\frac{\partial^2 T}{\partial z^2} - \frac{1}{\kappa}\frac{\partial T}{\partial z} = 0.$$

At $t = 0$ we will set $T = T_0$ for $z = 0$ and $T = 0$ for $z > 0$. Now consider that at any time t_p the temperature at

$$z_\mathrm{D} = 2(\kappa t_\mathrm{p})^{1/2}$$

is $1/e$ times that at $z = 0$.

The second case corresponds to the release of the quantity of heat $2\rho C T_0 (\pi\kappa)^{1/2}$ per unit area over the plane $z = 0$ at time $t = 0$. This could be a decent approximation to the case of laser heating of a semi-infinite metal irradiated by a short pulse [defined in terms of the relative values of beam radius and $(\kappa t_\mathrm{p})^{1/2}$] in a uniform beam where we wish to know a temperature on the beam axis. The quantity of heat per unit area would be the laser intensity I times the pulse duration $\mathrm{T_p}$ times the fraction absorbed or

$$I \mathrm{T_p} \alpha = 2\rho C T_0 (\pi\kappa)^{1/2}$$

where α is the absorptance of the metal. Then

$$T_0 = \frac{\alpha I \mathrm{T_p}}{2\rho C (\pi\kappa)^{1/2}}$$

and

$$T(z,t) = \frac{\alpha I \mathrm{T_p}}{2\rho C (\pi\kappa)^{1/2}} t^{-1/2} e^{-z^2/4\kappa t}$$

or

$$T(z,t) = \frac{\alpha I \mathrm{T_p}}{\rho C \pi^{1/2}} \frac{e^{-z^2/4\kappa t}}{(4\kappa t)^{1/2}} \quad \text{for } t > 0.$$

The important role of the quantity $z_\mathrm{D} = 2(\kappa t)^{1/2}$ in describing the process of laser heating is clear; z_D defined above is the thermal diffusion distance.

B. Absorption in a Very Thin Surface Layer

These are cases where

$$\alpha^{-1} \ll (\kappa \mathrm{T_{laser}})^{1/2}$$

where $\mathrm{T_{laser}}$ is the duration of the irradiation.

1. Uniformly Illuminated Surface

The incident intensity is given by

$$I_0(t) = \begin{cases} 0, & t < 0 \\ I_0, & t \geq 0 \end{cases}$$

and is assumed to be uniformly distributed. The temperature distribution on-axis is sought by assuming no heat diffusion in the x and y directions, even though this may be important in certain applications. The results of this special case are very useful since most often the on-axis temperature is the desired quantity.

The thermal diffusion equation with this heat source and the boundary condition

$$T(z,t) = 0 \quad \text{for } t < 0 \text{ and all } z$$

has the solution

$$T(z,t) = \left(\frac{2\alpha I_0}{K}\right)(\kappa t)^{1/2}\text{ierfc}\left[\frac{z}{2(\kappa t)^{1/2}}\right], \qquad (34)$$

where

$$\text{ierfc}(X) = \int_X^\infty \text{erfc}(X')\,dX'$$

and

$$\text{erfc}(X) = 1 - \text{erf}(X) = \frac{2}{\pi}\int_X^\infty e^{-(X')^2}\,dX'$$

and

$$\text{erf}(X) = \frac{2}{\pi}\int_0^X e^{-(X')^2}dX'.$$

The error function erf(X) has the following properties:

$$\text{erf}(0) = 0 \quad \text{erf}(\infty) = 1 \quad \text{erf}(-X) = -\text{erf}(X)$$
$$\text{erfc}(0) = 1 \quad \text{erfc}(\infty) = 0$$

and

$$\text{ierfc}(0) = 1/\sqrt{\pi}.$$

Thus the surface temperature is given by

$$T(0,t) = \frac{2\alpha I_0}{K}(\kappa t)^{1/2}\left(\frac{1}{\pi}\right)^{1/2}. \qquad (35)$$

It is proportional to $t^{1/2}$, and if there were no phase changes (i.e., melting and vaporization) the temperature would continue to increase. Note that the energy flux absorbed by the surface is

$$E = \int_0^t \alpha I\,dt = I_0 t\alpha_0,$$

but because of conduction the surface temperature increases more slowly, that is, as $t^{1/2}$. This means that a greater surface temperature can be achieved for a given laser pulse energy by shortening the pulse and increasing I_0.

Since we have mentioned pulses, let us model a simple pulse as one where

$$I = \begin{cases} 0, & t < 0 \\ I_0, & 0 \le t \le \text{T} \\ 0, & T < t. \end{cases}$$

During the interval 0 to T the temperature is as given above with the maximum temperature obtained at $t = \text{T}$. For $t > \text{T}$,

$$T(z,t) = \frac{2\alpha I_0}{K}\left\{(\kappa t)^{1/2}\,\text{ierfc}\left[\frac{z}{2(\kappa t)^{1/2}}\right] - [\kappa(t-\text{T})]^{1/2}\,\text{ierfc}\left[\frac{z}{2\kappa(t-\text{T})}\right]^{1/2}\right\}. \qquad (36)$$

Remembering that this problem is only a one-dimensional approximation to the actual case and that the assumption of an infinitely thick medium requires that the slab thickness L be

$$L > 2(\kappa\text{T})^{1/2}$$

allows one to use these results to calculate the temperature distribution achieved. As pointed out above, the result will only be correct near the beam axis. The cooling rate following the irradiation in this example is obtained by taking the time derivative of $T(z,t)$ in Eq. (36).

2. Uniformly Illuminated Circle of Radius A

The power is

$$P = \begin{cases} 0, & t < 0 \\ P_0, & t \ge 0. \end{cases}$$

Using cylindrical coordinates as shown in Fig. 7, one finds

$$T(r,z,t) = \frac{\alpha P_0}{2\pi AK}\int_0^\infty \frac{d\lambda}{\lambda}J_0(\lambda r)J_1(\lambda A) \times \left\{e^{-\lambda z}\,\text{erfc}\left[\frac{z}{2(\kappa t)^{1/2}} - \lambda(\kappa t)^{1/2}\right] - e^{\lambda z}\,\text{erfc}\left[\frac{z}{2(\kappa t)^{1/2}} + \lambda(\kappa t)^{1/2}\right]\right\}, \qquad (37)$$

where J_0 and J_1 are Bessell functions of the first kind. This expression may be evaluated numerically. It is far more useful in planning experiments to evaluate the result directly under the beam since this will be the hottest place. In other words, set $r = 0$ and $z = 0$, to find

$$T(0,0,t) = \frac{2\alpha P_0(\kappa t)^{1/2}}{\pi A^2 K}\left\{\frac{1}{\pi^{1/2}} - \text{ierfc}\left[\frac{A}{2(\kappa t)^{1/2}}\right]\right\}. \qquad (38)$$

Note that if $A \gg 2(\kappa t)^{1/2}$, then at those values of t one can treat the problem as the simple one in Section II,B,1. In other words, while the beam radius is greater than the thermal diffusion distance, the problem can be reduced to the simple case for the on-axis temperature. However, as time marches on this will break down and there will be differences. Most obviously, T does not become infinite but instead as $t \to \infty$

$$T(0,0,\infty) = \alpha P_0/\pi AK = T_{\max}. \qquad (39)$$

This reveals a curious point—the quantity P_0/A determines the maximum achievable temperature. If, for example, we wish to heat the material without melting the surface, then we must choose P_0/A such that $T_{\max} < T_{\text{melt}}$.

3. Uniformly Illuminated Rectangle

It is often the case that the focused laser beam is not circular. In fact, most excimer laser beams are rectangular. Assuming that the light distribution is uniform, we can examine the rectangular distribution. The solution on-axis and on the surface is

$$T(0,0,\infty) = [\alpha P/2K\pi la]\,[a\sinh^{-1}(l/a) + \sinh^{-1}(a/l)] \tag{40}$$

at late times. Here a is the width and l the length of the irradiated rectangle.

For example, if the width of the rectangle a and the radius in the circular case A were set equal and we asked what power would be needed to reach a specific temperature in each case we would have

$$\frac{P_{\text{circ}}}{P_{\text{rect}}} = \frac{1}{2l}\left[a\sinh^{-1}\left(\frac{l}{a}\right) + l\sinh^{-1}\left(\frac{a}{l}\right)\right]. \tag{41}$$

Even a square beam ($l = a$) requires more power to produce a specific temperature than a circular beam. This is obvious since a larger area is irradiated and the absorbed energy must heat up more material. To achieve the same maximum temperature, more input must be provided. If $l/a = 3$, then we see from Eq. (41) that the power must be increased by a factor of $\sim$2. This simple example demonstrates the importance of good beam quality and focusing optics in order to obtain maximum heating from laser irradiation.

4. Gaussian Beam Illumination

The Gaussian beam intensity distribution is

$$I = I_0 e^{-r^2/w^2}. \tag{42}$$

[Note: It is common to describe a Gaussian mode of a laser by the electric field distribution

$$E = E_0 e^{-r^2/w_0^2}$$

and then give

$$I = I_0 e^{-2r^2/w_0^2} = I_0 e^{-r^2/(w_0/\sqrt{2})^2}.$$

The beam parameter w used in this treatment is the radius at which the *intensity* has fallen to $1/e$ of its on-axis value. It is the radius at which the electric field has fallen to $e^{-1/2}$ of *its* on-axis value. If we used w_0 to describe the Gaussian, then it corresponds to the radius at which the intensity has fallen to e^{-2} of its on-axis value.]

The following expression can be obtained for the temperature distribution in a semi-infinite solid due to irradiation at the surface by an instantaneous ring source of radius r' and total deposited energy Q:

$$T_{\substack{\text{inst.}\\ \text{ring}}}(r,z,t) = \frac{Q}{4\rho C(\pi\kappa t)^{3/2}} \times \exp\left(\frac{-r^2 - r'^2 - z^2}{4\kappa t}\right)\mathcal{I}_0\left(\frac{rr'}{2\kappa t}\right). \tag{43}$$

In this expression $\mathcal{I}_0$ is the modified Bessel function of order zero.

For a Gaussian source we have

$$Q = q_0 e^{-r'^2 w^2} 2\pi r'\, dr',$$

where q_0 is the energy per unit area at the origin. Inserting this into Eq. (43), we integrate to obtain

$$T_{\substack{\text{inst.}\\ \text{Gaussian}}}(r,z,t) = \frac{q_0 w^2}{\rho C(\pi\kappa t)^{1/2}(4\kappa t + w^2)} \times \exp\left(-\frac{z^2}{4\kappa t} - \frac{r^2}{4\kappa t + w^2}\right). \tag{44}$$

If q_0 is replaced by $\alpha I_0 \mathrm{T_p}$, where $\mathrm{T_p}$ is the duration of a pulse, and

$$w \gg (4\kappa \mathrm{T_p})^{1/2},$$

we can use Eq. (44) for the temperature distribution following irradiation by a short pulse. (*Note: Short pulse is defined in terms of whether heat diffuses significantly with respect to the beam dimension during the pulse. If it does not, then the pulse is "short."*)

Rewriting Eq. (44) gives

$$T_{\substack{\text{inst.}\\ \text{Gaussion}}}(r,z,t) = \frac{\alpha I_0 \mathrm{T_p}}{\rho C(\pi\kappa t)^{1/2}}\left(\frac{1}{4\kappa t/w^2 + 1}\right) \times \exp\left[-\frac{(z/w)^2}{4\kappa t/w^2} - \frac{(r/w)^2}{4\kappa t/w^2 + 1}\right]. \tag{45}$$

Defining the unitless quantities

$$T' = w\rho C\pi^{1/2}T/2\alpha I_0 \mathrm{T_p} \qquad t' = 4\kappa t/w^2$$
$$z' = z/w \qquad r' = r/w$$

enables writing

$$T' = \left(\frac{1}{t'}\right)^{1/2}\frac{1}{t'+1}\exp[-z'^2/t' - r'^2/(t'+1)]. \tag{46}$$

For any z' and t', we see that T' decreases with r' as $e^{-r'/t'+1}$. Similarly for any t' and r', T' decreases as $e^{-z'^2/t'}$. Thus, having found the on-axis surface temperature, one can easily find the temperature at any point in the material!

For copper $\kappa \sim 1$ cm^2/sec and when $w = 0.5$ mm we would require

$$\mathrm{T} \ll 6 \times 10^{-4} \text{ sec}.$$

For steel $\kappa \sim 0.15$ cm^2/sec and for $w = 0.5$ mm we would require

$$T_p \ll 4 \times 10^{-3} \text{ sec.}$$

In the case of pulsed Nd: YAG lasers where $T_p \sim 10^{-5}$ sec, Eq. (44) is quite adequate. For CO_2 TEA laser pulses with $T_p \sim 10^{-6}$ sec, Eq. (44) is again acceptable. However, for discharge pulsed CO_2 lasers with $T_p \sim 10^{-3}$ sec or for cw lasers (or for smaller w) the noninstantaneous form must be used.

The case of cw Gaussian illumination can be treated by setting $q_0 = \alpha I_0(\mu)\, d\mu$ (where μ is a dummy variable for time) and integrating $T_{\substack{\text{inst.}\\ \text{Gaussian}}}$ from 0 to t. This gives

$$T_{\substack{\text{noninst.}\\ \text{Gaussian}}}(r, z, t) = \frac{\alpha I_0^{\max} w^2}{K} \frac{\kappa^{1/2}}{\pi} \int_0^t \frac{p(t-\mu)\, d\mu}{(\mu)^{1/2}(4\kappa\mu + w^2)} \times \exp\left(-\frac{z^2}{4\kappa\mu} - \frac{r^2}{4\kappa\mu + w^2}\right),$$

where $I_0(\mu) = I_0^{\max} p(\mu)$. Setting

$$\mu' = 4\kappa\mu/w^2, \quad z' = z/w, \quad r' = r/w$$

gives

$$T_{\substack{\text{noninst.}\\ \text{Gaussian}}}(r', z', t') = \frac{\alpha I^{\max} w}{K 2\pi^{1/2}} \int_0^{t'} \frac{p(t'-\mu')\, d\mu'}{(\mu')^{1/2}(\mu'+1)} \times \exp\left(-\frac{z'^2}{\mu} - \frac{r'^2}{\mu+1}\right)$$

and so for a dimensionless $T' = 2\pi^{1/2} K T/\alpha I_0^{\max} w$ we have

$$T' = \int_0^{t'} \frac{p(t'-\mu')\, d\mu'}{(\mu')^{1/2}(\mu'+1)} \times \exp\left(\frac{z'^2}{\mu} - \frac{r'^2}{\mu+1}\right) \quad (47)$$

At the surface $z' = 0$ and at the center $r' = 0$ and so

$$T' = \int_0^{t'} \frac{p(t'-\mu')\, d\mu'}{(\mu')^{1/2}(\mu'+1)}$$

For a cw laser or one where $T_p \gg w^2/4\kappa$

$$P(t'-\mu') = 1$$

and

$$T' = 2\tan^{-1}(t')^{1/2}$$

or

$$T(0, 0, t) = \frac{\alpha I_0^{\max} w}{K\pi^{1/2}} \tan^{-1}\left(\frac{4\kappa t}{w^2}\right)^{1/2}. \quad (48)$$

5. Non-Gaussian Beam Illumination

The polarization of laser and its mode structure affect the heat distribution during laser materials processing. The focal spot consists of several *TEMmn* modes for most high-power lasers, and therefore, the energy distribution at the focal spot is nonuniform, and localized hot spots are generated at the substrate surface. The laser beam profile is characterized by its transverse electromagnetic mode, which is denoted by *TEMmn*. *m* and *n* are the mode numbers in two perpendicular directions that are orthogonal to the direction of laser beam propagation. The values of *m* and *n* depend on the cavity design. For rectangular cavities, the amplitude of the standing wave pattern is given by the Hermite–Gaussian polynomial, and for cylindrical cavities it is given by Laguere–Gaussian polynomial. The intensity distribution is related to the square of the amplitude. Generally, circular output apertures produce laser beams of cylindrical symmetry whose mode structure is expressed in terms of cylindrically symmetric Laguerre–Gaussian modes. However, the presence of a Brewster window or of any other rectangular asymmetries in the mirror coatings or the laser gain medium or structural alignment of the laser cavity would generate rectangular beams whose mode structure is expressed in terms of rectangularly symmetric Hermite–Gaussian modes. It is common experimental observation that single higher order Hermite–Gaussian modes are more easily obtainable in stable laser resonators than are higher order Laguerre–Gaussian modes, even in resonators with apparently cylindrical symmetry. Both theoretical and experimental studies have confirmed that even very small deviations from cylindrical symmetry breaks the degeneracy of cylindrical Laguerre–Gaussian modes, transforming them into nondegenerate Hermite–Gaussian modes. For this reason, the Hermite–Gaussian polynomial is considered to express the laser irradiance.

The mode number affects the focusability of the laser beam. The higher the number of modes, the more difficult it is to focus the beam since the beam no longer comes from a single spot. With proper design of the resonator, higher order modes can be eliminated. In many applications, such as laser surface hardening and cladding, higher order modes are desirable for uniform heating. TEM00 mode is best suited for laser cutting, drilling, and welding applications.

The absorptivity of the material is affected by the polarization. If polarization is perpendicular to the scanning direction, a shallow but wider cut is obtained. The directional effects in laser cutting can be avoided by using circularly polarized light.

The propagation of heat in the workpiece is encountered in all laser materials processing. The heating of a workpiece with multimode laser beam is different from the heating with a *TEM00* mode beam. This is because the multimode beam produces multiple localized hot spots, whereas the *TEM00* mode beams generate a single central hot spot.

The heat conduction equation for a moving heat source can be written as

$$\kappa\left(\frac{\partial^2 T}{\partial x^2}+\frac{\partial^2 T}{\partial y^2}+\frac{\partial^2 T}{\partial z^2}\right)=-v\frac{\partial T}{\partial x}$$

for $-\infty < x < \infty, -\infty < y < \infty$ *and* $0 \le z < \infty$

under the quasi-steady-state condition. The dimensionless temperature $T(x, y, z)$ is defined as $T = (T_1 - T_0)/(T_m - T_0)$, where T_1, T_0 and T_m are the substrate, ambient, and melting temperatures respectively. This equation is written in a moving coordinate system such that the center of the rectangular laser spot coincides with the origin of the coordinate system. The x, y, and z coordinates are defined such that the x–y plane lies on the top of the workpiece and the positive z-axis is into the workpiece. Another set of axes x_1 and y_1 are so chosen that its origin lies at the beam center and the x_1 and y_1 axes are parallel to the length and width of the rectangular spot respectively. The laser beam is scanned along the x-axis at a constant velocity v. When the beam is rotated by an angle ϕ from the x-axis, the x_1 and y_1 axes are also rotated so that they subtend an angle ϕ with the x and y axes, respectively. This situation is equivalent to keeping the length and width of the laser spot parallel to the x and y-axes, respectively, and scanning the beam in the x_1 direction at a scanning angle $-\phi$.

This governing equation is solved under the following boundary conditions:

$$T \to 0 \; as \; x \to \pm\infty, y \to \pm\infty \text{ and } z \to \infty, \text{ and}$$

$$K\frac{\partial T}{\partial z} = -\frac{1}{T_m - T_0} I(x, y) \quad \text{at } z = 0.$$

The laser irradiance $I(x, y)$ is given by

$$I(x,y)=\sum_{m=0}^{M}\sum_{n=0}^{N}\frac{2AP_{mn}}{2^{m+n}\pi m!n!l_0 w_0}\times H_m^2\left(\frac{\sqrt{2}(x\cos\phi+y\sin\phi)}{l_0}\right)\times H_n^2\left(\frac{\sqrt{2}(y\cos\phi-x\sin\phi)}{w_0}\right)\times\exp\left(-2\frac{(x\cos\phi+y\sin\phi)^2}{l_0^2}\right)\times\exp\left(-2\frac{(y\cos\phi-x\sin\phi)^2}{w_0^2}\right),$$

where M and N are the largest mode numbers in the x_1 and y_1 directions, respectively. A is the absorptivity of the workpiece at the wavelength of the incident laser radiation. P_{mn} is the power ofthe laser of m-th and n-th order modes. l_0 and w_0 are respectively the half-length and half-width of the Gaussian mode of the beam. They are given by $l_0 = l_s\sqrt{2m+1}$ and $w_0 = w_s\sqrt{2m+1}$, where l_s and w_s are respectively the half-length and half-width of the rectangular laser spot. H_m and H_n are the m-th and n-th degree Hermite polynomials. The solution to the above governing equation is given by

$$T(x,y,z) = \frac{1}{2\pi}\int_0^{2\pi}\int_{-\infty}^{\infty} f(x-r'\cos\theta, y-r'\sin\theta)\times\exp(-br'\cos\theta)\frac{\exp(-b\sqrt{r'^2+z^2})}{\sqrt{r'^2+z^2}}r'\,dr'\,d\theta,$$

where $f(x,y)=\dfrac{1}{K(T_m-T_0)}I(x,y)$ and $b=\dfrac{v}{2\alpha}$.

6. Comparison of Results

The several expressions for the induced surface temperature are summarized below.

For the uniformly irradiated surface of a semiinfinite sample (Section II.B.1)

$$T_A(0,0,t)=\frac{2\alpha I_0}{K}\left(\frac{\kappa t}{\pi}\right)^{1/2}.$$

For the uniformly irradiated beam of radius A (Section II.B.2)

$$T_B(0,0,t)=\frac{2\alpha P_0}{\pi A^2 K}(\kappa t)^{1/2}\left[\frac{1}{\pi_{1/2}}-\text{ierfc}\frac{A}{2(\kappa t)^{1/2}}\right]$$

and

$$T_B(0,0,\infty)=\frac{\alpha P_0}{\pi AK}.$$

For the noncircular uniform beam (Section II.B.3)

$$T_C(0,0,\infty)=\frac{\alpha P_0}{2K\pi la}\left[a\sinh^{-1}\left(\frac{l}{a}\right)+l\sinh^{-1}\left(\frac{a}{l}\right)\right]$$

For the Gaussian beam (Section II.B.4) with $I = I_0 e^{-r^2/w^2}$

$$T_D(0,0,t)=\frac{\alpha I_0 w}{K\pi^{1/2}}\tan^{-1}\left(\frac{4\kappa t}{w^2}\right)^{1/2}$$

and

$$T_D(0,0,\infty)=\frac{\alpha I_0 w\pi^{1/2}}{2K}.$$

It is appropriate to ask how much difference there is between these results. For example, if a Gaussian beam were used but treated as if it were a circular beam of radius w with $I = I_0$, what would happen?

By comparing cases II.B.4 and II.B.2 at equilibrium it is clear that

$$\frac{T_D(0,0,\infty)}{T_B(0,0,\infty)} = \frac{\alpha I_0 w \pi^{1/2}/2K}{\alpha P_0/\pi w K} = \frac{\pi^{1/2}}{2} = 0.886$$

The error in approximating the Gaussian by a uniform circle with average intensity equal to the peak on-axis intensity of the Gaussian is only 11%. Considering the approximations involved in setting the thermal and optical properties equal to some average value, this is acceptable. A similar conclusion could be reached when comparing the other on-axis surface temperatures at other times. Thus, an acceptable estimate of the on-axis surface temperature produced by a beam with $I(r=0)=0$ is obtained from the simplest case with a finite beam radius, that is, case II.B.I.

7. Some Numerical Examples

Now consider some numbers to get a feeling for the scale of things and for when to include phase changes. The properties of three interesting materials are listed in Table I.

The other property needed to estimate the temperatures that can be obtained is the material's absorptivity for the laser in question. The carbon phenolic can be assumed with reasonable accuracy to absorb 100% of incident 1.06- and 10.6-μm laser light. The Al and 204 stainless steel absorptivities can be questioned as they will depend on surface finish and the presence of any molten material. Also, the absorptivity of a metal is expected to increase with temperature. Recent measurements for Al and 1016 steel show that at 10.5 μm the Drude model gives accurate values for both the absorptivity and its temperature coefficient. At 1.06 μm the experimental results do not agree with the Drude model. The absorptivity is higher than predicted. Furthermore, whenever even a small amount of any metal is melted ~20% of the incident laser light seems to be absorbed. This may be due to an unexpected increase in absorptivity at the melting temperature or to geometric considerations on melting. However, a fair estimate of the absorptivity of almost any metal exposed to 1.06- and 10.6-μm laser beams is ~10%. (You cannot go too far wrong with this estimate, particularly if a little melting occurs.)

For a Gaussian beam at the surface and on-axis

$$T(0,0,t) = \frac{\alpha I_0 w}{K\pi^{1/2}} \tan^{-1}\left(\frac{4\kappa t}{w^2}\right)^{1/2} = \frac{\alpha I_0}{K\pi^{1/2}} 2(\kappa t)^{1/2}$$

for small t or for

$$w \gg 2(\kappa t)^{1/2} = \frac{\alpha P_0}{K\pi w^2} 2\left(\frac{\kappa t}{\pi}\right)^{1/2}$$

and

$$T(0,0,\infty) = \frac{\alpha I_0 w \pi^{1/2}}{2K} = \frac{\alpha P_0}{2K\pi^{1/2} w}$$

Let us assume a 500-W beam focused to a spot with $w =$ 0.02 cm (200 μm). Then we have $T(0,0,\infty) = 353$°C for Al, 2712°C for 304 S.S., and 705, 237°C for carbon phenolic. Obviously, Al gets warm, 304 stainless steel melts, and carbon phenolic vaporizes. We can see from this the role of K and α. Note that more power or smaller w enters the answer linearly.

Now

$$\frac{T(0,0,t)}{T(0,0,\infty)} = \frac{4(\kappa t)^{1/2}}{w\pi}.$$

Let us assume that the approximation $w \gg 2(\kappa T)^{1/2}$ holds if

$$w/2(\kappa t)^{1/2} = 5.$$

For times such that this inequality holds we have

$$\frac{T(0,0,t)}{T(0,0,\infty)} \sim 0.3.$$

In other words, we reached one-third of the final temperature in the following times: Al, 5.4×10^{-6} sec; 304 S.S., 7.4×10^{-5} sec; and carbon phenolic, 1×10^{-3} sec.

C. Time Required to Achieve Melting or Vaporization

It is interesting to consider the time required to achieve melting or vaporization for a given intensity. Assuming that this can be accomplished while we are in the "short time" or diffusion-free range, we have

$$T(0,0,t) = \frac{2\alpha P_0}{K\pi^{3/2} w^2} (\kappa t)^{1/2}]$$

TABLE I Material Thermal Properties

Material	K (W/cm °C) Solid	K (W/cm °C) Liquid	C (J/g °C)	ρ (g/cm^3)	κ (cm^2/sec)	L_m (kJ/g)	L_v (kJ/g)	T_m[a] (°C)	T_v[a] (°C)
Al	2.0	1.0	1.0	2.7	0.74	0.4	11	640	2430
304 S.S.	0.26	0.26	0.6	8.0	0.054	0.27	4.65	1430	2980
Carbon phenolic	0.01	—	1.7	1.45	0.004		14.6		4000

[a] T_m and T_v are measured as the increase from 20°C need to melt or vaporize, respectively.

and so

$$t_m = \pi^3 w^4 K^2 T_m^2 / 4\alpha^2 P_0^2 \kappa \quad (49)$$

or

$$v = \pi^3 w^4 K^2 V / 4\alpha^2 P_0^2 \kappa.$$

This expression is quite useful in designing a process for a given laser. For example, if the laser is fixed in output capability and a certain material is to be treated, the only variable that can be adjusted to obtain melting is w. Thus it is crucial to prepare proper beam-handling optics. (Of course, it helps to have t_m or $v \propto w^4$.)

1. Rates of Heating and Cooling

Some elementary considerations are presented concerning pulsed heating and cooling. (These matter a great deal when considering studies of rapid resolidification, special alloying, annealing, and metastable phases.) We are concerned with order-of-magnitude estimates of the heating and cooling rates and for this purpose will assume no latent heat due to phase transitions. There are two limiting cases.

1. When α is very large or the optical absorption depth α^{-1} is very small compared to the thermal diffusion depth $(\kappa t_p)^{1/2}$. Here t_p is the duration of the irradiation.

In this case the energy absorbed per unit area (assuming a uniform beam) is $I_0 t_p$ and it is used to heat a layer $(\kappa t_p)^{1/2}$ thick. Thus

$$\Delta T = \alpha I_0 t_p / C\rho(\kappa t_p)^{1/2},$$

and the heating rate is

$$\Delta T / t_p = \alpha I_0 / C\rho(\kappa t_p)^{1/2}.$$

After the laser pulse it takes about the same time, t_p, for the heat from the layer to diffuse a distance $(\kappa t_p)^{1/2}$ into the material. During this time the temperature at the surface drops by an amount that is of order of magnitude equal to ΔT. Thus the cooling rate is also

$$\Delta T / t_p = \alpha I_0 / C\rho(\kappa t_p)^{1/2}$$

2. If α^{-1} is greater than $(\kappa t)^{1/2}$, the light is absorbed in the medium according to

$$I = I_0 e^{-\alpha z},$$

and a temperature distribution is created that is roughly given as

$$T(z) = (1 - R) I_0 (e^{-\alpha z}) t_p / C,$$

where R is the reflectivity at the surface. After the pulse, cooling will occur if heat diffuses a distance of $\sim 1/\alpha$, and so the cooling time is given approximately as

$$t_c = (\alpha^{-1})^2 / \kappa$$

Thus the cooling rate at the surface is

$$\Delta T(z) / t_c = (1 - R)\alpha^3 I_0 \kappa t_p / \rho C.$$

In many applications it is necessary to obtain $T = T_m$. By comparison it is clear that the cooling rate for case 1 will always be very high. In fact,

$$dT/dt = T_m / t_p,$$

and if one has sufficient energy to obtain T_m in, say, a 10^{-9} sec pulse, one will have

$$dT/dt \approx 10^{12} ^\circ\text{C/sec}.$$

III. RESPONSES OF MATERIALS

When laser light is absorbed by a material it results in localized heating and sometimes in a physical change in the heated matter. The change may be desired, as in the many materials processing applications that have been demonstrated, or undesired, as in the formation of laser-induced damage in optical components.

The variety of materials responses to laser heating is so great that not all can be covered here. The key to understanding this phenomenon is to recognize that laser heating is often very localized and very rapid. As estimated in Section II.C, the cooling rate for very localized surface heating can be very large. The result is an ability to rapidly heat and cool materials to produce desired changes. Such desired changes as hardening, annealing, melting, and vaporization are obtained routinely in laser materials processing.

Hardening steels and cast irons requires heating above the martensitic transition point and quenching the material. The laser does this by heating a surface layer (the material in which either the absorption occurred or the heat diffused during the irradiation) and then allowing it to quench by cooling conductively into the bulk. This process has a significant advantage over bulk heating and more conventional quenching methods in that there is very little workpiece distortion. That is, a laser-hardened sample is ready for use with little or no remachining.

In hardening large areas the laser beam can be spread out by defocusing or by scanning a focused beam across the surface. This is possible because in the hardening process one wishes to heat the metal and not to melt it. The overlapped trails of hardened material that are produced contain annealed metal in the regions where they intersect.

However, it is often possible to design the beam and scan parameters to minimize this effect.

Sometimes it is necessary to create alloys on the surface by local rapid heating and cooling. Again, the laser is an ideal tool for this type of controlled melting process. Use of this type of alloying can reduce the amount of expensive hard alloy required for a given purpose. One step beyond alloying is forming metastable forms of alloyed metals by very rapid melt resolidification. This technique, which owes its existence to the very high cooling rates discussed in Section II.C, is used to create and study certain glassy forms of metals. It is hoped that these corrosion- and wear-resistant forms can be reliably created by laser heating and cooling techniques.

The process of melting and vaporizing materials is the way in which lasers weld and drill or cut. These require more tightly focused light to achieve the necessary temperatures. When matter becomes molten, the vapor pressure above it can depress the surface and form a hole. The hole can be supported by the vapor pressure and, because light is more strongly absorbed in the hole, a deep hole is generated. This is the process of forming a "keyhole" to obtain deeppenetration drilling, cutting, or welding. In the welding process the atmosphere and irradiation conditions are chosen to encourage proper mixing of the materials to be welded. In drilling or cutting the conditions are chosen to enhance material removal from the irradiated zone; thus, auxiliary gas jets are used to blow away the molten material and sometimes to react with the heated matter to speed the reaction.

While laser treatment of metals was the first area of laser materials processing to receive major attention, it is by no means the only activity in which lasers are used to process materials. A great deal of work has gone into developing laser techniques for processing organic materials, composites, and ceramics. One type of processing, readily visible in many homes, is the use of lasers in carving very detailed images in wood. This type of artwork is made possible by the selective nature of laser heating and the ease with which optical imaging techniques and computer-controlled workpiece handling can be integrated.

Laser processing of semiconductors, in particular silicon, has generated very wide interest in the electronics industry. The application arose from the need to find efficient and effective ways to anneal ion-implanted silicon. The process of ion implantation creates structural damage in semiconductor material. Annealing an ion-implanted silicon wafer by conventional means requires heating in a convection furnace at temperatures over 1000°C for times of the order of $\frac{1}{2}$ hr. As a result, wafers are often warped and undergo some degree of chemical decomposition. With microelectronic devices dependent on precise electronic and physical properties these present serious problems. In the mid-1970s researchers in the Russia, and later in the United States and Europe, began to study the use of lasers to anneal the implanted surfaces of silicon wafers. While the work began as a way to treat damage due to ion implantation, it has been applied to such other problems as the removal of dislocation networks introduced by high-temperature diffusion and the reduction of misfit defects in epitaxial silicon on sapphire. Lasers have also been used to provide localized, controllable heat sources for epitaxial growth of evaporated silicon on crystalline silicon, for the growth of large-grain polysilicon from smallgrain polysilicon, and for the formation of metallic silicides and metal overlayers on silicon.

The interaction that makes this type of processing possible is again that of localized rapid heating and cooling. When sufficient heating has taken place there can be a solid phase rearrangement of atoms in the silicon that was damaged by the implantation or, if melting has been achieved, recrystallization can occur with the undamaged material serving as the seed. There is no debate that the processes of interaction that allow these effects to take place are thermal when cw lasers or lasers with pulse durations greater than 100 nsec are used. If pulses less than 100 nsec in duration are employed, solid phase rearrangements cannot take place quickly enough and the material must be melted by the irradiation. The process is similar to that of laser glazing of metals by very rapid melting and resolidification.

Initially there was some concern over the process of semiconductor annealing using pulses with durations in the subpicosecond range. This centered on the issue of whether there was time for thermal processes to operate or whether some other interaction had to be invoked to explain the observed phenomenon. Elegant experiments were performed to which the process of melting and resolidification with ultrashort pulse irradiation was monitored by Raman scattering. The interaction process was confirmed to be thermal in nature. This is a very important scientific result in laser–materials interactions research in that it confirms the fact that thermalization takes place in times as short as several femtoseconds.

Other laser–materials interactions applied in the electronics industry involve the unique ability of laser light to be focused to very small areas and to heat only the areas irradiated. This allows lasers to serve a wide range of resistor and capacitor trimming applications. Also, when properly focused, lasers are able to remove unwanted bridges in microcircuits. If the atmosphere above the substrate is selected to form desired deposits when the substrate or the gas is heated, then lasers can be used to personalize microcircuit masks and to repair damaged circuits. Work with ultraviolet lasers may make it possible

to process microcircuit components with submicrometer dimensions.

Lasers of sufficient intensity can cause very highly reflective or transparent materials to be damaged as a result of very high order processes that come into play when very high optical intensities are employed. In Section I.B. the process of nonlinear absorption and breakdown was described. These processes result in very rapid conversion of the nonabsorbing matter into matter that absorbs nearly 100% of the incident light. The absorption takes place locally in the irradiated region and the heated material melts and may vaporize. When melting or vaporization occurs in the bulk of a transparent substance severe cracking also takes place. In addition, some materials display slip banding after intense laser irradiation, indicating that they were heated to the point where plastic deformation could occur.

While breakdown due to nonlinear processes represents the intrinsic failure mechanism of nonansorbing materials, laser-induced damage is most often determined by the presence of some discrete inclusion or irregularity in the irradiated volume. Damage due to absorption at such defects can be detected as arrays of discrete, randomly located damage sites within the irradiated region. In thin films and on most bare surfaces this type of defect-determined damage dominates the interaction and response of the matter with laser light.

SEE ALSO THE FOLLOWING ARTICLES

ELECTROMAGNETICS • LASERS • PLASMA SCIENCE AND ENGINEERING

BIBLIOGRAPHY

Bass, M. (1983). *In* "Physical Processes in Laser Materials Interactions," pp. 77–116. Plenum, New York.

Bertolotti, M., ed. (1983). "Physical Processes in Laser Materials Interactions," pp. 175–220. Plenum, New York.

Brown, W. L. (1983). *In* "Laser Materials Processing" (M. Bass, ed.), pp. 337–406. North-Holland Publ., Amsterdam.

Chryssolouris, G. (1991). "Laser Machining Theory and Practice," Springer-Verlag, New York.

Duley, W. W. (1976). "CO_2 Lasers: Effects and Applications." Academic Press, New York.

Ready, J. F. (1971). "Effects of High Power Laser Radiation." Academic Press, New York.

Siegman, A. E. (1986). "Lasers," University Science Books, Mill Valley, California.

Steen, W. M. (1991). "Laser Material Processing," Springer-Verlag, New York.

Wooten, F. (1972). "Optical Properties of Solids." Academic Press, New York.

Yardley, J. T. (1985). *In* "Laser Handbook" (M. Bass and M. L. Stitch, eds.), Vol. 5, pp. 405–454. North-Holland Publ., Amsterdam.

Lasers

William T. Silfvast
School of Optics — CREOL, University of Central Florida

I. Laser History
II. Laser Gain Media
III. Laser Beam Properties
IV. Laser Linewidth
V. Laser Wavelengths
VI. Types of Lasers
VII. Laser Properties Associated with Applications

GLOSSARY

Absorption Extinction of a photon of light when it interacts with an atom and excites an internal energy state of that atom.
Emission Radiation produced by an atomic species when an electron moves from a higher energy level to a lower one.
Frequency Reciprocal of the time it takes for a light wave to oscillate through a full cycle.
Gain Condition that causes a beam of light to be intensified when it passes through a specially prepared medium.
Linewidth Frequency or wavelength spread over which emission, absorption, and gain occur in the laser amplifier.
Mode Narrow frequency beam of light that follows a unique path as it grows in a laser amplifier and emerges as a beam.
Photon Discrete quantum of light of an exact energy or wavelength.
Population inversion Condition in which more atoms exist in a higher energy state than a lower one, leading to amplification, or gain, at a wavelength determined by electron transitions between those states.
Wavelength Distance over which light travels during a complete cycle of oscillation.

A LASER is a device that amplifies, or increases, the intensity of light, producing a strong, highly directional or parallel beam of light of a specific wavelength. The word *laser* is an acronym for light amplification by stimulated emission of radiation. Stimulated emission is a natural process, first recognized by Einstein, that occurs when a beam of light passes through a medium and initiates or stimulates atoms of that medium to radiate more light in the same direction and at the same wavelength as that of the original beam. A specific laser device (see Fig. 1) consists of (1) an amplifying, or gain, medium that produces an increase in the intensity of a light beam, and (2) an optical resonator or mirror arrangement that provides feedback of the amplified beam into the gain medium for further amplification, thereby producing the beam-like and ultrapure frequency or coherent properties of the laser. The optical cavity, or resonator, which typically consists of two

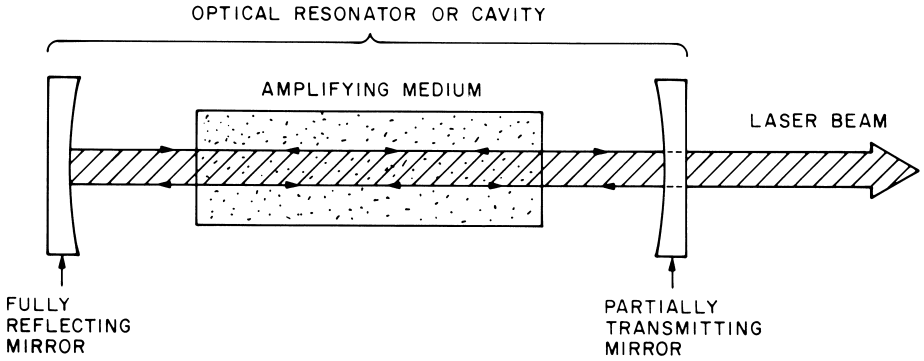


FIGURE 1 Laser components including amplifying medium, optical cavity, and laser beam.

highly reflecting mirrors arranged at opposite ends of an elongated amplifier, allows a strong beam of light to develop due to multiple reflections that produce growth of the beam as it bounces back and forth through the amplifier. A useful beam emerges from a laser either by making one of the laser mirrors partially transmitting or by using a mirror with a small hole in it. Some lasers have such a high gain that the intensity increase is large enough, after only a single pass of the beam through the amplifier, that mirrors are not necessary to achieve a strong beam. Such amplifiers produce a directional beam by having the gain medium arranged in a very elongated shape, causing the beam to grow and emerge from the amplifier in only the elongated direction.

Some of the unique properties of lasers include high-beam power for welding or cutting, ultrapure frequency for communications and holography, and an ultraparallel beam for long-distance propagation and extremely tight focusing. Laser wavelengths cover the far infrared to the near infrared, the visible to the ultraviolet, and the vacuum ultraviolet to the soft X-ray region. Laser sizes range from small semiconductor lasers the size of a grain of salt, for use in optical communications, to large solid-state and gas lasers the size of a large building, for use in laser fusion programs.

I. LASER HISTORY

Charles Townes was the first person to take advantage of the stimulated emission process to be used in the form of an amplifier by conceiving and constructing the first *maser* (an acronym for microwave amplification by stimulated emission of radiation). The maser produced a pure beam of microwaves that were anticipated to be useful for communications in a similar way to that of a klystron or a traveling-wave tube. The first maser was produced in ammonia vapor, and the inversion occurred between two energy levels that produced gain at a wavelength of 1.25 cm. In the maser, the radiation wavelengths are comparable to the size of the device, and therefore the means of producing and extracting the beam could not obviously be extrapolated to the optical spectrum in which the wavelengths of light are of the order of 100th the size of a human hair.

In 1958, Townes, along with Arthur Schawlow, began thinking about extending the maser principle to optical wavelengths. At that time they developed the concept of a laser amplifier and an optical mirror cavity to provide the multiple reflections thought to be necessary for rapid growth of the light signal into an intense visible beam. Townes later (1964) shared the Nobel Prize in Physics with A. Prokhorov and N. Basov of the Russia for the development of the maser–laser principle.

In 1960, Theodore Maiman of the Hughes Research Laboratories produced the first laser using a ruby crystal as the amplifier and a flash lamp as the energy source. The helical flash lamp surrounded a rod-shaped ruby crystal, and the optical cavity was formed by coating the flattened ends of the ruby rod with a highly reflecting material. In operation, an intense red beam emerged from the ends of the rod when the flash lamp was initiated.

Shortly after the ruby laser came the first gas laser, developed in 1961 in a mixture of helium and neon gases by A. Javan, W. Bennett, and D. Herriott of Bell Laboratories. At the same laboratories, L. F. Johnson and K. Nassau first demonstrated the now well-known and high-power neodymium laser. This was followed in 1962 by the first semiconductor laser demonstrated by R. Hall at the General Electric Research Laboratories. In 1963, C. K. N. Patel of Bell Laboratories discovered the infrared carbon dioxide laser, which later became one of the most powerful lasers. Later that year A. Bloom and E. Bell of Spectra-Physics discovered the first ion laser, in mercury vapor. This was followed in 1964 by the argon ion laser developed by W. Bridges of Hughes Research

Laboratories and in 1966 the blue helium-cadmium metal vapor ion laser discovered by W. T. Silfvast, G. R. Fowles, and B. D. Hopkins at the University of Utah. The first liquid laser in the form of a fluorescent dye was discovered that same year by P. P. Sorokin and J. R. Lankard of the IBM Research Laboratories, leading to the development of broadly tunable lasers. The first of the rare-gas-halide excimer lasers was first observed in xenon fluoride by J. J. Ewing and C. Brau of the Avco-Everett Research Laboratory in 1975. In 1976, J. M. J. Madey and co-workers at Stanford University developed the first free-electron laser amplifier operating at the infrared carbon dioxide laser wavelength. In 1985 the first soft X-ray laser was successfully demonstrated in a highly ionized selenium plasma by D. Matthews and a large number of co-workers at the Lawrence Livermore National Laboratory.

II. LASER GAIN MEDIA

A. Energy Levels and the Emission and Absorption of Light

All lasers (except the free-electron laser) result from electron energy changes among discrete energy levels of atomic species, including (1) individual atoms or ions, (2) small uniquely bonded groups of atoms (molecules), (3) periodically arranged groups of atoms (semiconductors or crystalline solids), or (4) randomly arranged groups of atoms (liquids and amorphous solid structures). All of these species contain a lowest energy level (ground state) in which the electrons reside at low temperatures, and a spectrum of higher Iying levels that are occupied when energy is pumped into the species either by irradiating it with light (optical pumping) or colliding the material with energetic particles such as fast electrons or fast atomic particles. Light originates from these species when electrons jump or decay from some of these high lying, or excited, energy levels to lower lying energy levels. The energy that is lost by the particular atomic material when the electron decays is given up in the form of a photon, a discrete particle of light. In order to satisfy the law of conservation of energy, the emitted photon must be of the exact energy corresponding to the energy difference between the higher lying level and the lower lying level, and the wavelength or frequency of the emitted photon is associated with that energy difference. Many photons together can form a beam or a wave of light.

Light, emitted when an electron decays, can occur either spontaneously, due to inherent interactions of the atomic structure, or by stimulated emission whereby the electron is forced or driven to radiate by an approaching photon of the appropriate energy or wavelength (see Fig. 2). Absorption, the opposite process of stimulated emission, occurs when an atom having an electron in a low-lying energy level absorbs and thereby eliminates an approaching photon, using the absorbed energy to boost that electron to a higher lying energy level.

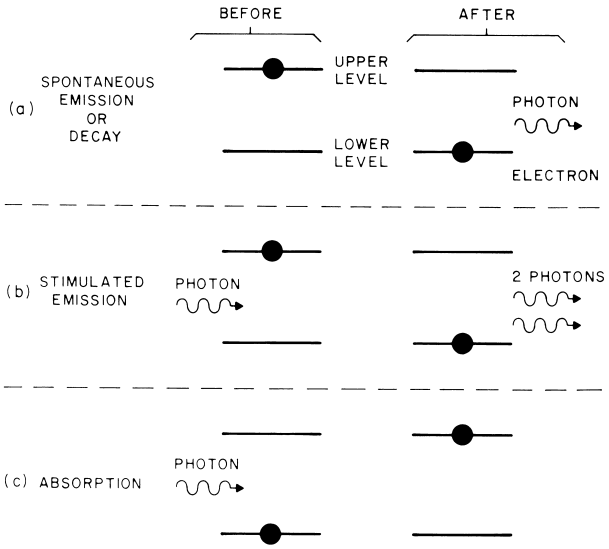


FIGURE 2 Electronic transitions between energy levels depicting (a) spontaneous emission, (b) stimulated emission, and (c) absorption.

B. Population Inversions

A laser amplifier is produced when conditions are created within the amplifying medium such that there are more atoms having electrons at a higher energy level than at a lower energy level for a specific pair of levels. This condition is known as a population inversion, since it is the opposite or inverse of almost all physical situations at or near thermal equilibrium in which there are more atoms with electrons in lower energy levels than at higher levels. Under the conditions of a population inversion, when a beam of light passes through the amplifier, more photons will be stimulated than absorbed, thereby resulting in a net increase in the number of photons or an amplification of the beam.

C. Excitation Mechanisms in Lasers (Energy Sources)

Since all laser emission involves radiation from excited states of atoms, the energy must be fed to those atoms to produce the excited states. This energy is provided in the form of highly energetic electrons (moving at rapid speeds of the order of 10^8 to 10^9 cm/see), energetic heavier particles such as protons, neutrons, or even other atoms, or

electromagnetic radiation (light) in the form of (1) a broad frequency spectrum of emission, such as a flash lamp, or (2) a narrow frequency spectrum provided by another laser. The most common excitation source is that provided by energetic electrons since they are easily accelerated by applying an electric field (voltage drop) to an amplifier. Electrons are typically used in most gas and semiconductor lasers, whereas light is most often used in liquid (dye) lasers and crystalline solid-state lasers. Electron excitation sources tend to be the most efficient since flash lamps are themselves generally excited by electrons before they are used to pump lasers. Heavier particles are generally less efficient as pumping sources since they are much more difficult to energize than either electrons or lamps.

D. Inversions in Gases, Metal Vapors, and Plasmas

Inversions in gases, metal vapors, and plasmas are generally produced by applying a voltage across the elongated gain region, thereby producing an electric field that accelerates the electrons within that region. These rapidly moving electrons then collide with the gas atoms and excite them to a wide range of excited energy levels. Some of those levels decay faster than others (primarily by spontaneous emission), leaving population inversions with higher levels. If the populations in the inverted levels are high enough, then the gain may be sufficient to make a laser. Typically if 1 in every 10^5 or 10^6 atoms is in a specific upper laser level, that will be a sufficient fraction to produce enough gain to make a laser. Most gas lasers have relatively low gains and therefore amplifier lengths of the order of 25 to 100 cm are necessary. Since spontaneous emission rates are much faster for shorter wavelength transitions, power input for short-wavelength lasers is significantly higher than for visible and infrared lasers. Typical gas pressures for gas lasers range from 1/100th to 1/1000th of an atmosphere, although there are some gas lasers that operate at atmospheric pressure and above in order to produce the necessary excited state populations for sufficient laser gain. These lasers require very closely spaced electrodes with transverse excitation in order to produce the necessary excitation at high pressure.

E. Inversions in Liquids (Dyes)

Most excited states of liquids decay so rapidly by collisions with surrounding atoms or molecules (10^{-13} sec) that it is difficult to accumulate enough population in an upper laser level to make significant gain. Also, since it is difficult to use electron excitation in liquids, the primary energy source is optical excitation, either by flash lamps or by other lasers. Fluorescing dyes are the best liquid media for lasers. The fact that dyes fluoresce suggests that their excited energy levels stay populated long enough to lose their energy by radiation of light rather than by collisions with surrounding atoms or electrons. To help establish the population inversion, the lower laser level of a dye decays very rapidly by collisions.

F. Inversions in Semiconductors

Inversions in semiconductors are produced when a *p-n* junction is created by joining two slightly different semiconducting materials (in a similar way to that of a transistor). The n-type material has an excess of electrons and the p-type material has an excess of holes (missing electrons). When they are joined, the excess electrons of the n-type material are pulled over into the p region and vice versa by their charge attraction, causing the electrons and holes in that region to recombine and emit recombination radiation. This neutralizes the junction region, leaving a small, inherent electric field to prevent further recombination of the remaining electrons and holes near the junction. If an external electric field is applied in the appropriate direction, by applying a voltage across the junction, more electrons and holes can be pulled together, causing them to recombine and emit more radiation but also to produce a population inversion. This inversion occurs on transitions originating from energy levels located just above the inherent energy band gap of the material. Semiconductor lasers operate in a similar way to light-emitting diodes except that the requirements for constructing the lasers are much more restrictive than the diodes, due primarily to the necessity for higher electric current densities that are essential to produce large gains, and also due to the need for better heat-dissipation capabilities to remove the heat produced by the higher current densities.

G. Inversions in Solids

Inversions in most solid-state lasers are obtained by implanting impurities (the laser species) within a host material such as a crystal or a glass in a proportion ranging from approximately 1 part in 100 to 1 part in 10,000. In most solid-state lasers the impurities are the form of ions in which the energy states are screened from the surrounding atoms so the energy levels are narrow, like those of isolated atoms or ions, rather than broad like those of liquids. In color center lasers the impurities are crystal defects produced by irradiating the crystal with X-rays. In solids, as in liquids, electrons cannot easily be accelerated by electric fields to excite the laser energy levels of the impurity species so the energy must be fed to the medium via flash lamps or other lasers. The input lamp energy occurs over a broad wavelength region to a large number of excited energy levels. These levels then decay to the upper laser level, which acts much like a temporary storage reservoir,

collecting enough population to make a large inversion with respect to lower lying levels that have a rapid decay to the ground state.

H. Bandwidth of Gain Media

The frequency or wavelength spectrum (gain bandwidth) over which gain occurs in a laser amplifier is determined by a number of factors. The minimum width is the combined widths of both the energy levels involved in the laser transition. These widths are due primarily to the uncertainty of the natural radiative decay time of the laser levels. This combined width can be increased by collisions of electrons with the laser states in high-pressure gas lasers, by interaction of nearby atoms and/or bonding electrons in liquids and solids, and by Doppler-shifted frequencies in most gas lasers. This gain linewidth of the laser amplifier is only part of the contribution to the linewidth of the laser beam. Significant line narrowing due to optical cavity effects will be described in the next section.

III. LASER BEAM PROPERTIES

A. Beam Properties

The beam properties of a laser, such as the direction and divergence of the beam and the wavelength or frequency characteristics that are not related to the bandwidth of the laser gain medium, are determined largely by the laser structure. The features of the structure affecting the beam properties include the width and length of the gain medium, the location, separation, and reflectivity of the mirrors of the optical cavity (if the gain is low and the gain duration long enough to make use of mirrors), and the presence of losses in the beam path within the cavity. These features determine unique properties of the laser beam referred to as laser modes.

B. Shape of the Gain Medium

If a laser gain medium were in the shape of a round ball, then stimulated emission would occur equally in all directions, and the only result might be a slight increase in the intensity of the light, for a slightly shorter duration than would occur if gain were not present, but the effect would probably not be noticeable to an observer. The goal of a laser designer is to cause most of the laser photons to be stimulated in a specific direction in order to produce a highly directional beam, instead of allowing those same photons to radiate in random directions by either spontaneous emission or stimulated emission (as was the case for the round ball). This is achieved by making the gain medium significantly longer in one dimension than in the other two.

C. Growth of the Beam and Saturation

When gain is produced in the amplifier and spontaneously emitted photons begin to be amplified by stimulated emission, photons that are emitted in directions other than the elongated direction of the amplifier soon reach the walls of the medium and die out. The photons that are emitted in the elongated direction continue to grow by stimulating other atoms to emit additional photons in the same direction until all of those photons reach the end of the amplifier. They then arrive at the mirror and are reflected back through the amplifier where they continue to grow. Finally, after a number of round trips, a beam begins to evolve. If the duration of the gain is long enough, amplification will lead to more photons in the beam than there are atoms available to be stimulated, and growth can therefore no longer occur. The beam is then said to be saturated and the beam power reaches a steady value, determined by the amount of energy being fed into the upper laser level by the pumping source. If the population inversion can be maintained on a continuous basis in the amplifier, the laser-beam output becomes steady and the laser is referred to as a continuous wave (cw) laser. If the gain only lasts for a short duration, the laser output occurs as a burst of light and the laser is referred to as a pulsed laser.

D. Optical Cavity (Optical Resonator)

The optical cavity or resonator is typically comprised of a mirror at each end of the elongated gain region (see Fig. 1). This resonator allows the beam to bounce back and forth or resonate between the mirrors, growing very rapidly as it does so. Although the first lasers used flat mirrors, as suggested in the original Schalow-Townes paper (1958), in 1961 Fox and Li suggested the use of slightly curved mirrors, especially for cavities where the amplifier was long and narrow, in order to reconcentrate the beam towards the central axis of the gain medium after each reflection from the mirrors. This was shown to reduce the diffraction losses of the narrow tube. Stable laser beam modes of lower loss are possible for slightly concave spherical mirrors than for flat mirrors if the separation between the mirrors is less than twice their radius of curvature. Mirror reflectivities of 99.9% at the laser wavelength, using multilayer dielectric reflective coatings, make possible laser operation under conditions of very low gain.

E. Stable and Unstable Resonators

The term *resonator* implies a wave that is in harmony or resonance with the device that is generating the wave, whether it be an organ pipe, a flute, or a microwave cavity. An optical wave can also have this property. The term *resonance* suggests that an exact integral number of half

wavelengths (a mode) of the wave fit between the mirrors of the resonator. A stable resonator refers to a mirror arrangement (usually one with a mirror at each end of the elongated cavity), producing modes that are continually reproduced during the duration that gain occurs in the amplifier (which could be a thousandth of a second or many days or longer). An unstable resonator is a mirror arrangement that is used to obtain modes when the amplifier gain is high and has a short duration (less than a millionth of a second) such that a normal mode would not have time to evolve. In that situation, the energy is extracted by using a mirror arrangement in which the beam begins to resonate in a small unstable region of the amplifier. Part of this beam is leaked into the larger portion of the amplifier, where it rapidly grows and extracts most of the amplifier energy in a few passes.

F. Laser Modes

Laser modes are wavelike properties of the beam of light that evolve while the beam passes back and forth through the amplifier, bouncing between the mirrors. The beam grows as long as the gain in the amplifier exceeds the losses within the cavity. The development of modes involves an attempt by competing portions of the light beam having slightly different frequencies, to fit an *exact* number of their waves into the optical cavity with the constraint that the oscillating electric field of the light beam is zero at each of the mirrors. This is much like a vibrating guitar string that is constrained at each end by the bridge and a fret, but is free to vibrate with as many nodes and antipodes in the region in between as it chooses. As an example, a laser mode of green light having a wavelength of *exactly* $5.0 \times 10\text{-}5$ cm will fit *exactly* 1,000,000 full cycles of oscillation between the laser cavity mirrors if they are separated by a distance of *exactly* 50 cm. Most lasers have a number of modes operating simultaneously, in the form of both longitudinal and transverse modes, which give rise to a complex frequency and spatial structure within the beam in what might otherwise appear as a relatively simple, pencil-like beam of light.

G. Longitudinal Modes

Each longitudinal mode is a separate light beam traveling along a distinct path between the mirrors and having an exact integral number of half wavelengths along that path. In the example of green light mentioned previously, three different longitudinal modes would have very slightly different wavelengths of green light (indistinguishable in color to the eye) undergoing respectively 1,000,000, 1,000,000.5, and 1,000,001 full cycles of oscillation between the mirrors while traveling exactly the same path back and forth through the amplifier (see Fig. 3). In this situation each mode would differ in frequency by exactly 300 MHz (3×10^8 Hz) as determined by the velocity of light (3×10^{10} cm/see) divided by twice the cavity length (2×50 cm). A gas laser amplifier having a relatively narrow gain width of 3 GHz could therefore fit 10 longitudinal modes within the gain bandwidth, whereas a liquid (dye) laser having a bandwidth covering up to one-fifth of the visible spectrum of light could have as many as 100,000 longitudinal modes all oscillating simultaneously if they are suitably locked together in phase.

H. Transverse Modes

Whereas longitudinal modes involve many light beams traveling exactly the same path through the amplifier, but differing in wavelength by an amount determined by the total number of wave half-cycles that fit between the mirrors, different transverse modes are represented by slightly different optical paths as they travel through the amplifier (Fig. 4). The beam that emerges from the laser then has a larger divergence as more transverse modes are allowed to operate. This divergence can be restricted by

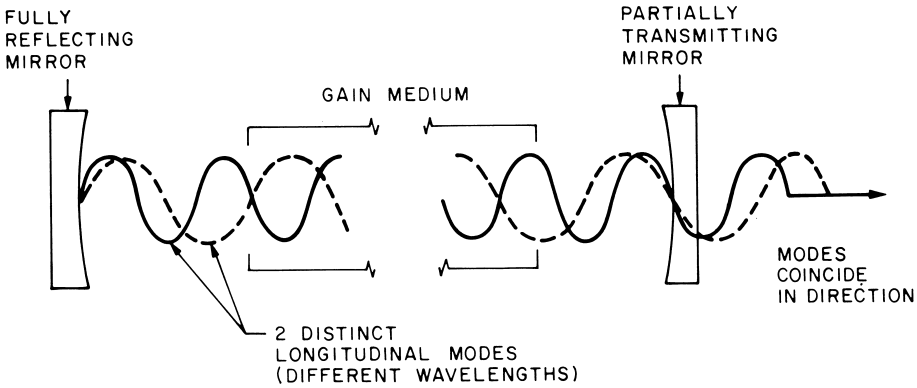


FIGURE 3 Two distinct longitudinal modes occupying the same spatial region of the laser optical cavity.

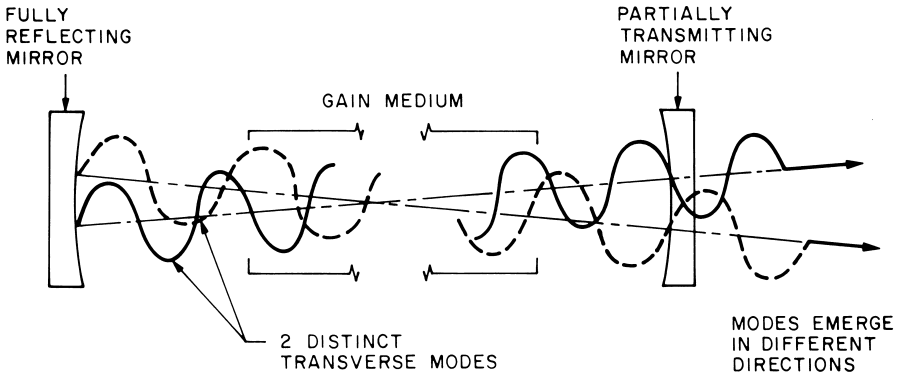


FIGURE 4 Two distinct transverse modes oscillating over different spatial regions of the laser optical cavity.

placing a small aperture within the laser cavity to prevent the more divergent transverse modes from developing. Each transverse mode traveling over its unique path could consist of several longitudinal modes oscillating along that path. In most instances, closely located transverse modes differ in frequency by a smaller value than do adjacent longitudinal modes that follow the same path through the amplifier.

IV. LASER LINEWIDTH

A. Linewidth of Laser Gain Media

All laser amplifiers have a finite frequency width or wavelength width over which gain can occur. This width is related to the widths of the energy levels involved in the population inversion. The linewidth is determined by both a homogeneous component and an inhomogeneous component.

1. Homogeneous Linewidth

The homogeneous component includes all mechanisms that involve every atom in both the upper and lower laser levels of the gain medium in identically the same way. These mechanisms include the natural radiative linewidth due to spontaneous emission, broadening due to collisions with other particles such as free electrons, protons, neutrons, or other atoms or ions, or power broadening in which a high-intensity laser beam rapidly cycles an atom between the upper and lower levels at a rate faster than the normal lifetime of the state. In each of these effects, the rate at which the process occurs determines the number of frequency components or the bandwidth (linewidth) required to describe the process, with faster processes corresponding to broader linewidths. The natural radiative linewidth is associated with the radiative decay of individual atoms. It is related to the radiative decay rates of both the upper and lower energy levels involved in the laser transition. Those rates are inherent to any specific set of levels and determined by the nature of the atomic structure for that species and electron configuration. The natural linewidth can vary anywhere from 10^4 Hz in the infrared to 10^{10} Hz in the soft X-ray spectral region, and the actual linewidth can increase to as high as 10^{12}–10^{13} Hz or more if collisional broadening occurs in high-density liquid or solid laser materials.

2. Inhomogeneous Linewidth

The inhomogeneous component of the broadening results from processes that affect different atoms in the upper and lower laser levels in different ways depending on unique characteristics of those atoms. The most common example of inhomogeneous broadening is Doppler broaden-ing or motional broadening that results from the random thermal motion of the atoms due to their finite temperature. In this effect, atoms traveling in one direction would see an approaching lightwave as having a specific frequency, and atoms traveling in the opposite direction would see that same lightwave as having a lower frequency. In a similar way, light emitted from each of those atoms would be seen by an observer as having different frequencies or wavelengths, even though each atom emitted the same frequency. This Doppler effect is similar to what happens when a train approaches with its horn blowing, and the pitch of the sound changes from a high to a lower tone as the train first approaches the listener and then moves away, even though the frequencies emitted from the horn remain the same.

B. Linewidth in Laser Amplifiers

The dominant broadening effect for most gas laser amplifiers is Doppler broadening, which produces a linewidth of the order of 10^9 to 10^{10} Hz for visible and shorter wavelength amplifiers. Molecules have linewidths that are determined not only by the jumping of electrons between

energy levels, as is the case of individual atoms, but also by the rotation of the atoms around a common center of gravity and by the vibrations of the atoms as though they were tied together by various springs. These rotations and vibrations produce a series of emissions of light over a relatively narrow range of wavelengths, and in some instances these emissions can overlap in frequency to produce broad emission lines. If no overlap occurs, then the emission associated with vibrational transitions consists of an equally spaced series of narrow lines.

Liquids have their energy levels broadened due to the interaction of the closely packed atoms of the liquid. Such dense packing leads to an overlap of the electronic emission from various closely spaced rotational and vibrational levels, thereby producing emission spectrum widths of the order of 10^{13} to 10^{14} Hz.

Emission linewidths on laser transitions in solid-state laser materials is determined both by the combined widths of groups of closely spaced energy levels associated with the transitions and by collisional broadening that merges those widths together to make a single larger energy band. This "band" can be relatively narrow for a solid material (of the order of 10^{11}–10^{12} Hz for the Nd laser) or relatively broad (10^{14} Hz for the Ti:sapphire laser). The linewidth depends upon the electronic arrangements of the bound electrons of the dopant ions with the atoms of the host material.

Emission linewidths in doped semiconductor laser materials is determined by the distribution of electrons in the conduction and valence bands of the semiconductor material. This is largely determined by the temperature and is of the order of 10^{13} Hz. An exception to this is in quantum well semiconductor laser material, in which the valence and conduction band energy levels are discrete and therefore 5–10 times narrower than ordinary doped semiconductor materials.

C. Line Narrowing Due to Cavity Effects

A laser cavity tends to select specific wavelengths (Fig. 5) within the normal gain bandwidth of the laser gain medium

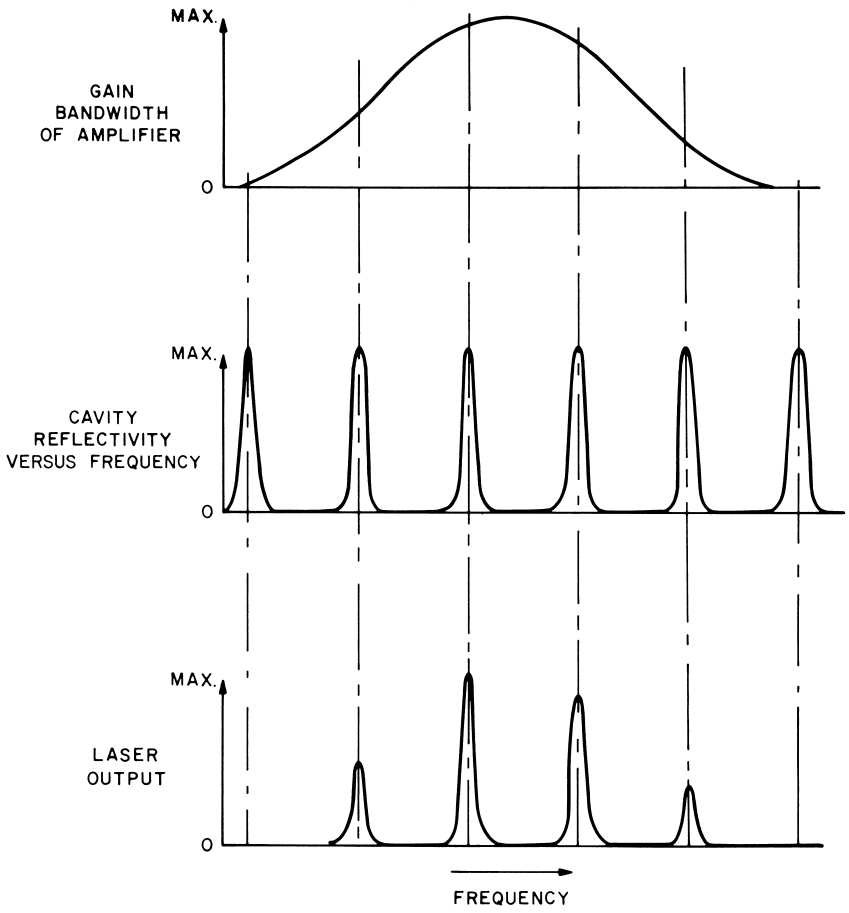


FIGURE 5 Laser output resulting from effects of laser amplifier gain bandwidth and laser optical cavity modes.

that have an exact integral number of waves that fit between the mirrors (modes). These modes, which (if there are more than one) are equally spaced in frequency or wavelength and tend to be amplified at the expense of other wavelengths that suffer losses by not exactly "fitting" between the mirrors. In principal these modes can "narrow up" to widths of the order of a few hertz, but cavity stability problems tend to keep them from going much below 1000 Hz unless extremely stable environments and rigid cavity structures are available. Typically a homogeneously broadened gain medium, such as a solid-state laser, will have only one of these modes, thus providing a single very narrow laser linewidth. A Doppler broadened line, on the other hand, can have many of such modes as shown in Fig. 5.

V. LASER WAVELENGTHS

A. Range of Wavelengths

Wavelengths of electromagnetic radiation covering the spectral region where lasers occur are referred to in terms of fractions of meters. The two most common units are the micrometer (μm) or 10-6 m, often referred to as microns, and the nanometer (nm) 10-9 m. Micrometers are used to designate the infrared region ranging from 0.7 μm to approximately 1,000 μm. Nanometers are used to cover the spectral range from the visible at 700 nm (0.7 μm) down to approximately 1 nm in the soft X-ray region. Each spectral region covers a specific wavelength range, although the boundaries are not always exact. The infrared is broken up into three regions. The far infrared ranges from about 15 to approximately 1,000 μm (approaching the microwave region). The middle infrared covers from 2–15 μm, and the near infrared ranges from 0.7–2 μm. The visible region includes the rainbow spectrum, ranging from the violet at 400 to 440 nm, the blue from 440–490 nm, the green from 490–550 nm, the yellow from 550–580 nm, the orange from 580–610 nm, and the red from 610 to 700 nm. The ultraviolet ranges from 400–200 rim with the 400–300-nm region termed the near ultraviolet and the 200–300-nm region the far ultraviolet. The vacuum ultraviolet covers the range from 100–200 nm, since it is the region where radiation can no longer transmit through the air (thus requiring a vacuum) because of absorption due to air molecules (mostly oxygen) and because transmission optics can still be used in this region. The shortest region is the extreme ultraviolet, which extends into the soft X-ray region, covering the range from 100–100 nm. In this region only reflection optics can be used and even then the reflectivities of the best materials are quite low (especially below 40 nm) when compared to those available in the visible. Free electron lasers have the potential for operating at all wavelengths and will therefore not be singled out for any specific wavelength in this section.

B. Infrared Lasers

The most powerful lasers occur in the middle and near infrared, but a large number of lasers have been produced in the far infrared. These include the discharge-excited water-vapor lasers with wavelengths ranging from 17–200 μm, the cyanide laser operating at 337 μm, and the laser-pumped gases of methyl fluoride (450–550 μm) and ammonia (81 μm). There are also two kinds of semiconductor lasers that operate in the far infrared. These are the quantum cascade lasers and the p-germanium laser. In the middle infrared, the carbon dioxide laser is the most significant laser operating primarily at 9.6 and 10.6 μm. The other strong middle infrared laser is the carbon monoxide laser which emits in the 5–6 μm region. There is also an important chemical laser, the hydrogen-fluoride laser, which operates at 3.7 μm. In the near infrared, many more lasers are available. The most significant solid-state lasers are the neodymium and the titanium-sapphire doped laser crystals and glasses. The Nd laser species is most often doped into either a yttrium-aluminum-garnet (YAG) host crystal, referred to as the Nd:YAG laser, or in a glass host called a Nd:glass laser, both operating at 1.06 μm. The titanium sapphire laser operates over a very wide wavelength range from 0.65 μm to over 1.0 μm and is therefore very suitable for use in a mode-locking cavity configuration to generate ultrashort laser pulses. Also the alexandrite laser is a solid-state laser tunable from 0.7–0.8 μm and color center lasers are tunable from 0.8–4.0 μm. Tunable dye lasers also extend into the infrared to about 1.5 μm. Semiconductor lasers operate over the range from 0.65 to approximately 1.70 μm; their wavelengths being extended in both directions with continuing research. These include Ga-As lasers in the 0.8 μm wavelength region, with wavelengths extended to as short as 0.63 μm with aluminum impurities added (AlGaAs). Also operating in the 1.55-μm region are the InP/InGaAsP heterostructure lasers that are important for optical fiber communications.

C. Visible Lasers

Visible lasers are primarily dominated by gas lasers and tunable dye lasers. The red helium-neon laser at 633 nm is one of the most commonly used lasers. The argon ion laser covers the blue and green spectrum with the most prominent wavelengths at 488 and 515 nm. The krypton ion laser covers the spectrum from green to red with some of the most prominent wavelengths at 521, 568, and 647 nm. The helium–cadmium laser operates in the blue at 442 nm. Tunable dye lasers operate over the entire visible

spectrum, generally pumped by a slightly shorter wavelength laser, such as the argon ion laser or the frequency doubled or tripled Nd:YAG laser. The rhodamine 6G dye has the lowest pump threshold for a dye and operates over a wavelength range from 570 to 630 nm.

D. Ultraviolet Lasers

There are fewer lasers available in ultraviolet primarily because pump thresholds become much higher than for visible lasers and because laser cavity mirror reflectivities are slightly lower and less durable. The two prominent continuous lasers are the argon ion laser, operating primarily at 351 nm and the helium cadmium laser at 325 and 354 nm. The rare-gas-halide excimer lasers produce high pulsed powers at 351 rim for xenon-fluoride, 308 nm for xenon chloride, and 248 nm for krypton fluoride. The pulsed nitrogen laser operating at 331 nm was extensively used for pumping dye lasers, but this has now largely been replaced with the frequency doubled and tripled Nd:YAG laser and the excimer lasers. Tunable dye lasers as short as 310 nm are available, but they are not easily pumped by commercial lasers and therefore frequency mixing and doubling of visible and infrared dye lasers is the most common technique for producing tunable laser radiation in this spectral region.

E. Vacuum Ultraviolet Lasers

The molecular hydrogen laser was the first laser developed in this spectral region, operating in the 120- and 160-nm ranges, but it never became a useful device. There are now only two readily available vacuum ultraviolet lasers. These are the argon–fluoride excimer laser, operating at 193 nm, and the fluorine molecular laser emitting at 157 nm. Much of the research requiring lasers in this spectral region is accomplished by frequency summing and mixing of various visible and ultraviolet lasers to produce coherent output, but such techniques are not readily available to the general laser community.

F. Extreme Ultraviolet and Soft X-Ray Lasers

Until 1984 there were no lasers in this spectral region. At that time, stimulated emission was reported in krypton at 93 nm. Then in 1985 several highly ionized atoms provided laser output in the soft X-ray end of this spectrum, including 24th ionized selenium at 21 nm, 5th ionized carbon at 18 nm, 29th ionized yttrium at 15 nm, and 32nd ionized molybdenum at 13 nm. These lasers are made by focusing a powerful laser onto a target material composed of the laser species. An elongated plasma is produced within which the laser gain occurs. They are not easily duplicated in other laboratories. More recently a discharge-pumped laser of this type has been produced in eighth ionized argon, with laser output at 46.9 nm. This laser has been miniaturized to table-top size and can be operated at a repetition rate of up to 10 Hz, a very high rate for this type of laser.

VI. TYPES OF LASERS

A. Gas Lasers

The most common types of gas lasers are the helium-neon laser, the argon and krypton ion lasers, the carbon dioxide laser, the rare-gas-halide excimer lasers, and the chemical lasers, most notably the hydrogen-flouride laser. Metal vapor lasers also fit into this category but are treated separately in Section V. With few exceptions, these lasers receive their energy input via collisions of gas atoms with high-energy electrons. This energy is provided by applying a high voltage between electrodes located within the gaseous medium in order to accelerate the electrons to the necessary high energies. In some instances the electrons first excite a storage level in a separate species within the gaseous medium rather than directly pumping the laser state. The energy is subsequently transferred from that storage level to the laser level of the lasing species by direct collisional exchange of energy.

1. Helium-Neon Laser

The helium–neon laser was the first gas laser. The original laser transitions were in the near infrared, but the most commonly used transition is the red laser at a wavelength of 632.8 nm. This laser is available in sizes ranging from approximately 10 cm in length to over 100. It has continuous power outputs ranging from less than 1 mW to over 100 mW and has a lifetime of 50,000 h for some commercial units. The excitation mechanism involves electrons colliding with helium atoms to produce helium metastable atoms, which then transfer their energy to neon laser levels. This laser is used in surveying, construction, supermarket checkout scanners, printers, and many other applications.

2. Argon and Krypton Ion Lasers

Argon and krypton ion lasers were discovered shortly after helium–neon lasers. They were the first lasers to operate in the green and blue regions of the spectrum, and some versions provide ultraviolet output. These lasers have the capability of producing more than 20 W of continuous power for the largest versions. The size of the laser tubes range from 50 to 200 cm in length, with a separate power supply. They are relatively inefficient and consequently require high-input power and water cooling for most units.

The high-power requirements, which put great demands on the strength of the laser discharge region, limit the lifetime of the high-power versions of these lasers. Some smaller, lower power versions of the argon ion laser are air cooled and offer lifetimes of 5000 hr. The excitation mechanisms for these lasers involve electron collisions first populating the ion ground states of the argon and krypton species, with subsequent electron excitation to the upper laser level. Applications include phototherapy of the eye, pumping dye lasers, printing, and lithography.

3. Carbon Dioxide Laser

The carbon dioxide lasers are some of the most powerful lasers, operating primarily in the middle infrared spectral region at a wavelength of 10.6 μm. They range from small versions with a few milliwatts of continuous power to large pulsed versions the size of large buildings producing over 10,000 J of energy. They are among the most efficient lasers (up to 30%) and can produce continuous output powers of over 100 kW in room-size versions. Small versions of these lasers are referred to as waveguide lasers because the excitation region is of a cylindrical shape small enough to guide the beam down the bore in a wave-guide type of mode. They can produce continuous power outputs of up to 100 W from a device smaller than a shoebox (with a separate power supply). The lasers typically operate in a mixture of carbon dioxide, nitrogen, and helium gases. Electron collisions excite metastable (storage) levels in nitrogen molecules, with subsequent transfer of that energy to carbon dioxide laser levels. The helium gas acts to keep the average electron energy high in the gas discharge region. This laser is used for a wide variety of applications including eye and tissue surgery, welding, cutting, and heat treatment of materials, and beam weapons. The white streamers are hot metal particles being ablated from the region of the hole.

4. Rare-Gas-Halide Excimer Lasers

The rare-gas-halide excimer lasers operate primarily in the ultraviolet spectral region in mixtures of rare gases, such as argon, krypton, or xenon with halide molecules such as chlorine and fluorine. They include the argon-fluoride laser at 193 nm, the krypton-fluoride laser at 248 nm, the xenon-chloride laser at 308 nm, and the xenon-fluoride laser at 351 nm. These lasers typically produce short pulses of energy ranging from tens of millijoules to thousands of joules, in pulse durations of 10–50 nsec and repetition rates of up to 1000 pulses per second. They range in size from an enclosure that would fit on a kitchen table top to lasers the size of a very large room. They are relatively efficient (1–5%) and provide useful energy in a wavelength region that has not had powerful lasers available previously. Their operating lifetimes are related to the development of discharge tubes, gas storage chambers, and gas pumps that can tolerate the corrosive halogen molecules that are circulated rapidly through the gain region. The laser species are mixed with helium gas to provide a total pressure of two atmospheres. Excitation occurs by electron dissociation and ionization of the rare gas molecule to produce Ar^+, Kr^+, or Xe^+ ions. These ions then react with the halide molecules, pulling off one of the atoms of that molecule to create an excited-state dimer (abbreviated as excimer) molecule. This excimer molecule then radiates rapidly to an unstable (rapidly dissociating) lower laser level. The very high operating pressures cause the molecules to react rapidly in order to produce the upper laser levels at a rate that can compete with the rapid decay of those levels. Applications include laser surgery, pumping of dye lasers, and lithography.

5. Chemical Lasers

In these lasers the molecules undergo a chemical reaction that leaves them in an excited state that has a population inversion with respect to a lower lying state. An example of this type of laser is the hydrogen–fluoride laser in which molecular hydrogen and molecular fluorine react to produce hydrogen–fluoride molecules in their excited state, resulting in stimulated emission primarily at 2.8 sum. There are no commercially available chemical lasers but they have undergone extensive development for military applications.

B. Metal Vapor Lasers

Metal vapor lasers are actually a type of gaseous laser since the laser action occurs in the atomic or molecular vapor phase of the species at relatively low pressures. However, these lasers have special constraints associated with having to vaporize a solid or liquid into the gaseous state either before or during the excitation and lasing process. These requirements, along with the problems associated with controlling the condensed vapors after they diffuse out of the hot region in some designs, and with minimizing the corrosive effects of hot metal atoms and ions in other designs, have caused them to be classified in a separate category. The two most well-known types of metal vapors lasers are the helium–cadmium ion laser and the pulsed copper vapor laser.

1. Helium–Cadmium Laser

The helium-cadmium laser operates primarily at three wavelengths, in the blue at 441.6 rim and in the ultraviolet

at 325.0 and 353.6 nm. It produces continuous power outputs of the order of 5–200 mW for the blue and 1–100 mW in the ultraviolet, in sizes ranging from 50–200-cm long. The cadmium vapor is obtained by heating the cadmium metal in a reservoir located near the helium discharge. The cadmium vapor then diffuses into the excited helium gas where it is ionized and the cataphoresis force on the cadmium ions causes them to move towards the negative potential of the cathode, thereby distributing the metal relatively uniformly in the discharge region to produce a uniform gain. Typical operating life for the laser is of the order of 5000 hr with the limiting factors being the control of the vapor and the loss of helium gas via diffusion through the glass tube walls. The laser levels are excited by collisions with helium metastable atoms, by electron collisions, and by photoionization resulting from radiating helium atoms. Applications include printing, lithography, and fluorescence analysis.

2. Copper Vapor Laser

This laser operates in the green at 510.5 nm and in the yellow at 578.2 am. It efficiently (2%) produces short laser pulses (10–20-nsec duration) of one mJ of energy at repetition rates of up to 50,000 times per second yielding average powers of up to 50 W. The size is similar to that of an excimer laser. Commercial versions of this laser are designed to heat the metallic copper up to temperatures of the order of 1600°C in order to provide enough copper vapor to produce laser action. The lifetime associated with operating these lasers at such high temperatures has been limited to 800 hr before servicing is required. The excitation mechanism is primarily by electron collisions with ground-state copper atoms to produce upper laser states. Applications of these lasers include uranium isotope enrichment, large-screen optical imaging, micromachining, and pumping of dye lasers. A gold vapor laser that is similar to the copper laser, but emits red light at a wavelength of 624.0 nm, is used for cancer phototherapy.

C. Solid-State Lasers

These lasers generally consist of transparent crystals or glasses as "hosts" within which ionic species of laser atoms are interspersed or "doped." Typical host materials include aluminum oxide (sapphire), garnets, and various forms of glasses, with the most common lasing species being neodymium ions and ruby ions (the first laser). In color-center lasers, the host is typically an alkali–halide crystal and the laser species is an electron trapping defect in the crystal. The energy input in all of these lasers is provided by a light source that is focused into the crystal to excite the upper laser levels. The light source is typically a pulsed or continuously operating flash lamp, but efficient diode lasers are also being used to pump small versions of neodymium lasers, and argon ion lasers are used to pump color-center lasers.

1. Neodymium Lasers

Neodymium atoms are implanted primarily in host materials such as yttrium-aluminum-garnet (YAG) crystals or various forms of glasses, in quantities of approximately one part per hundred. When they are implanted in YAG crystals, the laser emits in the near infrared at 1.06 μm with continuous powers of up to 250 W and with pulsed powers as high as several megawatts. The YAG crystal growth difficulties limit the size of the laser rods to approximately 1 cm in diameter. The YAG host material, however, has the advantage of having a relatively high thermal conductivity to remove wasted heat, thus allowing these crystals to be operated at high repetition rates of the order of up to 1000 pulses per second. Glass hosts also are used for Nd lasers in the l.0-μm wavelength region, but with a somewhat broader bandwidth than YAG. They can also be grown in much larger sizes than YAG, thereby allowing the construction of very large amplifiers. However, glasses have a much lower thermal conductivity, thus requiring operation at much lower repetition rates (of the order of one pulse every few minutes or less). Thus Nd:YAG is used for continuous lasers and relatively low-energy pulsed lasers (one J per pulse), operating at up to 10 pulses per second, whereas glass lasers exist in sizes up to hundreds of centimeters in diameter, occupy large buildings, and are capable of energies as high as 100 kJ per pulse for laser fusion applications. Neodymium lasers typically have very long lifetimes before servicing is required, with the typical failure mode being the replacement of flash lamps. Neodymium lasers are used for surgery applications, for pumping dye lasers (after doubling and tripling their frequencies with nonlinear optical techniques), as military range finders, for drilling holes in solid materials, and for producing X-ray plasmas for X-ray light sources, for laser fusion, and for pumping X-ray lasers.

2. Ruby Laser

The ruby laser, the first laser discovered, is produced by implanting chromium ions into an aluminum oxide crystal host and then irradiating the crystal with a flash lamp to excite the laser levels. Although ruby lasers were frequently used during the early days of the laser, the difficulties associated with growing the crystals, compared with the ease of making neodymium lasers, has led to their being used much less often in recent times.

3. Titanium–Sapphire Laser

The titanium-sapphire laser is the most widely used tunable solid-state laser. The range of possible wavelengths is from 0.660–1.18 μm. It therefore has one of the broadest bandwidths of any laser. It can be operated either cw or pulsed. The broad bandwidth makes it especially suitable for use in mode-locking to generate ultrashort light pulses. It can be either flashlamp pumped or laser pumped, the latter being the most common due to the relatively short upper laser level lifetime, which makes flashlamp pumping less efficient.

4. Color-Center Laser

Color-center lasers use a different form of impurity species implanted in a host material in quantities of 1 part per 10,000. In such lasers the laser species is generally produced by irradiating the crystal with X-rays to produce defects that attract electrons. These defect centers produce energy levels that absorb and emit light and are capable of being inverted to produce gain. Color-center lasers typically operate in the infrared from 0.8–4 μm and are tunable within that range by using different crystals having different emission wavelengths. Their tunability makes them attractive lasers for doing spectroscopy.

D. Semiconductor Lasers

Semiconductor or diode lasers, typically about the size of a grain of salt, are the smallest lasers yet devised. They consist of a *p-n* junction formed in an elongated gain region, typically in a gallium-arsenide crystal, with parallel faces at the ends to serve as partially reflecting mirrors. They operate with milliamps of current at a voltage of only a few volts. The entire laser package is very small and could be incorporated into an integrated circuit board if required. Heterostructure lasers, a more recently developed type of diode laser, include additional layers of different materials of similar electronic configuration, such as aluminum, indium, and phosphorous on the sides of the junction to help confine the electronic current to the junction region in order to minimize current and heat dissipation requirements. Semiconductor lasers range in wavelengh from 0.4–1.8 μm with typical continuous output powers of up to 100 mW. By constructing a row of *p-n* junctions next to each other, all of the separate gain media can be forced to emit together in a phased array to produce an effective combined power output of many Watts. Applications for semiconductor lasers are primarily in the communications field in which the near-infrared beams can be transmitted over long distances through low-loss fibers. In addition, they have recently found a large market as the reading device for compact disc players.

E. Liquid (Dye) Lasers

Dye lasers are similar to solid-state lasers in that they consist of a host material (in this case a solvent such as alcohol) in which the laser (dye) molecules are dissolved at a concentration of the order of 1 part in 10,000. Different dyes have different emission spectra or colors thus allowing dye lasers to cover a broad wavelength range from the ultraviolet (320 nm) to the infrared at about 1500 nm. A unique property of dye lasers is the broad emission spectrum (typically 30–60 nm) over which gain occurs. When this broad gain spectrum is combined with a diffraction grating or a prism as one of the cavity mirrors, the dye laser output can be a very narrow frequency beam (10 GHz or smaller) tunable over a frequency range of 10^{13} Hz. Frequency tuning over even larger ranges is accomplished by inserting different dyes into the laser cavity.

Dye lasers are available in either pulsed (up to 50–100 mJ) or continuous output (up to a few Watts) in tabletop systems that are pumped by either flash lamps or by other lasers such as frequency-doubled or tripled YAG lasers or argon ion lasers. Most dye lasers are arranged to have the dye and its solvent circulated by a pump into the gain region from a much larger reservoir, since the dye degrades slightly during the excitation process. Dyes typically last for 3 to 6 months in systems where they are circulated. Dye lasers are used mostly for applications where tunability of the laser frequency is required, either for selecting a specific frequency that is not available from one of the solid-state or gas lasers, or for studying the properties of a material when the laser frequency is varied over a wide range. Most of these applications are in the area of scientific experiments. Another large application of dye lasers is for producing ultrashort optical pulses by a technique known as mode locking. In this process, all of the longitudinal modes of a dye laser (as many as 10,000) are made to oscillate together (in phase), causing individual pulses as short as 50 fsec (5×10^{14} sec) to emerge from the laser, spaced at intervals of the order of 20 nsec. These short pulses are of interest in studying very fast processes in solids and liquids and may have applications for optical communications.

F. Free-Electron Lasers

Free-electron lasers are significantly different from any other type of laser in that the laser output does not occur from discrete transitions in atoms or molecules of gases, liquids, or solids. Instead, a high-energy (of the order of one million electron volts) beam of electrons is directed to pass through a spatially varying magnetic field. This field causes the electrons to oscillate back and forth in a direction transverse to their beam direction, at a frequency related to the magnet separation and spacing and also to

the energy of the electron beam. This oscillation causes the electrons to radiate at the oscillation frequency and to stimulate other electrons to oscillate and thereby radiate at the same frequency, in phase with the original oscillating electrons, thereby producing an intense beam of light emerging from the end of the device. Mirrors can be placed at the ends of the magnet region to feed the optical beam back through the amplifier to stimulate more radiation and cause the beam to grow. The free-electron laser, although still more of a laboratory curiosity than a useful device, can produce very high-power output over a wide range of wavelengths from the far infrared to the vacuum ultraviolet.

VII. LASER PROPERTIES ASSOCIATED WITH APPLICATIONS

Laser applications are now so varied and widespread that it is difficult to describe the field as a single subject. Topics such as surgery, welding, surveying, communications, printing, pollution detection, isotope enrichment, heat treatment of metals, eye surgery and reshaping, drilling, and laser art involve many different disciplines. Before these individual topics are reviewed, the various characteristics or features of lasers that make them so versatile will be summarized. These properties include beam quality, focusing capabilities, laser energy, wavelength properties, mode quality, brightness properties, and pulse duration.

A. Beam Quality

When a laser beam emerges from a laser cavity, after having evolved while reflecting back and forth between carefully aligned mirrors, it is highly directional, which means that all of the rays are nearly parallel to each other. This directional property, if considered as originating from a single longitudinal mode emerging from a laser cavity, would have such a low-beam divergence that a properly designed laser aimed at the moon would produce a spot on the moon only 5 miles in diameter after having traveled a distance of 239,000 miles! That is a very small divergence when compared to a flashlight beam, which expands to five times its original size while traveling across a room.

B. Focusing Capabilities

Because of the parallel nature of the beam when it passes through a lens (of good quality) all of its rays are concentrated to nearly the same point at the focus of the lens. Since diffraction properties of the light must be taken into account, the beam cannot be focused to an infinitely small point but instead to a spot of a dimension comparable to the wavelength of the light. Thus a green light beam could be focused to a spot 0.00005 cm in diameter, a size significantly smaller than that produced by focusing other light sources.

C. Laser Energy

The primary limitations on laser energy and power are the restrictions on the size of the amplifier and the damage thresholds when the beam arrives at the laser mirrors and windows. A 1-W continuous carbon dioxide laser beam, focused on a fire brick, will cause the brick to glow white hot and begin to disintegrate. A carbon dioxide laser has been made to continuously operate at a level 1,000,000 times that powerful! Similarly, a pulsed laser of 0.0005 J, when focused into the eye, can cause severe retinal damage. A pulsed neodymium laser has been made to produces a billion times that energy!

D. Wavelength Properties

The capability of generating a very pure wavelength or frequency leads to many uses. It allows specific chemical reactions to be activated, wavelength-sensitive reflective and transmissive properties of materials to be exploited, and certain atomic processes to be preferentially selected. Some of these processes can use the fixed frequencies of gas and solid-state lasers. Others require the broad tunability of dye and solid state lasers.

E. Mode Quality

The ultra-pure frequency available from a single-mode output of a laser not only makes possible the concentrated focusing capabilities mentioned, but it provides a very pure frequency that is capable of being modulated in a controlled way to carry information. Since the amount of information that is carried over an electromagnetic (light or radio) wave is proportional to the frequency of the radiation, a single-mode optical wave can carry over 10,000 times the information of a microwave and a billion times that of a radio wave. This transmission capacity is possible whether the optical wave is carried through space, through the atmosphere, or through a tiny glass fiber the size of a human hair.

F. Brightness Properties

The brightness properties are represented by the amount of light concentrated in a specific area during a definite amount of time in a specific wavelength or frequency interval. Such properties are applicable for use in holography and other coherent processes. The effective brightness temperature of a laser can be 102° to 103°C, much higher than any other human-made light source.

G. Pulse Duration

The duration of a laser beam can vary anywhere from a continuous (cw) beam that will last as long as the power is supplied to the amplifier and as long as the amplifier keeps producing gain (50,000 hr for some gas lasers and potentially many tens of years for some semiconductor lasers) to pulses as short as 5×10^{-15} sec for the shortest pulses from a pulse-compressed, mode-locked dye laser. Many applications require very short pulses, including high-speed digital signal transmission for communications and also surgical applications that would result in damage to surrounding tissue via heat conduction if longer pulses were used. Other applications require very long or continuous light fluxes to produce effects in materials that are relatively insensitive to the light beam but would be destroyed by shorter duration, higher-intensity beams.

SEE ALSO THE FOLLOWING ARTICLES

LASERS, CHEMICAL • LASERS, COLOR CENTER • LASERS, DYE • LASERS, FREE ELECTRON • LASERS, GAS • LASERS, NUCLEAR PUMPED • LASERS, OPTICAL FIBER • LASERS, SEMICONDUCTOR • LASERS, SOLID STATE • LASERS, ULTRAFAST PULSE TECHNOLOGY • LASERS, X-RAY

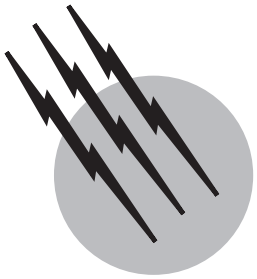
BIBLIOGRAPHY

Siegman, A. E., "Lasers," University Science Books, Mill Valley, CA, 1986.

Silfvast, W. T., "Lasers," McGraw-Hill "Handbook of Optics," Second Edition, Chapter 11, Volume I, New York, 1995

Silfvast, W. T., "Laser Fundamentals," Cambridge University Press, NY 1996

Steen, W. M., "Laser Materials Processing," Springer-Verlag, New York, 1991 (3rd printing 1996).



Lasers, Chemical

Glen P. Perram
Air Force Institute of Technology

I. Laser Fundamentals
II. The HF/DF Chemical Laser
III. The Chemical Oxygen–Iodine Laser
IV. Advanced Chemical Lasers

GLOSSARY

Beam quality Measure of a excellence of a laser beam, usually defined as the ratio of the actual spot size to the diffraction-limited spot size.
Chemical laser Laser operating on a population inversion produced (directly or indirectly) in the course of an exothermic reaction.
Gain medium The material in the laser system where the light is amplified.
Mass efficiency Laser power achieved per reagent flow rate.
Nozzle flux Laser power achieved per unit cross-sectional area of the nozzle assembly.
Optical resonator Optical extraction system (mirrors) which provides feedback of light to the gain medium.
Stimulated emission Fundamental process by which light can interact with matter to amplify the intensity of the electromagnetic field.

CHEMICAL LASERS are a class of laser device that transform the energy stored in chemical bonds into a nearly monochromatic beam of coherent electromagnetic radiation, or light. These chemical lasers are often large, high-power devices that integrate chemical delivery systems, a supersonic flow reactor, and an optical resonator. The chemical laser has been defined as a laser operating on a population inversion produced (directly or indirectly) in the course of an exothermic reaction.

Chemical lasers are normally divided into two classes: (1) continuous wave (cw) flowing systems and (2) various types of pulsed devices. These can be further subdivided into direct chemical pumping and energy transfer lasers, depending on how the lasing species receives its energy. The laser systems of greatest potential for scaling to high average powers are the cw class. We will therefore direct this discussion to cw systems offering the most promise, including the direct chemically pumped hydrogen (or deuterium) fluoride laser (HF/DF) and the energy-transfer-pumped chemical oxygen–iodine laser (COIL).

I. LASER FUNDAMENTALS

In 1917, Albert Einstein discovered the fundamental principal by which light can interact with matter to amplify the intensity of the electromagnetic field, termed stimulated emission. Indeed, the word *laser* is an acronym for light amplification by the stimulated emission of radiation.

The light from a laser differs in several significant ways from the light from an ordinary incandescent bulb.The electromagnetic waves emitted by an ordinary source have many different wavelengths, or colors, and the soure often appears "white." Furthermore, these waves are uncoordinated in that their crests and valleys exhibit no well-defined pattern with respect to each other. Finally, emission from individual points from within the ordinary source propagates in all directions. In contrast, laser light possesses a well-defined color (monochromatic), the waves are coordinated with crests and valleys in phase with each other (coherent), and the light propagates in a well-defined direction (as a beam).

The light emitted from an ordinary source is generated by spontaneous emission. If an atom or molecule is in an excited state (i.e., has excess internal energy E_2), one of the ways it can decay to a state of lower energy E_1 is by the emission of light (a photon). The frequency of the emitted light ν is related to the energy difference $E_2 - E_1$, by Planck's constant, $h = 6.6256 \times 10^{-34}$ J-sec:

$$\nu = (E_2 - E_1)/h. \tag{1}$$

The frequency and wavelength λ of light propagating in a medium with index of refraction n are related by the speed of light, $c = 2.9979 \times 10^8$ m/sec:

$$\lambda\nu = c/n. \tag{2}$$

Spontaneous emission occurs without external influence and is independent of other emission from within the sample.

Alternatively, laser light is produced when emission is stimulated by an incident photon. In this stimulated emission process, the newly emitted photon possesses the same energy, direction of propagation, and phase as the incident photon. An exact copy of the incident photon is generated and the intensity of the light is amplified as the wave continues to propagate through the medium. Both stimulated emission and absorption will occur between a pair of energy levels, yielding an intensity variation $I(z)$ as the wave propagates a distance z:

$$I(z) = I(z = 0)e^{\sigma(N_2 - N_1)z}, \tag{3}$$

where N_1 and N_2 are the number of atoms or molecules per unit volume with energy E_1 and E_2, respectively. The laser gain γ is specified by the product of the cross section for stimulated emission σ and the population inversion $\Delta = (N_2 - N_1)$:

$$\gamma = \sigma\Delta. \tag{4}$$

To achieve net amplification, $\gamma > 0$, and thus lasing, a population inversion, $N_2 > N_1$, is required. Normally, there are more atoms and molecules in the lower energy state and the rate of absorption is greater than the rate of stimulated emission. However, if there are more atoms or molecules in the state of higher energy, then stimulated emission will dominate and amplification of the incident light can be achieved. Thus, a population inversion between the two energy levels is required to produce a laser. This is a situation not normally found in nature. The laser medium must be "pumped" by an external energy source to selectively deposit energy in the upper laser level, E_2. A chemical laser achieves the population inversion from the products of a chemical reaction.

When the gain exceeds the total optical losses α due to mirror output coupling, scattering losses, and resonator losses, then net amplification and lasing is achieved. The threshold population inversion is defined as the inversion required for the gain to equal the losses:

$$\Delta_{\text{th}} = \gamma_{\text{th}}/\sigma = \alpha/\sigma. \tag{5}$$

The HF laser exhibits high gain, which enables the design of optical resonators with large output coupling. Low-gain lasers, such as the COIL device, suffer in efficiency when operated near threshold, but allow for longer gain media and greater geometrical scaling for high-power performance.

In was not until 1960 that T. H. Maiman used these ideas and demonstrated the first laser device, the ruby laser. A schematic of a typical laser is shown in Fig. 1, highlighting the three main components: (1) a source of energy to drive the system far from thermodynamic equilibrium, (2) a gain medium, where a population inversion exists and amplification occurs, and (3) the optical resonator, usually a set of mirrors that provides multiple passes of the light through the gain medium to establish an intense beam of radiation. The methods of pumping energy into the laser medium include optical electrical, gas dynamic, and, as discussed in this article, chemical means. The development of the laser represents a significant scientific and technical achievement. In 1964, the Nobel Prize in Physics was awarded to C. H. Townes, N. G. Basov, and A. M. Prokhorov for their fundamental work in quantum electronics, in particular leading to the construction of the laser.

II. THE HF/DF CHEMICAL LASER

A. Chemistry

J. C. Polanyi first proposed a chemical laser in 1960 by recognizing that many chemical reactions deposit excess energy in the stretching vibrations of the newly formed bond. If the chemical reaction favors the production of states with greater internal energy over states with less internal energy, then a population inversion can be achieved. A prime example is the reaction of atomic fluorine with molecular hydrogen to produce vibrationally excited hydrogen fluoride:

$$F + H_2 \rightarrow HF(v) + H. \tag{6}$$

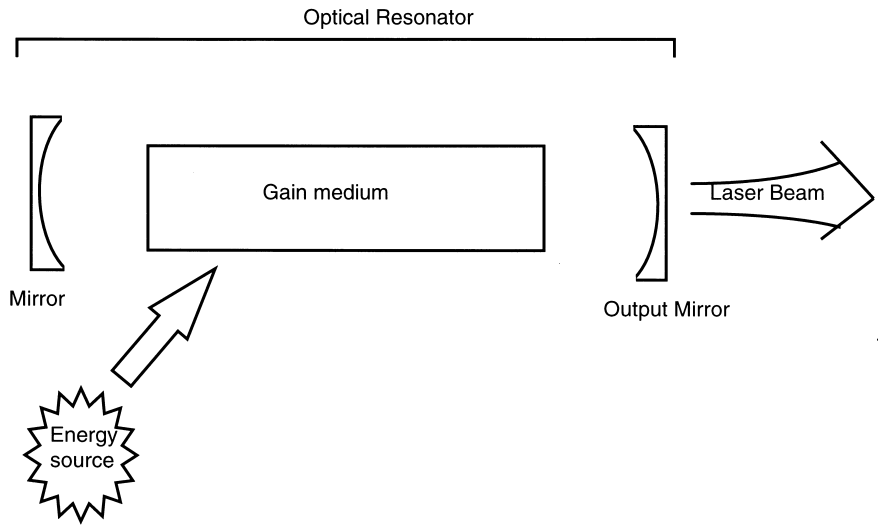


FIGURE 1 The major components of a laser include a gain medium where the light is amplified, an external energy source to produce the population inversion, and an optical resonator or laser cavity which provides feedback to the gain medium and establishes a well-defined laser beam.

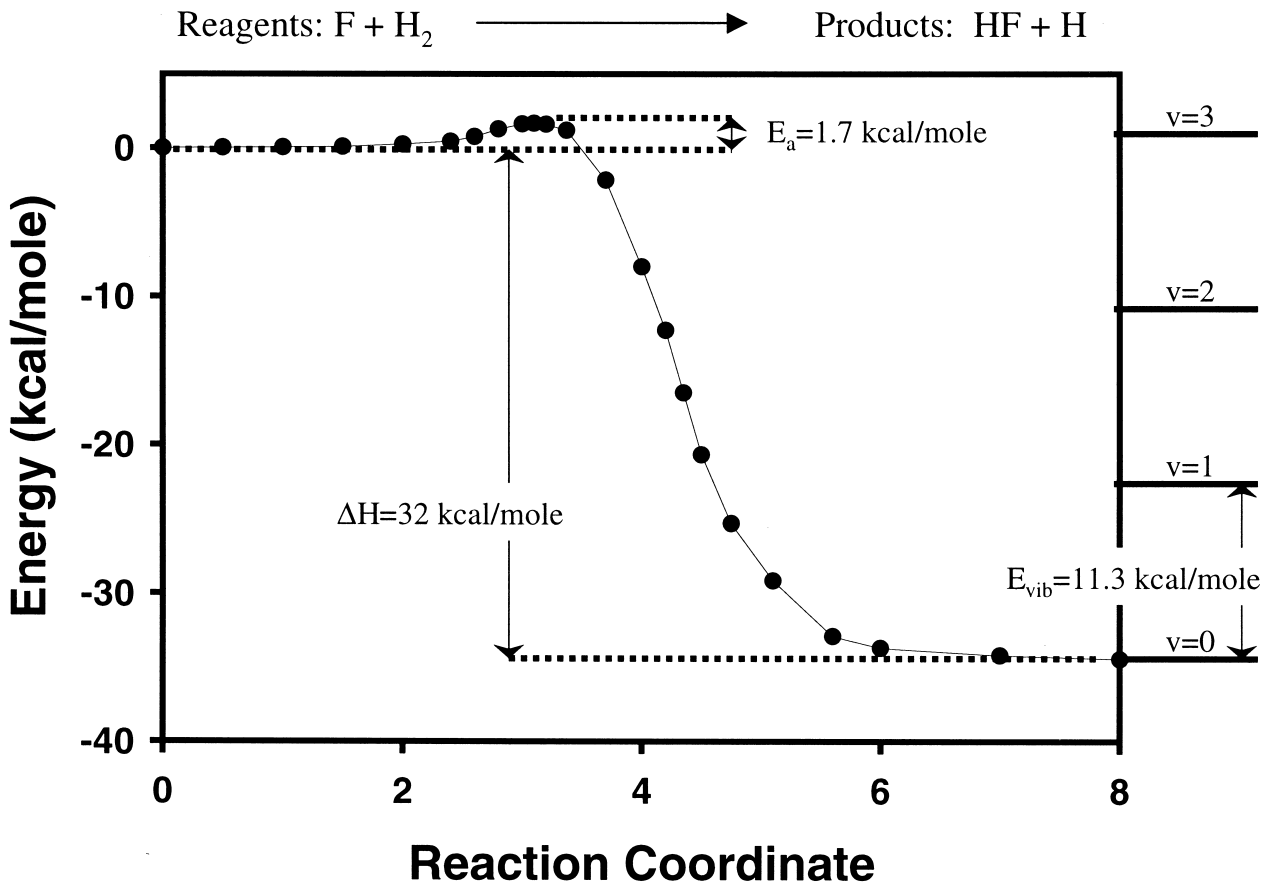


FIGURE 2 The reaction of atomic fluorine with molecular hydrogen produces vibrationally excited hydrogen fluoride. The reaction coordinate represents the breaking of the molecular hydrogen bond as the fluorine atom approaches and the subsequent formation of the HF bond as the hydrogen atom departs the collision center.

The energetics of this reaction is shown in Fig. 2. As the reaction proceeds (from left to right in Fig. 2), a small barrier of $E_a = 1.7$ kcal/mole is encountered, providing a rapid rate of reaction even at moderate temperatures. This reaction is 32 kcal/mole exothermic and the excess energy can be deposited into vibration of the newly formed HF bond. The energies for the first four vibrational levels of the HF molecule, $v = 0$–3, are shown on the right-hand side of Fig. 2 with a spacing of about 11.3 kcal/mole. The key to developing an HF laser based on this reaction is the relative rates for producing the various vibrational levels of HF. These relative rates have been measured as 12:20:6:1 for $v = 3{:}2{:}1{:}0$. Thus, 31% of the HF products will be found in the vibrational level $v = 3$. The most probable product is $v = 2$ at 51%. Only 3% of the reactions yield the lowest vibrational state, $v = 0$. Clearly a population inversion between vibrational levels in HF can be generated by this direct chemical reaction.

The first chemical laser was demonstrated in 1965 by Kasper and Pimentel by initiating a hydrogen–chlorine explosion with a flashlamp. By 1967, lasers based on the reaction of atomic fluorine with molecular hydrogen and other molecules had been developed. Several key experiments were performed in 1969, and lasing was achieved in an HF supersonic diffusion flame by Spencer and coworkers. Subsequent development of the HF chemical laser was rapid and by 1984 HF lasers with powers exceeding 1 MW had been developed.

The HF laser operates in the infrared region of the electromagnetic spectrum with a wavelength near 2.7 μm. The output power as a function of laser wavelength is shown in the simulation of Fig. 3. The HF laser can operate at several discrete wavelengths corresponding to different vibrational and rotational transitions. These transitions are labeled according to their vibrational (v) and rotational (J) quantum numbers. In Fig. 3, the $(2 \rightarrow 1)$ $P(3)$ transition is shown to involve lasing from the excited level $v = 2$, $J = 2$ to the lower level $v = 1$, $J = 3$. The earth's atmosphere is nearly opaque at the HF laser wavelengths, as seen in Fig. 4. However, by substituting deuterium for hydrogen in reaction (6), the DF laser can be made to lase at $\lambda = 3.6$–4.2 μm. The atmosphere exhibits a window with good transmission in this range, making the DF laser preferred for applications requiring atmospheric propagation.

B. Hardware

Continuous-wave HF/DF lasers are designed with elements best illustrated by considering a simple combustion-driven supersonic diffusion laser as an example. In this case, no external power supply is required; the laser operates purely on chemical energy. Figure 5 presents the main elements that make up this device. The laser combustor provides the appropriate thermodynamic conditions for the dissociation of the oxidizer (fluorine or fluorinated oxidizer). Diluent (helium) mixed with combustor fuel (deuterium, D_2) is introduced into the combustor through a conventional gas–gas injector. Additional diluent plus an excess of molecular fluorine, F_2, enter through impinging jets, providing the complete mixture of combustion gases, which undergo autoignition and rapidly achieve the desired equilibrium conditions. Combustion stagnation temperatures vary from 1500 to 2500 K, and stagnation pressures are in the range 30–300 psi in the subsonic flowing section. This short combustion zone has walls cooled with excess fuel or liquid coolant. The conditioned combustor gas containing the excess dissociated fluorine subsequently flows through an array of highly cooled small nozzles. These serve as the fine supersonic–supersonic mixing arrays introducing the oxidizer stream into the optical cavity. These supersonic mixing nozzles are designed for maximum efficiency at particular operating conditions that produce flow at a temperature and pressure substantially lower than that in the combustor (optical cavity static temperatures are 200–400 K and pressures are 2–15 torr). Another function of the nozzle is to "freeze" chemically the F-atom mole fraction at or near its combustor value. The supersonic mixing nozzles thus establish appropriate pressure, temperature, and composition for all the important chemical reactions to proceed rapidly in the laser cavity. The cavity mixing and chemical excitation reaction stages take advantage of the large cavity inlet velocity provided by the supersonic nozzles, which act to extend the lasing-zone length and fill the optical cavity.

The chemical reaction producing excited HF is severely limited by diffusion, and fast mixing is the key to high power and efficiency. Short, minimum-length nozzles are required to expand and accelerate the gas flow to provide supersonic mixing of the oxidizer–diluent mixture with the secondary fuel jet. These nozzles are decidedly different from wind-tunnel nozzles or rocket nozzles. The main object of a rocket nozzle is to produce a gas flow that is in equilibrium and in which all the energy introduced by the heater is recovered as kinetic energy and not frozen in the internal energy modes of the molecules. The high velocity of the supersonic jet is also an important feature of the chemical laser. This is because the very rapid deactivation of vibrationally excited HF and therefore the very short effective lifetime of the lasing species result in short inversion zones even at supersonic velocities.

A wide variety of laser optical systems are employed to extract a beam from the cavity. Typical of these is the unstable resonator in Fig. 5. In this system, there is a central region near the optical axis (of such a radius that its Fresnel number is near unity) within which all the laser radiation is closely coupled by diffraction. Therefore, the

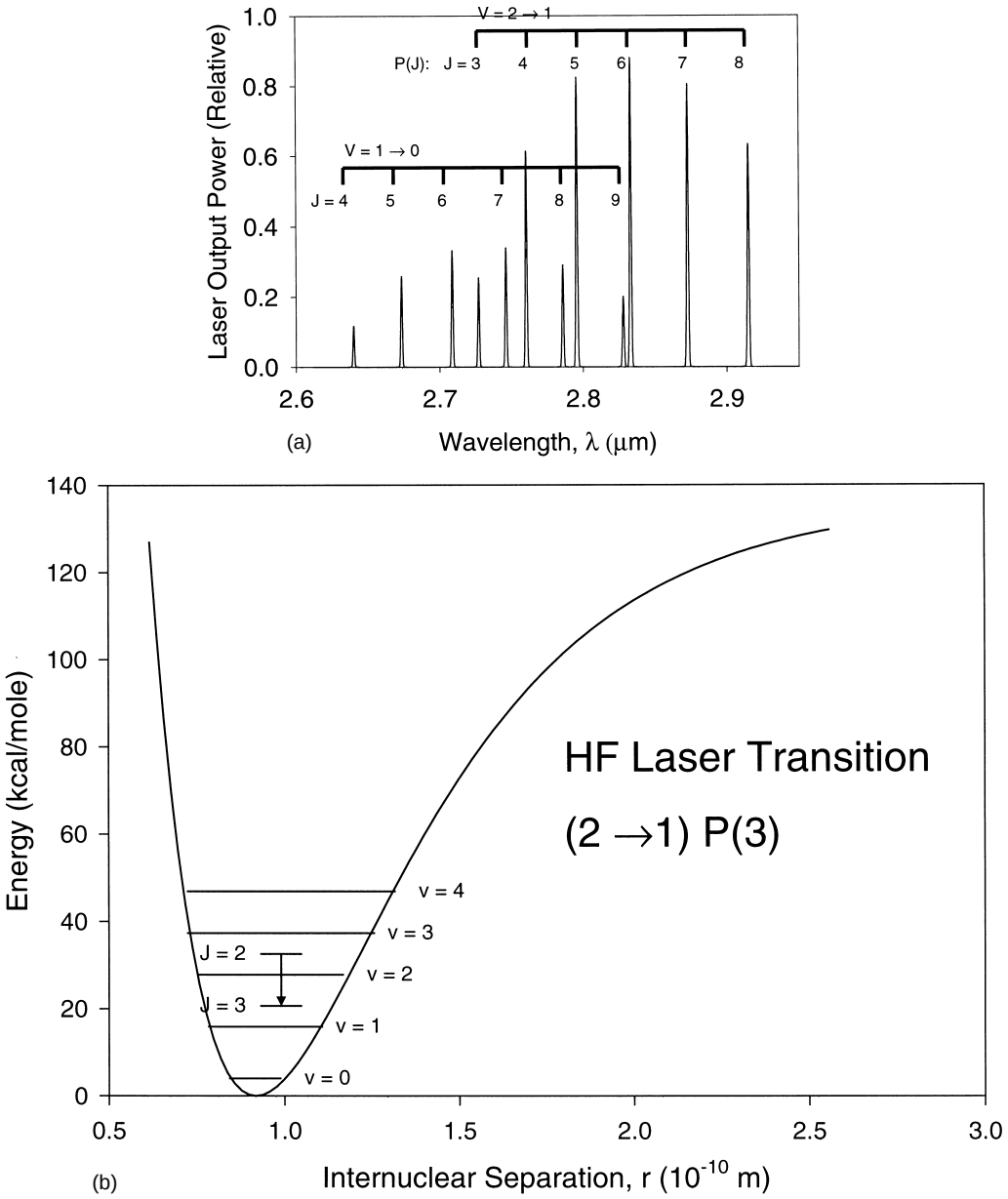


FIGURE 3 (a) Simulated spectrum of an HF laser showing several discrete laser wavelengths. (b) The transitions between ro-vibrational levels are labeled according to spectroscopic notation.

laser emission produced by this central region is coherent. The convex mirrors magnify the beam continually as it propagates through the optical cavity, and this central or uniphase mode spreads out to fill the entire laser medium. Eventually the radiation overflows around one mirror to form the laser output shown exiting the beam extraction mirror (scrapper mirror). One way to think of this unstable resonator is as a low-Fresnel-number oscillator region surrounded by a multipass amplifier. Since phase control of the output beam occurs within the central portion of the resonator, a high-quality output beam is usually obtained.

The output beam has the form of an annulus with a zero-intensity central region produced by the part of the beam that is blocked by the extraction mirror. Such an output beam, when passed through a conventional telescope designed to accommodate an unobscured beam, has an annular intensity distribution which departs from the ideal uniform intensity (such would give the minimum far-field beam width). In high-power lasers, it is frequently found that the annular beam departs so little in beam quality from the uniform beam that other factors, such as gain medium homogeneity, are more important than uniform intensity.

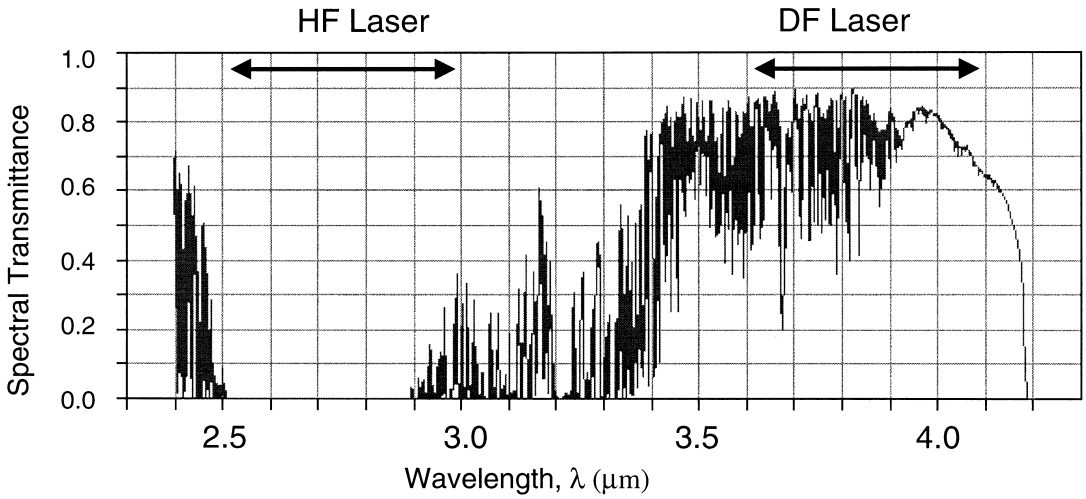


FIGURE 4 The atmosphere is opaque to the HF laser. However, the longer wavelength DF laser operates in an atmospheric window.

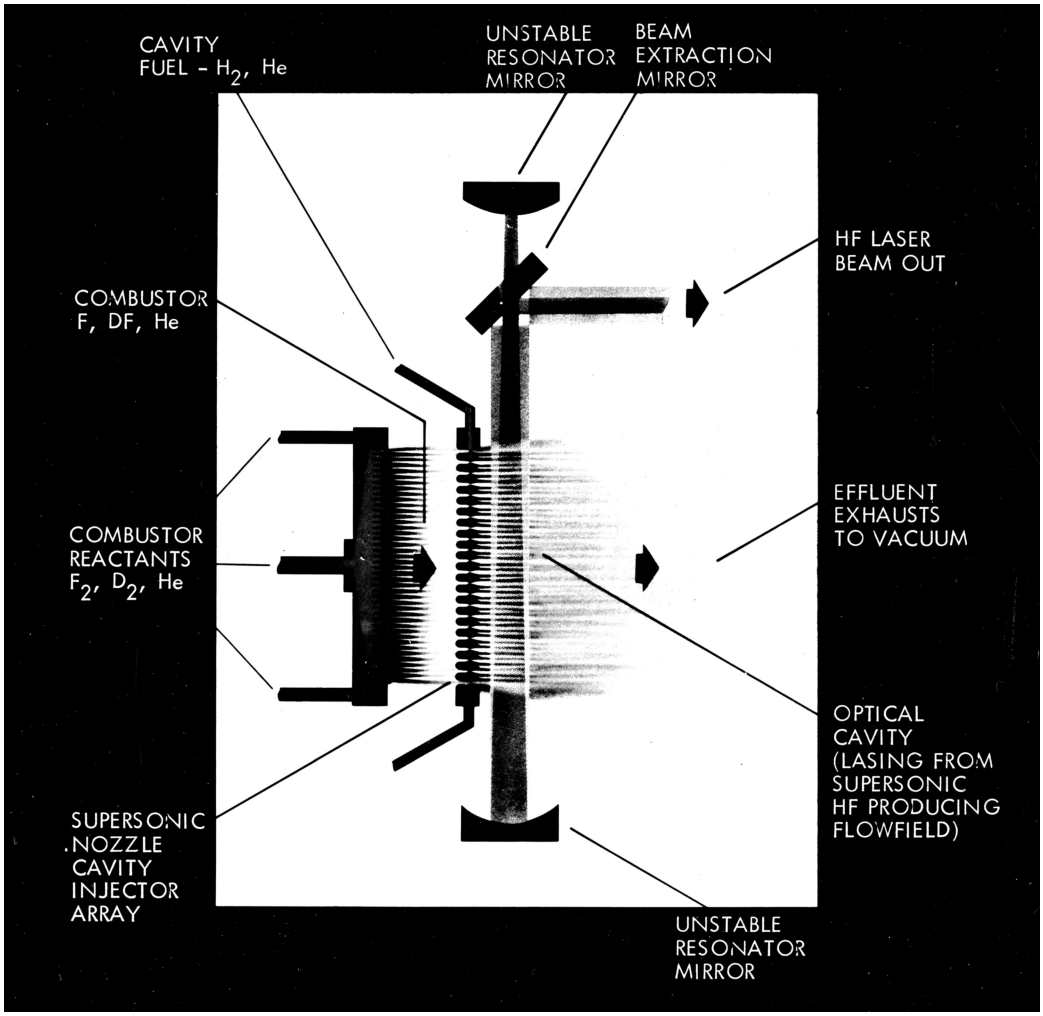


FIGURE 5 Schematic of a cw HF chemical laser. (Courtesy of TRW, Inc.)

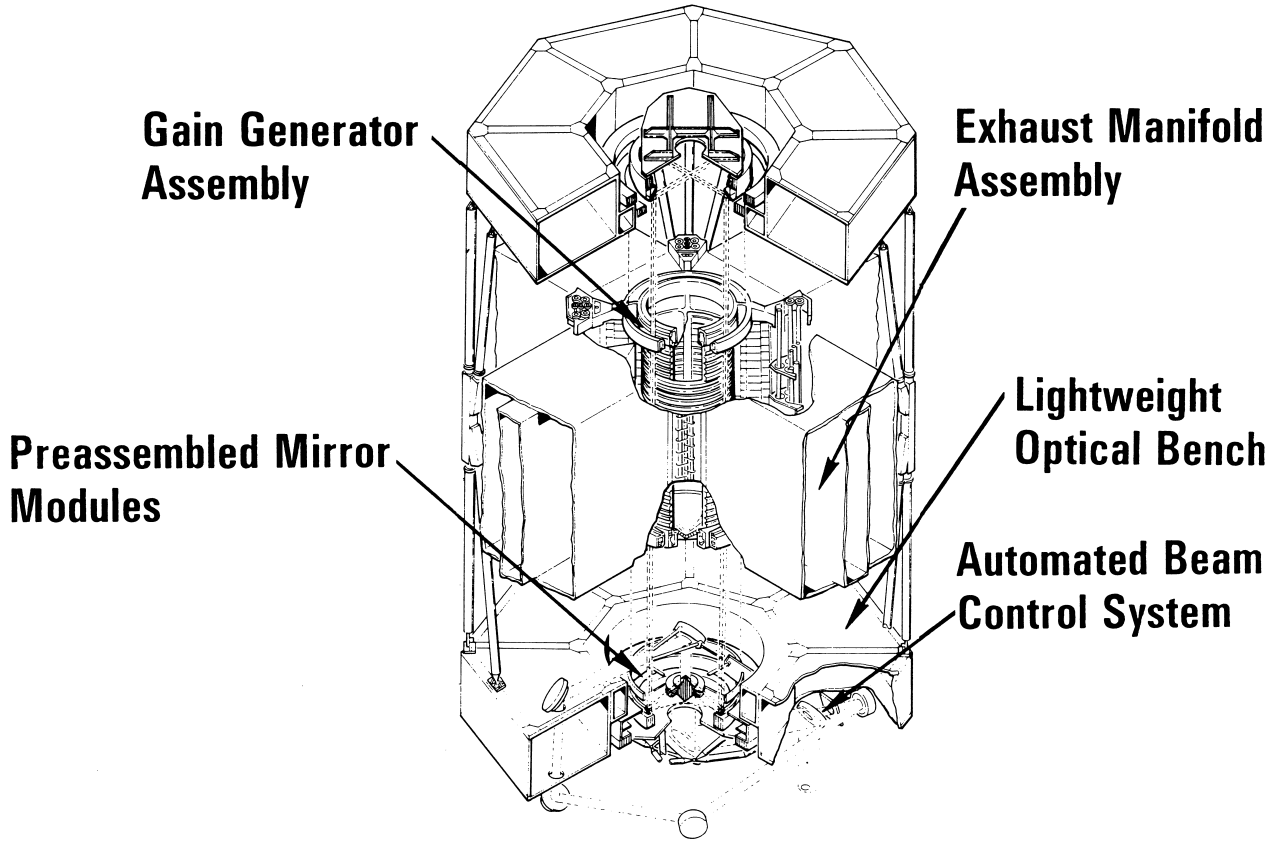


FIGURE 6 Alpha HF chemical laser. (Courtesy of J. Miller, TRW, Inc.)

In the final stage of any flowing chemical laser, the processing of the exhaust effluent must be considered. This is usually accomplished by a smooth transition of the cavity shroud into a diffuser designed to raise the pressure and accommodate the low entrance Reynolds numbers. Diffuser efficiency can be considered an integral part of the overall design of the laser system. A common normal-shock diffuser model is usually adequate to optimize the performance of this section. Pumping of this effluent is subsequently accomplished after recovery of the available pressure within the system.

C. High-Power Devices

Megawatt-class power levels were first achieved by the mid-infrared advanced chemical laser (MIRACL) in the early 1980s. The MIRACL device is a cw DF chemical laser with a linear nozzle configuration which produced a square 14 cm $\times$ 14 cm beam with outcoupled power in excess of 1 MW. A modest beam quality of two times the diffraction limit was achieved. The limit to scaling the gain length using a linear geometry is nearly reached in this device.

The Alpha cylindrical HF chemical laser is also a megawatt-class laser which wraps arrays of nozzle modules in the form of a large cylinder. The fluorine and diluent flow passes radially out around the annulus that holds the hypersonic hydrogen injectors, as shown in Fig. 6. Curvature in the radial direction on the annular mirror allows mode control and intensity levels to be controlled in the compacted leg. The cylindrical configuration is preferred for very high powers over that of conventional linear laser configurations. This provides many times the device output power without increasing the nozzle power flux, cavity injector length, or mode width. Such a resonator reduces effects from flow-field disturbances, diffraction losses, anomalous dispersion, and other optical loss mechanisms that tend to produce a poor-quality beam. In 1991, the Alpha laser achieved megawatt power using operating conditions consistent with the low-pressure environment required for space-based operation.

The tactical high-energy laser (THEL) uses a high-power chemical laser for protection of theater-level combat forces and marks the transition of high-power lasers to the battlefield. The THEL is designed to be a transportable, tactical-sized, DF laser system including the laser device,

TABLE I Chemical Laser Performance Parameters

Laser parameter	Symbol	Typical value	
		HF laser	COIL
Wavelength (μm)	λ	2.7	1.3
Gain (%/cm)	γ	20	1
Mass efficiency (kJ/kg)	σ_m	150	300
Nozzle flux (kW/cm^2)	δ	—	200
Beam quality	BQ	2	1.3
Saturation intensity (kW/cm^2)	I_{sat}	0.1	5

pointing and tracking system, beam director, and command and control system. In June 2000, the THEL successfully intercepted and destroyed a Katyusha rocket at the High Energy Laser Systems Test Facility (HELSTF), White Sands Missile Range, New Mexico.

The United States is currently pursuing the development of a space-based chemical laser (SBL). The Integrated Flight Experiment (IFX) is designed to validate emerging technologies, demonstrate systems integration, and develop the engineering designs required to develop an operational, treaty-compliant SBL system. A megawatt-class chemical laser in space should culminate in the lethal intercept of a boosting ballistic missile.

Several key parameters describing the general performance of chemical lasers are provided in Table I. The mass efficiency σ_m (kJ/kg) describes the laser power achieved per reagent flow rate and is particularly important for space-based applications, where the cost of delivering the fuel to orbit can dominant total system costs. Typical mass efficiencies for HF lasers are 150–300 kJ/kg. By comparison, dynamite (TNT) possesses an energy density of 9 MJ/kg. The nozzle flux parameter represents the laser power achieved per unit cross-sectional area of the nozzle assembly. This is a critical parameter for power scaling and specifies the size of a high-power chemical laser.

Beam quality is the parameter that describes the ability to propagate the laser as a narrow beam and to focus the beam to a small spot. The angular divergence of a laser beam $\theta_{1/2}$ is ultimately limited by diffraction,

$$\theta_{1/2} \sim (\lambda/D), \tag{7}$$

where D is the diameter of the limiting aperture or cavity mirror. However, such a diffraction-limited beam is not achieved, and the beam quality (BQ) is often reported as the ratio of the actual spot size to the diffraction-limited spot size. Chemical lasers have very good beam quality due to the transport of excess heat from the laser cavity via supersonic flow. Beam quality for a chemical laser is usually limited by the spatial uniformity of the gain media imposed by the supersonic mixing nozzles.

The ability of a laser to damage a distant target is often characterized by the source brightness, or power per unit solid angle of beam divergence. The brightness B for a diffraction-limited beam is approximately related to laser power P, operating wavelength λ, and diameter of the laser aperture D:

$$B = PD^2/\pi\lambda^2. \tag{8}$$

It is the power flux, or power per unit area, that causes damage due to rapid heating, and the area to which a laser beam can be focused is limited by diffraction, beam quality, and beam jitter. Device powers of multimegawatts and brightnesses of 10^{20}–10^{23} W/steradian are required for some military missions.

III. THE CHEMICAL OXYGEN–IODINE LASER

The chemical oxygen–iodine laser (COIL) was first demonstrated at the Air Force Weapons Laboratory in 1977. The COIL is the shortest wavelength, high-power chemical laser and the first chemical laser to operate on an electronic rather than vibrational or rotational transition.

A. Chemistry

To produce an efficient inversion between electronic states requires the chemical reaction selectively to produce a specific excited electronic state over all other states, particularly the ground electronic state. Reactions involving light atoms and molecules tend to preserve the total electronic spin. Thus, the generation of an electronic inversion is aided by a difference in electronic spin between the ground and excited electronic states. However, such transitions exhibit slow spontaneous emission rates, and thus very low stimulated emission cross sections and insufficient gain to build a laser. For example, the $a^1\Delta$–$X^3\Sigma$ transition in O_2 exhibits a particularly slow spontaneous emission rate of 2.58×10^{-4} sec^{-1}. However, these long-lived, or metastable, electronically excited states establish a large energy reservoir that could be used to pump a chemical laser. By collisionally transferring the energy stored in these metastable states to a suitable lasant species, a new class of chemical laser, the metastable energy transfer chemical laser, can be developed. The chemical oxygen–iodine laser is the prime example of such an energy transfer laser.

The first electronically excited state of molecular oxygen, $O_2(a^1\Delta)$, is produced by a two-phase reaction of chlorine gas and liquid basic hydrogen peroxide with near 100% efficiency:

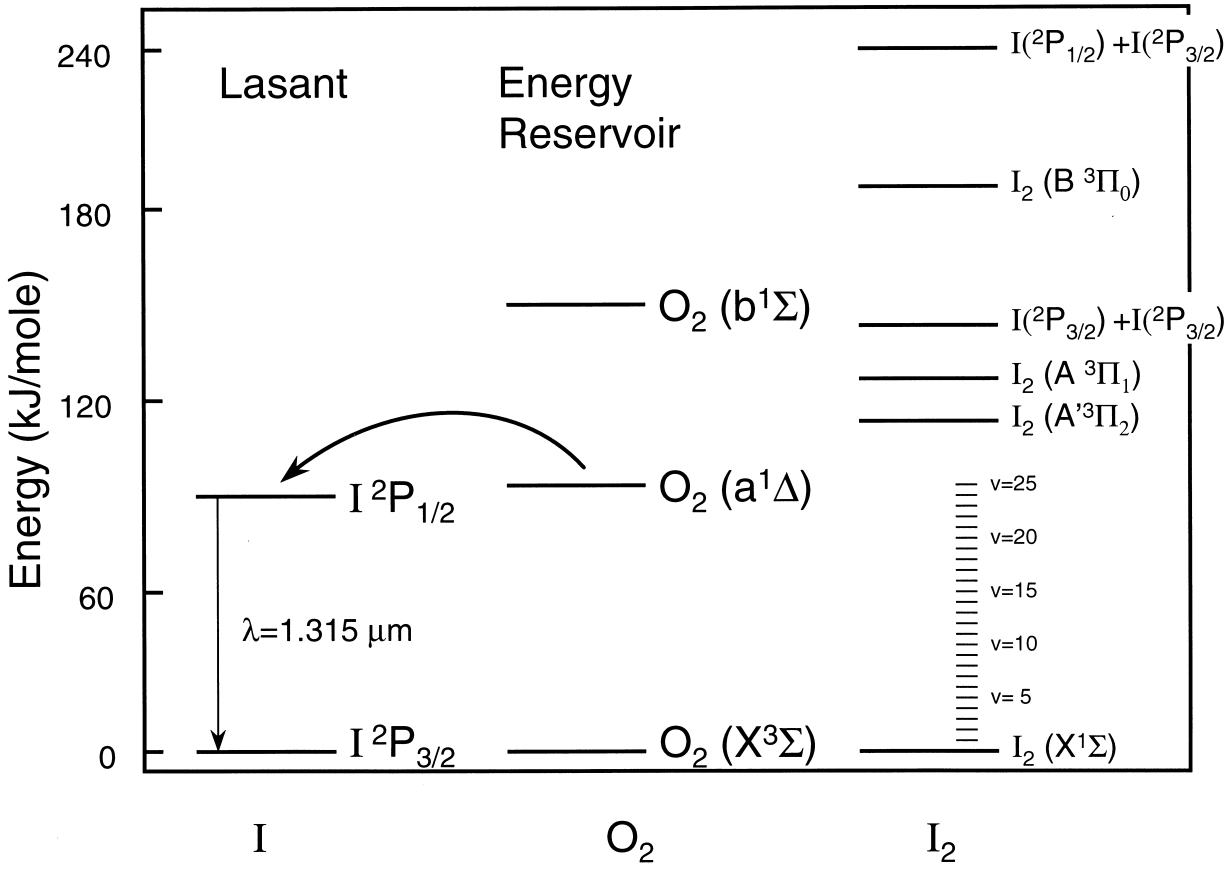


FIGURE 7 Energy-level diagram for the COIL illustrating the resonant energy transfer from the metastable energy reservoir $O_2(a^1\Delta)$ to the upper laser level of atomic iodine, $I(^2P_{1/2})$. The energy levels for several key states involved in the dissociation of molecular iodine are also provided.

$$Cl_2(g) + H_2O_2(l) + 2KOH(l) \rightarrow O_2(a^1\Delta) + 2KCl + 2H_2O. \quad (9)$$

Once produced, this singlet state of oxygen can be transported for considerable distances even at moderately high pressures, as the radiative lifetime is ~64 min and $O_2(a^1\Delta)$ is very resistant to collisional deactivation. Thus, $O_2(a^1\Delta)$ is a metastable energy reservoir. Energy densities of 10–20 kJ/m^3 are typically attained.

The chemical oxygen–iodine laser operates on an inversion between the $5^2P_{1/2}$ and $5^2P_{3/2}$ spin–orbit split states of atomic iodine at a wavelength of 1.315 μm. A near resonance exists between $O_2(a^1\Delta)$ and $I(^2P_{1/2})$ with an energy difference of only 3.3 kJ/mole, and laser pumping is achieved by collisional energy transfer from the $O_2(a^1\Delta)$ metastable energy reservoir, as shown in Fig. 7. The energy transfer is rapid and a near equilibrium between the upper laser level $I(^2P_{1/2})$ and $O_2(a^1\Delta)$ is quickly established:

$$O_2(a^1\Delta) + I(^2P_{3/2}) \leftrightarrow O_2(X^3\Sigma) + I(^2P_{1/2}). \quad (10)$$

The ratio of excited- and ground-state atomic iodine concentrations is approximately determined by the equilibrium constant for reaction (10), $K_{eq} = 0.75\,exp(402/T)$:

$$\frac{[I(^2P_{1/2})]}{[I(^2P_{3/2})]} = \frac{[O_2(a^1\Delta)]}{[O_2(X^3\Sigma)]} K_{eq}(T), \quad (11)$$

where the square brackets denote the concentration of the indicated species. Threshold for lasing (positive inversion) is achieved at $T = 295$ K when the ratio of $O_2(a^1\Delta)$ to total oxygen, or yield of singlet oxygen, is approximately 15%. The threshold is reduced considerably at lower temperatures due to the temperature dependence of the equilibrium constant. At $T = 160$ K, a temperature consistent with typical nozzle exit conditions, the threshold yield is reduced to 5%. The cross section for stimulated emission at room temperature is $\sigma = 7.4 \times 10^{-18}$ cm^2 and the gain in a COIL device is typically near $\gamma = 1\%/\text{cm}$.

The atomic iodine $5^2P_{1/2} \rightarrow 5^2P_{3/2}$ lasing transition is composed of six hyperfine components, as shown in Fig. 8. The nuclear spin is $I = 5/2$. The $5^2P_{1/2}(F' = 3) \rightarrow$

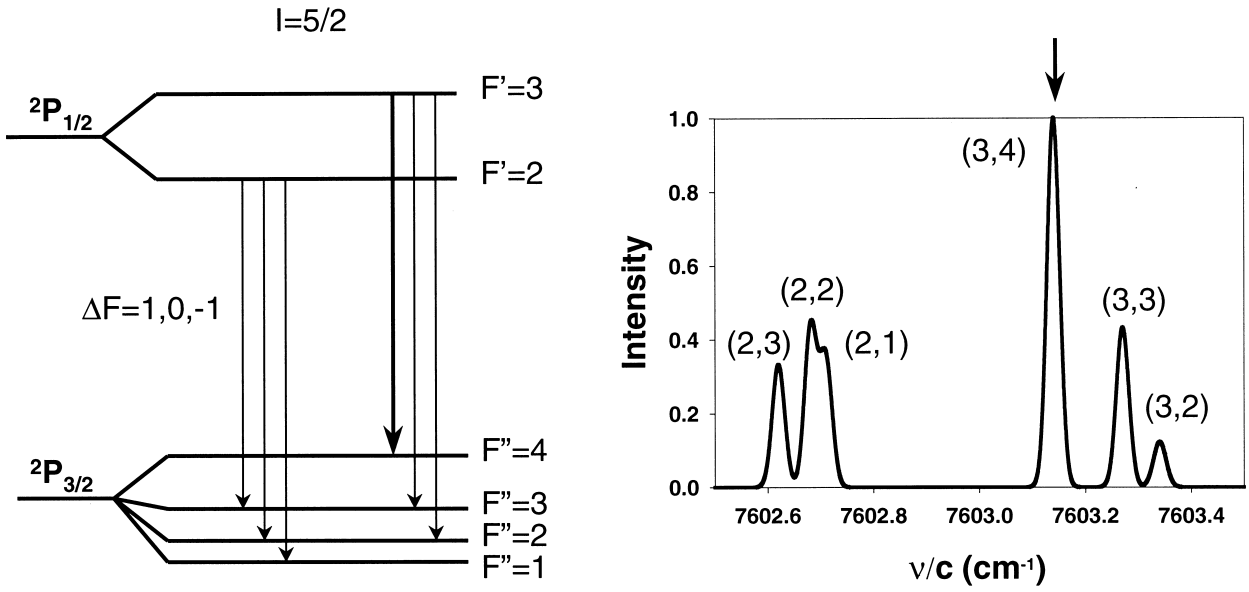


FIGURE 8 Hyperfine structure and spectrum of the I($^2P_{1/2} \rightarrow {}^2P_{3/2}$) transition in atomic iodine indicating the greatest gain is achieved for the $F' = 3 \rightarrow F'' = 4$ component near $\nu = 7603.2\ \text{cm}^{-1}$ or $\lambda = 1.315\ \mu\text{m}$.

$5^2P_{3/2}(F'' = 4)$ component provides the highest gain, and lasing occurs on this individual transition at $\lambda = 1.315\ \mu$m.

B. Hardware

A schematic diagram of a typical supersonic COIL device is shown in Fig. 9. The main components are (1) a gas–liquid reactor producing singlet oxygen, (2) the iodine supersonic mixing nozzles, (3) the gain region and optical resonator, and (4) the pressure recovery system.

The earliest singlet oxygen generators were of a sparger design, where chlorine gas is bubbled through the liquid hydrogen peroxide solution. The two-phase reaction requires large liquid surface-to-volume ratios for efficient

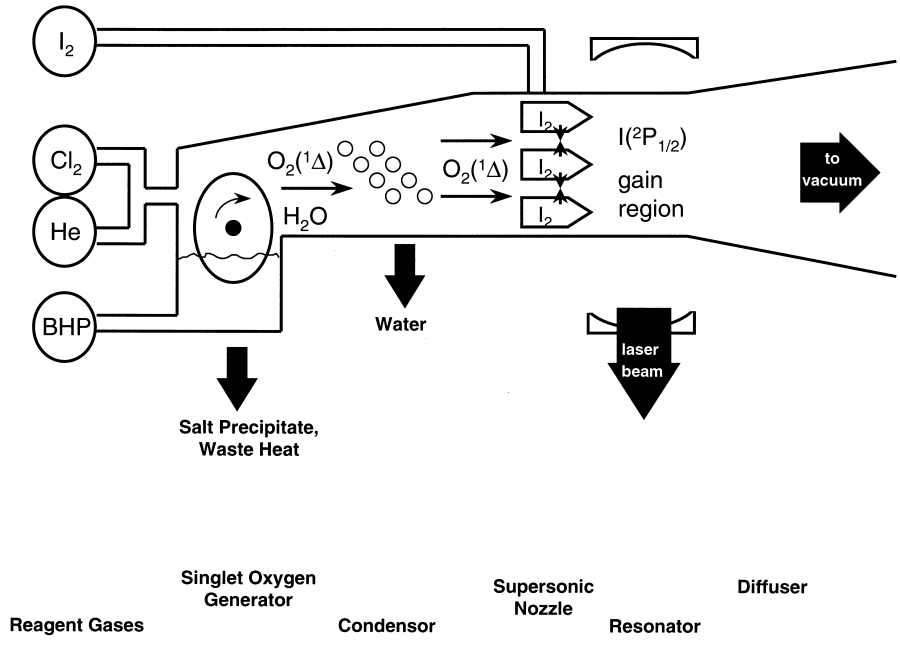


FIGURE 9 Schematic diagram of a typical supersonic COIL.

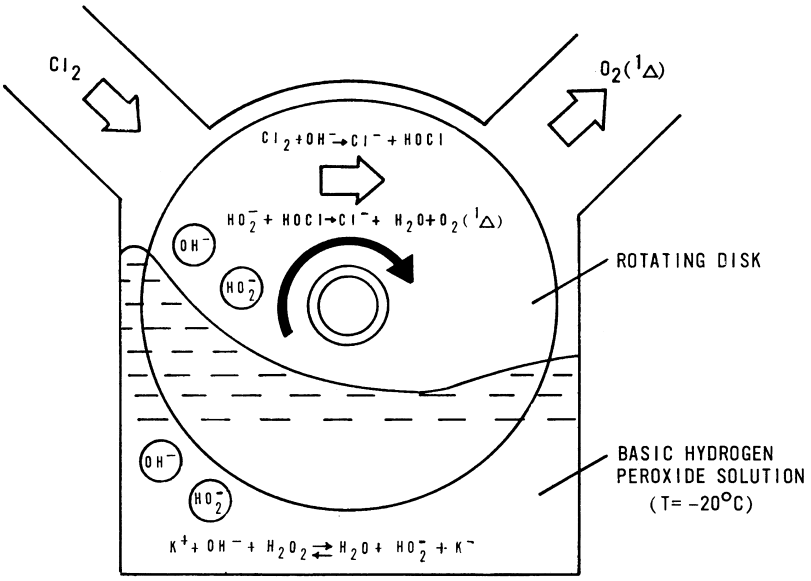


FIGURE 10 Rotating disk design for the chemical generation of singlet oxygen, $O_2(a^1\Delta)$, via the two-phase reaction of chlorine gas with liquid hydrogen peroxide.

operation, and the second generation of chemical generators were based on a wetted wall reactor, as shown in Fig. 10. By rotating a stack of disks through a pool of basic hydrogen peroxide and past a stream of chlorine gas, the reaction is confined to a thin film on each disk. The time required to diffuse the singlet oxygen from within the liquid phase to the gas-phase interface is reduced and the deactivation of $O_2(a^1\Delta)$ is minimized. Advanced generators involving sprays of droplets with near uniform size are now being used for COIL devices.

The effluent from the singlet oxygen generator includes $O_2(a^1\Delta)$ and water vapor. Water vapor is an efficient collisional quencher of the upper laser level, $I(^2P_{1/2})$, and reduces device performance. Thus, a water trap or condenser is often used to remove the water from the gas stream. The total pressure at the exit of the generator is typically 10–50 torr and usually includes helium as a diluent.

Molecular iodine is injected transverse to the primary oxygen flow near the throat of a supersonic nozzle, as shown in Fig. 11. The molecular iodine is rapidly dissociated to atomic iodine by the presence of the singlet oxygen in a complex, multistep mechanism. The flow is fully dissociated as it reaches the exit plane of the nozzle and before entering the gain region.

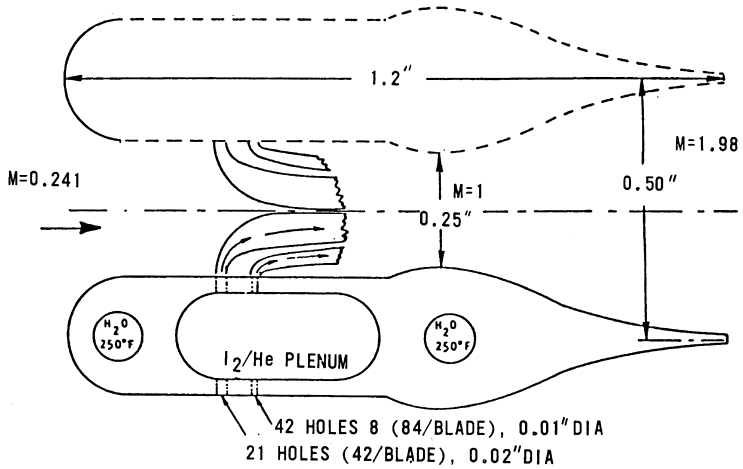


FIGURE 11 Supersonic mixing nozzle for a COIL where molecular iodine is injected transverse to the flow of singlet oxygen.

The iodine dissociation process is poorly understood and provides the greatest difficulty in modeling the gas-phase kinetics of the COIL device. The energetics of the dissociation process is shown in Fig. 7. The second electronically excited state of oxygen, $O_2(b^1\Sigma)$, is produced from the energy pooling reaction:

$$O_2(a^1\Delta) + O_2(a^1\Delta) \rightarrow O_2(b^1\Sigma) + O_2(X^3\Sigma) \quad (12)$$

and $O_2(b^1\Sigma)$ is sufficiently energetic to dissociate iodine. However, there is insufficient concentration of $O_2(b^1\Sigma)$ to explain the phenomenological dissociation rate, particularly when significant water vapor is present. A slow initiation of the dissociation process may begin by $O_2(a^1\Delta)$ excitation of vibrationally excited iodine, I_2^*, followed by a second $O_2(a^1\Delta)$ collision producing ground-state atomic iodine. The resulting iodine atoms are rapidly excited to the spin orbit state, $I(^2P_{1/2})$, by energy transfer. The dissociation then rapidly accelerates as $I(^2P_{1/2})$ replaces $O_2(a^1\Delta)$ as the dominant partner for production of I_2^*. This dissociation process removes at least 2, and sometimes as many as 6–18, $O_2(a^1\Delta)$ molecules per dissociated I_2 molecule, depending on the rate for deactivating the intermediate I_2^* and any quenching of $I(^2P_{1/2})$. In order to avoid consuming a large fraction of the energy stored in the metastable reservoir in this dissociation process, the ratio of molecular iodine to singlet oxygen flow rates is usually small, about 1%. During the lasing process, the iodine atoms will experience many energy transfer excitation and stimulated emission cycles to extract most of the energy stored in $O_2(a^1\Delta)$.

A simplified kinetic mechanism that retains the essential features of the iodine dissociation process is provided in Table II.

In the gain region, the molecular iodine is fully dissociated and a near equilibrium between excited iodine atoms and $O_2(a^1\Delta)$ is rapidly attained via the energy transfer reaction (10). Quenching of $I(^2P_{1/2})$ by H_2O drains energy from the singlet oxygen reservoir, until stimulated emission dominates the deactivation. Continuous-wave lasing action is achieved on a single hyperfine transition of the inhomogeneously broadened iodine atom at $\lambda = 1.315\,\mu$m.

C. High-Power Devices

Subsonic COIL devices were developed in the United States during the period 1977–1984. A rapid set of initial demonstrations took the subsonic COIL device from 4 MW to 100 W in just a few years. Scaling to 4.6 kW was achieved by 1982 with the 4-m-gain length COIL IV device.

An abrupt change in emphasis occurred with the first demonstration of a supersonic chemical oxygen–iodine laser in 1984. The reduction in device size, the lower cavity operating temperature, and the longer gain lengths were immediately recognized as critical advantages. The ReCOIL device, the first supersonic COIL, boasted 1.6 kW with only a 25-cm gain length.

The RotoCOIL device was based on the rotating-disk singlet oxygen generator design as shown in Fig. 10 and represents a third stage in COIL development emphasizing engineering demonstrations. By 1989, the supersonic RotoCOIL device was operated at 39 kW.

In 1996, the Air Force contracted to develop a megawatt-class chemical oxygen–iodine laser mounted aboard a modified Boeing 747-400F capable of destroying theater missiles during the powered boost phase. This new major weapons system, the airborne laser (ABL), is designed to destroy multiple tactical ballistic missiles, possibly carrying chemical and biological weapons, launched from random, previously unidentified sites over enemy territory

TABLE II Simplified COIL Kinetic Mechanism

Reaction	Rate coefficient (cm^3/molecule-sec)
$O_2(a^1\Delta) + O_2(a^1\Delta) \rightarrow O_2(b^1\Delta) + O_2(X^3\Sigma)$	2.7×10^{-17}
$O_2(b^1\Sigma) + H_2O \rightarrow O_2(a^1\Delta) + H_2O$	6.7×10^{-12}
$I_2(X^1\Sigma) + O_2(b^1\Sigma) \rightarrow I(^2P_{3/2}) + I(^2P_{3/2}) + O_2(X^3\Sigma)$	4.0×10^{-12}
$I_2(X^1\Sigma) + O_2(a^1\Delta) \rightarrow I_2^* + O_2(X^3\Sigma)$	7.0×10^{-15}
$I_2(X^1\Sigma) + I(^2P_{1/2}) \rightarrow I_2^* + I(^2P_{3/2})$	3.8×10^{-11}
$I_2^* + O_2(a^1\Delta) \rightarrow I(^2P_{3/2}) + I(^2P_{3/2}) + O_2(X^3\Sigma)$	3.0×10^{-10}
$I_2^* + H_2O \rightarrow I_2(X^1\Sigma) + H_2O$	3.0×10^{-10}
$I(^2P_{3/2}) + O_2(a^1\Delta) \rightarrow I(^2P_{1/2}) + O_2(X^3\Sigma)$	7.8×10^{-11}
$I(^2P_{1/2}) + O_2(X^3\Sigma) \rightarrow I(^2P_{3/2}) + O_2(a^1\Delta)$	2.7×10^{-11}
$I(^2P_{1/2}) + O_2(a^1\Delta) \rightarrow I(^2P_{3/2}) + O_2(b^1\Sigma)$	1.1×10^{-13}
$I(^2P_{1/2}) + H_2O \rightarrow I(^2P_{3/2}) + H_2O$	2.0×10^{-12}

at ranges of several hundred kilometers. Such technology offers to revolutionize warfare, as expressed by then Secretary of the Air Force, Sheila E. Widnall, in 1997: "It isn't very often an innovation comes along that revolutionizes our operational concepts, tactics, and strategies. You can probably name them on one hand—the atomic bomb, the satellite, the jet engine, stealth, and the microchip. It's possible the airborne laser is in this league." A fleet of seven airborne laser aircraft is anticipated to become operational by the year 2008.

IV. ADVANCED CHEMICAL LASERS

A. Short-Wavelength Chemical Lasers

Significant progress toward the development of chemically driven laser operating in the visible portion of the spectrum has been made over the past few years. The promise of very high brightness, high mass efficiency, and wavelength agility has justified a modest basic research program for several decades. For example, Eq. (8) suggests a chemical laser operating in the blue would possess a diffraction-limited brightness of almost 40 times that for an HF laser with similar power and aperture.

The premier candidate for a metastable energy reservoir to drive a visible chemical laser is $NF(a^1\Delta)$, which possesses an energy of 1.4 eV per molecule and has a radiative lifetime of 5.6 sec. Energy densities of greater than 6.7 J/liter have been demonstrated with near-unity chemical production efficiency. The two primary methods of production are the reaction of hydrogen atoms with NF_2 and the thermal decomposition of fluorine azide.

To extract the energy stored in $NF(a^1\Delta)$ in the form of laser power, energy transfer to lasant species such as BiF, $NF(b^1\Sigma)$, $IF(B^3\Pi)$, and $BH(A^1\Pi)$ has been studied. In a pulsed thermolysis apparatus, $BiF(A)$ concentrations of 1.8×10^{13} molecules/cm^3 have been achieved and a gain of 0.027%/cm demonstrated in the blue. Stimulated emission has been observed in a Mach 2.5 tabletop shock tube. However, saturated lasing has not been demonstrated.

B. All-Gas-Phase Chemically Pumped Iodine Laser

Another key metastable energy carrier is $NCl(a^1\Delta)$, which is isovalent with $NF(a^1\Delta)$ and $O_2(a^1\Delta)$. The discovery of the efficient energy transfer from $NCl(a^1\Delta)$ to atomic iodine opened the door to an all-gas-phase, chemically pumped atomic iodine laser operating at the same wavelength as COIL. Indeed, a subsonic transverse flow device employing hydrogen azide, HN_3, as the primary fuel has recently achieved lasing with a power of 180 mW.

The $NCl(a)$–I laser hardware appears similar to an HF laser in that the chemistry is driven by atomic fluorine. The chemical generation of excited iodine atoms is based on the following sequential reaction mechanism:

$$F + DCl \rightarrow DF + Cl \tag{13}$$

$$Cl + HI \rightarrow HCl + I\left(^2P_{3/2}\right) \tag{14}$$

$$Cl + HN_3 \rightarrow HCl(v) + N_3 \tag{15}$$

$$Cl + N_3 \rightarrow NCl(a^1\Delta) + N_2 \tag{16}$$

$$NCl(a^1\Delta) + I(^2P_{1/2}) \rightarrow NCl(X) + I\left(^2P_{1/2}\right). \tag{17}$$

In the first demonstration of lasing in this system the fluorine atoms were generated in a dc discharge. The efficiency and scaleability of this new chemical laser has yet to be established.

With the advent of the laser, a new source of light is available that possesses unique monochromatic, coherence, and propagation characteristics. When such laser devices are driven by chemical reactions, the power of the laser beam can be spectacular. Many scientific and technical challenges have been met in the past, three decades, integrating advances in chemistry, physics, fluid dynamics, optics, and engineering to develop these high-power chemical lasers. These devices are now finding important applications in industrial and military operations.

SEE ALSO THE FOLLOWING ARTICLES

LASERS • LASERS, GAS • LASERS, NUCLEAR PUMPED

BIBLIOGRAPHY

Basov, N. G. (1990). "Chemical Lasers," Springer, Berlin.

Bloembergen, N., and Patel, C. K. N. (eds.). (19) *Rev. Mod. Phys.* **59** (Part II).

Duffner, R. W. (1997). "Airborne Laser: Bullet of Light," Plenum Press, New York.

Forden, G. E. (19) *IEEE Spectrum* **34,** 40.

Gross, R. W. F., and Bott, J. F. (1976), "Handbook of Chemical Lasers," Wiley, New York.

Miller, J. (1987). *In* "Proceedings International Conference on Lasers, 1987," STS Press, McLean, VA.

Pimentel, G. C. (19) *Sci. Am.* **214,** 32.

Schawlow, A. L. (19) *Sci. Am.* **219,** 120.

Truesdell, K. A., and Lamberson, S. E. (19) "SPIE Proceedings International Symposium on 9th Gas Flow Chemical Lasers," p. 476.

Truesdell, K. A., Helms, C. A., and Hager, G. D. (19) "SPIE Proceedings 10th International Symposiun on Gas Flow Chemical Lasers," p. 217.

Verdyen, J. T. (1989) "Laser Electronics," Prentice-Hall, Englewood Cliffs, NJ.

Lasers, Color Center

Clifford R. Pollock
Cornell University

I. Introduction to Color Center Lasers
II. Physics of Laser-Active Color Centers
III. Formation of Color Centers
IV. Optical Gain from Color Centers
V. Color Center Laser Operation
VI. Future Developments

GLOSSARY

Cross-section A measure of the probability of a stimulated transition occurring. The larger the cross-section, the more likely a photon will interact with the system.
Free spectral range (FSR) The frequency difference between adjacent resonances of an optical cavity.
Interstitial The presence of an extra ion in the lattice, at a site not normally occupied in a perfect crystal.
Femtosecond (fsec) 10^{-15} second.
Quantum efficiency The ratio of the number of emitted photons to the number of absorbed photons.
Stokes shift The difference in wavelength between the luminescence and absorption band for a given optical center.
Threshold The point where a laser has just enough gain to begin oscillation.
Vacancy The absence of an ion at a lattice point that would normally be occupied.
Wavefunction The distribution of an electron in space. This distribution is generally localized around a potential well.

CERTAIN COLOR CENTERS in the alkali halides can be used to create broadly tunable, optically pumped lasers in the near infrared. These lasers may be operated in a continuous wave with output powers exceeding several watts, or mode locked to yield pulses of less than 100-femtosecond (fsec) duration. Laser-active color centers generally provide long-lived operation only when operated at cryogenic temperatures, although many of the centers can be stored at room temperature with no degradation. Color center lasers are unique in their ability to generate tunable radiation in the near infrared, spanning from 0.8 μm to 4 μm. This has made the color center laser useful for the study of guided wave optical devices, observation of ultrafast processes narrow bandgap semiconductors, and for high-resolution molecular spectroscopy.

I. INTRODUCTION TO COLOR CENTER LASERS

Color centers are simple point defects in crystal lattices, consisting of one or more electrons trapped at an ionic

vacancy in the lattice. These point defects are common in many crystalline solids and have been studied rather extensively in alkali halide crystals (e.g., NaCl, KCl, etc.). Certain color centers have optical absorption and emission bands that make them suitable as laser gain media. The structure and physics of these centers and transitions will be reviewed in the next section. Lasers based on these color centers are optically pumped, can be broadly tuned, share a similar cavity design with organic dye lasers, and can generate ultrashort pulses. The unique characteristic about a color center is the tuning range. At present, through the use of several different types of centers and host lattices, the entire region between 0.8 μm and 4.0 μm can be covered. Tunable coherent radiation over this wavelength region is important for studies in optical communication, ultrafast processes in semiconductors, and high-resolution molecular spectroscopy, to name a few applications. It is this unique combination of broad wavelength tuning range and continuous wave power that makes the color center laser useful.

The laser-active color centers possess two characteristics ideal for efficient, tunable operation. First, their absorption and emission bands are homogeneously broadened: all the excited centers can contribute energy to a given laser mode, and all the centers will be equally well pumped by a single-line laser operating within the pump absorption band. Homogeneous broadening allows efficient single-mode operation over the entire tuning range of the laser. Second, most of the transitions involved in laser emission are fully allowed. Such transition strengths when combined with broad homogeneous linewidths lead to large optical gains, quiet continuous wave (cw) operation, and the ability to generate ultrashort pulses ($<10^{-13}$ sec duration). The single-pass gain per unit length in a color center crystal containing a typical density of centers can exceed 50%/mm. Such gains are nearly two orders of magnitude larger than those obtained in solid-state lasers that use transition metal ion impurities for the gain medium.

Color center lasers can be operated in pulsed, mode-locked, or cw fashion. In the cw mode, output powers exceeding 5 W have been achieved in certain crystals. Typical output powers and tuning ranges for the various color center lasers are summarized in the following sections of this article. In pulsed operation, output powers of 100 MW have been achieved near 950 nm.

The single-frequency linewidth of the color center laser is truly exceptional in its purity. In single-mode cw operation, linewidths below 4 KHz have been achieved. This performance is attributable to the solid-state nature of the gain media: there are no moving parts in the laser cavity to perturb the phase of the laser. Such precise frequency definition, coupled with broad tunability, has made the color center laser a powerful tool for spectroscopy and metrology.

It is perhaps the mode-locked operation of the color center laser that has been the most spectacular. Taking advantage of the broad emission bandwidth, the laser output can be transformed into a train of tunable picosecond and femtosecond duration pulses. The femtosecond color center laser was one of the first tunable subpicosecond pulse source, and it is still used as a source with reasonable average output power (hundreds of milliwatts) and fast repetition rates ($\sim$100 MHz). Such short pulses are extremely useful for the investigation of ultrafast phenomena in semiconductor materials and the nonlinear characteristics of optical fibers.

The laser-active color centers are generally stable only when cooled to cryogenic temperatures. Although some center types allow for low-duty-cycle pulsed operation at room temperature, cw operation at room temperature usually leads to fading of the output power. At room temperature, the centers either thermally dissociate, or become mobile and transform into nonlasing centers through attachment to other defects. Thermal degradation can be minimized or eliminated at reduced temperatures, hence the crystals are usually anchored to a cold finger maintained at liquid nitrogen temperature (77 K). Additionally, the radiative quantum efficiency for some of these centers increases significantly at reduced temperatures.

Due to the relatively short radiative lifetime of the color center, intense optical pumping from a laser source is necessary to achieve efficient laser operation. The pump laser depends on the crystal, but is usually a Nd:YAG laser operating at 1.06 μm, or an Ar- or Kr-ion laser operating in the visible. Diode-pumping has been demonstrated on a few color centers, but to date the resulting output power has been only marginally useful due to the poor spatial overlap between the pump beam and the color center laser mode in the crystal. Enhancements in the brightness of available laser diode pump lasers could change this situation.

The color center laser cavity must contain dispersive elements, such as a prism, in order to facilitate tuning and line narrowing. Properly designed, a color center laser is usually capable of tuning over a range exceeding 25% of its central wavelength. Specific examples of the design and construction of the laser are described in this article.

II. PHYSICS OF LASER-ACTIVE COLOR CENTERS

A. Optical Emission Processes

While color centers exist in many different crystal lattices, most research to date has been done on point defects in alkali halide crystals. This review will concentrate on the alkali halide centers because they form the basis of practically all the useful color center lasers, and they are well understood. A representative sample of color centers in

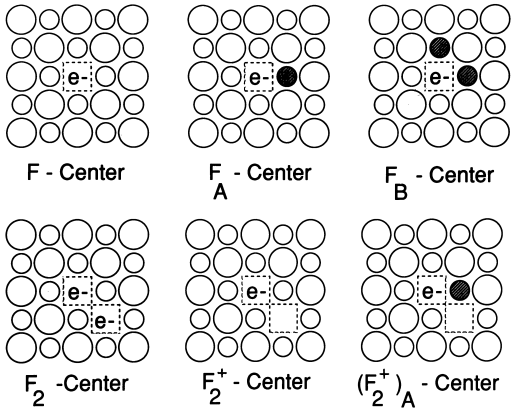


FIGURE 1 A representative sample of color centers in alkali halide crystals. The large and small circles represent negative and positive ions, respectively. Colored circles represent alkali impurities.

alkali halide crystals is shown in Fig. 1. All laser-active color centers involve anion (halide ion) vacancies. The F-center has the simplest structure, consisting of a single electron trapped at a vacancy surrounded by an essentially undisturbed lattice. If one of the neighboring alkali ions is a substitutional alkali impurity, say Li^+ in a KCl crystal, the center is called an F_A center. Similarly, the F_B center consists of an F-center beside two substitutional impurities. Two adjacent F-centers along a [110°] axis of the crystal form the F_2 center. Its ionized counterpart is called the F_2^+ center, which is an F_2 center with only one electron. Analogous to the F_A center, the $(F_2^+)_A$ center consists of an F_2^+ center adjacent to a substitutional alkali ion. Of the various centers described in Fig. 1, the F_A, F_B, F_2^+, and $(F_2^+)_A$ defects form color center lasers. Note that the simple F-center does not lase. The reason for this will be discussed below.

There are also several types of laser-active color center not shown in Fig. 1. These include the $(F_2^+)^*$ and $F_2^+{:}O^{2-}$ centers, which consist of F_2^+ centers associated with an as-yet unknown defect, and with an oxygen ion, respectively. Also missing is the Tl° (1) center, which is schematically similar to the F_A center but is functionally more like a neutral Tl atom perturbed by an adjacent vacancy, and the N_2-center, which consist of three F-centers in a trigonal arrangement. Many other alkali-halide color centers are not shown, such as the larger aggregate centers (F_4, etc.), but with few exceptions, these larger centers do not play a significant role in the physics of color center lasers. Laser-active color centers have also been reported in crystals other than the alkali halides, including the F^+ center in alkaline earth oxides, and the H_3 and N_3 centers in diamond. These lasers have not been widely reproduced or studied, and there are still questions concerning the physics of their operation, hence they will not be discussed in this article.

The optical absorption and emission of the F-center can be understood on a qualitative basis using a highly simplified quantum mechanical model, the particle-in-a-box. This model is based on the fact that F-centers are essentially electrons trapped in a three-dimensional square-well formed by the electrostatic potential of the surrounding positive ions. In this model, the energy between the ground state and first excited state is

$$E_{2p} - E_{1s} = 3h^2/8ma^2,$$

where 1s denotes the ground state, 2p the excited state, h is Planck's constant, a is the box dimension, and m is the mass of the electron. Note that the energy scales with the well dimension as a^{-2}. This model works amazingly well when applied to color centers. It is found experimentally that if distance a is taken as the nearest neighbor separation, the F-band energy can be related to a for most alkali halides

$$E_F = 17.7\, a^{-1.84},$$

where a is in angstoms, and E_F is in electron volts. This is known as the *Mollwo relation*. As will be illustrated with examples below, many color centers follow a similar relation between the lattice constant and the energies of their absorption and emission bands.

The model also predicts a strong coupling between the lattice dimensions and the transition energy. Lattice vibrations, called phonons, harmonically vary the dimensions of the square well with a period on the order of 10^{-13} sec. This causes the energy levels to harmonically vary on this same time scale. Since this perturbation occurs to all F-centers, and the perturbation rate is much faster than the excited state lifetime of the center, the absorption and emission bands are homogeneously broadened.

B. F-Centers

The F-center is the most fundamental color center defect in the alkali halide lattice. Although it is not laser-active, the optical properties of the F-center are important in understanding the laser physics of other color center lasers. The fundamental absorption band of the F-center, called the "F" band, corresponds to a transition from the 1s-like ground state to the 2p-like first excited state of the square-well potential. The F-band transition is very strong, and dominates the optical spectrum of the alkali-halide crystal. In fact, the term *F*-center comes from the German word *Farbe*, meaning color, and refers to the strong color imparted to the otherwise transparent alkali–halide crystals.

As noted above, the simple F-center is not suitable for lasing. After excitation to the first excited state, the F-electrons are very near the conduction band, and can be easily ionized by thermal or optical energy. The strong

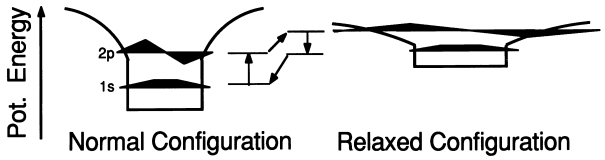


FIGURE 2 Potential well and wavefunction for the F center (a) before excitation (normal configuration) and (b) after excitation (relaxed configuration).The four-level energy scheme is shown between the wells. All energies are plotted with respect to the conduction band, hence the relaxed excited state appears to be at a higher energy than the first excited state.

possibility of self-absorption into the conduction band by photons emitted from other F-centers also exists, destroying the potential gain mechanism as well as introducing loss. These deleterious effects result from an anomalous spatial relaxation following optical excitation, as illustrated in Fig. 2. After excitation to the 2p state, the wavefunction of the electron spreads out. In response to this diffuse wavefunction, the lattice expands slightly, enlarging the dimensions of the potential well. This expanded well causes even further spreading of the excited-state wavefunction, leading to yet more expansion of the well, until finally lattice forces restrain the relaxed well. The resultant wavefunction extends out several lattice constants and is called the relaxed excited state (RES). Due to the poor spatial overlap of the RES and the ground state wavefunction, the emission dipole moment is relatively small, leading to a small gain cross-section. No laser has ever been made based on simple F-centers, and it appears unlikely that such a laser will ever exist. On the other hand, the F-center is a very useful building block for forming other varieties of color centers.

C. F_A(II) Centers and F_B(II) Centers

Recall that the F_A center consists of an F center located beside a substitutional impurity (see Fig. 1). F_A centers are classified into two categories depending on their relaxation behavior after optical excitation. Type I centers, denoted F_A(I), behave almost identically to the simple F-center just described. They display the same diffuse relaxed excited state, low-gain cross section, and self-absorption. The other (and rarer) class of F_A centers are called Type II. They are distinguished by a dramatically different relaxation process. After excitation with a photon in the visible region of the spectrum, F_A(II) centers relax to a double well configuration, as shown in Fig. 3. A neighboring anion moves between the impurity and neighboring cation, creating a double-well potential. The resulting electronic wavefunction remains localized, and the transition moment is very strong. Because energy is expended in moving the anion within the lattice, the resulting emission from this transition is dramatically Stokes-shifted out to the 2–3-μm region.

The presence of the impurity ion in the F_A(II) system causes the P_z orbital to be distinguished from the $P_{x,y}$ orbitals (see Fig. 3). Transitions of the P_z orbital result in a longer wavelength absorption band that is often well resolved from the main band. In laser applications, this additional band considerably enhances the probability that a convenient pump source will overlap with the absorption.

Upon relaxation, the anion separating the two wells can move back into its original location, or into the original vacancy. In the latter case, the F_A(II) center will have effectively rotated 90°. Such reorientation can lead to orientational bleaching of the band, where the center no longer absorbs light polarized along the original z-axis. In practice, optimum power is achieved from F_A(II) crystals if a crystal ⟨100⟩ axis lies at 45° to the laser and pump polarization, so that either center orientation will have a component of the dipole moment aligned with the stimulating and pump fields.

F_A(II) centers have been found only in Li-doped alkali halides, specifically KF, KCl, and RbCl. The Li is incorporated into the crystal during growth in concentrations approaching 0.02 mole %. Lasers have been built using the KCl and RbCl centers, and except for different tuning ranges, their characteristics are similar. The quantum efficiency, η, of the F_A(II) luminescence in KCl:Li is about 40% at T = 77 K, and decreases slowly with increasing temperature until it approaches zero at T = 300 K. Efficient operation of F_A(II) center lasers thus requires that the crystals be maintained at cryogenic temperatures during use. Liquid nitrogen (T = 77 K) is commonly used, being inexpensive and readily available. Coupled with the large Stokes shift, the low quantum efficiency leads to a maximum power conversion efficiency of about 10% for the F_A(II) laser.

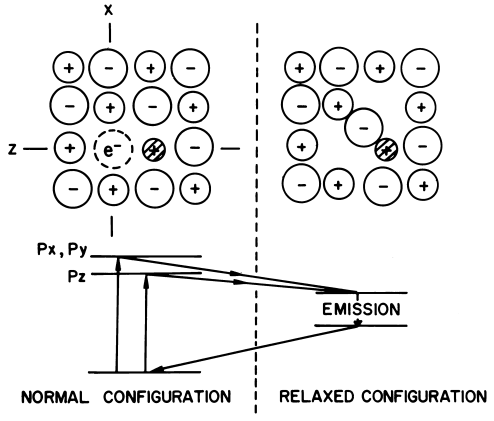


FIGURE 3 Structure and energy-level diagram of the F_A(II) center (a) before optical excitation (normal configuration) and (b) after optical excitation (relaxed configuration).

F_B centers involve an F-center beside two substitutional alkali impurities. Since the distribution of foreign alkali ions in an otherwise pure lattice is statistical, F_B centers are far fewer in number than F_A centers, and are obtained in substantial quantities only when the impurity dopant concentration is fairly high (approximately 1%). Like the F_A center there are two types of F_B center, also classified by their relaxation behavior. The F_B(I) center is formed when the two substitutional impurities lie along a common ⟨100⟩ axis. The F_B(I) center has optical properties similar to the F-center, which preclude the possibility of lasing action. The F_B(II) center is formed when the two alkali impurities are adjacent to one another along a ⟨110⟩ axis of the crystal (see Fig. 1).

The F_B(II) centers relax into a double-well configuration after excitation, similar to that shown in Fig. 3 for the F_A(II) center. Their emission is almost entirely quenched at temperatures of 4 and 300 K, but reaches a maximum around 100 K. Optimum laser performance can be obtained from the F_B(II) center with cw pumping, which raises the crystal temperature from 77 K to around 100 K. Pulsed operation is less effective in raising the crystal temperature, since the crystal has a chance to cool between pulses. One difficulty with F_B(II) centers is that they are accompanied by substantial quantities of F_A(I) and F_B(I) centers, all of which have overlapping absorption bands with the F_B(II) center. Consequently, the optical pump power suffers losses by these residual centers, diminishing the overall efficiency of the F_B(II) center laser.

The F_A(II) and F_B(II) are among the most stable color centers lasers presently known, providing tunable laser radiation in the 2.2–3.8-μm region. The operational lifetime of these lasers is almost entirely determined by secondary effects, typically crystal fogging due to small vacuum leaks or water desorption in the dewar. Figure 4 shows the power and tuning range of an optimized color center laser using F_A(II) and F_B(II) centers. Other relevant details of these and the other color center lasers are tabulated in Table I.

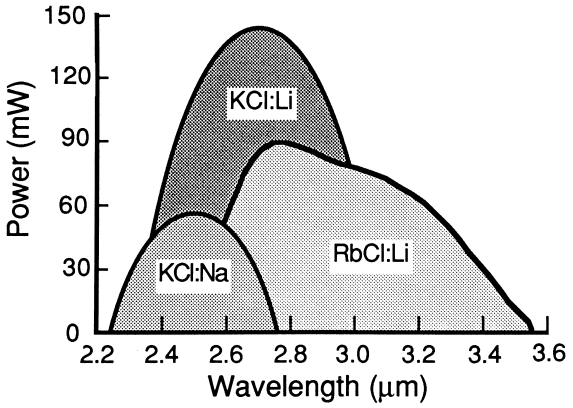


FIGURE 4 Tuning range and output power from an optimized color center laser using F_A(II) and F_B(II) centers. (Adapted from K. R. German, 1986).

D. F_2^+ Centers

A powerful but short-lived color center laser is based on the F_2^+ center. As indicated in Fig. 1, the F_2^+ center consists of two adjacent anion vacancies sharing one trapped electron. In contrast to the $F_{A,B}$(II) centers, the F_2^+ relaxation

TABLE I Performance of Common Color Center Lasers

Host lattice	Center	Pump wavelength (μm)	Tuning range (μm)	Maximum power (watts)	Operational lifetime
LiF	F_2^+	0.647	0.82–1.05	1.8	Days
NaF	$(F_2^+)^*$	0.87	0.99–1.22	0.4	Weeks
KF	F_2^+	1.06	1.22–1.50	2.7	Days
NaCl	F_2^+	1.06	1.4–1.75	1	Days
NaCl:OH	F_2^+:O^{2-}	1.06	1.42–1.85	5	Years
KCl:Tl	$Tl^\circ(1)$	1.06	1.4–1.64	1	Years
KCl:Na	$(F_2^+)_A$	1.32	1.62–1.95	0.05	Months+
KCl:K_2O	F_2^+:O^{2-}	1.32	1.7 –1.85	0.06	Months+
KCl:Li	$(F_2^+)_A$	1.32	2.0–2.5	0.4	Years
KI:Li	$(F_2^+)_A$	1.7	3.0–4.0	0.006 (pulsed)	Unknown
KCl:Na	F_B(II)	0.514	2.25–2.65	0.05	Years
KCl:Li	F_A(II)	0.514	2.3–3.0	0.1	Years
RbCl:Li	F_A(II)	0.647	2.6–3.9	0.7	Years
KCl	N	1.064	1.27–1.35	0.04 (pulsed)	Months+

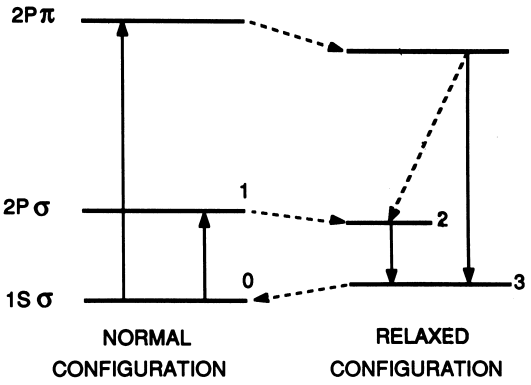


FIGURE 5 Energy-level diagram of the F_2^+ center. The dotted lines represent nonradiative transitions.

following excitation entails only a slight enlargement of the surrounding lattice.

The configuration of the F_2^+ center is suggestive of an H_2^+ ion imbedded in a dielectric continuum. The two lattice vacancies play the role of the protons. The energy levels of such an ion are related to the free space case by

$$E_{F2+}(r, K_o) = \left(1/K_o^2\right) E_{H2+}(R),$$

where $R = r/K_o$ is the proton separation in free space, r is the distance between the lattice vacancies, K_o is the dielectric constant of the lattice, and $E_{H2+}(R)$ is the energy function of the molecular hydrogen molecule in free space, which has been calculated for various states and separations R. Figure 5 shows an F_2^+ energy-level diagram. The levels are named after their molecular-ion counterparts. The left side of the energy diagram shows the RES levels. Comparison of measured ground and excited state levels of the F_2^+ center in several alkali halide lattices with predicted levels from the imbedded H_2^+ model show excellent agreement. Dotted lines indicate nonradiative transitions.

Since the alkali halides have a wide variety of lattice dimensions (r) and dielectric constants (K_o), one would expect the F_2^+ center to have a wide distribution of absorption energies in the various crystals. This is indeed the case, as shown in Fig. 6, where the absorption bands of

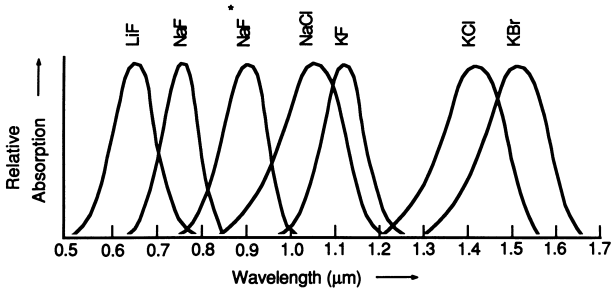


FIGURE 6 Normalized absorption bands of the F_2^+ center in various alkali halides.

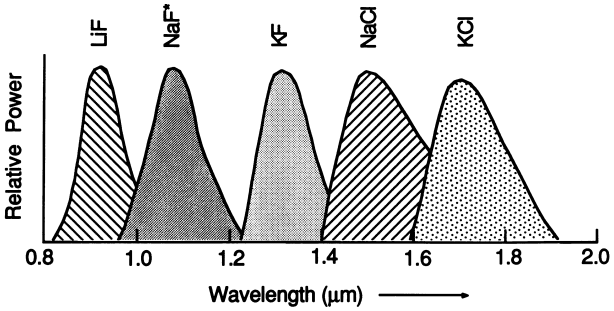


FIGURE 7 Normalized tuning curves for some of the known F_2^+ lasers.The NaF* curve is from an $(F_2^+)^*$ laser.

the F_2^+ center range from 0.67 μm in the tight LiF lattice, up to 1.5 μm in the much larger KBr lattice. Similar to the F-center, the broadening of individual absorption bands is caused by lattice phonons, and is homogeneous.

The transition of greatest interest for F_2^+ lasers is the $1S\sigma \rightarrow 2P\sigma$ transition between the ground state and first excited state. This transition has nearly ideal properties for laser operation: (1) the oscillator strength is large, allowing for a large gain cross-section; (2) the quantum efficiency is 100%; (3) the Stokes shift is just enough to prevent overlap of the emission and absorption bands, forming an ideal four-level system; and (4) there are no excited state absorptions that overlap with the emission. Lasers based on this transition are among the most powerful and efficient color center lasers ever made. Figure 7 shows normalized tuning curves of some common F_2^+ lasers. The tuning range extends from 0.8 and 1.9 μm, a range important for optical communication. (The NaF* band shown in Fig. 7 belongs to the $(F_2^+)^*$ center described below.)

F_2^+ center lasers have two drawbacks: (1) they display a slow fading of output power with extended operation, and (2) the crystals must be stored at cryogenic temperatures at all times. The mechanism for the decay is not totally understood, but it is probably associated with reorientation of the center. Excitation to the $2P\pi$ state is known to lead to a reorientation of the center's axis through nonradiative relaxation to one of the other ⟨110⟩ directions in the lattice. Figure 8 illustrates the reorientation of an F_2^+ center. Under the intense optical fields of laser operation, multiphoton excitation from the laser pump source can excite the $2P\pi$ state and lead to reorientation. Since there is nothing in the lattice to pin the center to one location, repeated flipping of the F_2^+ centers cause them to take a random walk through the crystal. Eventually, it is likely they will run into other centers or defects, forming larger F-center aggregates.

The requirement of cold storage arises because the F_2^+ centers are formed through radiation damage. (Coloration techniques are described briefly in Section III.) At elevated temperatures, the F-centers and radiation

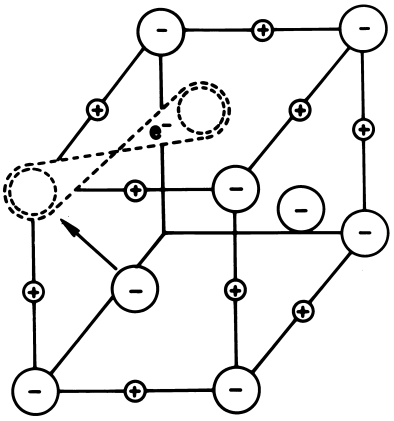


FIGURE 8 Schematic representation of the reorientation of an F_2^+ center. A neighboring anion moves into one of the two vacancies, causing an effective rotation of the center.

byproducts (such as interstitials) tend to annihilate one another through thermal motion. Storage at reduced temperature ($T < 170$ K) stops this aggregation. Cryogenic specimen dewars are generally used for long-term storage.

The operational lifetime of an F_2^+ laser depends on how long the crystal has been pumped. Under the best conditions, a single crystal can be made to operate for only several days. This decay, coupled with the awkwardness of creating and storing the active crystals, and the discovery of more stable color center lasers, has virtually eliminated the use of F_2^+ lasers. However, the F_2^+ center laser was the first powerful color center laser, and in the hands of skilled experimenters it has been used to generate tunable near infrared radiation.

E. Stabilized F_2^+ Centers

The F_2^+ centers can be associated with certain defects in the crystal lattice to form more stable color centers with output characteristics similar to the F_2^+ center. To date, four types of stabilized centers have been reported: the $(F_2^+)_A$, $(F_2^+)^*$, $(F_2^+)^{**}$, and $F_2^+{:}O^{2-}$ centers. The (F_2^+)A center is an F_2^+ center located beside an alkali impurity (see Fig. 1). The $F_2^+{:}O^{2-}$ center is an F_2^+ center beside a doubly negative anion impurity ion. The $(F_2^+)^*$ and $(F_2^+)^{**}$ centers are F_2^+ centers associated with as yet undetermined defect in the lattice, most likely an interstitial halogen defect.

These stabilized centers display most of the advantages of the F_2^+ center, such as large cross-section and unity quantum efficiency, but they offer a number of additional benefits as well. First, the centers are pinned at one point in the lattice, and as a result they display reduced or no fading. Second, new and useful tuning ranges can be created because the stabilizing defect slightly perturbs the energy levels. Finally, the room temperature shelf life of the centers is usually increased to essentially infinite length with no degradation of ultimate laser performance.

1. $(F_2^+)^*$ and $(F_2^+)^{**}$ Centers

The $(F_2^+)^*$ center appears in NaF after a radiation-colored F_2^+ center crystal is allowed to sit at room temperature for several days in the dark. The F_2^+ band disappears and is replaced by a new band at a longer wavelength. Pumping this new band produces emission that is shifted to longer wavelength relative to the F_2^+ emission band, but otherwise displays all the characteristics of an F_2^+ center. Strong evidence has been found linking the density of $(F_2^+)^*$ centers to the dosage of radiation used to introduce color centers.

The $(F_2^+)^{**}$ center is similar to the $(F_2^+)^*$, but is even further shifted in wavelength for both the absorption and emission band. In contrast to the $(F_2^+)^*$, the $(F_2^+)^{**}$ center is only generated in NaF crystals which contain OH^-, thus it is thought that the stabilizing defect for the "double star" center is an artifact of radiation-damaged OH^- in the crystal.

The "star" centers can be stored at room temperature, and in NaF they cover an important spectral window around 1.1 μm. Figure 7 includes a tuning curve of the NaF $(F_2^+)^*$ laser. Unfortunately, like the F_2^+ laser, the $(F_2^+)^*$ and ** centers also display a slow fading of output power with use. This fading is an order of magnitude slower than that of the F_2^+ laser, so they represent an improvement from the user point of view. To this date, the actual structure of the "star" centers is not known, although it is likely that they are identical to the $F_2^+{:}O^{2-}$ center described in section II.E.4, with the difference that the star centers are created with radiation damage and the $F_2^+{:}O^{2-}$ center is created through additive coloration.

2. $(F_2^+)_A$ Centers

A stable laser-based $(F_2^+)_A$ center has been demonstrated in several lattices, and currently is the only color center laser able to tune beyond 3.9 μm. $(F_2^+)_A$ center lasers combine the best characteristics of the F_A and F_2^+ laser: (1) they are operationally stable with no fading; (2) they can be stored at room temperature because they are additively colored (see section III); and (3) they are reasonably powerful.

The stability of the center arises from the trapping of the otherwise mobile F_2^+ center to one location in the lattice, thereby stopping the debilitating migration. The impurity also affects the energy levels of the F_2^+ center: the absorption bands are only slightly shifted, but the emission bands are shifted to longer wavelengths. The size of the shift depends on the substitutional impurity: a small ion in a large lattice causes a large shift, while a similar-sized

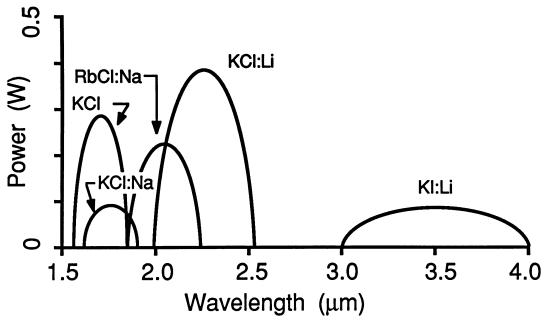


FIGURE 9 Tuning curves of several $(F_2^+)_A$ lasers. The data from the KI:Li laser are extrapolated from pulsed data.

substitutional ions cause a smaller shift. This new degree of freedom is a boon to the laser engineer: by choosing a suitable dopant, the energy bands of the F_2^+ center can be shifted to new wavelength ranges. Examples of these new ranges are shown in Fig. 9, where the laser tuning curves of several $(F_2^+)_A$ centers in KCl, KI, and RbCl are displayed. To illustrate the wavelength shift caused by the impurity, the emission band of the F_2^+ center in pure KCl is also displayed, along with the two known $(F2+)_A$ centers in KCl.

To date, $(F_2^+)_A$ lasers have been able to produce output powers of only several hundred milliwatts. This is a fraction of the power available from a comparable F_2^+ laser, and is due to the low density of $(F_2^+)_A$ centers that can be created in a given crystal. Typical $(F_2^+)_A$ crystals only absorb about 30% of the pump power, limiting the overall conversion efficiency. The low absorption is due to a low concentration of centers, which in turn appears to be due to a lack of suitable electron traps. The F_2^+ and $(F_2^+)_A$ centers are each ionized, so there must be one stable electron trap in the crystal for each active center. In the additively colored $(F_2^+)_A$ center crystals, the traps are other color centers, such as the F_A center, which traps two electrons to become a F_A^- center. (In the F_2^+ center crystal, artifacts from radiation damage such as interstitial ions serve this function.) Typical of most negatively charged color centers, the lowest lying energy level of this extra electron is energetically near the conduction band of the crystal, so that only a small amount of energy (thermal or optical) is needed to ionize the center, creating again a free electron. The F_A^- electron traps are subsequently slowly discharged by the intense pump light from the laser. To counter this, all $(F_2^+)_A$ lasers use an auxiliary ultraviolet light to continually reionize the $(F_2)_A$ centers to form $(F_2^+)_A$ centers. A dynamic equilibrium between $(F_2^+)_A$ and F_A^- centers thus exists in the crystal during lasing. In Section IV, an example of an $(F_2^+)_A$ laser is shown (see Fig. 13) that uses a Hg lamp to sustain the lasing population of $(F_2^+)_A$ centers. Typically, lasing ceases within 1 sec of blocking the UV illumination onto the crystal. In spite of the comparatively low power of the $(F_2^+)_A$ laser, for many applications a stable, reliable laser source with 200-mW output power is sufficient.

3. The F_2^+:O^{2-} Center

The F_2^+:O^{2-} center forms one of the most stable and powerful color center lasers yet discovered. The center is an F_2^+ center adjacent to a doubly negative substitutional anion impurity, usually O^{2-} although S^{2-} has also been used. A schematic representation of the model is shown in Fig. 10. The structure is distinct from the $(F_2^+)_A$ center because the perturbing impurity is located at an anion site.

This F_2^+:O^{2-} color center laser is operationally similar to the $(F_2^+)_A$ center in three ways: (1) the crystals are additively colored and can be stored at room temperature; (2) lasing is enhanced by an auxiliary light; and (3) the tuning range of each crystal is shifted to slightly longer wavelength than that of the F_2^+ center in the corresponding lattice. In addition, the center appears to be more robust than the $(F_2^+)_A$ center in terms of electron trapping and operating temperatures. Because the stabilizing impurity ion is doubly negative (as opposed to the singly negative anion it replaces), the overall F_2^+:O^{2-} center is "charge neutral," so there is little tendency for a free electron to be attracted to and neutralize the laser-active center, as is the case with other F_2^+ systems. The F_2^+:O^{2-} center essentially has its own "built-in" electron trap.

The laser has been demonstrated in NaCl doped with O^{2-} and S^{2-}, in KCl and RbCl doped with O^{2-}, and will probably be found in other lattices as well. Best results to date have been obtained with the NaCl system, which conveniently operates over the 1.4–1.8-μm range. Figure 10 shows the absorption, emission, and laser-tuning range

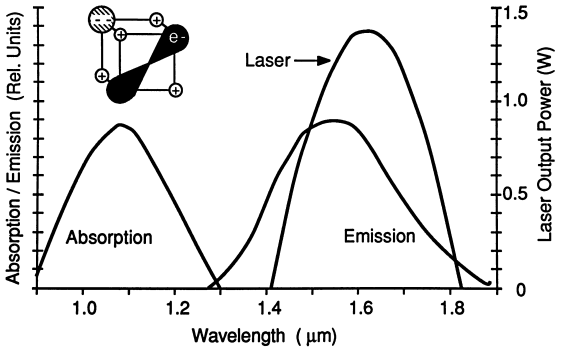


FIGURE 10 Tuning curve for a NaCl F_2^+:O^{2-} center laser. The absorption and emission bands are shown in relative units. The insert shows a schematic representation of the F_2^+:O^{2-} center. Note that the O^{2-} impurity is at an anion location, in contrast to the $(F_2^+)_A$ center.

of the NaCl system. Following additive coloration, the crystal is exposed to UV light for 30 min to aggregate F centers into F_2^+ centers. A stable absorption band forms near 1.09 μm, accompanied by a strong emission band centered at 1.6 μm. When cooled to 77 K and pumped by a cw Nd:YAG (Nd-doped yttrium-aluminum garnet) laser at 1.06 μm, the NaCl crystal has been made to lase with over 5 W of cw output power. Synchronous-pumping mode-locked operation of the NaCl laser generates 4-psec pulses typically. Using additive pulse mode-locking techniques (see Section V.c) 75-fsec pulses can be routinely generated with this laser.

Under the action of intense pumping, the $F_2+:O^{2-}$ center stays pinned to the O^{2-} ion, but it can still reorient in the lattice. Only centers that are aligned parallel to the pump polarization strongly interact with the radiation, so they are most likely to reorient. Unfortunately, once the center flips, it no longer is aligned with the driving polarization, so it no longer participates in the gain process. To counter this effect, an auxiliary light is simultaneously exposed onto the crystal during pumping. The auxiliary light acts to reorient the misaligned centers so that their dipole moment becomes aligned parallel to the pump- and laser-electric-field polarization. Commonly, the output of a HeNe laser or a frequency-doubled Nd:YAG laser are used. A few milliwatts of auxiliary power is more than enough to counter the effects of intense pumping.

F. Tl°(1) Center

The Tl°(1) center consists of a neutral Tl atom perturbed by the field of an adjacent single-anion vacancy. The superscript in the notation denotes that the Tl atom is neutral. The number in parenthesis represents the number of vacancies that are adjacent to the atom. The Tl°(1) center has been demonstrated to lase in KCl and KBr lattices doped with Tl. The latter crystal has proven to be very difficult to grow, so most commonly the Tl°(1) center is used in KCl. An F-center trapped beside the Tl ion will find that its electron spends most of its time on the Tl^+ ion, effectively forming a neutral Tl atom that is perturbed by an adjacent positive vacancy.

The transition involved in the Tl°(1) laser is not related to other color center transitions discussed so far. This laser transition occurs between the perturbed $^2P_{1/2}$ and $^2P_{3/2}$ states of the free Tl atom. This transition is normally parity forbidden, but the strong odd symmetry perturbation caused by the positive ion vacancy mixes the ^2P states with higher lying states and allows for a modest electric dipole to appear between these states. The absorption band caused by this transition in KCl is centered around 1.06 μm, and the emission band is centered at 1.5 μm. The Stokes shift comes about through lattice relaxation after excitation. The laser has a relatively narrow tuning curve, extending from 1.4–1.63 μm for the KCl host, and from 1.5–1.7 μm for the KBr host. This range is about half that expected for an F_2^+-type center. Output powers up to 1 W have been obtained from the KCl Tl°(1) laser. When mode locked, pulses as short as 9 psec have been directly generated.

The Tl°(1) center must be formed by radiation damage, usually using 2-MeV electron beams. Once formed, the center is operationally stable, but requires modest cooling (T < 0 C) for long-term storage. The laser properties of the crystal are destroyed if warmed above room temperature for short periods (days), or if pumped too hard in a laser. Due to these inconveniences, and since the NaCl $F_2+:O^{2-}$ offers more power, broader tuning, and easier handling, the Tl°(1) laser has been largely displaced by NaCl.

G. N Centers

In most alkali halides a pair of absorption bands exist at a wavelength slightly longer than the F_2 band, called the N_1 and N_2 bands. The structure of the N centers responsible for these bands has not been conclusively determined, even though the center has been the focus of many studies over the past two decades. The longer wavelength band, called the N_2 band, is usually attributed to the F_3 center in which the three-point defects form a triangle in the lattice.

In KCl, the N_2 band overlaps the 1.06-μm line of the Nd:YAG laser, which forms a convenient pump for the center. The emission, shown in Fig. 11, spans from 1.1–1.5 μm. Lasing in a pulsed mode has been obtained from this center over the 1.23–1.35-μm range. The laser only operates in a pulsed mode however. The pulsed operation is probably (but not yet conclusively) due to multiplet formation of the electrons in the excited state. Unlike all the other lasers described in this article, the N-center has more than one electron, which leads to possible triplet or multiplet formation upon relaxation. Such multiplet formation could lead to deleterious absorptions in the lasing region, and to a general depletion of laser-active centers. Nonetheless, the N-center laser represents the first of a new class of aggregate center lasers that may open the door to new systems and applications in the future.

H. Summary of Laser Performance

Table I lists relevant data concerning the performance of the more common color center lasers. Since the operational lifetime is sometimes an issue with color center lasers, the approximate useful period of a single crystal is also listed. Unless otherwise noted, the powers are for cw operation.

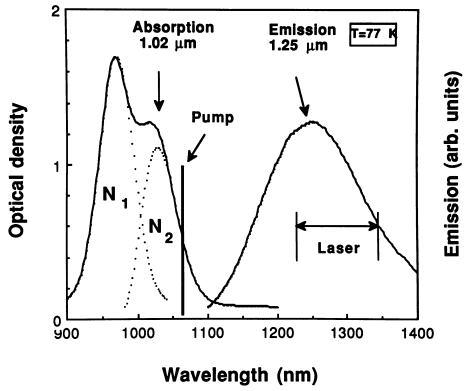


FIGURE 11 Absorption and emission bands for the N center in KCl. When pumped by a Q-switched 1.06-μm laser, the N_2 center lases from 1.23–1.37 μm.

III. FORMATION OF COLOR CENTERS

A question often asked is, How are these color centers formed? In this section, several proven techniques for forming laser-active crystals will be described. The first step in generating laser-active centers is to create a population of ordinary F-centers, either through additive coloration or radiation damage. Additive coloration is the preferred coloration technique because the F-centers produced are very stable, whereas radiation-damaged crystals usually require cryogenic storage at all times. However, some lasers, like the F_2^+ and the Tl°(1), can only be colored through radiation damage.

A. Additive Coloration

Additive coloration is achieved by diffusing a stoichiometric excess of alkali metal into an alkali halide crystal. This diffusion process can be achieved by placing the crystal in an alkali-metal vapor for a sufficient length of time. To get reasonable vapor pressures of metal requires temperatures on the order of 500–600°C. The equilibrium concentration for F-centers is directly proportional to the metal vapor density.

An excellent apparatus for coloring crystals is the heat pipe. A detailed description of the heat pipe would be out of place in this article, but an excellent review of the method is given by Mollenauer (see Bibliography). Briefly, the heat pipe maintains a zone of pure alkali metal vapor at a precisely controlled pressure. An uncolored crystal is lowered into the alkali vapor for 30–60 min, during which time excess alkali ions diffuse into the crystal until an equilibrium is established. To maintain charge neutrality, negative ion vacancies with electrons (F-centers) must diffuse into the crystal in equal concentration. The ultimate density of F-centers in the crystal is precisely controlled by adjusting the vapor pressure. The density of centers is often critical for optimized laser performance.

After coloration, the F-centers are converted into the desired laser-active center through controlled aggregation. The F-centers are made mobile by the illumination of the crystal with F-band light (light absorbed by the F-center). This optical excitation is usually carried out with nothing more sophisticated than an unfiltered light bulb. The light ionizes some of the F-centers, forming vacancies and free electrons. The free electrons become trapped elsewhere in the lattice, typically forming F^--centers. The vacancies wander through the lattice until they combine with another color center or with a substitutional impurity. Recapture of an electron (from one of the F^- centers) by the new center leads to the formation of an F_2 or F_A center, respectively.

The end product can be controlled through proper doping and temperature control during processing. If F_A or F_B centers are the desired end product, the host lattice should have a higher density of foreign ions than it has F-centers. The wandering anion vacancy will then most likely run into a trapping impurity before it runs into another F-center. Temperature control is critical: if $T < -50°C$, the vacancies become immobile, while at $T > 0°C$, the vacancies will not attach once they meet an impurity. If the crystal temperature is kept within these limits, an equilibrium population of F_A or F_B can be formed in about 30 min.

To create a more complicated center, such as an $(F_2^+)_A$, a multistep process is usually followed that must be empirically determined for each lattice and center. In KCl doped with Li, for example, a high density of F-centers is first put into the crystal through additive coloration. Aggregation at $T = -20°C$ then leads to the formation of large populations of F_2 centers as well as F_A centers. The binding energy for an F_2 center beside a substitutional impurity is less than the thermal energy of the lattice at these temperatures, so few F_{2A} centers are formed in this first step. The crystal is then cooled to $T = -70°C$ and illuminated with F-light. At this temperature, the F-centers are nearly immobile, but the F_2 centers are excited by the F-light, and relax by reorienting (see Fig. 8). Subsequent excitation–relaxation cycles lead to a random walk of the F_2 centers through the lattice, and eventually they are trapped at individual impurity ions, forming F_{2A} centers. Cooling to 77 K and exposing the crystal to UV light ionizes the F_{2A} centers, forming the desired $(F_2^+)_A$ centers. The free electrons are conveniently trapped by the residual F_A centers.

Both the F_A and $(F_2^+)_A$ centers dissociate if the crystal temperature is raised to ~300 K. In fact, if left at room temperature for a long time (days) the F-centers will form large aggregate centers called colloids. The colloids are easily dispersed by briefly annealing the crystal at the coloration temperature for about 2 min. Due to this ability to be regenerated, additively colored crystals are considered

to be room-temperature storable for essentially infinite periods, with the proviso that the crystal may require annealing and reaggregation prior to use.

B. Radiation Damage

Radiation damage is a simple technique for creating color centers. No matter what source of radiation is used, whether it be X-rays, γ-rays, or high-energy electrons, the primary effect is to produce electron-hole pairs. In the alkali halides, the incident radiation strips the electron from a negatively charged halogen ion. Eventually the electron returns home, with 5–10 eV of kinetic energy. The released energy, in combination with thermal energy from the lattice, is usually enough to eject a halogen atom into an interstitial position in the lattice, leaving the electron behind in the vacancy to form an F-center. If irradiation is carried out at low temperatures ($T < 170$ K), the vacancies and interstitials will remain frozen in place. But, if the temperature is raised sufficiently, the anion vacancies will migrate. Should it meet an interstitial, the two defects will annihilate each other, leaving a perfect lattice behind. Thus radiation-damaged crystals must be stored at cold temperatures during and after coloration.

Radiation damage is the only demonstrated way to make F_2^+ and Tl° (1) lasers. To form the F_2^+ center, the crystal is irradiated at about −100°C, then allowed to warm to room temperature for a few minutes, during which time thermally mobile anion vacancies aggregate with F-centers. This warm aggregation process conveniently occurs during the mounting of the crystal in the laser: the crystal must be warmed to room temperature to prevent condensation or frost forming on the crystal surfaces. After mounting, the crystal is cooled to 77 K for use. It must remain below 170 K from then on.

The Tl°(1) center is processed in a similar manner, except with less severe temperature restrictions. After irradiation, the crystal must be kept below 0°C, which can easily be achieved in a home freezer. The steps involved in warming the crystal to mount it are sufficient to aggregate the F-centers beside the Tl ions. White light at −10°C for 30 min completes the aggregation process. From then on, the crystal must be stored below 0°C to insure reliable laser operation.

IV. OPTICAL GAIN FROM COLOR CENTERS

A. Stimulated Emission Cross-Section

Color centers form an ideal four-level system for laser operation. The energy level diagram of the F_2^+ center in Fig. 5 shows a good example of this. The relevant energy levels for laser operation are labeled 0 through 3, representing the ground state (0), the first excited state (1), RES (2), and the lower laser level (3). The laser depends on stimulated transitions between levels 2 and 3.

Optical gain is usually defined in terms of a stimulated emission cross-section, σ, which can be calculated from readily measurable quantities:

$$\sigma = \lambda^2 \eta / \left(8\pi n^2 \delta\nu \tau_{\text{obs}}\right),$$

where τ_{obs} is the measured radiative lifetime of the center, η is the quantum efficiency of the emission process, n is the index of refraction, λ is the wavelength of emission, and $\delta\nu$ is the full width at half maximum (FWHM) of the luminescence. For color centers, η is defined as the number of quanta emitted per quanta absorbed. A value less than unity implies that a nonradiative process is occurring in addition to spontaneous emission. Table II lists typical values of τ, η, $\delta\nu$, and σ for a few laser active color centers.

The gain coefficient $\gamma(\text{cm}^{-1})$ for a laser medium is related to the cross-section by

$$\gamma = \sigma(n_2 - (g_2/g_3)n_3),$$

where n_2 and n_3 are the population densities of the relaxed-excited and relaxed-ground state of the transition, respectively, and g_2/g_3 is the ratio of degeneracies for the two states. The population of the upper laser level, n_2, depends on the pumping rate. For color center lasers, virtually all of the centers pumped from the ground state arrive in the relaxed excited state. If we assume cw pumping and no saturation of the ground state (i.e., $n_o \gg n_2$ at all times), simple rate equations show that the unsaturated population of n_2 is

$$n_2 = P\tau/h\nu,$$

where P is the power per unit volume absorbed by the crystal, $h\nu$ is the pump photon energy, and τ is the lifetime of the upper state. Due to the relatively short radiative lifetime of the color center, the pump intensity required to achieve a useful population in level 2 must be in the 10^5 W/cm^2 range. Such large intensities can most readily be achieved using a tightly focused pump laser beam. Most

TABLE II Laser Design Parameters

Center	λ_o (μm)	τ_{obs} (nsec)	η (%)	$\Delta\nu$ (THz)	σ (10^{-16} cm^2)	I_{sat} (kW/cm^2)
F_A(II)	2.7	200	40	15	1.7	9.4
F_2^+	1.5	80	100	30	1.6	45
$(F_2^+)_A$	2.3	170	100	20	2.7	8.2
F_2^+:O^{2-}	1.6	160	100	45	0.9	9.0
Tl°(1)	1.5	1600	100	15	0.2	21
N	1.3	210	~100	20	0.4	18

color center lasers are pumped by the Ar^+ ion or Nd:YAG laser.

The population of the lower laser level, n_3, decays rapidly into the normal ground state, 0, through lattice contraction. Like the $1 \rightarrow 2$ transition, the $3 \rightarrow 0$ transition is nonradiative, and occurs within 10^{-12} sec, so the residual population n_3 is essentially zero. The gain in a color center system thus depends primarily on the upper state population.

Energy extraction from a laser crystal involves the key process of stimulated emission. High intensity is needed to induce stimulated transitions and overwhelm the spontaneous emission from level 2. The characteristic intensity for a stimulated emission process is called the Saturation Intensity, I_{sat}, and is given by the equation

$$I_{sat} = h\nu/\sigma\tau.$$

At $I = I_{sat}$, the stimulated emission rate just equals the spontaneous emission rate. I_{sat} values for several types of color center are given in Table II. A laser is usually designed to operate with an intracavity intensity I in the range $I_{sat} < I < 10\, I_{sat}$. The lower limit of I is obviously for efficiency. The upper limit comes about from the reducing marginal increase in extraction efficiency with increasing intensity.

A linear color center laser cavity is schematically shown in Fig. 12. The cavity has two arms, one with a tight beamwaist at which the color center crystal is placed, and one with a nominally collimated beam where tuning elements may be inserted. The crystal is oriented at Brewster's angle to minimize reflective losses. The tight focus at the crystal establishes the high intensity needed for efficient saturation of the population inversion. The spot size at the beam focus, designated the beamwaist parameter ω_o, is typically in the range of 20–35 μm for color center lasers.

The beam size expands due to diffraction as the light propagates away from the waist. The characteristic distance over which the beam radius remains smaller than $\sqrt{2}\,\omega_o$ is called the confocal parameter, b, given by

$$b = 2\pi n\omega_o^2/\lambda,$$

where n is the index of refraction, and λ is the wavelength of the beam. Because efficient stimulated emission requires high intensity, the gain media outside of the confocal parameter makes little contribution to the overall power. Therefore, color center laser crystals are usually made no thicker than b, with typical values of b falling in the 1–3 mm range.

Pump absorption is determined by the Beer's law expression

$$P_{abs} = P_{in}(1 - e^{-\sigma_g N_o l}),$$

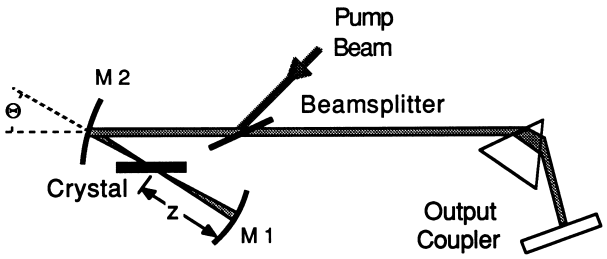


FIGURE 12 Schematic of a linear color center laser. The beamsplitter is dichroic, reflecting the pump wavelength and transmitting the color center wavelength. Tuning is accomplished with a prism by tilting the output coupler.

where σ_g is the ground-state absorption cross-section (which is approximately the same size as the emission cross-section for the laser-active color centers), N_o is the ground-state population, l is the crystal thickness, and P_{in} and P_{abs} represent the input and absorbed pump power, respectively. Efficient pumping demands that most of the pump power be deposited in the crystal, namely $e^{-\sigma_g N_o l} < 0.1$. However, if $N_o l$ is too high, the pump power will be absorbed in a thin region, causing severe local heating problems. Good results have been obtained with $N_o l$ set to give 80–90% absorption per pass.

A clever aspect of the folded cavity shown in Fig. 12 is that it is astigmatically compensated. A Brewster-angle crystal introduces a certain amount of astigmatism which is exactly opposite in sign to the astigmatism introduced by the off-axis folding mirror. The total astigmatism can be eliminated by choosing the proper angle of incidence, θ, of the beam on the folding mirror.

B. Color Center Laser Cavity

Since the color center laser is operated at cryogenic temperatures, the crystal must be enclosed in a vacuum chamber to provide thermal insulation and to prevent condensation on the crystal surfaces. The focusing optics, M_1 and M_2 of Fig. 12, are usually located inside the vacuum chamber to avoid the astigmatism effects of windows. These mirrors must be carefully prealigned prior to operation. The crystal is usually mounted on a translatable cold finger so that the position of the beamwaist on the crystal can be adjusted. In this way, any local defects or scratches in the crystal can be avoided.

The collimated arm, sometimes called the *tuning arm*, is directed out of the main vacuum chamber through a Brewster's angle window. Tuning elements can then be easily placed and adjusted in the tuning arm without disrupting the vacuum surrounding the crystal. The tuning arm is often separately evacuated as well in order to avoid absorptive losses from H_2O and CO_2 bands in the 2.7-μm region.

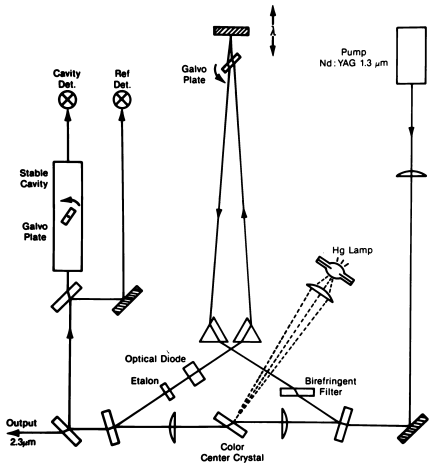


FIGURE 13 Actively stabilized color center ring laser. The wavelength is adjusted by translating the top mirror. The Hg lamp is for $(F_2^+)_A$ center operation.

C. Optical Pumping Techniques

Due to the thickness of the gain media, it is necessary to pump the crystal coaxially with the cavity mode to achieve reasonable efficiency. Several schemes have been employed for injecting the pump beam into the cavity. One method involves the use of dichroic mirrors, where one of the cavity mirrors is given a coating that is transparent to the shorter wavelength pump laser, but reflective for the color center laser wavelength. An example of this technique is shown in the ring laser of Fig. 13. In this example, the pump beam enters the cavity through an end mirror, and the cavity lens focuses both the laser mode and pump beam onto the crystal.

A second popular method is to use a Brewster's angle beamsplitter with a special coating that transmits nearly 100% of the color center laser radiation, yet reflects nearly all of the pump radiation. This technique is shown in Fig. 12. The injected pump beam is focused onto the crystal by the curved mirror M_1, and power that is not absorbed on the first pass through the crystal is reflected back by mirror M_2 for a second pass. One advantage of this scheme is that the mirrors can be coated with broadband enhanced silver, which permits one set of mirrors to operate over a broad wavelength range.

V. COLOR CENTER LASER OPERATION

A. Broadband Tuning Techniques

Color center emission bands are homogeneously broadened by lattice vibrations. This broadening leads to the wide tuning range of the color center laser, and its homogeneous nature allows for straightforward single-mode operation. However, this same broad tuning range makes it nearly impossible to control the output wavelength with only one tuning element. Generally, a hierarchy of dispersive elements are required for single-mode operation. These elements include a coarse tuner for general wavelength selection, an intermediate dispersive element such as an etalon for individual mode selection, and a cavity length adjustment for continuous frequency scanning of the individual mode. For mode-locked operation, however, it is only necessary to use a coarse tuner.

A Brewster angle prism is often used as a simple and inexpensive tuning element. The prism has the advantage of being low loss and monotonic in wavelength transmission; however, it has the disadvantages of having relatively low dispersion and deviating the direction of the beam as the wavelength is changed. The prism is generally not used in favor of other elements, such as the grating or birefringent tuner.

The Littrow-mounted diffraction grating has very high dispersion. The Littrow orientation causes the refracted beam to retroreflect along the incident beam. If the blaze angles are chosen correctly, gratings can exhibit first-order retroreflections of $>95\%$. A practical method of using a grating tuner is shown in Fig. 14. The first-order reflection feeds back into the laser and tunes it according to $\lambda = 2d \sin\theta$, where d is the groove spacing, and θ is defined in the figure. The zeroth-order reflection serves as the output coupling, and is bounced off a second mirror adjusted such that a constant deviation of 180° is achieved with respect to the input beam. A small pick-off mirror then directs the beam away from the cavity. The laser wavelength can be linearly scanned using a "sine bar," where motion in the x direction causes a linear change in $\sin\theta = x/h$. The dispersion of the grating is given by

$$\delta\lambda = 2d \cos\theta\, \delta\theta,$$

where $\delta\theta$ is the laser beam diffraction angle. It has been found that a laser ceases to operate when the returning beam is misaligned by the angle $\Delta\theta = \lambda/40\omega$, where ω is the spot size of the beam. Assuming a 1-mm beam spot size, $\lambda = 1.5\ \mu$m, and $d = 600\ \text{mm}^{-1}$, we find $\delta\lambda =$ 0.06 nm, which is about 30 times narrower than the bandwidth of a prism.

A moderately dispersive tuning element can be constructed using birefringent plates in the cavity placed at Brewster's angle to the beam. Tuning is accomplished by rotating the plates about their normal axis. A linearly polarized incident beam emerges in general as an elliptically polarized beam. This exit beam suffers substantial reflection losses at the Brewster surfaces of the cavity. However, for certain wavelengths the output polarization is linear

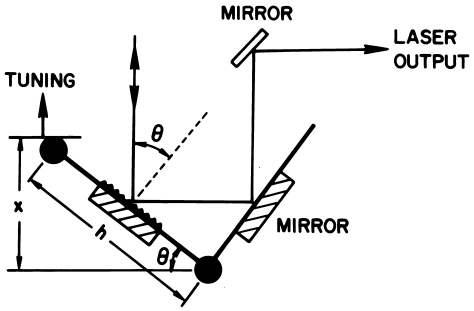


FIGURE 14 Littrow-oriented diffraction grating tuner for a tunable laser. The zeroth-order reflection is used for output coupling, while the first-order reflection is fed back into the laser.

and unrotated, and for these there is no loss. These eigenwavelengths vary with plate rotation angle θ as

$$\lambda = (\lambda_p/m)[1 - \cos^2\phi \sin^2\theta],$$

where $\lambda_p = (n_o - n_e)t / \sin\phi$, m is an integer, t is the plate thickness, and ϕ is Brewster's angle. The bandwidth $\delta\lambda$ is a complicated function of the number of polarizing surfaces in the cavity, but has a magnitude that falls somewhere between the dispersion of the prism and the grating. An advantage of the birefringent tuner is that it requires no displacement of the beam during tuning, in contrast to the prism or grating. One disadvantage of the birefringent tuner arises from the multiple orders that exist in a plate, making it possible for the laser to hop back toward line center when tuned to the extremes of the tuning curve.

B. Single-Mode Operation

It is straightforward to operate a color center laser in a narrowline single mode that can be tuned to any wavelength within the tuning range of the crystal. Such capability has proven useful for molecular spectroscopists, where the excellent frequency definition and tunability of the color center laser have allowed detailed studies of many important molecular species.

1. Standing Wave Linear Laser

Using a coarse tuning element, the homogeneously broadened standing wave laser oscillates in one primary mode, and one or two secondary modes caused by spatial hole burning. The frequency spacing of the secondary modes from the primary mode is given by

$$\Delta\nu_{hole} = c/4z,$$

where z is the distance between the end mirror and the active media (see Fig. 12). The physical mechanism causing oscillation at the hole-burning frequencies is illustrated in Fig. 15. The standing wave of the primary mode "burns" holes into the population inversion of the gain media, saturating the gain to the level where gain equals cavity loss. At the nodes of the standing wave, the gain is not saturated, and remains at values well above that needed to exceed threshold. A second mode that is spatially $\pi/4$ out of phase with the primary mode will perfectly overlap this periodic gain, and will see the excess gain. Unless mode-selective losses are included in the cavity, this second "hole-burning" mode will oscillate.

Single-mode operation in a linear laser is usually achieved by making the hole-burning spacing large (small z) and using a combination grating and etalon. An etalon is a Fabry-Perot optical resonator consisting of two plane parallel mirrors separated by a distance l. The etalon has a "picket-fence" of transmission peaks, equally spaced in frequency by the FSR:

$$FSR = c/2nl,$$

where n is the refractive index between the mirrors. The etalon selects one cavity mode of the laser, and the grating selects one order of the etalon. The optimum FSR is chosen to be twice the hole-burning mode spacing, $\Delta\nu_{hole}$. The single-mode output power is approximately 70% that of the multimode laser. This power drop is due to the loss of the hole-burning mode.

Scanning the etalon and coarse tuner in tandem, the laser will tune in steps of the cavity FSR, $c/2l_c$, where l_c is the cavity length. Typical cavity FSRs are on the order of 300 MHz. Tuning the laser frequency between adjacent cavity modes can be achieved by smoothly increasing the cavity length by $\lambda/2$. This is usually accomplished with a piezoelectric transducer (PZT)-driven cavity mirror or a tilted Brewster angle plate. To achieve smooth continuous scans, the three tuning elements must be synchronized.

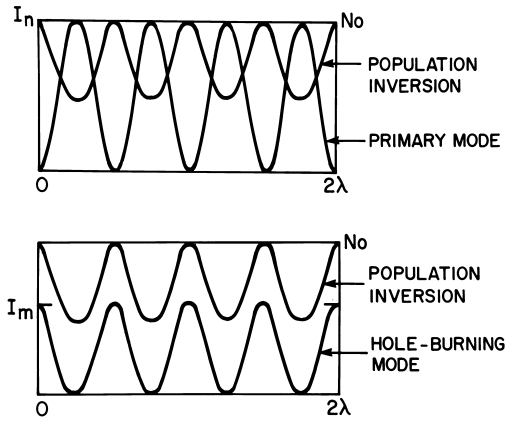


FIGURE 15 Plot of the spatial gain profile in the crystal for a standing wave laser. The antinodes of the standing wave saturate the gain, while at the nodes the gain remains high.

2. Traveling Wave Ring Laser

The ring laser overcomes the hole-burning limitations of the standing wave laser by operating in a traveling wave mode, allowing uniform saturation of the gain media. This has two advantages: first, single-mode operation is more efficient, since one mode can extract all the power, and second, because there is no gain for other modes, low finesse optics can be used for tuning.

Figure 13 shows a schematic of a frequency stabilized ring laser. Instead of curved mirrors, two antireflection-coated lenses are used for focusing the cavity mode into the color center crystal. In this scheme, the lenses serve both as focusing elements and as vacuum barriers for the evacuated region surrounding the crystal. Astigmatism is compensated by tilting the lenses at a small angle to the beam axis.

A unique property of ring lasers is that the direction of the traveling wave is not determined unless some form of biasing is introduced. Commonly, a combination of a *Faraday effect* rotator and an optically active plate is employed to force oscillation in only one direction. This combination is called an *optical diode*. The Faraday effect is due to magnetically induced birefringence that nonreciprocally rotates the polarization of transmitted light through a small angle

$$\theta = \mathrm{VlH},$$

where V is the Verdet constant of the material, l is the length of the device, and H is the strength of the magnetic field inside the material. At wavelengths longer than 1.5 μm, YIG (yttrium-iron-garnat) is used for the Faraday rotator because of its large Verdet constant. YIG becomes lossy below 1.5 μm, so other Faraday rotators similar to YIG must be used for the shorter wavelength applications. Following the Faraday device, an optically active plate is used as a reciprocal rotator. Depending on the direction of travel, the reciprocal device will either rotate the polarization to a further degree, or for oppositely directed waves, rotate the polarization back to its original state. Light whose polarization has been rotated will suffer loss at polarizing Brewster surfaces, while unrotated light will see no additional loss. This small direction-dependent loss is enough to force the laser to oscillate in one direction only. Typical rotation values are $\sim$2%.

The linewidth of the single-mode laser is determined by fluctuations in the optical path length of the resonator. These fluctuations are caused primarily by mechanical instabilities of the optical mounts, and by temperature variations of the color center crystal. The mechanical structure of the laser cavity is perturbed by floor vibrations and acoustics, both of which can be minimized with proper design to yield a laser linewidth on the order of 10 KHz. Pump-power amplitude noise is a major source of frequency instability. Changes in pump power cause temperature variations that alter the index of refraction of the crystal, effectively modulating the cavity length. The magnitude of this thermal tuning has been measured to be around 20 KHz per milliwatt change of pump power. For example, a pump power of 1 W with 1/2% amplitude noise will have a minimum linewidth of 100 KHz.

Active frequency stabilization can be applied to color center lasers to yield spectacularly narrow linewidths. The ring laser in Fig. 13 shows one scheme for active frequency stabilization. A portion of the output is directed through a thermally and acoustically isolated Fabry-Perot optical resonator. Comparing the frequency of the laser with the stable transmission passband of a reference Fabry-Perot cavity yields an error signal that is proportional to the frequency excursion. This elctronically amplified error signal is used to alter the laser cavity length through a PZT-driven mirror and a Brewster angle plate. The PZT mirror compensates for small high-frequency excursions, while the Brewster plate is galvonometrically tilted to control low-frequency noise and drift. The laser in Fig. 13 achieved a stabilized linewidth of less than 4 KHz.

C. Mode-Locked Operation

Color center lasers are relatively easy to mode-lock by synchronous pumping. In view of the large homogeneous gain bandwidth, the ultimate limit on pulse width with mode-locked color center lasers are under 1 psec. Mode-locking refers to the phase-locking of hundreds of adjacent longitudinal cavity modes. In synchronous pumping, the cavity length L of the color center laser is adjusted such that the cavity round-trip time, 2L/c, corresponds exactly to the period between pulses of the mode-locked pump laser. The dramatic temporal gain modulation produced by such pumping excites many sidebands, whose frequency separation corresponds to the color center cavity mode spacing. The sidebands build up in intensity through the homogeneous gain of the medium.

In practice, synchronous mode locking is achieved by pumping with a mode-locked laser, such as an ion laser or Nd:YAG laser, and adjusting the cavity length of the color center laser until the above length condition is satisfied. It is generally necessary to add a low-dispersion tuning element to the mode-locked cavity in order to obtain transform-limited pulses (i.e., the product $\Delta\nu\Delta\tau$ is minimized, where $\Delta\tau$ and $\Delta\nu$ represent the FWHM of the temporal width and bandwidth of the pulse, respectively). With no tuning element, the pulses will have excess frequency bandwidth, but will not necessarily be any

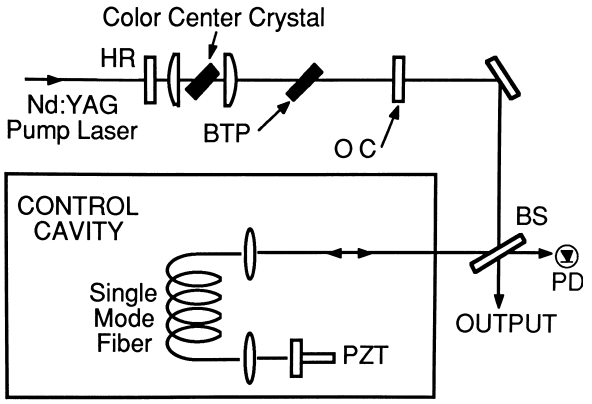


FIGURE 16 Additive pulse mode-locked laser. The output of the color center laser is coupled to an optical fiber. Self-phase modulation in the fiber adds bandwidth to the pulse, which is coupled back to the laser. The combination produces femtosecond-duration pulses. (OC, output coupler; BS, beam splitter; BTP, birefringent tuner plate; HR, high reflector; PZT, piezotransducer).

shorter in duration. Such nontransform-limited pulses can lead to anomalous results and excess dispersion in applications. A prism or single-plate birefringent tuner is thus often employed in mode-locked color center lasers. Typically, synchronous pumping mode locking has generated 5–15 psec pulses from the various color center lasers.

Several new mode-locking techniques have recently been introduced that provide subpicosecond pulses from color center lasers. The first method uses passive mode-locking through the introduction of a suitable semiconductor saturable absorber in the laser cavity. To date, 200-fsec pulses have been generated at fixed wavelengths in the 1.5- and 2.7-μm region.

The second technique is called additive pulse modelocking (APM). Figure 16 shows a schematic representation of an additively mode-locked laser. The laser consists of two coupled cavities, one with the color center gain media, and the other containing a single-mode optical fiber. A portion of the mode-locked output is coupled onto the optical fiber, where it experiences a nonlinear effect called self phase modulation (SPM). SPM effectively adds bandwidth to the pulse by red-shifting the leading edge, and blue-shifting the trailing edge of the pulse. The two cavities are made equal length so that when the broadened pulse is coupled back into the laser it interferes with the cavity pulse. The two pulses destructively interfere in the wings, leading to a shorter pulse being reinjected back into the gain media. This shortening process repeats with every round trip until a bandwidth limitation is reached, such as due to a tuning element. To date, 75-fsec pulses have been generated that are transform limited and tunable from 1.48 to 1.75 μm, making this the first tunable femtosecond source. APM techniques are now being extended to other sources, such as Ti:sapphire and Nd:YAG lasers.

VI. FUTURE DEVELOPMENTS

The color centers unique ability to generate broadly tunable light in the infrared insures its continued usefulness, especially as a research tool for molecular spectroscopy. Over the past decade new solid-state laser materials have been developed that can provide broadly tunable coverage across the 1.1–1.6-μm region (notably Cr-doped forsterite and Cr-doped YAG). The use of new systems such as Cr-doped ZnSe has shown promise as a tunable laser in the 2.3–2.8-μm region. Like color center lasers, these require optical pumping, but they do not require cryogenic operation. It is clear that the color center laser is slowly being displaced by other systems in the near infrared. Two niches have not been filled yet by other systems. The first is the pulse amplifier that uses the extremely high gain that can be obtained for a color center crystal. Short, intense pulses used for nonlinear studies of solitons, optical fibers, and semiconductors are still generated using NaCl amplifiers. Second, the spectral region from 2.8–4 μm is a rich region for molecular spectroscopy due to the C-H stretch vibration, and currently the color center laser is the only high-resolution source for this type of work.

There is effort being put into stabilizing the F_2^+ center using various dopants, and this may ultimately result in powerful, room-temperature operational lasers in the 0.8–1.2-μm region, a region where there are presently few tunable sources. It is likely that stable, room-temperature color center lasers will be developed to fill this region in the next few years.

SEE ALSO THE FOLLOWING ARTICLES

CRYOGENIC PROCESS ENGINEERING • LASERS • LASERS, DYE • LASERS, EXCIMER • LASERS, ULTRAFAST PULSE TECHNOLOGY

BIBLIOGRAPHY

Farge,Y., and Fontana, M. P. (1979). "Electronic and Vibrational Properties of Point Defects in Ionic Solids," North-Holland Pub. Co., Amsterdam.

Geogiou, E. T., Pinto, J. F., and Pollock, C. R. (1987). "Optical properties and formation of oxygen-perturbed F_2^+ color center in NaCl," *Phys. Rev.* B **35,** 7636.

Geogiou, E. T., Carrig, T. J., and Pollock, C. R. (1988). "Stable, pulsed, color center laser in pure KCl tunable from 1.23 to 1.35 μm," *Opt. Lett.* **13,** 987.

German, K. R. (1986). "Optimization of F_A(II) and F_B(II) Color Center Lasers," *J. Opt. Soc. Am. B: Opt. Phys.* **3,** 149.

Ippen, E. P., Haus, H. A., and Liu, L. Y. (1989). "Additive pulse mode locking," *J. Opt. Soc. Am. B: Opt. Phys.* **6,** 1736.

Mollenauer, L. F. (1979). Color Center Lasers, *in* "Methods of Experimental Physics, Vol. 15B," C. L. Tang, ed., Academic Press, New York.

Mollenauer, L. F. (1985). Color Center Lasers, *in* "Laser Handbook, Vol. 4," M. Stitch, ed., North-Holland Pub. Co., Amsterdam.

Mollenauer, L. F., and Stolen, R. H. (1984). "The Soliton Laser," *Opt. Lett*. **29,** 13.

Mollenauer, L. F., Vieira, N. D., and Szeto, L. (1983). "Optical properties of the Tl^o(1) center in KCl," *Phys. Rev*. B **27,** 5332.

Pinto, J. F., Georgiou, E. T., and Pollock, C. R. (1986). "Stable color center laser in OH-doped NaCl tunable from 1.41 to 1.81 μm," *Opt. Lett*. **11,** 519

Lasers, Dye

T. F. Johnston, Jr.

Coherent, Inc.

F. J. Duarte

State University of New York, Brockport

I. Introduction
II. Quantum Chemistry and Physics of Dyes
III. Short-Pulse Dye Lasers
IV. Continuous-Wave, Single-Frequency Dye Lasers
V. Ultrashort-Pulse (Mode-Locked) Dye Lasers
VI. Specialized Dye Lasers
VII. Recent Developments

GLOSSARY

Amplified spontaneous emission (ASE) Background light from dye fluorescence, which in high-gain short-pulse dye lasers limits the achievable gain and must be suppressed between amplifier stages to leave a spectrally pure laser output.

Autocorrelation, background-free Method of measuring mode-locked laser pulse lengths, in which the pulse train is split into two beams, one of which is delayed, and in which the two beams are focused to overlap in a crystal whose phase-matching conditions permit frequency summing only when one photon is taken from each beam. The summed or doubled power, versus path delay, is proportional to the autocorrelation in time of the pulse with itself and so gives the pulse length, if the functional form of the pulse shape is known.

Brewster's angle That singular angle of incidence (the angle between the ray direction and the normal to the surface of a dielectric plate of index n) given by $\tan\theta = n$ for which there is 100% transmission (no reflection) of light whose electric field vibrates in the plane of incidence. This low-loss property makes Brewster angle windows and entrance faces very common in intracavity laser optics.

Cavity dumper Device for coupling single pulses out of the resonator of a mode-locked laser, consisting of an acousto–optic cell situated at a beam focus in the cavity. A transducer bonded to the cell receives a short burst of radiofrequency (rf) energy timed to the pulse, setting up sound waves that diffract the pulse at a slight angle to dump it out of the cavity. This is used both to reduce the pulse repetition rate to a manageable value and increase the energy per pulse.

Chromophore Series of alternating single and double bonds in a dye molecule (be it linear, branched, or cyclic linkages), which through chemical resonance give rise to laser states, strong visible absorption and color, and delocalized mobile electrons, making the dye molecule a good "antenna" for visible light.

Dispersive oscillator A tunable laser oscillator that incorporates dispersion elements such as prisms and gratings.

Etalon Fabry–Perot or parallel-plate interferometer of fixed plate separation (often the coated, parallel sides of a plate of glass), which when tipped at an incidence angle greater than the beam divergence acts in transmission as an intracavity frequency (or longitudinal mode) selecting filter.

Free spectral range (FSR) Frequency spacing between adjacent transmission peaks of a parallel-plate interferometer and, by extension, between orders of other multiply ordered frequency filters.

Jitter Average residual frequency excursion or linewidth of a single-mode laser, due to residual environmental perturbations of the cavity length.

Linewidth equation An equation used to quantify the spectral width of the radiation emitted by a pulsed tunable laser. The linewidth is equal to the divergence of the radiation divided by the total intracavity dispersion of the oscillator.

Mode locking Formation of a high-peak-power, repetitive train of short-output pulses from a laser of cavity length d by the imposition of constant or "locked" phases for the difference or beat frequencies (nominally $c/2d$) between adjacent longitudinal modes. Here c is the speed of light, and a modulator (mode locker) or saturable absorber placed in the cavity forces the phase locking. The Fourier transform to the time domain of this locked-mode spectrum gives a pulse of period $2d/c$, the round-trip transit time for a circulating pulse in the cavity, and pulse length equal to the inverse of the locked bandwidth.

Modes, longitudinal and transverse Terms for the discrete frequencies and associated intensity distributions supported by a laser resonator. At a mode frequency, an integral number of wavelengths fits in the path distance of a round trip in the resonator, for then the multiply reflected waves from many transits add up in phase, and oscillation will be fed back there. The integral number of on-axis wavelengths defines the *longitudinal mode* number q, and there are intensity distributions possible peaking off-axis, adding path length, altering the frequency, and producing interference nodes in the two directions transverse to the axis, whose integer numbers m, n label the *transverse mode*. As light waves are transverse electromagnetic waves, by analogy with the designations for modes in microwave cavities, the full designation for a laser cavity mode is TEM_{mnq}, although by convention the q is often supperessed, the lowest-order transverse mode (a single spot of Gaussian intensity profile) being designated the TEM_{00} mode.

Multiple-prism dispersion The cumulative dispersion of an intracavity multiple-prism assembly. This can be augmented or reduced according to the configuration of the prismatic array. For narrow-linewidth tunable oscillators (of the multiple-prism grating class) this dispersion is often chosen, by design, to be very low or close to zero. In that case the design is said to be "achromatic." For femtosecond lasers the multiple-prism array is configured to yield negative dispersion.

Narrow-linewidth emission Laser emission that is highly coherent or of extremely high spectral purity. Single-longitudinal-mode emission is inherently narrow-linewidth.

Phase matching Process of adjusting the extraordinary refractive index relative to the ordinary refractive index of a nonlinear crystal either by changing the temperature (temperature-tuned matching) or the direction of propagation of a laser beam relative to the crystal's optic axis (angle-tuned matching), to make the indices of the fundamental beam and its orthogonally polarized harmonic precisely match, within a few parts per million. Then the harmonic frequencies radiated from the microscopic volume elements along the path of the fundamental beam all add up in phase in the forward direction to produce a macroscopic harmonic beam.

Saturation intensity Intensity producing a sufficiently rapid stimulated transition rate to reduce the level population difference to half its initial value (and thus reduce the absorption or gain on the transition to half its initial value); the scale parameter in saturated absorption or gain saturation.

Singlet and triplet states Terms referring to the net or total electronic spin of a molecular state, a singlet being a spin of zero, and a triplet a spin of one quantum unit of angular momentum. (In a magnetic field the unity spin can be oriented parallel, antiparallel, or perpendicular to the field, giving rise to the term "triplet.") The distinction is important in dyes that have a singlet ground state, because the optical selection rules forbid strong transitions where the spin changes (spin is conserved), making the excited triplet state metastable and troublesome in designing dye lasers.

Solid-state dye laser A dye laser that uses a gain medium in the solid state. This medium can be a dye-doped polymeric matrix or dye molecules in a crystalline configuration.

Spatial hole burning Gain saturation by a single-frequency standing wave field proceeds primarily

locally at the antinodes of the standing wave, reducing the gain there or "burning spatial holes," a process of importance in mode selection in single-frequency lasers.

Spectral condensation Efficient spectral narrowing that occurs when a wavelength-selective filter is added to a dye laser cavity; the oscillating linewidth may be reduced many orders of magnitude, while the output power or energy is reduced typically by less than half.

Synchronous pumping Method of mode-locking a laser (here the dye laser) by gain modulation or repetitive pulsed pumping, where the round-trip transit time in the cavity of the pumped laser is adjusted to be precisely equal to (synchronous with) the period of a mode-locked pumping laser (here, usually an argon laser or doubled Nd:YAG laser).

Tunable laser A laser, such as a dye laser, capable of continuously tuning the wavelength (or frequency) of its output radiation. Continuous tuning means the smooth change of wavelength across the spectrum according to the controlled physical tuning of the oscillator.

Tunable laser oscillator A laser resonator, or cavity, integrating intracavity frequency selective elements such as gratings, etalons, or prisms, or a combination of these. Usually this definition applies to laser devices capable of narrow-linewidth, or single-longitudinal mode, emission.

Wave number The number $1/\lambda$ of wavelengths in a centimeter (units of cm^{-1}), a convenient spectroscopic unit of frequency because the free spectral range of an interferometer of plate spacing t (cm) is just $1/2t$ (cm^{-1}). To convert the wave number of a photon to its energy in joules, multiply by $hc = 1.99 \times 10^{-23}$ (h is Planck's constant, c the speed of light).

DYE LASERS are the light energy convertors of the scientific world, converting input pump light from flashlamps or fixed-wavelength pump lasers into tunable-output wavelengths, broad pumping bandwidths into narrow-linewidth outputs, and short pumping pulses into ultrashort-output pulses. To make these conversions they use optically excited molecular antennas—the conjugated double bonds of the strongly absorbing and flourescing organic substances commonly called dyes, the same substances that color clothes, photographic emulsions, and Easter eggs. Among the widely variable and highly desirable properties of the output beams from dye lasers are the broadest range of wavelengths from any laser (ultraviolet to infrared), the narrowest linewidths of common spectroscopic sources (10^2 Hz at center frequencies near 10^{15} Hz), and the shortest electrical pulses directly generated in a human-made device (10^{-14} sec).

I. INTRODUCTION

A. Dye Laser Types

Laser action in dyes takes place between the broad, diffuse energy bands characteristic of complex molecules in condensed liquid solvents. The tunability of the dye laser—the ability to smoothly vary the output wavelength over a broad range, much as the receiving wavelength of a radio may be continuously tuned—is a result of these broadened electronic states. A large number of organic dyes exist (more than 200 are actively in use as laser dyes), each fluorescing in a specific wavelength region, giving overlapping tuning ranges from which the laser's broad overall spectral coverage is constructed. The tunability of the dye laser is the basic attribute from which its other desirable properties are derived.

The diffuseness of the energy bands is evidence of the short lifetimes of the laser states, due both to the strongly allowed nature of the laser electronic transition and to the frequent collisions of each dye molecule with its neighbors. To invert such states takes large pumping rates, so far only achieved by optical means, with input light from a high-intensity flashlamp or a second, pumping laser. Pump lasers occur most commonly in scientific applications (the focus of this article), since with them the widest range of dye laser output properties are provided; flashlamps were important in the development of dye laser and are still the most economical pump sources.

The three main branches of the family of scientific dye lasers are distinguished by the types of their pump sources and by the applications appropriate to their resultant outputs. Pulsed dye lasers take advantage of the active laser medium being a liquid, to scale to large pumped volumes readily cooled by flowing the dye. From pulsed pump laser inputs, they produce high-output energies per pulse, typically tens of millijoules in a 10-nsec pulse, in a beam a fraction of a centimeter across. The resultant peak intensities, exceeding a megawatt per square centimeter, are large enough that the nonlinear terms may dominate in the response function of a dielectric sample placed in the beam. The intense electric field time modulates the index of refraction of the sample, to produce new light frequencies or other effects in the interaction revealing the structure of matter, and making up the body of applications termed nonlinear optics.

Continuous-wave (cw) dye lasers, pumped by continuous output lasers, take advantage of the efficient spectral narrowing that occurs in a dye laser when a dispersive element is added to the feedback cavity for wavelength control. In the first observation of this effect, a diffraction grating was substituted for an end mirror in an early pulsed dye laser. The term "spectral condensation" was coined to

describe the result, that the bandwidth of the output beam shrank to 1% of the original bandwidth, while retaining 70% of the original output energy. The spectral filters for today's cw dye lasers are designed to select and allow oscillation only on a single mode of the dye laser cavity. The narrowness of the spectrum of output frequencies is then determined by how stable is the frequency of that cavity mode, and frequency stabilization servos are employed to produce condensation factors smaller than 10^{-8}. These output bandwidths are less than 10^{-3} of the Doppler broadening width of typical spectral line in a gas. When incident on a gaseous sample, this light will produce a velocity-selected, "saturated" absorption—an absorption only by those atoms with the proper velocity component to be Doppler-shifted into resonance with the frequency of the laser beam. The selected absorption can be read with a second, pobe laser beam. By looking inside the Doppler width, spectral resolutions are reached that are finer by factors of 10^{-1} to 10^{-3} than was possible before, making up the body of applications termed high-resolution laser spectroscopy.

Mode-locked dye lasers take advantage of the broad spectrum over which the dye molecules can lase, to assemble a short pulse of light from oscillations at many dye cavity mode frequencies (tens of thousands of modes are made to oscillate with their phases locked together). The dye cavity or feedback resonator nominally supports oscillations only at the cavity mode frequencies $q(c/2d)$ (where q is a large integer of order 10^6, c is the speed of light, and d is the optical path distance between the cavity end mirrors). These are just the frequencies where the multiply reflected waves, generated in multiple transits of the cavity, reinforce and add up in phase. The frequency differences between adjacent cavity modes are all nominally $c/2d$, but ordinarily the phases of these difference (or mode-beating) frequencies are random. If the phases can be made constant, or *locked*, the Fourier sum of the phase-locked frequencies will produce a repetitive, high-peak-power spike of duration the inverse of the locked bandwidth, and of period of $2d/c$, (the cavity round-trip transit time) corresponding to a short pulse of light circulating back and forth in the cavity. Phase- or mode-locking is produced in two different ways. In passive mode-locking (which uses a cw pump laser), an absorber is introduced into the dye cavity (generally, this is a second dye), whose absorption lessens (saturates) at high input intensities. This favors a form of oscillation that maximizes the peak power incident on the absorber, to produce the greatest saturation, and pulse formation and mode-locking ensue. In active mode-locking, most commonly today of the form called synchronous mode-locking, the dye laser is pumped with a mode-locked pump laser. The optical cavity length of the dye laser is adjusted to precisely equal that of the pump laser, matching the transit times in the two cavities. This favors the formation of that short pulse in the dye cavity that optimally saturates the repetitive, pulsed gain from the pump pulse, and again mode-locking ensues. With the short pulses assembled in this way, it is possible (by saturation and probe techniques) to measure molecular reorientation times and trace energy flow paths in the photochemistry of molecules of biological or chemical interest. This body of applications is termed time-domain laser spectroscopy.

The dye laser has well fulfilled the spectroscopist's wish, expressed soon after the laser was invented, for an almost infinitely malleable laser whose wavelength, bandwidth, or pulse length could be varied as desired. Since its invention, the dye laser has evolved with or caused each advance in pump laser technology, in a 20-year record of increasing malleability. Practical, commercial dye laser systems convert 20–50% of the input light energy into variable, controlled-output light energy—an efficiency record that does not leave much room for other competing nonlinear optical processes to displace the use of dyes. As long as this historical trend of coevolution persists, it is likely that the dye laser will remain the premier instrument bringing a gleam to the spectroscopists' eye.

B. Historical Overview

The evolution of the scientific dye laser family tree is outlined in Fig. 1, which shows the year of introduction of each of the major ideas to be discussed in this article. The figure may serve as a guide in further literature study, to help find an early paper where the idea (then new) is thoroughly discussed. On the margins, indications are given of the pump lasers predominately in use, and on the right edge of each branch the performance specifications available by year are listed. For the linewidths (in hertz) shown for the cw branch, and pulse lengths (in seconds) shown for the mode-locked branch, these specification represent the best performance published through that year, as researchers generally could reproduce these results once the details were published. The average power specification (in watts) given for the pulsed branch is more problematical. For some experiments the energy per pulse is most important, in others it is the pulse repetition rate, and researchers (whose lasers may excel in only one regime, and not the other) do not always report both. Therefore this specification axis gives the average power available from commercial pulsed lasers by year. Flashlamp pumped dye lasers are obviously pulsed lasers, but are shown as a transition leading into the column to which they gave rise, the cw branch. The years 1966–1970, when the three main types were differentiated, are discussed here. A brief sketch is then given of the trends each branch subsequently followed, with more details in later sections.

Several publications in the first years of the 1960s suggested that lasing might be possible in optically excited

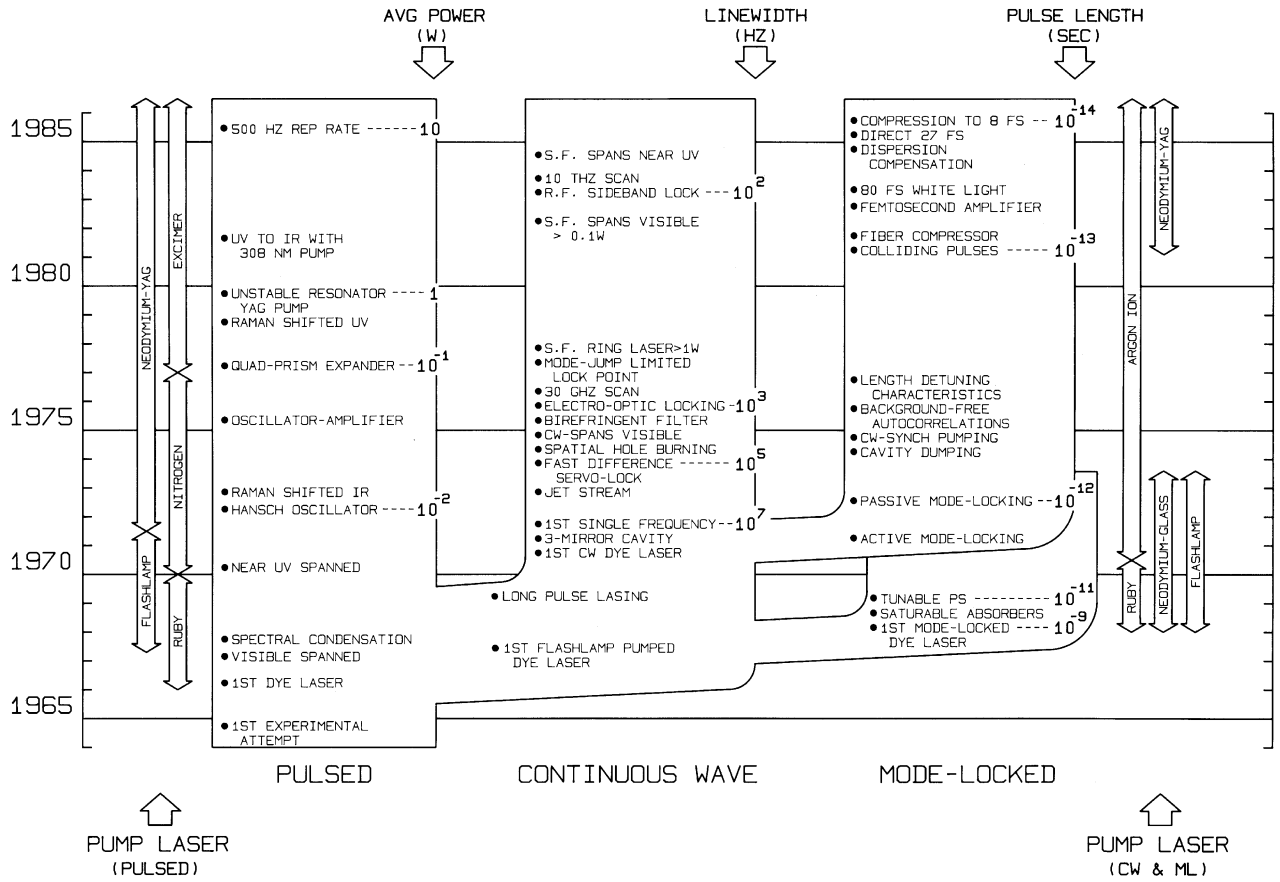


FIGURE 1 The historical development of the three main types of dye lasers used as scientific instruments today. The year of introduction in the literature is shown of ideas discussed in this text, along with an indication of the types of pump lasers predominately in use and the output specifications available from each type of dye laser.

organic dyes. The first experimental attempt to demonstrate a dye laser was in 1964, where a dye cell positioned between two cavity mirrors and filled with a solution of the dye perylene was illuminated by a flashlamp. This dye is now known to have high losses from excited-state absorptions, and this attempt was ended when there was no clearly positive result.

The first certain dye laser action was in 1966 in the work of Sorokin and Lankard at the IBM Research Center, Yorktown Heights, New York. They were interested in the spectral properties of the phthalocyanine dyes, a class of dyes then in use in *Q*-switch solutions for ruby lasers. (The normal temporally spiking output from a ruby laser may be converted to a single giant pulse, by adding a saturable absorber dye cell, or "*Q*-switch," to the ruby cavity.) When the dye chloroaluminum phthalocyanine was excited by the output of a giant-pulse ruby laser at 694 nm, Sorokin and Lankard observed an anomalous darkening of their spectral plate at 755 nm. This, they suspected, was due to stimulated emission as this wavelength coincided with a peak in the dye fluorescence curve, and they quickly confirmed this by adding resonator mirrors around the dye cell to generate an intense, well-collimated beam. By the end of the year, the same discovery had been made independently three more times, again in the United States by Spaeth *et al.*, in Germany by Schafer *et al.*, and in Russia by Stepanov *et al.*, all workers also studying *Q*-switch dyes stimulated by ruby laser giant pulses.

Dyes fluoresce and lase only at wavelengths longer than their excitation wavelength; thus, to extend the infrared emission of these first dye lasers to the visible required a shorter-wavelength pump source than the direct ruby laser. This was obtained by doubling the output frequency of ruby or of neodymium–glass (Nd:glass) lasers in the nonlinear crystals potassium dihydrogen phosphate (KDP) or ammonium dihydrogen phosphate (ADP), to give 347-nm (near ultraviolet) or 532-nm (green) pumping beams. Nearly two dozen new lasing dyes were quickly demonstrated with these new pump sources, spanning the visible spectrum with pulsed dye outputs by the end of 1967. These dye lasers were tuned by varying the concentration of the dye in the cell.

The spectral condensation effect was discovered that same year. A diffraction grating in the Littrow orientation

replaced one mirror of a two-mirror cavity, to reflect a dispersed fan of color back along the general direction of the incident light. Only a small band of wavelengths lay in the solid angle of the modes of the dye cavity defined by the grating and the second mirror. Not only was most of the original energy retained in the resulting narrow band output but also the wavelength of this output could be tuned (over seven times the previous range) by merely rotating the grating, a vast improvement in ease of tuning.

The requirements for successful flashlamp pumping of a dye could be determined from measurements on laser-pumped dye lasers, and this significant step followed in 1967 as well, making dye lasers accessible to many more scientists. What had not been sufficiently appreciated in 1964 was that the absorption overlapping the lasing wavelengths, arising from the excited, metastable triplet state of the dye molecule, could be largely avoided if the rise time of the flashlamp pulse were fast enough (less than a few tenths of a microsecond instead of the typical 10 μsec). Then lasing could occur before the build-up of an appreciable triplet population. Perylene was made to lase in 1972 with 20-nsec pump pulses from a *Q*-switched, doubled ruby laser.

The large-gain bandwidths demonstrated by the tuning ranges of these dyes prompted efforts to mode-lock the dye laser to produce tunable ultrashort pulses. Mode-locking was achieved in 1968 first by synchronously pumping with a mode-locked, doubled, Nd:glass laser, and then by introducing an absorber dye cell in the cavity of a flashlamp-pumped dye laser. The pulse lengths of these first systems were reported as being "detector limited," less than a half nanosecond, the rise time of a fast oscilloscope. By the end of that year the pulses had been shown by the two-photon fluorescence technique to be 10^{-11} sec in length, and the substitution of the diffraction-grating end reflector had shown these short pulses to be tunable.

There also was interest in generating longer pulses from dye lasers to approach cw operation, by removing the limits imposed by the absorbing triplet states. Careful measurements on long-pulse flashlamp-pumped dye lasers led to the conclusion that in some dyes, notably the rhodamines, the triplet population could be sufficiently quenched (returned to the ground state) by chemical interaction with oxygen from the air dissolved in the dye mix, that the residual triplet absorption did not in principle prevent cw operation. (Other chemical triplet-state quenchers were subsequently discovered.) What terminated the lasing of the dye before the end of the pump pulse, it was then realized, were the losses due to optical inhomogeneities in the heated dye. Peterson, Tuccio, and Snavely, at the Kodak Research Laboratories (Rochester, NY), designed a dye cell and pumping geometry to deal with these thermal problems after carefully evaluating the pump power that would be required to exceed cw laser threshold. They used a 1-W, 514-nm argon laser pump beam critically focused into a near-hemispherical dye cavity 4.5 mm long with a minimum transverse mode diameter, or waist, of only 11 μm. Water was used to dissolve the dye, because of its excellent thermo-optic properties, and the solution was circulated in a fast flow of 4 m sec^{-1} velocity across the pumped spot. Each dye molecule effectively saw a flash, since the transit time through the pumped volume was only 3 μsec. With this apparatus they successfully operated in 1970 the first cw dye laser, in rhodamine 6G dye.

Thus by the end of 1970 the three main branches existed. Developments in the following years carried the pulsed-branch lasers through a sequence of adaptions to ever more versatile and powerful pump lasers. First the nitrogen gas laser, with a 337-nm pumping wavelength, gave access to dyes lasing from the near ultraviolet (UV), through the blue and into the red, with one pump laser. This was then supplanted by the doubled and tripled Nd:YAG pump laser, giving higher pump energies per pulse (hundreds instead of a few millijoules) and the excimer pump laser, giving higher pulse repetition frequencies (hundreds instead of tens of pulses per second). In cw dye lasers, the technology first became easier with the introduction of the three-mirror cavity and free-flowing dye jet stream, which eliminated the critically toleranced and damage-prone dye cell, and then became harder again as frequency servo and control electronics of increasing sophistication were brought in to scan the single-frequency oscillation of ever narrower residual linewidths, over ever broader scan widths. Mode-locked lasers adopted the cw-type pumping format once this became available. The steady conditions allowed for a better characterization and definition of the regime where the mode-locked circulating pulse compresses on each transit of the cavity, and shorter pulses were produced. In this regime, the brevity of the pulse length is limited by the residual pulse-stretching effects in all of the optical elements the pulse sees, and the subsequent history is one of finding these and eliminating them ever since.

It is clear from Fig. 1 that the dye laser is still in a state of rapid evolution. What is also important in the figure, but is not perhaps immediately evident, is that the history of the dye laser gives a splendid example of the way a rich technology evolves. A new idea brings additional new ideas by extrapolation (and laser-pumped dye lasers led to flashlamp-pumped ones, which ushered in cw dye lasers). Improvements make the end device more useful, which leads to a component development allowing further improvements (excimer lasers, argon lasers, and Nd:YAG lasers have all experienced a decade of improving specifications through the demands made for better dye pump lasers). Performance only gets better, as to be adopted,

each new idea must produce results that surpass the old. After grasping more of the details of Fig. 1 (in what follows), it is worth looking again on the figure from these points of view.

II. QUANTUM CHEMISTRY AND PHYSICS OF DYES

A. Conjugated Double Bonds as Chromophores

The fundamental properties of a dye laser derive directly from the molecular structure of the dyes. Figure 2 shows the chemical structure in solution of the chloride salt of the most common laser dye, rhodamine 6G. Two equally probable forms [Figs. 2(a) and (b)] are shown, differing only in the location of the trivalent nitrogen atom. One form is obtained from the other by transfrerring an electron from one side of the molecule to the other, along the connecting chain of conjugated carbon bonds (alternating single and double bonds) in the ring structures. To maintain the correct +4 valence of each carbon atom in this transfer, the locations of four of the six double bonds in the three connecting rings must also be changed. In classical theories of dyes, the chain of shifted bonds was called the chromophore, and the light absorption and color of the dye were attributed to this oscillating dipole formed by the transfer of charge—the chromophore was a "molecular antenna." These theories arose from the empirical fact that essentially all organic dyes were found to contain conjugated double bonds, or, equivalently, were represented by a chemical "resonance" between nearly identical structures, differing only in the placement of the double bonds.

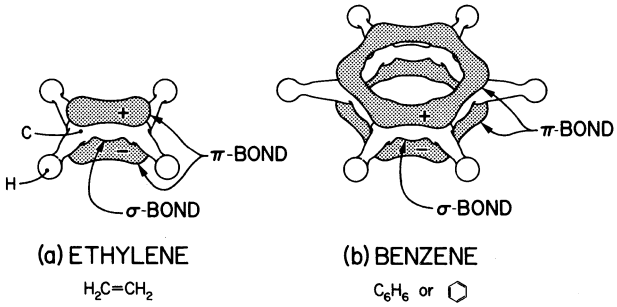


FIGURE 3 Chemical symbols and schematic three-dimensional representations of the electron density distributions, for (a) a double bond system, ethylene, and (b) a conjugated double bond system, benzene. The shapes show the surfaces of constant electron density except for distortions necessary for clarity in depicting the overlapping σ-bond and π-bond electron clouds.

The important features of this type of bonding are best grasped by building up to a complicated dye structure like Fig. 2, through the simpler molecules shown in Fig. 3. Ethylene, $H_2C{=}CH_2$, illustrates the character of the double bond in carbon (Fig. 3a). To "prepare" the carbon for bonding, three of the four $2s^2 2p^2$ outer electrons are hybridized into three sp^2 trigonal orbitals, having lobes of the wave function lying 120° apart in a plane (like a clover leaf). The fourth electron remains as a p electron with a wave-function node at the nucleus and the positive- and negative-phase lobes of the p orbital projecting perpendicularly from this plane. The first bond of the double bond is formed by overlapping one sp^2 orbital from each carbon atom, making a bond rotationally symmetric about the nuclear axis, called the σ bond. The remaining four sp^2 orbitals in ethylene are filled by making σ bonds to hydrogen atoms. The second bond in the double bond is formed by overlapping the two remaining p orbitals. The

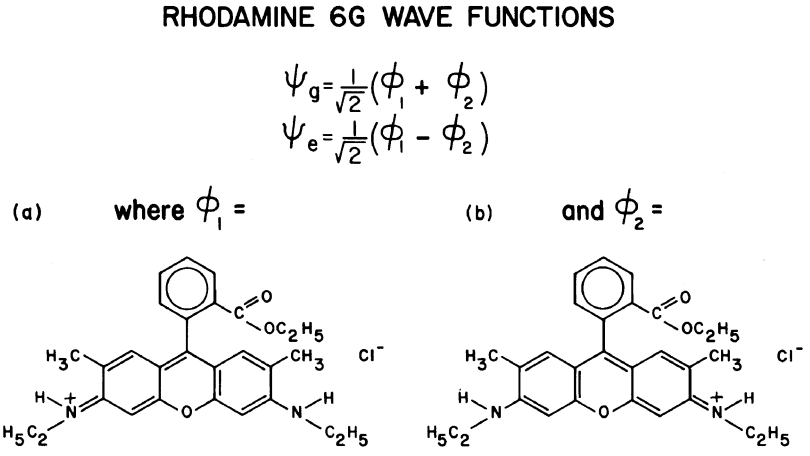


FIGURE 2 Chemical symbols for two "resonant" forms of the structure of the rhodamine 6G dye molecule and the wave functions to which this resonance gives rise.

molecular π bond thus formed retains the node, giving a pair of sausage-shaped lobes (of opposite phases) floating above and below a nodal plane containing all the nuclei and the center lines of all the σ bonds (Fig. 3a). The binding energy of the π bond is greatest when the atomic p orbitals overlap most (when their lobe axes are parallel). This explains the absence of free rotation about the double bond, and this and the sp^2 hybridization explain why the characteristic forms forms of dye molecules are planar structures built up from zigzag chains and hexagonal rings with 120° bond angles.

As a prototype conjugated molecule, consider benzene, C_6H_6 (Fig. 3b), six sp^2 carbons joined in a plane. X-Ray data show that all the carbon bond lengths in benzene are equal, and of intermediate length (0.139 nm) between a single (0.154 nm) and double (0.134 nm) carbon bond. By definition, there are only enough p-orbital electrons (one per conjugated atom) to make a second bond over half of the σ-bond carbon linkages. What happens, by symmetry, is that the six available p electrons are *shared equally* by the six carbons, linking them by six "half-bonds." They form two ring-shaped lobes or π-bond electron clouds in which the π electrons freely move, floating above and below the molecular skeleton of σ bonds lying in the nodal plane as shown in Fig. 3b. The characteristic of the conjugated bonding found in dyes is that it gives delocalized, mobile π-orbital electrons.

In quantum mechanics this resonance or sharing between possible structures is described by taking for the wave function of the actual structure a linear combination of the wave functions for the resonant possibilities. The energy of the system is calculated with the combined wave function, and the linear coefficients are determined as those that they give stationary values for the system's energy. For example, in combining the equally likely structures in Figs. 2a and b, this produces from the old wave functions Φ_1, Φ_2, each associated with the energy E, the two new sum and difference wave functions Ψ_g, Ψ_e of the figure, one associated with an energy less than E and one greater than E. The square of the absolute value of the wave function gives the spatial electron density in the molecule, In the additive combination $\Phi_1 + \Phi_2$, there results an averaging of the positions of the double bonds of the resonant forms, essentially placing an extra half bond everywhere along the chromophore. This reduced spatial specificity, by the uncertainty principle, implies a lower electronic kinetic energy, and the additive combination Ψ_g is the lower-energy state. The amount the energy is lowered over the nonresonant state is called the delocalization energy. Since the molecular energy is lowered, this combination it is termed a *bonding* state, or π state. In the subtractive combination $\Phi_1 - \Phi_2$, the nodes and antinodes of the wave function along the chromophore are emphasized, corresponding to localized electron density peaks and higher kinetic energy for the π electrons, and this is the higher-energy (excited) state. Since the molecular energy is increased, this is termed an *antibonding* state, or π^* state. Electronic states of stable molecules usually correspond to having more electrons in bonding states than in antibonding states.

By similar quantum-mechanical rules, starting with N conjugated atoms in a dye chromophore, each contributing one electron to the π system, it can be shown there will result N separate energy states for π electrons, each capable of holding up to two electrons (in a filled state the electrons must be paired due to the Pauli exclusion principle, with antiparallel spins). Half of these states will be bonding states and the other half-antibonding states. The π states in large unsaturated molecules are usually the highest-lying bonding levels and the π^* states the lowest-lying antibonding levels. Thus the normal electronic ground state for a dye uses the available N electrons, to just fill the $\frac{1}{2}N$ bonding π states with two paired electrons each. This gives a resultant zero spin to the system, and the ground state in dyes is a singlet state. The lowest-lying empty state is the first antibonding π^* state, the $\frac{1}{2}N + 1$ level of the system. The characteristic absorption and fluorescence of dyes, which is the working electronic transition in the dye laser is this $\pi(\frac{1}{2}N) \rightarrow \pi^*(\frac{1}{2}N + 1)$ transition.

The energies of these levels, and thus the center wavelength for that dye in a laser, may be calculated successfully with a surprisingly simple model, called the free-electron model. This assumes that the σ-bond framework of the conjugated chain provides a line or confining box of constant electrostatic potential along which the π electrons move freely. The important variables are the size of the box (number of conjugated atoms) and its topology (whether the bonding network is linear, branched, or ring structures). At the ends of the chromophore, the constant potential gives way to a steeply rising Coulomb potential, so the π electrons have the properties found in the quantum-mechanical problem of "a particle in a one-dimensional box." Their wave functions are sinusoidal standing waves with an integral number of half-wavelengths fitting into the length of the box, with their energies increasing quadratically with their quantum number (the number of half wavelengths). The electrons with more antinodes in their wave functions thus have larger kinetic energies. From these solutions, correct first-order values are obtained not only for the wavelength of the peak of the dye absorption (which in creases with the length of the chromophore in a linear molecule) but also for many other properties. Some of these are the oscillator strength of the transition (which is large, of order unity), the polarization dependence of absorption (which peaks for the incident electric field aligned along the chromophore), and

the shift of peak absorption wavelength with atomic substitution in the molecule, or with change of solvent.

It was shown already that the ground state for a dye molecule has zero spin and is a singlet. In the first excited state there are two unpaired electrons (one in the $\frac{1}{2}N$ bonding level, and the other in the $\frac{1}{2}N+1$ antibonding level). The Pauli exclusion principle does not restrict the spins of these electrons to be antiparallel, since they are in different levels, and a triplet state (parallel spins, for a total spin of unity) is possible in the excited electronic state. The free-electron wave functions can be used to show that due to spatial correlation effects between these unpaired electrons, the energy of the excited triplet state will lie below that of the singlet level. This gives rise to the problem of accumulation of dye molecules after optical pumping in the metastable triplet level, which was of such great importance in the historical development of dye lasers. (The triplet state is metastable or requires a collision to deactivate it, because a change in spin is not allowed in a radiative transition).

To summarize this, the quantum theory substantiates the classical view of mobile electrons, free to move along the chain of conjugated atoms in dyes, giving rise to oscillating dipoles and strong absorption and fluorescence. These are shown to be π-bond electrons, arising from the "resonance" or ambiguity in placement of the double bonds in the chain, which splits the electronic energy into a set of delocalized bonding levels and a set of localized antibonding levels. Laser operation and the strong color of the dye are due to the transition between the highest member of the low-lying set and the lowest member of the high-lying set, and are seen to be phenomena deeply imbedded in the quantum properties of matter.

B. Dye Absorption and Fluorescence Spectra Derived from Potential-Energy Curves

The basic spacing of the levels discussed above is about 20,000 cm^{-1} in rhodamine 6G dye. These levels are broadened by molecular vibration, rotation, and collisions with solvent molecules. Vibration produces sublevels, spaced by up to several hundreds of wave numbers, which in some dyes are sufficiently resolved to give secondary peaks in the absorption spectrum. Rotational and collisional (homogeneous) broadening smears these sublevels into overlapping bands to make the tunable continuum.

Vibrational broadening in molecular spectra is interpreted with a potential energy diagram such as is shown, somewhat schematically, for rhodamine 6G in Fig. 4a. A potential diagram has a clear interpretation for a simple diatomic molecule, where the structural coordinate R (the horizontal axis) is the separation between the two atomic nuclei. The characteristic shape of the potential energy curve is a dip or potential well, as the nuclei approach from infinite separation, by the amount of the binding energy (reached at separation R_0) before rising steeply as the nuclear Coulomb repulsion takes over at smaller separations. The ground-state separation R_0, at the potential minimum, fixes the normal linear dimension in a diatomic molecule. In a complex dye molecule, there are many nuclear separations that change in various directions with different modes of vibration, and the concept of a potential energy depending on a structural coordinate must be generalized. The ideas are the same as in the diatomic case, but the diagram becomes somewhat schematic. For Fig. 4a, the dimension for the structure coordinate axis was chosen (arbitrarily) as the distance between the nitrogen atoms in rhodamine 6G. The measured spectra (for wavelengths less than 700 nm) for rhodamine 6G are plotted in Fig. 4(c), and the schematic nature of the potential diagram is unimportant in this discussion since it was constructed to be consistent with these spectral data.

Each electronic state (ground state S_0, first excited singlet S_1, triplet T_1, and so forth) has a separate potential curve $E(R)$. As the molecular dimensions generally swell in the excited states, larger values for the structural coordinate R are found at the potential minima of the higher curves (a fact having important consequences for laser operation).

Vibrational energy in a molecule is quantized, giving a set of discrete levels shown in Fig. 4a as horizontal lines drawn between the boundaries of the potential wells. The square of the wave function for each of these levels gives the probability the nuclei will be found at a given value of R; three of these, for energies E_0, E_2, and E_3, are shown. These probabilities peak near the potential boundaries that correspond to the classical turning points in the vibratory motion, where the nuclear velocity is near zero and reverses direction. The intermediate points are passed through relatively quickly; the vibrating nuclei are likely to be found near their maximum or minimum positions.

Light absorption or emission is governed by the Franck–Condon principle, which states that the change in electronic energy occurs so much more rapidly than the vibrational motion that neither the position nor the momenta of the nuclei can change appreciably. Thus, on the potential-well diagram, the absorption (or emission) of a photon is represented by a vertical line, which, by the weighting of coordinate values, begins and ends near a potential-curve boundary for the strong transitions. (Transitions to endpoints well away from the curve boundaries are prevented by the second part of the Franck–Condon principle—there can be no discontinuous changes in nuclear kinetic energy.) The length of the vertical line is given by the photon energy (which matches the difference in electronic energy)

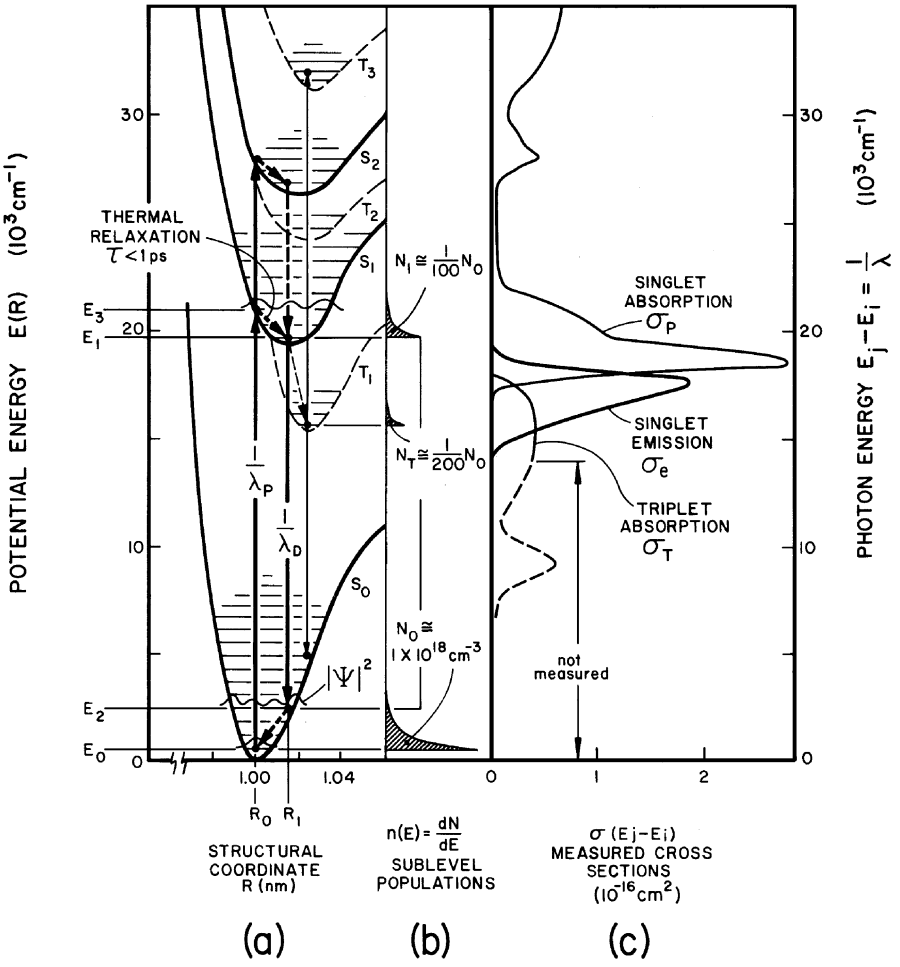


FIGURE 4 Schematic potential-energy wells constructed to be consistent with the measured spectra for the rhodamine 6G dye molecule. (a) The potential energy versus generalized structural coordinate (nuclear separation) for the first three singlet and first three triplet energy bands. Laser light absorption and emission (at pump and dye laser wavelengths) are indicated by heavy arrows, and nonradiative (collision) processes by dashed arrows. (b) Thermal equilibrium sublevel population distributions in these six bands (only the lowest three are appreciably populated). (c) Absorption and emission cross sections measured for this dye (except that the triplet absorption between 5000 and 14,000 cm^{-1} is inferred from spectra on similar dyes).

hc/λ, where h is Planck's constant, c the velocity of light, and λ the photon wavelength, or, if energy is measured in wave numbers as in Fig. 4, the length is directly $1/\lambda$. The line strength for the transition is given by quantum mechanics as an overlap integral, calculated with the two end-point vibrational wave functions.

The intensity of the transition is also weighted by the sublevel population density $n(E)$, the number of molecules per unit volume in the initial vibrational sublevel at energy E. The collisional and rotational broadening at room temperature spans the vibrational level spacing, so this sub-level population density has the form appropriate for Boltzmann (thermal) equilibrium (Fig. 4b) or $n(E)$ decreases sharply (falling exponentially) with energy above the bottom of the potential well. The product of this Boltzmann factor and the quantum-mechanical line strength is taken and plotted versus the energy difference of the potential curves, to produce the dye absorption and fluorescent emission curves (Fig. 4c).

These spectra are presented as cross sections for molecular absorption σ_p, or its inverse, stimulated emission σ_e. These are defined by Beer's law, which for example in the case of absorption of pump light from the ground state is

$$I_p(z) = I_p(0)\exp(-N_0\sigma_p z). \qquad (1)$$

Here a plane light wave of uniform intensity I_p (watts per square centimeter) and wavelength λ_p (nanometers) has been assumed, incident on a sample of dye of thickness z (centimeters) with a ground-state dye concentration of N_0 (molecules per cubic centimeter). A cross section has units

of centimeters squared and represents in the case of σ_p, for example, the absorbing area per molecule in the ground state. Because these cross sections are defined by an average over the Boltzmann population distribution in the initial state, they are temperature and solvent-dependent.

The total molecular population density N is the sum over all the populated states, which is for Fig. 4 where only the ground state, first singlet, and first triplet are appreciably populated:

$$N = N_0 + N_1 + N_T. \tag{2}$$

(Each of these state populations is the integral of the sublevel population density $n(E)$, over the energy range of several kT that the level populations are spread, k being Boltzmann's constant, T the absolute temperature, and $kT = 200\ \text{cm}^{-1}$ at room temperature. In Fig. 4b the energy spread is exaggerated to about twice this range, to allow the exponential fall-off to show). The assumption of a wave of uniform intensity here and in what follows simplifies the equations, and real devices with nonuniform waves can be treated later by integrating these results over the intensity variation within the active volume.

The two most striking features of the curves of absorption and emission versus photon energy are their asymmetric shapes and the fact that these are nearly mirror images of each other (Fig. 4c). These are consequences of the offset $R_0 < R_1$ between the potential minima of the S_0 and S_1 singlet states. The absorption from the ground state (labeled $1/\lambda_p$ in Fig. 4a) proceeds from the populated sublevels near the bottom of that well, to high-lying sublevels at points of minimum R along the steep inner boundary of the upper potential well. Hence, the σ_p spectrum shows a sharp rise, then falls with a more gradual tail, with increasing photon energy. Conversely, the fluorescence decay (5-nsec time constant) or stimulated transition (labeled $1/\lambda_D$ in Fig. 4a) arises from populated sublevels near the bottom of the upper well. These transitions go to the high-lying sublevels at points of maximum R along the descending outer boundary of the ground-state potential curve. Hence the σ_e spectrum rises sharply, then falls with a more gradual tail, with decreasing photon energy.

Other processes shown in Fig. 4a may be listed that compete with the main dye laser cycle of absorption at λ_p and emission at λ_D. Higher-energy pump photons can excite π electrons to the second singlet state directly ($S_0 \rightarrow S_2$, depicted as an extension of the arrow for the $1/\lambda_p$ transition, and as the second, higher-energy peak of σ_p). This excitation relaxes to S_1 by fast *internal conversion* processes (shown as dashed lines; these are collision-induced energy transfers on a picosecond time scale, without radiation). Internal conversion on $S_2 \rightarrow S_1$ allows pumping of the rhodamines and other visible dyes by ultraviolet pump lines such as the 308-nm line from the xenon chloride excimer laser. Internal conversion on $S_1 \rightarrow S_0$, which reduces laser efficiency, is also possible.

Some collisions produce *intersystem crossing*, converting the excited population in S_1 to that in the triplet state T_1 (indicated by a light dashed arrow). The triplet is a long-lived metastable state, as the radiative decay to the ground state $T_1 \rightarrow S_0$ is forbidden by selection rules that do not allow a change in spin in an electronic transition. Interactions with solvent molecules and torsional bending of the dye molecule produces a weak mixing of spin states, allowing a weak, delayed phosphorescence on this transition (thin solid arrow downward in Fig. 4a). Because this state is long-lived (microsecond to millisecond lifetimes), even a slow pump rate from S_1 due to collisions can result in an appreciable T_1 population, and thus also $T_1 \rightarrow T_3$ absorption, which overlaps the lasing wavelengths (thin solid arrow upward in Fig. 4a and absorption spectrum σ_T in Fig. 4c). Estimates of the triplet population can be made by various means, among them by noting the power increase in a dye laser by the addition of the liquid cyclooctatetrane (COT) to the dye mixture. This is a molecule with acceptor states lying below the first triplet level of many visible dyes, which by collisions relaxes (or quenches) the dye triplet population back to the ground state, reducing the deleterious absorption. From a typical cw laser output power increase of +30% in rhodamine 6G, this method estimates the T_1 state population to be half that of S_1. Another well-known triplet-state quencher is oxygen (O_2) dissolved from the air into the dye solution. These processes are described by a set of quantum efficiencies (branching ratios) φ, φ_T, φ_{IC} whose sum is unity, and that give the probabilities for a molecule put into S_1 to decay by each of the three channels. For good laser dyes, the fluorescence efficiency $\varphi = 1 - \varphi_T - \varphi_{IC}$ will exceed 85%, making these competing pathways unimportant (so long as the triplet absorption is quenched). With these definitions, if τ is the spontaneous emission lifetime for fluorescent $S_1 \rightarrow S_0$ decay, the total decay rate out of the first singlet state is $1/\tau\varphi$.

C. Spectral Condensation, Gain Saturation, and Laser Output Power

Referring to Fig. 4a, dye molecules optically pumped from the ground state to E_3 above the singlet potential minimum thermally relax to E_1 by collisions giving up heat energy to solvent molecules in a time on the order of a picosecond. This relaxation is faster than the optical absorption or stimulated emission rates in actual dye lasers. It is this thermal equilibrium process that, by maintaining the population distribution among the sublevels of each state at the Boltzmann values despite the perturbing optical transitions, allows the pumped dye molecules to exhibit gain and

work as a laser medium. The fluorescent decays labeled $1/\lambda_D$ into the ground state are into high-lying depopulated vibrational sublevels (near E_2), which remain so because of the ground-state Boltzmann equilibrium. This greatly reduces the pumping rate needed to establish gain in the dye. A pump rate sufficient to produce a singlet population N_1 of only 1% of that of the ground state N_0 is all that is needed to establish a population inversion between sublevels $n(E_1) > n(E_2)$, and gain in rhodamine 6G dye (see Fig. 4b). The thermal equilibrium process works in the laser for the pump absorption labeled in $1/\lambda_p$ as well, by rapidly removing population from the terminal sublevel E_3 to prevent stimulated emission at λ_p back to the ground state. This makes it possible to totally invert the working states in a dye laser (i.e., put all of the dye molecules into the excited state, $N_1 \gg N_0 \approx 0$). The Boltzmann equilibrium creates a classical four-level laser system out of the two broadened states.

It is also responsible for the spectral condensation effect. When broadband, nonselective feedback is provided in the dye laser, stimulated emission occurs from all occupied sublevels in the upper band S_1. The higher-lying sublevels (in the high-energy tail of the population distribution) emit primarily to higher sublevels in the ground state, (i.e., at longer wavelengths) than do those molecules near the potential minimum E_1. In this case the spectral bandwidth of the emission is broad, 10–20 nm in a pulsed dye laser. When a spectrally selective element is added to the dye feedback cavity, the threshold is raised for all emissions but those in the narrow selected wavelength band, say, the wavelengths emitted from level E_1, which are all that lase. The population in the entire upper state S_1 drains through this channel as the thermal collisions maintaining the Boltzmann equilibrium attempt (successfully) to fill the hole in the depleted sublevel population distribution. The same argument holds if the selective feedback is tuned for lasing at shorter or longer wavelengths, but the higher energy of the draining sublevel permits access only to the higher-energy tail of the Boltzmann distribution, and the process is not as efficient.

How much less efficient may be seen by calculating the steady-state laser output power for the case of a spectrally narrow tunable filter with center wavelength λ_D selecting the oscillation wavelength. The idea of gain saturation is the basis for an output power calculation. Assume uniform plane waves once again, with later integrations over the active volume accounting for intensity spatial variations in real devices. Initially, before lasing begins, the population inversion and the (small signal) gain per pass $k_0 L$ are at a maximum. Here $k_0 L$ is the exponential gain constant in Beer's law, and L is the dye optical thickness along the resonator axis.

This initial gain must decrease, or saturate, because as the lasing intensity I_e builds up, it causes stimulated transitions at rates comparable to the spontaneous fluorescent transition rate $1/\tau\varphi$, and the population inversion is depleted. Thus the gain constant $k(I_e)L$ is a decreasing function of I_e. In the steady state after the buildup transient, the laser intensity inside the resonator is the intensity that decreases the gain per pass to just balance the cavity losses, as then the laser intensity suffers no net change in a cavity round trip. For a ring (or traveling-wave) laser, the dye gain is traversed once per round trip of the resonator (of mean reflectivity r), so that balance is achieved if $re^{kL} = 1$, or

$$k(I_e)L = \ln(1/r). \tag{3}$$

This is the implicit equation for the steady-state laser intensity. Once it is solved for I_e, the output power P_{out} is given by

$$P_{out} = t\left(\pi w_D^2\right) I_e. \tag{4}$$

Here πw_D^2 is the area of the laser beam of (uniform) radius w_D in the dye medium. The resonator losses consist of the output coupling loss t (transmission of the output mirror) and the dissipative losses a due to light absorption and scatter in the intracavity optics, and the insertion losses of the wavelength selective filter. As t and a are usually small, $\ln(1/r) = \ln[1-a-t]^{-1} = a + t$ in Eq. (3).

The form of the gain decrease may be found by solving rate equations for the level populations, which balance in the steady state the rate of transitions into S_1 from pump absorptions ($S_0 \rightarrow S_1$) against the rate out of S_1 due to stimulated laser transitions and spontaneous fluorescent decays. The broadening of the laser states is included by the Boltzmann average in the cross-section definitions, so the rate equations deal with the total state populations N_1 and N_0. (Care must be exercised in the definitions and averaging to arrive at a consistent set of rate equations. Additional terms describing intersystem crossing, triplet and excited singlet state absorptions, and so forth, are not included here. In particular, it is assumed that the triplet population is quenched, $N_T = 0$.)

The essential physics is illustrated with just these four terms, in the time-independent, plane-wave, longitudinal pumping case. (The pump light propagates along the same axis as the dye laser beam.) The gain constant is given by

$$\begin{aligned} k(I_e)L &= (N_1\sigma_e - N_0\sigma_a)L \\ &= [N_1(\sigma_e + \sigma_a) - N\sigma_a]L, \end{aligned} \tag{5}$$

since $N = N_1 + N_0$ by Eq. (2). Here σ_a is the singlet absorption cross section, evaluated at the lasing wavelength λ_D. The steady-state balance of the transition rates into and out of the upper laser state gives a value $N_1(I_e)$ for

that population density that is dependent on the laser intensity I_e. When used in Eq. (5), there results a saturated gain equation of the form expected for a homogeneously broadened laser,

$$k(I_e)L = \frac{k_0 L}{1 + I_e/I_s}, \tag{6}$$

where the initial or small signal gain is

$$k_0 L = NL\left(\frac{\sigma_e\sigma_p I_p - \sigma_a(hc/\lambda_p)(1/\tau\varphi)}{\sigma_p I_p + (hc/\lambda_p)(1/\tau\varphi)}\right). \tag{7}$$

[Equation (7) corrects an error in the corresponding equation in the Johnston article.] The saturation intensity, which by Eq. (6) is the value of I_e dropping the gain to half its initial value, is

$$I_s = \frac{\lambda_p}{\lambda_D}\left(\frac{\sigma_p I_p + (hc/\lambda_p)(1/\tau\varphi)}{\sigma_e + \sigma_a}\right). \tag{8}$$

Setting the saturated gain equal to the resonator losses (asumed small) as in Eq. (3), solving Eq. (6) for I_e, and putting this into Eq. (4) gives the expected form for the output power (valid above threshold, when the expression in the last bracket is positive):

$$P_{out} = t(\pi w_D^2) I_s\left(\frac{k_0 L}{a+t} - 1\right). \tag{9}$$

The output increases linearly with the excess above unity or threshold value of the small-signal gain-to-loss ratio. Since the denominator of Eq. (7) is the same as the numerator of Eq. (8), it is convenient to consider the form the output power takes when the laser is pumped well above threshold (where these factors cancel in the product $I_s k_0 L$), and as a further simplification, at dye wavelengths in the middle of the tuning range or longer where the ground-state reabsorption may be neglected.

Then

$$P_{out} \approx \left(\frac{t}{a+t}\right)\frac{\lambda_p}{\lambda_D}(NL\sigma_p)\rho P_{in} \tag{10}$$

for $k_0 L/(a+t) \gg 1$ and $\sigma_a = 0$. Here $\rho = w_D^2/w_p^2$, the ratio of dye beam area to pumped area, is introduced and $\pi w_p^2 I_p$ is recognized as the input power in the longitudinal pumping case.

As a first use of these results, note the direct connection they establish between the experimental behavior of a dye laser and the underlying molecular physics. In using the laser, the ease of operation is greatest when well above threshold, and differences in behavior at opposite ends of the tuning range of a dye are immediately apparent. On the short-wavelength end of the range, there is a wide tolerance of output to mirror misalignment, dirty optics, etc., which persists even as the short-wavelength oscillation limit is approached and the output abruptly dies. The opposite is true as the long-wavelength limit is approached, where cavity alignment and losses are critical, and there always seems to be some output available at even longer wavelengths if only the critical adjustments can be maintained. The excess gain may be thought of as a measure of the tolerance to cavity losses, while by Eq. (9) the saturation intensity is a weighting factor, giving the value of this gain in determining output power. The behavior just described is thus due to the mirror-image and asymmetry properties of the absorption and emission cross sections. The short-wavelength end of the range is near the peak where the emission cross section σ_e is slowly varying, and the tuning limit is determined by the sharp rise in absorption cross section σ_a with decreasing λ_D. The rise in σ_a both decreases the saturation intensity [Eq. (8)] and adds a cavity loss term due to laser photon reabsorption in the dye itself [Eq. (7)]. The long-wavelength end of the range is determined by the asymmetric slow falloff of σ_e with increasing λ_D. This drops the gain toward the threshold (resonator) loss value, there is negligible reabsorption ($\sigma_a = 0$), and the rise in I_s [with $1/\sigma_e$, Eq. (8)] allows the output power to persist. An example where the $k_0 L$ and I_s parameters were measured as a function of λ_D in a cw dye laser is given in Section IV.

This linking of the macroscopic laser properties with the microscopic dye ones may be enhanced by further consideration of these results. In the gain expressions Eqs. (5) and (7) neglect the reabsorption term in σ_a, as is permissible in the middle of the tuning range. Then the small signal gain per pass of the dye medium is

$$k_0 L = N_1 L\sigma_e = NL\sigma_e\left(\frac{\sigma_p I_p}{\sigma_p I_p + (hc/\lambda_p)(1/\tau\varphi)}\right), \tag{11}$$

which is the product of the gain a uniform beam traversing the medium would see if all the dye molecules were in the upper state, at total inversion (NL being the molecular density in the beam cross section, σ_e the emitting area per molecule), times the fraction of molecules in the upper state (bracketed term). The interpretation of the bracketed term comes from Eq. (11), or directly from the rate equation and particle conservation equation from which it is derived.

The term $(hc/\lambda_p)(1/\tau\varphi)$ is the pump power dissipated (by all spontaneous decay processes) per molecule in the upper laser S_1. It is useful to estimate from Eq. (11) the minimum pump intensity needed to reach threshold in a dye laser. This will be when the pump power density absorbed by all ground-state molecules, $N_0\sigma_p I_p$, balances against the power density dissipated by upper-state decays, $N_1(hc/\lambda_p)(1/\tau\varphi)$, to give that excited-state fraction producing gain equal to the resonator losses.

From Eq. (11), this is

$$I_p^{th} = \frac{(1/\sigma_p)(hc/\lambda_p)(1/\tau\varphi)}{[NL\sigma_e/(a+t)]-1}. \quad (12)$$

For typical values $\sigma_p = \sigma_e = 1 \times 10^{-16}$ cm^2, $\lambda_p = 500$ nm, $\tau\varphi = 5 \times 10^{-9}$ sec, $N = 1 \times 10^{18}$ cm^{-3}, $(a + t)/L = 10$ cm^{-1}, this gives a threshold pump intensity of 90 kW cm^{-2}, clearly showing the need for bright flashlamps or a laser to pump a pulsed dye laser, and a coherent pump beam focusable to a small spot to pump a cw one. Note that in the example given, the pump power absorbed per ground-state molecule, $\sigma_p I_p$, is one-ninth the spontaneously dissipated power per molecule in the upper state. Gains and pump intensities 10 times the threshold value or more are realized in practical dye lasers.

The interpretation of the saturation intensity result, Eq. (8), contains a subtlety. In the conservative two-state system under discussion, a molecule removed from the upper state by laser-stimulated emission at the rate $\sigma_e I_e(hc/\lambda_D)$ per molecule must appear in the lower state. There it immediately is subjected to a pump rate (per molecule) of $\sigma_p I_p/(hc/\lambda_p)$ returning it to the upper state. Thus for stimulated emission to produce a reduction of the small-signal upper-state population by half, it must be at a transition rate per molecule equal to the sum of the spontaneous decay rate plus the return rate, yielding Eq. (8). This makes the saturation intensity a linear function of the pump intensity; at high pump rates, the dye bleaches, the small-signal gain saturates at the total inversion value $NL\sigma_e$, and the output power increases with pump rate solely through the I_s term in Eq. (9).

Turning now to Eq. (10), the first term is recognized as the coupling efficiency $\eta_C = t/(a + t)$ found in all lasers, expressing the fact that a finite circulating power must be left in the resonator to stimulate emission (not all of the power represented by the initial population inversion can be coupled out of the resonator). High-gain dye lasers with optimized output transmission t have $\eta_C \approx 0.8$. The next term is the Stokes efficiency $\eta_S = \lambda_p/\lambda_D$, accounting for the power lost (for a quantum efficiency of unity) due to the dye photon being of lower energy than the pump photon. The term $NL\sigma_p$ is the fraction of the pump light absorbed in the dye in this uniform plane-wave case, where this factor must be small to maintain the assumption of uniformity. When the pump intensity is allowed to vary along the axis of propagation, this becomes the fraction of pump intensity absorbed in the transit of the dye medium, $\eta_A = [1 - \exp(-N\sigma_p L)]$ by Eq. (1). The ratio ρ of lasing to pumped areas accounts for any pumped volume left unutilized and is replaced by an average value sometimes called the filling factor η_F, equal to a volume integral of this ratio over the pump and dye beam intensity variations in describing actual devices. The threshold term may be restored to Eq. (10), so that in its general form this may be rewritten to give the slope efficiency or ratio for conversion of pump power in excess of threshold to tunable dye laser output power:

$$P_{out}/(P_{in} - P_{th}) = \eta_C \eta_S \eta_A \eta_F. \quad (13)$$

Typical values, in the case of the rhodamine 6G cw single-frequency ring laser at the peak of the tuning curve and 6 W of input power, for the last three efficiency factors are $\eta_S = 0.85$, $\eta_A = 0.80$, and $\eta_F = 0.80$, giving a theoretical slope efficiency of 44%. Experimental values for this case reach 35%.

The question of the efficiency of the dye laser convertor and of spectral condensation in particular at wavelengths away from the maximum emission cross section is answered by Eq. (13). As long as the dye is pumped hard enough to create a gain well above threshold, the lasing conversion process occurs at fixed slope efficiency essentially independent of center wavelength or lasing bandwidth.

The most serious omission in the treatment given so far is the neglect of the triplet-state terms in Eqs. (7)–(13). The power loss due to triplet-state absorptions is treated by Peterson, who shows that the laser types are usefully separated by a demarcation time of $(K_{ST})^{-1} \approx 100$ nsec in rhodamine 6G, the inverse of the intersystem crossing rate filling the triplet state from the first excited singlet. This order of time is required to build up the equilibrium triplet population. For short-pulse dye lasers (laser-pumped, pulsed lasers) the pump pulses are typically 10 nsec in length and the triplet absorption terms can be neglected. The largest effect is for the cw type, where, for example, in the rhodamine 6G ring laser considered earlier, inclusion of triplet losses would reduce the theoretical small signal gain by 34%, increase the saturation intensity by 17%, and reduce the output power by 30%. To compute these changes requires knowledge of the triplet absorption cross section σ_T, which is known in this case but has not been measured for most other dyes.

In short-pulse dye lasers there is insufficient time for triplet-state buildup and the results above are usable without correction. In flashlamp-pumped dye lasers, not only are triplet terms significant, but often the full time-dependent rate equations are needed, requiring numerical integrations in modeling their behavior. In mode-locked dye lasers the resonator losses are deliberately time-dependent, requiring further discussion (see Section V).

The dye lifetime, or stability against photo decomposition, is another important dye property not in these equations. All dye molecules eventually decompose under repeated pump excitations. In the more stable molecules this occurs after 10^6–10^7 photon absorptions, corresponding

to about 10^4 J absorbed energy per cubic centimeter of dye solution, with perhaps a minimum of 1% of this lifetime required for a dye to be usable. In practice, a large volume of a liter or more of dye is circulated through a laser to give several hundred hours of operation between dye changes. Dye concentrations are adjusted to give the desired pump beam absorption depth, in matching the pumped volume to the dye beam diameter (fixed by the dye resonator). Dye concentrations are generally limited to less than 10^{-2} mol/liter, where the limit of solubility for some dyes occurs, and where aggregation (dimer formation) in others alters the dye spectra.

The main intent of the remainder of this article is to give a picture of the art of dye laser design in the three main scientific instrument categories, at the 1986 stage of their evolution.

III. SHORT-PULSE DYE LASERS

A. Status of Flashlamp-Pumped Dye Lasers

The high threshold pump intensity given by Eq. (12) means that flashlamps for dye laser pumping must be driven hard, blackbody temperatures above 6000 K being required, which results in short-lived, frequently replaced lamps. Much of the lamp radiation is in a broad ultraviolet band at these plasma temperatures, leading to relatively rapid photodecomposition of the dye if the lamp light is unfiltered, and inefficient operation if it is filtered. Lamp pulse lengths of 1–50 μsec are long enough that triplet-state quenching and thermal lensing of the dye must be dealt with. These pulses are too long (compared to the 5- to 30-nsec pulse lengths from pump lasers) to reach the peak dye powers most useful in nonlinear optics and available from laser-pumped dyes. These are the disadvantages of flashlamp-pumped dye lasers.

Nevertheless, because the flashlamp laser has the economy of eliminating the pump laser, many special-purpose lamp designs and dye lasers to go with them have been developed to fill special-purpose needs. There is also a considerable history of contributions to the dye laser art made with flashlamp pumping, ranging from the first demonstration of lasing in a large number of dyes, to clarifying the mechanisms of pulse formation in mode-locked dye lasers. But because of the disadvantages already listed, there are today few flashlamp-pumped scientific dye lasers, and no widespread general example of this type in the same sense that these exist for the three types represented as the three main branches (Fig. 1) of the dye laser family tree. The review article by Peterson expertly and extensively covers many flashlamp systems. Here only the short-pulse category of pulsed dye lasers will be discussed.

B. Short-Pulse Oscillators

Dye fluorescent lifetimes are too short (4–8 nsec) to provide energy storage, and for a high small-signal gain, a high-intensity pump pulse (but not a high energy one) is required. Short pump pulses (less than 100-nsec duration) offer the necessary intensity without adding extensive cooling demands and have the advantages of being too fast for triplet-state absorption or thermal lens distortions to develop in the dye. (The index of refraction of a liquid is dominated by its density, and a relatively high change of refractive index with temperature is a consequence of a relatively high rate of thermal expansion—which has a time constant on the order of 10 μsec, however, to develop.) The dye in a short-pulse system is simply circulated or stirred fast enough (about 1 m/sec) to remove energy-deposition temperature gradients in the few milliseconds between shots.

In the early development of pulsed dye lasers these reasons made the nitrogen (N_2) lasers then available attractive as dye pumps. The nitrogen laser has 4- to 10-nsec pulse lengths, 100-kW peak powers, a near-UV wavelength of 337 nm, and a repetition rate of 100 Hz, high enough to complete spectroscopic experiments in hours, (unlike the alternative doubled ruby laser pump at 1 Hz). The design principles for a narrowband spectroscopic short-pulse dye laser were first worked out by Hänsch in 1972 with this pump laser (Fig. 5a). The dye laser market thus established stimulated development of the more powerful Nd:YAG and excimer lasers for pumping use, which were to later displace the nitrogen laser (Fig. 1).

Hänsch realized that the low ratio of pump pulse length to round-trip transit time in the laser cavity dictated the design of the short-pulse oscillator. In a cavity long enough (here 40 cm) to accommodate a tuning device, a circulating, spontaneously emitted photon has only a few round trips during the excitation to replicate in an avalanche to $\sim 10^{15}$ photons (0.3 mJ at 600 nm), an output representing about 0.5 quantum efficiency of conversion of input photons. This necessitates a high gain in the dye, but nitrogen laser pumping permits this, as unsaturated gains of 10^3 in a few millimeters path had already been demonstrated. (Such gains correspond to nearly total inversion in Eq. (11). To achieve them, Hänsch transversely pumped an $L = 1$-cm-long dye cell transversely (from the side) as shown in Fig. 5a. The nitrogen laser beam was focused into a narrow line of 0.15 mm width along the inner cell wall, with a dye concentration adjusted for a symmetrical 0.15-mm absorption depth). The output linewidth in such an avalanche is only a little smaller than the narrowest filter in the cavity, but very narrow filters (even though inevitably lossy) could be used because of the high gain, which also gives a high transmission for the output

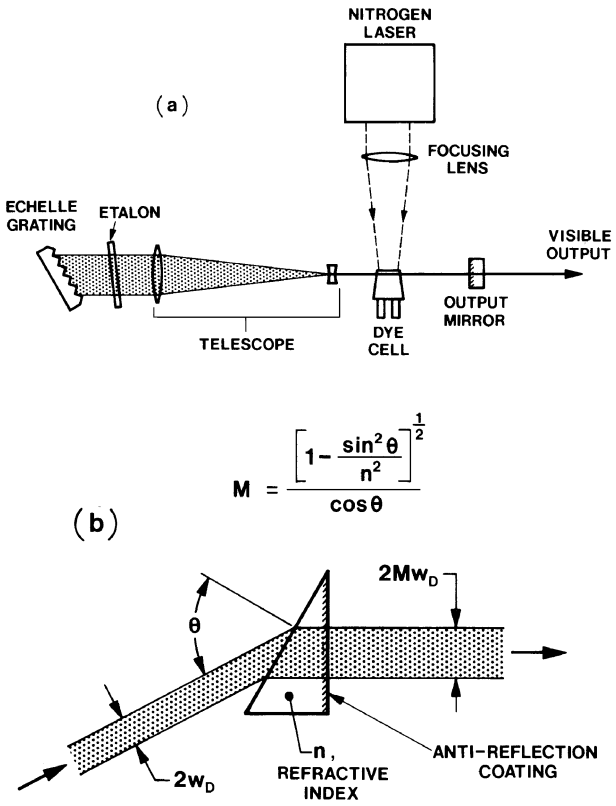


FIGURE 5 The orginal short-pulse tunable dye oscillator. (a) The layout by Hänsch, with an intracavity telescope to increase the resolution of the grating. (b) One-dimensional beam expansion in a prism with high incidence angle on the input face and normal incidence on the exit face. A series of four prisms, arranged for no net dispersion, makes up the "quad-prism expander" that has replaced the Hänsch telescope. [Adapted with permission from Klauminzer, G. K. (1977). *IEEE J. Quantum Electronics* **QE-13,** 103. ©1977 IEEE.]

coupler. This widens the cavity mode widths to be comparable to their spacing, eliminating discrete longitudinal mode structure and allowing the laser to be smoothly tuned without servo control of the cavity length. The laser's frequency stability is determined just by that of the narrowest spectral filter. This short-pulse oscillator is essentially a few-pass amplifier of filtered spontaneous emission.

An appropriate filter set in this design is a grating in Littrow mounting (equal incidence and diffraction angles) plus a high-finesse etalon. An etalon is a pair of parallel reflecting surfaces (often the two sides of a plate of glass), which makes a wavelength filter by interference of the multiply reflected beams. The relative narrowness or half-transmission width of the filter ratioed to the spacing of adjacent orders, called the finesse, usually is determined by and increases with the reflectivity of the two surfaces. The grating resolution is determined by the number of lines illuminated, times the grating order. To increase the number of lines (or equivalently, to decrease the beam divergence on the grating), the cavity contains a beam-expanding telescope of magnification 25 (Fig. 5a). This also reduces the intensity incident on the grating to avoid damaging its aluminum coating, and reduces the etalon losses that are due to walk-off or incomplete overlap of the interfering beams. A coarsely ruled or echelle grating (600 lines/mm) is actually an advantage in that it allows the grating to work over a large surface area at high incidence angles, while keeping only one dye tuning range per grating order. With this design, Hänsch achieved 5-nsec pulses of 0.1-mJ energy and 0.1-cm^{-1} spectral width without the etalon, and 0.01-mJ, 0.01-cm^{-1} pulses with the etalon, tunable with several dyes over the visible spectrum.

The basic Hänsch design is used today with one major change—the telescope is replaced by a set of half-prism beam expanders (Fig. 5b). (As the exit face is used at normal incidence, this is "half" of a normal prism, used with symmetrical entrance and exit angles.) To increase resolution, expansion is required only in the dispersion plane of the grating; by expanding in one dimension only with prisms, the grating tilt in the orthogonal plane is made noncritical. Typically four prisms are used in pairs with substractive dispersions to give an in-line achromatic arrangement, constituting a quad-prism expander (see Fig. 6 in the next section). The magnification given by a single prism (Fig. 5b) is large only at very high incidence angles, creating a large entrance-face reflection loss. By using several prisms, the incidence angle can be brought nearer to Brewster's angle to drastically reduce this loss; to produce the 66 times expansion of Fig. 6 with four symmetrical prisms, the incidence angle would be about 74° and the transmission 70% (through all four prisms) for the favored polarization direction lying in the incidence plane. The prisms automatically polarize the dye laser beam and are more economical than an achromatic telescope. Most importantly, large beam expansions are possible in a shorter length, allowing more transits during the excitation pulse and more efficient extraction of the oscillation input energy.

Since prism beam expanders and subsequently multiple-prism arrays have become rather important in many areas of optics, it is appropriate to consider them in some detail and to mention a few relevant historical aspects. The prism as a beam expander was first depicted by Newton, in his book *Opticks*, in 1704. He also considered prism arrays to control dispersion. Prism pairs applied to beam expansion were introduced by Brewster, in 1813. The prism was first used as a beam expander in a dye laser by Myers in 1971 and by Stokes *et al.* in 1972. Beam expanders comprised of prism pairs and several prisms were introduced independently to tunable lasers by Kasuya *et al.* and Klauminzer in the late 1970s.

The dispersion theory of generalized multiple-prism arrays was introduced by Duarte and Piper in 1982. In this

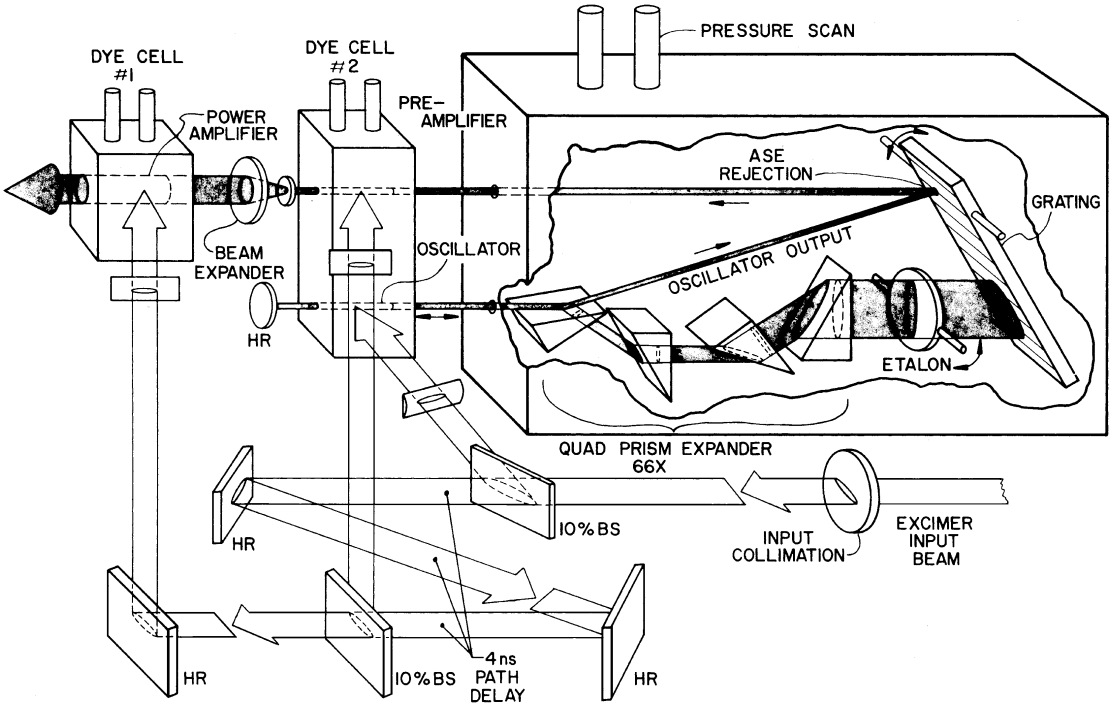


FIGURE 6 Schematic diagram of a narrow-linewidth short-pulse dye laser of current design, for pumping by an excimer laser. Two stages of amplification and features increasing resolution and rejecting amplified spontaneous emission are shown. Highly reflecting mirrors are labeled as HR and beamsplitters as BS. (Placement of dye cells, optics, and angles between beams are altered for purposes of clarity.) (Courtesy of Lambda Physik, Göttingen, West Germany.)

theory the laser linewidth, which determines the spectral purity of the laser emission, is given by

$$\Delta\lambda \approx \Delta\theta\,(\nabla_\lambda \Phi)^{-1}, \tag{14}$$

where $\Delta\theta$ is the beam divergence, $\nabla_\lambda = \partial/\partial\lambda$, and $\nabla_\lambda \Phi$ is the overall dispersion. The cumulative single-pass generalized multiple-prism dispersion at the mth prism, of a multiple-prism array is given by

$$\nabla_\lambda \phi_{2,m} = \mathcal{H}_{2,m} \nabla_\lambda \mathrm{n}_m + (k_{1,m}\, k_{2,m})^{-1} \times \left(\mathcal{H}_{1,m} \nabla_\lambda \mathrm{n}_m \pm \nabla_\lambda \phi_{2,(m-1)}\right). \tag{15}$$

In this equation, $\mathcal{H}_{1,m} = \tan\phi_{1,m}/\mathrm{n}_m$, $\mathcal{H}_{2,m} = \tan\phi_{2,m}/\mathrm{n}_m$, $k_{1,m} = (\cos\psi_{1,m}/\cos\phi_{1,m})$, and $k_{2,m} = (\cos\phi_{2,m}/\cos\psi_{2,m})$. Here, $k_{1,m}$ and $k_{2,m}$ represent the physical beam expansion experienced by the incidence and the exit beams, respectively. Also, $\phi_{1,m}$ is the angle of incidence, and $\psi_{1,m}$ the corresponding angle of refraction, at the mth prism. Similarly, $\phi_{2,m}$ is the exit angle, and $\psi_{2,m}$ the corresponding angle of refraction, at the mth prism. In general, Duarte and Piper in 1984 showed that for a pulsed multiple-prism grating oscillator, the return-pass dispersive linewidth is given by

$$\Delta\lambda \approx \Delta\theta\,[RM\,\nabla_\lambda \Theta_\mathrm{G} + \mathrm{R}\nabla_\lambda \Phi_\mathrm{P}]^{-1}, \tag{16}$$

where M is the overall beam magnification and R is the number of return-cavity passes. The grating dispersion in this equation, $\nabla_\lambda \Theta_\mathrm{G}$, can be either from a grating in Littrow or near grazing-incidence configuration. This equation includes the return-pass contribution to dispersion from the multiple-prism beam expander, which was given explicitly by Duarte in 1985 as

$$\nabla_\lambda \Phi_\mathrm{P} = 2M \sum_{m=1}^{r} (\pm 1)\, \mathcal{H}_{1,m} \left(\prod_{j=m}^{\mathrm{r}} k_{1,\mathrm{j}} \prod_{j=m}^{\mathrm{r}} k_{2,\mathrm{j}} \right)^{-1} \nabla_\lambda \mathrm{n}_m + 2 \sum_{m=1}^{r} (\pm 1)\, \mathcal{H}_{2,m} \left(\prod_{j=1}^{m} k_{1,\mathrm{j}} \prod_{j=1}^{m} k_{2,\mathrm{j}} \right) \nabla_\lambda \mathrm{n}_m. \tag{17}$$

Here, $M = M_1 M_2$, where M_1 and M_2 are the total beam magnification factors given by

$$M_1 = \prod_{m=1}^{r} k_{1,m} \tag{18}$$

$$M_2 = \prod_{m=1}^{r} k_{2,m}. \tag{19}$$

In this regard, it should be indicated that beam magnification factors of up to 100, and beyond, have been reported in the literature.

At this stage it should be mentioned that pulse compression in femtosecond lasers also use intracavity prisms. Newton's concept of addition and subtraction of dispersions, in arrays integrated by isosceles prisms, is rather applicable to prism sequences as deployed in femtosecond lasers. Prisms were first used to subtract dispersion, and hence compress laser pulses, by Dietel *et al.* in 1983 and then by Fork *et al.* in 1984. An extension of Eq. (15) applicable to generalized multiple-prism pulse compression schemes was given by Duarte in 1987.

C. Excimer Laser-Pumped Dye Lasers and Oscillator/Amplifier Configurations

As more powerful short-pulse pump lasers were developed, the dye-laser designs evolved to make use of them. The limit on the small-signal gain per pass in a short-pulse oscillator is due to the growth of amplified spontaneous emission (ASE). This will be greatest for wavelengths near the peak of the emission cross section σ_e, and the intensity reached in a single pass must be kept well below one saturation intensity I_s [which by Eq. (8) varies inversely with σ_e] to prevent depletion of the gain at the desired, fed back wavelength λ_D. Calculation of the growth of ASE for a cylindrical geometry shows this will be true if a conservative rule of thumb is followed, that the gain per pass $G = (\exp k_0 L)$ be kept below $(L/w_p)^2$, the square of twice the aspect ratio of the pumped volume (w_p is the pump beam radius). (The spectral separation between the ASE and λ_D at the ends of the dye tuning range means the ASE can be spectrally filtered outside the resonator here; in midtuning range, the heavy gain saturation at λ_D reduces the ASE.)

High gains in one pass thus require a long, thin pumped volume, leading to small total volumes and ineffective use of a powerful pump laser. A better route to high-output energies in a tunable narrowband source is to use an oscillator–multiple-stage amplifier configuration. Here the inefficiency of the oscillator (from lossy spectral filters and the cavity length to contain them) can be ignored, as most of the pump energy is put into the following amplifier cells. By spectral filtering between stages, and timing (with path delays) the arrival of the pump pulse at the amplifier cell, ASE can be minimized. The amplifiers work in the partially saturated regime, which smooths out intensity fluctuations. The beam can be expanded between stages (with the dye concentration adjusted in each cell) to achieve a large active volume and a balance between amplifier extraction efficiency and absorption losses from the excited singlet state, a problem in high-gain systems with a large percentage of inverted molecules.

These features are evident in the modern design for a short-pulse, narrow-linewidth dye laser shown in Fig. 6. (Other designs also in use are the Littmann grating mount, a high-incidence grating with back-up mirror, and sometimes three amplifier stages.) This shows an excimer laser-pumped laser, but the designs for doubled Nd:YAG laser pumping are very similar. The 308-nm line from the xenon chloride excimer is used to pump some 20 dyes covering 332–970 nm. In the longer-wavelength dyes, this is by absorption to higher singlet states, which nonradiatively relax to the upper laser state (see Section II.B). (There is usually reduced photochemical stability with UV pumping of low Stokes efficiency dyes, however.) Pump beams of 100-W average power (0.4-J energy in 28-nsec pulses at 250-Hz rate) are available.

An excimer is molecular species that is stable only in the excited state. In a XeCl excimer laser, a high-current pulsed discharge in a mixture of xenon and chlorine gas makes Xe^+ and Cl^-, which attract to form the excited $XeCl^*$ molecule, the upper laser state. After radiative decay, the molecule dissociates back to Xe and Cl, so there is no laser lower-state bottleneck to be concerned with in this laser. The discharge channel in an excimer laser is rectangular, giving an oblong cross section to the pumping beam (entering from the right in Fig. 6), making transverse pumping the most convenient. Ten percent of this beam is split off and focused into the dye cell (number 2) of the Hänsch type oscillator whose resonator mirrors are the grating on the right and the high reflector to the left of the dye cell. In the design shown, the first prism of the 66× quad-prism expander is placed at a higher incidence angle to make the reflection off its input surface serve as the oscillator output coupler. This output beam is directed at the grating at a small incidence angle, smaller than the angular width of the diffraction lobe for the unexpanded beam. This second diffraction off the grating thus acts as a spectral filter (that tracks automatically), reducing the amount of oscillator ASE that enters the preamplifier stage. An intracavity etalon with a finesse of 30 and order spacing of 1 cm^{-1} gives an 0.03-cm^{-1} linewidth, scannable either by microprocessor-controlled tilting of the grating and etalon (30-cm^{-1} range before resetting the etalon order) or by changing the gas pressure in the sealed oscillator housing. The 4-nsec path delay of the pump beam, before the 10% split off to pump the preamplifier dye, allows an appropriate oscillator buildup time before gating "on" the amplification of the next stage. The remaining 80% of the pump beam is weakly focused into the enlarged active volume of the final power amplifier stage, which is filled by the appropriately enlarged dye beam (about eight diameters expansion). This dye cell has

a separate dye circulation system to allow the concentration to be adjusted for a longer absorption depth.

At 90-W average 308-nm input power, this system has generated, at a 250-Hz pulse rate with fast flows of coumarin 102 dye to avoid burning the dye cells, an average output power of 16 W at 475 nm. This is at 0.2-cm^{-1} linewidth (no etalon), with less than 1% ASE and a beam divergence of twice the diffraction limit. At 22-nsec pulse length this is 2.9-MW peak power. With an etalon giving 0.04-cm^{-1} linewidth, the average power was 13 W. At the time this article is being written, only a few dyes have been characterized at this high an input power, an example of a point in the evolutionary race between short-pulse pump and dye lasers where the pump laser is a little ahead.

D. Dye Lasers Pumped and Mixed with Harmonics of the Nd:YAG Laser

For the highest peak powers (at lower-pulse repetition rates and average powers), the dye laser is pumped (and mixed) with harmonics of the *Q*-switched neodymium–YAG laser. Higher peak powers result from a more efficient pump wavelength (green, 532 nm) for rhodamine dyes, coupled with a shorter pump pulse length (of 5–9 nsec).

Pump energy is stored in the optically excited solid-state YAG laser rod over the upper-state lifetime of 200 μsec, and an intracavity electrooptic *Q*-switch dumps this stored energy at a fixed repetition rate, usually 10 Hz, to form these short pulses. (The fixed rate allows the thermal lens in the thermally loaded rod to be accounted for in optimizing the YAG resonator design.) High-power YAG systems consist of an oscillator rod followed by a larger-diameter amplifier rod. For dye pumping, the fundamental wavelength of 1.064 μm must be doubled in a phase-matched nonlinear crystal to the green wavelength. For shorter-wavelength spectral coverage, the green beam is sum-frequency mixed with the unused portion of the fundamental in a second nonlinear crystal to generate a 355-nm ultraviolet pump beam, the third harmonic. Mixing is like doubling, but with photons of unequal energies. Amplitude noise on the fundamental is carefully controlled as the noise level multiplies in these nonlinear conversions.

The same processes of doubling and of mixing with the 1.06-μm fundamental beam, are performed with the strong outputs of the green pumped dyes as well as to achieve spectral coverage into the ultraviolet. These nonlinear conversions are most efficient for beams in the lowest-order transverse intensity profile, or TEM_{00} Gaussian mode. It was the development of low-order-mode YAG lasers, based on unstable resonators with diffraction or polarization output coupling, that made pumping dyes with YAG lasers practical. This is also the reason for the main difference in dye laser design for use with YAGs, namely, that for the green-pumped dyes the final power-amplifier stage is pumped longitudinally, with the round, Gaussian profile pump beam folded in and overlapped along a common axis or at a small angle with the dye beam. The dye laser then replicates the low-order mode of the pump beam, enhancing the subsequent nonlinear conversions.

A 10-Hz YAG–dye system with 532-nm pump pulses of 360-mJ energy produces 120-mJ output pulses at 560 nm with rhodamine 6G dye, in a 5-nsec pulse of 0.3-cm^{-1} linewidth (no etalon), and less than 1% ASE. That is a 33% conversion efficiency, a 24-MW peak power, and 1.2-W average power with a beam divergence of four times the diffraction limit. With an etalon, 90-mJ pulses are produced at 0.05-cm^{-1} linewidth. Linewidths increase in the nonlinear conversions with the number and bandwidths of the photons involved.

E. Short-Pulse Dye Laser Tuning Spectrum

Dye output tuning curves for the two pump sources are compared in Fig. 7. This shows the output energy per pulse versus dye and output wave number, with only that portion shown of a dye's tuning range where the output exceeds that of the adjacent dyes. To obtain these results, the pump systems are configured to maximize the energy per pulse. This is at low repetition rates ($<$10 Hz), giving 0.5 J energy per 308-nm pump pulse for the XeCl system, and for the YAG system, 0.36 J per 532-nm pump pulse, 0.15 J per 355-nm pump pulse, and the undoubled portion of the initially 0.8-J, 1.06-μm mixing pulse. At higher rates both systems show a decreasing energy per pulse (but an increasing average power) as the lasers must dissipate more heat. The YAG system reaches a plateau of constant average power at about 15 Hz and is not designed to be used above 30 Hz. The excimer system pulse energy decreases more slowly, to 80% of the single pulse energy at a 250-Hz repetition rate, giving the excimer system the capability for nearly an order of magnitude higher average power. The difference in efficiency of the pump wavelength and in the shorter pulse width gives the YAG system the capability in the rhodamine dye region for nearly an order of magnitude higher peak power, and in this sense the two pumping systems are complementary.

Other features of Fig. 7 are also readily related to differences in the pump sources. For the strong green-absorbing dyes, the lower conversion efficiency of the excimer system is nearly accounted for by the lower Stokes efficiency η_S, which makes the 0.5-J UV input equivalent to 0.29 J of green input. Conversely, the UV excimer efficiently pumps the dyes lasing in the blue, where conversion to the third harmonic cuts the YAG pump power to less than half. In the red output region where there are common dyes,

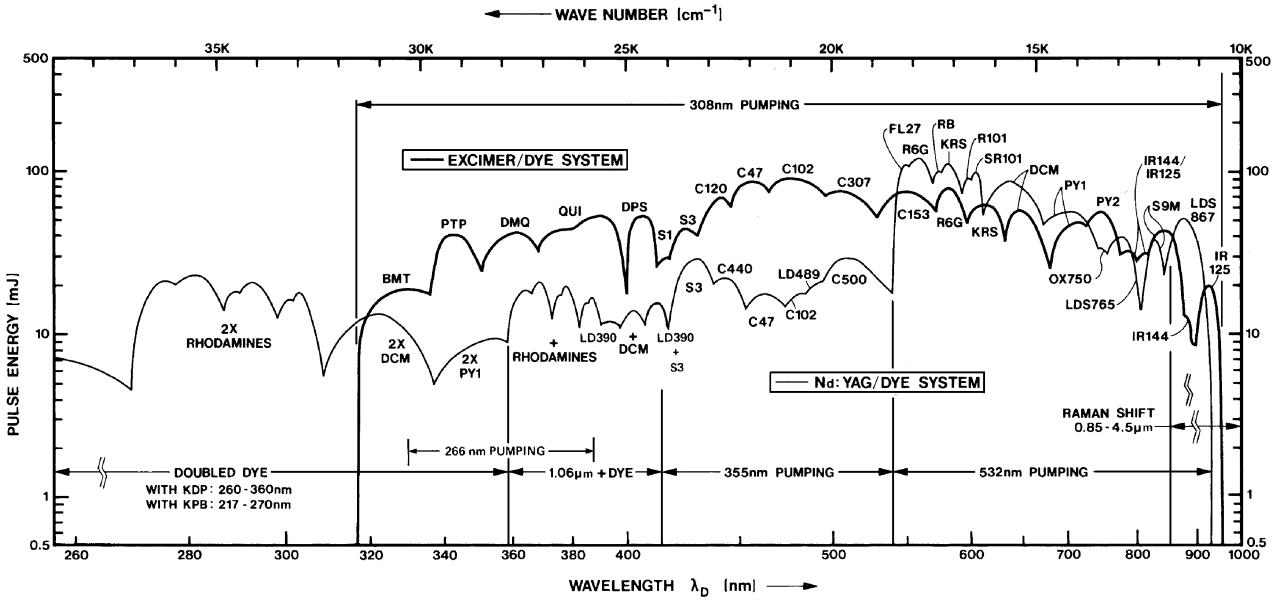


FIGURE 7 Output tuning curves in energy per pulse at low repetition rates for excimer laser-pumped (heavy line) and Nd:YAG laser-pumped (light line) short-pulse dye lasers. Only the upper portion of a dye's tuning curve (above that of the adjacent dyes) is shown. Alternative pumping schemes to those shown are also possible for the YAG system, with the 266-nm fourth harmonic to pump dyes with outputs of 334–387 nm and with the 1.06-μm fundamental to pump dyes with outputs of 1.09–1.32 μm. (Excimer data, courtesy of Lambda Physik, Göttingen, West Germany; Nd:YAG data, courtesy of Quanta-Ray, Mountain View, CA.)

notice the shift to longer wavelengths for the excimer system. This is due to the overlapping σ_a absorption at the higher dye concentration, adjusted for a short transverely pumped absorption length and a more weakly absorbed pump wavelength. (The dye codes either are listed in the next section or are cross-referenced in laser dye manufacturers' literature.)

The YAG system propagates its strength in the red lasing dyes (using appropriate crystals and steering optics) by sum-frequency mixing these outputs with the YAG fundamental to cover the 358- to 413-nm span and doubling these dyes for the 270- to 358-nm span. Thus the characteristic pattern of the red dye tuning peaks appears three times in the YAG/dye spectrum. In both systems, doubling the coumarin and stilbene dye outputs reaches shorter UV wavelengths (only the YAG results are shown in Fig. 7), to 217 nm with two crystals. Output wavelengths are extended into the infrared to about 4.5 μm with both systems by directing the dye beam into a Raman shifter, a long (1 m) high-pressure (30 atm) cell filled with hydrogen gas. The hydrogen molecules are driven in coherent vibrations by the high-intensity input light scattered in the Raman process, and they radiate coherent beams shifted in wave number from the input beam in increments of the hydrogen vibrational level spacing of 4155 cm^{-1} at conversion efficiencies exceeding 20% for the first Stokes wave (light frequencies downshifted by one vibrational spacing) and 2% for the first anti-Stokes wave (upshifted frequencies, used to extend outputs deeper into the ultraviolet). In the days before lasers, the Raman effect was considered a weak process, as hours of exposure time with a several-kilowatt mercury arc source were required to register a Raman-shifted line on a spectrographic plate. With little more electricity taken from the wall than required by the old arc source, the short-pulse dye laser system produces, by stimulated Raman scattering, Raman-shifted beams of a megawatt peak power. Thus the wavelength extension techniques employed with short-pulse dye lasers serve quite well as examples of the nonlinear optics applications for the beams these systems generate.

IV. CONTINUOUS-WAVE, SINGLE-FREQUENCY DYE LASERS

A. Dye Cells, Dye Jets, and Focusing Cavities

The first continuously operated dye laser focused the pump light from an argon ion laser into a cell of flowing dye to a spot diameter of around 10 μm, an active cross-sectional area of only 10^{-6} cm^2. By Eq. (12), this brought down the pump power needed to be comfortably above threshold to less than a watt, and the dye flow took care of cooling and

triplet-state relaxation by exchanging the dye in this tiny active volume every few microseconds.

Focusing the coherent pump beam to this diameter was no problem, but building a stable dye laser resonator with a minimum area or beam waist diameter this small to overlap the pump spot was another matter. The necessity for matching the two beam diameters comes from a calculation of the filling factor η_F of Eq. (13) for the cw case of longitudinal pumping. A Gaussian TEM_{00} mode pump beam of $1/e^2$-radius w_p is assumed, whose axis and waist location overlaps that of a TEM_{00} mode dye beam of radius w_D. As previously defined, the ratio of beam areas is $\rho_0 = w_D^2/w_p^2$, where in this case the subscript zero has been added to indicate that the beam radii are to be taken as their minimum values, at their waists. The results are that the incremental gain constant for the dye beam is reduced from the plane wave result of Eq. (7) (where I_p is taken as the peak pump intensity at the center of the Gaussian profile) by a factor of $1/(1+\rho_0)$, the threshold pump power in Eq. (13) is increased by a factor $(1+\rho_0)$, and the filling factor η_F in Eq. (13) becomes $\rho_0/(1+\rho_0^2)^{1/2}$. The output power is still a linear function of the input power above threshold, as given by Eq. (13), but the slope of this conversion increases only asymptotically with ρ_0, while the threshold power subtracted from the pump power increases linearly. Consequently there is in practice an optimum value of $\rho_0 \approx 1.2$–1.3 (giving $\eta_F \approx 0.8$), namely, that the dye beam diameter should be only slightly larger than the pump beam diameter.

The necessary small beam diameter was achieved with a resonator consisting of a flat and curved mirror (radius R), whose mirror separation d was critically adjusted around $d \approx R$ (but with $d < R$) to be near the stability limit. A resonator is said to be stable when multiply reflected paraxial rays stay within the structure; for this case, as d approaches R, a multiply reflected ray makes increasingly steep angles with the axis as it propagates, until at $d > R$ the ray walks out of the structure, as a ray tracing in the unfolded resonator (an equivalent sequence of lenses) will readily show. As this unstable condition is approached, the beam diameter at the flat mirror shrinks toward vanishing, and at the curved mirror it blows up to exceed the mirror aperture diameter. The first cw dye laser used $R = 4.5$ mm, and the spacing difference $R - d$ was set to 20 μm, to within ± 0.3 μm to keep the dye beam area within $\pm 20\%$ of the desired size. If the curved-mirror radius were made larger in such a resonator (to accommodate a wavelength tuning device), the tolerance on the required mirror spacing would become proportionally smaller. Clearly, a less critical assembly than a two-mirror resonator is needed for a cw dye laser.

A three-mirror or focusing resonator consisting of two curved mirrors and a flat provides this (Fig. 8b). This has a long arm of length $d \approx 0.6$ m where the beam is collimated and of a relatively large diameter $2(2d\lambda/\pi)^{1/2} \approx 1$ mm. The beam is then brought to a small focus of diameter $R(\lambda/2\pi d)^{1/2} \approx 40$ μm by the internal focusing mirror of radius $R \simeq 0.1$ m. A rhodamine 6G wavelength of $\lambda = 0.6$ μm has been assumed here.

(These dimensions are appropriate to the cw laser discussed later, which is pumped with a 6- to 24-W input and so works with a proportionally larger focal area than the earlier cw laser.) A new problem is introduced with this solution, which is fortunately readily solved. A spherical mirror used off-axis like the internal folding mirror in Fig. 8b generates an astigmatic focus, as shown in Fig. 8a. The rays lying in the incidence plane see the mirror's curvature foreshortened into a tighter arc and come to an out-of-plane line focus (the tangential focus) ahead of the on-axis focal length; rays lying in a plane orthogonal to the incidence plane see no foreshortening and reach a line focus (the sagittal focus) lying in the incidence plane beyond

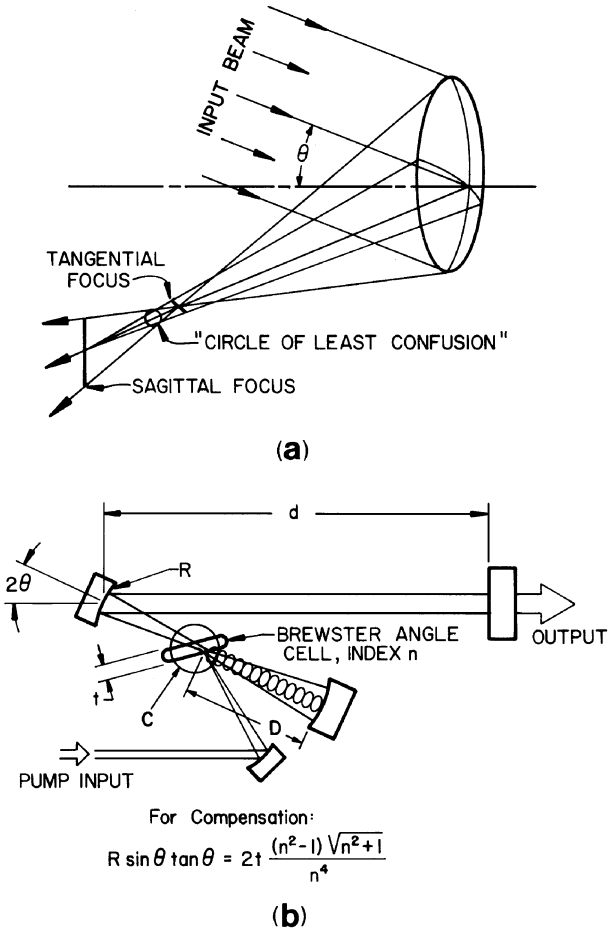


FIGURE 8 (a) Astigmatism of an off-axis spherical mirror and (b) its compensation by a Brewster plate in a three-mirror focusing cavity, as originally used for standing-wave cw dye lasers.

the on-axis focal length. In between the line foci, the beam shape becomes round (the "circle of least confusion") but at a considerably larger focal area.

Fortunately, an astigmatism of the opposite sense results when a flat plate is placed in the focused beam in the same plane of incidence but at an angle off normal incidence. In the dye laser this will be the dye cell, and it can compensate the fold-mirror astigmatism and restore the small (round) focus. The incidence angle chosen for the plate will be Brewster's angle θ_B (where $\tan \theta_B = n$, the index of refraction of the plate) for low loss. [Minimizing the nonuseful resonator losses a in Eq. (9) is very important in cw dye lasers that operate at low gain; this accounts for the abundance of Brewster angle faces found on cw intracavity wavelength control devices—which then serve to polarize the beam as well.] For a dye cell of thickness t, the correct fold angle 2θ between the collimated and focused beams that achieves compensation is calculated in Fig. 8b; for a 3-mm-thick glass cell in the example above, the result is a 9° fold angle.

The stability limits for a three-mirror resonator are computed as in the two-mirror, case after forming from the combination of the output and fold mirror an optically equivalent single mirror, which has a small radius of curvature $R' = R^2/2(2d - R) = 4.5$ mm in the example. In aligning the dye laser, as the spacing between the two curved mirrors of Fig. 8b is varied plus or minus half this distance off the central focus position, the dye spots on the curved mirrors blow up to fill the mirror aperture as the resonator reaches its limits of stable focus. The spot at the focus, in the dye, correspondingly shrinks toward vanishing, which gives a way to adjust ρ_0 downward and experimentally find the optimum value. Since the resonator for a laser must be stable in both the tangential and sagittal planes, if the astigmatism compensation is not correct, this range of stable focus adjustment is reduced by the amount of residual astigmatism (distance between the line foci of Fig. 8a), and the output beam is not round (except for the one adjustment where the circle of least confusion lies in the plane of the dye). For the resonator of the example, assuming perfect compensation, the adjustment range over which the dye beam area stays within ±20% of the central focus size is ±1.5 mm, a manageable tolerance and a vast improvement over the two mirror case. Note in Fig. 8b that the pump beam enters the dye cell at a slight angle to the dye beam (called noncollinear pumping), which is permissible as by the Beer's law absorption profile a good overlap of the two beams is only essential in the first third of the path length in the dye, where over half of the pump beam is absorbed (for a total absorption of 80–85% as normally employed).

The improved ease of operation made by the focusing resonator soon led from milliwatt-level to watt-level cw

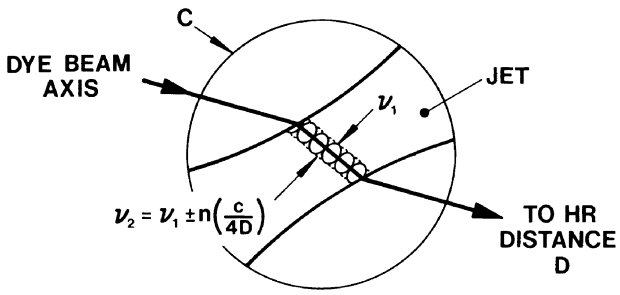


FIGURE 9 Enlarged view of the jet, replacing the dye cell inside the circle C of the three-mirror cavity of Fig. 8b, showing the standing waves of the first frequency to lase locally depleting the gain or "burning spatial holes." This leaves gain regions unaddressed and allows a second frequency to lase.

output powers (at higher input powers), accentuating another problem. No matter how clean or carefully filtered the dye solution was made, something that would burn and stick always seemed to pass through the focused pump spot after a few hours of operation, forcing shutdown and realignment of the system on a clean position on the dye cell. The answer to this was to eliminate the dye cell! It was realized that good optical quality in the region of the dye was needed over an area only a few times larger than the small area of the focus. Experimentation with free jets of dye dissolved in a viscous solvent like ethylene glycol, squirted at 10 m sec^{-1} from a nozzle with a slotted tip and caught in an open tube returning the solution to a fluid pump, showed this jet stream to give adequate surface flatness, stability, and dye velocity. (The flat jet actually necks down near the middle of its cross section as shown in Fig. 9, and is translated during set-up to find the mid position where the two surfaces are parallel.) A jet and catch tube are shown in Fig. 10. In addition to solving the burning problem, nozzles are much more economical than optical quality cells. Astigmatism can be so reduced by closing up the fold angle (possible because of the small jet dimensions, only 4 mm wide by 0.1 mm thick in section at the incidence plane) that compensation is no longer necessary.

B. Spatial Hole Burning and Ring Cavities

To fulfill its promise as a light source for high-resolution spectroscopy, the cw dye laser had to be operated with a single-frequency output (i.e., in a single longitudinal and transverse mode), which led to the development of tunable intracavity selective filters or "filter stacks" capable of supressing oscillation in all but one such resonator mode. These are considered in the next section; here the phenomenon of spatial hole burning is discussed. This limits the single-frequency power obtainable in the resonator of Fig. 8b to small values.

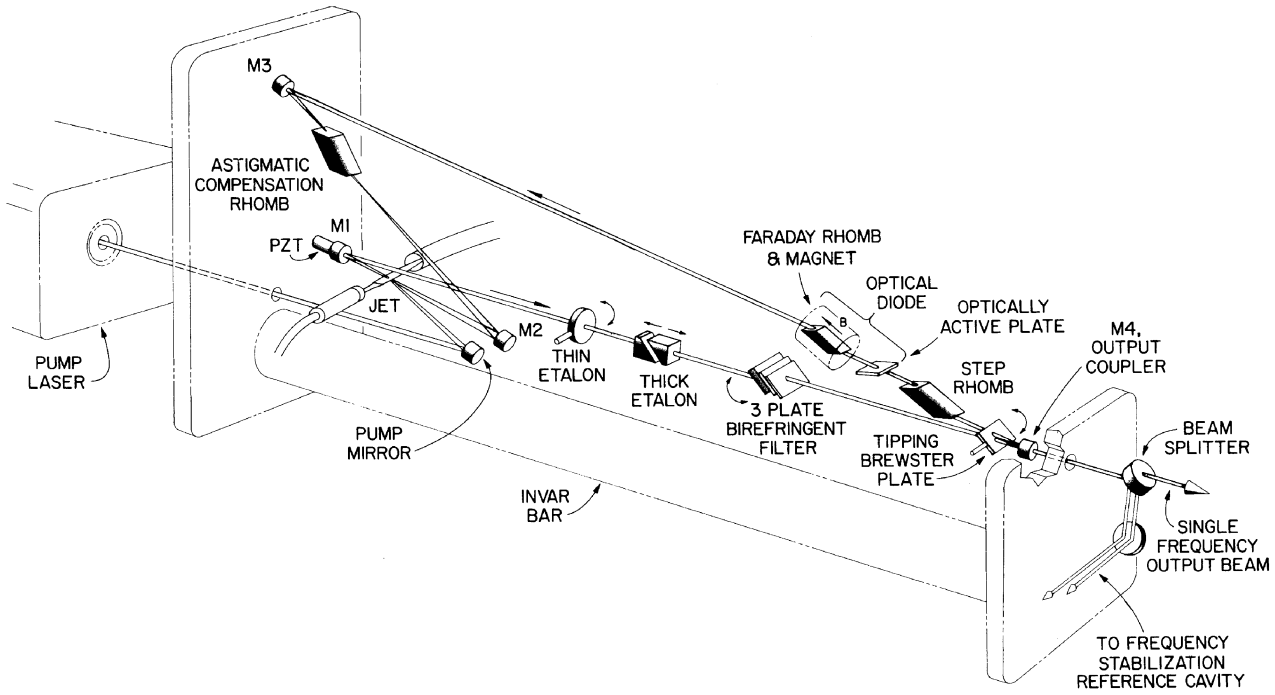


FIGURE 10 Tunable single-frequency ring (traveling-wave) dye laser of current design. Spherical folding mirrors M1 to M3 and flat output coupler M4 fold the beam in a figure-eight path, which is traversed in the direction shown by arrows due to the biasing action of the optical diode element. Elements in the horizontal arm lying between M1 and M4 constitute the tunable filter stack, and the output frequency is controlled and the linewidth narrowed by a frequency stabilization servo (whose elements are not shown). (Courtesy of Coherent, Inc., Palo Alto, CA.)

The three-mirror cavity shown is called a standing-wave or linear resonator, because the right-going and left-going dye beams in the resonator coincide and transform into each other at the normal incidence reflections at the two end mirrors. The superposition of these two traveling waves produces standing waves in the cavity that have nodes at the reflecting end surfaces. The allowed resonant frequencies or longitudinal modes of the cavity are precisely those that fit an integer number of half wavelengths into the optical path between the two end mirrors, as this condition gives constructive interference of the multiply reflected waves and results in the standing waves shown over length D in Fig. 8b. The oscillation in this resonator is at a single frequency, say the frequency ν_1. By counting half-wavelengths from the nearest end mirror (the curved mirror at distance D from the jet), it is seen that inside the jet (see the magnified view of Fig. 9, inside the circle C of Fig. 8b) the standing waves occupy only part of the pumped volume. A second cavity mode with a frequency near $\nu_2 = \nu_1 \pm n(c/4D)$, where n is an integer, will have antinodes where the first mode has nodes and hence will not be suppressed by the gain saturation due to the first mode—the two modes in the same jet interact with spatially separated dye molecules. The first mode is said to "burn spatial holes" in saturating the gain medium, and the second mode utilizes the gain left between the holes made by the first. When a standing-wave, initially single-frequency dye laser is pumped well above threshold in an attempt to increase the single-frequency output power, at a fairly low input power a second frequency ν_2 begins to oscillate and destroys the single-frequency operation.

This limit is removed by going to the contemporary form of single-frequency dye laser in a traveling-wave or "ring" resonator (Fig. 10). The optical elements of the laser are shown here in their actual positions, with their support structure (invar bar and end plates) indicated in phantom lines. The dye beam is folded in a loop by mirrors M1 to M4, avoiding normal incidence end reflections and standing waves. (Actually, the folded path is more of a figure eight, to keep the fold angles on spherical mirrors M1 to M3 and their astigmatism small and compensable by the thick Brewster plate, or compensating rhomb, between M2 and M3.) Since the pumped volume is wholly utilized by a traveling wave, the first direction to oscillate (right- or left-going) will suppress the other possible one, and a single-frequency (s.f.) ring laser runs stably and naturally in one direction only. But cw dye lasers are subject to having their oscillation very briefly but often interrupted and restarted (due to microscopic particles or bubbles in the jet stream passing through the focal volume). A means

to bias favorably one direction is required, or the output beam switches randomly and uselessly in exit angle from the output coupler.

The biasing device is called an "optical diode" and consists of a Faraday rhomb (a glass plate in a magnetic field) and an optically active plate. Considerations of time-reversal symmetry in Maxwell's equations dictate that any one-way or nonreciprocal light wave device be based on the Faraday effect, or the rotation of the direction of a linear polarization in transit of a material in a longitudinal magnetic field. The sense of the rotation depends only on the sign of the magnetic field B; for instance, for B pointing away from M4 in Fig. 10, the electric field of *either* a right-going or left-going beam is tipped away from the observer in a transit of the Faraday rhomb. When this is combined with a *reciprocal* polarization rotator (the Brewster incidence crystal quartz plate cut with optic axis along the ray propagation direction, to show optical activity), the forward or favored wave suffers substractive rotations, resulting in no tip of polarization out of the incidence plane of the remaining Brewster plates in the cavity and no reflection loss there. The backward wave suffers additive rotations and subsequent reflection losses; experimentally, a small differential loss of only 0.4% between these waves was found to be sufficient to always select the forward wave. The Faraday and optically active rotations are made to match over the desired spectral region by appropriate choice of magnetic field, rhomb material, and quartz plate thickness.

In addition to giving an order of magnitude increase in single-frequency output power, ring cavities enhance the intracavity second-harmonic generation process. By having all of the fundamental intracavity power traveling in one direction, the power of the doubled output, which extends single-frequency spectroscopy into the near ultraviolet, is increased an automatic factor of 4.

C. Tunable Filter Stacks for Single-Frequency Lasers

Most of the remaining intracavity elements of Fig. 10 comprise the single frequency filter stack, a series of frequency filters of progressively narrower bandwidths, with overlaid center frequencies that tune together, designed to allow oscillation in but a single longitudinal mode of the ring cavity. (The transverse TEM_{00} dye mode is selected by the overlap with a TEM_{00} mode pump beam; this is called "gain aperturing.") Aside from the coarsest element (the lasing bandwidth of the dye itself), these filters all use the constructive interference between a first beam, and a second beam split off and delayed by a path length of an integer number of wavelengths, to generate a transmission peak. Each integer labels an order number of the multiply peaked filter, the spacing in frequency between orders is called the free spectral range (FSR), and the bandwidth of a transmission peak is given as the full width at half-maximum (FWHM). To minimize the intracavity losses [a in Eq. (9)], the FWHM of each element in the stack is made just narrow enough to select a single order of the next finer element of the stack, and spectral condensation plays a considerable role in this selection process.

The filter stack must be designed to span an enormous range. The two parts of Fig. 11 showing the filter stack at low and high resolution are therefore drawn with the expansion of the horizontal scale of 300 times between them directly indicated to help visualize this. The parameters of the figure are appropriate for rodamine 6G dye at its peak 580-nm wavelength. The gain bandwidth of this dye, determined largely by the emission cross section $\sigma_e(\lambda_D)$ derived in Fig. 4c, is about 3×10^3 cm^{-1}, making the curvature of the gain curve (scaled up here slightly to fit a 0–100% axis) just noticeable on the 250-cm^{-1} span of Fig. 11a.

The FSR of the birefringent filter or BRF, the next finer element in the stack, is chosen to be 3×10^3 cm^{-1} so that only the sixth order falls within the region of strong lasing of this dye. Birefringent crystal quartz plates oriented at Brewster's incidence are each cut with their optic axis lying in the face of the plate but rotated about 45° out of the incidence plane. The linearly polarized dye beam is split upon entry into the crystal into orthogonally polarized components, the extraordinary and ordinary rays. These recombine at the exit face back into light linearly polarized in the incidence plane for the central pass frequency, but for other frequencies into elliptically polarized light (which is partially reflected at intracavity Brewster surfaces). Like an etalon where $2nt$ is the second beam path delay (n is the index, t the physical thickness) producing a multipeaked filter of FSR $= 1/2nt$ wavenumbers, this arrangement may be thought of as a two-beam interferometer with an $(n_0–n_e)t$ path delay. In crystal quartz the ordinary–extraordinary index difference is about 0.01, so the birefringent filter acts like an etalon of thickness 1% of the physical thickness and is a convenient way to make a low-resolution interference filter. (An etalon with a thickness of only 1 μm would be required to give the same FSR.) Three plates with precisely aligned axes and thicknesses in the exact ratios 1:4:16 produce an intracavity FWHM of 56 cm^{-1}, about half the width the same filter would produce extracavity between full polarizers. Gain saturation and spectral condensation are very effective in a cw dye laser, and this filter alone in the cavity ("broadband operation") will give an output spectral width of 0.03 cm^{-1}. The filter is tuned by rotation of the three plates together about their face normal, which varies the extraordinary index (and second beam path delay) while

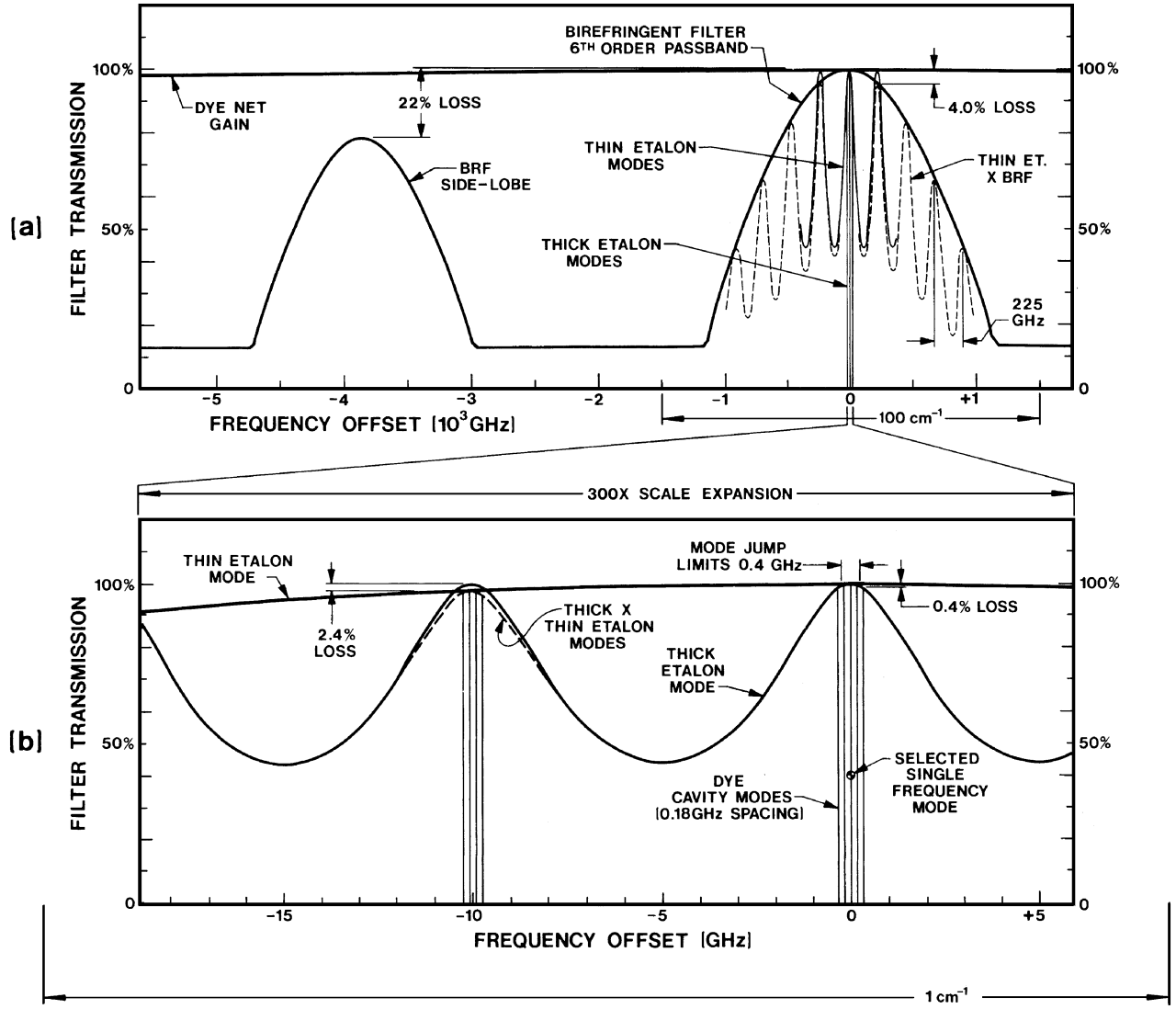


FIGURE 11 Transmission functions of the filters in the tunable single frequency ring laser filter stack (a) at low resolution and (b) at 300 times higher resolution. Solid lines are for a filter element used alone, and dashed lines are for the composite stack. The curvature of the dye tuning curve is barely noticeable in (a) and the separation of the selected mode to the next adjacent cavity mode is barely visible in (b), emphasizing the enormous span of frequency over which the filter stack must work to select one mode out of 400,000. (Courtesy of Coherent, Inc., Palo Alto, CA.)

keeping Brewster's low-loss angle of incidence. Side lobes or frequency leaks in the filter transmission function correspond to polarizations that emerge from one or more (but not all) of the plates near the low-loss linear polarization. A BRF is not as positive as a grating for rejecting frequencies, because of these side lobes, but it is much lower in loss for the selected frequency (0.1 instead of 2–10%), which is what is important in cw dye lasers. It was the introduction of the low-loss BRF tuner in 1974 that allowed the cw dye laser to first span the visible with the low cw UV power then available to pump the blue lasing dyes.

A pair of low-reflectivity etalons (finesse of 1.8) reduces the BRF oscillating bandwidth to that of a single longitudinal cavity mode, the "thin" etalon of 7.5 cm^{-1} FSR or $\frac{1}{2}$-mm physical thickness, and the "thick" etalon of 0.33-cm^{-1} FSR and 10-mm physical thickness. Two etalons of low finesse give less loss for the selected mode than a single high-finesse etalon, due to a more complete overlap of the interfering beams (less "walk-off" loss). In Fig. 11a the thin-etalon transmission functions for the three central orders are shown as solid lines, and the products of the BRF and thin etalon functions (the composite filter) for several more orders are shown as dashed. The

BRF suppresses the thin-etalon orders adjacent to the selected one by adding an extra 4% loss at the frequency offset of these orders. In Fig. 11b, the frequency scale is expanded 300 times to show the selection of a thick-etalon mode by the selected mode of the thin etalon. Here the curvature of the thin-etalon function is just noticeable, and a central thick-etalon mode is selected by an extra 2.4% loss at adjacent orders. Finally, the thick etalon selects a single mode of the ring cavity to complete the filter stack, by inserting an additional 0.4% loss at adjacent cavity modes split off from the central one by a $c/P = 0.18$ GHz frequency offset (P is the path length around the ring perimeter). The total absorption and scatter loss per round trip at the selected mode frequency for the whole filter stack of Fig. 11, the a of Eq. (9), is 2.5%, with the walk-off loss of the thick etalon at 0.7% the largest component.

The center frequencies of this stack of filter elements must track together during scanning to tolerances that maintain proper mode selection. An offset of a coarse element, no greater than one-third of the FSR of the higher-resolution element whose order it selects, is acceptable. The basic continuous scan width of the system is 1 cm^{-1} (30 GHz); longer scans are pieced together from 1-cm^{-1} segments, and over this width the tracking tolerances and means of achieving them are as follows. The BRF need not be tuned at all but is stepped between segments. The thin etalon is tilt tuned with a rotational galvanometer and must stay within 3 GHz (or 10% of the scan range) of the linearly scanning frequency. This is done with a preprogrammed square-root drive to the galvo (the path-length change and hence center-frequency offset of a tilted etalon goes as the square of the tilt angle away from normal incidence). The thick etalon must track to 60 MHz or 0.2% of the scan range, requiring a servo to meet this tolerance. The thick etalon is made up of two Brewster-surfaced prisms (essentially a solid etalon with a thin slice of air in the middle as shown in Fig. 10), so that its thickness may be tuned by piezoelectric translation. A small dither at a rate of a few kilohertz is added to the thick etalon drive voltage, generating a small-amplitude modulation of the output beam. This is phase-sensitively-detected, giving a discriminant signal to lock the thick-etalon center frequency to the scanning cavity mode frequency (standard dither- and-track circuit). To scan the cavity mode, a Brewster plate (on another rotational galvanometer) is tipped slightly ($\pm 2^\circ$) off Brewster's angle. This introduces only slight reflection losses ($<0.4\%$) and linearly varies the path length through the plate by 120 μm, enough to scan the cavity frequency over the desired 1-cm^{-1} range. To avoid misaligning the ring cavity, this tipping plate is mounted near the beam vertex at the output coupler M4 and is traversed by both beams, in the design of Fig. 10. It was the introduction of the tipping Brewster plate as the cavity-mode scanning element that first allowed a 1-cm^{-1} or 30-GHz continuous scan range, due to the improvement in linearity and repeatability of this method over the older use of a mirror mounted on a piezoelectric element for the long-range scan.

D. The Narrowest Linewidths—Frequency Stabilization Servomechanisms

The intrinsic linewidth of the oscillation at the frequency of the selected single dye cavity mode, limited by spontaneous emission noise, can be found by the well-known Schawlow–Townes formula to be something less than 10^{-3} Hz. The observed linewidth is much larger and is determined not by this intrinsic width but by the stability of the cavity mode frequency given by $\nu_q = q(c/P)$ (where q is an integer of the order 3×10^6). The optical path P around the ring is subject to random environmental perturbations, a one-wavelength or 0.6-μm change in P, changing the cavity frequency by c/P or 0.18 GHz. The effective linewidth or frequency jitter produced by these perturbations will be (without stabilization) about 15 MHz rms, due to environmental noises, vibrations, and particularly due to the jet stream, whose thickness varies with pressure surges in the dye fluid circuit. This linewidth is reduced with a frequency stabilization servo that locks the dye cavity mode frequency to that of a stable reference interferometer, transferring the frequency stability properties of the output beam from that of the generating cavity, to that of the reference. The reference cavity, which can be compact and need not enclose a jet, is built to be stable by employing low-expansion materials in a temperature-stabilized housing, isolating the structure from vibrations, and sealing it against air-pressure changes.

A frequency servo consists of three parts: a *discriminator*, or error signal that changes sign when the perturbed output frequency passes through that of the reference; gain to boost this signal; and path-length transducers in the dye cavity driven by the amplified signal to bring the output frequency back to the reference value. The responses of the various transducers in this loop are included in the (complex) loop gain $M(f)$, which varies with the servo response rate or Fourier frequency f (the inverse of servo response time). The loop gain can be measured by breaking the loop, terminating the ends with the same impedances that were there before the gap, injecting a sinusoidal signal at frequency f into one end, and measuring the signal [$M(f)$ times as large] returned by the loop to the other end. Circuit analysis of the servo shows that a free-running or open-loop frequency deviation of the dye beam is reduced in magnitude when the loop is closed by the factor $1/[1 - M(f)]$. It is thus desirable to keep $M(f)$ large at high response rates f to correct excursions in P

that occur even at rapid rates, to most reduce the output frequency jitter.

In doing this, however, the servo design must manage the phase shifts from all elements in the loop to keep the servo stable. The phase of $M(f)$ must be controlled over the full servo bandwidth where the magnitude of M is greater than unity, in order to keep the feedback negative and corrective and avoid oscillations of the loop, which correspond to a zero in the denominator of the feedback factor. All stable servos thus roll off the loop gain with increasing Fourier frequency to get and keep M less than 1 at high rates where phase shifts become uncontrollable. The two sources of phase shift of main concern are transducer resonances and the roll-off rate of M with f.

All path-length transducers have some form of inertia, and thus, when driven at a sufficiently high Fourier frequency f, they will resonate or reach large excursions at a drive frequency where the phase of their response is 90° relative to their low-f response. This resonance may span a narrow range of response rates, be highly nonlinear, and be critically dependent on small dimensions in the transducer (like the wall thickness of a piezoelectric element). Thus an inverse electrical filter (or "trap") added to the loop can generally only null out a transducer resonance to first order, and a servo designer's rule of thumb is to use gain roll-off to provide attenuation and a stability margin at the resonance by keeping the unity-gain frequency considerably smaller than the frequency of the lowest transducer resonance (no larger than one-third).

This upper bound on the unity-gain frequency is especially restrictive when phase shift due to the smooth, controlled roll-off of the loop gain itself is taken into account. This phase shift may be derived by interpreting mathematically the requirement of causality, that the response of the loop not precede its cause. The simplest case (most often used in practice) is for a constant roll-off slope S on a plot of log M versus log f, where S is constant over a minimum seven-octave span around unity gain (i.e., over frequencies from $\frac{1}{10}$ to 10 times the unity-gain frequency). Then servo theory gives that this phase shift is proportional to S and reaches 180° (instability) at a slope of 12 dB/octave. A second designer's rule of thumb is thus to keep the roll-off rate less than 10 dB/octave, a painfully slow rate that limits the loop gain over much of the working servo bandwidth at Fourier frequencies below the unity-gain frequency, but such is the price for stability.

To summarize the basic design of a servo loop; the unity-gain frequency is usually dictated by the lowest resonant frequency of the path-length transducers; the loop gain is made to rise at 10 dB/octave from unity gain toward lower response rates, reaching a maximum loop gain at a corner frequency determined by the drift rate of the reference [where generally $M(f)$ is leveled off, since it is pointless to provide more gain, and a tighter lock, good only at response rates where the reference is in error].

Usually, it is desirable to provide several length-control elements of different response rates in the dye cavity because the slow transducers will have larger correction ranges than the fast ones, and there is a corresponding correlation of speed with range among the forces perturbing the cavity length. The system of Fig. 10 connects the dc to 100-Hz components of the error signal to the galvo of the tipping Brewster plate, the slow (800-Hz first resonance) but long-range (30-GHz) transducer, to correct for room-temperature and air-pressure changes. With a crossover network it connects the 0.1-Hz to 10-kHz error information to the mirror M1 mounted on the piezoelectric transducer (PZT), the fast (30- to 60-kHz first resonance) but limited-range (1-GHz) element, to correct for bubbles or pressure fluctuations in the dye jet. The crossover network allows the first design rule to be followed in both the slow loop and the fast loop, to keep the servo stable. A convenient check that the servo is operating in the ring laser of Fig. 10 is to lightly push on a mirror end plate with a finger; when the servo is connected, the reflected spot off the tipping Brewster plate will reveal a proportional plate rotation, countering the direction of push and keeping P constant.

Frequency servos designed for dye lasers must address the fact that the dye oscillation is subject to brief (but frequent) momentary interruptions, or frequency excursions too fast for the servo transducers to follow due to bubbles in the jet stream. If the frequency servo can be made to automatically and rapidly relock the laser frequency to the original lock point after such an interruption, then the effect of the brief excursion can be made insignificant in the use of the laser, since this is generally to acquire time-averaged spectroscopic signals. The problem posed for the servo designer, since the laser oscillation may jump to a new dye cavity mode upon interruption (a "mode jump"), is how to unambiguously steer this arbitrary new frequency back to that of the original reference cavity mode.

Fortunately, the highest-resolution filter in the filter stack (the thick etalon) operates through a "mode-jump cycling" effect at optical rates as an optical frequency limiter and allows a relocking servo to be built. Figure 12 illustrates the relocking concept used for the laser and filter stack of Fig. 10 and 11. A discriminator is generated here by the "fast differencing" method (also called a "fringe-side" discriminator). This, one of the simplest discriminators, was the earliest type to be widely used, and is found today in commercial s.f. ring systems. A beamsplitter picks off from the output beam of Fig. 10 two low-intensity beams, one of which is transmitted through the reference cavity, the other attenuated an adjustable amount, and then

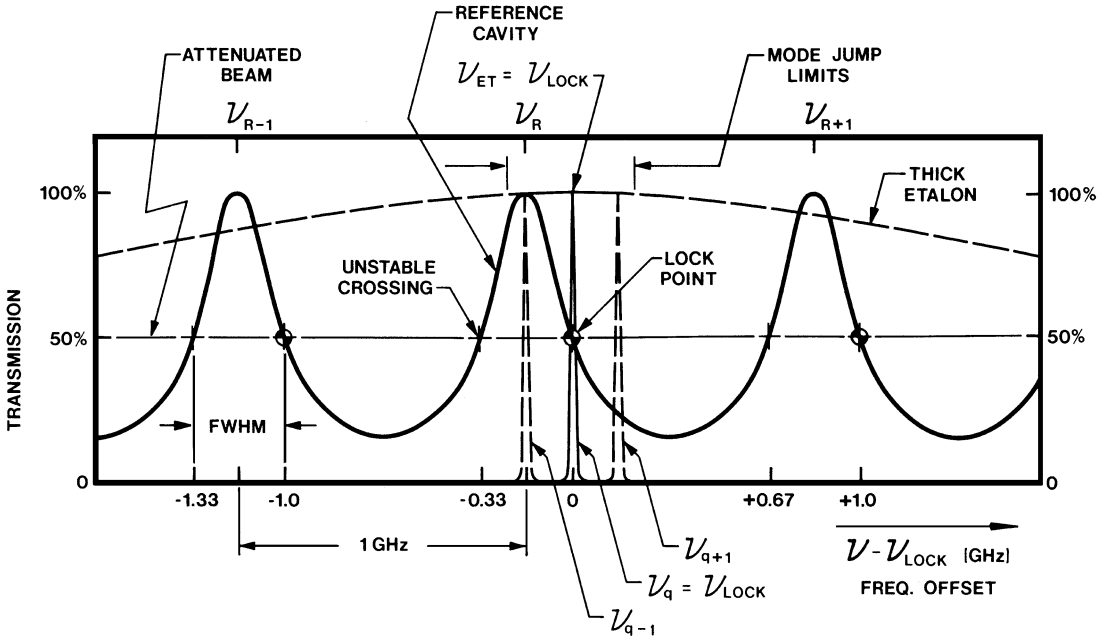


FIGURE 12 Design for an automatic relocking servo with the "fringe-side" (fast difference) discriminator used with the ring laser of Fig. 10. The mode-jump limiting frequencies defined around the lock point (the zero-crossing frequency with a negative slope) must lie within that point's acquisition range (the frequency span between positive-slope zero crossings) to insure relocking to the original reference cavity mode after a momentary interruption of lasing. (Courtesy of Coherent, Inc., Palo Alto, CA.)

the beams are detected with photocells connected to the input terminals of a fast differencing amplifier. The attenuation is adjusted to put the zero point of this difference signal at about the half-transmission point of a reference interferometer transmission peak (or half-way up the "side of the fringe"), as shown as a horizontal line labeled "attenuated beam" in Fig. 12. This level signal (independent of optical frequency) functions to prevent the mapping of intensity fluctuations into the difference, or frequency, error signal. The difference between this line, and the reference cavity transmission signal (at the negative slope zero crossing at 0 GHz frequency offset) is negative for frequencies above the zero crossing and positive below that lock point; taking the difference has generated a discriminator. There are additional potential lock points (negative-slope zero crossings) at -1.0 GHz, $+1.0$ GHz, and so forth, corresponding to adjacent reference cavity modes. These are separated by positive-slope, unstable zero crossings (at -1.33 GHz, -0.33 GHz, $+0.67$ GHz, and so forth) representing the boundaries or limits of the acquisition range of each stable lock point. The "mode-jump cycling" effect is used to contain the oscillating dye frequency within the acquisition range of the original lock point (of the reference cavity mode labeled ν_R) despite the interruption. Then when the perturbation has subsided, since there is only the original stable lock point (and no crossings to adjacent lock points) within the frequency range accessible to the dye oscillation, the servo will automatically relock as desired.

To understand mode-jump cycling, consider Fig. 12 and imagine the dye cavity modes ν_{q-1}, ν_q, ν_{q+1}, etc., being swept rapidly up in frequency by a fast perturbation, so rapidly that the thick-etalon tracking circuit cannot respond and the etalon peak is frozen in place. Originally the oscillating qth mode (at ν_q) is at the frequency of the etalon peak, but when this mode reaches an offset from the peak of $+\frac{1}{2}(c/P)$, it suffers a loss due to the etalon equal to that of the (nonoscillating) $(q-1)^{st}$ mode at $-\frac{1}{2}(c/P)$ offset, and a mode-jump of the oscillation backward by $-c/P$ in frequency becomes possible. Due to the hysteresis introduced by gain saturation, the original oscillation will generally persist out to offsets greater than $\pm\frac{1}{2}(c/P)$ (both signs occur if downsweeps are considered as well as upsweeps). Nevertheless, observations have shown there are definite limits for the actual optical frequency change after the fast perturbation has subsided, regardless of the number of wavelengths of path change the perturbation contained. Measurements for the ring laser of Fig. 10 over the visible dyes (and over the gain variation within the tuning range of each dye) give a maximum mode-jump limit span of less than $3(c/P)$ around the etalon peak (see Fig. 12). Once these limits are known, a relocking servo is

constructed by choosing the parameters of the reference cavity such that when the system is normally locked and the thick-etalon frequency is aligned with the lock frequency, only one lock point is accessible from within this mode-jump limit span. For the fast-differencing discriminator of Fig. 12, the nearest unstable crossing point is at one fringe width (or FWHM, the full width at half maximum of the reference cavity peak) lower in frequency than the lock point. For symmetrical mode-jump limiting frequencies about the lock point, a relocking design is thus achieved with a reference cavity for which the FWHM is greater than half the mode-jump limit span.

A logical problem may exist in this relocking design, in that at the first turn-on of the servo, before the thick etalon frequency and reference cavity lock points are aligned, there may be no lock point within the mode-jump limit span. (For example, the etalon frequency could be at +0.33 GHz upon turn-on in Fig. 12). What happens then is that the fast length control element (the PZT mounted mirror) is driven to a limit of its range by mode-jump cycling and saturates, or ceases momentarily to function (until a lock is acquired, and the charge is bled from the capacitance of the PZT drive circuit). The servo response rate thus momentarily drops to that of the slow loop (the galvo response rate), which by design is comparable to the response rate of the thick-etalon dither-and-track circuit. Thus, the thick-etalon frequency can move on initial turn-on to allow acquisition of a lock point.

With other discriminators the details are different but the ideas are the same. A logical trail must be built, such that recovery from a momentary interruption leads only into the acquisition range of the original reference cavity mode. For servos with unity-gain frequencies in the megahertz range (attained with electro-optic pathlength transducers), the bubble interruption may last several servo response times. This makes it necessary to clamp various transducer drive voltages within certain bounds and wait out the interruption. The response times and sequencing of elements during the recovery from the clamped state must then be tailored to stay within the desired acquistion range.

Two other methods are also commonly employed for generating frequency discriminators. The first, due to Hänsch and Couillaud, generates a symmetrical, dispersion-shaped discriminator with relatively large signal amplitudes at large offsets from the lock point, which results in prompt relocking. It derives from a frequency-dependent elliptical polarization in duced in the beam reflected from a reference cavity containing a Brewster plate with an incidence plane inclined to the plane of the linear input polarization. The second method, developed for lasers by Drever, Hall, Bjorklund, and their coworkers, is termed the "rf sideband lock," because the input beam to the reference cavity is modulated at a radio frequency with an electro-optic phase modulator. The upper and lower out-of-phase rf sidebands that this generates are split off from the central carrier by the modulation frequency, chosen to be greater than the width of the reference cavity resonance. When the laser (carrier) frequency is on resonance with a reference cavity mode, there is a buildup inside the cavity of standing-wave energy at the carrier frequency but not at the side band frequencies. A discriminator is formed by heterodyning in a fast detector the fields reflected from the cavity, and phase-sensitively detecting the beat frequencies against the rf modulation source. Analysis of this signal shows it to be the optical analog of the Pound or "magic-tee" stabilizer for microwave sources, and to have the property of acting as an optical phase detector on resonance due to the phase memory of the buildup field leaking out of the cavity.

The linewidths produced by these systems are a function of the speed of the path-length transducers and the resultant unity-gain frequency or Fourier bandwidth of the servo. The commercial system of Fig. 10 with a tipping Brewster plate and a PZT mounted mirror as transducers, using a fast-differencing discriminator, achieves a 5- to 10-kHz servo bandwidth and an optical frequency jitter of 200–500 kHz rms measured over a 1-sec sampling time with a 10-kHz detector bandwidth. The laboratory system with the best reported performance adds an intracavity AD^*P electrooptic phase modulator to the list of pathlength transducers (a crystal whose index of refraction depends on an applied high voltage) and, using an rf sideband lock with a 40-MHz rf frequency, achieves a servo bandwidth of 3–4 MHz. The resulting jitter was so small that special techniques were required to measure it. A helium–neon laser was similarly rf sideband locked to an adjacent mode of the same reference cavity, so that problems with drift and isolation of the reference would not mask the measurement of the locked frequency jitter. The beat note between these two independently locked lasers was examined and found to be less than 100 Hz in width—only 2 parts in 10^{13} of the dye laser's output frequency. Such systems offer enormous opportunities for precision measurements in physics and other sciences.

The ability of a relocking system to rapidly and automatically recover from a momentary perturbation is what makes electronic control of an s.f. dye laser effective. This ability is put to good use in the most recently evolved version of this instrument, where the laser is coupled to a wavemeter and computer-driven over a 10-THz or 333-cm^{-1} scan width. Such a long scan is done by piecing together many shorter scans stored in memory, where the endpoints of the short scans are abutted by the computer to the wavelength precision of the coupled wavemeter to present seamless output spectra. All of the filter stack

elements must be moved to cover this long scan, and reset each in turn after each has been scanned over its free spectral range. To do this, the computer is simply programmed to introduce a "momentary perturbation" of its own by clamping the drive signal for each element at its nominal value, opening the loop, resetting the appropriate element, and again closing the loop. The system then reacquires the original lock point just as though this had been an external perturbation.

E. Single-Frequency Dye Laser Output Tuning Spectrum

The argon and krypton ion pump lasers for the s.f. cw dye laser offer a wide span of pump wavelengths, and the dye laser performs best where the absorption maxima of efficient dyes can be matched to strong pump lines. The overlapping tuning ranges of the 11 combinations giving the highest reported output powers over the fundamental tuning range (the visible spectrum) in a single-frequency laser are shown in Fig. 13. Additional data for this figure are listed in Table I. All of the tuning curves are for the ring laser of Fig. 10, with the exceptions of the short-wavelength half of the stiebene 1 curve and all of the IR 140 curve. For these two ends of the fundamental spectrum, the laser was set up in a standing-wave (three-mirror) configuration, as a ring configuration was not needed for these low-gain combinations and the elimination of the optical diode losses gave better performance there.

Table I gives the pump powers used for these curves. The pump wavelengths or bands (groups of lines that lase simultaneously) from the argon ion laser are identified as UV for the ultraviolet (333- to 368-nm) pump band, BG for the blue-green (455- to 514-nm) band, B for the blue (488-nm) wavelength, and G for the green (514-nm) wavelength; these last two are the two strongest individual lines and are often used singly. Similarly, for the krypton ion laser, V indicates the violet (407–415 nm) band, R the red (647- to 676-nm) band, and IR the infrared (753- to 799-nm) band. The prime example of a strong combination is the high-gain dye rhodamine 6G, which is efficiently pumped either by the blue-green argon band or single-line (for the best dye transverse mode control, important in intracavity doubling) with the 514-nm line, the highest-power line of all ion laser outputs. A single-frequency stabilized output of 5.6 W has been attained with this dye.

The dye laser was fully optimized for these fundamental tuning curves, making the available input power be the limit to the output reached. With the weaker pump lines, the radius of the fold mirror (M1 of Fig. 10) was decreased to tighten the dye and matching pump focal areas and keep the gain well above the 2.5% s.f. cavity losses. (The 10-cm radius for the red dyes became 7.5 cm for the stilbenes, coumarins, and S9M, and 5.0 cm for IR140). The transmission of the output coupler was also optimized across the span of each dye (the breaks on several of the tuning curves are places where a different output coupler was

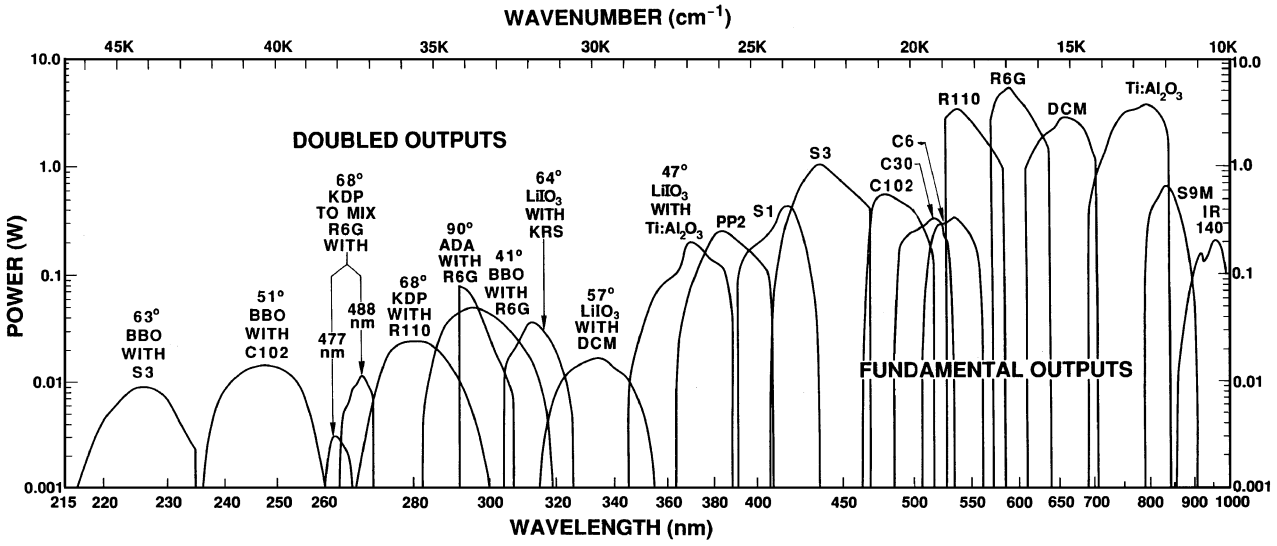


FIGURE 13 Output tuning curves for the single frequency laser, pumped with the ion laser inputs listed in Table I. The fundamental outputs are extended into the ultraviolet by adding nonlinear crystals to the cavity to produce doubling and mixing. For discussion of the updated results in the media PP2, S1, S3, Ti:Al_2O_3, and doubling with BBO, see the discussion of recent developments at the end of the article. [UV data and Ti:Al_2O_3 results courtesy of Coherent, Inc., Palo Alto, CA, and visible data adapted with permission from Johnston, T. F., Jr., Brady, R. H., and Proffitt, W. (1982). *Appl. Optics* **21,** 2312.]

TABLE Ia Fundamental Single-Frequency Outputs

Gain medium symbol	Molecular weight (amu)	Common name	Peak output λ_D (nm)	Peak output P_{max} (W)	$n\pi w_D^2$ (10^{-6} cm^2)	Optimum coupling t_0 (%)	Gain $k_0 L$ (%)	I_s (MW/cm^2)	Pump laser P_{in} (W)	Pump laser λ_p^a
PP2	542	Polyphenyl 2	383	0.25	5.7	1.9	8	2.2	3.4	SUV
S1	569	Stilbene 1[b]	415	0.42	6.2	2.7	11	1.8	6.0	UV
S3	562	Stilbene 3	435	1.0	6.5	3.9	18	2.0	7.0	UV
C102	255	Coumarin 102	477	0.58	7.1	2.9	12	2.4	4.8	V
C30	347	Coumarin 30	518	0.38	7.7	2.4	10	2.1	4.6	V
C6	350	Coumarin 6[c]	535	0.35	8.0	3.8	16	0.8	6.0	B
R110	367	Rhodamine 110[c,d]	540	3.6	14	10	60	0.6	23	BG
R6G	479	Rhodamine 6G[c,d]	593	5.6	16	11	70	0.8	24	BG
DCM	303	Dicyano-methylene[c,d]	661	2.9	17	5.1	23	1.6	20	BG
Ti:Al_2O_3	—	Titanium-sapphire single crystal	790	3.6	94	5.3	23	0.4	20	BG
S9M	529	Styryl 9M	835	0.65	12	4.5	21	0.6	7.5	G
IR 140	779	Infrared dye 140[b]	960	0.2	6.2	~1	—	—	3.0	IR

TABLE Ib Mixed and Doubled Single-Frequency Outputs

Crystal symbol	Match angle	Crystal name	Peak output λ_{UV} (nm)	Peak output P_{max} (mW)	Gain medium	Ion laser mixing line (nm)	Pump laser P_{in} (W)	Pump laser λ_p^a
BBO	62°	Beta-barium borate	228	9	S3	—	7.1	UV
BBO	53°		248	15	C102	—	5.5	V
KDP	80°[e]	Potassium dihydrogen phosphate	262	3	R6G	477	7.5	BG
KDP	74°[e]		268	10	R6G	488	7.5	BG
KDP	68°		280	24	R110[c]	—	6.8	G
BBO	41°	Beta-barium borate	295	48	R6G[c]	—	8.5	G
ADA	90°	Ammonium dihydrogen arsenate	292	77	R6G	—	6.0	G
$LiIO_3$	65°	Lithium iodate	313	36	Kiton Red S	—	7.5	G
$LiIO_3$	57°		332	16	DCM	—	7.5	BG
$LiIO_3$	47°		370	200	Ti:Al_2O_3	—	20	BG

[a] Codes for the pump wavelengths are explained in the text, except for SUV, which is the 300–336 nm pump band.
[b] The IR 140 curve used a standing wave laser configuration; all others are for a ring laser. (The standing-wave portion of the S1 curve in the earlier edition is here replaced by ring laser data.)
[c] The triplet-state quencher COT was added here to the dye mix.
[d] Special water-based solvent AMX used (see the text).
[e] The crystal for R110 dye (cut for a match angle of 68° at Brewster incidence) was used here at large tip angle.

substituted). For the BG pumped dyes, a special dye solvent (called AMX by Johnston) consisting of three parts ammonyx LO and one part ethylene glycol, chilled to 10°C, replaced the normal ethylene glycol jet. This reduced the optical distortion of the jet from thermal lensing, which otherwise would limit the output at these input powers of 20 W or more. The ammonyx LO is mostly water, which has a low refractive index change with temperature and a high specific heat, yet due to the remaining component (a soap, lauryl amine oxide) has a high viscosity at this temperature and can form a stable jet. The additive COT (cyclooctatetraene), which relaxes the triplet state of the dye and enhances output, was used with this AMX solvent and in three other cases identified by footnote *c* in Table I. Small amounts of dissolving agents permitted dye concentrations sufficient for 80–90% pump beam absorption in the 0.1-mm jet thickness. For all of the Fig. 13 dye recipes, the observed locked jitter was less than 0.5 MHz

rms, although when COT is added it must be fresh (stored in a vacuum ampule or nitrogen-purged container to avoid contact with oxygen, which turns it oily and immiscible). The vertical drops at the ends of some of the tuning curves are birefringent filter "break points," where the oscillation jumps from the main BRF transmission lobe to a side lobe closer to the peak of the dye gain curve. Tuning can be extended slightly beyond these breaks, by substituting a cavity optic coated to have a high transmission at the side-lobe wavelength.

Just as in the short-pulse dye laser case, these fundamental outputs are extended into the ultraviolet by doubling and mixing. In the short-pulse case, because of the high peak powers, high conversion efficiencies for the nonlinear processes are reached easily and the problem there is one of limiting the peak powers to less than the damage thresholds of the nonlinear crystals. The converse problem exists in the s.f. cw case, that of how to raise the fundamental power incident on the crystal sufficiently to reach a usable conversion efficiency. This is solved by placing the nonlinear crystal inside the dye laser cavity and replacing the output coupler with a high reflector, to reach circulating fundamental powers in the range 20–60 W and s.f. ultraviolet outputs of several milliwatts. The circulating intensity I_e at the focus in the dye jet is proportional (neglecting the threshold term) to the inverse of the cavity dissipative losses a, as shown by solving Eqs. (4) and (9) for I_e and setting $t = 0$, to give $I_e = I_s k_0 L/a$. The crystal absorption and scatter losses contribute to a and must be quite low (less than 1%) to reach the desired circulating powers. Therefore, high-quality, well-polished crystals are used in the cavity at Brewster's incidence angle.

The high-power dyes R110 to LD700, for which doubling is attractive, span a broad fundamental wavelength range, best covered without gaps by angle-tuned phase matching. Angle matching produces a lower conversion efficiency to the ultraviolet than the alternative, temperature-tuned matching (with a fixed matching angle of 90°). But unlike this alternative, angle matching is insensitive to temperature gradients in the crystal and requires no slow, high-hysteresis temperature servo. (Temperature-tuned matching was the first method used for intracavity doubling in s.f. ring lasers, and it is still employed to double the long-wavelength half of the rhodamine 6G spectrum. Matching this half spectrum is done by heating an ADA crystal substituted for the astigmatic compensation rhomb of Fig. 10, from room temperature to ~100°C, producing the results in Fig. 13 and Table I).

The angle-matched conversion efficiency is lower because at input beam propagation directions other than 90° (or 0°) to the optic axis the harmonic beam is doubly refracted, or separates ("walks off") at a slight angle from the fundamental beam. This separation that develops between the beams effectively limits the interaction length, or distance over which the harmonic radiation is most efficiently produced, to the distance it takes for harmonic light generated at the crystal input face to be displaced by a beam diameter from the fundamental. This interaction length is the beam diameter divided by the walk-off angle. The conversion efficiency still rises with lengths longer than this, but the rate of increase slows and most of the additional harmonic power goes to make the UV beam progressively more elliptical and distorted in transverse mode profile.

Thus the angle-matched nonlinear crystals used with the dye laser, which must generate a clean beam profile to be usable in spectroscopy, are chosen to have a length of one interaction distance. This choice has the interesting consequence that the nonlinear crystal need not be placed in a beam focus in the cavity. If the crystal is placed in the collimated arm of the cavity where the fundamental beam is of large diameter (0.7 mm), there is no loss in conversion efficiency if the crystal length is scaled with the beam diameter to still be one interaction distance. This amounts to 1-cm-long $LiIO_3$ crystals and 4-cm-long KDP crystals for placement between M3 and the optical diode in the upper arm of the cavity of Fig. 10. This placement is chosen because of its mechanical convenience and because in a parallel beam the crystal may be made to produce negligible interference with the single-frequency filter stack. The doubled beam is coupled out of the resonator with a Brewster-angle dichroic beam splitter.

For most angular orientations in the cavity, the nonlinear crystal would act as a fourth plate of the birefringent filter, and when angle-tuned would surely disrupt the s.f. filter stack. This is avoided if the dye beam sees only one (pure) refractive index, which for the two crystals above is the pure ordinary index. The crystal is oriented so that the electric field of the refracted dye beam is perpendicular to the crystal optic axis, then the dye beam is "resolved" as a pure ordinary wave, and the crystal acts then as would a plate of glass instead of as a doubly refractive material. Setting up this perpendicular condition uses one of a total of two angular degrees of freedom; orienting the crystal for Brewster incidence (to maintain the high circulating fundamental beam power) uses the other. One of these conditions must be relaxed to permit angle-phase matching; for this adjustment the crystal is rotated about an axis parallel to the electric field direction, maintaining the "no-disruption" perpendicular condition. This gives smooth tuning (normal action of the s.f. filter stack) but produces cavity reflection losses that increase from zero as the square of the crystal tip angle away from Brewster's incidence. Several different cuts of crystal are used to minimize these reflection losses, each cut putting the Brewster incidence no-loss point at a wavelength in mid-

tuning range of each dye to be doubled, with the results shown in Fig. 13 and Table I. These ultraviolet doubled-dye outputs scan over twice the range that the fundamental is scanned and have twice the jitter.

Tunable ultraviolet s.f. outputs extending to even shorter wavelengths, which still take advantage of the strong red dyes, can be generated by nonlinear mixing of the dye beam with a shorter-wavelength, single-frequency beam from an ion laser, to make the sum frequency. Mixing is just like doubling, but with unequal energies for the input photons; this merely requires slightly different phase-matching angles. For the results of Fig. 13, an ion laser cavity was folded to intersect and include in the ion beam cavity, the 68° KDP crystal in the dye cavity. The collimated-arm crystal placement permits this; the ion and dye beams must be collinear inside the crystal for the summing interaction, but at the Brewster angle faces of the crystal the two wavelengths are refracted differentially and separate with an angle between them of five times the beam divergence. The throw distance to the nearest dye cavity mirror in the collimated arm is large enough for the ion beam to clear the edge of the dye mirror, and on the other side of the crystal for the dye beam to clear the edge of an inserted ion cavity end mirror. The only common element in the two cavities is the KDP crystal. The ion laser is operated in a standing-wave cavity as the phase-matching condition with the traveling wave dye beam picks out the desired running wave ion beam component in the summing interaction. The mixed and doubled dye outputs of Fig. 13 allow single-frequency, high-resolution spectroscopy over the range 260–400 nm. Sum-frequency mixing has been used to generate even shorter UV tunable s.f. outputs, down to 194 nm, in scattered wavelength bands appropriate for particular experiments. The beam at the sum frequency has the scan range of the dye laser and a jitter equal to the sum of the ion laser and dye laser jitter.

F. Gains and Saturation Intensities for Common Dyes

The tuning curve data of Table I may be analyzed to give gains and saturation intensities for these dyes, under the listed pumping conditions. This permits an experimental check on the uniform plane-wave theoretical analysis of Section II.

The analysis uses the optimum output coupling values $t = t_0$ determined experimentally with the tuning curves and the estimate of the single-frequency cavity dissipative losses $a = 2.5\%$ for the ring laser of Fig. 10. Differentiating Eq. (9) with respect to t to find the output mirror transmission that maximizes the output power gives

$$t_0 = a\left[(k_0 L/a)^{1/2} - 1\right] \tag{20}$$

from which

$$k_0 L = a[(t_0/a) + 1]^2. \tag{21}$$

The gain values computed from Eq. (21) are listed in Table Ia for all the dyes (except IR 140, where lasing was too close to threshold to determine an optimum coupling). The output power with this optimum coupling is

$$P_{\max} = n\pi w_D^2 I_s\left(t_0^2/a\right), \tag{22}$$

which shows that the maximum output is proportional to the inverse of the cavity losses a. The factor of $n = 1.4$, the index of refraction of the dye jet (at λ_D), did not appear in Eq. (9) for the plane-wave, longitudinally pumped case. It is appropriate in expressions like Eqs. (9) or (22) containing a focal area in a Brewster angle dye medium (the present case), because a beam of cross-sectional area πw_D^2 (normal to the beam direction) is expanded in area upon refraction at the Brewster surface to $n\pi w_D^2$. The prism expanders of Figs. 5 and 6 illustrate this expansion. [Note also that this factor of n was erroneously omitted in Eq. (6) of Johnston *et al.*, which resulted in saturation intensities reported there being 1.4 times too large.] Table Ia lists the correct beam focal areas. When the output power was measured using an optimum output coupler, inverting Eq. (22) gives the saturation intensity

$$I_s = \frac{P_{\max}}{n\pi w_D^2}\left(\frac{a}{t_0^2}\right), \tag{23}$$

and if a nonoptimum transmission t was used, Eq. (9) (including the factor of n) is solved for I_s, yielding the values in Table Ia.

The same analysis can be applied over a dye's full tuning range, to show experimentally the wavelength dependence of the gain and saturation intensity. This was done for the high-power tuning curve of rhodamine 6G, by determining the optimum transmission at 7 points over the tuning range (Fig. 14b). The Fig. 14a tuning curves show as well the high conversion efficiency from broadband output (~2 GHz linewidth without etalons) to single-frequency operation of a ring laser. These curves were taken with a single output coupler to produce smooth full-range tuning data (this coupler gave a slightly lower peak output than the one that gave the peak power listed in Table Ia). Also, the t_0 values plotted were chosen (within the range of experimental error) to give smooth curves for $k_0 L$, I_s [labeled EXPT in Fig. 14c and d] when put into Eq. (21) or Eq. (23).

These experimental values measure the response of the dye medium to changes in resonator loss, and as such represent the average over the active volume of the microscopic inversion variations in the dye. A direct comparison with theory would thus involve a (numerical) integration of the plane-wave equations over the focal volume,

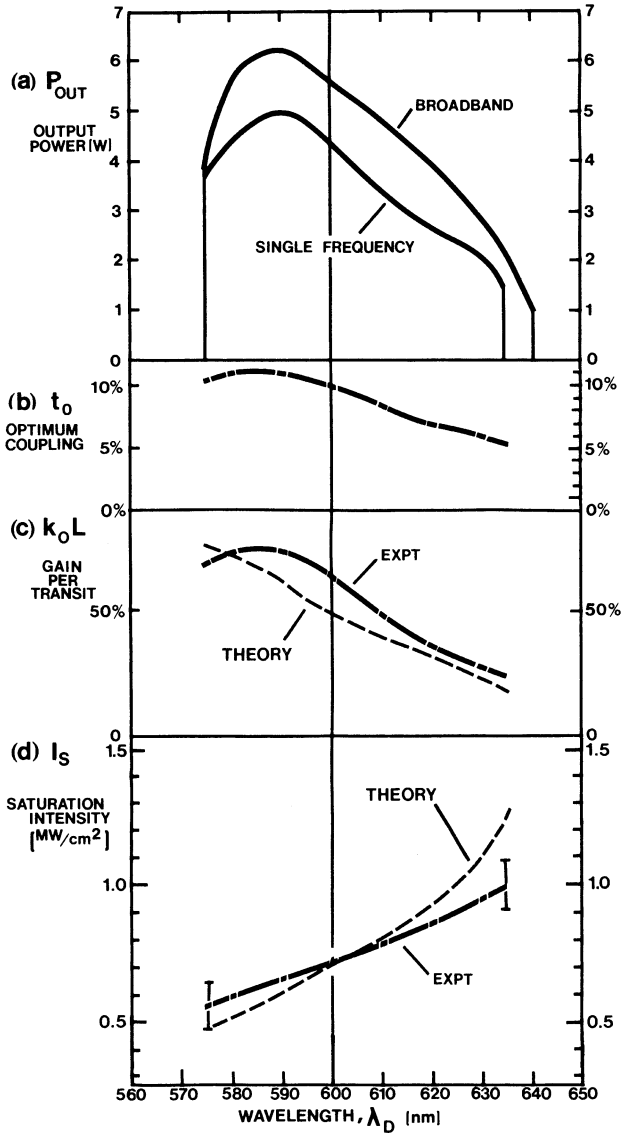


FIGURE 14 Comparison of the gain and saturation intensities derived from measurements of optimum coupling (EXPT), with the theoretical plane-wave expressions using spectroscopic cross sections (THEORY) for rhodamine 6G dye in AMX solvent at 24 W pumping. [Adapted with corrections and permission from Johnston, T. F., Jr., Brady, R. H., and Proffitt, W. (1982). *Appl. Optics* **21,** 2311.]

a procedure more complex than is warranted in view of the moderate accuracy of the measured t_0 values (about $\pm 10\%$ of their means). A simpler approximate treatment is to first establish upper bounds for I_s and $k_0 L$ by putting into the plane-wave Eqs. (8) and (11) the maximum pump intensity reached in the experiment (the intensity at the center of the Gaussian mode profile, on the input side of the jet). This saturation intensity upper bound is then reduced by an estimate of the ratio of the average pump intensity to this maximum value. This gain upper bound is then reduced by the factor $1/(1+\rho_0)$ quoted earlier (in Section IV.A) derived from the overlap integral at the focus in longitudinal pumping of a dye beam of Gaussian waist radius w_D and pump beam of Gaussian waist radius w_p, where $\rho_0 = w_D^2/w_p^2$. The upper bounds are evaluated for $\lambda_D = 590$ nm, at the peak of the experimental gain curve.

The stimulated pumping rate $(\lambda_p/hc)\sigma_p I_p$ (transitions per second) appearing in Eq. (8) is replaced for a multiline pump beam by the sum of such terms over the wavelengths present. When weighted by the appropriate pump cross sections (Fig. 4), the 24-W BG pump is found to be equivalent to 15 W of 514-nm pumping. To convert power to intensity, note that a TEM_{00} mode beam of total power P and Gaussian $1/e^2$-radius w has a maximum intensity on axis of $2P/\pi w^2$ and an average intensity of *half* this, if the beam cross-sectional area is taken to be πw^2. The calculated pump focal area in the jet was $n\pi w_p^2 = 6.7 \times 10^{-6}$ cm^2, giving $\rho_0 = 2.3$, and a maximum pump intensity of $I_p = 4.0$ MW/cm^2. The spontaneous decay rate for rhodamine 6G is $1/\tau\varphi = 0.2$ GHz; the green-equivalent pump intensity above yields a maximum pumping rate of 1.6 GHz, and Eq. (8) gives an upper-bound saturation intensity of $I_s = 4.0$ MW/cm^2. Because this pump rate is much larger than the spontaneous decay rate, the saturation intensity is proportional to pump intensity and the upper bound can be cut in half to account for the radial, off-axis falloff of the pump intensity. It is cut again by a factor ranging from 0.32 to 0.39, the ratio of the maximum to the average on-axis intensity, to account for exponential absorption along the axis to a 5–10% transmission through the jet. This gives theoretical I_s estimates for the cw laser case of 0.6–0.8 MW/cm^2. These agree satisfactorily (within the error estimates) with both experimental values at this wavelength, 0.8 MW/cm^2 from the data of Fig. 13 and 0.65 MW/cm^2 from the data of Fig. 14.

The gain upper bound, for $N = 1.3 \times 10^{18}$ cm^{-3} (from the estimated dye concentration of 2×10^{-3} M), $L = 0.012$ cm, and a maximum inversion fraction (calculated from the relative transition rates above) of 0.88, is given by Eq. (11) as 210% per pass. Applying the Gaussian beam reduction factor of 0.30 for the 2.3 focal area ratio gives a theoretical gain estimate of 64% per pass, which agrees satisfactorily with the two experimental results of 70 and 73%.

The large inversion fraction implies that the dye jet was bleached on the pump input side, violating the Beer's law absorption profile assumed in deriving both reduction factors and showing the roughness of the present approximations. Thus, the best test of the plane-wave theory is probably the comparisons of the predicted and observed wavelength dependencies of Fig. 14(c) and (d), which

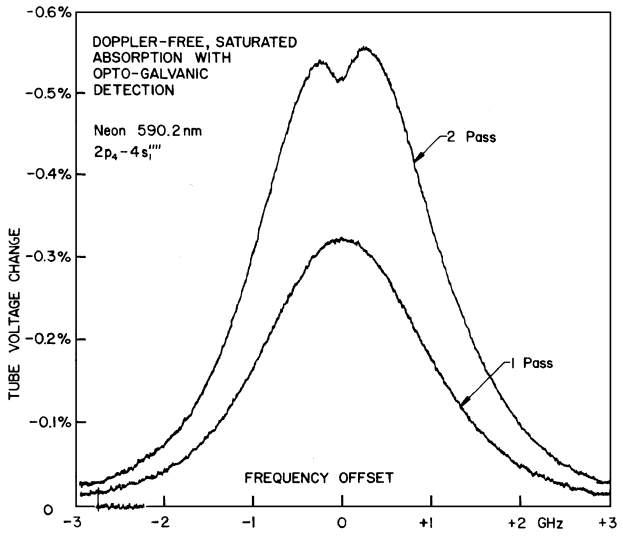


FIGURE 15 Doppler-free and Doppler-limited laser spectroscopy compared on a neon absorption line, using opto-galvanic detection. The change in voltage across a helium–neon discharge tube was recorded as the single frequency of a coaxial dye laser beam (passed either once or twice through the discharge) and was scanned through the 590.2-nm resonance. [Adapted with permission from Johnston, T. F., Jr. (1978). *Laser Focus* **Mar.,** 58–63.]

involve fewer assumptions. Here the theoretical curves are scaled from the $\lambda_D = 590$ nm results according to Eqs. (7) and (8) by the measured cross sections of Fig. 4. Using the smaller of the two theoretical saturation intensity values gave the best overlap with the experimental curve. The plane-wave analysis is seen to predict satisfactorily the variation of I_s and $k_0 L$ over the tuning range observed in an actual cw single-frequency dye laser.

Figure 15 is an example of sub-Doppler saturated absorption spectroscopy, the basic application of the scanning single-frequency dye laser. This example was chosen to display the connection of laser spectroscopy with the older, Doppler-limited spectroscopy, and consequently is at a low resolution (not at all near the 1-MHz limit imposed by the frequency jitter of this stabilized laser). The laser beam was passed (through end windows) along the axis of a narrow-bore discharge tube run at constant current in a mixture of helium and neon gas. As the frequency of the dye laser scanned through the frequencies of neon spectral lines, power was absorbed from the laser beam, and proportional changes in the discharge power were recorded. This has been appropriately named optogalvanic (current-producing light) detection and is a detection scheme with a wide dynamic range and a good signal-to-noise ratio that is often used with spectral calibration lines for dye laser wave-meters.

The lower trace of Fig. 15 shows the absorption signal when the beam was passed once through the tube and shows a classical Doppler line profile, obtainable before the existence of lasers by high-resolution interferometry. This Gaussian line shape reveals the velocity distribution of the moving atoms in the gas and not the underlying homogeneous width of the spectral line (due to the finite lifetimes of the absorbing states). For laser frequencies below line center (negative offsets in Fig. 15), atoms with an on-axis velocity component along the laser beam (but in an opposed direction) see the laser frequency up-shifted into resonance and absorb the light. Since the atoms in the gas are moving in all directions with a finite average speed, the most likely axial velocity component in the distribution is zero, and the absorption increases (according to the Gaussian weighting function) as the laser frequency moves toward line center. The process is reversed as the laser frequency scans above line center, where the laser light is absorbed by a decreasing number of atoms seeing a down-shifted laser frequency.

If now the transmitted beam is retroreflected to make a second pass in the opposite direction through the tube, the absorption is doubled as expected for off-line center laser frequencies (as shown in the upper trace of Fig. 15), as both a right-going velocity group and a left-going velocity group of atoms have the correct Doppler shifts for absorption from one of the beams. But when the laser scans to line center, the two distinct velocity groups coalesce into a degenerate "zero-axial-velocity" group and there is a saturated absorption dip of width equal to the homogeneous width, marking the center of the line. The appearance of the dip requires sufficient laser power to perturb the velocity profile; a saturation intensity for absorption on the spectral line must be reached. Gas-phase spectral lines are typically 10^{-5} times narrower than the tuning range of a dye and the saturation intensities are correspondingly smaller than those of dyes, so the required intensities are readily reached (40 mW was the input here) even in an unfocused dye laser output beam. The improvement in ability to locate the central position of the line in Fig. 15 is only about an order of magnitude but is still enough to show, by the asymmetry of the dip, the presence of an isotope shift for the 9% of ^{22}Ne in the natural neon gas fill, to the high-frequency side of the line center of the main (91%)^{20}Ne component. For molecular absorption lines, the improvement in resolution reaches a factor of 10^3. This ability to see inside the classical Doppler limit throughout the dye tuning spectrum has revolutionized modern spectroscopy.

V. ULTRASHORT-PULSE (MODE-LOCKED) DYE LASERS

Of the three scientific dye laser types, the mode-locked branch is evolving most rapidly at present, as shown by

the density of accomplishments in recent years in Fig. 1. This section covers the basics of the mode-locking of dye lasers.

A. Pulse Formation by Phase-Locking of Cavity Modes

In TEM_{00} transverse mode operation, the mode frequencies ν_q of a standing-wave dye laser with resonator optical path length d are given by the longitudinal-mode values $\nu = q(c/2d)$. Here the integer q (of order 10^6) differs by 1 for adjacent modes, making the beat or difference frequencies between nearest neighbors in multimode oscillation all nominally $c/2d$. The beats are generally not precisely at this value, because of mode pulling and pushing effects. Associated with a gain peak at some center frequency there is an index of refraction variation given by the Kramers–Kronig relations tending to pull the mode frequencies slightly toward the center frequency; mode frequencies are slightly repelled from an absorption dip. The homogeneous-width "holes" in the gain curve caused by the competition between modes in a multimode laser thus normally cause a slight random shifting of nearest-neighbor beat frequencies away from the nominal value. This situation is best described by saying that the phases of the beat frequencies are normally random and uncorrelated.

By Fourier's theorem, however, if the phases could all be made constant (or "locked"), then the sum of the many mode amplitudes spaced regularly in frequency by differences of $c/2d$ would produce a periodic function, of period $T = 2d/c$. If initially these locked-mode amplitudes were phased for maximum constructive interference at some point in the cavity (producing a high peak power there), at time intervals of T they would return to this state at that location. The laser's steady multimode output would be converted to a repetitive train of pulses, with an interpulse spacing of T, corresponding to a circulating pulse striking the output coupler at intervals of the cavity round-trip transit time T. The width τ_D (FWHM) of this dye output pulse by Fourier's theorem would be given by a constant divided by the width $\Delta\nu$ (FWHM) of the spectral distribution of amplitudes of the locked modes, the constant being always of order unity but dependent on the form of the pulse shape.

To produce this locking of phases requires a small gain (or loss) term sufficient to overcome the random mode pulling effects; to establish the constructive initial phase distribution, this term should favor a high peak power pulse. Such a term is provided in the argon ion and Nd:YAG pump lasers by placing at one end of their resonators an acoustooptic loss modulator, or mode-locker. A sinusoidal radiofrequency voltage drives an acoustic wave in the mode-locker cell, giving a diffraction loss of the laser beam varying in time at a frequency precisely adjusted to equal the mode spacing frequency of the resonator. In frequency-domain terms, the sidebands introduced on a longitudinal mode carrier frequency by this amplitude modulation produce zero beats with the adjacent mode frequencies to lock the modes. In time-domain terms, the same result occurs because the formation of a short pulse allows oscillation with the least loss, as the pulse can pass through the loss modulator at the times of zero diffraction loss. Because input energy (other than that of the laser beam itself) is required to drive the mode-locker, this is called "active" mode-locking. Typically, a long pump laser is used giving a mode spacing of 76 MHz and a period of 13 nsec. The green pumping pulses from a mode-locked argon laser are about 120 psec long and from a mode-locked, doubled YAG laser are about 70 psec long, the difference reflecting the different gain bandwidths for these lasers.

The dye laser is similarly mode-locked by pumping synchronously (with the cavity round-trip transit times matched as will be described) with a mode-locked pump laser, producing repetitive gain modulation. Alternatively, mode-locking is produced (with constant cw pumping) by adding a second dye jet to the cavity, containing a saturable absorber dye. This adds a loss term that is least (saturates the most) for a high peak-power circulating pulse. This method is called "passive" mode-locking since the absorber is driven by the laser energy itself. Both mode-locking methods may be used together in a "hybrid" mode-locked laser as shown in Fig. 16.

To match the dye cavity length to that of the long pump laser, while keeping the dye laser compact, the beam in Fig. 16 is folded in a zig-zag path by the two flat mirrors. The focal diameters produced in the gain jet by the mirrors M1 and M2 and in the absorber jet by mirrors M3 and M4 are given as in the cw case by the three-mirror analysis of Fig. 8. A translation stage (not shown) on the output end of the laser allows precise length adjustment. In the figure, a cavity-dumper assembly (which can be translated) is shown in place of an output mirror. This device is a time-gated output coupler, driven in synchronism with the mode-locker of the pump laser, and is often used to reduce the repetition rate of the train of output pulses while increasing the energy per pulse. An acoustic wave is turned on to generate a diffraction grating in the Brewster cell at the M5 and M6 focus for several nanoseconds at the time of passage of the light pulse. The interference of the diffracted and direct beams returned by M6 is phased in the second diffraction to couple nearly 100% of the pulse energy into the diffracted beam direction, where it is directed out of the laser by the take-out prism. Compared to a transmitting mirror, the use of all high reflectors and this large coupling factor in cavity-dumped operation gives 10

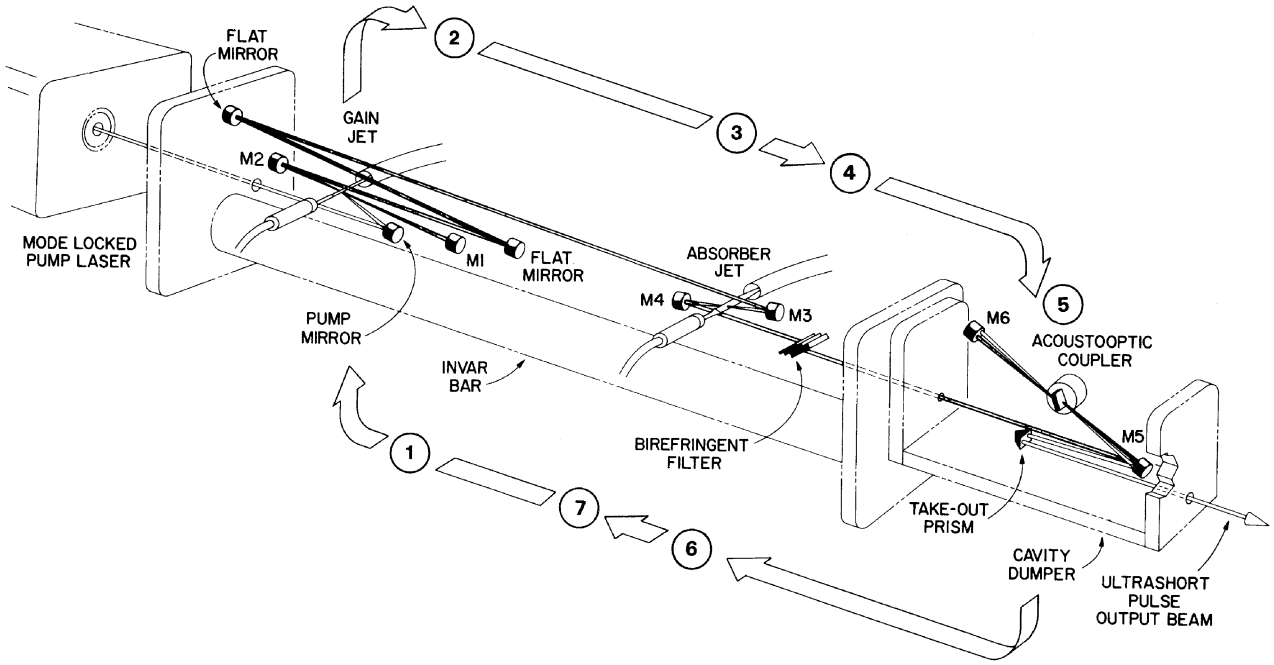


FIGURE 16 Ultrashort pulse dye laser of current design for "hybrid" mode-locking (employing both synchronous pumping and a saturable absorber). Shown in place of an output mirror is a cavity dumper assembly, which is driven in synchronism with the circulating pulse to couple out a reduced number of pulses, at higher energy per pulse. (Courtesy of Coherent, Inc., Palo Alto, CA.)

or more times the output energy per pulse, at 1/20 the repetition rate.

B. Pulse-Shaping by Saturable Gain and Absorption

The number of modes phased together in the dye laser is so large (10^4 in a picosecond pulse) that the process is most conveniently described in the time domain as the shaping of the pulse in passage through the elements of the cavity. A picosecond pulse is spatially only 0.3 mm long and interacts sequentially with each element. In Fig. 17, the circulating pulse is moving to the right (time increases to the left), and the effect of each (numbered) passage in Fig. 16 is shown schematically.

In passage through the gain jet (1, 2) the leading edge of the pulse is amplified, which increasingly depletes the gain over the remainder of the pulse, since on the time scale of the dye pulse the rates of pumping transitions and spontaneous decays are slow. This steepens the leading edge and advances the timing of the peak of the pulse. For good mode-locking, the focal area in the absorber jet is made smaller than in the gain jet (note the shorter-focal-length mirrors in Fig. 16) to ensure faster saturation in the absorber (3, 7) to further steepen the leading edge (and slightly retard the peak). The uniform, linear loss of an output coupler (5) leaves the pulse shape unchanged, but the attenuation of the trailing edge here, in combination with the growing middle and attenuated leading edge, means a net pulse shortening in passage through these three elements. What keeps the pulse width from collapsing to zero is the finite spectral bandwidth of the frequency filter used in the cavity to tune the center wavelength of the pulse. In Fig. 16, this is shown as a three-plate birefringent filter (whose transmission function is shown in Fig. 11), but broader two- and one-plate filters are also used for shorter pulses. By Fourier's theorem, the clipping by the filter of the highest and lowest frequencies in the pulse spreads the pulse shape and balances the compression.

The frequency filter functions in the mode-locking process to clean up the frequency noise and give transform-limited pulses. Pulse formation starts from noise bursts, spontaneous amplitude fluctuations that are limited in rate only by the dye gain bandwidth and that can be quite rapid. These are amplified and filtered in the "good mode-locking" regime, but can persist as temporal substructure in a nonoptimum pulse. The quality of a mode-locked pulse train is judged by measuring both its temporal and spectral widths and by checking that the product of these is close to the Fourier transform constant (implying a consistent pulse shape).

The pulse-shaping process just described applies directly to passive mode-locking, where the gain gradually recovers with cw pumping over a cavity transit time, and

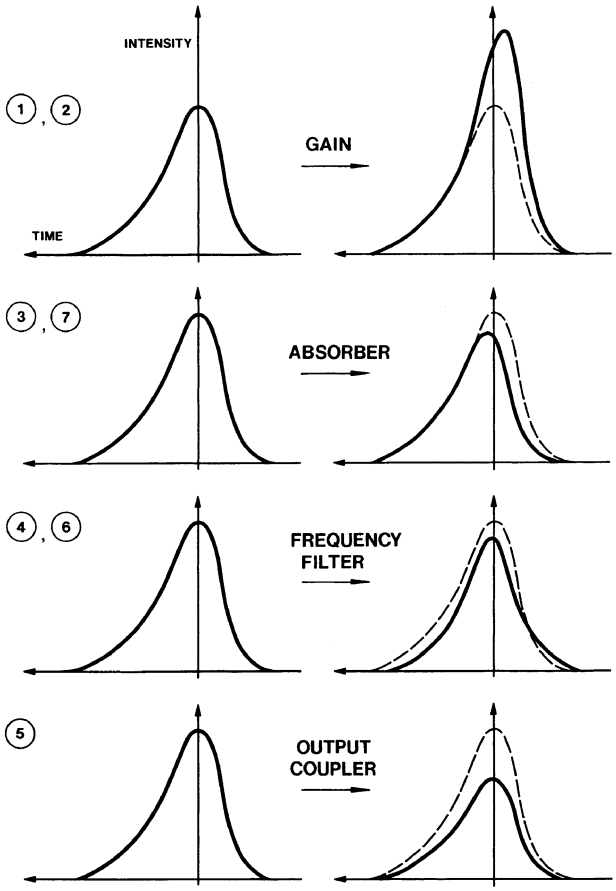


FIGURE 17 Schematic diagram of the shaping of the pulse that occurs in the transit of the cavity elements (indicated by the circled numbers) of the laser of the preceding figure. The peak of the pulse is advanced by the interaction with the gain jet and retarded by the absorber; the pulse width is increased in the filter, and left unchanged (though the amplitude is attenuated) at the output coupler.

the gain seen by the dye pulse does not depend critically on its arrival time back at the gain jet. By contrast, in synchronous mode-locking, the gain rises as the integral of the pump pulse over the ~100-psec pulse width, and the shaping of the dye pulse critically depends on its arrival time (relative to the pump pulse) at the gain jet. This is discussed in Section V,C, after the method of measuring pulse shape is explained. To conclude here note that the smaller linear losses in cavity-dumped operation during the pulse shaping process accounts for output pulses 1.5–2 times greater in pulse width observed in that mode.

To measure the shape of the dye pulse requires optical techniques, since the $\sim 10^{14}$-Hz bandwidths involved far exceed the capabilities of conventional circuitry. While a variety of techniques has been developed, the standard today is "background-free autocorrelation"—an average (over interference terms) of the overlap of the pulse train with a delayed sample of itself is recorded as a function of the path delay. The setup is a modified Michelson interferometer where the incoming pulse train is divided at a beamsplitter into two beams, which follow different paths down the two arms of the interferometer and emerge as parallel separate beams. The length of one arm can be varied in a precise manner to give a calibrated path delay. A lens focuses the parallel beams to cross at the same spot on a thin nonlinear crystal oriented to phase match for second harmonic generation only when one photon is taken from each of the two crossing beams. An aperture can be centered to pass only the doubled light as this beam emerges between the two crossing beams, and together with a UV transmitting filter over the photomultiplier detector gives a "background-free" signal. Mathematically this is proportional to the autocorrelation of the pulse shape with itself. By its nature, an autocorrelation signal must be symmetric about zero path delay, and to deconvolve this signal trace to give the width of the original pulse, a prior knowledge of the functional form of the pulse shape is required. Consequently, the autocorrelation width itself is often quoted, which is adequate for many experiments. Table II gives the relationship of the pulse width τ_D, to the width of the autocorrelation trace τ_C (both FWHM) for several functional forms. The squared hyperbolic secant form (sech^2) is the shape predicted for dye laser pulses in first-order, linearized mode-locking theories. The Fourier transform "pulse widthspectral width" product $\tau_D\,\Delta\nu$ used to check for incomplete mode-locking is also listed. Another indication of partial mode-locking is the appearance of a "coherence spike" at zero time delay on the autocorrelation trace, where the noisy substructure in a pulse will have its maximum self overlap. For dye pulses of a few picoseconds or longer, where some dispersion in the autocorrelator apparatus can be tolerated, a convenient instrument is the "rapid scanning" type using a rotating glass block to generate a repetitive path delay scan at 60 Hz, permitting display of the trace on an oscilloscope. For subpicosecond pulse lengths, a stepping

TABLE II Autocorrelation Widths and Spectral Bandwidths for Several Transform-Limited Pulse Shapes

Functional form of pulse shape	Ratio of pulse to autocorrelation width, τ_D/τ_C	Fourier transform product, $\tau_D\Delta\nu$
Square	1	0.886
Gaussian	0.707	0.441
sech^2	0.648	0.315
Lorentzian	0.500	0.221
One-sided exponential	0.500	0.110
Symmetric two-sided exponential	0.413	0.142

motordriven translation stage with micrometer resolution gives the calibrated path delay and a thin doubling crystal (100 μm KDP) gives adequate detection bandwidth. A more recent technique that has become widely used in measuring ultrashort pulses involves sampling in the time-frequency domains using a method known as frequency-resolved optical gating, or *FROG* for short.

C. Pulse-Shape Dependence on Cavity Length in Synchronous Mode-Locking

In synchronous pumping, mismatching the two cavity lengths varies the relative arrival times at the gain jet of the dye and pump pulses, which has a strong pulse-shaping effect. The nature of this effect was revealed in a clever experiment by Firgo, Daly, and Mahr, using what they called an "optical up-conversion light gate." They used rhodamine 6G dye with 514-nm argon laser pumping in a three-mirror cavity equivalent to Fig. 16 without an absorber jet or cavity dumper. Using a lens that collimated fluorescent light from the gain jet, they collected a sample of both the dye and pump beams reflected from the same point at the jet and sent this "signal" beam into one arm of an autocorrelator, with the dye output pulse train incident along the other, variable path-delay arm. After being collinearly combined at a beamsplitter, all of these beams were focused into an angle-matched ADP crystal, and the sum-frequency ("up-converted") UV light from any pair of one of the three signals and the dye beam was detected through a filter monochrometer. This cross correlation of the dye pulse with either the argon pump pulse or the fluorescence signal became in this way a waveform sampling apparatus with high time resolution. All three signals originated at the same point on the jet, and the spatial arrangement of the experiment preserved the relative signal timing as shown in Fig. 18. (This is a fine example of time-domain spectroscopy, the main application of the mode-locked dye laser.)

The chronology of the pulse-forming process demonstrated here is as follows. The gain (and fluorescence) is expected to rise as the time integral of the pump pulse, since the dye's spontaneous decay time far exceeds the time scale of the figure. Indeed, the fluorescence signal closely follows the integrated argon pulse (computed from the measured pump pulse shape) except in three places. First, there is a sharp drop in fluorescence coinciding with the time the dye pulse strikes the jet at 300 psec, due to the gain saturation (loss of upper-state population) in amplifying the dye pulse. The dye pulse here was traveling in a direction equivalent to that from M2 to M1 [passage (1)] in Fig. 16. Second, a fluorescence dip occurs 300 psec later, the propagation time from the jet to M1 and back, so this drop is due to the (reduced) amplification of the dye

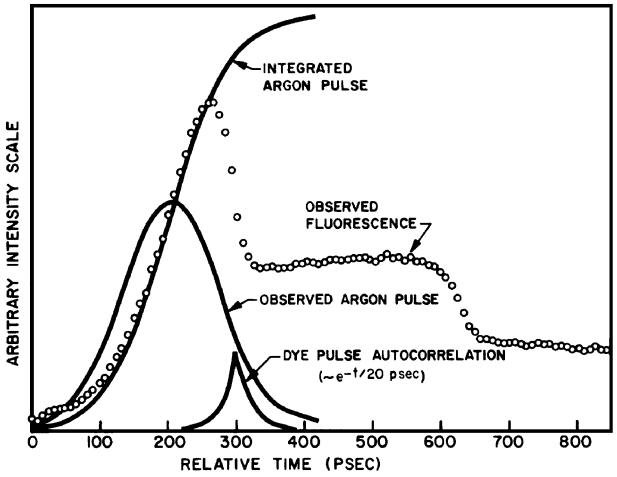


FIGURE 18 Results of the optical up-conversion light gate experiment of Mahr *et al.* that measured the arrival time of the dye pulse relative to the pump pulse in synchronous pumping and demonstrated that synchronism is maintained by pulse shaping over the mode-locking range of cavity length. [Reprinted with permission from Frigo, N. J., Daly, T., and Mahr, H. (1977). *IEEE J. Quantum Electronics* **QE-13,** 103. © 1977 IEEE.]

pulse on passage (2). Third, at zero time there is a residual fluorescence signal, due to incomplete spontaneous decay of the inversion produced 10 nsec earlier by the preceding pump pulse and left after the two passes of the preceding dye pulse.

The effect on these signals of mismatching the two cavity lengths was then studied. It was found that the optimum dye cavity length (that giving the shortest pulses) was slightly less than the length setting of Fig. 18. For this optimum length the dye pulse moved some 80 psec closer to the peak of the pump pulse, close enough that there was some gain recovery (rise in observed fluorescence) after the first passage of the dye pulse. This gain rise followed the shape of the integrated argon pulse. The cavity length change that caused this 80-psec time shift was equivalent to only a 0.25-psec cavity transit time change. The circulating dye pulse satisfies two conditions: that its round-trip transit time is precisely equal to the period of pumping pulses (after accumulated advance and retardation adjustments due to pulse shaping), and the "steady-state" condition, that the shape of the pulse is reproduced at each location in the cavity in each transit. To accommodate the shorter cavity length, the pulse position moved down the integrated fluorescence curve to a position of less gain. The pulse shape evolved into a shorter shape having 0.25 psec less net advance per round trip in the cavity. For a further 0.25-psec shortening of the transit time, the dye pulse position moved another 30 psec down the gain curve, and at this position the gain recovery was large enough after the first passage that a second dye pulse began to form.

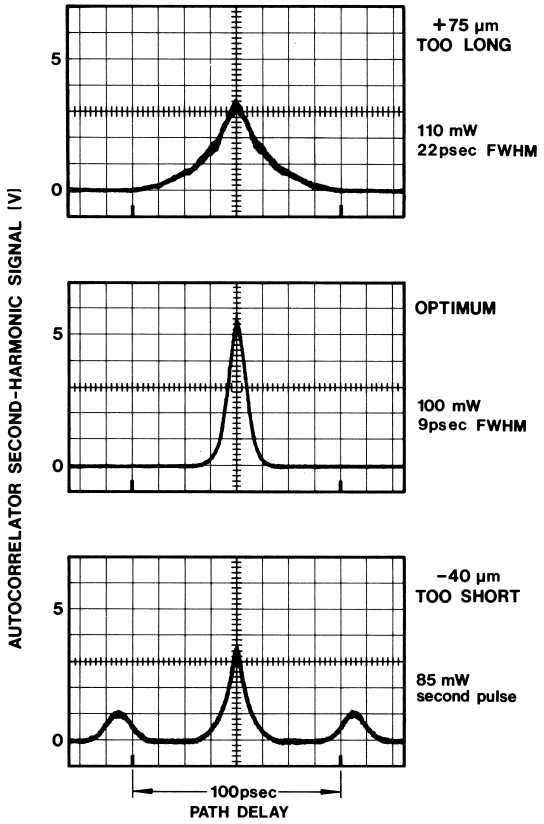


FIGURE 19 Autocorrelation traces taken during setup of the cavity-dumped laser of Fig. 16, showing changes in the pulses with cavity length adjustment. A three-plate birefringent filter and a 1-W 532-nm pump beam were used without an absorber jet for this data. (Courtesy of Coherent, Inc., Palo Alto, CA.)

Examples of this typical pulse-shape behavior for three cavity length settings are shown in Fig. 19, taken with the laser of Fig. 16 in 3.8-MHz pulse repetition rate (cavity-dumped) operation, with rhodamine 6G dye, a three-plate birefringent filter, and 1 W of doubled YAG pumping. The vertical axis is the signal from a rapid scanning autocorrelator, and the horizontal is the calibrated path delay. A display like this is used in setting up the laser to locate the optimum cavity length setting. In this laser, the absorber jet would then be turned on, the three-plate filter exchanger for a one-plate, and the new optimum cavity length setting would be found, yielding shorter pulses of 1-psec autocorrelation trace width, 18-nJ energy per pulse, and 18-kW peak dye power.

D. The Shortest Pulses—Colliding-Pulse Ring Lasers and Pulse Compression

This picture of pulse formation in mode-locked dye lasers was sufficient for the generation of subpicosecond pulses and stood as the state of the art until the early 1980s, when two ideas were introduced, one that was new and one that then became widely appreciated, to permit the generation of even shorter pulses. The new idea was that of colliding-pulse mode-locking, due to Fork, Greene, and Shank. Here cw pumping of a ring laser was used, with a thin saturable absober jet position at one-quarter the perimeter of the ring away from the gain jet. Two pulses circulate in this cavity, one clock-wise and one counterclockwise, and collide in the absorber jet to produce interference fringes and a high degree of saturation there. The spacing of the two jets in the cavity ensures that each beam sees the gain jet at half a transit time after the last passage of a dye pulse, to give equalintensity beams for maximum interference in the absorber. This laser produced 90-fsec pulse widths—a pulse only 0.03 mm long.

The second idea that awoke the scientists' imagination was the demonstration by Nikolaus and Grischkowsky of the compression by a factor of 12 of a 5-psec pulse to 450 fsec with a fiber-grating pulse compressor (Fig. 20). These experimentors realized that the spatially uniform (non-Gaussian) intensity distribution that results when a light beam is coupled into a single-mode optical fiber would result in spatial uniformity of the nonlinear optical effect known as self phase modulation (SPM). The refractive index of the fiber material is time-modulated by the large electric field amplitude of the cavity-dumped dye pulse coupled into the fiber. This phase modulation increases the bandwidth of the pulse. (In fact, SPM has been used to generate a femtosecond white-light continuum by the focusing of ultrashort dye pulses into dielectric media.) In traveling through the fiber, the ordinary group velocity dispersion for this broadened spectral content linearly chirps the pulse—the redder wavelengths emerge first, followed linearly in time (in a longer pulse) across the pulse spectrum by the bluer wavelengths. A dispersive

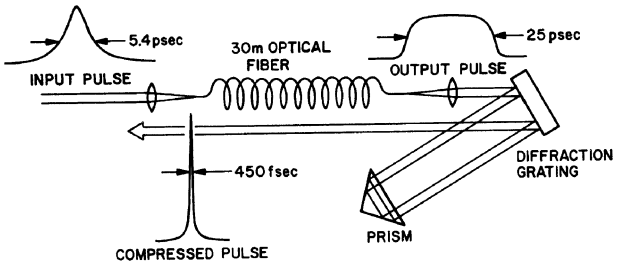


FIGURE 20 The pulse compression experiment of Grischkowsky *et al.*, in which self-phase modulation and group velocity dispersion of a pulse in an optical fiber are balanced to produce a linear frequency chirp in the output pulse. The two passes off the diffraction grating constitute a dispersive delay time, which compresses this pulse to one-twelfth the width of the input pulse. [Reprinted with permission from Nikolaus, B., and Grischkowsky, D. (1983). *Appl. Phys. Lett.* **42**, 1.]

delay line (like the two passes off the grating of Fig. 20) provides a path length proportional to the wavelength and compensates this chirp to produce a compressed pulse.

These results showed that femtosecond-width pulses were possible in mode-locked dye lasers, revitalized the search for the processes limiting the pulse lengths, and led to the present evolutionary burst in this field. Dispersion-compensating delay lines were added intracavity in the dye laser. This looks like the quad-prism expander of Fig. 6 when placed in-line in an arm of a ring laser (only with "full" Brewster prisms instead of the "half" prisms in that figure), or equivalently is a pair of prisms placed in front of an end mirror of a linear laser. One prism is laterally displaced to adjust the relative spectral delay to compensate linear chirp from other intracavity elements. To date, pulses have been generated of 55-fsec width from a linear hybrid mode-locked laser, 27-fsec width from a colliding-pulse ring laser, and the shortest of all (as of 1985), a train of pulses of 8-fsec width was produced by compression from a 40-fsec source.

The molecular relaxation times measured with these ultrashort pulses in time-domain spectroscopy are as diagnostic of the system studied to a chemist or biologist as are atomic energy levels to a physicist. By modeling the chemical interaction and matching the calculated decay times to the observed ones, the actual molecular structure is revealed, and these new laser tools should keep these fields of study active for many years to come.

VI. SPECIALIZED DYE LASERS

The majority (in numbers) of dye lasers in existence is represented by the three scientific types already discussed, but these by no means represent the diversity of ways that lasers with a dye gain medium have been built. In this concluding section, four additional dye lasers are briefly discussed as a better indication of the breadth of this laser type and of its potential impact on future technology.

Vapor-phase dye lasers were initially investigated in the hope that the efficiency of the dye laser (plus pumping system) could be improved with excitation by a discharge directly in the dye vapor. This proved not to be the case. While molecular densities comparable to solution dye lasers were reached in the vapor state (by heating to $\sim 400^{\circ}$C nonionic dyes sealed in silica glass cells), and spectra comparable to dye spectra in solution were observed (with slight broadening and a shift to shorter wavelengths), the inevitable result of running a discharge in the vapor was the rapid decomposition of the dye molecules. Several vapor-phase dyes were made to lase as short-pulse oscillators by optical pumping with a pulsed nitrogen laser, with results comparable to the same dyes in solution. The initial goal, however, is still well worth emphasizing. The applications of most lasers are limited by their high consumption of input power (their inefficiency), and the dye laser is one of relatively few lasers with an intrinsically efficient inversion mechanism (described in Section II). Thus, means to improve efficiency by eliminating another laser as the pump source, as in these vapor-phase experiments, are well worth seeking.

The distributed feedback dye laser first demonstrated the tunable, narrow-band output from a very compact structure that would be desired in a light source for integrated optical devices. The optical feedback in these mirrorless lasers is provided by Bragg scattering from a periodic spatial variation of the refractive index or the gain itself in a thin-film gain medium. The first such laser used a holographic phase grating, which was exposed and developed in a dichromated gelatin film on a glass substrate, before the film was dyed by soaking in a solution of rhodamine 6G. This was transversely pumped with a nitrogen laser to produce a 630-nm beam (of 0.05-nm linewidth) from the $0.1 \times 0.1 \times 10$-mm total laser volume. To show a distributed feedback dye laser that was externally tunable, a dye cell was side-pumped with intersecting, interfering, 347-nm beams (from a beam-split, doubled ruby laser) to produce fringes and a periodic gain variation in the sheet of dye. The rhodamine 6G dye output wavelength tuned 70 nm for a 48–56° change in incidence angle for each of the two pump beams. In the tiny volumes of these lasers, if the dye molecules are immobilized in gelatin or plastic, the small number of molecules are recycled so frequently that photodecomposition limits the dye lifetime to a few seconds or less. The actual distributed feedback lasers that are candidate sources for integrated optical circuits, the inheritors of these dye laser results, will be analogs fabricated in semiconductor materials.

On the other end of the size and output power scale are the specialized dye laser oscillator and amplifier devices built for the atomic vapor laser isotope separation (AVLIS) process planned as the prime means in the United States for the enrichment of uranium-235 for nuclear reactor fuel rods near the turn of this century. In this process, the uranium raw stock is vaporized in a vacuum with an electron beam, and the streaming atomic vapor is exposed to dye laser radiations tuned to selectively photoionize (in three absorption steps) only ^{235}U atoms. These charged ions are deflected out of the stream by electrostatic fields to achieve separation. (Unlike the ^{20}Ne to ^{22}Ne isotope shift of Fig. 15, the ^{235}U to ^{238}U isotope shift for spectral lines of interest is greater than a Doppler width.) The general scale and specifications of the three laser beams (at three visible wavelengths) required in the separator are about the same as the design goal reported for the selective step, namely,

$\sim$1 J/pulse, $\sim 10^4$-Hz pulse repetition rate (or $\sim$10-kW average power), in about a 1-GHz bandwidth locked onto the center of the uranium line. This high average power will be reached with massive banks of copper-vapor lasers (a short-pulse laser like the nitrogen laser but with emissions in the green and yellow) as pumps for short-pulse dye oscillator and amplifier chains. The copper-vapor lasers have a few percent electrical-to-optical efficiency, which is important in the economics of the process, and the yellow-green pump lines are efficient both in pumping red dyes and in giving long dye photodegradation times. Thus, dye lasers in about 2 decades may be responsible for a sizable fraction of the world's electric power generation.

Continuous-wave dye lasers have also been the basis in experimental investigations for the cure of cancer (Fig. 21). In this work, termed photodynamic therapy, the patient with a tumor is first injected with a drug (hermatoporphyrin derivative, HpD) that has the property of being selectively retained in rapidly dividing tissue (the tumor) and of photodegrading to release free oxygen when exposed to a measured dose of light of the proper color. After a waiting period of 2–3 days for the drug to clear from healthy tissue, the patient is given the light dose through fiber optic guides slipped into needles implanted into the tumor. The released oxygen burns up the tumor but has no effect (and leaves a sharp line of demarcation) on the healthy tissue. In principle, the drug can be activated by ordinary light sources, filtered to give the proper wavelengths (corresponding to an appropriate absorption peak). In practice, the only activating wavelength with enough penetration depth for coverage of a few cubic centimeters volume per implant in most bodily tissue is the one in the red, which requires a 630 ± 5 nm wavelength. A laser source is dictated by the convenience of delivery through fibers (a coherent source focuses efficiently into the fiber's input end) and by the large total dosage of $\sim$40 J cm^{-3} needed in this narrow bandwidth. The cw dye laser is the laser of choice for this application. Experimental medicine using the tunability of the dye laser to selectively absorb in different tissues without drugs is also being done in photocoagulation inside the eye and in other laser surgical procedures. These examples show that the future importance of the dye laser can be expected to extend much beyond today's useage as a research tool.

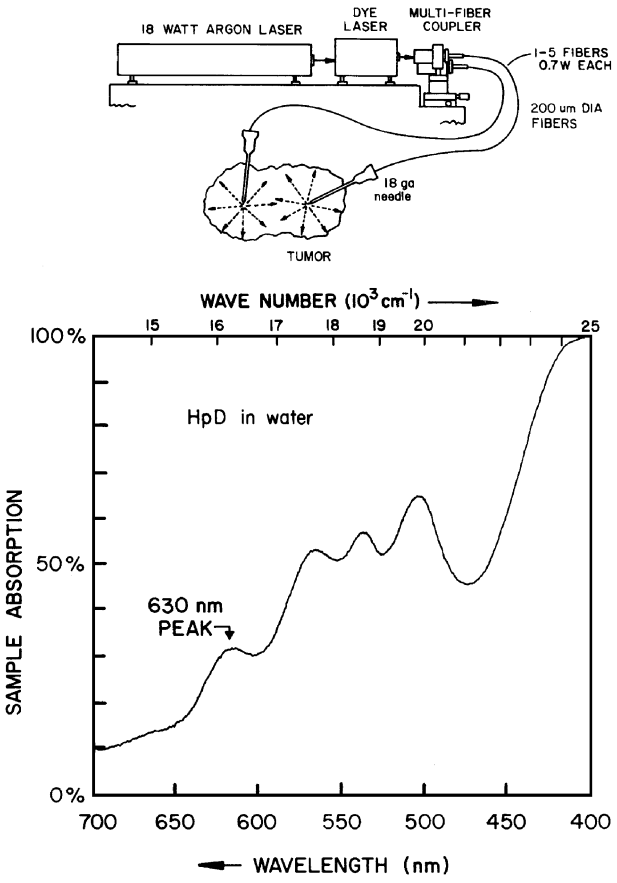


FIGURE 21 Schematic diagram of the destruction of cancerous tumors by the activation of the experimental drug HpD (hermatoporphyrin derivative). A dose of dye laser light, tuned to a specific absorption band of the drug, is administered through fiber optic guides to photodecompose the drug and kill the tumor. The absorption spectrum shown is for an early form of the drug in water; the peaks shift somewhat in tissue and with the newer forms of the drug. (Courtesy of Coherent, Inc., Palo Alto, CA.)

VII. RECENT DEVELOPMENTS

Two new crystals have recentaly become available in high optical quality form to the benefit of the tunable lasers discussed in this article. The first, beta-barium borate (BBO), is a doubling crystal with transparency to slightly below 200 nm that has extended tunable outputs down into this wavelength region. The second, titanium:sapphire (Ti:Al_2O_3), is an optically pumped crystal which has been substituted for the dye medium in single-frequency and mode-locked laser cavities to produce tunable outputs at higher power in the near infrared than previously available. Increased powers from the pump lasers in all three types of scientific dye lasers have boosted tunable outputs as well, particularly in the case of the UV pump lines from argon ion lasers. The pump powers of the standard UV-line mix now are twice as great, and shorter UV pump lines have become available, making possible cw tunable lasing in the UV directly from new dyes. In ultrafast lasers the shortes-pulse record (done by extracavity pulse compression) is now 6 fsec; tains of pulses of 29-fsec duration have been generated directly in the hybrid mode-locking geometry by the addition of prism pairs to the laser cavity

to compensate the intracavity group-velocity-dispersion; and commercial systems now attain and maintain sub-100-fsec pulse lengths by incorporating servo systems in their design.

A. Single-Frequency and Short-Pulse Lasers

The resulting changes in wavelength coverage for the single-frequency lasers have been illustrated in the tuning spectrum of Fig. 13. Beginning on the long wavelength end of the spectrum, the solid-state crystal titanium-doped sapphire has produced the highest reported single-frequency output over the 690- to 840-nm fundamental tuning range, with a peak of 3.6 W at 790 nm (replacing the 2.0-W peak at 740 nm of the dye LD700). This is a crystalline optically pumped gain medium and not a dye, but due to the homogeneous line broadening at the active Ti^{3+} ion sites, titanium:sapphire shows spectral condensation and tunability like a dye and is used as an alternate gain medium in the same laser as shown in Fig. 10 (now termed a Ti:dye laser). The 2-cm-long crystal replaces the astigmatic compensation rhomb in the cavity (just as this position was used earlier for the temperature-tuned ADA doubling crystal, see Section IV.E), and the dye jet is turned off. The argon ion laser pump beam is folded to enter the cavity through the mirror M2, after striking a 30-cm radius fold mirror placed behind M2, to focus into the crystal in the colinear pumping geometry needed with a long gain medium. The same single-frequency filter stack and frequency servo system used with a dye produce with the crystal <500 kHz rms frequency jitter as described in the text. Titanium:sapphire is pumped by the blue-green band of lines from the argon ion laser. By clamping the crystal in a water-cooled block it withstands a full 20-W pump power, which accounts for the high output power (the LD700 dye used a red pump beam from a krypton ion laser, for which only 5-W pump power is available). The tunability of this crystal extends to 1000 nm and it would be expected with additional work to replace as well the next two longer wavelength dyes.

Improvements in argon ion lasers have raised the UV-band pump power to 7 W, and added a short-ultraviolet wavelength (300–336 nm) pump band (SUV in Table I) of about 3 W power. The first improvement has increased by a factor of three the blue wavelength outputs from the stilbene 1 and stilbene 3 dye lasers to 0.42 and 1.0 W, and the second has permitted single-frequency ultraviolet lasing from the dye polyphenyl 2 over the range 364–408 nm with a 0.25-W peak output at 383 nm. Other UV-lasing dyes are under development. An output comparable to the directly UV-lasing dyes was produced by doubling titanium:sapphire radiation intracavity in an angle-matched lithium iodate crystal as described in Section IV.E. This high output (0.20 W at 370 nm with 345- to 388-nm tunability) was due to the low crystal absorption of $<1\%$ at the fundamental wavelengths and to the ability of this crystal to sustain high pump powers.

Advantage could be taken of the improved output powers in the blue dyes as the angle-matched crystal beta-barium borate for doubling this region became available at the same time. When used in the cavity (Fig. 10) of the stilbene 3 and coumarin 102 dye lasers, this extended the single-frequency tuning spectrum over most of the range from 215 to 260 nm. The small 235- to 238-nm gap of Fig. 13 is expected to be closed by doubling in stilbene 3 using the technique explained in Section IV.E for pushing the tuning range past a tuning curve break point. In R6G dye BBO doubling has produced five times more second harmonic power than the earlier angle-matched result with KDP, and it is anticipated that the tuning curves remaining in the figure between the doubled R6G and C102 results will be superceded by BBO doubling results when this crystal is tried in the intervening dyes. To be useful for cw intracavity doubling a crystal must have an absorption of $<1\%$ over the length of the crystal (one aperture length of 7 mm here); the excellent results of Fig. 13 demonstrate that the crystals available now are of this high quality. This crystal also has a high damage threshold, permitting it to be used in doubling the blue wavelength outputs from short-pulse dye lasers, which has been done to produce ten times the previously available deep UV pulse energies.

B. Ultrashort-Pulse Lasers

An important recent result in the area of ultrafast lasers was the demonstration of pulses of 29 fsec duration directly genrated in a hybrid mode-locked linear dye cavity (like Fig. 16 of the text), to which a set of four prisms were added and arranged to compensate group-velocity-dispersion as discussed in Section V.D. This nearly matches the shortest pulse (27 fsec) so far generated directly in a colliding pulse mode-locked cavity and favors the use of hybrid mode-locking which has produced considerably higher output power (350 versus 50 mW) and considerably greater tunability. But since the pulse-width, shape, and noise of the pulse train from a synchronously mode-locked cavity are strong functions of the cavity length as discussed in Section V.C, for reproducible sub-100-fsec pulses the cavity length must be held stable to within about 100 nm. Holding this tolerance requires a cavity length servo. From among several possibilities, commercial systems use a length-discriminant based on the output power of the dye laser, which rises linearly through the cavity length region that gives the shortest pulses. With an autocorrelator to identify the desired pulse width and shape, the optimum

cavity length and corresponding output power is found and a proportional reference voltage subtracted from the detected power to generate the discriminant. The advantages of this error signal are that it is reproducible over the whole tuning range and in different dyes, and its use simultaneously stabilizes the output power. However, it presumes that the only source of power fluctuation is the dye cavity length. In order to make this assumption true, a fast amplitude noise reduction servo based on acoustooptic diffraction (called a "noise eater") is used on the pump beam, and the pointing fluctuations of the pump beam are stabilized by sensing its positional error at the dye laser and feeding back adjustments to steer the pump laser output mirror. Routine optimization of these interacting servo systems is done by microprocessor control. In generating sub-100 fsec pulses a preference has come about for the doubled Nd:YLF (neodymium–yttrium–fluoride) laser as a pump source instead of the doubled Nd:YAG laser, due primarily to the lower noise characteristics of the former host crystal.

The new shortest pulse record of 6-fsec (compared to the former 8-fsec result) was produced from a train of 50-fsec pulsewidth amplified pulses, linearly chirped in a 0.9-cm length of single-mode quartz fiber, then compressed in a dispersion compensating delay line similar to that discussed in Section V.D except that the two pairs of prisms were preceded by two pairs of gratings in a similar folded geometry. The extra degree of freedom given by adjustment of both gratings and prisms permitted compensation of both the square and cubic terms in the Taylor expansion of the optical phase versus optical frequency offset from the center of the pulse spectrum, with the shorter compressed pulse being the result. The ultrashort pulse (ultrafast) laser area continues to be one of rapid evolution with the results discussed in this section merely suggestive of the types of many such changes occurring in this field.

C. Solid-State Dye Lasers

The first solid-state dye lasers were reported in the scientific literature by by Soffer and Macfarland in 1967 and by Peterson and Snavely, in 1968. However, problems with material homogeneity kept these lasers in the archives for more than 25 years. In the early to mid 1990s new forms of improved solid-state gain laser media brought renewed attention and intense activity to the field. The research has focused in the development and demonstration of new polymer matrices with a high degree of optical homogeneity. One type of widely used solid-state dye laser gain media is dye-doped modified polymethyl methacrylate (MPMMA). The word modified here means PMMA of a high-purity form yielding an extremely homogeneous optical medium. An alternative to dye-doped MPMMA is dye-doped 2-hydroxyethyl methacrylate: methyl methacrylate (HEMA:MMA). Also, researchers have used a solid-state gain medium where the dye rather than being diffused into the matrix, it is copolymerized with HEMA:MMA.

Taking advantage of these new and improved materials, Duarte, in the mid to late 1990s, demonstrated a new class of very-compact narrow-linewidth multiple-prism grating tunable laser oscillators (Fig. 22). These tunable laser oscillators are capable of extremely pure high-power laser emission across the visible spectrum. The degree of spectral purity is characterized by single-longitudinal-mode emission at laser linewidths in the $350\ \text{MHz} \leq \Delta\nu \leq 375\ \text{MHz}$ range. The temporal pulse emission from these lasers has been shown to be near-Gaussian with durations in the 3- to 7-nsec regime. Other

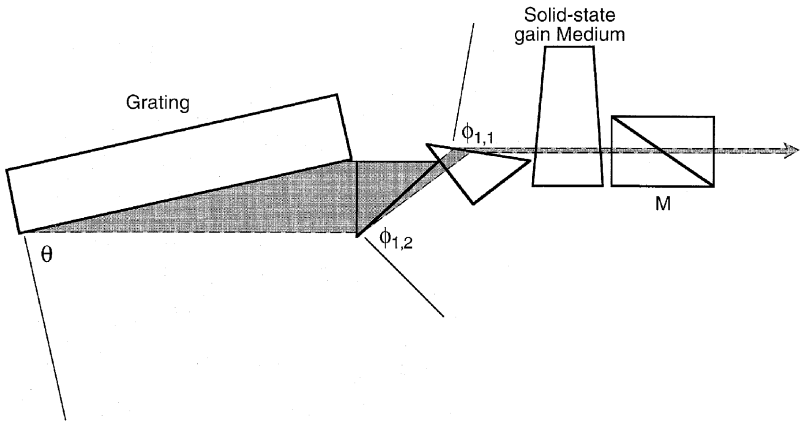


FIGURE 22 Multiple-prism Littrow (MPL) grating solid-state dye laser oscillator. The cavity length is about 75 mm and the excitation is performed semilongitudinally at a $\sim 3^\circ$ angle relative to the optical axis [From Duarte, F. J. (1999). *Appl. Opt.* **38,** 6347–6349.]

developments in this field include the successful demonstration of solid-state distributed feedback dye lasers. Research in the area continues to develop new laser gain matrices with improved thermal characteristics.

SEE ALSO THE FOLLOWING ARTICLES

LASERS • LASERS, COLOR CENTER • LASERS, FREE-ELECTRON • LASERS, ULTRAFAST PULSE TECHNOLOGY • NONLINEAR OPTICAL PROCESSES • POTENTIAL ENERGY SURFACES • QUANTUM CHEMISTRY

BIBLIOGRAPHY

Bradley, D. J., and New, G. H. C. (1974). "Ultrashort pulse measurements," *Proc. IEEE* **62,** 313–345.

Couillaud, B., and Fossati-Bellani, V. (1985). "Mode-locked lasers and ultrashort pulses," *Lasers and Applications* Jan., 79–83, Feb., 91–94.

Diels, J. C., and Rudolph, W. (1996). "Ultrashort Laser Pulse Phenomena," Academic Press, New York.

Drever, R. W. P., Hall, J. L., Kowalski, F. V., Hough, J., Ford, G. M., Munley, A. J., and Ward, H. (1983). "Laser phase and frequency stabilization using an optical resonator," *Appl. Phys. B* **31,** 97–105.

Duarte, F. J. (1995). "Solid-state dispersive dye laser oscillator: very compact cavity," *Opt. Commun.* **117,** 480-484.

Duarte, F. J., and Piper, J. A. (1982). "Dispersion theory of multiple-prism beam expander for pulsed dye lasers," *Opt. Commun.* **43,** 303–307.

Duarte, F. J., and Hillman, L. W. (1990). "Dye Laser Principles with Applications," Academic Press, New York.

Fork, R. L., Shank, C. V., Yen, R., and Hirlimann, C. A. (1983). "Femtosecond Optical Pulses," *IEEE J. Quantum Electron.* **QE-19,** 500–505.

Hänsch, T. W. (1972). "Repetitively pulsed tunable dye laser for high resolution spectroscopy," *Appl. Opt.* **11,** 895–898.

Ippen, E. P., and Shank, C. V. (1978). "Sub-picosecond spectroscopy," *Phys. Today* May, 41–47.

Johnston, T. F., Jr., Brady, R. H., and Proffitt, W. (1982). "Powerful single-frequency ring dye laser spanning the visible spectrum," *Appl. Opt.* **21,** 2307–2316.

Mason, S. F. (1970). Color and the electronic states of organic molecules. *In* "The Chemistry of Synthetic Dyes," Vol. 111 (K. Venkataraman, ed.), pp. 169–221, Academic Press, New York.

Peterson, O. G. (1979). Dye lasers. *In* "Methods of Experimental Physics," Vol. 15A (L. Marton, ed.), pp. 251–359, Academic Press, New York.

Ryan, J. P., Goldberg, L. S., and Bradley, D. J. (1978). "Comparison of synchronous pumping and passive mode-locking of CW dye lasers for the generation of picosecond and subpicosecond pulses," *Optics Commun.* **27,** 127–132.

Scavennec, A. (1976). "Mismatch effects in synchronous pumping of the continuously operated mode-locked dye laser," *Optics Commun.* **17,** 14–17.

Schäfer, F. P. (ed.) (1990). "Dye Lasers," Vol. 1 of "Topics in Applied Physics," 3rd ed., Springer-Verlag, Berlin.

Shank, C. V. (1975). "Physics of dye lasers," *Rev. Modern Phys.* **47,** 649–657.

Sorokin, P. (1969). "Organic lasers," *Sci. Am.* Feb., 30–40.

Valdmanis, S. A., and Fork, R. L. (1986). "Design considerations for a femtosecond pulse laser balancing self phase modulation, group velocity dispersion, saturable absorption, and saturable gain," *IEEE J. Quantum Electron.* **QE-22,** 112–118.

Lasers, Excimer

Minoru Obara
Keio University

Masayuki Kakehata
Electrotechnical Laboratory

I. Brief Historical Overview
II. Basic Principles
III. Excimer Laser Technology
IV. Applications and Trends

GLOSSARY

Excimer Molecule that is strongly bound in an excited state but normally has a dissociative ground state. The excited state has a very short lifetime of less than 10 nsec. The term excimer comes from "excited dimer." Excimer also includes exciplex (excited complex).

Pumping Means by which an inversion is produced in a laser system.

Tunable Describes a laser having an output wavelength that can be adjusted within a gain bandwidth.

Vacuum ultraviolet (VUV) Region of the spectrum with wavelength less than 200 nm.

STRONG LASER EMISSION has been obtained from Xe_2, Kr_2, Ar_2, F_2, ArF, KrF, XeF, KrCl, XeCl, and XeBr excimers. Strong output wavelength of rare gashalide lasers spans the spectrum from the near ultraviolet to the vacuum ultraviolet. Excimer lasers have been operated only in a pulsed mode. Their laser pulse length extends from sub-picoseconds to a microsecond time domain. Excimer lasers may be pumped in a pulsed mode by electric discharges, intense electron beams, proton beams, or optical sources. These lasers are wavelength tunable over wavelength ranges of several nanometers.

I. BRIEF HISTORICAL OVERVIEW

A. History

The first demonstration of an excimer laser was made by Basov, Danilychev, and Popov in 1971. They showed stimulated emission of Xe_2 at 172 nm using electron-beam pumping of cryogenic liquid Xe.

The first lasing of a rare gas-halide (RGH) excimer (XeBr) was reported by Searles and Hart in 1975. Shortly thereafter, lasing from XeF was reported by Brau and Ewing. Both XeBr and XeF lasers were pumped by intense electron beams. Excimers shown in Table I were reported to lase. In addition to electron-beam pumping, researchers have also employed volume-uniform avalanche discharges with X-ray, UV, or corona preionizations, electron-beam controlled discharges, and proton beams successfully to pump a variety of excimer lasers.

Table I lists excimer lasers. The strong output wavelength of excimer lasers spans the spectrum from the near UV to the vacuum ultraviolet (VUV). Under normal

TABLE I A List of Excimer Lasers

Excimer	Wavelength (nm)	Author	Reported year
Xe_2	172[a]	Basov, Danilychev, and Popov	1971
Kr_2	146	Hoff, Swingle, and Rhodes	1973
Ar_2	126	Hughes, Shannon, and Hunter	1974
F_2	157	Rice, Hays, and Woodworth	1977
ArCl	175	Waynant	1977
ArF	193	Hoffman, Hays, and Tisone	1976
KrF	248(275)[b]	Ewing and Brau	1975
KrCl	222(240)	Murray and Powel	1976
XeBr	282(300)	Searles and Hart	1975
XeCl	308(345)	Ewing and Brau	1975
XeF	351,353(460)	Brau and Ewing	1975

[a] Wavelength due to $B \rightarrow X$ transition having the largest gain coefficient.

[b] Wavelength due to $C \rightarrow A$ transition in parenthesis.

operating conditions, excimer lasing occurs on the wavelength based on the $B \rightarrow X$ transition described in Section I.B.

In addition to the diatomic RGHs, triatomic RGH excimers can provide tunable coherent photon sources in the visible to the UV region of the spectrum. Table II shows a list of triatomic RGH excimers.

In addition to the triatomic RGHs, a new broad band emission from a fouratomic RGH excimer, Ar_3F at (430 ± 50) nm was reported by Sauerbrey *et al.* in 1986. A fouratomic Ar_3F RGH excimer emission (not lasing) was detected from electron-beam pumped Ar/F_2 and Ar/NF_3 mixtures.

B. Spectroscopic Features of Excimer Lasers

Figure 1 shows the schematic potential curves for the RGH excimers. The upper laser level is an ionically bound state while the ground state is covalently bound. The upper laser level is formed via ionic or neutral reactions. At close internuclear separation, the potential energy curve splits

TABLE II A List of Triatomic RGH Excimers[a]

RGH	Wavelength (nm)
Ar_2F	285 ± 25
Ar_2Cl	245 ± 15
Kr_2F	420 ± 35
Kr_2Cl	325 ± 15
Kr_2Br	~ 318
Xe_2F	610 ± 65
Xe_2Cl	490 ± 40
Xe_2Br	440 ± 30

[a] Wavelength and tuning range are based on the data by F. K. Tittel *et al.*

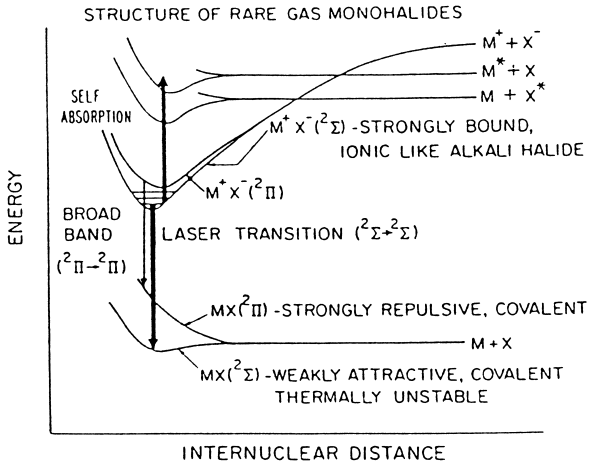


FIGURE 1 Schematic potential energy diagram of rare gas halides. [Reproduced with permission from Ch. A. Brau (1984). Rare gas halogen excimers, *in* "Excimer Lasers" (Ch. K. Rhodes, ed.), Springer-Verlag, Berlin and New York.]

into the $^2\Sigma$ and $^2\Pi$ states as shown in Fig. 1. By convention the $^2\Sigma$ state is referred to as the B state. Under normal high-pressure operating conditions both the B ($^2\Sigma$) and C ($^2\Pi$) state are collisionally mixed.

The ground-state manifold consists of two states, of which the $^2\Sigma$ state has the lowest energy, and is referred to as the X state. This X state is generally nearly flat or weakly bound with the exception of an XeF excimer having a strong bound state with a 1065-cm^{-1} binding energy. The other manifold is the $^2\Pi$ state, which is always repulsive as shown in Fig. 1. This $^2\Pi$ state is referred to as the A state.

Figures 2 and 3 show the potential curves of KrF and XeCl excimers, respectively.

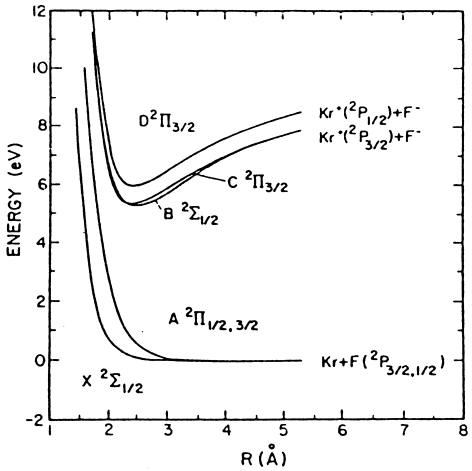


FIGURE 2 Potential energy diagram of the KrF excimer. [Reproduced with permission from Ch. A. Brau (1984). Rare gas halogen excimers, *in* "Excimer Lasers" (Ch. K. Rhodes, ed.), Springer-Verlag, Berlin and New York.]

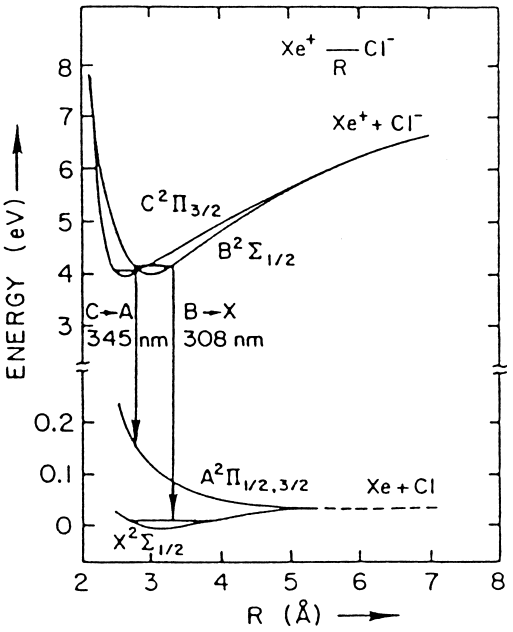


FIGURE 3 Potential energy diagram of the XeCl excimer. [Reproduced with permission from D. L. Huestis, G. Marowsky, and F. K. Tittel (1984). Triatomic rare-gas-halide excimers, *in* "Excimer Lasers" (Ch. K. Rhodes, ed.), Springer-Verlag, Berlin and New York.]

The emission spectrum of RGH excimers consists of two bands such as B $(^2\Sigma) \rightarrow X(^2\Sigma)$ and $C(^2\Pi) \rightarrow A(^2\Pi)$. The $B \rightarrow X$ transition has a larger stimulated-emission cross section than that of the $C \rightarrow A$ transition, indicating that the $B \rightarrow X$ transition usually gives intense lasing. The $C \rightarrow A$ band consists of relatively broad continua, which is attributed to the repulsive structure of the A state.

The stimulated-emission cross section of the $B \rightarrow X$ bands of RGHs like XeCl and KrF may be written as

$$\sigma = \frac{1}{4\pi c\tau}\sqrt{\frac{ln2}{\pi}}\frac{\lambda^4}{\Delta\lambda}$$

assuming that near the line center the line shape is approximately Gaussian, where c is the speed of light, λ the wavelength at the line center, τ the lifetime of excimers, and $\Delta\lambda$ the bandwidth of the spectrum.

II. BASIC PRINCIPLES

The kinetic processes involved in a RGH laser are very complicated compared to those of discharge-pumped CO_2 lasers, because neutral reactions and ionic reactions, two-body and three-body reactions, superelastic reactions, and absorption reactions are all responsible for the RGH laser. The kinetic processes involved in the individual RGH lasers are to some extent similar with the exception of those of the XeF laser. Therefore, the kinetic processes for only the discharge-pumped XeCl laser and the e-beam-pumped KrF laser are presented here as an example.

A. Discharge-Pumped XeCl Lasers

A typical gas mixture for a self-sustained discharge-pumped XeCl laser is 3-atm mixture of Xe/HCl/Ne = 1/0.1 – 0.2/balance(%). The electron energy distribution in the discharge mixture can be calculated using a Boltzmann equation code, an example of which is shown in Fig. 4. The case is treated of the above mixture pumped by a 100-nsec discharge pulse at an excitation rate of 3 MW/cm^3.

Formation reactions for XeCl(B) are shown in Fig. 5. The percent contribution of the XeCl formation is varied with the HCl concentration. Dominant formation reactions are $Xe^+ + Cl^- \rightarrow XeCl(B)$ (ion recombination eaction) and $NeXe^+ + Cl^- \rightarrow XeCl(B) + Ne$. A little contribution comes from $Xe^* + HCl(v) \rightarrow XeCl(B) + H$ and $Xe_2^+ + Cl^- \rightarrow XeCl(B) + Xe$. Over 23% of the electrical energy deposited into the discharge can be supplied to form XeCl(B).

Relaxation reactions for XeCl(B) are shown in Fig. 6. About 65% of the formed XeCl(B) contributes to the stimulated emission as an intracavity laser flux due to its large stimulated-emission cross section. About 30% of the XeCl(B) is collisionally quenched and spontaneous emission is negligible during lasing. Collisional quenching processes concerned with the XeCl(B) excimer are shown in Fig. 7. At this high excitation rate of 3 MW/cm^3, a large fraction is occupied by the discharge electron, which is called a superelastic collision process.

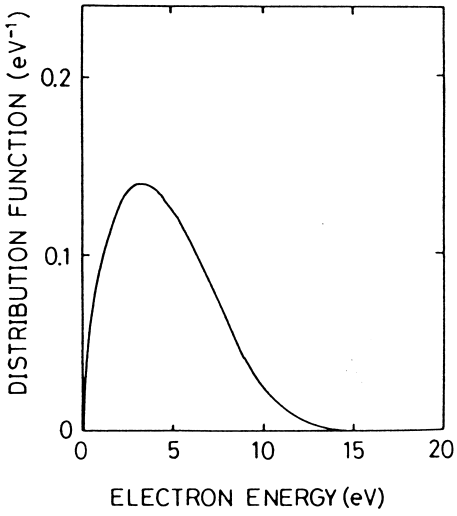


FIGURE 4 Electron energy distribution in the 3-atm mixture of Xe/HCl/Ne = 1/0.1/98.9(%) pumped at 3 MW/cm^3. $E/N = 2.9 \times 10^{-17}$ V cm^2.

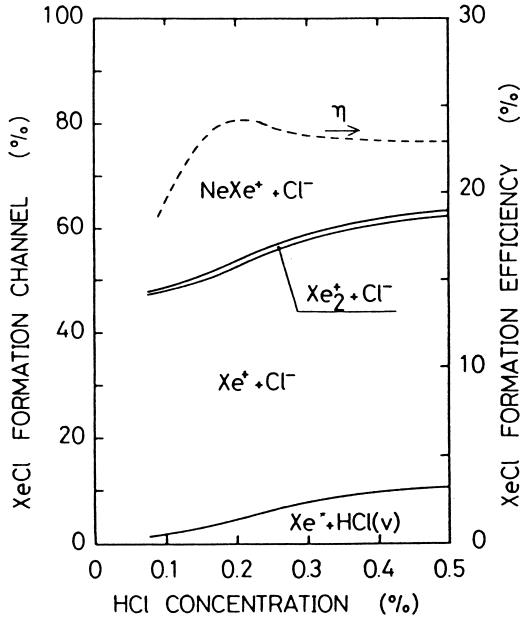


FIGURE 5 Variation of percent contribution of the XeCl* formation channels and the XeCl* formation efficiency with the HCl concentration. The gas mixture is X% HCl/1% Xe/(99-X)% Ne. [Reproduced with permission from M. Ohwa and M. Obara (1986). *J. Appl. Phys.* **59**(1), 32.]

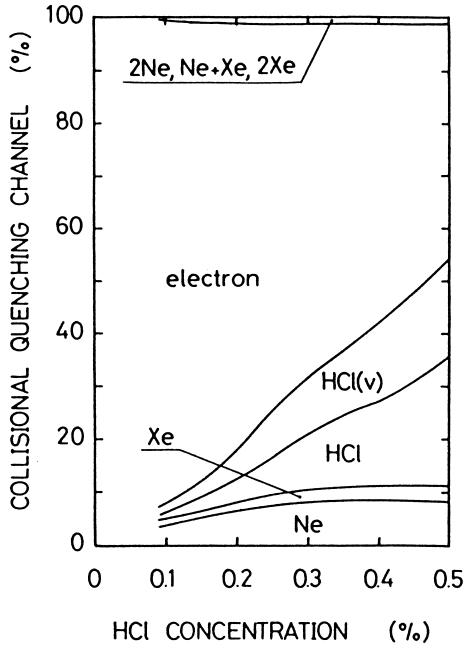


FIGURE 7 Variation of percent contribution of the XeCl* collisional quenching channels with the HCl concentration. [Reproduced with permission from M. Ohwa and M. Obara (1986). *J. Appl. Phys.* **59**(1), 32.]

All the intracavity laser flux cannot be extracted because the RGH laser mixture contains many absorbers at the laser wavelength. The percent contribution of the absorption channel is shown in Fig. 8. Main absorbers seem to be Cl^- and Xe_2^+. The photon extraction efficiency, defined as the ratio of the extracted laser energy to the intracavity laser energy, is in excess of 70% here (see Fig. 6). This means that over 70% of the intracavity laser energy can be

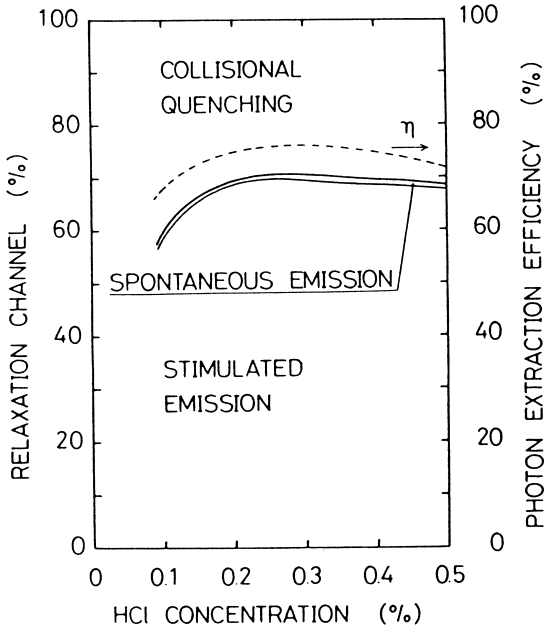


FIGURE 6 Variation of percent contribution of the XeCl* relaxation channels and the photon extraction efficiency with the HCl concentration. [Reproduced with permission from M. Ohwa and M. Obara (1986). *J. Appl. Phys.* **59**(1), 32.]

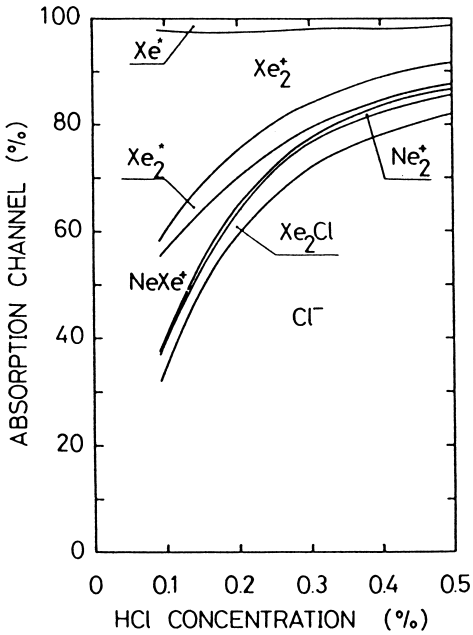


FIGURE 8 Variation of percent contribution of the absorption at 308 nm with the HCl concentration. [Reproduced with permission from M. Ohwa and M. Obara (1986). *J. Appl. Phys.* **59**(1), 32.]

extracted as a laser output. Here, the maximum extraction efficiency can be written as

$$\eta_{\max} \simeq (1 - \sqrt{\alpha/g})^2$$

where g and α are small-signal gain and absorption coefficients, respectively. Therefore, the efficiency is increased with increasing values of g/α.

If the mixing ratio of Xe/HCl/Ne is varied, the electron energy distribution in the discharge plasma changes. As a result, formation of precursors Xe^+, Xe^*, Ne^+, and Ne^* is greatly affected. If helium is used as a diluent gas in place of Ne, the electron temperature also changes, resulting in a different pathway for the XeCl(B) formation and a less effective formation than that of the Ne diluent.

B. Electron-Beam Pumped KrF Lasers

Kinetic processes concerned with the e-beam-pumped KrF laser using a typical mixture of $Kr/F_2/Ar$ are described here.

High-energy electrons of 300 keV to 2 MeV generated from a pulsed relativistic e-beam generator are deposited into the $Kr/F_2/Ar$ mixture to produce the precursors Ar^+, Ar^*, Kr^+, and Kr^*. A typical e-beam current density is from 10 A/cm^2 to 1000 A/cm^2, and its excitation rate ranges from 0.1 MW/cm^3 to 10 MW/cm^3, approximately, depending on the mixture pressure. The high-energy electrons can rapidly be thermalized in He, Ar, and Kr gases in the nano-second time range, and as a result the secondary electrons reach a thermalized energy of around 1 eV. The initially formed Ar^+ and Ar^* are used to form ArF^*, and then KrF^* is formed via the reaction $ArF^* + Kr \rightarrow KrF^* + Ar$. Here Kr^+ is formed via the charge transfer reaction $Ar^+ + Kr \rightarrow Kr^+ + Ar$, and Kr^+ is also used to form KrF^* via $Kr^+ + F^-$. The negative ion F^- is formed via an electron attachment process with the thermalized slow electron $e + F_2 \rightarrow F^- + F$.

The percent contribution to the KrF^* formation as a function of excitation rate is shown in Fig. 9. The $Kr/F_2/Ar$ mixture of 1.5 atm is treated as a typical system. The KrF^* excimer is mainly formed throuh $Kr^+ + F^-$, $ArF^* + Kr$, and $Kr^* + F_2$. The KrF^* formation efficiency reaches roughly 20%.

The percent contributions of the KrF^* relaxation channels are shown in Fig. 10. Sixty to eighty percent of the KrF^* excimers can contribute to the stimulated emission as an intracavity laser flux, depending on the excitation rate. Other relaxation processes are by a slow electron, F_2, Kr, and Ar. In these collisional relaxation reactions the reaction $KrF^* + Kr$ forms the Kr_2F^* trimer, and the reaction $KrF^* + Ar$ forms the $ArKrF^*$ trimer.

About 85% of the intracavity laser flux can be extracted as a laser output energy, as shown in Fig. 11. The rest of the energy is lost in the cavity by absorptions due to F_2, F^-, Ar_2^+, Kr_2^+, and Kr_2F^*, etc.

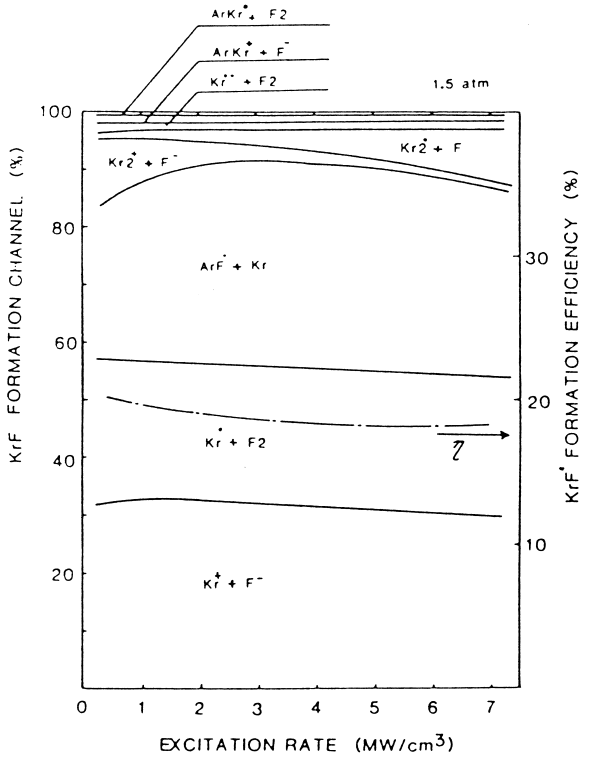


FIGURE 9 Variation of percent contribution of the KrF^* formation channels and the KrF^* formation efficiency with the excitation rate for the 1.5-atm mixture with argon diluent pumped by a 70-nsec e-beam. [Reproduced with permission from F. Kannari *et al.* (1983). *IEEE J. Quantum Electron.* **QE-19**(2). Copyright © 1983 IEEE.]

III. EXCIMER LASER TECHNOLOGY

Typical gas mixtures of rare gas-halide lasers are high-pressure (approximately 1–5 atm) rare gas mixtures containing a small amount of halogen donor. High excitation-rate pumping of several 100 kW/cm^3 to several MW/cm^3 is necessary to efficiently produce high gain of the excimer laser of interest in relation to the absorption. For this purpose, pulsed e-beam pumping and self-sustained discharge pumping have been mainly employed, while proton-beam pumping, nuclear pumping, e-beam sustained discharge pumping, and rf discharge pumping have also been tried.

Shown in Fig. 12 is a plot of the laser output energy as a function of the pulse repetition rate for various reported RGH lasers. The oblique lines show the average RGH laser output power (laser energy per pulse multiplied by pulse repetition rate). It is understood that the recent RGH laser power level lies in the power region of 1 kW, although the laser power level is rapidly being increased. Discharge-pumped RGH lasers capable of producing laser output up to several hundreds W are now commercially available. Region A in Fig. 12 shows the

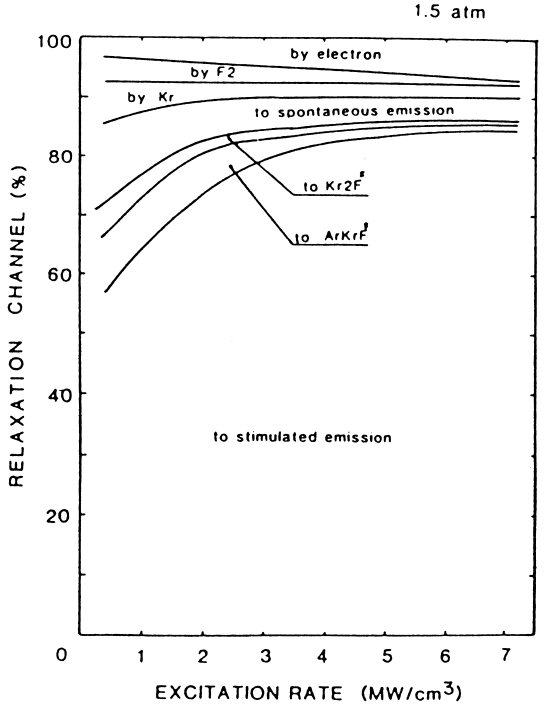


FIGURE 10 Variation of percent contribution of the KrF* relaxation channels with the excitation rate for the 1.5-atm mixture. [Reproduced with permission from F. Kannari *et al.* (1983). *IEEE J. Quantum Electron.* **QE-19**(2). Copyright © 1983 IEEE.]

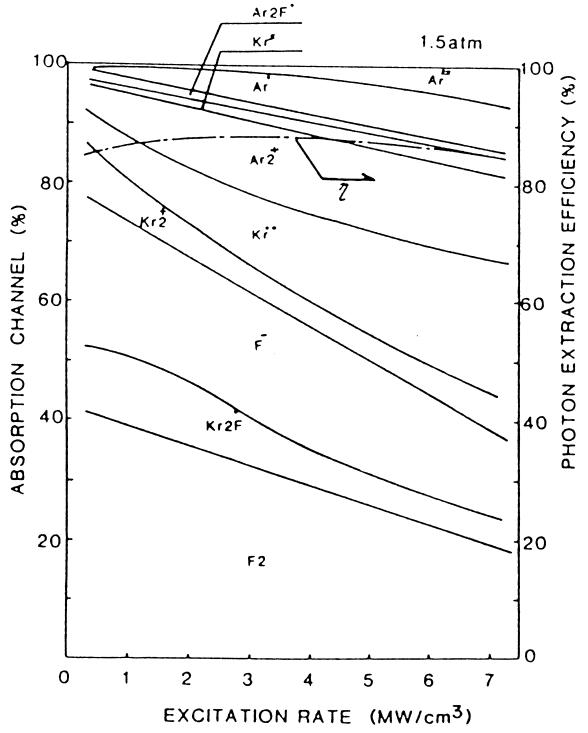


FIGURE 11 Variation of percent contribution of the 248-nm absorption channels and the photon extraction efficiency with the excitation rate for the 1.5-atm mixture. [Reproduced with permission from F. Kannari *et al.* (1983). *IEEE J. Quantum Electron.* **QE-19**(2), 232. Copyright © 1983 IEEE.]

regime in which the average laser power is gained not by pulse energy but by pulse repetition rate. These lasers are predominantly discharge pumped. Region B indicates the regime in which the average power is increased by both pulse energy and pulse repetition rate. Finally, Region C shows the regime in which low-repetition-rate or single-pulse lasers are operated with high-energy laser output. These lasers are mainly pumped by intense relativistic e-beams. The high-repetition-rate technology for these lasers is not a current issue, but energy-scaling technology for single pulse operation is currently a rather urgent issue. The lasers concerned in this region are e-beam-pumped high-energy KrF and XeF lasers for inertial confinement fusion applications and for military applications, respectively. The world's largest KrF laser reported to date is the system called LAM (large aperture module), developed at Los Alamos National Laboratory, and is used as an inertial confinement fusion driver. In 1985 this laser could produce ~10 kJ per pulse output. The highest-energy XeCl laser of 60-J level output pumped by discharges was developed at the US Naval Research Laboratory in 1985.

A. Electron-Beam Pumping

The pulsed-power technology involved in the efficient and spatially uniform generation of intense (high current-density) relativistic e-beams from relatively large aperture diodes has advanced recently for single shot operation. The e-beam generation technology is so scalable in energy that high-energy excimer lasers are more readily realizable than discharge-pumped lasers.

The relevant components of a cold cathode-type e-beam diode suited for excimer laser pumping are shown

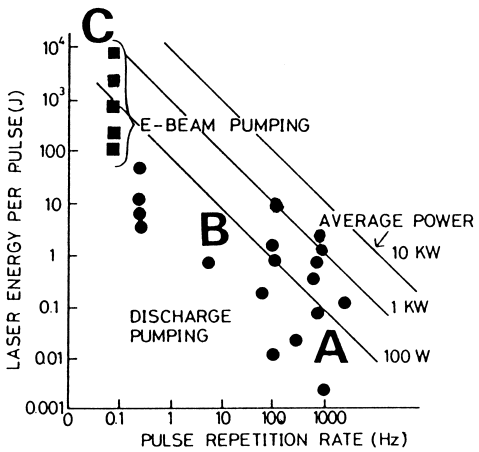


FIGURE 12 Plots of the RGH laser energy as a function of pulse repetition rate for various RGH lasers reported.

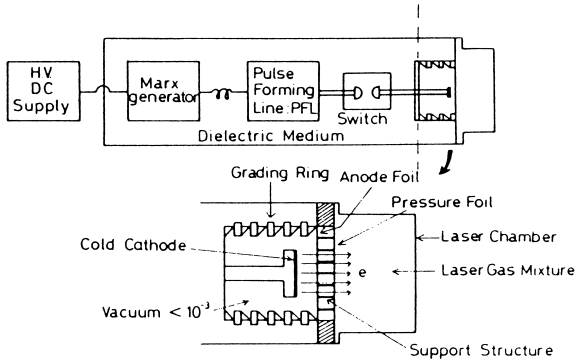


FIGURE 13 Schematic diagram of a typical e-beam pumping system for excimer lasers.

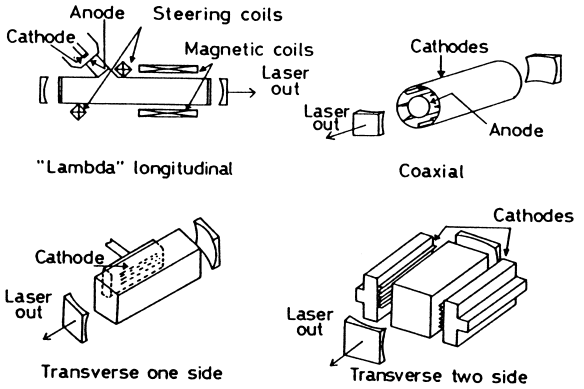


FIGURE 14 Schematic of four major layouts for e-beam pumped excimer lasers.

schematically in Fig. 13. The e-beam diode consists of a cold cathode to which the pulsed high voltage (negative polarity) of 30 to 1500 nsec is applied, a thin foil anode, and a pressure foil separating the high-pressure laser gas mixture from the vacuum diode chamber. These foils are made of a half-mil to three-mil-thick titanium foil or aluminized polyimide (e.g., KAPTON®) film. The support structure is placed between the foils like a Hibachi assembly to withstand the pressure of the laser gas mixture. The diode chamber is typically evacuated to low pressures less than 10^{-3} torr. An intense relativistic e-beam with a current density of 10 to 1000 A/cm^2 and an electron energy of 0.3 to 2 MeV has been used to pump excimer lasers. This high current-density e-beam can be generated only from a cold cathode-type diode made of carbon felt, or multi-blades of tantalum or titanium foils, which can enhance the local electric field strength up to typically 1 MV/cm. The operational characteristics of the cold cathodetype diode approximately obey the Child–Langmuir law (space-charge-limited electron flow). Therefore, the e-beam current density can be approximately written as

$$J_{\mathrm{eb}} = \frac{\sqrt{2}}{9\pi}\left(\frac{e}{m_0}\right)^{1/2}\frac{V^{3/2}}{d^2} = 2.3\times 10^3\frac{V^{3/2}}{d^2}\ \mathrm{A/cm^2},$$

where J_{eb} is the current density in A/cm^2, V the applied voltage in MV, and d is the anode–cathode spacing in cm. Here e and m_0 are the electron charge and mass, respectively. This expression is accurately the nonrelativistic expression. In the region of $V > \sim 0.5$ MV some relativistic correction is required. Thermionic electron emitters (hot cathode) are not used for direct e-beam pumping of excimer lasers (except for indirect e-beam pumping, such as an e-beam-sustained discharge pumping), due to their inherent characteristic of low current-density e-beam generation.

The axial magnetic field is in some cases applied to ensure uniform e-beam generation if total e-beam current is larger than the critical current for self pinching, and also to improve the e-beam utilization.

A variety of e-beam-pumped excimer laser layouts have been successfully tried. Four major geometries are shown schematically in Fig. 14. The merits and demerits of these layouts should be discussed in terms of the energy scaling, the aspect ratio of the pumped region, uniform pumping, e-beam utilization, and output performance of an available pulsed high-voltage generator. To improve the e-beam energy deposition into the laser mixture, the external magnetic field is applied in some cases as shown in Fig. 14. The depth of the pumped region is determined by the electron penetration depth at a given accelerating voltage of electrons. The variation of the electron range in one atmospheric pressure of various rare gases commonly used for RGH lasers is shown in Fig. 15. As can be seen in Fig. 15, a higher-Z rare gas can deposit more e-beam energy because

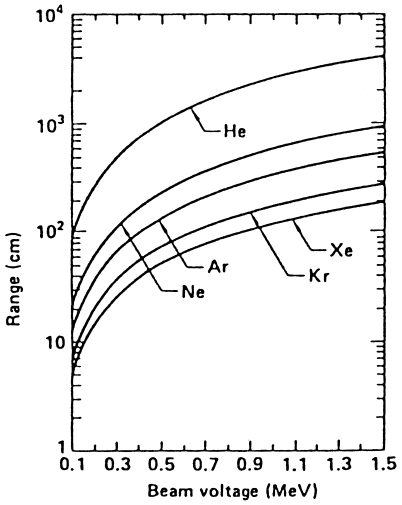


FIGURE 15 Berger and Seltzer range for commonly used rare gases at 1 atm. [Reproduced with permission from J. J. Ewing (1979). Excimer lasers, *in* "Laser Handbook" Vol. 3 (M. L. Stitch, ed.), North-Holland Publ., Amsterdam.]

of its shorter propagation distance. Actually, a typical KrF mixture is a 2–3 atm mixture of $Kr/F_2/Ar = 10/1/89(\%)$. In 1983, low pressurization of the KrF laser mixture was proposed and successfully tried by increasing a higher Z krypton content than argon or by using an argonless mixture of Kr/F_2 to keep the e-beam deposition constant.

Under the high-excitation-rate pumping, the precursor for a RGH excimer is a rare gas ion, as mentioned previously. Some typical values of the average energy (eW_i) to produce an ion pair of rare gas by an e-beam are shown in Table III along with their ionization potentials. If the energy deposited into the rare gas mixture is calculated, or measured, the net rate of ion-pair production can be calculated. The secondary electrons can rapidly cool down to around 1 eV in pure rare gases such as He, Ar, Kr, etc. In Ne gas the electrons cannot be thermalized rapidly. The e-beam energy deposited into the rare gases can be measured by a calorimetric and pressurejump method. This energy measurement is required to estimate the intrinsic efficiency for e-beam-pumped lasers, defined as a ratio of the laser output energy to the deposition energy. Experimentally, intrinsic efficiencies for KrF and XeCl lasers are typically 6–12% and around 5%, respectively. As a single shot, or low-repetition-rate device several companies can supply complete e-beam accelerators for excimer laser pumping.

The world-wide KrF lasers pumped by e-beams are shown in Fig. 16. It is understood that the scaling-up of the laser energy is done experimentally by increasing laser pulse width (in other words, pumping pulse width). These large-scale KrF laser systems have all been constructed to be used as a power amplifier in a fusion laser system. The NIKE KrF laser system in Naval Research Laboratory can produce 4–5 kJ of UV light out of the large amplifier in a 4-ns pulse. The SPRITE laser employed a four-sided transverse excitation scheme, while RAPIER and LAM employed two-sided transverse excitation schemes as shown in Fig. 14. The repetitive operation of the e-beam-pumped RGH laser is mainly limited by the heating of the anode foil. By specially cooling the anode foil, the repetitive operation is realizable at thermal loadings

TABLE III Comparison of Average Energy to Ionize Rare Gases and Their Ionization Potentials

Rare gas	eW_i (eV)	Ionization potential (eV)
He	42	24.6
Ne	36	21.6
Ar	26	15.8
Kr	24	14.0
Xe	22	12.1

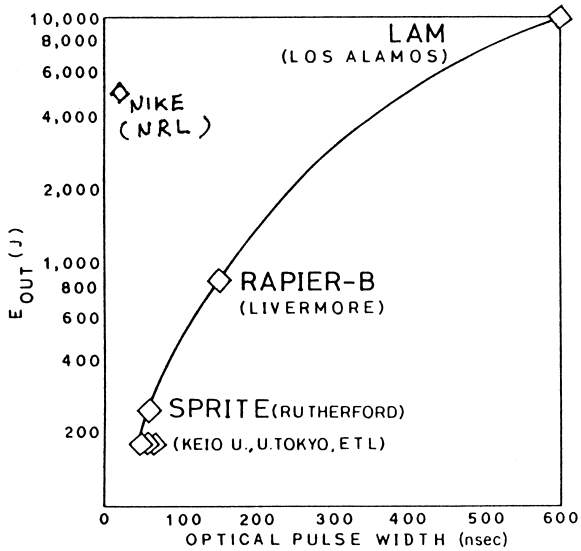


FIGURE 16 World-wide KrF lasers pumped by e-beams. [Adapted with permission from L. A. Rosocha (1985). A short-pulse multikilijoule KrF inertial fusion laser system, presented at CLEO'85 (Conference on Lasers and Electro-Optics), May, LA-UR-85-1506.]

of up to several hundreds of watts per square centimeter onto the anode foil.

B. Discharge Pumping

The discharge pumping technology is well suited to excite high-repetition-rated excimer lasers, which can operate so far with laser output energies of several milijoules per pulse to a joule level at a repetition rate up to several kHz.

Applying an X-ray preionization to a high-pressure mixture of Xe/HCl/Ne, an energy scaling study of discharge-pumped XeCl lasers has also been done since 1979 and in 1985 a high-energy laser of 60-J level output has been successfully demonstrated as a single pulse device at the U.S. Naval Research Laboratory.

A pumping rate on the order of 1 GW per liter of discharge volume is necessary to efficiently produce excimer laser gains. The discharge resistance of the excimer discharge load is typically around 0.2 ohms. Therefore, the typical voltage of 20 kV gives discharge current as high as 100 kA. This high current cannot be switched by thyratron switches due to the heavy loadings of the switch. Hence, the primary low power and long pulse is produced in a primary circuit and then this pulse is compressed in the secondary circuit into the secondary high power and short pulse, which can efficiently pump the excimer lasers. Discharge pumping circuits developed so far are mainly classified into capacitor transfer circuit, pulse forming line (PFL) circuit, magnetic pulse compressor circuit, and

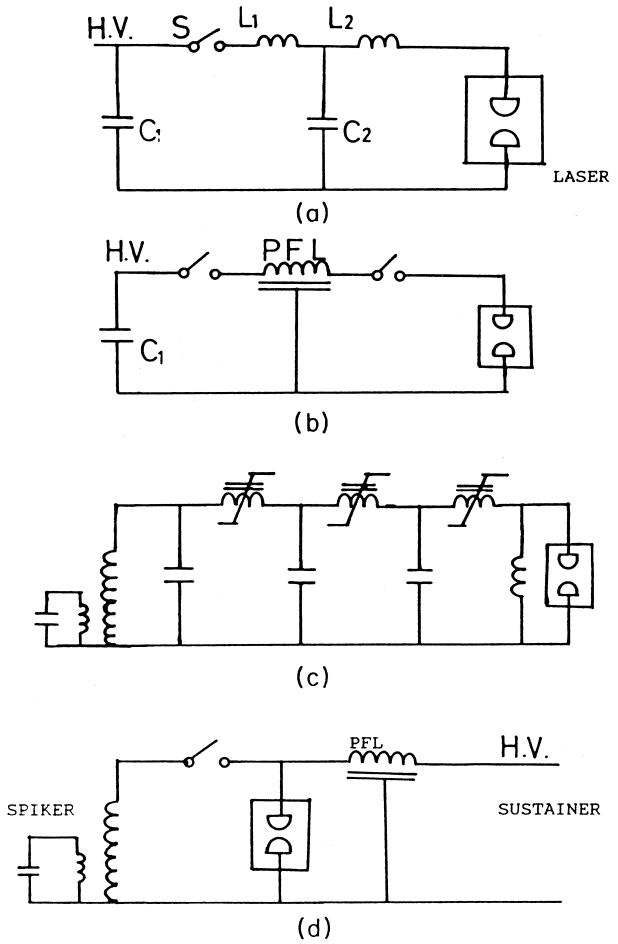


FIGURE 17 Schematic of widely used RGH laser excitation circuits. (a) Charge transfer circuits, (b) pulse forming line (PFL) circuit, (c) magnetic pulse compressor circuit, and (d) spiker sustainer circuit.

spiker sustainer circuit, which are shown schematically in Fig. 17.

1. Pumping Circuits

a. Charge transfer circuit. The charge transfer circuit is a circuit that is widely used in relatively small-scale repetition-rated RGH lasers and commercially available RGH lasers. Among this type of excitation circuit, the resonant charge transfer mode ($C_1 = C_2$ in Fig. 17) is mostly employed, because the charge transfer efficiency from C_1 to C_2 is maximized. Figure 18 shows a set of typical values for this type of excitation circuit used to pump RGH lasers. When operating the circuit shown in Fig. 18, the operating characteristics of the XeCl laser can be numerically analyzed by computer simulation, the results of which are shown in Fig. 19. A 4-atm mixture of Xe/HCl/Ne = 1.3/0.1/98.6(%) is assumed as an optimum

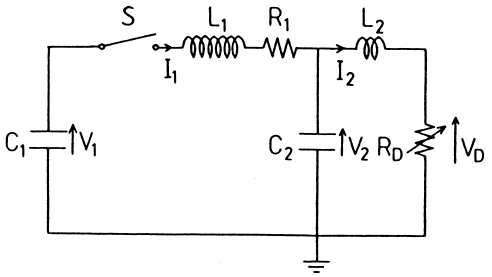


FIGURE 18 Equivalent circuit of the charge transfer circuit. $C_1 = 60$ nF, $L_1 = 200$ nH, $R_1 = 0.2\ \Omega$, $C_2 = 60$ nF, $L_2 = 5$ nH, and V_D = laser discharge voltage. Discharge volume is 2.0(H) × 1.0(W) × 6.0(L) = 120 cm^3.

gas mixture. Peak value of the primary current I_1 is less than 5 kA, which is within the current ratings of thyratrons widely used for RGH lasers. While peak value of the secondary current I_2 is increased up to about 18 kA. This increase is attributed to the fact that L_2 is much less than L_1 as shown in Fig. 18. The corresponding excitation rate is ~1.6 MW/cm^3, which gives a specific laser energy of 3 J/liter. The excessive excitation rate will result in the low laser efficiency mainly due to the excimer deactivation by discharge electrons.

Using this type of excitation circuit, a maximum laser efficiency of nearly 3% for both XeCl and KrF lasers was obtained with output energies of around 300 mJ. It is noted

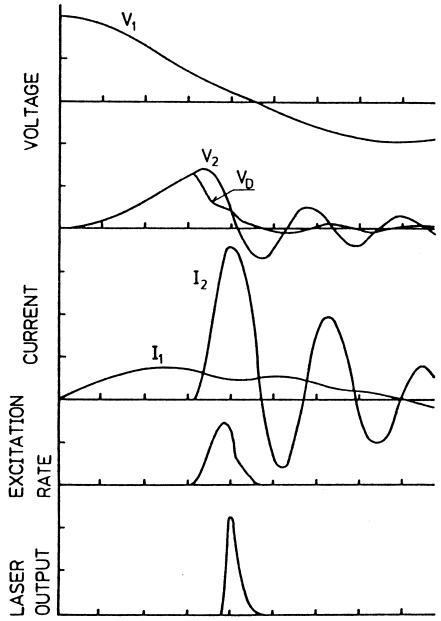


FIGURE 19 Temporal histories of voltage, current, excitation rate, and XeCl laser output at a charging voltage of 20 kV. The inductance L_2 of the secondary loop is 5 nH. Each ordinate is as follows: voltage; 10 kV/div.; current, 10 kA/div.; excitation rate, 1 MW/cm^3/div.; and laser output; 5 MW/div.

that laser efficiency and output energy are experimentally traded off. This is attributed to the fact that the breakdown voltage is determined not by the charging voltage V_1 but by the laser gas mixture. This laser is mostly equipped with the UV spark array preionization, which is directly and simultaneously driven by the main discharge circuit because of its simplicity.

b. Pulse-forming line (PFL) circuit. The PFL circuit uses a low-impedance (typically less than 1 ohm) PFL consisting of solid or liquid dielectric materials in place of capacitor C_2 in the capacitor transfer circuit in Fig. 17. Solid dielectric materials include Mylar® (polyamide) and epoxy sheets, while liquid materials are deionized water or deionized ethylene glycol. Co-axial-type, parallel-plate-type, and Blumlein-type PFLs were used for the RGH laser excitation.

The advantage of this circuit is that it makes it possible to inject a quasi-rectangular waveform pulse into a discharge load. The pulse duration and output impedance are simply selected by changing the length and geometry of the PFL, respectively. This pumping system is well suited for high-energy XeCl lasers of in excess of several joules per pulse.

Figure 20 shows a schematic of a PFL discharge-pumped XeCl laser together with measured and theoretical time histories of discharge voltage and current. The scaling-up of the PFL circuit is readily realizable. However, establishment of large-volume uniform self-sustained discharge in a RGH laser mixture is a state of the art technology. As an output switch in this scheme, rail gaps, UV-laser-triggered rail gaps, or magnetic switches, which will be described below, are used instead of thyratrons.

In 1985 a maximum output of 60 J in an XeCl laser was obtained with X-ray preionization at the U.S. Naval Research Laboratory. An X-ray preionized XeCl laser of 100 W at 100 Hz is commercially available that is pumped by a PFL discharge.

c. Magnetic pulse compressor (MPC) circuit. A magnetic switch (MS) is a new switch consisting of a magnetic core made of ferromagnetic materials, being completely different from a gas-discharge switch such as thyratron. A magnetic core is completely saturated in an ON phase. If these MSs are connected in series, the primary pulse is successively compressed only by decreasing the saturated inductance of the magnetic coils. A three-stage MPC is schematically shown in Fig. 17. A typical MPC circuit is shown in Fig. 21, together with each voltage and current. An 8-μsec pulse is compressed to a 100 nsec pulse. The MS is a long-life or endless-solid-state switch because it experiences no erosion, and it can act reliably at high-

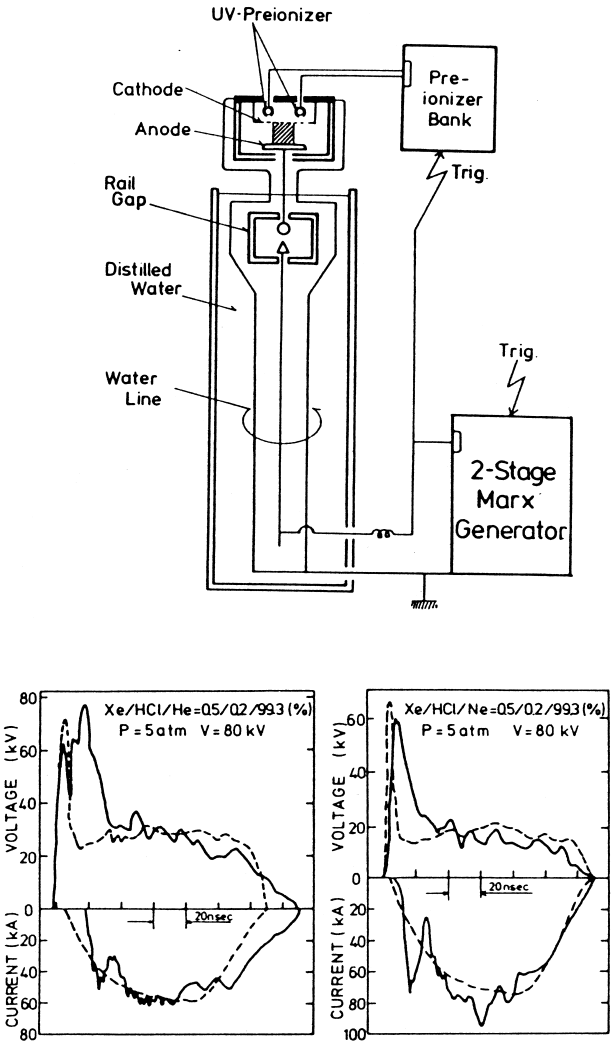


FIGURE 20 Schematic of a PFL discharge pumped XeCl laser (top) together with measured (solid line) and theoretical (dashed line) time histories of the discharge current and voltage (bottom). [Reproduced with permission from H. Hokazono *et al.* (1984). *J. Appl. Phys.* **56**(3), 680.]

repetition frequency. An XeCl laser operating at 150 W in average laser output at 500 Hz is commercially available.

d. Spiker sustainer circuit. The spiker sustainer circuit is an advanced excitation circuit for RGH lasers. The low-impedance PFL sees initially the discharge load being an open load so that the reflection of the voltage pulse may occur due to the impedance mismatch. To eliminate this unfavorable voltage reflection, the high-voltage high-impedance pulser initially breaks down the laser gas mixture and then the other power supply maintains the discharge plasma to condition the electron energy distribution for efficient pumping of the RGH laser. In this scheme, the former is called a spiker, the latter being a sustainer. This

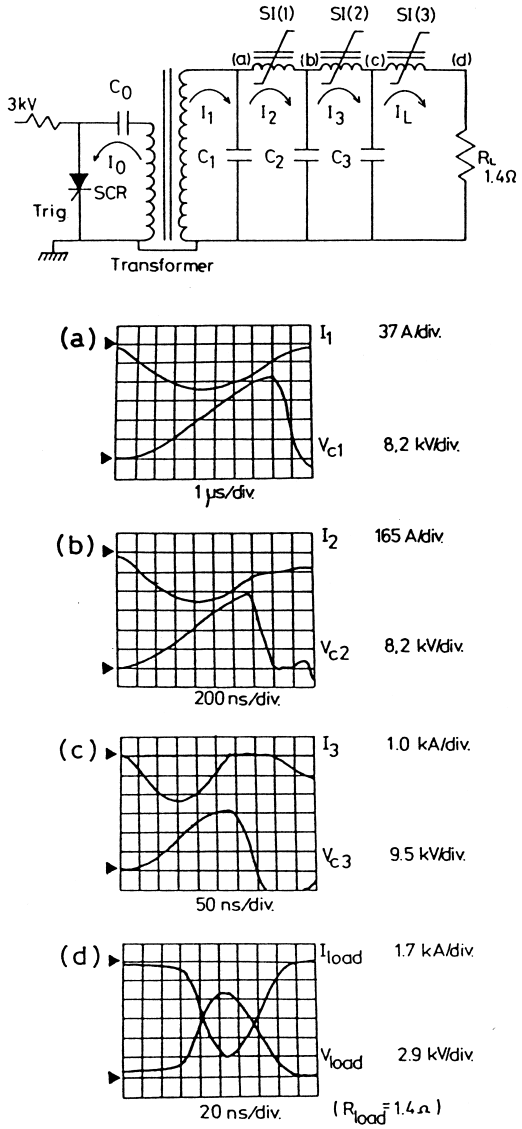


FIGURE 21 A typical MPC circuit (top) together with each voltaghe and current (bottom). $C_0 = 2.5$ μF and $C_1 = C_2 = C_3 =$ 13.6 nF. (a) First stage, (b) second stage, (c) third stage, and (d) load. [Courtesy of Tsutomu Shimada.]

scheme is more complicated than the PFL circuit alone, but higher overall electrical efficiency is expected, because of sufficient impedance matching. Using this scheme, an XeCl laser of 4.2 J has been realized with a high efficiency of 4.2%.

2. Preionization Technology

To initiate a volumetrically uniform avalanche discharge in the 2–4 atm rare gas–halogen mixture, preionization of the high-pressure mixture prior to the initiation of the main discharge is indispensable. Spatial uniformity of the

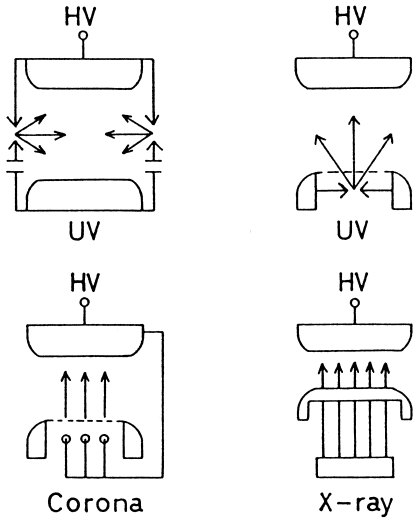


FIGURE 22 A variety of preionization technologies developed to date. UV and corona preionization are suitable for high repetition rate and relatively small lasers. The X-ray preionization is suitable for large-scale discharge lasers.

preionization in the rare gas-halogen mixture is the most important issue. Especially, its uniformity perpendicular to the discharge electric field is of importance. The preionization electron number density on the order 10^6 to 10^{12} #/cm^3 is experimentally utilized, and is dependent on the preionization strength employed and gas mixtures.

A variety of preionization technologies developed so far are shown in Fig. 22. The simplest and most convenient preionization technology is a UV photo-preionization using a photo-electron emission process. The UV photons are generated by the use of a pin-arc discharge and a dielectric surface discharge, both of which are induced in the laser gas mixture, and UV RGH laser beams. UV preionization via pin-spark discharge is widely used in commercially available RGH lasers.

X-ray preionization technology was successfully applied in 1978 to high-pressure RGH lasers, and has been used preferably in large-scale RGH laser devices. A high-energy XeCl RGH laser of 60-J level output employed the X-ray preionization. By masking the X-ray beams by Pb plates as shown in Fig. 23 to determine the preionized volume, only the preionized (X-ray irradiated) volume can be discharge pumped with less effect of the nonuniformity of the electric field induced by main electrode geometries than that with noncollimated UV preionizers.

The effect of the preionization electron number density is large and it is experimentally revealed that the RGH laser energy and laser pulse width are increased in logarithmic form with increasing initial electron number density. An example for the KrF laser is shown in Fig. 24.

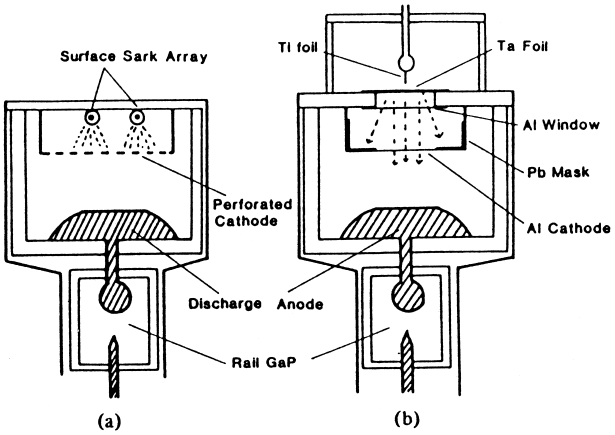


FIGURE 23 Comparison of (a) noncollimated UV and (b) collimated X-ray preionizers. [Reproduced with permission from K. Midorikawa *et al.* (1984). *IEEE J. Quantum Electron.* **QE-20**(3), 198. Copyright © 1984 IEEE.]

3. Long Pulse Operation

Long pulse operation of the RGH lasers pumped by a self-sustained discharge is interesting in view of gaseous electronics and ultrashort pulse generation via mode-locking technology. Many RGH lasers have so far been operated in a short pulse (10–50 nsec) regime. In 1985, long pulse oscillations of 1.5-μsec XeCl lasers were demonstrated by conditioning the discharge in a sophisticated manner. Long pulse RGH lasers with reduced intensity can efficiently deliver their energies through quartz fibers.

4. High-Repetition-Rate Operation

High-repetition-rate operation of RGH lasers is desirable for high average-power generation. For this purpose thyratrons have been commonly used as a switching element.

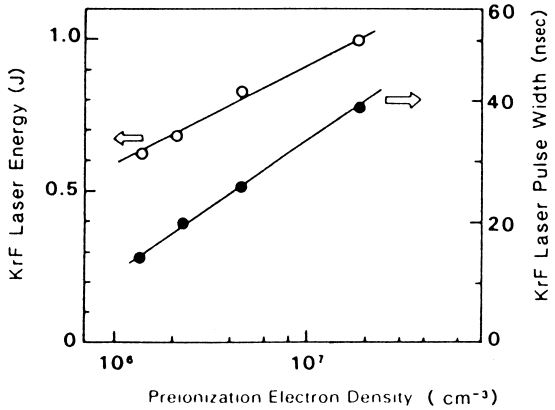


FIGURE 24 Effect of preionization electron number density on KrF laser energy and pulse width. [Reproduced with permission from K.cxcx Midorikawa *et al.* (1984). *IEEE J. Quantum Electron.* **QE-20**(3), 198. Copyright © 1984 IEEE.]

TABLE IV Relationship of High Repetition Rate RGH Lasers and the Switching Elements Used

Year reported	RGH	Repetition rate (Hz)	Average power (W)	Switch[a]
1976	KrF	20	16×10^{-3}	G
1977	XeF	200	50×10^{-3}	T
	KrF	1000	40	T
1979	KrF	1000	55	T
	XeF	2000	24	$4 \times$ T
1982	XeCl	750	3.5	T + M
1983	XeCl	1500	130	T + M
	XeCl	300	120	$2 \times$ T
	XeCl	400	180	G
1984	XeCl	500	150	T + M
	XeCl	250	40	S + M
1986	XeCl	500	300	T + M
1995	XeCl	800	2100	$2 \times$ T + M
1998	XeCl	1000	1000	$2 \times$ T + M

[a] G: Spark Gap; T: Thyratron, M: Magnetic Switch, S: Semiconductor Switch.

The operational performance of the used switch is one of the repetition-rate limiting issues.

The relationship of high-repetition-rate RGH lasers to the switching elements used is shown in Table IV. Because of rapid progress in both the pulsed power technology involved in the modulator and laser gas purification, operations at repetition rates of up to 2 kHz and an average laser power of up to 2.1 kW have been demonstrated separately. Using a single thyratron an average laser power of up to 55 W is obtained and an average laser power of in excess of 300 W is obtained using multi-thyratrons in parallel, or a thyratron with a magnetic assist. The magnetic assist means that using a saturable inductor in series with a thyratron switch the current-rise rate through the thyratron is reduced so as to decrease the energy dissipated in the thyratron. The allowable rise rate of the current for thyratron is less than 10^{11} A/sec, while that of a gas-insulated spark gap reaches 10^{13} A/sec. However, their operational lifetime becomes extremely short under the heavily loaded conditions required for efficient RGH laser excitation.

For attaining the nearly endless lifetime of a RGH laser exciter, an all-solid-state circuit is the state of the art and appears to be promising. At present, commercially available high-power semiconductor switches have been well developed, but maximum ratings such as hold-off voltage, peak current, and current-rise rate cannot fulfill the switching requirements necessary for efficient RGH laser excitation. Therefore, additional voltage transformer circuits and a magnetic pulse compression circuit are needed, as shown previously in Fig. 21.

In addition to the high-repetition-rate exciter, gas purification and aerodynamic technologies such as a fast gas

circulation system and an acoustic damper at repetition rates exceeding the multi-kilohertz range are required to realize long-life high-repetition-rated operation of the RGH lasers. A long operational shot life of in excess of 10^8 shots has been demonstrated to date.

C. Control of Laser Properties

1. Narrow Linewidth

The fluorescence from the rare gas-halides with bound lower states (e.g., XeF and XeCl) typically contains line structure originating from the transitions from the lowest vibrational level of the upper laser state to various vibrational levels of the ground electronic state. When these lasers are operated without wavelength tuning elements in the resonator, the laser emission consists of radiation on the two or three strongest vibrational transitions. For the rare gas-halides with unbound lower electronic states (e.g., KrF and ArF), the fluorescence shows a weakly structured band. The untuned laser emission typically consists of a band approximately 100 cm^{-1} wide near the peak of the fluorescence.

The linewidth of the RGH lasers can be reduced and the tunability of the RGH lasers can be realized using gratings, prisms, etalons, or some combination of these elements in the laser resonator. The lowest level of frequency selection is usually obtained with a grating used in a Littrow geometry at one end of the resonator or with one or more prisms for the tuning element. This technique provides tunability over the emission band of the RGH lasers and realizes selection of single vibrational transition in lasers such as XeF or XeCl. Narrow linewidths of 80 cm^{-1} for KrF and 5 cm^{-1} for ArF have been demonstrated with two prisms in the resonator.

Narrower linewidths of RGH lasers are needed especially for the photolithography of the ultra-large-scale integration (ULSI) devices. With intracavity etalon, either alone or with a grating, narrower linewidths are achievable. A combination of three etalons has been used in a XeCl laser to produce a linewidth of 220 MHz and single-mode operation with a linewidth of 30 MHz. Narrower linewidths can also be obtained by expanding the intracavity laser beam cross section to cover the large number of grooves on the grating, which is used especially in dye lasers. This configuration includes grazing-incidence gratings and Littrow gratings with a prism beam expander. Linewidths of 0.3 cm^{-1} for KrF and 0.5 cm^{-1} for ArF and XeCl have been achieved by the use of a grazing-incidence grating.

2. Low Divergence

Because of the high gain and wide gain bandwidth of RGH lasers, high-Fresnel-number stable resonators, which are commonly used to efficiently extract the available laser energy, produce a laser output with high spatial divergence. A low divergence laser beam from RGH lasers can be obtained by using a confocal unstable resonator with a large magnification or by simply using a low-Fresnel-number stable resonator. The use of an unstable resonator can efficiently extract laser output with high beam quality, while the use of low-Fresnel-number stable resonators results in a substantial decrease in the available laser energy. Rare gas-halide laser beams with divergencies within a factor of one or two of their diffraction limits are obtainable without any difficulty by the use of high-magnification confocal unstable resonators. Typical output coupling methods of confocal unstable resonators are schematically shown in Fig. 25.

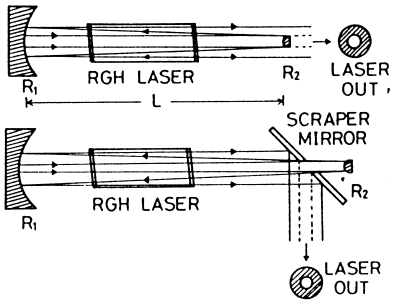


FIGURE 25 Typical output coupling methods for confocal unstable resonators.

3. Injection Locking

For some applications, it is necessary to produce RGH lasers at a reasonable output power level with low spatial divergence and narrow bandwidth. To fulfill these requirements, use of either a master oscillator/power amplifier system or an injection-locked resonator is the most convenient method. Figure 26 shows a schematic of an injection-locked resonator. In this scheme the slave oscillator is an unstable resonator that can control the output energy and spatial mode quality. Input pulse can be made by a master oscillator typically consisting of a cavity with a line-narrowing function. An input power level on the order

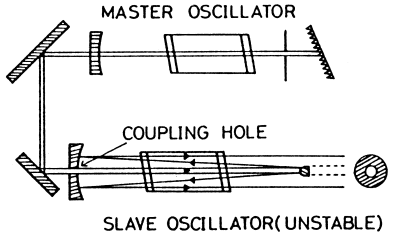


FIGURE 26 Schematic of an injection-locked resonator (unstable).

TABLE V Ultrashort Pulse RGH Lasers

Excimer	Bandwidth (cm^{-1})	Bandwidth-limited pulsewidth (fsec)	Saturation energy (mJ)	Present status	
				Output energy	Pulsewidth
XeCl	120	88	1.35	300 mJ	310 fsec
			(<3 psec)	12 mJ	160 fsec
KrF	325	32	2.0	1.5 J	390 fsec
				4 mJ	45 fsec
ArF	400	26	3.0	30 mJ	10 psec

of 10^{-3} or less of the output laser level is effective. When a phase-conjugate mirror employing an effective stimulated Brillouin scattering (SBS) material in the UV wavelength (such as C_6H_{14} or SF_6) is used as a back mirror of an amplifier, phase front distortion induced on the laser beam in the amplifier can be eliminated, and then the highly controlled properties of the input laser can be realized on the amplified beam. High-power laser extraction is more readily achievable as a by-product with injection locking than without injection because the injected seed pulse assists the build-up of the intracavity flux and reduces energy loss. This effect is more remarkable in shorter duration lasers.

In injection-locked RGH lasers, an input pulse can be made by an entirely different kind of laser. When radiation generated from a visible dye laser chain pumped by a flashlamp, an Ar ion laser, a frequency-doubled YAG laser, or an RGH laser is injected into the RGH amplifier after one or more stages of frequency conversion, a highly controlled laser pulse with desirable spectral, spatial, and temporal beam quality can be amplified to high energy level. The use of a cw single-mode dye laser provides a narrow band seed signal whose wavelength is tunable over the bandwidth of the RGH laser. A system of this type has also been adapted for an e-beam pumped XeF(C-A) laser, and the extremely wide bandwidth of 80 nm was continuously tuned with a linewidth of less than 0.001 nm and an intrinsic efficiency of 1.5%. Beam divergencies within a factor of three of the diffraction limit were achieved.

IV. APPLICATIONS AND TRENDS

A. Ultrashort Pulse Excimer Laser

The broad bandwidth afforded by RGH lasers makes these systems promising candidates for constructing ultrashort pulse laser systems. Although direct generation of ultrashort RGH laser pulses by using either active or passive mode locking is limited to its pulsewidth on the order of or slightly shorter than 1 nsec due to the short gain duration, successful ultrashort pulse generation using visible dye lasers and their wavelength conversion technologies have made it possible to generate high-power subpicosecond UV lasers using RGH gain media for an amplifier. The shortest pulses and highest powers achieved by various RGH lasers are listed in Table V. The highest laser peak power of 4 TW in a 390 fsec pulse has been obtained by amplifying a frequency tripled short pulse dye laser generated in a synchronously pumped dye laser stage by three discharge pumped KrF amplifiers and an e-beam pumped KrF amplifier. The shortest RGH laser pulse of 45 fsec was obtained by amplifying a frequency doubled short pulse dye laser which was generated in a distributed feedback dye laser cavity pumped by a cavity quenched dye laser pulse.

Since the storage times of RGH lasers are typically 1–2 nsec, saturation fluence is typically 1–2 mJ/cm^2, which is much lower than the values of other high-power lasers. Therefore, an amplifier with large aperture is required to amplify a laser pulse to very high energy levels. In such a system, amplified spontaneous emission (ASE) generated from the final large aperture amplifier tends to reduce the aspect ratio of the output laser pulse. At 248 nm wavelength, acridine dye dissolved in ethanol acts as an effective saturable absorber.

When focusing a subpicosecond RGH laser pulse generated in these systems, intensity reaches well over 10^{17} W/cm^2. High-order harmonic generation of these short pulse RGH lasers can decrease their wavelength down to the XUV wavelength region. A 14.6 nm pulse was obtained by 17th-harmonic generation with a picosecond KrF laser. An ultrashort UV probe continua was also generated through self-phase modulation (SPM) induced in high peak power RGH lasers focused into high-pressure gases. A TW-level RGH laser will be able to generate continuum pulse in the XUV spectral range.

B. Excimer Laser Lithography

The photolithography has been supporting mass production of the ULSI semiconductor devices. From 1960s through the middle of 1990s, mercury lamps (g-line at 436 nm and i-line at 365 nm) were used as light sources. Along the evolution of electronics and computers, the number of transistors in a chip became larger and the

capacity of a memory such as DRAM has been increased by a factor of four in every three years.

The resolving power of the photolithography is expressed as:

$$R = k\lambda/NA, \tag{1}$$

where k is the process factor ($\sim$0.4 for mass production), λ the wavelength of a light source, and NA is the numerical aperture of the projection system ($\sim$0.6 for a standard system). To make the feature size of the line pattern smaller, the wavelength of the light sources should be shorter. From Eq. (1), assuming $k = 0.4$ and $NA = 0.6$, the resolving power is 0.29 μm for $\lambda = 436$ nm (g-line); 0.24 μm for $\lambda = 365$ nm (i-line); 0.17 μm for $\lambda = 248$ nm (KrF laser); 0.13 μm for $\lambda = 193$ nm (ArF laser); and 0.11 μm for $\lambda = 157$ nm (F_2 laser). Merits of the excimer laser lithography are the shorter wavelength as well as the higher intensity than the mercury lamps.

Equation (1) is valid for a monochromatic light source, and the resolution becomes worse for a larger spectral-width light source. Therefore, the line width of light sources should be narrowed. For periodic line patterns, resolution enhancement techniques (RET), which utilize interference of coherent light can make the feature size smaller than the resolution limit. The phase-shifting masks or new optical geometries have been developed.

The advantages of the excimer laser over the lamps are the shorter wavelength and higher intensity with narrower linewidth.

The first laser lithography for mass production is KrF excimer laser lithography. Since the middle of 1990s, it is used for mass production of 256-Mb DRAM with a feature size of 0.25 μm. The linewidth of $\sim$0.6 pm with spectral narrowing technique is widely used. The KrF laser operates at several kHz, with an average output power of several tens watts, and a gas lifetime of over 1×10^8 shots.

The next light source will be ArF laser, which will be used for 0.13-μm 4Gb-DRAM production. As for ArF lasers almost the same laser performance is required as the KrF laser, and the ArF laser development is almost completed.

There are several candidates for the post-ArF laser lithography: 157-nm F_2 excimer laser lithography, 13.5-nm extreme ultraviolet (EUV) lithography, e-beam projection, X-ray lithography, and so on. The most direct extension of ArF laser lithography is probably to the F_2 laser lithography because it uses almost the same technique. However, in order to reduce the strong optical absorption by molecular oxygen, the optical path should be evacuated or purged by molecular nitrogen or rare gases. The first demonstration of lithography using F_2 laser was demonstrated by contact lithography in 1983. The first projection lithography by F_2 laser demonstrated a 80-nm feature size using resolution enhancement techniques without spectral narrowing techniques in 1997. Development of F_2 laser device for lithography is very rapid. The requirements for laser lithography will be that a repetition rate is several kHz, average power is over several tens watts, linewidth is less than 0.2 pm, and gas lifetime is over 1×10^8 shots. In 1999, F_2 laser with several kHz, 20 W, and 2.5×10^7 shots lifetime was commercially available.

The F_2 laser lithography will be the last laser lithography for mass production because there are no high-power discharge-pumped lasers with shorter wavelength.

C. Molecular Fluorine Excimer Laser

The molecular fluorine (F_2) laser is one of the high-power excimer lasers and a discharge-pumped excimer laser in the VUV region of the spectrum. Among the discharge-pumped lasers that can generate laser output energies over 10 mJ/pulse, it has the shortest wavelength, around 157.6 nm. The laser spectra ($D'^3\Pi_{2g} \rightarrow A'^3\Pi_{2u}$) consist of some vibrational-rotational transition lines. The emission lifetime of the upper laser level is 3.7 ns, and the spectral width of the spontaneous emission is $\sim$0.3 nm. The potential curves are shown in Fig. 27. Lasing at 156.7 nm, 157.5 nm, and 157.6 nm is reported using

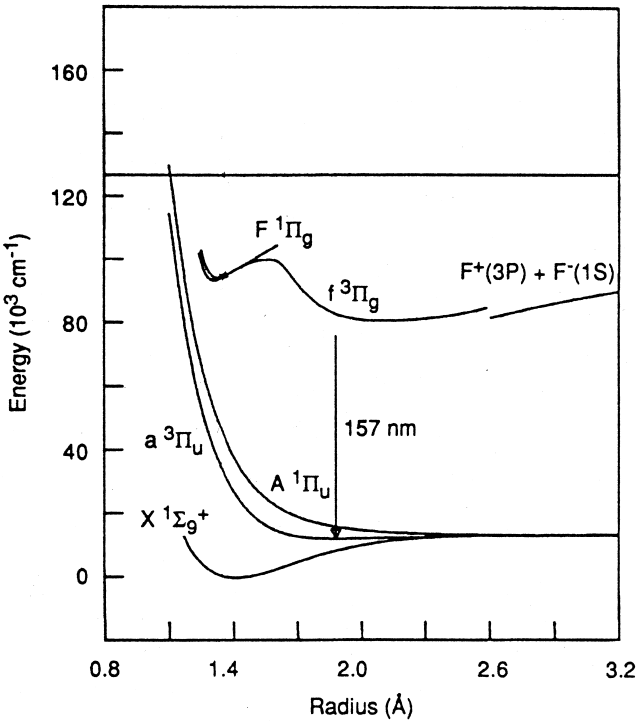


FIGURE 27 Potential energy diagram of the F_2 excimer. [Reproduced with permission from G. W. Faris, M. J. Dyer, D. L. Huestis, and W. K. Bischel (1992) *J. Chem. Phys.* **97**(9), 5964.]

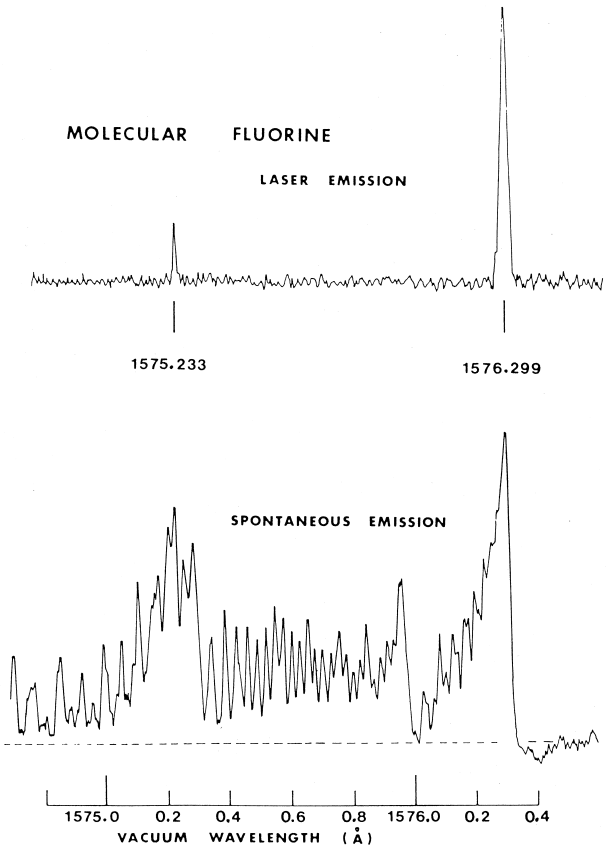


FIGURE 28 Spectra of spontaneous and stimulated VUV emission of F_2 excimer. [Reproduced with permission from T. J. Mckee (1985) *Can. J. Phys.* **63**(2), 214.]

e-beam pumping, while lasing at 157.5 nm and 157.6 nm is reported using discharge pumping. The laser and spontaneous emission spectra are shown in Fig. 28. The linewidth of the lasing spectra is $0.01 \sim 0.05$A.

The first lasing is demonstrated by Rice et al. with e-beam pumping in 1977 and the first discharge-pumped lasing is by Pummer et al. in 1979. The laser gas mixture usually consists of $<1\%$ F_2 and He buffer. Because of the short emission lifetime of the upper laser level, fast pumping with high peak power is required for producing a large laser gain. At the beginning of the discharge-pumped F_2 laser research, the same device as the RGH excimer lasers was used, and the laser performance was not as good as the RGH lasers. Improvement of the laser performance with higher-pressure gas mixture was demonstrated theoretically by Ohwa et al. in 1987 and experimentally by Kakehata et al. The output energy was increased over 100 mJ/pulse, and the specific output energy reached $\sim$2.9 J/liter at the gas pressure around 10 atm.

In order to avoid optical absorption by molecular oxygen, the laser systems have to be evacuated or purged. This difficulty had prevented scientists from measuring the laser parameters. Injection seeding with unstable amplifier is also demonstrated to reduce the beam divergence. Recently, a further improvement in the output energy is reported of up to 240 mJ/pulse, and longer pulse operation ($\sim$70 ns) with X-ray preionized prepulse-main pulse pumping scheme is demonstrated. Ultrashort pulse amplification of several-hundreds-femtosecond pulses is also reported by Momma et al. in 1993. Wavelength conversion by stimulated Raman scattering in H_2 is demonstrated.

Other applications of F_2 lasers are optical pumping of Br and OCS Raman lasers, NO laser, and solid state Nd:LaF_3 laser. Using the high photon energy and short wavelength, photolithography, laser chemical vapor deposition (CVD), laser ablation, etc. are demonstrated.

Recently, an extensive effort has been made to improve the laser characteristics for the industrial application for laser lithography. Improvement in the high-repetition rate frequency, high average power, narrow linewidth, and operational long gas lifetime continues.

SEE ALSO THE FOLLOWING ARTICLES

LASERS • LASERS, COLOR CENTER • LASERS, GAS • LASERS, NUCLEAR PUMPED

BIBLIOGRAPHY

Bass, M., and Stitch, M. L. (1985). "Laser Handbook," Vol. 5, North-Holland, Amsterdam.

Midorikawa, K., Obara, M., and Fujioka, T. (1984). X-ray preionization of rare-gas-halide lasers, *IEEE J. Quantum Electron* **QE-20**(3), 198–205.

Ohwa, M., and Obara, M. (1986). Theoretical analysis of efficiency scaling laws for a self-sustained discharge-pumped XeCl laser. *J. Appl. Phys.* **59**(1), 32–34.

Rhodes, C. K., Egger, H., and Pummer, H. (1983). "Excimer Lasers-1983," American Institute of Physics, Vol. 100, New York.

Rhodes, C. K. (1984). "Excimer Lasers," 2nd ed., Springer-Verlag, Berlin and New York.

Stitch, M. L. (1979). "Laser Handbook," Vol. 3, North-Holland, Amsterdam.

Ohwa, M., and Obara, M. (1987). Theoretical evaluation of high-efficiency operation of discharge-pumped vacuum-ultraviolet F_2 laser, *Appl. Phys. Lett.* **51**(13), 958–960.

Kakehata, M., Hashimoto, E., Kannari, F., and Obara, M. (1990). Development of a high power vacuum ultraviolet F_2 laser (157 nm) excited by an intense electric discharge, "Proceedings of the International Conference on LASERS'90," 457–461.

Lasers, Free-Electron

H. P. FREUND
Science Applications International Corp.

R. K. PARKER
Naval Research Laboratory

I. Introduction
II. General Principles
III. Experiments and Applications
IV. Summary and Discussion

GLOSSARY

Amplifier A device used to increase the amplitude of some input signal. The term *superradiant amplifier* is often used to describe a free-electron laser that is configured to amplify noise (i.e., random fluctuations) in the beam rather than an externally supplied signal. In contrast to oscillator configurations, there is no reflection of the signal, and the superradiant amplifier is a single-pass device.

Beat wave A composite wave formed by the superposition of two waves having different angular frequencies (ω_1, ω_2) and wavenumbers (k_1, k_2). Beat waves form at the sum and difference frequencies ($\omega_1 \pm \omega_2$) and wavenumbers ($k_1 \pm k_2$), and are characterized by amplitudes proportional to the product of the amplitudes of the original waves.

Compton regime This denotes operation in a free-electron laser in which the dominant mechanism is stimulated Compton scattering. In this process, the wiggler field (which appears to be a backwards-propagating electromagnetic wave in the rest frame of the electrons) produces secondary electromagnetic waves by scattering off the electrons.

Electrostatic accelerators These include Van de Graaf and Cockcroft–Walton configurations. Van de Graaf accelerators accelerate charges by passing a moving belt through the coronal discharge from an array of points. The charge is then carried away by the belt to a field-free region where it can be extracted from the accelerator. The maximum voltages that can be achieved with electrostatic accelerators are in the range of from 10–30 MV.

Free-streaming The free and unrestrained propagation of particles without hindrance by external forces. This is often used in a context that denotes trajectories that are unbounded.

Gain A measure of the amplification of the input signal in an amplifier. The gain of an amplifier is often measured in decibels (dB), which is ten times the common logarithm of the ratio of the output power of the amplifier to the input drive power.

Linear induction accelerators (induction linacs) Linear accelerators that operate by inducing an electromotive force in a cavity through a rapid change in the magnetic field strength. In effect, the electron beam acts as the analog of the secondary winding in a transformer.

Larmor rotation The circular rotation of charged particles in a uniform magnetic field. Also called *cyclotron* rotation.

Master oscillator power amplifier (MOPA) A shorthand notation for a power amplifier in which the source of the input drive signal is an external, or *master*, oscillator.

Microtrons Cyclic accelerators in which the particles execute circular motion in a uniform magnetic field that carries the beam through an rf accelerating cavity, one in each cycle.

Modulators A pulsed voltage source that includes an energy storage element (such as a capacitor bank) and a switching system to discharge the energy through some load. See *Pulse line accelerators*.

Oscillator A device used to generate periodic (i.e., oscillatory) signals without the necessity of an external drive signal. Free-electron laser oscillators are constructed by the insertion of the wiggler–electron beam system within a reflecting cavity in which the signal makes many passes through the interaction region. As a result, the signal grows from noise and is amplified during each pass. As a result, the signal can grow to high intensities without the need for a strong input (driving) signal.

Phase space The multidimensional space formed by the position and momentum of a particle (or of an ensemble of particles).

Phase velocity The speed of propagation for a single wave of angular frequency ω and wavenumber k. The phase velocity is determined by the ratio ω/k and defines the speed of a point of constant phase of the wave.

Ponderomotive wave A slowly-varying wave formed by the beating of two waves.

Pulse line accelerators These draw an intense current from a diode by means of a transmission line from a high-voltage capacitor bank. The capacitor bank and transmission line form a distributed capacitance network. The diode acts as a load through which the energy is discharged. See *Modulators*.

Radio-frequency linear accelerators (rf linacs) A linear accelerator that employs radio frequency (rf) cavities for electron acceleration. The particles are accelerated in cylindrical cavities that require a high power source for the rf fields. The rf fields in these cavities may be either traveling or standing waves. In the case of traveling wave configurations, which are most often employed for electron accelerators, the phase velocity if the rf fields must be synchronized with the desired electron velocity.

Raman regime This denotes operation in a free-electron laser in which the dominant mechanism is a three-wave scattering process. It occurs when the electron beam density is high enough that the longitudinal (i.e., electrostatic) waves driven by the beam exert a greater force than the ponderomotive wave. In this process, as opposed to stimulated Compton scattering, the wiggler field scatters off the beam-driven longitudinal waves to produce the secondary electromagnetic waves.

SASE Self-amplified spontaneous emission.

Separatrix The line or surface in the phase space of a particle that distinguishes between two classes of trajectories as, for example, between bounded and unbounded motion.

Storage rings A toroidal configuration in which beams of electrons and/or positrons circulate for periods of the order of several hours. Short bunches of electrons and positrons are injected into the torus and guided around the ring by a system of bending and focusing magnets. Due to the curved trajectories of the beam, the particles lose energy during each circuit by means of synchrotron radiation. In order to compensate for this loss, an rf-accelerating cavity is included in the ring. The ring itself may be circular or polygonal; however, the straight sections in polygonal rings facilitate the insertion of the wiggler magnets required for free-electron lasers.

Wiggler The periodic magnet used in free-electron lasers to generate an undulatory motion in the electron beam. The term *wiggler* is specifically used for the magnets employed to generate coherent radiation in free-electron lasers, as opposed to the similar periodic magnets (referred to as *undulators*) employed in synchrotron light sources to generate spontaneous, or incoherent, radiation.

Undulator see *Wiggler*.

Untrapped trajectories Unbounded trajectories of particles that are under the influence of some external force. Examples include parabolic and hyperbolic trajectories of bodies subject to a central gravitational force or the unrestrained motion of a circular pendulum.

IN ITS FUNDAMENTAL concept, the free-electron laser is an extremely adaptable light source that can produce high-power coherent radiation across much of the electromagnetic spectrum. In contrast, gas and solid-state lasers generate light at well-defined wavelengths corresponding to discrete energy transitions within atoms or molecules in the lasing media. Dye lasers are tunable over a narrow spectral range but require a gas laser for optical pumping and operate at relatively low power levels. Further, while conventional lasers are typically characterized by energy conversion efficiencies of only a few percent, theoretical calculations indicate that the free-electron laser is capable of efficiencies as high as 65% in some spectral bands, and efficiencies of 40% have been demonstrated in the laboratory.

Applications of free-electron lasers to date range from experiments in solid-state physics to molecular biology. Several studies have examined the applicability of free-electron lasers to perform basic scientific research and have found wavelength ranges in which the free-electron laser has no effective competition (10 microns to several millimeters excepting several fixed points where existing lasers can operate). In addition, the free-electron laser is under consideration as the basis for the next generation of synchrotron light source operating in the X-ray region. Early free-electron lasers were built around available electron accelerators. However, free-electron lasers have begun to come into their own as accelerators are designed using laser-driven photo-injectors to produce high-quality electron beams, and user facilities are set up so that researchers in other disciplines can take advantage of this new source of intense and tunable light.

I. INTRODUCTION

The free-electron laser was first conceived almost five decades ago and has since operated over a spectrum ranging from microwaves through the ultraviolet. In a free-electron laser, high-energy electrons emit coherent radiation, as in a conventional laser, but the electrons travel through a vacuum instead of remaining in bound atomic states within the lasing medium. Because the electrons are free-streaming, the radiation wavelength is not constrained by a particular transition between two discrete energy levels. In quantum mechanical terms, the electrons radiate by transitions between energy levels in the continuum and, therefore, radiation is possible over a much larger range of frequencies than is found in a conventional laser. However, the process can be described by classical electromagnetic theory alone.

The radiation is produced by an interaction among three elements: the electron beam, an electromagnetic wave traveling in the same direction as the electrons, and an undulatory magnetic field produced by an assembly of magnets known as a *wiggler* or *undulator*. A schematic illustration of these three elements is shown in Fig. 1 for a planar wiggler geometry. The wiggler magnetic field acts on the electrons in such a way that they acquire an undulatory motion. The acceleration associated with this curvilinear trajectory is what makes radiation possible. In this process, the electrons lose energy to the electromagnetic wave that is amplified and emitted by the laser. The tunability of the free-electron laser arises because the wavelength of light required for the interaction between these three elements is determined by both the periodicity of the wiggler field and the electron beam energy.

Although the basic principle underlying the free-electron laser is relatively simple, the practical application of the concept can be difficult. In 1951, Hans Motz of Stanford University first calculated the emission spectrum from an electron beam in an undulatory magnetic field. At the time, coherent optical emission was not expected due to the difficulty of bunching the electron beam at short wavelengths; however, it was recognized that microwave amplification through stimulated emission of radiation (maser) operation was possible. Experiments performed by Motz and co-workers shortly thereafter produced both incoherent radiation in the blue-green part of the spectrum and coherent emission at millimeter wavelengths. The free-electron maser was invented independently by Robert Phillips in 1957 in search of higher power from microwave tubes. The term *ubitron* was coined at this time as an acronym for undulating beam interaction. Over the succeeding 7 years, Phillips performed an extensive study of the interaction and pioneered many innovative design concepts in use today. Whereas the original microwave

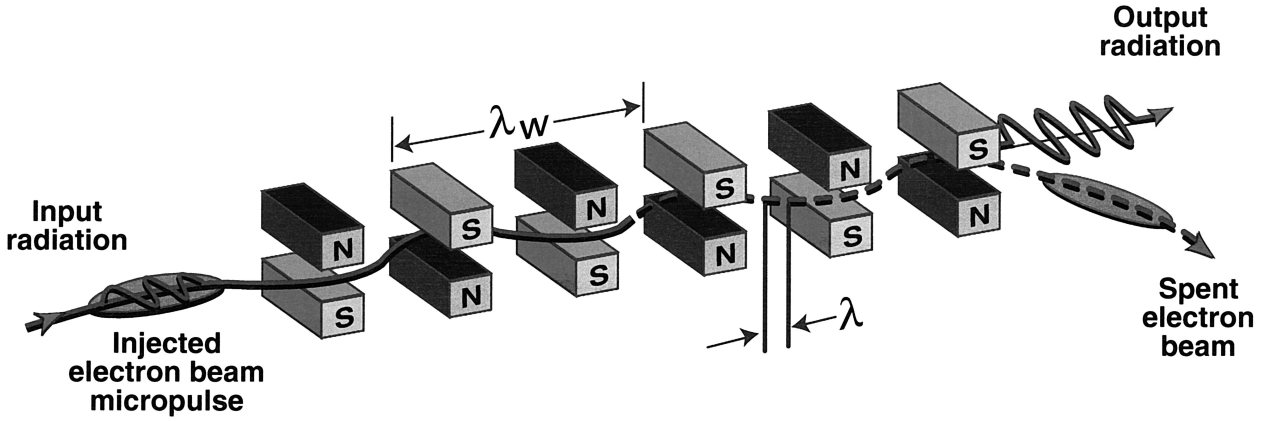


FIGURE 1 Schematic illustration of the interaction between the beam and the wiggler in an free-electron laser with a planar wiggler. [From H. P. Freund and G. R. Neil, "Free-Electron Lasers: Vacuum Electronic Generators of Coherent Radiation," Proc. IEEE, vol. 87, pp. 782–803, 1999].

experiment at Stanford observed an output of 1–10 watts, Phillips achieved 150 kW at a 5-mm wavelength. However, the full potential of the free-electron laser was unrecognized, and the ubitron program was terminated in 1964 due to a general shift in interst from vacuum electronics to solid-state physics and quantum electronics.

The resurgence of interest in the concept that began in the mid-1970s followed two paths. The term *free-electron laser* was coined in 1975 by John Madey to describe an experiment at Stanford University that used an electron beam from a radio-frequency linear accelerator (rf linac). This experiment produced stimulated emission in the infrared spectrum at a wavelength of 10.6 μm using an electron beam from a radio-frequency linear accelerator (rf linac).

In parallel with the work at Stanford, experimenters at several laboratories began work on microwave free-electron lasers, successors to the ubitron. Those projects were located at the Naval Research Laboratory, Columbia University, the Massachusetts Institute of Technology, Lawrence Livermore National Laboratory, TRW, and the École Polytéchnique in France. They differed from the original work by Phillips by using intense relativistic electron beams with currents of the order of a kiloamperes and voltages in excess of a megavolt. The principal goal of these efforts was the production of high absolute powers. The results ranged from a peak power of the order of 2 MW at a wavelength of 2.5 mm at Columbia, through 70 MW at a 4-mm wavelength at the Naval Research Laboratory, to a maximum power figure of 1 GW obtained by Livermore at an 8-mm wavelength. This latter result represents an efficiency (defined as the ratio of the output radiation power to the initial electron beam power) of 35%, and was made possible by the use of a nonuniform wiggler field.

Since that time numerous free-electron lasers have been built to operate over spectral ranges extending from the infrared through the ultraviolet, and the free-electron laser has made the transition from that of laboratory curiosity to experimental user facility. A list of the presently available user facilities is given in Table I. Observe that the rf linac predominates, but that storage rings and electrostatic accelerators are also in use. In addition, a high average-power free-electron maser based on an electrostatic accelerator is under development at FOM in the Netherlands for use in heating experimental plasma fusion reactors.

More generally, free-electron lasers have been constructed over the entire electromagnetic spectrum. This spectral range is summarized in Fig. 2, in which we plot the peak power of a sample of conventional lasers, microwave tubes, and free-electron lasers as a function of wavelength. At wavelengths above 0.1 mm, free-electron lasers already either match or exceed power levels obtainable from conventional technology. At shorter wavelengths, conventional lasers can be found with higher powers than is currently available from free-electron lasers. However, free-electron laser technology is rapidly maturing, and this situation is likely to change in the future.

There are two prime goals for free-electron laser development at the present time: high average powers and ultrashort wavelengths. High average power work is exemplified by the development of a free-electron laser at

TABLE I Summary of Currently Available FEL User Facilities

Country	Institution[a]	λ (μm)	τ_p (ps)	E_b/I_b (MeV/A)	P_{peak} (MW)	P_{avg} (W)	Accelerator
USA	Stanford University	19–65	1–5	20/14	.3	.4	rf linac
		3–10	0.7	37/10	10	1.2	rf linac
	Vanderbilt University	2–10	2	43/50	10	10	rf linac
	Duke University	3–10	3	44/20	2	3	rf linac
		0.3	10	1000/350	1000	0.1	Storage ring
	Jefferson Accelerator Facility	3–6.6	1–2	42/50	17	1700	rf linac
	UCSB	150–2000	6×10^6	6/2	0.004	0.08	Van de Graf
Japan	FELI	5–22	1.7	33/42	5	2	rf linac
		1–6	1.7	75/50	5	0.5	rf linac
		0.23–1.2	1.7	165/60	5	0.5	rf linac
		20–80	1.7	30/40	5	1	rf linac
		40–100	3–5	20/40	(2)	(1)	rf linac
France	LURE	0.3	20	800/10	12	0.8	Storage ring
		1.8–17.5	1.5–6	70/80	10	9	rf linac
Netherlands	FOM	5–35	0.5–10	45/70	2	1	rf linac
		25–110	1–10	25/70	2	0.5	rf linac
China	Beijing	10–16	4	30/14	20	2	rf linac

[a] UCSB, University of California, Santa Barbara; FELI, LURE, ; FOM.

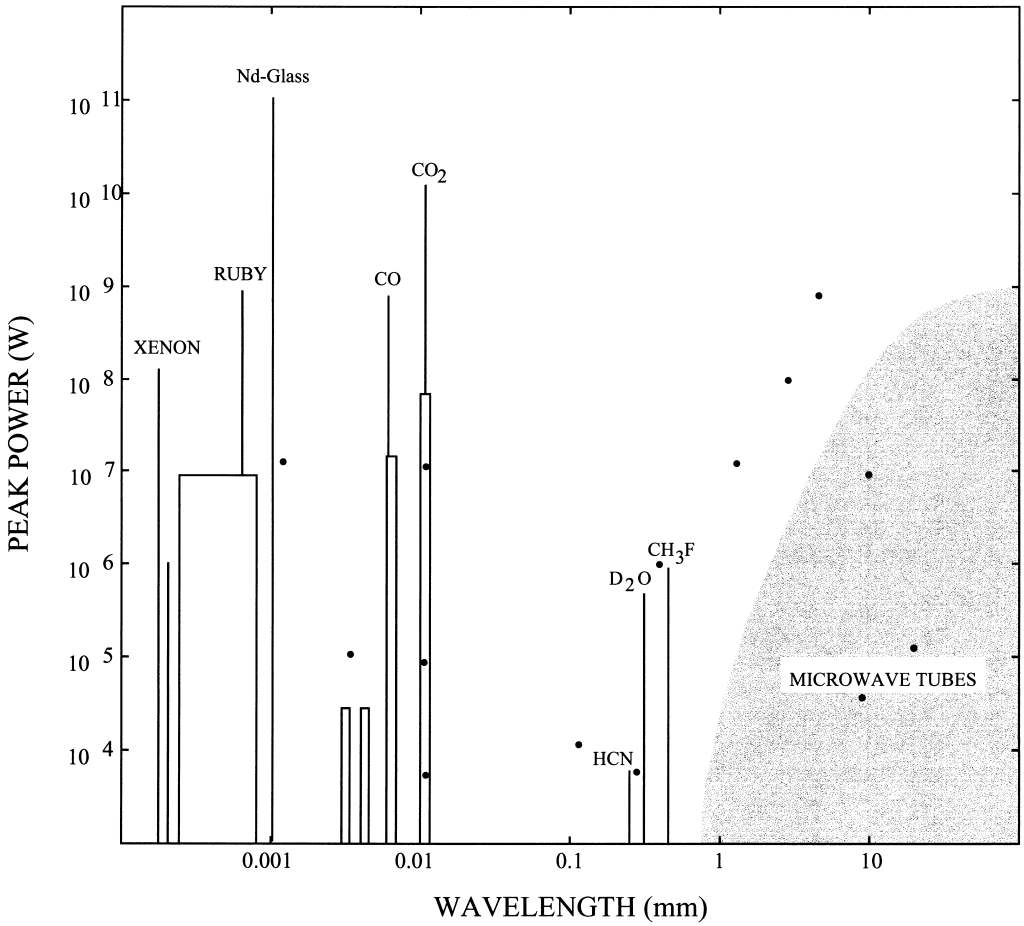


FIGURE 2 An important criterion in the judgment of radiation sources is the peak power available at specific wavelengths. A comparison of free-electron lasers, represented by dots, with conventional lasers and microwave sources is shown.

the Thomas Jefferson National Accelerator Facility (see Table I). This device is based upon a superconducting rf linac with energy recovery, and it has achieved the highest average power recorded thus far from a free-electron laser: 1.7 kW at a wavelength of 3 μm. The short-wavelength effort is directed toward the design of a fourth-generation light source to replace the incoherent synchrotron light sources currently in use for the production of intense X-rays. The design work on such a source is international in scope and includes collaborations between the Stanford Linear Accelerator Center, Argonne National Laboratory, Brookhaven National Laboratory, and the Lawrence Berkeley and Livermore National Laboratories in the United States and at the TESLA Test Facility at the Deutsches Elektronen-Synchrotron in Germany.

II. GENERAL PRINCIPLES

An electron beam that traverses an undulatory magnetic field emits incoherent radiation. Indeed, this is the mechanism employed in synchrotron light sources. In conventional terminology, the periodic magnetic field in synchrotron light sources is referred to as an *undulator*, whereas that used in free-electron lasers is called a *wiggler*, although there is no fundamental difference between them. It is necessary for the electron beam to form coherent bunches in order to give rise to the stimulated emission required for a free-electron laser. This can occur when a light wave traverses an undulatory magnetic field such as a wiggler because the spatial variations of the wiggler and the electromagnetic wave combine to produce a beat wave, which is essentially an interference pattern. It is the interaction between the electrons and this beat wave that gives rise to the stimulated emission in free-electron lasers.

This beat wave has the same frequency as the light wave, but its wavenumber is the sum of the wavenumbers of the electromagnetic and wiggler fields. With the same frequency, but a larger wavenumber (and thus a shorter wavelength), the beat wave travels more slowly than the light wave; for this reason it is called a *ponderomotive*

wave. Since the ponderomotive wave is the combination of the light wave and the stationary (or magnetostatic) field of the wiggler, it is the effective field experienced by an electron as it passes through the free-electron laser. In addition, since the ponderomotive wave propagates at less than the speed of light *in vacuo* it can be in synchronism with the electrons that are limited by that velocity. Electrons moving in synchronism with the wave are said to be in *resonance* with it and will experience a constant field—that of the portion of the wave with which it is traveling. In such cases, the interaction between the electrons and the ponderomotive wave can be extremely strong.

A good analogy to the interaction between the electrons and the ponderomotive wave is that of a group of surfers and a wave approaching a beach. If the surfers remain stationary in the water, the velocity difference between the wave and the surfers is large, and an incoming wave will merely lift them up and down briefly and then return them to their previous level. There is no bulk, or average, translational motion or exchange of energy between the surfers and the wave. But if the surfers "catch the wave" by paddling so as to match the speed of the wave, then they can gain significant momentum from the wave and be carried inshore. This is the physical basis underlying the resonant interaction in a free-electron laser. However, in a free-electron laser, the electrons amplify the wave, so the situation is more analogous to the surfers "pushing" on the wave and increasing its amplitude.

The frequency of the electromagnetic wave required for this resonant interaction can be determined by matching the velocities of the ponderomotive wave and the electron beam. This is referred to as the phase-matching condition. The interaction is one in which an electromagnetic wave characterized by an angular frequency ω and wavenumber k and the magnetostatic wiggler with a wavenumber k_w produce a beat wave with the same frequency as the electromagnetic wave, but a wavenumber equal to the sum of the wavenumbers of the wiggler and electromagnetic waves (i.e., $k + k_w$). The velocity of the ponderomotive wave is given by the ratio of the frequency of the wave to its wavenumber. As a result, matching this velocity to that of the electron beam gives the resonance condition in a free-electron laser

$$\frac{\omega}{k + k_w} \cong v_z$$

for a beam with a bulk streaming velocity v_z in the z-direction. The z-direction is used throughout to denote both the bulk-streaming direction of the electron beam and the symmetry axis of the wiggler field. The dispersion relation between the frequency and wavenumber for waves propagating in free space is $\omega \cong ck$, where c denotes the speed of light *in vacuo*. Combination of the free-space dispersion relation and the free-electron laser resonance condition gives the standard relation for the wavelength as a function of both the electron beam energy and the wiggler period

$$\lambda \cong \frac{\lambda_w}{2\gamma_z^2},$$

where $\gamma_z = (1 - v_z^2/c^2)^{-1/2}$ is the relativistic time dilation factor, which is related to the electron streaming energy, and $\lambda_w = 2\pi/k_w$ is the wiggler wavelength. The wavelength, therefore, is directly proportional to the wiggler period and inversely proportional to the square of the streaming energy. This results in a broad tunability that permits the free-electron laser to operate across virtually the entire electromagnetic spectrum.

How does a magnetostatic wiggler and a forward-propagating electromagnetic wave, both of whose electric and magnetic fields are directed transversely to the direction of propagation, give rise to an axial ponderomotive force that can extract energy from the electron beam? The wiggler is the predominant influence on the electron's motion. In order to understand the dynamical relationships between the electrons and the fields, consider the motion of an electron subject to a helically symmetric wiggler field. An electron propagating through a magnetic field experiences a force that acts at right angles to both the direction of the field and to its own velocity. The wiggler field is directed transversely to the direction of bulk motion of the electron beam and rotates through 360° in one wiggler period. An electron streaming in the axial direction, therefore, experiences a transverse force and acquires a transverse velocity component upon entry into the wiggler. The resulting trajectory is helical and describes a bulk streaming along the axis of symmetry as well as a transverse circular rotation that lags 180° behind the phase of the wiggler field. The magnitude of the transverse wiggle velocity, denoted by v_w, is proportional to the product of the wiggler amplitude and period. This relationship may be expressed in the form

$$\frac{v_w}{c} \cong 0.934 \frac{B_w \lambda_w}{\gamma_b},$$

where the wiggler period is expressed in units of centimeters, B_w denotes the wiggler amplitude in Tesla, and

$$\gamma_b = 1 + \frac{E_b}{m_e c^2},$$

denotes the relativistic time dilation factor associated with the total kinetic energy E_b of the electron beam (where m_e denotes the rest mass of the electron, and $m_e c^2$ denotes the electron rest energy).

Since the motion is circular, both axial and transverse velocities have a constant magnitude. This is important

because the resonant interaction depends upon the axial velocity of the beam. In addition, since the wiggler induces a constant-magnitude transverse velocity, the relation between the total electron energy and the streaming energy can be expressed in terms of the time dilation factors in the form

$$\gamma_z \cong \frac{\gamma_b}{\sqrt{1 + 0.872 B_w^2 \lambda_w^2}}.$$

As a result, the resonant wavelength depends upon the total beam energy, and the wiggler amplitude and period through.

$$\lambda \cong \left(1 + 0.872 B_w^2 \lambda_w^2\right) \frac{\lambda_w}{2\gamma_b^2}.$$

It is the interaction between the transverse wiggler-induced velocity with the transverse magnetic field of an electromagnetic wave that induces a force normal to both in the axial direction. This is the ponderomotive force. The transverse velocity and the radiation magnetic field are directed at right angles to each other and undergo a simple rotation about the axis of symmetry. A resonant wave must be circularly polarized with a polarization vector that is normal to both the transverse velocity and the wiggler field and which rotates in synchronism with the electrons. This synchronism is illustrated in Fig. 3, and is maintained by the aforementioned resonance condition.

In order to understand the energy transfer, we return to the surfer analogy and consider a group of surfers attempting to catch a series of waves. In the attempt to match

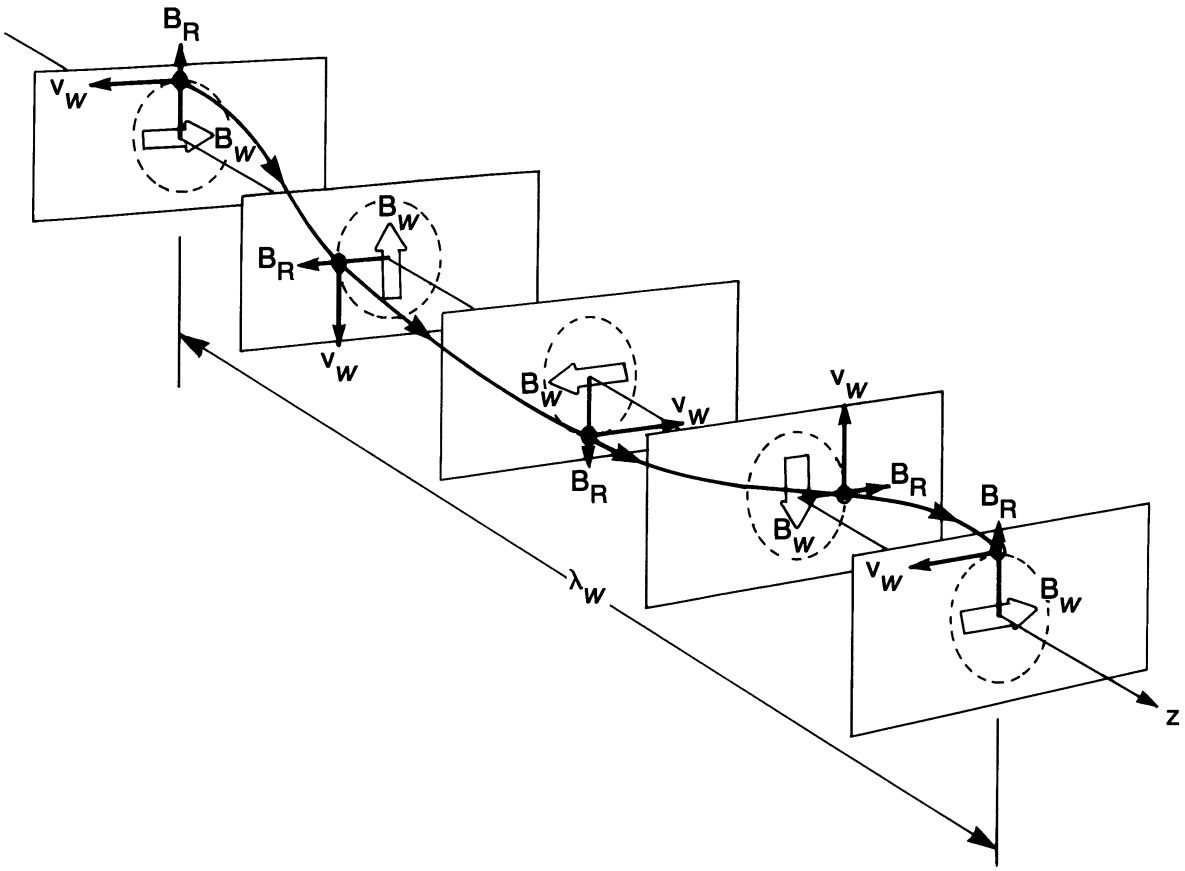


FIGURE 3 The electron trajectory in a helical wiggler includes bulk streaming parallel to the axis of symmetry as well as a helical gyration. The vector relationships between the wiggler field $\mathbf{B}_w$, the transverse velocity $\mathbf{v}_w$, and the radiation field $\mathbf{B}_R$ of a resonant wave are shown in the figure projected onto planes transverse to the symmetry axis at intervals of one-quarter of a wiggler period. This projection of the orbit is circular, and the transverse velocity is directed opposite to that of the wiggler. A resonant wave must be circularly polarized, with a polarization vector that is normal to both the transverse velocity and the wiggler field and that rotates in synchronism with the electrons. The electrons then experience a slowly varying wave amplitude. In addition, the transverse velocity and the radiation magnetic field are directed at right angles to each other and undergo a simple rotation about the axis of symmetry. The interaction between the transverse velocity and the radiation field induces a force in the direction normal to both, which coincides with the symmetry axis.

velocities with the waves, some will catch a wave ahead of the crest and slide forward while others will catch a wave behind the crest and slide backward. As a result, clumps of surfers will collect in the troughs of the waves. Those surfers who slide forward ahead of the wave are accelerated and gain energy at the expense of the wave, while those that slide backward are decelerated and lose energy to the wave. The wave grows if more surfers are decelerated than accelerated, and there is a net transfer of energy to the wave. The free-electron laser operates by an analogous process. Electrons in near resonance with the ponderomotive wave lose energy to the wave if their velocity is slightly greater than the phase velocity of the wave, and gain energy at the expense of the wave in the opposite case. As a result, wave amplification occurs if the wave lags behind the electron beam.

This process in a free-electron laser is described by a nonlinear pendulum equation. The *ponderomotive phase* $\psi\,[=(k+k_w)z-\omega t]$ is a measure of the position of an electron in both space and time with respect to the ponderomotive wave. The ponderomotive phase satisfies the circular pendulum equation

$$\frac{d^2}{dz^2}\psi = K^2 \sin\psi,$$

where the pendulum constant is proportional to the square root of the product of the wiggler and radiation fields

$$K \approx 8.29\frac{\sqrt{B_w B_R}}{\gamma_b}.$$

Here K is expressed in units of inverse centimeters and the magnetic fields are expressed in Tesla. The pendulum equation can be reduced to

$$\frac{1}{2}\left(\frac{d\psi}{dz}\right)^2 + U(\psi) = H,$$

where H has the form of the Hamiltonian or total energy of the system, and

$$U(\psi) = K^2 \cos\psi,$$

is the ponderomotive potential. The electron trajectories through the wiggler, therefore, may be expressed as

$$\frac{d\psi}{dz} = \pm\sqrt{2H - 2K^2\cos\psi}.$$

Observe that the first derivative of the phase [i.e., $d\psi/dz = (k+k_w)-\omega/v_z$] is a measure of the electron-streaming velocity; hence, this equation effectively describes the electron velocity as a function of the phase of the ponderomotive wave.

There are two classes of trajectory: trapped and untrapped. The *separatrix* describes the transition between trapped and untrapped orbits, and occurs when $H = K^2$. Hence, the separatrix is defined by a pair of curves in the phase space for $(d\psi/dz, \psi)$

$$\frac{d\psi}{dz} = \pm 2K \sin(\psi/2).$$

Free-streaming, untrapped orbits are characterized by $H > K^2$ and occupies that region of the phase space outside of the separatrix,

$$\frac{d\psi}{dz} > |2K \sin(\psi/2)|.$$

The trapped orbits are those for which $H < K^2$ within the bounds of the separatrix. The free-streaming orbits correspond to the case in which the pendulum swings through the full 360° cycle. The electrons pass over the crests of many waves traveling fastest at the bottom of the troughs and slowest at the crests of the ponderomotive wave. In contrast, the electrons are confined within the trough of a single wave in the trapped orbits. This corresponds to the motion of a pendulum that does not rotate full circle, but is confined to oscillate about the lower equilibrium point. The dynamical process is illustrated in Fig. 4, and is one in which the pendulum constant evolves during the course of the interaction. Electrons lose energy as the wave is amplified; hence, the electrons decelerate and both the pendulum constant and separatrix grow. Ultimately, the electrons cross the growing separatrix from untrapped to trapped orbits.

In typical operation, electrons entering the free-electron laser are free-streaming on untrapped trajectories. The gain in power during this phase of the interaction increases as the cube of the distance z in the case of relatively low current operation in which the total gain is less than unity. This case is often referred to as the low-gain Compton regime, and the power gain in decibels is given by

$$G(\text{dB}) \approx 1.09\frac{\omega_b^2}{c^2 k_w^2}\frac{v_w^2}{c^2}(k_w z)^3 F(\theta),$$

where

$$\omega_b = \sqrt{\frac{4\pi e^2 n_b}{\gamma_b m_e}}$$

is the so-called *electron plasma frequency*. Here e represents the charge of the electron, and n_b denotes the bulk density of the electron beam. The spectral shape of this gain is determined by

$$F(\theta) = \frac{d}{d\theta}\left(\frac{\sin\theta}{\theta}\right)^2,$$

where $\theta = (\omega/v_z - k - k_w)z/2$. This spectral function is shown in Fig. 4 and exhibits an maximum at $\theta \approx -1.3$

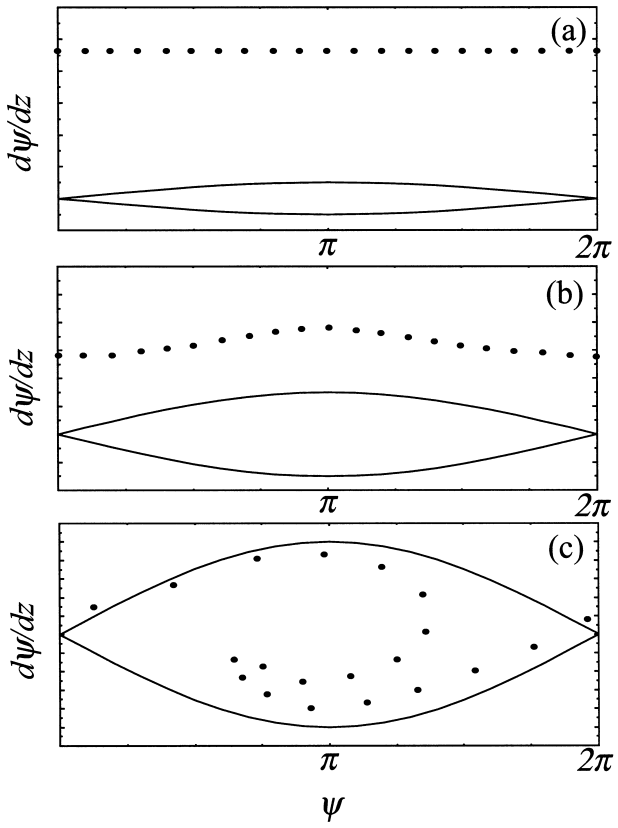


FIGURE 4 The evolution of the phase space distribution of an electron beam is illustrated schematically for three general stages of the interaction. In the initial state (a), a monoenergetic electron population is located far from the separatrix. As the interaction progresses (b), the electrons lose energy to the wave and the separatrix grows. Observe that the electrons are still on untrapped trajectories; however, the coherent bunching mechanism characteristic of the interaction has begun. In the final stage (c), the bulk of electrons have crossed the separatrix and become trapped. Electron trajectories after this point circulate about within the separatrix, alternately gaining from and losing energy to the wave. Saturation occurs when the net rate of energy transfer vanishes. Observe that all electrons do not become trapped.

for which $F \approx 0.54$; hence, the peak gain occurs at a wavelength

$$\lambda \approx \frac{\lambda_w}{2\gamma_z^2\left(1 + \dfrac{2.6}{k_w z}\right)}.$$

This regime is relevant to operation at short wavelengths in the infrared and optical spectra. These experiments typically employ electron beams generated by radio-frequency linear accelerators, microtrons, storage rings, and electrostatic accelerators in which the total current is small.

In contrast, experiments operating at microwave and millimeter wavelengths often employ high current accelerators, and the wave amplification is exponential. In these cases, two distinct regimes are found. The high-gain Compton (sometimes called the *strong-pump*) regime is found when

$$\frac{\omega_b}{ck_w} \ll \frac{1}{16}\gamma_z^2\left(\frac{v_w}{c}\right)^2.$$

In this regime, the maximum gain in the signal power over a wiggler period is given by

$$G(\text{dB}) \approx 37.5\left(\frac{v_w^2}{c^2}\frac{\omega_b^2}{c^2k_w^2}\right)^{1/3},$$

and is found at the resonant wavelength.

The opposite limit, referred to as the collective Raman regime, is fundamentally different from either the high or low-gain Compton regimes. It occurs when the current density of the beam is high enough that the space–charge force exceeds that exerted by the ponderomotive wave. The gain in power over a wiggler period in this regime varies as

$$G(\text{dB}) \approx 27.3\frac{v_w}{c}\sqrt{\gamma_z\frac{\omega_b}{ck_w}}.$$

In this regime, the space-charge forces result in electrostatic waves that copropagate with the beam and are characterized by the dispersion relations

$$\omega = k_{sc}v_z \pm \frac{\omega_b}{\gamma_z},$$

which describe the relation between the frequency ω and wavenumber k_{sc}. These dispersion relations describe positive- and negative-energy waves corresponding to the "+" and "−" signs respectively. The interaction results from a stimulated three-wave scattering process. This is best visualized from the perspective of the electrons, in which the wiggler field appears to be a backwards-propagating electromagnetic wave called a *pump* wave. This pump wave can scatter off the negative-energy electrostatic wave (the *idler*) to produce a forward-propagating electromagnetic wave (the *signal*). The interaction occurs when the wavenumbers of the pump, idler, and signal satisfy the condition $k_{sc} = k + k_w$, which causes a shift in the wavelength of the signal to

$$\lambda \approx \frac{\lambda_w}{2\gamma_z^2\left(1 - \dfrac{\omega_b}{\gamma_z ck_w}\right)}.$$

Observe that the interaction in the Raman regime is shifted to a somewhat longer wavelength than occurs in the high-gain Compton regime.

Wave amplification can saturate by several processes. The highest efficiency occurs when the electrons are trapped in the ever-deepening ponderomotive wave and undergo oscillations within the troughs. In essence, the

electrons are initially free-streaming over the crests of the ponderomotive wave. Since they are traveling at a velocity faster than the wave speed, they come upon the wave crests from behind. However, the ponderomotive wave grows together with the radiation field, and the electrons ultimately will come upon a wave that is too high to cross. When this happens, they rebound and become trapped within the trough of the wave. In analogy to the oscillation of a pendulum, the trapped electrons lose energy as they rise and gain energy as they fall toward the bottom of the trough. As a result, the energy transfer between the wave and the electrons is cyclic and the wave amplitude ceases to grow and oscillates with the electron motion in the trough. The ultimate saturation efficiency for this mechanism can be estimated from the requirement that the net change in electron velocity at saturation is equal to twice the velocity difference between the electron beam and the ponderomotive wave. This results in a saturation efficiency of

$$\eta \approx \frac{\gamma_b}{2(\gamma_b - 1)} \left(\frac{v_w^2}{2c^2} \frac{\omega_b^2}{c^2 k_w^2} \right)^{1/3}$$

in the high-gain Compton regime, and

$$\eta \approx \frac{\gamma_b}{\gamma_b - 1} \frac{\omega_b}{\gamma_z c k_w}$$

in the collective Raman regime.

However, this process places stringent requirements on the quality of the electron beam. The preceding formulae apply to the idealized case of a monoenergetic (or cold) beam. This represents a theoretical maximum for the gain and efficiency since each electron has the same axial velocity and interacts with the wave in an identical manner. A monoenergetic beam is physically unrealizable, however, and all beams exhibit a velocity spread that determines a characteristic temperature. Electrons with axial velocities different from the optimal resonant velocity are unable to participate fully in the interaction. If this axial velocity spread is sufficiently large that the entire beam cannot be in simultaneous resonance with the wave, then the fraction of the electron beam that becomes trapped must fall. Ultimately, the trapping fraction falls to the point where the trapping mechanism becomes ineffective, and saturation occurs through the thermalization of the beam. Thus, there are two distinct operating regimes: the *cold* beam limit characterized by a narrow bandwidth and relatively high efficiencies, and the *thermal* regime characterized by a relatively broader bandwidth and sharply lower efficiencies.

The question of electron beam quality is the most important single issue facing the development of the free-electron laser at the present time. In order to operate in the cold beam regime, the axial velocity spread of the beam must be small. It is convenient to relate the axial velocity spread to an energy spread to obtain an invariant measure of the beam quality suitable for a wide range of electron beams. In the case of the low-gain limit, this constraint on the beam thermal spread is

$$\frac{\Delta E_b}{E_b} \ll \frac{1}{N_w},$$

where ΔE_b represents the beam thermal spread, and N_w is the number of periods in the wiggler. In the high-gain regimes, the maximum permissible energy spread for saturation by particle trapping is determined by the depth of the ponderomotive or space-charge waves, which is measured by twice the difference between the streaming velocity of the beam and the wave speed (Fig. 5). The maximum permissible thermal spread corresponds to this velocity difference, and is one-half the saturation efficiency for either the high-gain Compton or collective Raman regimes, in other words,

$$\frac{\Delta E_b}{E_b} \ll \frac{1}{2}\eta.$$

Typically, this energy spread must be approximately 1% or less of the total beam energy for the trapping mechanism to operate at millimeter wavelengths and decreases approximately with the radiation wavelength. Hence, the requirement on beam quality becomes more restrictive at shorter wavelengths and places greater emphasis on accelerator design.

Two related quantities often used as measures of beam quality are the *emittance* and *brightness*. The emittance measures the collimation of the electron beam and may be defined in terms of the product of the beam radius and

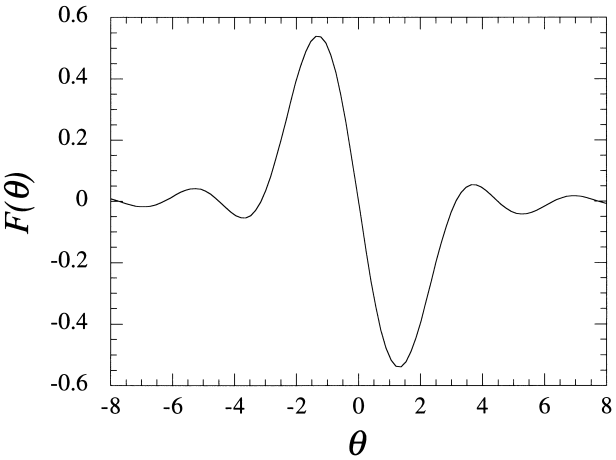


FIGURE 5 The spectral function measures the strength of the interaction as a function of the detuning of the wave from resonance with the electron beam. The maximum gain is found at $\theta \approx -1.3$. In contrast, the wave is damped for $\theta \approx 1.3$.

the average pitch angle (i.e., the angle between the velocity and the symmetry axis). It describes a random pitch angle distribution of the beam which, when the velocities are projected onto the symmetry axis, is equivalent to an axial velocity spread. In general, therefore, even a monoenergetic beam with a nonvanishing emittance displays an axial velocity spread. The electron beam brightness is an analog of the brightness of optical beams, and is directly proportional to the current and inversely proportional to the square of the emittance. As such, it describes the average current density per unit pitch angle, and measures both the beam intensity and the degree of collimation of the electron trajectories. Since the gain and efficiency increases with increasing beam current for fixed emittance, the brightness is a complementary measure of the beam quality. While it is important to minimize the emittance and maximize the brightness in order to optimize performance, both of these measures relate to the free-electron laser only insofar as they describe the axial velocity spread of the beam.

Free-electron laser efficiencies can reach approximately 12%, however, significant enhancements are possible when either the wiggler amplitude or period are systematically tapered. The free-electron laser amplifier at Livermore, which achieved a 35% extraction efficiency, employed a wiggler with an amplitude that decreased along the axis of symmetry, and contrasts with an observed efficiency of about 6% in the case of a uniform wiggler. The use of a tapered wiggler was pioneered by Phillips in 1960. The technique has received intensive study of late, and tapered wiggler designs have also been shown to be effective at infrared wavelengths in experiments at Los Alamos National Laboratory and using a superconducting rf linac at Stanford University.

The effect of a tapered wiggler is to alter both the transverse and axial velocities. Since the transverse velocity is directly proportional to the product of the amplitude and period, the effect of gradually decreasing either of these quantities is to decrease the transverse velocity and, in turn, increase the axial velocity. The energy extracted during the interaction results in an axial deceleration that drives the beam out of resonance with the wave; hence, efficiency enhancement occurs because the tapered wiggler maintains a relatively constant axial velocity (and phase relationship between the electrons and the wave) over an extended interaction length. The pendulum equation is modified in the presence of a tapered wiggler and has the form

$$\frac{d^2}{dz^2}\psi = K^2(\sin\psi - \sin\psi_{res}),$$

where the resonant phase ψ_{res} is determined by the wiggler taper. The wiggler taper can be accomplished either through the amplitude or period. However, it is technically simpler to taper the amplitude, and most tapered-wiggler experiments have employed this approach. In this case, the resonant phase varies as

$$\sin\psi_{res} = \left(\frac{v_w}{v_z}\right)^2 \frac{k + k_w}{K^2}\frac{1}{B_w}\frac{d}{dz}B_w.$$

Integration of the tapered-wiggler pendulum equation results in an equation similar to that found in the case of a uniform wiggler, with a ponderomotive potential

$$U(\psi) = K^2(\cos\psi + \psi\sin\psi_{res}).$$

If the wiggler amplitude is a decreasing function of axial position, then $\sin\psi_{res} < 0$, and the average potential decreases linearly with the ponderomotive phase. The difference between the ponderomotive potential for a uniform and tapered wiggler is illustrated in Fig. 6. As a result, the motion is similar to that of a ball rolling down a *bumpy* hill that accelerates as it falls.

The enhancement in the tapered-wiggler interaction efficiency is proportional to the decrement in the wiggler field of ΔB_w, and satisfies

$$\Delta\eta \approx \frac{0.872 B_w^2\lambda_w^2}{1 + 0.872 B_w^2\lambda_w^2}\frac{\Delta B_w}{B_w}.$$

In practice, a tapered wiggler is effective only after the bulk of the beam has become trapped in the ponderomotive wave. In single-pass amplifier configurations, therefore, the taper is not begun until the signal has reached saturation in a section of uniform wiggler, and the total extraction efficiency is the sum of the uniform wiggler efficiency and the tapered wiggler increment. Numerical simulations indicate that total efficiencies as high as 65% are possible under the right conditions.

Once particles have been trapped in the ponderomotive wave and begin executing a bounce motion between the troughs of the wave, then the potential exists for exciting secondary emission referred to as sideband waves. These sidebands arise from the beating of the primary signal with the ponderomotive bounce motion. This bounce motion is at the pendulum period, which depends weakly upon the depth of the ponderomotive well and amplitude of the bounce motion. For deeply trapped particles oscillating near the bottom of the well, the pendulum equation can be approximated as a harmonic oscillator with a wavenumber equal to the pendulum constant K. The sideband wavenumbers are shifted upward and downward from the radiation wavenumber by this value (i.e., $k_\pm = k \pm K$, where $k_\pm$ denotes the sideband wavenumber). Note that the bounce period is, typically, much longer than the radiation wavelength (i.e., $K \ll k$), and these sidebands are found at wavelengths close to that of the primary signal.

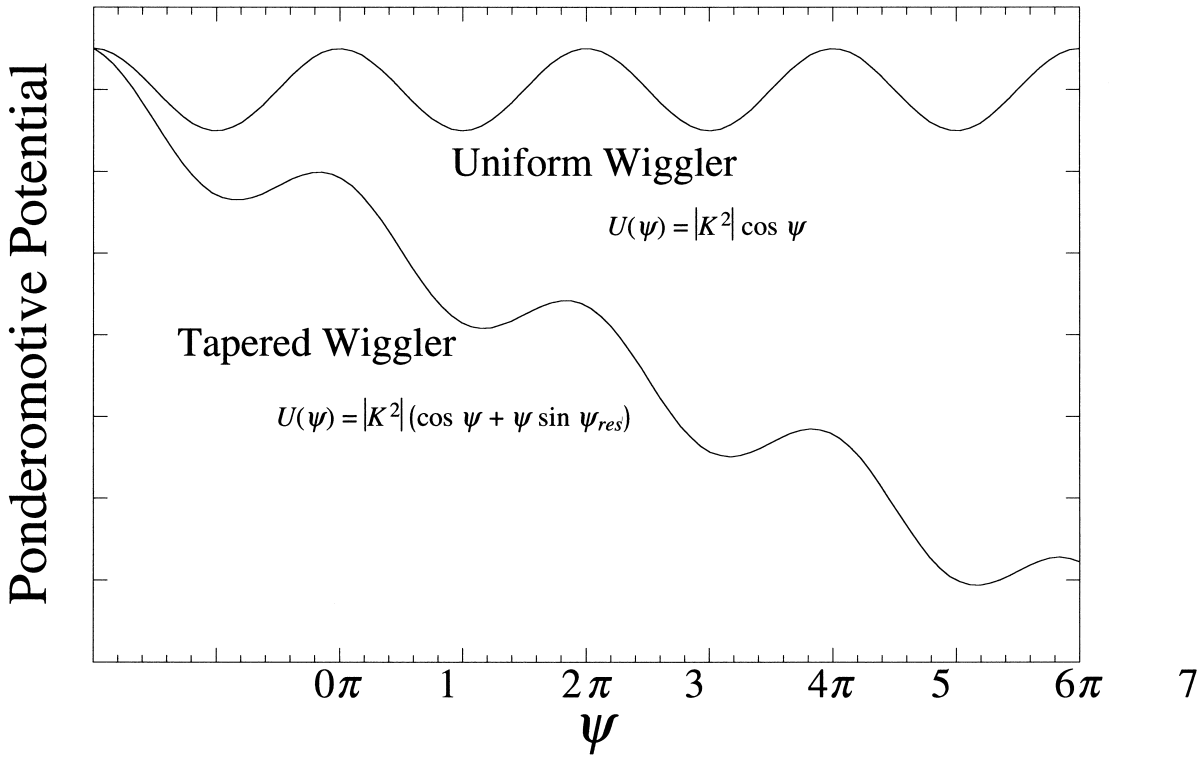


FIGURE 6 A comparison of the ponderomotive potentials for uniform and tapered wigglers. The extraction efficiency of the free-electron laser can be enhanced by introducing a taper in the wiggler field. In particular, a gradual reduction in either the amplitude or period of the wiggler will reduce the magnitude of the transverse velocity and accelerate the electrons in the axial direction. This is illustrated by the effect of the taper, for $\sin \Psi_{res} < 0$, on the ponderomotive potential. In such a potential, an electron behaves in a manner analogous to a ball rolling down a *bumpy hill*, which accelerates as it falls.

The difficulty imposed by the presence of sidebands is that they may compete with and drain energy from the primary signal. This is particularly crucial in long tapered wiggler systems that are designed to trap the beam at an early stage of the wiggler, and then extract a great deal more energy from the beam over an extended interaction length. In these systems, unrestrained sideband growth can be an important limiting factor. As a result, a great deal of effort has been expended on techniques of sideband suppression. One method of sideband suppression was employed in a free-electron laser oscillator at Columbia University. This experiment operated at a 2-mm wavelength in which the dispersion due to the waveguide significantly affected the resonance condition. As a consequence, it was found to be possible by proper choice of the size of the waveguide to shift the sideband frequencies out of resonance with the beam. Experiments on an infrared free-electron laser oscillator at Los Alamos National Laboratory indicate that it is also possible to suppress sidebands by (1) using a Littrow grating to deflect the sidebands out of the optical cavity or (2) changing the cavity length.

The foregoing description of the principles and theory of the free-electron laser is, necessarily, restricted to the idealized case in which the transverse inhomogeneities of both the electron beam and wiggler field are unimportant. This is sufficient for an exposition of the fundamental physics of the free-electron laser. In practice, however, these gradients can have important consequences on the performance of the free-electron laser. It is beyond the scope of this article to delve into these subjects in depth, and we will present only a simple sketch of the types of effects that are to be encountered. The most important effect found if the wiggler field varies substantially across the diameter of the electron beam is that the electron response to the wiggler will vary as well. In practice, this means that an electron at the center of the beam will experience a different field than an electron at the edge of the beam, and the two electrons will follow different trajectories with different velocities. As a result of this, the wave-particle resonance that drives the interaction will be broadened and the gain and efficiency will decline. In essence, therefore, the transverse wiggler inhomogeneity is manifested as an effective beam thermal spread. The bounded nature

of the electron beam also affects the interaction, since wave growth will occur only in the presence of the beam. Because of this, it is important in amplifier configurations to ensure good overlap between the injected signal and the electron beam. Once such overlap has been accomplished, however, the dielectric response of the electron beam in the presence of the wiggler can act in much the same way as an optical fiber to refractively guide the light through the wiggler.

III. EXPERIMENTS AND APPLICATIONS

There are three basic operating modes for free-electron lasers: amplifiers, oscillators, and self-amplified spontaneous emission (SASE). In an amplifier, the electron beam is injected into the wiggler in synchronism with the signal to be amplified. The external radiation source that drives the amplifier is referred to as the master oscillator and can be any convenient radiation source, such as a conventional laser or microwave tube. As a consequence, this configuration is often referred to as a master oscillator power amplifier (MOPA). Because amplification occurs during one pass through the wiggler, MOPAs require intense electron beam sources that can operate in the high gain regime. Oscillators differ from amplifiers in that some degree of reflection is introduced at the ends of the wiggler so that a signal will make multiple passes through the system. The signal is amplified during that part of each pass in which the radiation copropagates with the electrons and allows for a large cumulative amplification over many passes even in the event of a low gain per pass. Oscillators are typically constructed to amplify the spontaneous (i.e., shot) noise within the beam, and no outside signal is necessary for their operation. However, a long pulse accelerator is required because a relatively long time may be required to build up to saturation. In SASE mode, shot noise in the beam is amplified over the course of a single pass through the wiggler and, like amplifiers, high current accelerators are required. Since the shot noise present in the beam is generally broadband, the radiation from SASE configurations is typically characterized by a broader bandwidth than a MOPA.

The optimal configuration and type of accelerator used in a free-electron laser design depends upon the specific application, and issues such as the electron beam quality, energy, and current are important considerations in determining both the wavelength and power. In general, however, each accelerator type is suited to the production of a limited range of wavelengths, as shown in Fig. 7. In addition, the temporal structure of the output light from a free-electron laser corresponds to that of the electron beam. Thus, either a pulsed or continuous electron beam source will give rise to a pulsed or continuous free-electron laser. Free-electron lasers have been constructed using virtually every type of electron source, including storage rings, radio-frequency linear accelerators, microtrons, induction linacs, electrostatic accelerators, pulse line accelerators, and modulators. Since the gain in a free-electron laser increases with current but decreases with energy, accelerators producing low-current–high-energy beams are generally restricted to the low-gain regime. Accelerator types that fall into this category include rf linacs, microtrons, storage rings, and electrostatic accelerators. However, designs are under study in which the beam pulses produced in rf linacs are compressed to extremely high peak currents for use as x-ray sources operating in SASE mode. Intense beam accelerators such as induction linacs, pulse line accelerators, and modulators are suitable electron beam sources for high-gain systems.

Storage rings are typically characterized by multiple electron pulses continuously circulating through the ring. Each pulse is several nanoseconds in duration, and the output light from a free-electron laser driven by a storage ring is a continuous stream of picosecond bursts. In addition, while storage rings produce high-quality and high-energy beams of low to moderate currents, the electron pulses are recirculated through both the ring and the wiggler, and the stability of the ring is disrupted by the extraction of too much energy. Hence, storage rings are feasible for applications that require uniform and continuous short wavelength radiation sources but do not demand high-output powers.

The first successful operation of a storage ring free-electron laser was at the Université de Paris Sud at Orsay. This experimental configuration was that of an oscillator, and made use of the ACO storage ring, which operates at energies and average currents in the range of 160–224 MV and 16–100 mA. The laser was tuned across a broad band of the visible spectrum, but was first operated at wavelengths in the neighborhood of approximately 0.65 μm. The peak output power from the oscillator was 60 mW over the 1 nanosec duration of the micropulses, which corresponds to an intracavity power level of 2 kW. The average power extracted from the system was typically of the order of 75 μW. Higher harmonic emission was also detected in the ultraviolet; however, this posed a problem because radiation at these wavelengths resulted in the ultimate degradation of the optical system. An ultraviolet free-electron laser oscillator has been also built using the VEPP storage ring at Novosibirsk. This experiment employed an optical klystron configuration in which the wiggler was composed of two distinct sections separated by a diffusive drift space. In this configuration, the first section operates as a prebuncher for the electron beam that subsequently enhances the gain in the second wiggler

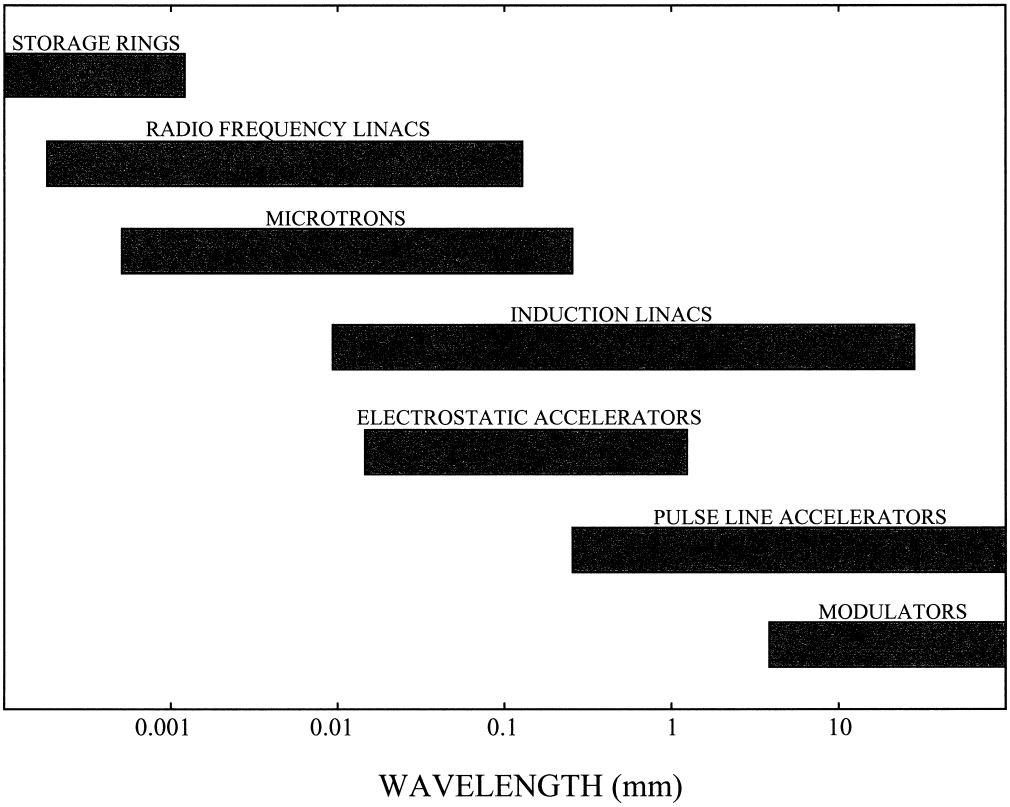


FIGURE 7 The limits on the wavelengths possible with different accelerators depends both on the electron-beam energies that may be achieved as well as on the state of wiggler development. The wavelength ranges that may be accessed by accelerators given the present state of accelerator and wiggler technology are shown.

section. Operating the storage ring at 350 MV and a peak current of 6 A, experimenters were able to obtain coherent emission at wavelengths as short as 0.3 microns and average output powers as high as 6 mW. More recently, a storage ring-driven free-electron laser has begun operation at Duke University using an optical klystron undulator (see Table I).

Radio-frequency linacs employ a series of cavities that contain rapidly varying electromagnetic (rf) fields to accelerate streams of electrons. The beams they produce are composed of a sequence of macropulses (typically of microseconds in duration), each of which consists of a train of shorter picosecond pulses. Microtrons produce beams with a temporal structure similar to that of rf linacs, but unlike the rf linac, they are composed of a single accelerating cavity coupled to a magnet that causes the electron beam to recirculate through the cavity many times. The output light from a free-electron laser built with these accelerators, therefore, is similar to that from a storage ring. Experiments at Stanford University and Boeing Aerospace have also demonstrated the feasibility of the rf linac to produce visible light, and an ultraviolet free-electron laser was built at Los Alamos National Laboratory. In addition, rf linacs and microtrons are suitable for high-power free-electron lasers since the electrons are not recirculated and the energy extraction is not limited by the disruption of the beam.

Free-electron lasers based upon rf linacs have demonstrated operation over a broad spectrum, extending from the infrared through the visible, and are the most common accelerator used in free-electron lasers at the present time. A schematic illustrating the configuration of an rf linac-based free-electron laser is shown in Fig. 8. As shown in the figure, in its simplest form the device consists in (1) an electron gun/injector which feeds (2) the rf accelerator that brings the beam up to the design energy and injects it into (3) the wiggler, which is within (4) an optical cavity, after which the spent beam is shunted off into (5) the beam dump.

The initial experiments conducted at Stanford University resulted in (1) an amplifier that operated at a wavelength of 10.6 microns, with an overall gain of 7%, and (2) a 3.4 μm oscillator that produced peak and average output powers of 7 kW and 0.1 mW, respectively. In collaboration

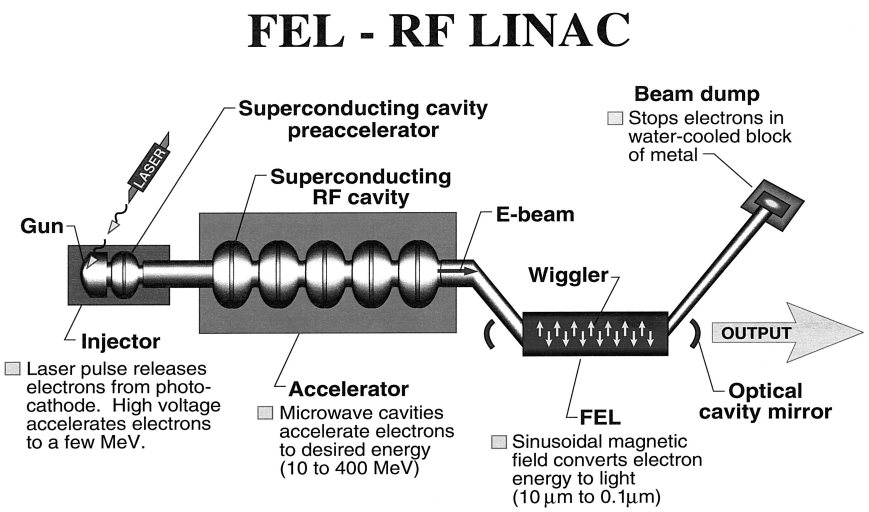


FIGURE 8 Schematic illustration of a rf linac-driven free-electron laser oscillator [from H. P. Freund and G. R. Neil, "Free-Electron Lasers: Vacuum Electronic Generators of Coherent Radiation," Proc. IEEE, vol. 87, pp. 782–803, 1999].

with TRW, the superconducting rf linac (SCA) at Stanford University has been used to drive a free-electron laser oscillator that has demonstrated efficiency enhancement with a tapered wiggler, operation at visible wavelengths, and beam recirculation. The significance of the latter point is that the overall wall-plug efficiency can be enhanced by recovery and reuse of the spent electron beam subsequent to its passage through the free-electron laser. The tapered wiggler experiment operated in the infrared at a wavelength of 1.6 μm and peak power levels of 1.3 MW. This yields a peak extraction efficiency of approximately 1.2%, which constitutes an enhancement by a factor of three over the efficiency in the case of an untapered wiggler. Operation at visible wavelengths was also found at 0.52 μm and peak power levels of 21 kW. The superconducting technology embodied in the SCA can enable the rf linac to further compete with storage rings by operating in a near steady-state mode. Both energy recovery and enhancement of the extraction efficiency by means of tapered wiggler were also demonstrated at Los Alamos National Laboratory. Starting in 1981 with a tapered wiggler free-electron laser amplifier that obtained an extraction efficiency of 4% at a wavelength of 10.6 μm, researchers went on to (1) extend that to a 5% extraction efficiency in an oscillator configuration, and (2) to demonstrate a 70% energy recovery rate with beam recirculation.

The limitations that storage rings, rf linacs, and microtrons impose on free-electron laser design stems from restrictions on the peak (or instantaneous) currents that may be obtained and that limit the peak power from a free-electron laser. High peak powers may be obtained by using induction linacs, pulse line accelerators, or modulators that produce electron beams with currents ranging from several amperes through several thousands of amperes and with pulse times ranging from several tens of nanoseconds through several microseconds. Induction linacs operate by inducing an electromotive force in a cavity through a rapid change in the magnetic field strength. In effect, the electron beam acts as the secondary winding in a transformer. For example, the Advanced Test Accelerator (ATA) built at Livermore is an accelerator of this type and achieved energies and currents as high as 50 MW and 10 kA over a duration of 50 nsec. At lower energies, pulse line accelerators and conventional microwave tube modulators are available. Pulse line accelerators produce beams with energies up to several tens of megavolts, currents of several tens of kiloamperes, and pulse times up to 50–100 nsec. As a result, pulse line accelerators and modulators have been applied exclusively to microwave generation.

Amplifier experiments employing induction linacs were performed at the Naval Research Laboratory, Lawrence Livermore National Laboratory, and at the Institute for Laser Engineering at Osaka University in Japan, and have demonstrated operation from the microwave through the infrared spectra. A superradiant amplifier at the Naval Research Laboratory employed a 650 kV/200-A electron beam and produced 4 MW at a wavelength of 1 cm. The Livermore experiments have employed both the Experimental Test Accelerator (ETA) and the ATA. The free-electron laser amplifier experiment at an 8-mm wavelength was conducted with the ETA operating at approximately 3.5 MV and a current of 10 kA. However, due to beam quality requirements, only about 10% of the beam was usable in the free-electron laser. This ETA-based MOPA has been operated in both uniform and tapered wiggler

configurations. In the case of a uniform wiggler, the measured output power was in the neighborhood of 180 MW, which corresponds to an extraction efficiency of about 6%. A dramatic improvement in the efficiency was achieved, however, using a tapered wiggler. In this case, the total output power rose to 1 GW for an efficiency of 35%. The ATA has been used for a high-power MOPA design at a wavelength of 10.6 μm. Experiments are currently in progress at both Livermore and the Institute for Laser Engineering at Osaka on the use of induction linacs in high-power MOPAs at frequencies above 140 GHz for the purposes of radio frequency heating of magnetically confined plasmas for controlled thermonuclear fusion.

An important consideration in the construction of free-electron lasers with the intense beams generated by these accelerators is that an additional source of focusing is required to confine the electrons against the self-repulsive forces generated by the beam itself. This can be accomplished by the use of additional magnetic fields generated by either solenoid or quadrupole current windings. Quadrupole field windings were employed in the 8-mm amplifier experiment at the Lawrence Livermore National Laboratory; and the interaction mechanism in a free-electron laser is largely, though not entirely, transparent to the effect of the quadrupole field. In contrast, the solenoidal field has a deep and subtle effect on the interaction mechanism. This arises because a solenoidal field causes precession, called Larmor rotation, about the magnetic field lines that can resonantly enhance the helical motion induced by the wiggler. This enhancement in the transverse velocity associated with the helical trajectory occurs when the Larmor period is comparable to the wiggler period. Since the Larmor period varies with beam energy, this relation can be expressed as

$$B_0 \approx 1.07 \frac{\gamma_b}{\lambda_w},$$

where the solenoidal field B_0 is expressed in Tesla, and the wiggler period is in centimeters. For fixed beam energies, therefore, this resonant enhancement in the wiggler-induced velocity requires progressively higher solenoidal fields as the wiggler period is reduced.

The effect of a resonant solenoidal magnetic field is to enhance both the gain and saturation efficiency of the interaction. This was demonstrated in a superradiant amplifier experiment using the VEBA pulse line accelerator at the Naval Research Laboratory. In this experiment, the output power from the free-electron laser was measured as a function of the solenoidal field as it varied over the range of 0.6–1.6 Tesla. The beam energy and current in the experiment were 1.35 MV and 1.5 kA respectively, and the wiggler period was 3.0 cm. It should be remarked that due to the high current in the experiment, the beam was unable to propagate through the wiggler for solenoidal fields below 0.6 Tesla. The magnetic resonance was expected for a solenoidal field in the neighborhood of 1.3 Tesla, and the experiment showed a dramatic increase in the output power for fields in this range. Other experiments that have demonstrated the effect of a solenoidal field have been performed at the Massachusetts Institute of Technology, Columbia University, and the École Polytéchnique in France.

The maximum enhancement in the output power in the experiment at the Naval Research Laboratory was observed for solenoidal fields slightly above the resonance. In this case, the nature of the interaction mechanism undergoes a fundamental change. In the absence of a solenoidal field (or for fields below the magnetic resonance), the axial velocity of the electrons decreases as energy is lost to the wave, while the transverse velocity remains relatively constant. In contrast, the result of the strong solenoidal field is to cause a negative-mass effect in which the electrons accelerate in the axial direction as they lose energy to the wave. The bulk of the energy used to amplify the wave is extracted from the transverse motion of the electrons. Computer simulations of free-electron lasers operating in this strong solenoidal regime indicate that extremely high extraction efficiencies (in the neighborhood of 50%) are possible without recourse to a tapered wiggler field. However, operation in this regime is precluded below submillimeter wavelengths. The reason for this is that the solenoidal field required for this magnetic resonance varies directly with the beam energy and inversely with the wiggler period. Because of this, impractically high fields are required for wavelengths in the far infrared and below.

It is important to bear in mind that the high peak versus average power and oscillator versus MOPA distinctions between the different aforementioned accelerator technologies are becoming blurred by advances in the design of both rf and induction linacs. On the one hand, the development of laser-driven photocathodes at Los Alamos National Laboratory has dramatically increased both the peak (of the order of 400 A) and average currents achievable with rf linacs. As a result, a high-gain MOPA experiment is under design at Los Alamos National Laboratory. It is significant in this regard that a collaboration between Stanford University and Rocketdyne Inc, has already achieved MOPA operation using the Mark III rf linac. On the other hand, new induction linacs are under development that may be fired repetitively at the rate of up to several thousand times per second. Successful completion of these development programs will enable high average power free-electron lasers to be constructed using this technology.

Although their average power is lower than that of linacs, electrostatic accelerators can produce continuous

electron beams using charge recovery techniques. In such a process, the electron beam is recirculated through the wiggler and back into the accelerator in a continuous stream. Using this technology, the electrostatic accelerator holds promise as an electron beam source for a continuous or long-pulse free-electron laser. However, given practical restrictions on the size of such accelerators, which limit energies and currents, electrostatic accelerators have been restricted to the construction of free-electron laser oscillators that operate from the microwave regime through the infrared spectrum.

The breadth of free-electron laser experiments includes many different wiggler configurations, and virtually every type of accelerator in use today. The wiggler has been produced in planar, helical, and cylindrical forms by means of permanent magnets, current carrying coils, and hybrid electromagnets with ferrite cores. An example of a bifilar helical wiggler that was used at the Naval Research Laboratory is shown in Fig. 9. Helical wiggler fields are produced by a current-carrying bifilar helical coil in which field increases radially outward from the symmetry axis and provides magnetic focusing to confine the beam against the mutually repulsive forces between the electrons. In a planar wiggler both the transverse and axial components of the velocity oscillate in synchronism with the wiggler. As such, the interaction is governed by the average, or root-mean-square, wiggler field. Because of this, planar wigglers require a stronger field to produce the same effect as a helical wiggler. This is compensated for, however, by the ease of adjustment allowed by a planar design, in which the strengths or positions of the individual magnets can be altered to provide either a uniform or tapered field. In contrast, the only adjustment possible for a bifilar helix is the strength of the field.

One practical constraint on the development of free-electron laser oscillators at the present time is mirror technology for infrared and shorter wavelengths and relates to both reflectivity and durability. The reflectivity is important since the net gain of an oscillator decreases as the mirror losses increase, and oscillation is possible only if the amplification due to the free-electron laser interaction exceeds the losses at the mirrors. The reflectivity is a measure of this loss rate, and must be kept sufficiently high that the energy losses at the mirrors do not overwhelm the gain. The issue of durability relates to the power level that any specific mirror material can endure without suffering optical damage. In this sense, optical damage refers to a decrease in the reflectivity. Note that the extreme case of complete burn out of the mirrors might be described as a catastrophic decrease in the reflectivity. Problems exist in finding materials with a sufficiently high reflectivity and durability to enable operation in the infrared and ultraviolet spectra, and even the visible presents problems. For example, the visible free-electron laser oscillator at the Université de Paris Sud experienced mirror degradation due to harmonic emission in the ultraviolet. At extremely high power levels, solutions can be found through such techniques as the grazing incidence mirrors used by Boeing Aerospace in which the optical beam is allowed to expand to the point where the power density on the mirrors is low. In the infrared, oscillator experiments at Los Alamos National Laboratory originally employed a dielectric mirror material with a high reflectivity at low power levels. However, recent observations indicate that

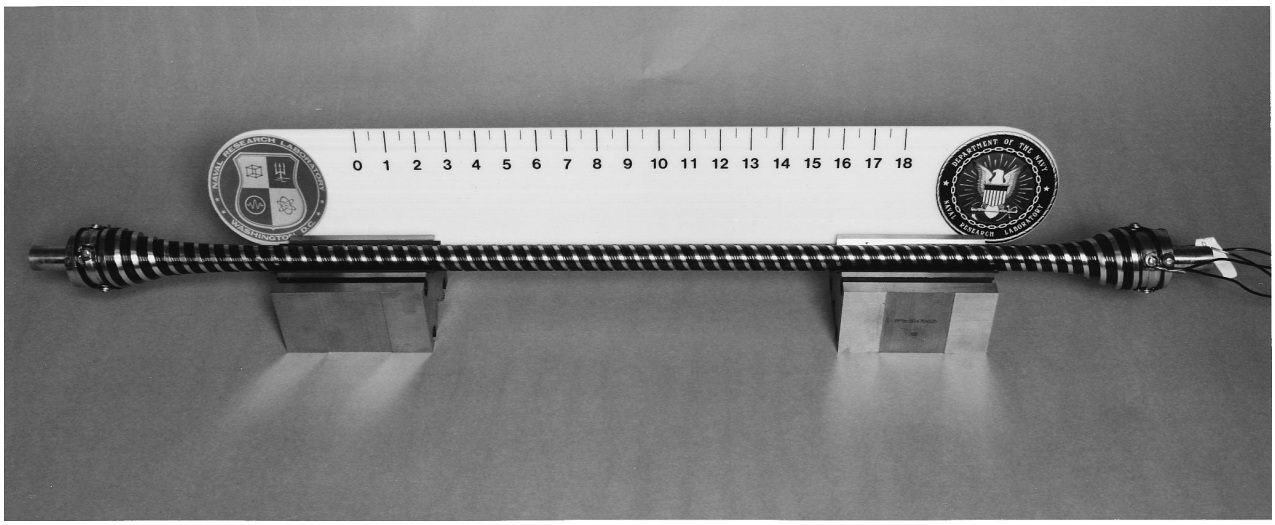


FIGURE 9 One of the principal wiggler designs in use is the bifilar helical current winding. The example shown here exhibits flares at both the entrance and exit to the wiggler in order to facilitate the transition of the electron beam both into and out of the wiggler field. The scale shown is in centimeters. (Photo courtesy of Dean Pershing.)

nonlinear phenomena occur in this material at high power levels, which effectively reduces the reflectivity, and that the use of copper mirrors substantially improves performance. An additional problem occurs at high power levels due to thermal distortion of the optical surface. In order to combat this problem, actively cooled mirrors are under development (Fig. 10).

There has been substantial growth in the applications of the radiation produced by free-electron lasers. It is impossible to cover the full breadth of this activity, but we will try to cover many broad categories. Several studies have examined the applicability of free-electron lasers to perform basic scientific research and have found wavelength ranges in which the free-electron laser has no effective competition (10 μm to several millimeters excepting several fixed points where existing lasers can operate). In other regions it may still be the source of choice by virtue of other characteristics: short pulse length, micropulse energy/peak power, ability to synchronize for pump probe efforts, bandwidth, beam quality, and, most fundamentally, tunability.

High average-power free-electron masers operating in the millimeter wavelength range are being developed for plasma heating. The free-electron laser can also excel as a research tool in this range because it produces picosecond pulses of high-intensity radiation. Typical studies include measurements of principle excitations in condensed matter systems, where it is possible to access the principal excitations such as plasmons, phonons, magnons, and intersubband transitions. Direct linear probing of defect modes and buried interfaces with bond specificity is possible. Measuring energy distributions and line shapes probes mode coupling and energy dissipation into the electron or phonon continuum of the substrate. Low-frequency modes in large biomolecules such as nucleic acids and proteins can be excited in FIR for similar studies of energy flow.

Most existing free-electron lasers operate in the infrared regime. For many uses at wavelengths less than around 10 μm, existing tabletop lasers can satisfy most researchers' needs. Exceptions that drive users to free-electron lasers occur when they need high peak powers, have a particular need to exploit the wide tunability of the free-electron laser, or the possibility of chirped pulses. Many of the researchers synchronize the free-electron laser output with another radiation source for measurements in this wavelength range. A particularly powerful technique has been sum frequency generation (SFG). The versatility of this approach is demonstrated by the

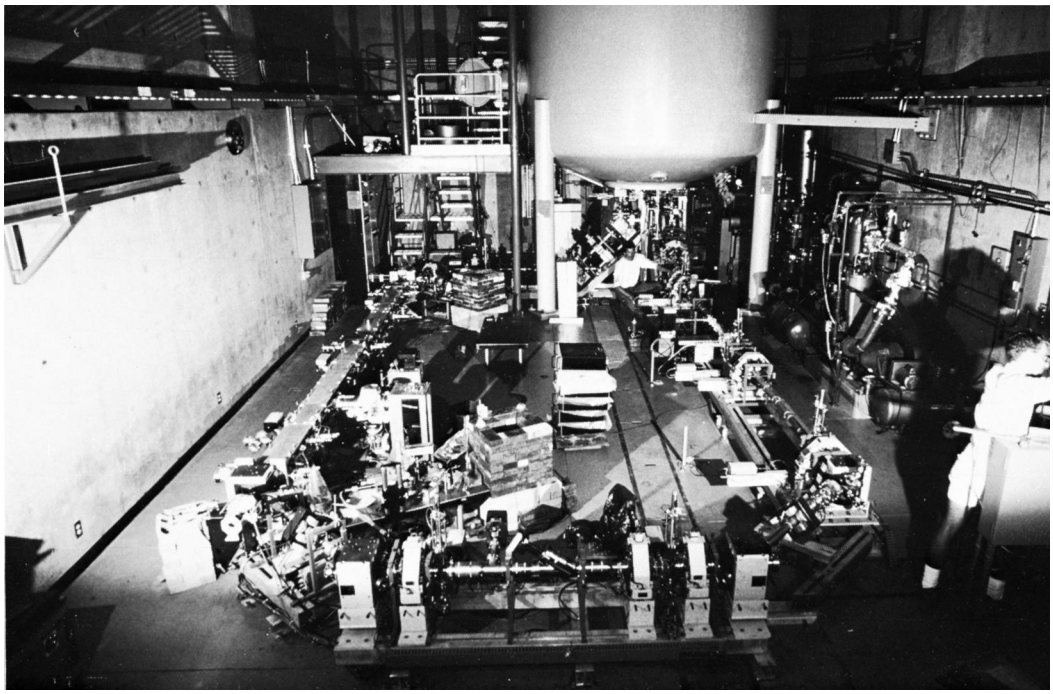


FIGURE 10 The first free-electron user facility was established at the University of California, Santa Barbara. This device operates in the far infrared and employs an electrostatic accelerator, as shown in the background. A magnetically focused system of electron-beam optics is used to transport the beam from the accelerator through the wiggler (in foreground) and back to the accelerator. (Photo courtesy of Luis Elia.)

variety of applications demonstrated. Early examples include sum-frequency measurements on the surface of Pt in methanol using 5-μm free-electron laser pulses and a synchronized laser. In other tests the free-electron laser was used for pump probe observation of coherent transient grating effects of narrow gap semiconductors and third-order nonlinearity coefficients and electron relaxation times in GaAs/AlGaAs quantum wells. More recent activities include using SFG to identify bonding and density monitoring when putting polyurethane on float glass, the interaction of fullerene deposited on gold to show the surface interaction modifies the fullerene geometry, and vibrational dynamics of CO at an electrochemical interface.

Studies in kinetics in the infrared include vibrational energy transfers in molecules. This opens up a new class of experiments to study mode-selective chemistry that requires high-power short pulses that excite molecular vibrations.

A group at Stanford studied second-order nonlinear susceptibility of the conduction band and valence band quantum well (QW) structures extracted from the interference between second harmonic fields of QWs and GaAs substrate as determined by the azimuthal dependence of the second harmonic power. This is the first demonstration of difference frequency generation of mid-IR in any QW. Groups at Stanford also studied vibrational dynamics in glass-forming liquids. These are the first vibrational photon echo experiments and first comprehensive temperature-dependent pump probe measurements on any condensed matter system.

Among the extensive studies carried out at the FELI facility in Japan are resonant excitations of molecular vibrations, band discontinuities of semiconductor heterojunctions, and isotope separation. The user facilities at this lab have become among the most productive in the world, providing over 2000 hr of beam time in the last year.

Studies in medical applications at Vanderbilt University are leading to the first use of an free-electron laser in human neurosurgery by taking advantage of particular absorption bands to produce an exceptionally fine cut in bone and tissue. FELI has also ablated and hardened dental materials and studied photodynamic therapy. The tunability, power, and pulse variability of the FEL has made it an efficient biophysical research tool.

Commercial applications have been proposed that use high average power free-electron lasers to heat the surface of polymers for enhancements to the surface morphology. This uses infrared at 5.8 to 6.2-μm wavelengths where high absorption results from carbonyl-related molecular absorption bands. By enhancing surface roughness in polyester and nylon fibers, the fabrics can be made softer, hydrophilic, and the material more readily accepts dyes.

Although existing lasers in the visible and ultraviolet spectral ranges satisfy most users, there are applications that apply the unique advantages of the free-electron laser. Time-resolved polarized fluorescence decays of the reduced nicotinamide adenine dinucleotide coenzyme NADH in aqueous solution using single-photon counting were performed at the user facility in France in the 350–430-nm range. The advantage of using a free-electron laser on a storage ring is the natural synchronization with synchrotron emission in bending magnets. This was used to study relaxational dynamics of the excited state. In other experiments, researchers measured the surface photovoltage effect on Si (111) and the resulting modification of electronic band bending using time-resolved photoemission.

Proposals for commercial and other applications are under study, for example, power beaming to satellites. The purpose is to provide enhanced light intensity to increase the efficiency of solar panels. Proponents claim an improvement of 1000% more electricity per panel than from the sun alone.

In the deep ultraviolet regime, the free-electron laser competes with synchrotron sources. Typically, synchrotrons provide average, not peak, power and the pulses are 10s of picoseconds in duration at 1-nsec intervals as opposed to the few picosecond pulses of free-electron lasers. Free-electron laser oscillators also provide a narrow linewidth; hence, when peak power at a particular wavelength is desired, the free-electron laser will hold the edge, while the synchrotron will have the advantage for average powers of wide bandwidth.

As mentioned previously, the free-electron laser is under study as a so-called fourth-generation light source in the x-ray regime. The goal here is to produce a coherent x-ray source that greatly exceeds the current (third-generation) incoherent synchrotron sources by many orders of magnitude in brightness. These devices would operate in the SASE mode due to the absence of either suitable x-ray sources to drive an amplifier and the mirrors needed for oscillator configurations. A design study at the Stanford Linear Accelerator Center will serve as an illustration of the concept that is referred to as the Linac Coherent Light Source (LCLS). The LCLS is based on using an electron beam with an energy of 15 GeV and beam pulses that have been compressed to peak currents of up to 3400 Å with a duration of 100 fsec. The short electron pulse length is achieved by a magnetic bunching chicane and a correlated energy spread on the beam. At the present time, there are unanswered questions about the growth of emittance in such chicanes due to coherent synchrotron emission effects that could lead to performance problems. However, efforts are underway to answer the outstanding questions from both theoretical and

experimental standpoints. A peak output power of 10 GW is indicated at a wavelength of 1.5 A. A similar concept, but based on advanced superconducting structures, is under construction at Deutsches Elektronen-Synchrotron in Germany. The device is referred to as the TESLA Test Facility (TTF), and the plan is to reach a wavelength of 65 Å using a 1-GeV/2500-A electron beam. The anticipated peak output power is in the range of 2–3 GW. The advantage of a superconducting rf linac is that the TTF has a high duty factor for high average brilliance. A comparison of the average brilliance versus photon energy produced in a wide range of synchrotron light sources and in the LCLS and TTF is shown in Fig. 11.

IV. SUMMARY AND DISCUSSION

The aforementioned discussion includes a necessarily abridged list of recently conceived applications of the free-electron laser. The fundamental principles of the free-electron laser are understood at the present time, and the future direction of research is toward evolutionary improvements in electron beam sources (in terms of beam quality and reliability) and wiggler designs. The issues, therefore, are technological rather than physical, and the free-electron laser can be expected ultimately to cover the entire spectral range shown. In this regard, it is important to recognize that the bulk of the experiments to date have been performed with accelerators not originally designed for use in a free-electron laser, and issues of the beam quality important to free-electron laser operation were not adequately addressed in the initial designs. As a consequence, the results shown do not represent the full potential of the free-electron laser, and significant improvements in performance can be expected.

The two principal thrusts of free-electron laser development at the present time are towards high average powers and ultrashort wavelengths. At the forefront of the high average power development work is activity at the Jefferson Laboratory using a superconducting rf linac work that achieved an average power of 1.7 kW at a wavelength of 3 μm. Future developments at this facility are directed towards still higher average powers and still shorter wavelengths, with the ultimate goal of producing a high average power ultraviolet free-electron laser. An estimate of the rate of development of this technology can be obtained from a historical survey of the development rates of other vacuum electronic devices, such as magnetrons, klystons, traveling wave tubes (TWTs), gyrotrons, and free-electron lasers. With such devices, the average power density (the product of the average power and the square of the operating frequency) is a useful figure of merit that provides an insightful basis for comparing device types and portrays trends in the rate of technological development. The physical significance of $P_{avg}f^2$ as a figure of merit derives from the fact that the maximum power that can be transported through a device is proportional to the circuit cross-sectional area that is inversely proportional to the operating frequency. Upper limits on the power density, therefore, are considered to be intrinsic to a particular device configuration. The evolution of this figure of merit over time for these devices is shown in Fig. 12.

It is seen in the figure that the history of vacuum electron devices can be viewed as successive waves of scientific innovation driven by technical opportunity coupled with market demand. At the outset, early gridded tubes were overtaken by the development of the magnetron under the impetus of wartime radar demands. In the post-war period, high-power liner beam tubes such as the klystron and coupled-cavity traveling-wave tube moved to the fore in response to accelerator and radar requirements. Magentrons, klystrons, and traveling-wave tubes can all be referred to as "slow-wave" sources since the structures (or circuits) are required to support waves that are sub luminous (i.e., where the phase velocity is less than the speed of light *in vacuo*). Gyrotrons and free-electron lasers are "fast-wave" sources in which the phase velocity of the wave is supraluminous. The advantage of fast-wave sources is that the wave is not supported by the circuit; hence, the structure can be opened up to larger dimensions. This, in turn, eases fundamental limitations on the power densities. Indeed, it is clear that gyrotrons and free-electron lasers surpass the slow-wave sources by the mid-1970s. However, although gyrotrons have been effective in achieving high-average powers at millimeter wavelengths, the solenoidal magnetic fields they require for shorter wavelengths become prohibitively high. With these transitions, the envelope of vacuum device performance, as defined by the straight line in Fig. 12, continues to grow exponentially at an approximate rate of more than an order of magnitude per decade. This development rate has been sustained for the past 60 years.

The ultrashort wavelength work is represented by the x-ray sources under development at the Stanford Linear Accelerator Center and at the Deutsches Elektronen-Synchrotron. Here, the goals are the production of x-ray sources with a high average brilliance at wavelengths in the vicinity of 1–2 Å. One problem that must be faced, however, is that the wakefields induced by the high peak current pulses in the accelerator and beam transport system can result in serious degradation in beam quality. The minimization of these effects is an important consideration in the overall design of the beam lines in these free-electron lasers.

In conclusion, free-electron laser research and development is international in scope with a wide range of experiments and user facilities either in operation or in the planning stages throughout the United States, Europe,

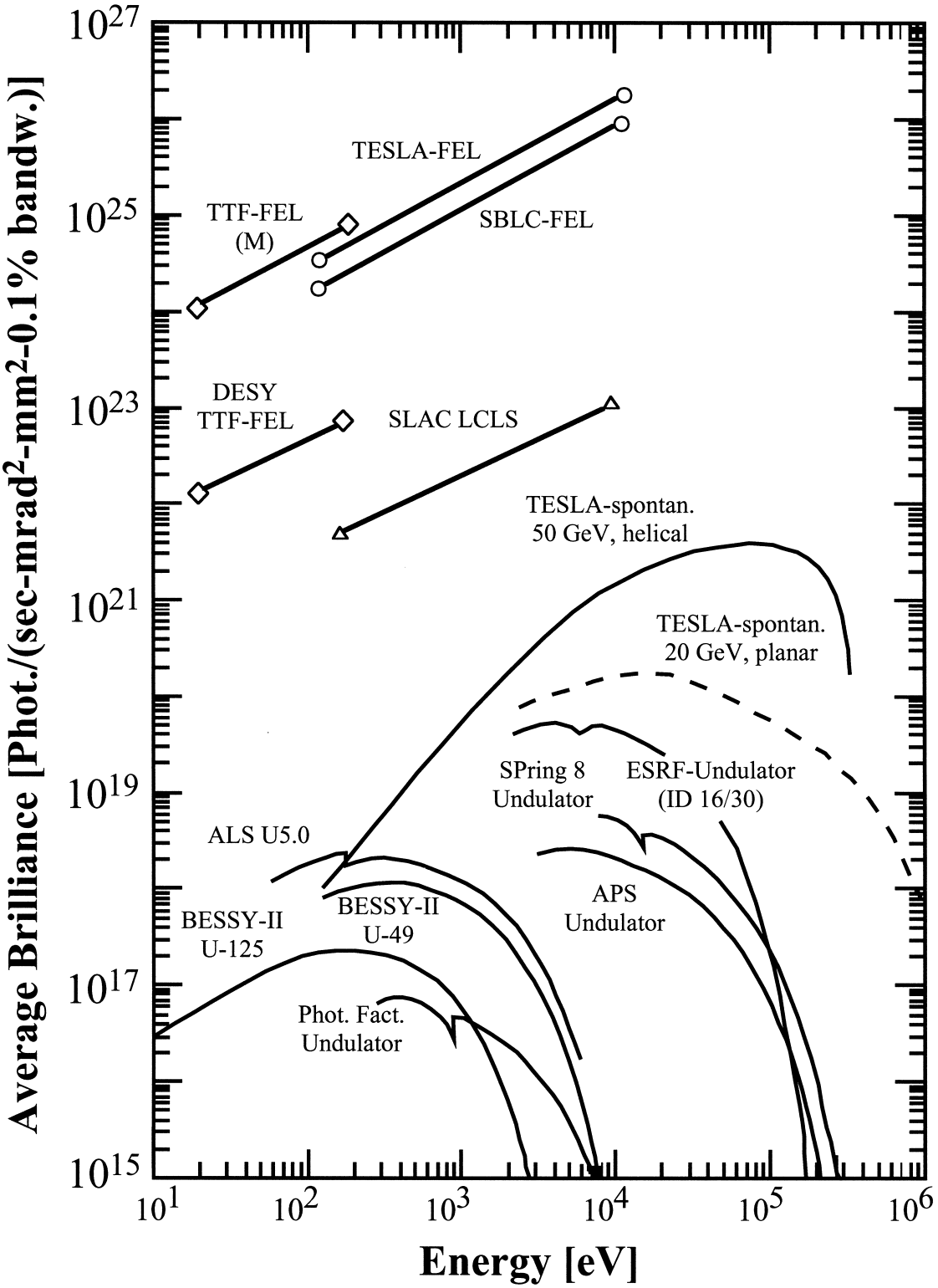


FIGURE 11 The average brilliance predicted for the Linac Coherent Light Source (LCLS) and TESLA Test Facility (TTF) free-electron laser (FEL) facilities in comparison with present-day synchrotron light sources. APS, ESRF, ALS, SBLC.

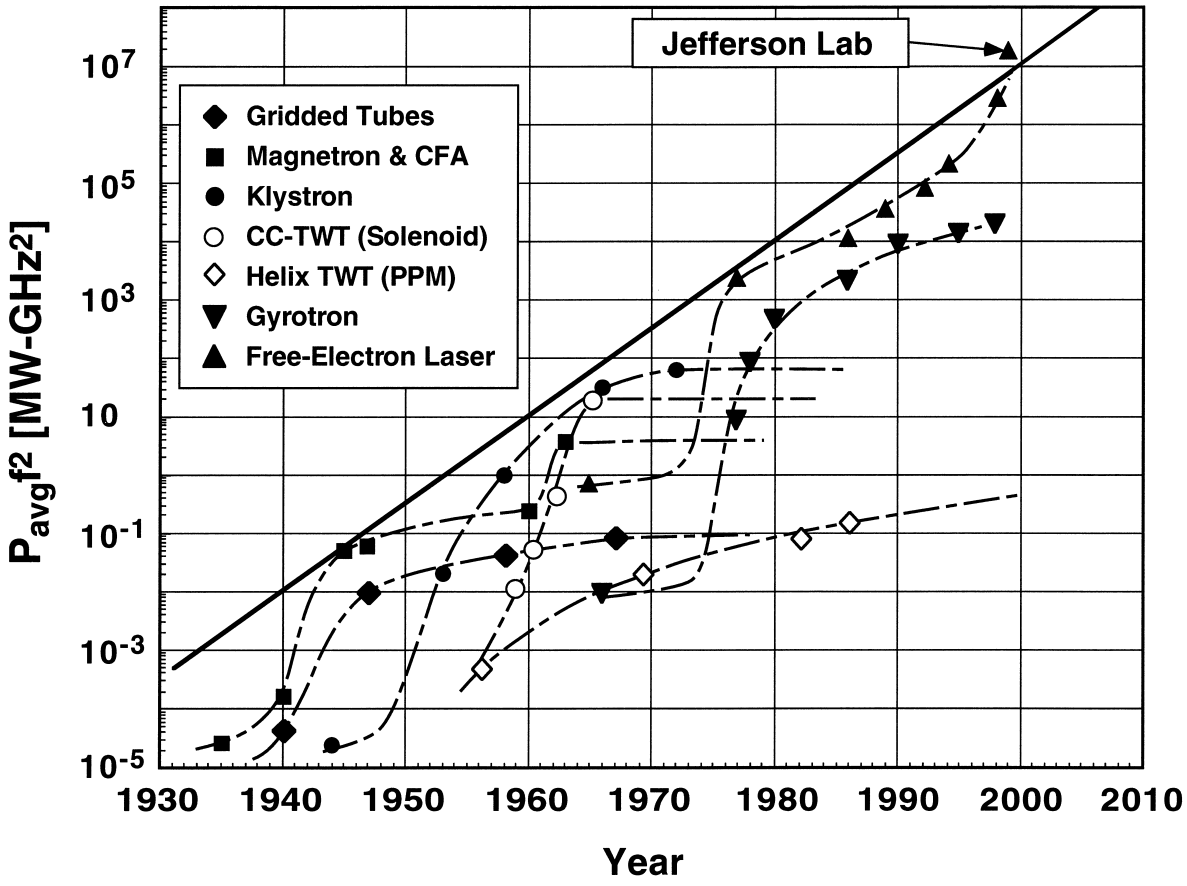


FIGURE 12 Progress in the development of vacuum electronic sources as measured by the growth in the product of the average power and the square of the frequency versus time. TWT, CFA, CC-TWT,

Japan, China, and the Soviet Union. These projects include a wide range of designs based upon all of the accelerator types and cover a spectral range extending from microwaves through X rays. The technology has begun to make the transition from development in the laboratory to active exploitation over a range of applications and still greater advances are likely over the next decade.

SEE ALSO THE FOLLOWING ARTICLES

LASERS • LASERS, DYE • LASERS, GAS • LASERS, SOLID-STATE • LASERS, X-RAY

BIBLIOGRAPHY

Deacon, D. A. G., Elias, L. R., Madey, J. M. J., Ramian, G. J., Schwettman, H. A., and Smith, T. I. (1977). First Operation of a Free-Electron Laser, *Phy. Rev. Lett.* **38**(16), 892–894.

"Free Electron Lasers and Other Advanced Sources of Light" (1994). Report of the Committee on Free Electron Lasers, National Research Council, National Academy Press, Washington, DC.

Freund, H. P., and Antonsen, T. M. Jr. (1996). "Principles of Free-Electron Lasers," Chapman & Hall, London.

Freund, H. P., and Parker, R. K. (1989) "Free-Electron Lasers," *Scientific American* **260,** 84–89.

Jacobs, S. F., Pilloff, H. S., Sargent, M., Scully, M. O., and Spitzer, R. (eds.) (1980). "Physics of Quantum Electronics: Free-Electron Generators of Coherent Radiation," Addison-Wesley Publishing Co., Vol. 7, 1980.

Jacobs, S. F., Pilloff, H. S., Sargent, M., Scully, M. O., and Spitzer, R. (eds.) (1982). "Physics of Quantum Electronics: Fre tron Generators of Coherent Radiation," Vols. 8 and 9, Addison-Wesley Publishing Co.

Marshall, T. C. (1985). "Free-Electron Lasers," Macmillan Publishing Co., New York.

Martellucci, S., and Chester, A. N. (eds.) (1983). "Free-Electron Lasers," Plenum Press, New York.

Motz, H. (1951). "Applications of the Radiation From Fast Electron Beams," *J. Appl. Phy.* **22**(5), 527–535.

Phillips, R. M. (1988). "History of the Ubitron," *Nuclear Instruments and Methods of Physics Research* **A272**(No. 1), 1–9.

Roberson, C. W., and Sprangle, P. (1989). "A Review of Free-Electron Lasers," *Physics of Fluids B* **1**(1), 3–42.

Sessler, A. M., and Vaughan, D. (1987). "Free-Electron Lasers," *American Scientist* **75,** 34–43.

Lasers, Nuclear Pumped

David A. McArthur
Sandia National Laboratories (retired)

I. Examples of Nuclear-Pumped Lasers
II. Characteristics of Nuclear-Pumped Lasers
III. Basic Physics of Reactor-Pumped Lasers
IV. Reactor-Pumped Lasers

GLOSSARY

Fast burst reactor (FBR) Produces neutron pulses about 0.1–1 msec in width, by rapid mechanical assembly of the critical mass in a neutron-free environment.
Nuclear device Nuclear explosive which produces a single short pulse of intense gamma and fast-neutron radiation.
Nuclear device pumped laser (NDPL) Laser pumped by the very high radiation fluxes obtainable near an exploding nuclear device.
Range Mass per unit area of a particular material which is sufficient to absorb all the kinetic energy of a charged particle having a well-defined initial energy (usually expressed in milligrams per square centimeter).
Reactor pumped laser (RPL) Laser pumped by a portion of the leakage neutrons or gamma rays produced by a pulsed or steady-state nuclear reactor.
Substrate Mechanical support for a thin coating of fissile material in direct contact with the laser medium or nuclear flashlamp medium.
TRIGA Zirconium-hydride-moderated research reactor with inherently safe operating characteristics which can be operated in a variety of modes (pulsed, steady-state, or multiple pulses).
Volumetric pump rate The rate of energy deposition per unit volume of laser medium.

This work was supported by the U.S. Department of Energy under contract DE-AC04-76PD00789.

NUCLEAR PUMPED LASERS (NPLs) are optical lasers operating at wavelengths ranging from the ultraviolet through the infrared which are excited directly or indirectly by high-energy charged particles resulting from nuclear reactions. This definition includes optically pumped laser media in which the pumping light source is produced by nuclear reactions, but it excludes lasers in which nuclear radiation is used merely to pre-ionize the laser medium to control a discharge which provides the primary excitation.

Ideas for NPLs were proposed shortly after the invention of the laser. Two broad categories of NPLs have been demonstrated, those pumped by nuclear explosives (NDPL) and those pumped by specialized (usually pulsed) research reactors (RPLs). Table I summarizes basic data on representative NPLs.

I. EXAMPLES OF NUCLEAR-PUMPED LASERS

A. Nuclear Device-Pumped Lasers

In 1974 two independent experiments in pumping of laser media by nuclear explosives were reported briefly.

Lawrence Livermore Laboratory reported measurements of optical gain and directionally enhanced light emission in high-pressure Xe gas, and Los Alamos Laboratory simultaneously reported a chemical HF laser initiated by radiation. These were the first NPLs reported and demonstrated that nuclear radiation could produce lasing action.

In these studies the laser media were pumped at extremely high volumetric pump rates with an intense short pulse of gamma rays, resulting in laser pulses of about 10 nsec width. The laser cells themselves were small but efficiencies were reasonable. Great care was exercised to minimize radiation damage to the laser optics and to construct an optical apparatus which was insensitive to radiation damage. It is likely that the laser apparatus was totally destroyed by the blast or at least made inaccessible by radioactivation. Russian laboratories have reported on these NDPL experiments in somewhat greater detail in recent years (Bonyushkin, 1996). Other countries with nuclear weapons programs and access to underground test facilities may have the capability to perform similar experiments.

Although many laser media require such high pump rates that they can only be pumped with the high radiation fluxes from a nuclear device, the difficulty and cost of protecting the laser apparatus from the shock and blast effects of the nuclear explosion may limit NDPLs to single-shot systems. Since much of the cost of a large laser system may be associated with the laser optics, routine destruction of the laser may not be acceptable. The catastrophic effects of an accidental explosion of the nuclear device would also limit deployment of NDPLs to areas far from human activity. In light of these limitations, it is likely that NDPLs could only be deployed in space or in underground locations below the earth's surface. Because of the highly specialized nature of NDPLs and the small amount of published information, further discussion will emphasize the reactor pumped laser.

B. Reactor Pumped Lasers

The more typical NPL reported has been pumped by a pulsed or steady-state nuclear reactor. Reactor pumped lashhers (RPLs) are in a sense "laboratory" laser devices, even though in practice there are only a few laboratories with the required reactor facilities. The relatively few workers in the RPL field is probably related to the scarcity of these facilities. In these RPL experiments an intense pulse of neutrons triggers a nuclear reaction, which emits heavy charged particles into a gaseous laser medium. This results in excitation processes qualitatively similar to those produced by an intense electron beam.

The first clear case of reactor pumped lasing was reported by Sandia National Laboratories in 1975 and occurred in CO gas cooled to 77 K (see Table I). Shortly

TABLE I Representative Nuclear Pumped Lasers (See attached corrections)

Laser	Wavelength (10^{-6} m)	Laboratory[a]	Date	Pump source[b]	Target nucleus	Volumetric pump rate (MW/m^3)	Laser energy[b] (J)	Laser efficiency[b] (%)
Xe_2^*	0.17	LLNL	1974	NDPL	gamma	1.5×10^9	N/A	N/A
HF	2.7	LANL	1974	NDPL	gamma	8.0×10^8	2.9	6.5
CO	5.4	Sandia	1975	FBR	^{235}U	1000	0.02	0.5
Xe	3.5	LANL	1975	FBR	^{235}U	200	10^{-6}	10^{-4}
N	0.86, 0.94	UI	1976	TRIGA	^{10}B	3.3	10^{-5}	10^{-4}
Hg^+	0.615	UI	1977	FBR	^{10}B	300	2×10^{-7}	10^{-6}
Xe	2.65	NASA	1980	FBR	$^{235}U/UF_6$	25	10^{-3}	0.05
Ar	1.79	NASA	1981	FRB	^{3}He	2100	0.12	0.02
Xe	1.73	ARSRIEP	1982	N/A	^{235}U	10	2–3	2–5
Cd^+	0.4416	ARSRITP	1982	N/A	^{3}He	1.9	5×10^{-6}	0.007
Ne	0.585	MEPhI	1990	N/A	^{3}He	100	3×10^{-4}	0.1
I	1.31	UI	1993	TRIGA	^{3}He	40	10^{-4}	<0.1
Ar	1.27	ARSRIEP	1991	FBR	^{235}U	30	3×10^{-2}	1.1

[a] LLNL, Lawrence Livermore National Laboratory, Livermore, California; LANL, Los Alamos National Laboratory, Los Alamos, New Mexico; Sandia, Sandia National Laboratory, Albuquerque, New Mexico; UI, University of Illinois, Champaign, Illinois; NASA, NASA Langley Research Center, Hampton, Virginia; ARSRIEP, All-Russia Scientific-Research Institute of Experimental Physics, Nizhni Novgorod, Russia; ARSRITP, All-Russia Scientific-Research Institute of Technical Physics, Snezhinsk, Russia; MEPhI, Moscow Engineering Physics Institute, Moscow, Russia.

[b] N/A, not available.

thereafter, Los Alamos Scientific Laboratory reported lasing in He/Xe gas mixtures, and in 1976 the University of Illinois reported lasing in Ne/N_2 gas mixtures at the much lower volumetric pump rates obtainable with their pulsed TRIGA reactor. However, in 1992 workers at Russia's Arzamas-16 Laboratory (ARSRIEP) reported that unpublished work had demonstrated lasing using only fission fragment excitation, and that this work predated the first published RPLs reported in 1975. In all of these feasibility experiments, only a very small fraction of the leakage neutrons from a reactor facility was used to excite the laser media.

Figure 1 shows a representative reactor pumped laser experiment. The laser apparatus consists of a rectangular stainless steel chamber located adjacent to a pulsed reactor (the Sandia National Laboratories SPR-II fast burst reactor). The chamber is fitted with high-reflectivity mirrors which reflect the laser beam in a folded path through the spatial region which contains the highest neutron flux. The chamber is filled with laser gas, and the laser gas along the "folded" optical path is excited by fission fragments emitted from the inner surfaces of hollow cylindrical ceramic tubes.

A slab of hydrogen-containing plastic is placed between the reactor and the chamber to slow the fast neutrons produced by the reactor (the process of "neutron moderation"). The slow neutrons are more readily absorbed by the uranium coatings on the inside of the cylinders, triggering the fission reaction. Fission fragments (energetic, highly charged heavy ions) emerge from the coatings to excite the laser gas. The excitation energy is stored in the ^{235}U nucleus, and the function of the neutron is to trigger the release of this fission energy. The energy deposition originates at the spatial location where the neutron is absorbed by the uranium nucleus and extends along the path

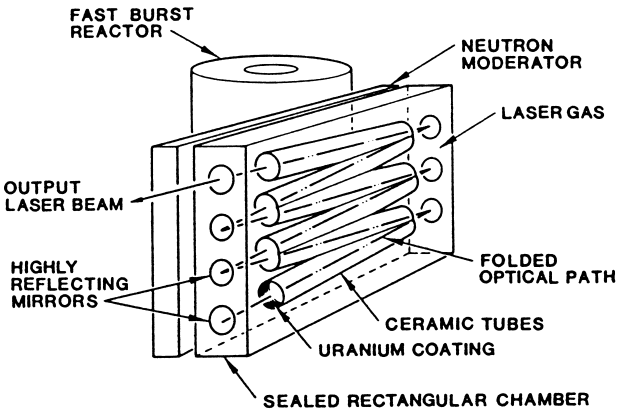


FIGURE 1 Reactor pumped laser apparatus for use with a fast burst reactor (FBR).

of the fission fragment until it comes to rest. The volumetric pump rate is easily varied over a wide range by varying the energy released in the reactor pulse.

Figure 2a shows the time dependence of the fission fragment pulse which excites the laser gas, along with the laser pulse from a CO laser gas mixture. The laser action begins at a low excitation power but peaks before the excitation power peaks, and lasing ceases before the end of the excitation pulse. The volumetric pump power is about 1000 W/cm^3, and the total energy deposited is about 0.2 J/cm^3.

Termination of the laser pulse before the end of the excitation pulse is commonly seen for reactor pumped lasers. It may be related to gas heating which reduces the optical gain, to optical distortion of the laser medium by gas heating, or to creation of chemical species which absorb the laser light. In early measurements of optical gain in CO gas, severe focusing of a probe laser beam by optical aberrations in the excited medium was observed. More recent measurements have revealed details of this focusing tendency in the excited medium, and this is an active area of current research directed toward improving reactor pumped laser performance. In early experiments, evidence of chemical decomposition of the CO gas was also observed. A yellowish film (probably C_3O_2) accumulated on the laser mirrors. C_3O_2 is a well-known decomposition product observed when CO is irradiated with alpha particles.

Figure 2b shows a laser pulse at about 1.8-μm wavelength which is obtained with the apparatus shown in Fig. 1 when the CO laser mixture is replaced by a He/Ar mixture. The fact that lasing occurs over most of the excitation pulse indicates that laser pulse termination is not a property of the laser apparatus itself but results from differences between the two laser gas mixtures.

Figure 2c shows the excitation pulse and the resulting laser pulse at 1.7-μm wavelength from an Ar/Xe laser gas mixture excited by a reactor analogous to a pulsed TRIGA reactor (the Sandia National Laboratories ACRR). In this case, lasing continues over most of the excitation pulse, which is much longer than the pulse width of the reactor used to obtain the data of Figs. 2a and 2b.

II. CHARACTERISTICS OF NUCLEAR-PUMPED LASERS

A. Pumping Mechanisms

There are two types of pump sources: a source of intense gamma radiation or a source of intense neutron flux. The gamma rays can irradiate the laser medium directly and produce excitation. The neutrons ordinarily excite the

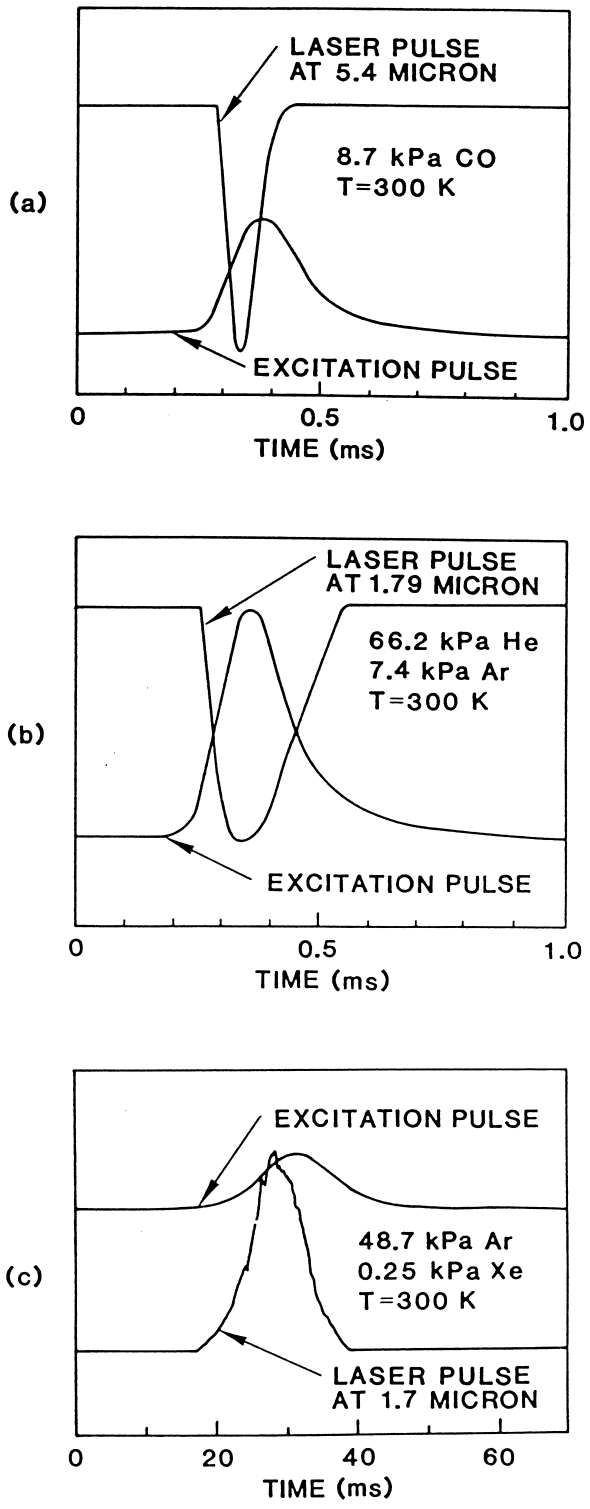


FIGURE 2 Representative reactor pumped laser signals: (a) CO RPL excited by Sandia Labs SPR-II (FBR); (b) Ar RPL excited by Sandia Labs SPR-II (FBR); (c) Xe RPL excited by Sandia Labs ACRR (analogous to pulsed TRIGA reactor).

laser medium indirectly by irradiating other "target" nuclei which undergo nuclear reactions and are the direct agents in exciting the laser medium. At present there are two practical sources: pulsed research reactors and nuclear devices.

The energetic charged particles which excite a nuclear pumped laser medium derive from three main sources: Compton-scattered electrons produced by energetic gamma rays; fission fragments (particles having atomic masses of about 95 and 140 amu, initial energies of 65 to 100 MeV, and initial charges of about +20e); and charged particles such as protons, tritons, and alpha particles produced in exothermic nuclear reactions involving absorption of a neutron. Laser action excited solely by fast neutrons has also been reported.

Table II contains basic data for several nuclear reactions which may be used to pump an RPL. The energy released varies over more than two orders of magnitude, with the largest energy release provided by fissioning of ^{235}U. Among the reactions in Table II, the fission reaction is unique in being able to produce more neutrons to sustain the nuclear reactions.

The charged particles are sufficiently energetic that excitation of a broad range of atomic and molecular states is energetically allowed. There is also, in general, no mechanism for selective excitation of only a few excited states. This selectivity of excitation, which has frequently been used to produce laser action, is not generally available in nuclear pumped lasers.

An important practical limitation is that each laser medium requires a minimum volumetric pump rate to begin lasing and a somewhat greater volumetric pump rate to convert the pumping power to laser light efficiently. One limitation of RPLs is that reactors tend to produce lower volumetric pump rates (no more than 3000 MW/m^3) than can be obtained with some conventional laser excitation methods. This limit is set by temperature limitations of the reactor structure itself. In the case of nuclear explosive devices used as pump sources, much higher pump rates can be obtained, since the survival of the nuclear device and its immediate surroundings is not required.

B. Potential Advantages as Large Lasers

For laser applications that require a very large laser or a laser remote from conventional electrical or chemical power sources, a nuclear pumped laser may be an effective alternative to more conventional lasers. Commercial, power reactors routinely operate for months at total fission powers of 3000 MW, with corresponding electrical power

TABLE II Nuclear Reactions for Pumping RPL

Target nucleus	Cross-section[a] (10^{-24} cm^2)	Energy release (MeV)	Particles released	Particle energy (MeV)	Range in 1 atm air (cm)
^{3}He	5330	0.76	^{1}H	0.6	1.0
			^{3}H	0.2	0.2
^{10}B	3838	2.79	^{4}He	1.8	1.0
			^{7}Li	1.0	0.5
^{235}U	585	165	^{93}Kr	100	2.3
			^{140}Xe[b]	65	1.8

[a] Thermal neutron energy.
[b] Representative nuclei from a broad spectrum of fission products.

outputs of over 1000 MW. Although a large RPL is not expected to resemble a power reactor in detail, conversion of a small fraction of such available fission energy to laser light would produce an extremely powerful, long-lived laser.

Such potential high output powers are possible because of the extremely compact and lightweight energy storage of the fissionable nucleus. Alternatively, the fuel for an electrical power source or the reactants for a powerful chemical laser must be stored in large tanks, in the form of highly corrosive or explosive liquids or gases with an energy storage per unit mass of about 1000 kJ/kg. By contrast, the energy stored in enriched uranium oxide (a dense, refractory ceramic) considerably exceeds 10^{10} kJ/kg. The uranium "fuel" of an RPL thus represents a huge, compact, stable, self-sufficient energy source, even allowing for incomplete uranium fuel utilization.

Another potential advantage of the RPL is that the energy is deposited in the laser medium based on its density, and the initial steps of the energy deposition process are largely independent of details of the laser medium condition (such as electron density in the medium or chemical composition of the medium).

The long absorption length of the neutron also allows the excited laser medium to be scaled to volumes of about 100 m^3, while remaining well within the bounds of current reactor technology. This scaling process involves replication of mechanical reactor substructures, such as plates of ceramic neutron-moderator material coated with very thin uranium coatings. Proliferation of complex, relatively fragile high-voltage apparatuses is not required, nor is it necessary to separate the laser gas from adjacent high-vacuum regions with thin, relatively fragile foils. Although the close spacing of the RPL structure presents problems with the extraction of a high-quality optical beam, several potential solutions are being studied. These unusual scaling laws may be of significant advantage in constructing very large lasers.

III. BASIC PHYSICS OF REACTOR-PUMPED LASERS

A. Nuclear Aspects

1. Reactor Characteristics

A nuclear reactor normally is designed to operate for many years at its maximum output (pulsed or steady-state); so the practical limitation on the lifetime of an RPL is typically related to other factors, such as limitations on coolant supply or degradation of laser optics.

Fast burst reactors (FBRs) have the advantage of producing high pump power in the RPL in a short pulse which does not overheat the laser medium. FBRs are also typically housed in laboratory facilities which allow convenient access to the volume around the reactor for setup of laser apparatuses. However, FBRs have very low pulse repetition rates because of the need to cool a large compact mass of reactor fuel between pulses.

Another relatively common research reactor, the pulsed TRIGA reactor, operates by pulsing from a very low initial power level using control rods. The pulsed TRIGA reactor can be pulsed more frequently because it typically is located at the bottom of a tank of water which acts as both coolant and shield. The neutron spectrum of a TRIGA reactor also has a much lower average energy, which reduces or eliminates the need for neutron-moderator material in the laser apparatus.

The practical laboratory sources of intense neutron fluxes are pulsed or steady-state reactors fueled primarily by uranium. Since each ^{235}U fission event releases about 168 MeV of fission fragment energy and about 2.4 neutrons, the energy cost per neutron created in a reactor is at least 70 MeV/neutron. In fact, because some of the neutrons are absorbed in the reactor fuel to sustain the reaction, the energy cost per neutron is greater than 70 MeV per neutron. Thus the use of low-energy-release reactions from Table II severely reduces the efficiency of a large RPL

system. However, for demonstration experiments any of the nonfission reactions in Table II may be used, in some cases with great practical advantages compared to use of the fission reaction.

There is typically some unavoidable waste energy deposited in the reactor structure itself, and the maximum temperature which it can withstand usually limits the obtainable pump rate in RPLs. For typical research reactors, pump rates of perhaps 3000 MW/m^3 can be obtained in pulses with minimum widths ranging from 0.2 to 10 msec. The neutron flux from high-flux steady-state reactors limits the laser pumping power per unit volume to about 10 MW/m^3.

2. Substrate Pumping

The least impact on the laser medium conditions is obtained by separating the target nucleus from the laser medium by using substrate pumping. In substrate pumping the target nuclei are chemically bound into a solid compound that is coated onto the surface of a substrate in contact with the laser gas. From Table II, the maximum path lengths of the charged particles in air at 1 atm pressure are typically a few centimeters. If substrate pumping is used, the pumping of the laser medium will be quite inhomogeneous if the source substrates are too far apart (so that the charged particles do not even reach the center of the laser medium, for example). A typical optimum separation between source substrates is approximately equal to the charged particle range. The energy deposition is also more uniform if the substrates are in the form of large parallel plates.

In all RPL experiments reported to date, a substrate has been used to support the target nuclei (with the exception of ^{3}He). This method has thus far been limited to excitation of gaseous laser media because of the extremely short range of charged particles in liquids and solids. The target nucleus compound is usually chosen for its chemical and mechanical stability and its adherence to the substrate material. The compound is deposited on the substrate in a thin layer, since the charged particles produced in the coating typically have ranges of a few milligrams per square centimeter of the coating material.

A portion of the energy of the charged particles is deposited in the coating and substrate material. This produces waste heat which does not contribute to laser excitation. For a given coating material and thickness, a constant fraction of the total energy created in the coating is emitted from the coating and is available to excite the laser medium. The output energy per unit area of coating increases as the slow neutron flux increases. For a typical uranium oxide coating of 1 μm thickness, the coating efficiency is about 32%. For neutron fluences at the edge of the core of the Sandia Laboratories ACRR reactor facility, volumetric energy depositions of 1.5 J/cm^3 and volumetric pump rates of 200 W/cm^3 are readily obtained for typical laser gases at about 1 atm pressure.

As the coating thickness increases, the coating efficiency decreases and the flux of energy from the coating approaches a constant value. In experiments it may be useful to maximize this energy flux to obtain the maximum excitation with a given fixed neutron flux. However, in a practical RPL system it would be important to maximize efficiency while maintaining the necessary pumping power and overall reactor characteristics desired.

3. Homogeneous Pumping

Dispersing the target nucleus throughout the laser medium (by choosing a gaseous compound containing the target nucleus and mixing it with the laser gas) normally produces the greatest spatial uniformity of pumping. Homogeneous pumping also uses essentially all the nuclear reaction energy at high pressures of the laser medium, since no energy is lost in a coating. However, the presence of the target nucleus in its host molecule may interfere with the processes which produce gain in the laser medium.

Experiments performed as part of the NASA RPL program have studied gaseous/homogeneous pumping to the greatest degree, culminating in a 1 kW laser (Table I). Development of a large, efficient RPL system based on ^{3}He is, however, limited by the very small energy release and the very large neutron absorption cross section of the ^{3}He reaction (Table II). If the ^{3}He partial pressure in the laser mixture is several atmospheres, the neutron absorption length shortens to a few centimeters, and the uniformity of pumping will be limited by the neutron absorption length itself.

If a laser medium tolerant of gaseous uranium compounds could be developed, homogeneous pumping would be preferable to any other scheme. The report of lasing at 2.65 μm in Xe in the presence of a small pressure of $^{235}UF_6$ is therefore of considerable interest (Table I). Researchers at Los Alamos National Laboratory have also proposed liquid RPL media containing uranium compounds.

Russian workers have also studied the potential for liquid homogeneous pumping in considerable detail, by measuring the optical effects of adding uranium compounds to liquid neodymium laser media, and the fluorescence efficiency of liquid laser media excited by fission fragments. However, reactor-pumped lasing in these liquid media has not yet been reported.

B. Laser Medium Aspects

The charged particles give up their kinetic energy to the laser medium by producing both ionization and excitation

of the laser medium. Low-energy secondary electrons produce further ionization and excitation in the medium. The number of ions and excited atoms/molecules can be estimated by using measured properties of the laser gas, such as the average energy required to create an electron–ion pair. However, there is no obvious mechanism to produce highly selective excitation directly. Thus, population inversion relies on selective kinetic and photophysical processes which preferentially populate certain states.

The excitation of high-lying excited states may stimulate high chemical reactivity of the laser medium, making it necessary to replace the laser gas periodically or scrub it.

For substrate pumping, the relatively short range of charged particles requires a separation between the substrates of a few centimeters. This results in large, slablike gain regions separated by opaque barriers. Extraction of a laser beam with good beam quality and high efficiency from this type of system is the most challenging aspect of large-system RPL design. Relatively little work has been published thus far to deal with this optical complexity, which must be weighed against the potential advantages of the simpler RPL excitation structure.

Finally, the relatively low pumping power densities obtainable with reactor pumping requires that the laser medium possess long lifetimes for retaining excitation. The CO laser was initially chosen as a likely candidate for reactor pumping, based on the long lifetimes of the upper energy levels of the laser transitions (∼30 msec). The Xe laser may also have a long effective lifetime for excitation because of energy recycling through a high-lying metastable level.

C. Medium Inhomogeneity

For the relatively long energy deposition times involved with typical pulsed reactors, the laser medium can move in response to spatial variations in the energy deposition. This gives rise to nonuniformities in the laser medium density and gradients of the index of refraction which cause distortion of the laser beam (tilt, focusing, defocusing, or higher order aberrations). To produce a very high-quality laser beam, these distortions must be corrected.

In general, there are two possible causes of inhomogeneous pumping: charged-particle-range effects and effects from absorption of neutrons by the target nucleus. Charged-particle-range effects are a problem with substrate pumping, and neutron-absorption effects are a problem with homogeneous pumping.

For substrate pumping (assuming efficient usage of the fission fragment energy), the spatial scale of the nonuniform energy deposition is of the order of the separation between adjacent substrates. If the substrates are in the form of large parallel plates, a nearly one-dimensional density variation results. Wavefront distortion measurements have been reported for this excitation geometry. For intense pulsed excitation, the density can vary by as much as a factor of two, producing an effective focal length of less than 1 m for a cell only 30 cm long. Such time-dependent focusing can have significant effects on the modal stability of a laser. In fact, resonator stability transitions have been observed in the atomic Xe laser when energy depositions approach 1 J/cm^3.

United States and Russian workers have studied the resulting optical distortions and the associated effects on the lasing process. Combined wall and volume pumping has also been suggested as a means of reducing this problem [Petra (1998)].

D. Contrasts with Conventional Lasers

The RPL differs from electrically, optically, or chemically excited lasers in several respects.

1. The maximum excitation power per unit volume in the laser medium is lower than the comparable maximum for electrically excited lasers (assuming that specialized RPL designs based on existing reactor technology are used). However, the energy deposition per unit volume of laser medium is at least comparable to that obtained with electrical or electron-beam excitation.
2. The geometry of the laser excitation region is complex if substrate pumping is used, which demands accurate relative phasing of beams from many independent gain regions to produce a single coherent laser beam.
3. If substrate pumping is used, the excitation process itself gives rise to variations in laser medium density which severely affect beam quality unless the resulting laser beam aberrations are corrected.
4. The laser optics must operate reliably in a relatively severe radiation environment, which consists primarily of high-energy gamma rays and fast neutrons.
5. The construction of the laser typically involves robust mechanical apparatus rather than electrical or chemical apparatus.

IV. REACTOR-PUMPED LASERS

A. Technological Progress

About 50 RPLs have been discovered since the first report of an RPL in 1975. These are summarized in review articles mentioned in the bibliography.

In 1977 several concepts for constructing large RPL systems were analyzed and reported by Sandia workers. The concepts were based solely on substrate pumping and described specialized nuclear reactor designs which could

excite large volumes of laser gas at efficiencies approaching the coating efficiency. These relatively efficient system concepts tended to resemble the TRIGA or ACRR reactors more than the FBR. These initial calculations showed that with a high intrinsic laser efficiency, perhaps 10% of the total reactor energy might be obtained as laser light, using current or near-term reactor technology. However, the problem of extracting a high-quality optical beam from the complex reactor structure was not addressed.

In recent years, Sandia workers have measured the small signal gain and efficiency of the more efficient noble-gas RPL excited by fission fragments. The influence of gas temperature on the efficiency of the noble-gas RPL has been measured (Hebner, 1995). Possible applications of large, steady-state RPL have been investigated, such as power beaming to space.

Experimental reseach directed toward discovery of new RPLs has continued at the University of Illinois. System studies of possible large-scale RPL applications and detailed measurements of radiation-induced absorption in optical materials have been reported. Concepts for using nuclear-pumped flashlamps for industrial applications and for laser excitation have been explored [Prelas (1995)], and the first nuclear-flashlamp-pumped laser (iodine at 1.31 micron) has been reported.

The NASA program at Langley Research Center was broad and productive: New RPLs were discovered, homogeneous pumping with ^{3}He was developed, the volume limits of RPLs were studied, excitation processes were investigated theoretically, and a reactor pumped amplifier was demonstrated.

In 1979 researchers in the Russia reported near-infrared RPLs using mixtures of noble gases pumped at very low volumetric pump rates, with surprisingly high efficiencies and high energy outputs (Table I). The combination of relatively short wavelength, high efficiency, relatively high pressure, and stable laser gases may make these lasers attractive for RPL applications.

Moderate-efficiency lasing was observed by Russian researchers in Cd vapor at 0.442 to 0.54 μm wavelength using very low volumetric pump rates. Thus, it is possible that high-power visible RPLs with reasonable efficiency can be developed.

Beginning in the early 1990s, Russian workers began to report on extensive RPL programs at the VNIIEF, VNIITF, and IPPE laboratories. These research programs covered a wide range of technologies that would be beneficial in developing practical RPL systems. New gaseous RPL laser lines at wavelengths that transmit well through the atmosphere were sought, and their efficiencies measured. The effects of uranium-compound addition and fission fragment excitation of liquid neodymium laser media were experimentally investigated; the aprotic neodymium liquid laser media which had been optically pumped with reasonable efficiency were studied in detail (Seregin, 1999). The nature of optical distortions in a gaseous flowing RPL was investigated (Borovkov, 1995). Designs for large RPL systems were proposed (Dyachenko, 1996), and system concepts for producing electrical power were considered. Short-wavelength RPLs were reported at 0.308-μm wavelength in XeCl*, and at 0.391-μm on wavelength in N_2^+. RPL scale-up experiments using a specialized double fast-burst reactor (BARS-6) were performed (Kononov, 1997), and results on the neutronic properties of a large RPL experiment were reported (Gulevich, 1998). A flowing RPL using mixtures of noble gases, lasing for up to 2 sec was experimentally demonstrated (Sinyanskii, 1999); this represents a significant step toward a large steady-state RPL at a convenient wavelength. Many of these results are described in recent review articles (Karelin, 1997); Sinyanskii, 1998; Magda, 1998a; Magda, 1998b). Finally, the usefulness of pulsed electron beams as simulation sources for investigating potential RPL media was studied (Fedenev, 1998).

Chinese researchers have reported lasing in noble gas mixtures excited by a fast-burst reactor, but have not followed up this work with publications in English. Japanese workers have investigated the usefulness of the 0.585-μm RPL line as a diagnostic for neutron flux in nuclear reactors (Yamanaka, 1997).

Finally, alternate forms of substrate pumping were considered in some detail. The energy escape fraction is largest from small spheres of uranium metal or oxide. Such spheres cannot be incorporated directly into the laser medium because of the optical scattering loss they introduce, even if they were overcoated with a thin reflective coating. However, concepts have been developed to use such reflective aerosols to pump a fluorescer medium, which produces intense incoherent light usable for optically pumping a laser medium.

B. Radiation Effects on Optics

One concern in early research on RPLs was the question of degradation of optics by the intense radiation field of the reactor environment. The radiation environment of the reactor consists of broad spectra of both neutrons and gamma rays ranging in energy up to several MeV.

In early experiments at Sandia in the FBR environment, it was found that no visible optical damage occurred to several convenient window materials, such as optical-grade NaCl, BaF_2, and CaF_2. Metallic mirrors showed almost no measurable change in reflectivity, perhaps because they are already highly conductive. In more recent experiments, dielectric-coated mirrors made of elements with low neutron absorption have also demonstrated very

little change in reflectivity in intense pulse radiation fields. Schlieren-grade fused silica also displays relatively low radiation-induced absorption for visible and near-infrared wavelengths, and is thus a convenient and versatile RPL optical material.

However, materials containing elements such as ^{10}B or ^{6}Li (such as borosilicate glass) should be avoided in optical components because of their large neutron-absorption cross sections. In these nuclei, absorption of a neutron creates tracks of intense ionization in the host material and atomic displacements which can act as defect centers for optical absorptions. Therefore, a basic step in optical design for RPLs is to avoid neutron-absorbing elements in window materials, coating materials, and substrate materials for mirrors.

Optical components can be protected from bombardment by low-energy neutrons by relatively thin sheets of materials such as Cd metal or B_4C suspensions in plastics. Very energetic neutrons cause damage by displacing lattice atoms in elastic collisions. A low-mass recoil atom may have enough energy to create ionization tracks in the material. High-energy neutrons can be moderated to low energies (for absorption in Cd or B) by scattering from shielding materials containing high densities of hydrogen (such as high-density polyethylene).

Gamma rays produce energetic Compton-scattered electrons which may cause significant damage even in optical materials which do not contain neutron-absorbing nuclei. The energetic Compton electrons degrade in energy by producing ionization in the optical material so that a weak transient conductivity is induced by the radiation. Complete shielding of the gamma rays requires thick layers of high-density materials such as Pb, but this is usually not practical in the vicinity of the laser apparatus.

C. Development Toward Applications

The unique characteristics of nuclear pumped lasers include the following: Essentially infinite energy storage; relatively straightforward scaling of the volume of excited laser medium; extremely high total excitation power with current reactor technology; no requirement for pulsed power systems or high-voltage electrical power; and robust mechanical construction.

However, these applications must be realized in the presence of the unique disadvantages of nuclear pumping: Low pump power per unit volume (for RPLs, not NDPLs); intense radiation fields; extraction of the laser beam from the complex reactor structure (for substrate pumping); and the health, safety, and proliferation concerns of nuclear technology in general.

Lasers with moderate total energies have found numerous practical applications. It may be that as applications for very large lasers emerge, the relatively robust "mechanical" nature of NPLs will have important system advantages.

SEE ALSO THE FOLLOWING ARTICLES

ATOMIC AND MOLECULAR COLLISIONS • LASERS • LASERS, EXCIMER

BIBLIOGRAPHY

Bonyushkin, E. K., *et al*. (1996). "Ignition Experiment Design Based on Gamma-Pumping Gas Lasers," *AIP Conf. Proc.* (*USA*) **369,** 2, 906.

Borovkov, V. V., *et al*. (1995). "Characteristics of Gas Dynamics of Flow Lasers Excited by Fission Fragments of Uranium Nuclei," *Quantum Electronics* **25,** 1149.

Dyachenko, P. P. (1996). "Concept of Nuclear Reactor Pumped Laser for ICF," *AIP Conf. Proc.* (*USA*) **369,** 2, 894.

Fedenev, A. V., and Tarasenko, V. F. (1998). "Simulation of NPL in Experiments with E-Beam Pumping," *Laser and Particle Beams* **16,** 327.

Gulevich, A. V., *et al*. (1998). "Fast Two-Core Pulse Reactor System with a Thermal Subcritical Laser Module: Analysis of Startup Results," *Trans. Am. Nucl. Soc.* (*USA*) **78,** 193.

Hebner, G. A. (1995). "Gas Temperature Dependent Output of the Atomic Argon and Xenon Lasers," *IEEE J. Quant. Electron.* **31,** 1626.

Karelin, A. V., *et al*. (1997). "Nuclear-Pumped Lasers and Physical Problems in Constructing a Reactor-Laser," *Quantum Electronics* **27,** 375.

Kononov, V. N., et al. (1997). "Nuclear Pumped Lasing Experiments on Fast Burst Reactor BARS-6," *ICENES '96 Proceedings*, **1,** 336.

Magda, E. P. (1998a). "Review: Nuclear Pumped Lasers, Problems and Perspectives," *Proc. SPIE—Int. Soc. Opt. Eng.* (*USA*) **3343** (1–2), 158.

Magda, E. P. (1998b). "Review: Powerful Nuclear Pumped Lasers," *Proc. SPIE—Int. Soc. Opt. Eng.* (*USA*) **3574,** 93.

Petra, M., and Miley, G. H. (1998). "Thermal Blooming in NPLs," *Proc. SPIE—Intl. Soc. Opt. Eng.* (*USA*) **3343** (1–2), 242.

Prelas, M. A. (1995). "Lasers with Combined Nuclear Pumping," *Laser Part. Beams* **13,** 351.

Seregin, A. A., *et al.* (1999). "Model of a Nuclear Pumped Liquid Laser," *Quantum Electronics* **29,** 406.

Sinyanskii, A. A., and Krivonosov, V. N. (1998). "Powerful Nuclear-Optical Converters for Applications in Scientific Investigations and National Economy," *Proc. SPIE—Int. Soc. Opt. Eng.* (*USA*) **3574,** 440.

Sinyanskii, A. A., and Melnikov, S. P. (1999). "Development of Continuous Nuclear-Laser Setups in VNIIEF," *Proc. SPIE—Int. Soc. Opt. Eng.* (*USA*) **3686,** 43.

Yamanaka, N., *et al*. (1997). "Analysis of a Nuclear Pumped Lasing Mechanism for Application to Nuclear Radiation Measurements," *Nuclear Instruments and Methods in Physics Research, Section A* **389,** 513.

Lasers, Optical Fiber

Graham E. Town
University of Sydney

I. Fiber Laser Fundamentals
II. Fiber Laser Components
III. CW Fiber Lasers
IV. Pulsed Fiber Lasers
V. Other Fiber Lasers
VI. Applications of Fiber Lasers
VII. Future Developments

GLOSSARY

Birefringence Polarization dependence of the propagation constant.
Chirp Variation of carrier frequency, or wavelength, with time.
Coherent Highly correlated.
Dispersion Wavelength dependence of the propagation constant.
Gain (or loss) coefficient The rate of change of amplitude of a wave with propagation distance.
Grating Periodic variation in refractive index or wave impedance, causing wavelength-dependent coupling between propagating waves, and often used as an optical filter.
Intensity noise Random variation in optical intensity, relative to an expected intensity.
Jitter Random variation in pulse arrival time, relative to an expected arrival time.
Loop mirror All-fiber Sagnac interferometer which functions as a mirror.
Mode Fundamental field distribution satisfying the wave equation, subject to boundary conditions.
Mode locked laser Optical oscillator in which the longitudinal cavity modes are locked in phase, generating a periodic (i.e., pulsed) optical output.
Optical amplifier A device with optical gain.
Optical oscillator An optical system with gain and feedback producing a coherent optical output (often referred to as a laser).
Propagation constant The rate of change of phase of a wave with propagation distance.
Pump Excitation used to induce gain in an optical amplifier.
Quantum efficiency Ratio of the energy of emitted photons to the energy of absorbed photons.
Saturable absorber Device with intensity-dependent transmissivity or reflectivity.
Self-phase modulation (SPM) Self-induced change in phase of an optical wave, caused by intensity dependence of the propagation constant or refractive index.
Slope efficiency The efficiency with which pump power above threshold is converted to output signal power.
Transform limited pulse Unchirped optical pulse.
Waveguide Structure used to confine electromagnetic

fields in two spatial dimensions, allowing wave propagation in the third dimension.

EARLY IN THE DEVELOPMENT of lasers it was realized that there would be significant advantages if the lasing action could be confined in an optical waveguide. Experiments in the mid-1960s showed great promise, demonstrating optical gain and lasing in multimode silica optical fibers, but it was not until the mid-1980s that the benefits became manifest through the convergence of glass laser technology with low-loss single mode fiber fabrication technology. The subsequent development of fiber lasers has been greatly assisted by the development of associated technologies, such as in-fiber Bragg grating filters, and high-power diode laser pump sources. Optical fiber lasers are now an important laser technology offering some unique capabilities, and an attractive alternative to other laser technologies in many applications.

The primary difference between fiber lasers and bulk lasers is the confinement of the lasing action in an optical fiber waveguide. In contrast to optical amplification in bulk media where the pump is focused into the active medium, in fiber amplifiers the pump spot size and interaction length are independent; the fiber waveguide allows a very high pump intensity to be maintained over a very long interaction length. In any optical amplifier the gain coefficient scales with the pump intensity, and the total gain and efficiency with absorbed pump power. Consequently fiber amplifiers may be realized with very high total gain using low pump power; 30–40 dB gain is typical. With feedback applied, fiber lasers or optical oscillators result with low threshold and high quantum efficiency.

The flexibility in laser design provided by the waveguide structure also allows fiber lasers to be realized using nonoptimal components, such as gain media with weak pump absorption, on transitions with low quantum efficiency, in cavities with high loss, etc. For example, in rare earth doped silica fibers a large number of lasing transitions between 0.38 and 3.9 μm are possible using a wide variety of pump wavelengths. A number of useful pump wavelengths are available from semiconductor diode lasers, allowing the fabrication of compact all-solid-state lasers in which the nonideal characteristics of semiconductor lasers are converted into the almost ideal characteristics of fiber lasers. Fiber lasers can generate higher output power with superior noise performance compared to semiconductor lasers. Furthermore, silica-based fiber lasers generally have broad homogeneous linewidths, and so are well suited to tunable and/or sub-picosecond pulse generation. Additional advantages of diode-pumped fiber lasers are their compact and mechanically robust construction, their diffraction-limited output, and their efficient heat dissipation. Lastly, they are readily coupled with low loss to other optical fibers and fiber systems.

The following sections provide a comprehensive review of optical fiber laser technology, commencing in Section I with a brief introduction to fiber laser fundamentals. Section II describes the components often used in fiber lasers, emphasizing those properties relevant to the performance of fiber lasers. Sections III–V review the typical performance characteristics of the main classes of fiber laser, covering continuous wave lasers, pulsed lasers, and other fiber laser sources. The article concludes in Sections VI and VII with a brief review of some fiber laser applications, and an overview of past and possible future developments in fiber laser technology.

I. FIBER LASER FUNDAMENTALS

Optical oscillators, commonly referred to as lasers, are optical feedback systems in which two fundamental conditions must be satisfied for stable oscillation at the lasing wavelength; the round-trip gain must be unity, and the round-trip phase must be an integer multiple of 2π radians. Consequently there are two fundamental elements in any optical fiber laser, i.e., a source of optical gain, and an optical feedback path; either or both may be in fiber. For the purposes of this article optical fiber lasers will be defined as lasers in which the optical gain element is an optical fiber amplifier. Extra elements may be added to facilitate coupling of optical energy in and out of the laser cavity, or to control the temporal and spectral characteristics of the laser.

A simple all-fiber laser is shown schematically in Fig. 1. A pump laser provides energy to the fiber amplifier. For the laser to exceed threshold and generate a coherent optical output, the amplifier must produce sufficient gain to overcome cavity losses, including the output coupling loss. The absorbed pump power required to reach threshold is approximately

$$P_{\text{th}} = \alpha_{\text{cav}} a_p h \nu_{\text{p}} / (\sigma_{\text{s}} \tau_1),$$

in which $\alpha_{\text{cav}} = \alpha_{\text{int}} - \ln(1 - T)$ and represents the cavity's total internal and outcoupling loss per round trip, a_p the mode area of the pump, h is Plank's constant, ν_{p} the pump frequency, σ_{s} the transition cross-section, and τ_1 the fluorescence lifetime or population relaxation time. In fiber lasers, the major loss is often the wavelength nonselective outcoupling of a fraction, T, of the optical power circulating in the laser cavity. In doped fiber lasers the fluorescence lifetimes are often long, and the pump mode area small. The latter factors result in fiber lasers requiring

potentially very low pump power (e.g., less than a few milliwatts) to reach threshold.

In theory, the lasing wavelength will be that which receives the most gain and therefore is the first to achieve lossless propagation around the cavity, i.e., unity loop-gain, and which simultaneously satisfies the loop-phase condition, i.e., an integral number of wavelengths around the loop. Fiber lasers are often long (i.e., meters), and consequently there are usually many closely spaced wavelengths capable of satisfying the loop-phase condition.

As with all lasers, once sufficient pump power is applied for the laser to reach threshold, further increase in pump power does not increase the amplifier gain, but rather is converted to output signal at the lasing wavelength, with some energy lost as heat and noise. The efficiency with which pump power above threshold is converted to output signal power is called the slope efficiency, approximately given by

$$\eta_s = P_{out}/(P_{abs} - P_{th}) = T/\alpha_{cav} \,.\, a_s/a_p \,.\, \nu_s/\nu_p,$$

in which P_{out} is the signal output power, P_{abs} is the absorbed pump power, and T the output coupling ratio. The slope efficiency is readily maximized in fiber lasers by minimizing the outcoupling and cavity losses.

The laser shown in Fig. 1 contains only a few elements. While it would work in practice, better performance is usually obtained by adding other elements to the cavity to control the temporal, spectral, and directional characteristics of the light generated. In the following section the characteristics of typical components used in fiber lasers are reviewed, and their effect on laser performance highlighted.

II. FIBER LASER COMPONENTS

This section provides a brief review of the types and characteristics of optical fiber amplifiers, step index fiber waveguides, and fiber components and often used fiber lasers. The effect of the various fiber component characteristics on fiber laser performance is highlighted.

A. Optical Fiber Amplifiers

One or more of the following three physical processes are commonly used to produce optical gain in silica-based optical fibers: (1) stimulated emission from excited rare earth ion dopants; (2) stimulated Raman scattering; or (3) stimulated Brillouin scattering. Optical gain may also be produced in polymer fibers by stimulated emission from excited dye dopants, however, at the time of writing polymer fiber amplifiers were in the relatively early stages of development, and so will not be considered further. Detailed information regarding optical fiber amplifiers may be found in the article entitled "Lasers, Solid State" (found in this volume), so the following sections serve to review and highlight characteristics of fiber amplifiers which are important to the performance of optical fiber oscillators.

1. Rare-Earth-Doped Fiber Amplifiers

Rare-earth-doped silicate fiber amplifiers were the first fiber lasers demonstrated, and are the now the most common class of optical amplifier used in fiber lasers. Lanthanide ions are readily incorporated into silicate glass fibers using well-established metal-oxide chemical vapor deposition (MOCVD) and solution doping methods, providing a convenient source of bidirectional and almost polarization-independent optical gain, readily interfaced to other optical fibers and optical pump sources with low loss. The performance characteristics of rare earth doped fibers may be tailored by modifying the host glass composition, for example, using codopants to improve the rare earth solubility or broaden the lasing transitions. The fiber design may also be optimized for specific applications, such as low noise or high power capability. The key characteristics are summarized in the following sections.

a. Lasing wavelengths. A wide variety of lasing wavelengths are accessible using rare earth doped silicate fibers, as shown in Fig. 2. Most activity has focused on rare earth dopants in silicate fibers with lasing transitions in the low-loss windows of optical fiber communication systems (i.e., erbium around 1.55 μm, neodymium around 1.06 μm, and 1.35 μm, and praseodymium around

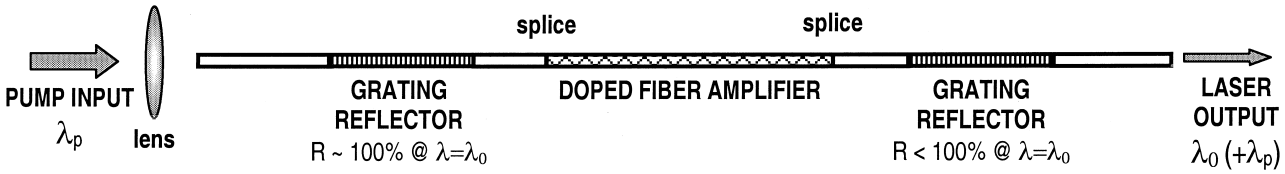


FIGURE 1 Schematic of a simple all-fiber laser oscillator. Two Bragg grating reflectors form a standing wave cavity containing an optical fiber amplifier. The optical pump is coupled into one end of the fiber cavity through a Bragg grating, which reflects only the lasing wavelength. Similarly, the laser output may be taken from either end of the cavity, through a Bragg grating with less than 100% reflectivity at the lasing wavelength.

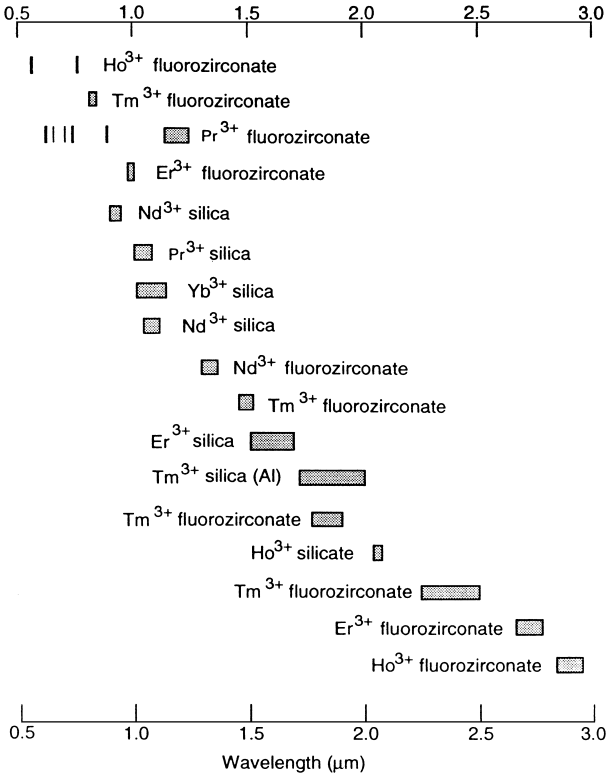


FIGURE 2 Reported tuning ranges of lanthanide-ion glass lasers [From Weber, M. J. (1999). "Handbook of Laser Wavelengths," CRC Press, Boca Raton, FL.]

1.3 μm). Nevertheless, interest in other dopants with other lasing transitions is increasing, e.g., for applications in sensing. Also, lasing wavelengths at both shorter and longer wavelengths than those shown in Fig. 2 are possible in low-phonon energy glasses (see Section III.D).

b. Pump sources. Laser diode pump sources are available at a number of convenient wavelengths for exciting rare earth ions. The absorption bands for rare earth ions in glass can be relatively wide, giving some leeway in the exact choice of pump wavelength, though varying the exact pump wavelength may also vary the gain spectrum. Erbium may be pumped as a 3-level system around 810, 980, or 1480 nm. While higher power pump sources are available around 810 nm, the latter two pump wavelengths are preferable as they do not suffer from excited state absorption, which limits the lasing efficiency at 1.55 μm. Neodymium may also be pumped around 810 nm, but as a 4-level system with lasing transitions around 1060 and 1350 nm, with the latter affected by excited state absorption. Ytterbium, commonly used as a codopant with erbium or thulium, has a strong absorption peak at 975 nm, and a very broad absorption band of $\pm$100 nm around 930 nm.

c. Gain bandwidth and homogeneity. Rare earth doped silicate glass amplifiers typically have large gain bandwidths (i.e., tens of nanometers), a consequence of the amorphous glass host and the presence of codopants such as germanium and aluminium. The broadband gain is provided by a mixture of homogeneously and inhomogeneously broadened emission transitions of the dopant ions, depending on the glass composition. For 3-level systems, such as in erbium fiber amplifiers, the gain spectrum also depends on the pump power and amplifier length. In general the homogeneity of the gain has a significant effect on the performance of fiber amplifiers in the saturated regime, as in fiber lasers.

In erbium-doped germanosilicate fiber, the approximate homogeneous and inhomogeneous linewidths at room temperature are, respectively, $\lambda_{\text{hom}} = 4$ nm and $\lambda_{\text{inh}} = 8$ nm, while in aluminosilicate fibers they are $\lambda_{\text{hom}} = \lambda_{\text{inh}} = 10$ nm. The linewidths in neodymium-doped germanosilicate glass are similar, but unlike erbium are usually dominated by the inhomogeneous component. Phosphorus codoping reduces the homogeneous linewidths by about a factor of two with respect to the latter figures. The inhomogenous linewidths are mostly independent of temperature, however, the homogeneous linewidths reduce with temperature to about 1 nm at 77 K.

d. Gain efficiency. In fiber amplifiers with only moderate dopant ion concentrations (i.e., to avoid ion-ion interactions), pumped at wavelengths which avoid pump upconversion, and which are long enough to absorb most of the pump power, then unsaturated gains of up to a few decibels per milliwatt of pump power may be achieved. The maximum achievable gain is generally limited by the buildup of amplified spontaneous emission noise, which can saturate the amplifier.

e. Gain dynamics. The gain dynamics in homogeneously broadened doped-fiber amplifiers depends primarily on two time constants, i.e., the fluorescence or excited state lifetime, τ_1, and the dipole relaxation or dephasing time, τ_2. The fluorescence time depends on the dopant and host material; for erbium in silicate glass $\tau_1 \sim 12$ ms, and for neodymium $\tau_1 \sim 500$ μs. The dipole relaxation time is typically 100 fs. Amplifier performance is largely independent of signal modulation period τ, for $\tau_2 < \tau < \tau_1$. For long pulses the gain may become time-dependent, and for very short pulses gain dispersion may become significant, but usually doped-fiber amplifiers simply respond to average signal power, even when saturated.

f. Saturation power. The signal saturation power of rare-earth-doped-fiber amplifiers, and hence maximum

output power of fiber lasers, depends primarily on the pump power coupled into the fiber amplifier. The output power of single mode semiconductor pump lasers is limited, however, multimode diode arrays can be assembled with total output powers exceeding hundreds of watts. While the latter cannot be coupled efficiently into a single mode fiber core, they can be coupled into a multimode cladding surrounding the amplifier core. Fiber amplifiers with a second cladding for pump guidance are referred to as cladding-pumped fiber lasers. In such fibers the pump absorption length is increased because of the poor overlap between pump and doped core, however, the total pump power is increased, resulting in an increased saturation power. For example, cladding-pumped fiber lasers have been realized with continuous wave (CW) output power over 100 W. Fiber nonlinearities such as stimulated Raman scattering (SRS) can also limit the maximum power level achievable in fiber amplifiers. Special fiber designs which maximize the fiber core area, and hence minimize the optical intensity, can alleviate problems associated with fiber nonlinearity, however, the advantage of such techniques is limited by trade-offs with increased fiber bend loss.

g. Noise performance. The noise figure of an optical fiber amplifier (i.e., the ratio of the signal-to-noise ratios at input and output) ultimately depends on the inversion level that can be achieved in the fiber amplifier. For example, the noise figure of erbium-doped-fiber amplifiers pumped at 980 nm can approach the quantum limit of 3 dB. In laser oscillators the inversion, and hence the spontaneous emission noise within the homogeneous linewidth, are clamped at their threshold levels. Furthemore, long fluorescence times reduce the effect of external noise sources (e.g., pump intensity noise) at high frequencies.

2. Stimulated Raman Amplifiers

Stimulated Raman scattering is a nonlinear effect in which energy is coupled from an optical pump beam to longer wavelengths via vibrational modes or thermal phonons of the molecules in the glass medium. In some situations SRS can limit the amount of power that can propagate without unwanted loss in optical fibers and fiber amplifiers. For SRS to produce gain, the pump power must be greater than a threshold value, i.e., about 1 W in typical optical fibers. The Raman gain spectrum depends on the fiber composition and temperature, but in general the Raman gain increases approximately linearly from zero at the pump wavelength to a peak at about 450/cm (i.e., approximately 50 nm at 1.06 μm) on the long wavelength side of the pump. Because of the amorphous nature of glass the gain spectrum is broad, typically 40 nm, and comparable to doped-fiber amplifiers.

Stimulated Raman scattering may occur in either the forward or backward direction relative to the pump wave, therefore the gain produced is bidirectional. Raman gain is polarization-sensitive, however, this is rarely a problem in practice, as the relative polarization state of the pump and signal can vary rapidly during propagation in standard fibers, resulting in the same average gain for both orthogonal polarizations, with a 3-dB reduction in the theoretical gain. Fiber Raman amplifiers potentially have a number of desirable features, including low noise figure (about 3 dB), high gain (e.g., $>$40 dB), and high output power capability (e.g., $>$1 W). The primary advantage of SRS over the other types of optical amplifier is that broadband optical gain may be generated at any wavelength, provided a suitable pump source is available.

The primary limitation of Raman fiber amplifiers is the relatively low gain efficiency (typically 10 dB/W/km), requiring high pump powers and/or long interaction lengths to achieve significant gain. Also, pump noise can cause variations in gain and modulate the signal. Consequently, fiber Raman amplifiers have largely been replaced by doped-fiber amplifiers.

3. Stimulated Brillouin Amplifiers

Stimulated Brillouin scattering (SBS) is similar to SRS in that energy is transferred from an optical pump beam to longer wavelengths through interaction with the glass medium, except that acoustic phonons are involved, and hence the frequency shift is small, about 11 GHz, and the bandwidth very small, typically 50 MHz. Phasematching conditions for SBS result in unidirectional gain, i.e., in the backward direction relative to the pump, however, the gain is polarization-independent. Because of the narrow bandwidth, the gain efficiency is large (e.g., 6 dB/mW in a typical fiber), and the threshold pump power low (e.g., 1 mW), however, long interaction lengths are required to achieve large gain, the noise figure is poor, and the saturated output power low (e.g., 1 mW). Like SRS, SBS can limit the power that can propagate in an optical fiber without unwanted loss or onset of lasing, but is generally only a problem for sources with narrow linewidth, i.e., comparable to the Brillouin gain linewidth.

B. Optical Fibers

Detailed information on optical fiber waveguides may be found in the article "Optical Fibers, Fabrication and Application" (Volume 11). The following sections therefore summarize those properties of optical fibers which can have a significant impact on the performance of fiber lasers. The characteristics may be classed as linear, nonlinear, and other environmental and material-dependent effects.

1. Linear Properties of Optical Fiber Waveguides

Optical fibers are highly transparent over specific wavelength ranges in the near infrared, a property which makes them attractive for applications such as telecommunication. The linear losses are typically in the range 0.2–1 dB/km, i.e., insignificant compared to losses incurred in coupling to and from fibers, or the gain that can be provided by optical amplifiers, hence the losses in passive optical fibers may usually be ignored in determining the performance of fiber lasers.

Optical fibers used in lasers are usually designed to guide a single mode. Step index fibers are single mode if the normalized frequency parameter satisfies the following constraint,

$$V = 2\pi \frac{\rho}{\lambda} \sqrt{\left(n_{co}^2 - n_{cl}^2\right)} \leq 2.405$$

in which ρ is the core radius, n_{co} and n_{cl} the core and cladding refractive indices, respectively. The intensity distribution of the fundamental guided mode is approximately Gaussian, with an RMS width or spot size which depends on wavelength. At short wavelengths (i.e., close to $V = 2.4$), the intensity distribution is well confined within the core, and consequently the effective index of the guided mode is close to the core refractive index. At long wavelengths the guided mode is only weakly bound to the core, with the intensity distribution extending well into the cladding, and the effective index is then close to the cladding refractive index. The spot size, or mode field radius, w_s, for $1.2 < V < 3$ is well approximated by

$$w_s \sim \rho\,(0.65 + 1.619\,V^{-3/2} + 2.879\,V^{-6}).$$

The choice of parameters made in designing an optical fiber affects more than the number of propagating modes; trade-offs between dispersion and losses due to bends and splices are also involved. The bend losses in optical fibers with $V > 1$ are usually negligible over short lengths, hence in the following we concentrate on dispersion.

The dependence of spot size and effective refractive index of the guided mode on wavelength results in waveguide dispersion, i.e., different guided wavelengths travel at different velocities within the fiber. The total dispersion experienced by light guided in an optical fiber is determined by the sum of material and waveguide dispersion, and is dominated by material dispersion except at wavelengths where the latter is small. Dispersion causes temporal broadening of an optical pulse during propagation, and the development of a "chirp" in the optical carrier frequency across the pulse. The chirp may be positive or negative, depending on the sign of the dispersion, and is often an important factor in the performance of pulsed fiber lasers. The dispersion length, i.e., a characteristic length over which dispersive effects become significant for pulse evolution, may be defined as $L_D = \tau_o^2/|\beta_2|$, in which τ_o is the initial RMS pulse width in intensity, and β_2 the group velocity dispersion.

In standard step index fibers the material and waveguide dispersions cancel at about 1.3 μm. At longer wavelengths the total dispersion is usually negative, and referred to as anomalous, (i.e., if $\beta_2 < 0$, shorter wavelength components in an optical signal travel faster than longer wavelengths). At shorter wavelengths the total dispersion is of opposite sign, and referred to as normal. Typical values of dispersion in commonly available step index fibers are $\beta_2(1\,\mu\text{m}) = +25\,\text{ps}^2/\text{km}$, and $\beta_2(1.55\,\mu\text{m}) = -20\,\text{ps}^2/\text{km}$. Specially designed fibers can have significantly modified dispersion characteristics. For example, dispersion-shifted fiber may have a zero-dispersion wavelength of 1.55 μm, and dispersion-flattened fiber typically has small dispersion over a wide range of wavelengths, e.g., from 1.3 to 1.55 μm.

Birefringence is another linear property of optical fiber waveguides that often plays an important role in fiber laser performance. Birefringence may be defined as the dependence of the propagation constant of a guided mode on its state of polarization. The magnitude of birefringence may be characterized by the beatlength between orthogonal polarization states, or the period over which the state of polarization evolves, i.e., $L_B = \lambda/B$, where for linear birefringence $B = |n_x - n_y|$. Birefringence may be caused by noncircular symmetry of the fiber core (i.e., form birefringence), and/or an asymmetrical stress across the fiber (through the stress-optic effect). The latter type of birefringence may be induced by squeezing or bending an optical fiber. Bending a silica fiber on radius R typically causes birefringence $B = -0.093(\rho/R)^2$. For example, bending a 125-μm diameter fiber on a radius of 2.5 cm induces a birefringence with beatlength $L_B \sim 2.5$ m. Highly birefringent fibers (e.g., with $L_B < 5$ mm) are often referred to as polarization maintaining (PM) fibers, because bends have little effect on their birefringence or the polarization of light propagating through the fiber. Fiber lasers constructed from highly birefringent fibers are less sensitive to bending and other stress-inducing perturbations in their environment.

2. Nonlinear Properties of Optical Fiber Waveguides

While many optical materials, including silica, are commonly regarded as linear media, their response to light is in fact slightly nonlinear, i.e., dependent on the local light intensity. Nonlinearity becomes significant in optical fibers if the guided light has a high intensity, or propagates over large distances; either situation is quite possible in optical fibers, due to their small guiding cross section and

low loss. The main nonlinear response is the Kerr effect, which is responsible for effects such as four-wave mixing, cross-phase modulation (CPM), and self-phase modulation (SPM). The latter effects may be regarded as due to an intensity-dependent component of the refractive index, i.e., $n = n_0 + n_2 I$, in which n_2 is the nonlinear index coefficient. In silica fibers, the nonlinear index is typically $n_2 = 3 \times 10^{-20}$ m^2/W.

The nonlinear index results in the accumulation of an intensity-dependent phase during propagation, $\phi_{\text{SPM}} = (2\pi/\lambda)\, n_2 I z$. For optical pulses the intensity is a function of time, and hence different parts of the pulse will accumulate different nonlinear phase during propagation, resulting in a chirp developing in the optical carrier across the pulse. From the nonlinear index of silica, a nonlinear coefficient for optical fiber waveguides may be defined, $\gamma = (n_2\omega_0)/(cA_{eff})$, in which ω_0 is the optical frequency, and $A_{eff} = \pi w_s^2$ is the effective area of the core, where w_s is the spot size defined earlier. In typical step index optical fibers the nonlinear coefficient is in the order of $\gamma \sim 10$ W^{-1} km^{-1}. From the nonlinear coefficient the nonlinear length of a waveguide may be defined, $L_{NL} = (\gamma P_0)^{-1}$, in which P_0 is the peak power of the optical pulse. The nonlinear length is a characteristic length over which fiber nonlinearity has a significant effect on pulse evolution.

If the nonlinear length and dispersion length are equal, and the dispersion is anomalous, then the effects of SPM and dispersion may cancel, resulting in optical solitons, i.e., pulses which can propagate in a lossless dispersive nonlinear media without temporal broadening. The balance between nonlinearity and dispersion plays a significant role in pulse formation and compression in short pulse fiber lasers. Cross-phase modulation and SPM also have important applications in realizing ultrafast all-optical-fiber switches and saturable absorbers, enabling the generation of sub-picosecond optical pulses in fiber lasers. The latter functions may also be realized using a related nonlinear effect, nonlinear birefringence, associated with the fact that the nonlinear index change $n_2 I$ is birefringent, i.e., it depends on the polarization of the light, and can modify the coupling between different polarization states in linearly birefringent fiber.

3. Other Properties of Optical Fiber Waveguides

Optical fiber lasers are generally long, and therefore susceptible to changes in their environment. The most significant effects are changes with temperature and stress. Optical fibers have low thermal mass, i.e., heat is rapidly transferred to or from the core through the fiber's cladding and surface, with little thermal resistance. The change in effective index of silica fiber with temperature is approximately -1.1×10^{-5} C^{-1}, and the thermal expansion coefficient is $+5.2 \times 10^{-7}$ C^{-1}, consequently the fractional change in optical path length per unit change in temperature is approximately -10.5×10^{-6} C^{-1}. For example, a one degree change in temperature would change the optical path length in a 10-m long optical fiber by 105 μm, i.e., many wavelengths. By varying the axial strain on the fiber, its length at constant temperature may be varied, or alternatively the effects of temperature change may be compensated. Silica fiber can usually safely withstand tensile strains up to 1%, or slightly more in compression.

Another linear effect often encountered in interfacing to optical fibers is Fresnel reflection, which occurs at discontinuities in the effective refractive index, for example, at the glass-air interface at the end of an optical fiber. The refractive index of silica glass in the infrared is approximately 1.5, and consequently 4% (i.e., 0.18-dB) of the incident optical power is reflected from any bare fiber end. Fiber amplifiers can have very high gain and so stray reflections from fiber ends or uncoated optical components are often enough to cause unwanted or uncontrolled oscillation. There are a number of ways to prevent unwanted reflections from fiber ends: (1) if coupling to other optical elements, the end of the fiber can be cleaved at an angle to the fiber axis or (2) if not coupling to other elements the incident power can be radiated, either by placing the fiber end in index matching fluid, or by bending the fiber in a small loop.

C. Optical Fiber Components

The following sections describe the most commonly used fiber-based components used in fiber lasers and other optical fiber systems.

1. Fiber Splices and Connectors

Optical fibers must often be joined, either permanently or temporarily, with minimal coupling loss and back-reflection. Depending on the performance required, one of three methods are commonly employed to join optical fibers: (1) fusion splicing, (2) mechanical splicing, or (3) butt-coupling mechanically aligned fibers terminated with connectors. The fiber ends to be joined must be clean, accurately aligned, and cleaved or polished, so the fiber end-faces are flat and parallel to each other.

Fusion splicing involves heating the fibers to be joined to melting point and forcing them together. The resulting joint is strong and permanent, and provided the fibers were accurately aligned the losses are generally very low (e.g., $<$0.1 dB). The primary disadvantage is the cost of fusion splicing equipment. Mechanical splicing is used in situations where a less permanent joint is required between

bare fibers, and slightly higher loss can be tolerated. The fiber ends are guided together, and their claddings aligned and held together mechanically in a small and relatively inexpensive splicing capsule. Finally, in situations where a fiber is often being disconnected and connected, and extra expense and slightly extra loss can be tolerated (e.g., 0.5 dB), it is convenient to use fibers terminated with one of a variety of commercial connectors. Feedthrough adapters facilitate joining fibers with connectors to one another by aligning and butting their end-faces together. Connectors are available with angled end-faces to prevent reflections occurring. Splices based on mechanical alignment of fiber claddings may have significant losses if the fiber cores are not concentric with their claddings.

2. Directional Couplers

Fiber couplers are useful for splitting or combining light propagating in optical fibers with minimal loss. Light from one of the input waveguides is coupled between the two waveguides, and leaves the device divided between the two output waveguides. The coupling ratio depends on the physical design of the device and wavelength. The two primary applications of directional couplers in fiber lasers are (1) tapping out (or injecting in) a fraction of the optical power circulating in a laser cavity and (2) combining or separating different wavelengths in the laser cavity. The latter device is called a wavelength selective coupler (WSC), and is often useful for injecting the pump wavelength into the cavity without coupling the signal out.

3. Grating Filters

Optical fiber grating filters have rapidly become a key signal processing element in optical fiber systems, including fiber lasers. The filters are usually fabricated by exposing the core of a germanosilicate optical fiber to a periodic UV illumination, which causes a small but permanent periodic increase in the refractive index. In Bragg gratings the periodically varying index acts as a distributed reflector, causing contradirectional coupling between forward and backward propagating modes around a central wavelength, λ_0, determined by the period of the index change, $\Lambda = \lambda_0/(2n)$. The periodic index change may also be designed to cause coupling from the core to other modes and directions, for example, to cladding or radiation modes.

The spectral response of Bragg gratings (i.e., their transmissivity or reflectivity as a function of wavelength) may usually be determined to very good accuracy using coupled mode equations. In a uniform grating of length L, in which the magnitude and period of the index change do not vary along the direction of propagation, the coupled mode equations may be solved analytically to give the ratio of reflected to incident power as a function of wavelength, i.e., the reflectivity

$$R = \kappa^2/(\delta^2 + \beta^2 \coth^2(\beta L)),$$

in which the coupling coefficient, κ, is related to the peak index variation experienced by the guided mode, Δn, by $\kappa = \pi.\Delta n/\lambda$, the effective coupling coefficient is $\beta = \sqrt{\kappa^2 - \delta^2}$, and the detuning from Bragg resonance is $\delta = 2\pi n/\lambda - \pi/\Lambda$. Typical reflectivity spectra are plotted in Fig. 3 for different values of κL as a function of normalized detuning, δ/κ. The main results are (1) the filter's fractional bandwidth is approximately $\Delta\lambda/\lambda_0 = \Delta n/n$ (i.e., about 10^{-3} in lossless gratings), (2) the phase response is a linear function of frequency around λ_0 (i.e., the temporal response is delayed in time by an amount related to the effective length of the grating), and (3) the spectral response has significant sidelobes.

The spectral response of a uniform grating is not ideal for most applications, but provides a useful reference for understanding more complicated structures. For example, apodization of the coupling coefficient toward the ends of the grating can eliminate the sidelobes. Also, if the grating period is linearly chirped, then the phase response will be quadratic, i.e., dispersive. Chirped gratings are useful for controlling the dispersion in optical fiber systems. Lastly, more complicated structures and spectral responses may be synthesized using more than one grating, for example, two gratings can form an in-fiber Fabry-Perot cavity, at least over the wavelength range that both gratings are reflective.

4. Fiber-Coupled Components

Many useful functions are difficult or impossible to realize within optical fibers, e.g., broadband amplitude and phase

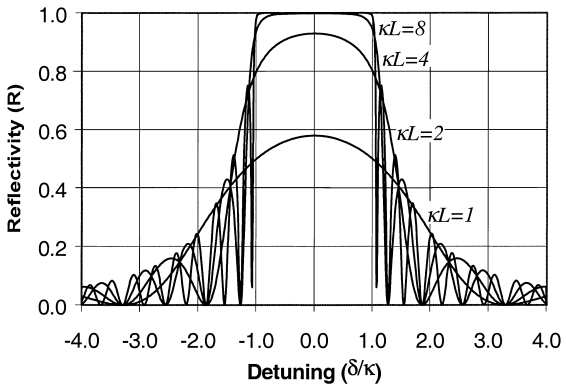


FIGURE 3 Reflectivity of uniform Bragg grating filters as a function of normalized detuning. A periodic change in refractive index causes a spectrally dependent coupling between forward and backward propagating waves in an optical fiber.

modulation, and nonreciprocal devices such as Faraday rotators and optical isolators, etc. In these situations it is necessary to use bulk optical devices which have been coupled to optical fibers. Insertion losses are typically from 1–3 dB or more, depending on the coupling mechanism. Manufacturing is often expensive, however, the advantages of having a packaged and robust device, with the convenience and flexibility of optical fiber input and output, are significant. The same considerations apply in coupling optical fibers to semiconductor devices, such as laser diodes and photodiodes.

5. Polarization Controllers

Cylindrically symmetric optical fibers are theoretically not birefringent, however, in practice bends and twists induce birefringence in optical fibers, which can modify the state of polarization of light in an unpredictable way as it propagates through the fiber. The performance of many devices used in optical fiber lasers (e.g., electro-optic modulators) is polarization-dependent, so it is often necessary to control the polarization state of light at some point in the fiber and/or compensate for unwanted birefringence.

Theoretically, any state of polarization of light may be transformed to any other state by passage through a quarter-wave plate and a half-wave plate, with their optical axis at some angle to each other and the input. Polarization transformation, and therefore polarization control, may be achieved in optical fiber by inducing birefringence equivalent to a quarter- and half-wave plates (e.g., by bending or squeezing the fiber, as described in Section II.B.1), and then rotating the birefringent sections with respect to each other. Three sections are often used in practice to provide extra flexibility.

6. Mirrors

Mirrors are used to reflect a fixed proportion of incident light, preferably without dependence on polarization or wavelength. There are three types of mirror commonly used in optical fiber systems (1) Bragg gratings (which have finite bandwidth, as described previously), (2) metal deposited onto the fiber end (only useful if no connection to another fiber is necessary), and (3) loop mirrors.

Loop mirrors are formed by looping the output ports of a directional coupler back to each other, as shown in Fig. 4. The magnitude and phase of the light reflected back into the input port depends on the amount of coupling, K; for a lossless coupler and fiber the reflectivity is given by $R = 4K(1 - K)$, and any power not reflected is transmitted. For example, if the power coupling ratio was 50%, then the loop mirror reflectivity is unity, though birefringence in the loop can change it from being normally reflective to normally transmissive.

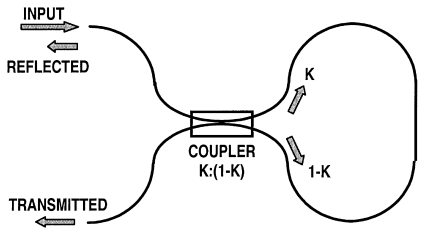


FIGURE 4 Schematic of a fiber loop mirror. The input is divided into two counterpropagating waves by the fiber coupler, which are recombined in the same coupler after propagation around the fiber loop. The output exits one or both ports in the backward direction, with proportion in each determined by the power-coupling ratio K.

The loop mirror is a nonresonant Sagnac interferometer, i.e., waves propagating around the loop in opposite directions usually see the same optical path length. Rotation of the loop, or the acquisition of a nonreciprocal phase shift (e.g., due to loop asymmetry and SPM), unbalances the phase shift seen by counterpropagating waves and causes a change in reflectivity. The latter effect is used in nonlinear loop mirrors, described next.

7. Nonlinear Mirrors and Switches

Self-phase modulation in optical fibers may be used to realize artificial ultrafast saturable absorbers, in which loss or transmission depends on instantaneous optical intensity. Such devices are extremely useful for sub-picosecond optical pulse generation and switching in optical fiber, and for mode locking fiber lasers. There are two approaches to realizing the fast saturable absorber function in optical fibers: (1) the nonlinear loop-mirror and (2) nonlinear polarization rotation. Both methods involve splitting the input into two unequal parts, which then propagate through a length of fiber and by SPM acquire phase shifts proportional to their instantaneous intensities. Recombining the two components causes constructive or destructive interference, depending on the instantaneous intensity and relative phase shift.

In the nonlinear fiber loop mirror the pulses propagate in opposite directions around a Sagnac interferometer. Low intensities are relatively unaffected by SPM, and are therefore reflected (as in a linear loop mirror), however, high intensities acquire a relative phase difference and are transmitted. In nonlinear polarization rotation the two waves propagate on orthogonal axes of a birefringent fiber, and the intensity-dependent phase difference accumulated by SPM and CPM manifests as an intensity-dependent state of polarization, or an intensity-dependent transmission through a polarizer. In both cases the resulting transmissivity is a periodic function of increasing intensity.

III. CW FIBER LASERS

The ideal continuous wave (CW) laser would have a single stable longitudinal mode with vanishingly narrow linewidth, moderate output power, and quantum limited noise. It is possible for fiber lasers to approximate the latter ideal very closely, as outlined in the following sections.

A. Long Cavity CW Lasers

All-fiber lasers may be configured in one of two fundamental forms: the standing wave or Fabry-Perot cavity, e.g., as in Fig. 1, or the ring cavity, e.g., as in Fig. 5. The low pump absorption coefficient and gain per unit length of fiber amplifiers usually necessitates a long fiber amplifier and therefore a long cavity. Consequently the mode frequency spacing in fiber lasers, $\Delta\nu = c_0/(n \cdot L_c)$, is usually relatively small. For example, in a standing wave cavity with physical length 10 m, the round trip length is $L_c = 20$ m, and the mode spacing approximately $\Delta\nu = 10$ MHz.

In continuous wave (CW) lasers the close mode spacing facilitates mode hopping, which together with environmental changes and perturbations is the main problem in realizing fiber lasers with narrow linewidth and low noise. Polarization mode competition is another potential cause of lasing instability, but can be avoided by using polarization maintaining fibers throughout the cavity. As discussed in Section III.E, inhomogeneous gain also tends to promote lasing on multiple wavelengths. Any or all of the following three fundamental techniques may be used to promote single-longitudinal mode operation of fiber lasers: (1) avoidance of spatial holeburning; (2) use of narrowband feedback; and (3) coupling two or more cavities.

Spatial holeburning is a periodic variation in intensity, and therefore of local gain, which occurs in the optical fiber amplifier due to a standing wave pattern set up by counterpropagating modes. The periodic variation in gain can result in reduced gain for the lasing mode, and thereby contributes to mode hopping instabilities. Spatial holeburning may be prevented in standing wave cavities by using a Faraday mirror (i.e., a polarization rotating mirror) at one end of the cavity, or in ring cavities by incorporating an optical isolator to ensure unidirectional lasing, as shown in Fig. 5.

The use of narrowband feedback, for example, a long Bragg grating reflector at one or both ends of a standing wave cavity, substantially reduces the number of modes that can exceed threshold to within a narrow bandwidth defined by the grating. However, this technique is unlikely to ensure single-longitudinal mode lasing unless the cavity is short (refer to Section III.B).

Coupling two or more cavities together promotes lasing only of those modes which are common to both cavities. By the Vernier principle, common modes will be widely spaced in frequency if the coupled cavities are of the similar length. An elegant example of this technique is the Fox-Smith resonator, shown in Fig. 6a, in which mirrors 1 and 3 and mirrors 2 and 3 define two coupled cavities. A similar approach is to form two colinear standing wave cavities using only three Bragg grating mirrors, as shown in Fig. 6b; the central mirror is common to both cavities, and determines the coupling between them. The latter technique has the added advantage of using narrowband feedback. Ring lasers can also be coupled to other cavities using fiber couplers, and/or use an intracavity etalon to provide mode selectivity, as shown in Fig. 5.

B. Short Cavity CW Lasers

Distributed Bragg reflector (DBR) and distributed feedback (DFB) fiber lasers have relatively short cavities, e.g., 10 cm or less, facilitated by the use of heavily doped-fiber amplifiers. The aim is to provide increased cavity mode

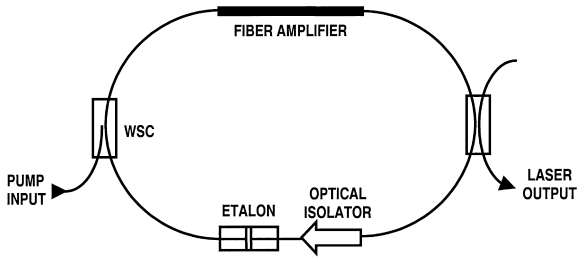


FIGURE 5 Schematic of a continuous wave (CW) fiber ring laser. The laser shown contains an isolator to force unidirectional operation and avoid spatial holeburning, and uses an intracavity etalon to promote lasing on a single longitudinal mode of the laser cavity. The pump for the fiber amplifier is coupled into the cavity using a wavelength selective coupler (WSC) and the laser output taken via another directional coupler.

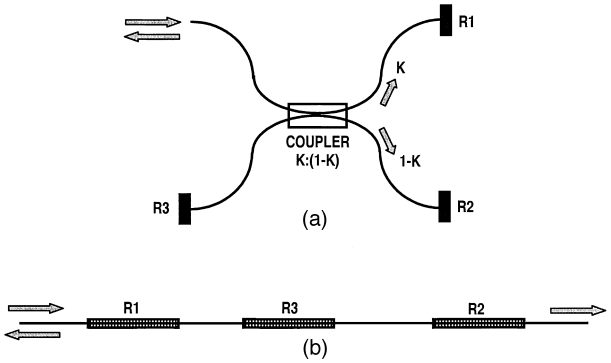


FIGURE 6 Schematic of typical fiber compound cavities used to promote lasing on a single longitudinal mode common to both cavities. (a) A Fox-Smith compound cavity using fiber couplers, (b) standing wave compound cavity, using Bragg gratings.

spacing, which combined with the wavelength selectivity of fiber Bragg gratings prevents lasing of more than a single mode of the cavity. In theory the laser linewidth is then limited by spontaneous emission noise to around 1 Hz (i.e., the Schawlow-Townes limit), however, environmental effects and gain dynamics have a significant effect on the linewidth achievable in practice.

Distributed Bragg reflector lasers use one or two Bragg gratings as end reflectors in a short Fabry-Perot cavity, with amplifying fiber between. Distributed feedback lasers comprise a single grating written into a fiber amplifier, in which a $\lambda/4$ phase shift around the middle of the grating defines a single lasing mode. In practice the phase shift is distributed, and can be formed by localized UV processing or heating of the fiber. Tuning the size of the phase shift tunes the lasing wavelength within the small range defined by the grating bandwidth. Distributed Bragg reflector lasers can be constructed to emit from either one or both ends of the cavity, whereas DFB lasers emit from both ends. Both DBR and DFB lasers may be coupled to other cavities to improve the suppression of unwanted modes, e.g., by adding an extra grating at one end of the DBR or DFB structure to form a Fabry-Perot cavity and provide frequency selective feedback. Measured linewidths of short-cavity fiber lasers generally fall between 1 kHz and 200 kHz.

Fabricating short fiber lasers presents several problems, particularly for erbium-doped fiber lasers. First, the pump absorption coefficient must be very high if sufficient gain to exceed threshold is to be generated in a short length of fiber, while also providing a useful amount of output power at the lasing wavelength. Increased pump absorption (e.g., 120 dB/m) has been obtained using increased dopant concentration and/or codoping techniques, however, neither method is ideal. Simply increasing the rare-earth dopant concentration facilitates ion-pair interactions such as cooperative upconversion, which reduces the amplifier gain efficiency and can cause laser instability (as per Section V.B, on chaotic lasers). Using erbium/ytterbium codoped fiber reduces upconversion, but is usually accompanied by phosphorus codoping which can significantly reduce the photosensitivity of the fiber, and makes DFB grating fabrication difficult. The reduction in photosensitivity associated with phosphorus may be circumvented by hydrogenating the fiber, however, this increases the loss and birefringence in the fiber following UV exposure, which reduces the lasing efficiency and facilitates lasing on two polarization modes. A promising technique for overcoming the latter problems is to write the gratings in codoped amplifying fiber containing a photosensitive ring surrounding the doped core, thereby separating the gain and grating regions in the fiber.

Another difficulty often encountered in short fiber lasers is instability, causing increased noise or self-pulsing behavior. Causes of instability include cooperative upconversion in highly doped fiber amplifiers, spatial holeburning in the gain medium due to interference between forward and backward traveling waves, external reflections, and temperature changes in the doped fiber due to absorbed pump power which shifts the wavelength of grating filters, etc. In linearly birefringent cavities, lasing on two polarization modes may occur and cause instability through polarization mode competition. Instabilities can sometimes be avoided by optimizing the pump power, for example, pumping the laser just above threshold avoids lasing on two polarization modes, however, a low pump power increases the likelihood of self-pulsation in the presence of ion-clustering. Due to the short length and low absorbed pump power in DFB and DBR fiber lasers, it is generally necessary to use a master-oscillator power-amplifier (MOPA) configuration to obtain significant output power from short-cavity fiber lasers, as shown in Fig. 7. Using the latter approach narrow linewidth lasers may be fabricated with output powers well in excess of 100 mW, and with relative intensity noise (RIN) less than -150 dB/Hz (above 10 MHz).

C. Tuning Techniques for CW Fiber Lasers

Rare-earth-doped and Raman fiber amplifiers have gain-bandwidths of up to several tens of nanometers, enabling lasing within a wide range of wavelengths. Tuning

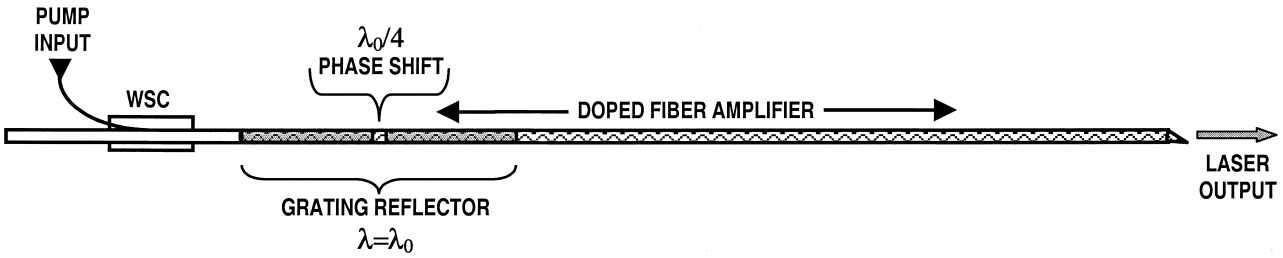


FIGURE 7 Schematic of a distributed feedback (DFB) fiber laser, including a power amplifier to boost the laser's output power.

between different wavelengths is often desirable, and may be achieved in a number of ways, depending on the cavity configuration. Fiber pigtailed bulk devices, such as dielectric-stack tunable filters and acousto-optic tunable filters, are easily incorporated into long-cavity lasers. Alternatively the laser can be coupled to an external bulk grating for tuned feedback. Tuning of Bragg grating filters is often the only alternative for tuning DBR and DFB lasers, and has the advantage of incurring minimal loss. In-fiber grating filters are commonly tuned using stress or temperature.

Temperature tuning of UV-written gratings is possible over a restricted range of wavelengths. Combining the change in volume and the change in index that both result from a change in temperature, the Bragg wavelength of a fiber grating changes with temperature at approximately 0.01 nm/μm/$^\circ$C, however, the maximum tuning range is limited by thermal annealing and eventual erasure of the grating. Tuning of in-fiber gratings by stretching or compression can achieve larger tuning ranges with higher speed; up to approximately $\pm$20 nm is possible before fiber failure occurs. A preferable approach for obtaining large tuning ranges with fiber gratings is to use two comb filters with periodic spectral response (e.g., Fabry-Perot resonators, or sampled gratings) as reflectors in the cavity. If the two combs have a slightly different spectral period, then tuning one comb by a small amount can change the wavelength in common significantly, by a Vernier effect.

Both long- and short-cavity fiber lasers are readily tuned without noticeable mode hopping. In a long-fiber cavity the modes are very closely spaced, hence many modes are close to threshold, and tuning between them makes little difference in laser output power. In a short-fiber cavity the modes are widely spaced, and the one lasing mode may be continuously tuned. Underdamped relaxation oscillations are usually encountered when tuning fiber lasers due to their slow gain dynamics. A simple way to damp the oscillations and reduce settling time to much less than a millisecond is to add a nonlinear element (e.g., a nonlinear loop mirror) to prevent large excursions in instantaneous optical power within the cavity.

D. Fiber Lasers in Low-Phonon Energy Hosts

Low-phonon energy glasses containing heavy metal fluorides (e.g., zirconium, barium, lanthanum, etc.) are commonly known as ZBLAN glasses. These glasses were originally of interest for their low infrared transmission loss, however, their low-phonon energy has become very useful for realizing rare-earth-doped fiber lasers at wavelengths that are not possible in silica glass fibers. While most activity has concentrated on CW lasers, a number of pulsed fiber lasers (discussed in Section IV have also been demonstrated using rare-earth-doped ZBLAN fiber amplifiers

The probability of nonradiative decay through multiphonon processes decreases rapidly as the number of phonons required to bridge the energy gap between two states increases. The maximum phonon energy in ZBLAN glass is approximately half that of silica, and consequently, the excited state lifetimes of rare-earth ions in ZBLAN fibers may be orders of magnitude longer than in silica, especially for transitions between widely separated states. Because the host is glass, the absorption and emission bands remain broader than in crystalline hosts, allowing some freedom in the choice of pump wavelength, and the possibility of limited tunability of the emission wavelength. Energy transfer between different rare-earth codopants can be used to provide extra flexibility in pumping schemes. The reduced nonradiative decay rates make a number of lasing transitions accessible that are not feasible in silica fibers, including efficient upconversion and long wavelength lasing.

1. Upconversion Fiber Lasers

Upconversion is the generation of light at shorter wavelengths than supplied by the pump. It occurs in rare-earth-doped materials through the absorption of two or more photons by a rare-earth ion before radiative emission occurs. In the absence of cooperative upconversion effects (i.e., energy transfer between closely spaced ions), long excited state lifetimes and high pump fluxes are required for efficient upconversion to occur. The latter conditions are difficult to achieve in bulk without cooling the host material, but are relatively easy to achieve in low-phonon energy glass fibers, even at room temperature.

Upconversion lasers were originally of interest because of their potential applications as visible lasers. With the development of periodically poled nonlinear materials for optical frequency conversion, upconversion lasers have become less important, although ZBLAN fiber lasers may generate light at several wavelengths simultaneously from one pump wavelength. Using laser diode pump sources, upconversion lasers have been demonstrated in rare-earth-doped ZBLAN fiber at several wavelengths between 381 and 910 nm, sometimes with high power ($>$100 mW) and slope efficiencies up to approximately 30%. For example, the generation of 230 mW of blue light at 481 nm has been achieved by three-stage pumping of thulium-doped ZBLAN at 1123 nm from an Nd:YAG laser. A potential problem with short wavelength upconversion lasers is the generation of color centers in the fiber, which cause loss at the lasing wavelength, though the loss is usually reversible.

2. Long Wavelength Lasers

The low nonradiative decay rates of rare-earth ions in low-phonon energy glasses allows efficient lasing between relatively closely spaced transitions, i.e., at long wavelengths. The longest wavelength reported lasing in silica fiber is 2.1 μm (holmium or thulium), while in ZBLAN fiber many transitions have been observed lasing above 2 μm (in thulium, erbium, and holmium) with the longest wavelength at 3.95 μm (holmium). High output power is possible using the same techniques used to scale output power in silica fiber lasers, i.e., cladding pumped fiber geometries and/or energy transfer between codoped ions. Using the latter techniques more than 1.5 W has been obtained at 2.7 μm from an erbium/praseodymium doped fiber laser.

E. Multiwavelength CW Lasers

The term "multiwavelength" is here used to denote lasers which simultaneously generate more than one coherent wavelength in a single cavity, but with only limited coherence between the different wavelengths. Advantages of multiwavelength lasers include the reduced hardware requirements compared with combining the outputs of several different lasers, and avoidance of high peak optical powers which may exist in mode locked lasers.

It has already been mentioned that ZBLAN fiber lasers can generate more than one wavelength on different transitions. In this case the lasing wavelengths are defined by specific energy transitions of the rare-earth ions involved, are usually widely spaced, and are likely to have unequal output powers. Similarly, silica fiber doped with more than one rare-earth species, and pumped with a common source, can lase at more than one wavelength simultaneously.

A very interesting form of multiwavelength CW fiber laser is the cascaded Raman laser, in which each lasing wavelength is used to pump a longer wavelength laser in the same length of fiber. A series of Fabry-Perot cavities are formed one inside the other using pairs of Bragg gratings, as shown in Fig. 8. The outermost pair of gratings are tuned to the longest lasing wavelength, one SRS Stokes shift from the next pair of gratings (i.e., located inside the first pair) which form the cavity to generate the pump for the longer wavelength. Another pair of gratings is located inside the latter pair, and so on, until the innermost cavity in which the Raman gain is pumped by a semiconductor pump source, launched into the fiber through the surrounding gratings. Using highly reflective gratings the latter laser construction can generate gain and/or optical output with excellent efficiency at any wavelength desired from a single high power pump. For example, 1.5-W output power has been generated at 1485 nm by cascading down from a 6-W ytterbium fiber laser pump source at 1117 nm. Similarly it is possible to generate up to 40 dB of low noise gain at 1300 nm by cascading down from a 1064-nm pump.

A similar cascade of wavelengths can be generated in hybrid Brillouin/erbium-doped fiber lasers. The lasing wavelengths are defined by a single narrow-linewidth seed laser, injected into a ring cavity containing a doped-fiber amplifier. The latter serves to amplify the Brillouin wavelengths generated to above threshold, while not exceeding threshold itself. In this way several lasing wavelengths may be generated with narrow linewidths (i.e., $<$10 kHz), separated by the SBS Stokes shift (i.e., approximately 11 GHz or 0.08 nm).

Multiwavelength lasing can also be achieved using rare-earth-doped fiber amplifiers, however, unlike the latter cascaded-gain lasers, specialized techniques are generally required to avoid cross-saturation effects in the normally homogeneous gain medium which usually prevents all but one wavelength from exceeding threshold at a time. In any rare-earth-doped multiwavelength laser there are usually two elements; a filter to define the wavelengths to be generated, and a mechanism for multiplexing different wavelengths in the gain medium. In-fiber resonators, or comb filters of one form or another, are well suited to defining the lasing wavelengths. The gain may be multiplexed between the lasing wavelengths by (1) cooling the fiber, thereby reducing the homogeneous broadening (e.g., to approximately 1 nm at 77 K), and leaving the fiber largely inhomogeneously broadened or (2) spatially multiplexing different wavelengths in, for example, a twin-core

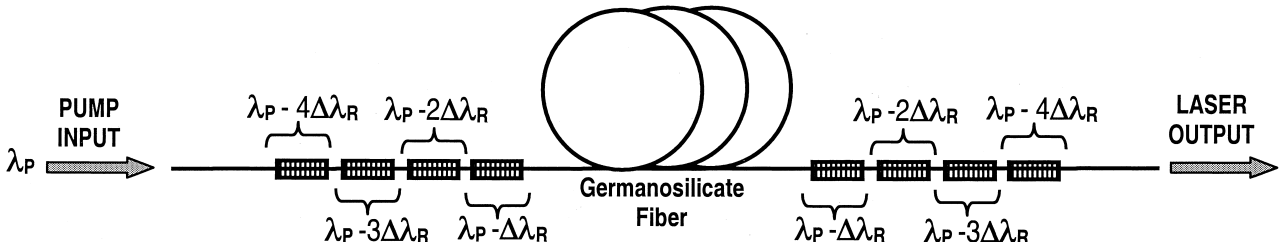


FIGURE 8 Schematic of an optical fiber cascaded Raman laser.

erbium-doped fiber. An alternative technique to promote simultaneous multiwavelength lasing is to individually adjust the loss and equalize the gain for each wavelength of interest in the cavity, however, this approach is cumbersome and not robust.

IV. PULSED FIBER LASERS

The broad gain bandwidth and high saturation flux of fiber lasers makes them well suited to generation of short and ultrashort optical pulses with high peak powers. Furthermore, the high gain of fiber amplifiers allows the incorporation of relatively lossy elements into the laser cavity, such as amplitude or phase modulators for mode locking, without significant deterioration in performance. Consequently most techniques used for pulse generation in bulk lasers have been applied successfully in fiber lasers, however, several factors influence pulse formation and steady-state operation in fiber lasers which are not always relevant in bulk lasers, such as distributed gain, various orders of dispersion, and nonresonant nonlinearities. The latter effects are often advantageous and usually fundamental to pulsed fiber laser performance. For example, fiber nonlinearity and dispersion can together provide intracavity soliton-like pulse compression, allowing shorter pulses to be generated than usually possible using active mode locking. Recent advances in pulsed fiber lasers have been assisted by developments in associated technologies, such as fiber Bragg gratings for filtering and dispersion control, semiconductor saturable absorbers for reliable mode locking, high power pump diode laser arrays for high output power, and periodically poled nonlinear media for frequency conversion.

A. Q-Switched Fiber Lasers

Q-switched fiber lasers are useful for generating moderately short optical pulses with high energy and high peak power. Pulse generation in Q-switched lasers is achieved by rapid changes in the intracavity loss, so that the pumped cavity suddenly goes from below threshold to well above threshold while below threshold energy is stored in the inverted gain medium, which is rapidly dumped once the cavity goes above threshold. A short high-energy pulse then builds up from spontaneous emission noise during several round trips of the cavity, quickly saturating the amplifier gain and driving the inversion back below threshold, after which the process may be repeated. Switching of the cavity loss may be achieved using either passive or active means, e.g., slow saturable absorbers based on dye-doped polymer can be placed in the cavity to initiate passive Q-switching, and acousto-optic modulators are commonly used for active Q-switching. For maximum pulse energy, repetition rates are typically less than 1 kHz.

The long excited state lifetimes and small gain cross-sections of rare-earth-dopants, together with the possibility of using long lengths of doped-fiber to store energy, make fiber lasers ideal for generating Q-switched pulses with large energy and peak power. The main limitation on energy storage in a fiber laser, and therefore on Q-switched pulse energy, is amplified spontaneous emission noise, which depletes the inversion; in erbium-doped fibers, the theoretical limit is about 1 mJ. Another limitation on pulse energy not associated with the fiber is the Q-switch extinction ratio, which must be very large to prevent premature lasing when the gain medium is well inverted. The main limitations on minimum pulse duration and peak pulse power are the Q-switching time and cavity lifetime. The primary trade-off is that for a given dopant density a short cavity and/or a large outcoupling reduces the cavity lifetime, and therefore the pulse duration, but also reduces the pulse energy.

Typical Q-switched rare-earth-doped fiber lasers generate pulses with hundreds of microjoules of energy, tens of nanoseconds duration, and a few kilowatts peak power. Recent results indicate that with careful design ytterbium-doped, Q-switched fiber lasers can generate pulse energies larger than 1 mJ, or peak powers up to 10 kW. At such power levels intracavity supercontinuum generation may be observed, due to the onset of cascaded Raman scattering.

B. Mode Locked Fiber Lasers

Mode locked fiber lasers can deliver much shorter optical pulses, with much higher repetition rate, similar peak power, and much better coherence than Q-switched fiber lasers, but at the expense of lower pulse energy. Mode locked fiber lasers, as the name implies, generate short optical pulses by locking modes of the laser cavity together in phase. Mode locked fiber lasers therefore require a nonlinear or time-varying element in the laser cavity to generate frequency components at one or more harmonics of the cavity fundamental frequency. The harmonics injection lock neighboring modes of the laser cavity and determine the pulse repetition rate.

In mode locked fiber lasers significant nonlinearities and dispersion often exist in the laser cavity, and in some cases the pulse may undergo considerable spectral or temporal evolution each round-trip of the cavity. In these lasers mode locking is better understood in the time domain, whereby stable pulse generation simply requires pulse self-replication every round trip of the laser cavity. In the latter view the pulses generated in mode locked fiber lasers may be regarded as stable modes of their laser cavity.

There are a wide variety of mode locking techniques and cavity designs used in mode locked fiber lasers, however, their performance can largely be measured in terms of the following parameters; pulse duration, peak pulse power or intensity, pulse energy, repetition rate or average power, and pulse time-bandwidth product. Pulse noise and jitter are also important measures of laser performance. The performance parameters depend largely on the mode locking technique used and on the components used for temporal and/or spectral control within the laser cavity, as described in the following sections.

1. Active Mode Locking

Active mode locking utilizes a time-varying element within the laser cavity, such as an integrated electro-optic amplitude or phase modulator driven by an external frequency reference, to promote phase locking of the cavity modes. The advantage of active mode locking is that pulses are generated at a stable repetition rate. The schematic of a typical actively mode locked fiber laser is shown in Fig. 9.

For mode locking to occur, the modulator must be driven at a frequency which is almost exactly at the cavity fundamental frequency (i.e., typically within parts per million, depending on the strength of the mode coupling mechanism), or a harmonic of it. Consequently small changes in optical path length within the cavity (e.g., due to temperature changes) can be detrimental to pulse generation, and must be tracked by the modulator drive frequency. For example, in a regeneratively mode locked system the modulator drive signal is derived by detecting the cavity harmonic frequency of interest and amplifying it in an external circuit. Alternatively, the cavity length may be stabilized by varying the strain on a section of fiber within the cavity. The latter technique may be applied in conjunction with a phase-locked loop circuit to mode lock the laser to a fixed external reference frequency. In practice, a polarization controller is also required to adjust the polarization of the optical signal entering the modulator, which is usually polarization sensitive. Alternatively, the laser cavity may be constructed using polarization maintaining (i.e., highly birefringent) optical fiber, which reduces the sensitivity of the laser to environmental effects and provides superior stability.

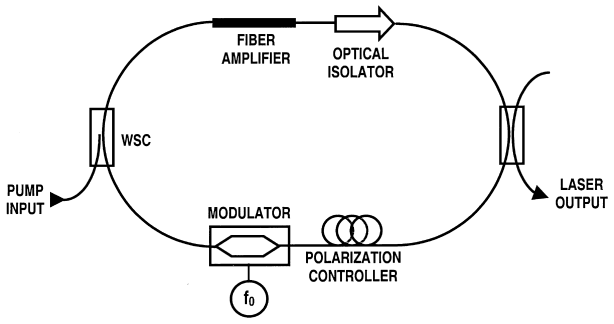


FIGURE 9 Schematic of an actively mode locked fiber laser. The modulator drive frequency must be a harmonic of the cavity fundamental frequency and determines the pulse repetition rate.

As fiber lasers are generally relatively long (e.g., with fundamental frequency of a few megahertz), harmonic mode locking is required to generate pulses at high repetition rates. Repetition rates of tens of gigahertz have been demonstrated, effectively filling the cavity with many pulses per round trip, however, there is usually a reduction in pulse peak power and some penalty in noise performance and stability. Linear analysis of homogeneous lasers mode locked by loss modulation (e.g., with an electro-optic amplitude modulator) predicts transform limited (i.e., unchirped) Gaussian pulse generation with fullwidth at half-maximum (FWHM) intensity given by

$$\Delta\tau = \frac{\sqrt{2\ln 2}}{\pi}\sqrt{\frac{1}{f_m \Delta\nu}\sqrt{\frac{2g}{\delta_{am}}}},$$

in which f_m is the modulation frequency, $\Delta\nu$ the gain bandwidth, δ_{am} the amplitude modulation index and g the saturated gain. However, in fiber lasers with net anomalous dispersion and nonlinear SPM, intracavity pulse compression may occur, resulting in hyperbolic secant (i.e., sech^2) shaped pulses with considerably shorter duration (e.g., to about 1 ps) than could be achieved in a linear cavity. The modulation index is then effectively increased, to $\delta = 0.07\gamma L P_0/(f_m \Delta\tau)^2$, where γ is the nonlinear coefficient defined in Section II.B.2, P_0 is the pulse peak power, and L the effective cavity length.

Similarly, mode locking a linear cavity with a phase modulator would normally result in slightly chirped pulse generation, but in a fiber laser the effect of cavity dispersion and nonlinearity may modify the time-bandwidth product of the generated pulses. For example, net normal dispersion together with SPM tends to lengthen pulses generated in fiber lasers, and can result in strongly chirped pulses with time-bandwidth products orders of magnitude larger than the transform limit. The same general principles apply in passively mode locked lasers.

2. Passive Mode Locking

Passive mode locking is the method of choice for generating the shortest possible pulses in the simplest possible laser system. Picosecond pulses are readily produced in passively mode locked fiber lasers using various rare-earth-doped fiber amplifiers, and in some cases pulses shorter than 100 fs have been obtained, utilizing the whole gain-bandwidth of the fiber amplifier. Pulse generation in

passively mode locked fiber lasers arises from balances between temporal and spectral gain and loss mechanisms and between nonlinearity and dispersion. Assuming the pulse does not change significantly in temporal or spectral width during propagation around the laser cavity, pulse generation may be modeled using the Ginzburg-Landau equation, which in general has chirped hyperbolic secant-shaped solutions.

If the pulses are unchirped, they reduce to standard soliton solutions of the anomalous dispersion regime, with $P(t) = P_0\ \mathrm{sech}^2(t/\tau_0)$, in which $P_0 = |\beta_2|/(\gamma\tau_0^2)$ is the pulse peak power, β_2 the average cavity dispersion (controllable using chirped Bragg gratings), and γ the nonlinearity coefficient defined previously. For a fundamental soliton the pulse energy is $E_0 = 2P_0\tau_0$, and the fullwidth pulse duration at half-maximum intensity is given by $\Delta\tau = 1.763\tau_0 = 3.53\,|\beta_2|/(\gamma E_0)$, i.e., for a fixed peak power the pulse width reduces with the square root of anomalous dispersion. The FWHM of the pulse power spectrum is $\Delta\nu = 1.763/(\pi^2\tau_0)$, and so the time-bandwidth product for an unchirped soliton is $\Delta\nu\Delta\tau = 0.315$. For example, with typical values of fiber nonlinearity and dispersion at $\lambda = 1.55\ \mu$m (mentioned in Sections II.B.1 and 2), if the pulse peak power is $P_0 = 10$ W, then the pulse duration is $\Delta\tau = 0.88$ ps, the spectral width is $\Delta\nu = 357$ GHz, and the pulse energy $E_0 = 17.6$ pJ. Higher energies would require chirped pulses.

A simple passively mode locked fiber laser is shown in Fig. 10. It is often called a figure-8 fiber laser, for obvious reasons, however, it is actually a ring laser incorporating a nonlinear amplifying loop mirror (NALM). The NALM is a nonlinear loop mirror with gain, in which the fiber amplifier is usually placed at one end of the loop. Although the input signal is split 50:50, the asymmetrically placed amplifier causes counterpropagating signals to have different intensities for most of their time in the loop, resulting in a difference in nonlinear phase being accumulated. The counterpropagating waves are recombined in the 50:50 coupler, with the nonlinear phase difference determining whether the loop is reflective (at low powers, with zero nonlinear phase difference) or transmissive (at high powers, with π nonlinear phase difference). The NALM is a nonlinear interferometer, and is therefore an example of an additive pulse mode locking system, though it behaves like a fast-saturable absorber with femtosecond response time. The fast saturable absorber action causes pulses to be shortened in time, with their low intensity wings being rejected, while the high intensity peaks are transmitted and amplified. Provided there is sufficient gain, pulses may build up out of noise within the laser cavity.

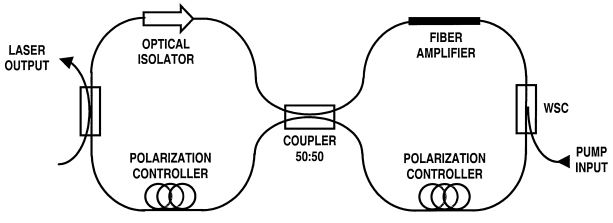


FIGURE 10 A figure-8 passively mode locked fiber laser. The laser is a ring cavity (i.e., the left-hand side of coupler) into which is incorporated a nonlinear amplifying loop mirror, or NALM (i.e., the right-hand side of coupler). The NALM acts as a fast saturable absorber.

The typical spectral output of the figure-8 laser is shown in Fig. 11, which shows a broad spectral peak due to the chirped soliton pulse, and a number of resonant sidebands which contain energy radiated from the soliton due to perturbations received as it propagates around the laser cavity. The cavity length must usually be kept at least a factor of eight times smaller than the soliton period, $z_0 = \pi L_D/2$, or else the periodic pulse perturbations can lead to pulse instability.

Another type of passively mode locked fiber laser is shown in Fig. 12. In this case nonlinear polarization rotation is used to provide the intensity-dependent loss or fast saturable absorber action. Nonlinear polarization rotation can also be used to mode lock relatively short standing wave fiber cavities; the shortest pulses ever generated in a fiber laser, i.e., less than 40 fs, used such an arrangement. Another feature of the laser shown is that while the average dispersion in the cavity is low, the cavity contains two sections with large but opposite dispersion, which cause the pulse to undergo order-of-magnitude changes in temporal and spectral width as it propagates around the cavity. The pulse returns to minimum temporal (and maximum spectral) width midway through each of the dispersive

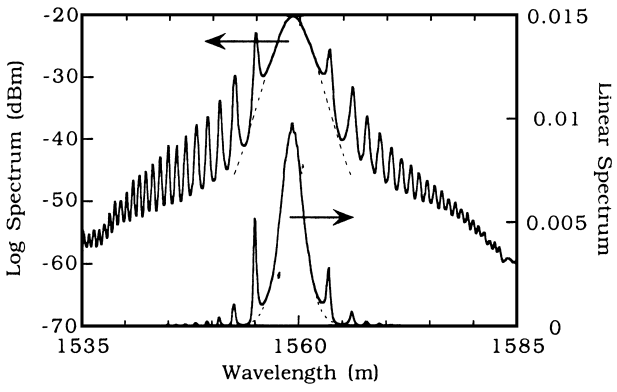


FIGURE 11 Typical spectral output from a figure-8 laser, showing spectral sidebands associated with a dispersive wave background resulting from periodic pulse perturbations. [From Duling, I., and Dennis, M. (1995). Modelocking of all-fiber lasers. *In* "Compact Sources of Ultrashort Pulses" (I. Duling, ed.), Cambridge University Press, London.]

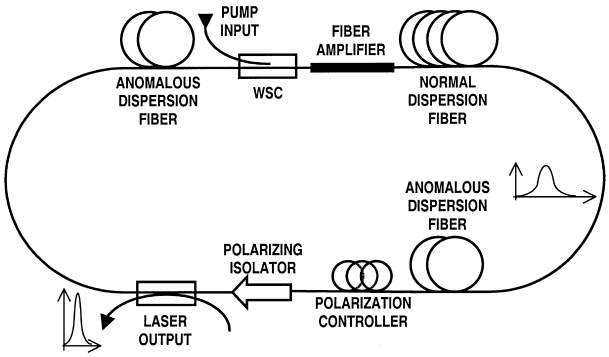


FIGURE 12 Schematic of a stretched pulse fiber laser with balanced dispersion, passively mode locked using nonlinear birefringence, and a polarizing isolator.

fiber sections. The main advantage of this approach is that the peak power of the soliton is reduced on average compared to a transform limited pulse. This allows the generation of pulses with larger energy and shorter duration than would otherwise be possible in a standard soliton laser. In any case, pulse amplification may be performed outside the laser cavity to further boost the pulse power and energy, as described in Section IV.B.6.

The primary disadvantages of passively mode locked lasers are the lack of timing reference, and the fact they are not always self-starting, though these drawbacks can be alleviated by incorporating an etalon (for repetition rate stabilization) or a semiconductor saturable absorber (to initiate pulsing and reduce the self starting threshold). Indeed, semiconductor saturable absorbers may be used to passively mode lock any laser system, however, the actions of nonlinearity and dispersion are necessary for the generation of ultrashort pulses. Intracavity bandpass filters are also beneficial, especially in soliton lasers, to reject noise outside the pulse bandwidth, to stabilize the pulse wavelength and reduce jitter, and to damp soliton interactions (which limit the maximum repetition rate achievable).

3. Soliton Mode Locking

Soliton mode locking may be defined as a subclass of mode locking techniques in which intracavity nonlinearity and dispersion are dominant elements in pulse formation within the laser cavity, with active or passive loss modulation providing only the mechanism for pulse initiation and stabilization. For example, ultrashort soliton generation is possible in lasers incorporating only a slow saturable absorber (e.g., a semiconductor saturable Bragg reflector), in which the recovery time is many times longer than the duration of the pulses generated; the pulses are contained in time against one edge of a broad temporal window. In an almost dual situation, pulse generation in sliding frequency soliton lasers (discussed in Section V.A.3.) occurs in the frequency domain, with spectral confinement being provided by a bandpass filter, and in which the pulse spectrum is much narrower than the filter passband.

4. Hybrid Mode Locking

Hybrid mode locking techniques combine the best features of other mode locking methods. For example, an active mode locking element may be inserted into a passively mode locked fiber laser cavity to produce an ultrashort pulse laser with well-defined repetition rate. Similarly, semiconductor saturable absorbers may be inserted into passively mode locked fiber lasers to initiate soliton generation, resulting in an ultrashort pulse laser with low self-starting threshold.

5. Noise and Jitter in Mode Locked Fiber Lasers

Intensity noise in mode locked fiber lasers arises from a number of sources; primarily pump power and wavelength variation and noise, environmental perturbations, and competition between supermodes of the laser cavity (i.e., groups of modes spaced at the pulse repetition rate). Quantization of pulse energy in soliton lasers can also lead to pulses dropping in and out with changes in pump power. In rare-earth-doped fiber lasers, pump noise and other perturbations affecting the output power are effectively low pass filtered by the long fluorescence time of the upper lasing level. On the other hand, the long fluorescence time means that the fiber amplifier responds only to the average power in the laser cavity, providing no stabilization against variations in pulse amplitude due to beating between neighboring supermodes. Nevertheless, various active and passive cavity stabilization techniques have been used to reduce the intensity noise to the shot noise limit, or around –150 dB/Hz. For example, passive coupled cavities such as intracavity etalons can be used to suppress all but one supermode, stabilizing the repetition rate and averaging pulse-to-pulse power variations. Nonlinear all-optical limiting mechanisms, such as used in passively mode locked fiber lasers, are also beneficial in reducing intensity noise. With stabilization, the RMS intensity noise in mode locked fiber lasers is typically less than 0.5%.

Jitter in mode locked lasers is temporal uncertainty in the arrival time of an optical pulse relative to an expected reference. Jitter can be particularly troublesome in telecommunication systems and in broadband sampling systems. The fundamental source of pulse jitter is uncertainty in the optical carrier frequency, or wavelength, caused by noise added in the amplification process. By propagating in a dispersive medium, the wavelength

uncertainty is translated into arrival time uncertainty. It has been shown experimentally that pulse jitter in mode locked fiber lasers may be reduced to almost quantum limited levels by reducing the total cavity dispersion. In soliton lasers, interactions between closely spaced pulses and dispersive wave radiation in the cavity are other potential sources of timing instability and jitter. The latter are readily limited by ensuring large pulse spacings and by using frequency-guiding filters to stabilize the soliton wavelength. In rare-earth-doped fiber lasers the best phase noise and jitter performance occurs on time scales shorter than the upper lasing level fluorescence time, which tends to filter out high frequency pump noise. For example, the RMS timing jitter in a 10-GHz harmonically mode locked erbium-doped fiber laser generating 1 ps pulses and optimized for low intracavity dispersion was measured to be less than 10 fs in the bandwidth 100 Hz to 1 MHz; the jitter was largely limited by microwave synthesizer phase noise.

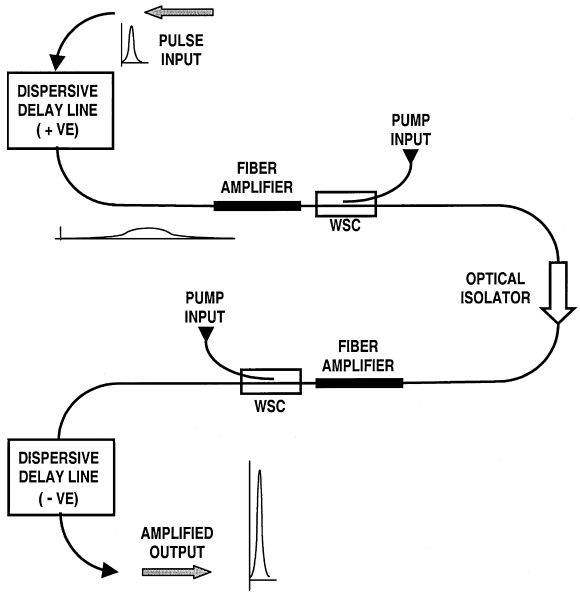


FIGURE 13 Schematic of a typical chirped pulse amplification (CPA) setup.

6. Specialized Techniques

A number of techniques have been developed for generating large pulse energies and for increasing the average and peak pulse powers generated by mode locked fiber lasers to extend their range of applications. Techniques for external pulse compression and frequency conversion have also been developed to extend pulse durations and wavelengths beyond the gain-bandwidth limit of mode locked fiber lasers.

a. Chirped pulse amplification. The energy of pulses generated in mode locked fiber lasers is usually limited to less than 1 nJ, with peak power in the order of kilowatts. In many applications higher pulse energy and/or peak power is required, however, simply amplifying the laser output is not feasible, as fiber nonlinearities can cause pulse distortion and pulse breakup long before the saturation fluence of fiber amplifiers is reached. The solution is to use chirped pulse amplification, a technique originally developed for radar systems to avoid problems caused by high peak powers in the generation of broadband and high energy radio pulses. The basic idea is to spread the frequency components of the pulse in time before amplification, thereby reducing the peak power while maintaining the full spectral bandwidth. The amplified pulse may then be passively recompressed using an equal and opposite amount of dispersion. A typical chirped pulse amplification system is shown schematically in Fig. 13.

Depending on the pulse energies present, either chirped Bragg gratings, lengths of highly dispersive fiber, or bulk diffraction gratings may be used as dispersive delay lines for pulse spreading and recompression. Short Bragg gratings are limited to working with moderate pulse energies, i.e., up to about 100 nJ, before nonlinear effects modify their performance, but are capable of producing large delays in relatively short gratings (i.e., 100 ps/cm) over a finite bandwidth up to hundreds of nanometers. Fibers with large mode areas are advantageous for working with high energy pulses, to minimize the impact of nonlinear effects and to maximize the amplifier saturation fluence. By appropriate choice of components, chirped pulse amplification schemes can deliver high pulse energies at low repetition rates and low average power, or alternatively moderate pulse energies at high repetition rate and high average power. For example, pulse energies of tens of microjoules are theoretically possible, and peak pulse powers of up to 500 kW in 5 ps pulses have been demonstrated in all-fiber chirped pulse amplification systems at megahertz repetition rates.

b. Pulse compression. Nonlinear effects in optical fibers may be used to generate new frequency components from high power pulses, resulting in optical pulses with spectral widths much larger than the gain-bandwidth of optical fiber amplifiers. In many cases the spectrally broadened pulses may be compressed in dispersive delay lines, resulting in femtosecond pulse sources. A variety of techniques have been demonstrated, including (1) supercontinuum generation in low dispersion fiber, (2) spectral broadening in normal dispersion fiber followed by dispersive pulse recompression, and (3) high order soliton compression in anomalous dispersion fiber.

c. Frequency conversion. While a wide range of output wavelengths are accessible using different rare-earth-doped mode locked fiber lasers, it is nevertheless desirable to be able to produce ultrashort pulses at other wavelengths. Optical frequency conversion is required for such purposes, and periodically poled (i.e., quasi-phase-matched) lithium niobate (PPLN) has proven itself to be an almost ideal medium, providing exceptionally high conversion efficiencies for frequency doubling of femtosecond optical pulses from fiber lasers. Large conversion efficiencies (10–50%) are achieved by accessing the large nonlinearity of lithium niobate, which is possible only by using quasi-phase-matching. For example, fiber lasers at 780 nm with 20 mW average power and 100 fs pulse duration are available using a femtosecond erbium-doped fiber laser frequency-doubled in quasi-phase-matched lithium niobate.

C. Multiwavelength Pulsed Lasers

Multiwavelength lasing in pulsed fiber lasers is sometimes desirable for the same reasons mentioned previously with respect to CW fiber lasers (Section III.E). While mode locked lasers could be considered to be multiwavelength, they generate high repetitive peak power due to enforced phase coherence between the lasing modes, and the lasing wavelengths are restricted in that they must be harmonically related. Therefore the term "multiwavelength" is here used to denote lasers which generate pulses (i.e., groups of mode locked wavelengths) simultaneously at different carrier wavelengths in a single cavity, but with only limited coherence between the different carrier wavelengths.

Achieving multiwavelength pulsed operation in fiber lasers presents some interesting challenges. Similar techniques to those used in CW lasers may be used to define the lasing wavelengths and to overcome gain cross-saturation. Temporal-spectral multiplexing may also be used in pulsed lasers to reduce the effects of gain cross-saturation. For example, a multiwavelength Q-switched laser has been demonstrated using an electrically tuneable filter to modulate the loss in a cavity with lasing wavelengths defined by a comb filter. By sweeping the tunable filter, pulses with different wavelengths were generated at different times. Other approaches must be used to promote multiwavelength lasing in mode locked fiber lasers, as pulses at all wavelengths must be synchronized at the mode locking element. Inhomogeneous broadening of the gain medium facilitates multiwavelength lasing in mode locked lasers, as in CW lasers. It has also been demonstrated that temporal-spectral multiplexing can be used in conjunction with intracavity dispersion to promote multiwavelength lasing in harmonically mode locked fiber lasers.

V. OTHER FIBER LASERS

In this section we briefly review fiber lasers which do not fall neatly into either the CW or mode locked categories, but which, nevertheless, generate distinctive and useful optical outputs.

A. Frequency Shifting Lasers

Lasers incorporating a frequency-shifting element, such as an acousto-optic modulator, in the cavity are interesting in that they may behave in a variety of ways, depending on the specific conditions present. A typical cavity configuration, including spectral filtering, is shown schematically in Fig. 14. In general, the frequency-shifted feedback prevents coherent feedback at zero modulation frequency, thus preventing CW lasing, however, the optical output may be coherent at other frequencies, or incoherent. In neither case is there necessarily any mechanism to lock the "modes" of the cavity in phase. The following cases will be described in turn: (1) low coherence broadband output; (2) periodic linear pulse generation; and (3) spontaneous nonlinear pulse generation.

1. Low Coherence Output

If the cavity may be regarded as linear (i.e., the peak intensity within the cavity is insufficient for the Kerr effect to be significant), and the frequency shift per round trip is not harmonically related to the cavity fundamental frequency, then the laser will generate a broadband output with low coherence, seeded by the amplifier's spontaneous emission noise.

2. Mode Locked Pulse Generation

If the cavity may be regarded as linear, and the frequency shift per round trip is harmonically related to the cavity fundamental frequency, then a comb of "modes" may be generated within the cavity. While the phase relationship

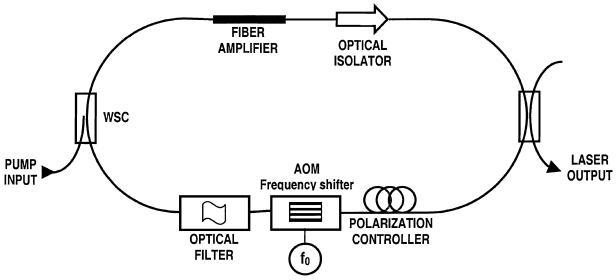


FIGURE 14 Schematic of a frequency-shifted feedback laser, using an acousto-optic modulator (AOM). The laser may generate an incoherent or pulsed output, depending on the frequency shift.

between the shifted wavelengths is not under direct control, the result is a periodic output from the laser, with the repetition rate determined by the frequency-shift applied per round trip.

3. Spontaneous Pulse Generation

If the cavity is nonlinear (i.e., the peak intensity within the cavity is sufficient for SPM to have an effect), and incorporates a bandlimiting element (e.g., a filter, or bandlimited gain), then nonlinear pulses (i.e., solitons) may be generated spontaneously. The solitons may be regarded as nonlinear modes of the cavity, i.e., they are the only self-replicating pulses which may circulate in the cavity with no net loss. Whereas any wavelengths corresponding to CW lasing or linear pulses are rapidly attenuated by being progressively shifted further into the stopband of the intracavity filter, SPM allows solitons to generate energy at wavelengths in the passband of the filter, thereby compensating for filter losses. The sliding-frequency laser has much in common with sliding-frequency and sliding-guiding-filter soliton transmission systems.

Perturbation analysis has shown that the pulse width, in picoseconds, of solitons generated in a frequency-shifted laser is given by

$$\Delta\tau = \frac{1.44}{\sqrt[3]{\pi \cdot 10^{-6} \cdot f_s \cdot \Delta\nu^2}},$$

in which f_s is the frequency shift per round trip, in megahertz, and $\Delta\nu$ is the 3-dB optical bandpass filter bandwidth, in teraradians per second. Somewhat surprisingly, theoretical and experimental results indicate that the pulse temporal characteristics are independent of the cavity dispersion.

Pulses with 10-ps duration are readily produced in sliding-frequency lasers, though much shorter pulses are difficult to generate in practice without the use of hybrid mode locking techniques. Nevertheless, the sliding-frequency soliton laser has a number of advantages over other soliton sources, including a low self-starting threshold, and the suppression of pulse interactions and low intensity background radiation by the filter. Lastly, because pulse generation is controlled in the frequency domain rather than the time domain, the use of a comb filter in the cavity allows soliton generation at multiple wavelengths.

B. Chaotic Fiber Lasers

It has been known since 1975 that chaotic behavior is possible in some lasers under specific conditions, due to the similarity between the Lorenz equations (which predict chaos in fluids) and the semiclassical Maxwell-Bloch equations describing single mode lasers including the effects of gain dynamics. Due to the relatively long lifetime of the metastable state in rare-earth-doped fiber amplifiers, and the largely homogeneous gain, fiber lasers studied to date often exhibit damped relaxation oscillations, but usually not chaotic behavior. Nevertheless, other mechanisms capable of producing chaotic behavior in fiber lasers have since been discovered, analogous to a driven second-order system, or to third and higher order dynamic systems. Such chaotic behavior is characterized by a seemingly nondeterministic output in the time-domain, and by significantly broadened bandwidth in the frequency domain.

A fiber laser may behave as a driven second-order system by introducing a time-dependent parameter into the laser cavity, such as loss modulation, gain (or pump) modulation, phase modulation, etc. For example, driving the mode locking element in a mode locked fiber laser at slightly below the cavity fundamental frequency has been observed to result in chaotic behavior, characterized by severe amplitude jitter on the optical pulses generated.

A fiber laser may behave as a third or higher order dynamic system by: (1) adding a saturable absorber into the cavity or (2) coupling two or more cavities together. For example, using highly doped optical fiber amplifiers exhibiting upconversion effectively adds a saturable absorber into the laser cavity, resulting in self-Q-switching and chaotic oscillation when just above threshold. The period of the self-pulsing in the latter case increases with the pump rate, but is typically in the kilohertz range for erbium-doped amplifier systems. The instability can be enhanced by gain competition and coupling between, for example, forward and backward propagating waves in a bidirectional ring laser. Lastly, under some circumstances the Kerr nonlinearity and dispersion of fiber in the cavity may have significant effects, giving rise to complicated dynamics with significantly shorter time constants, including chaotic behavior.

C. Superfluorescent Sources

A superfluorescent source is effectively an open-loop laser cavity, i.e., with no feedback. A typical example of a double-pass backward-pumped source is shown schematically in Fig. 15. The output consists of spontaneous emission noise which has been increased in power by passage through the amplifying fiber once or twice, ideally converting the input pump power to a broadband output with

FIGURE 15 Schematic of a double-pass superfluorescent laser, for producing a broadband incoherent output.

low coherence. The advantage of superfluorescent sources over semiconductor devices such as light-emitting and superluminescent diodes is their higher fiber-coupled output power and superior spectral stability.

The main difficulty encountered in optimizing superfluorescent fiber sources is that the output bandwidth reduces (i.e., coherence increases) as the pump and output power is increased. The latter may be understood by considering the fiber amplifier as a series of gain elements with finite bandwidth; increasing the gain of all the elements also narrows the end-to-end bandwidth of the amplifier, and hence narrows the FWHM bandwidth of any amplified spontaneous noise passing through the system. A secondary problem is the peak emission wavelength changes with temperature and/or pump power. The latter problems may largely be overcome by optimizing the cavity configuration, the amplifying fiber length, and the pump power for specific operating conditions. Some spectral filtering may also be included in the system. Using the latter techniques superfluorescent sources using neodymium- or erbium-doped fiber amplifiers have produced output powers over 100 mW and FWHM bandwidths over 30 nm (erbium) or 60 nm (neodymium), with wavelength stability of a few parts per million per degree Celsius.

VI. APPLICATIONS OF FIBER LASERS

The wide range of performance characteristics of various fiber lasers leads them to be useful in a wide variety of applications. This section briefly reviews only a few applications of fiber lasers, highlighting the particular properties which make them attractive in specific applications.

A. Sensing

The sensitivity of systems which detect changes in phase or wavelength ultimately depends on the linewidth of the source, i.e., the narrower source linewidth the better. Similarly, the sensitivity of sensing systems which detect changes in intensity may depend on the source intensity noise. Fiber lasers can have close to quantum-limited linewidth and intensity noise, making them excellent sources for many sensing applications.

1. Strain and Temperature Sensing

Localized strain and/or temperature sensing is possible with very high sensitivity using two-mode DFB lasers, in which changes in temperature and/or strain cause a change in the wavelength difference, or beat frequency, between two simultaneous but orthogonal polarization modes of the laser. By measuring both the beat frequency and absolute wavelength of one polarization mode, the effects of temperature and strain may be separated and measured independently, overcoming the problem of cross sensitivity of fiber Bragg gratings to temperature and strain. Accuracy of simultaneous strain and temperature measurements using birefringent DFB lasers has been reported to be better than $\pm 3\,\mu\varepsilon$ and $\pm 0.04^{\circ}$C, respectively. Dual polarization mode DFB lasers can also be used to measure transverse force, which changes the fiber birefringence. In principle, wavelength division multiplexing techniques may be used to perform sensing at multiple locations in the one fiber, e.g., using a single pump source to pump several lasers with different output wavelengths.

2. Gas Sensing

Many molecules have fundamental rotational or vibrational resonances in the mid-infrared wavelength range, which can cause strong optical absorption at wavelengths longer than 2 μm. Long wavelength fiber lasers in low-phonon energy glasses, such as discussed in Section III.D.2, are therefore of interest as sources for gas detection systems. For example, thulium-doped fiber lasers are tunable around 2.3 μm, and of particular interest for hydrocarbon sensing. Some reports indicate that extremely high detection sensitivity of methane (50 ppm meters) is possible.

3. Rotation Sensing

Optical fiber gyroscopes are essentially a Sagnac interferometer or fiber loop mirror, in which light propagates in opposite directions around a fiber loop. Rotation in the plane of the loop causes a slight phase difference between the counter propagating waves, and thereby causes a change in loop reflectivity. Rotation rates much less than one degree per hour are detectable. The properties required of optical sources for use in high performance gyroscopes are broad bandwidth (i.e., >10 nm), good wavelength stability (i.e., <1 ppm), and high power. The broad bandwidth is required to reduce coherence errors due to Rayleigh backscattering, and good wavelength stability (e.g., with variations in temperature and pump power) is required to avoid changes in scale factor or sensitivity. Optical fiber superfluorescent sources are attractive for use in fiber gyroscopes, providing larger output power and significantly better wavelength stability than superluminescent semiconductor diodes.

B. Optical Fiber Telecommunications

The primary advantage of fiber lasers over semiconductor lasers for applications in telecommunications is the

reliability and low loss achievable in coupling to the transmission fiber. Furthermore, fiber lasers have narrower linewidths than semiconductor lasers, useful in high bit rate telecommunications where the source must often be externally modulated to minimize dispersive effects. Lastly, fiber lasers may be tuned over a wider range of wavelengths than semiconductor DFB lasers, which is desirable for applications in wavelength division multiplexed systems.

1. Digital Telecommunications

Optical solitons can behave as natural optical bits of information in fiber transmission systems, propagating over large distances almost unaffected by chromatic and polarization mode dispersion, and retaining their particle-like nature through significant perturbations, making them ideal candidates for use in all-optical switching systems. Furthermore, ultrashort transform limited pulses generated in harmonically mode locked fiber lasers have excellent stability in both wavelength and intensity. All the latter properties are required to maintain usefully low bit error rates in high-bit-rate optical fiber telecommunication systems. The prevalence of wavelength division multiplexed communication systems, i.e., using many sources at lower bit rates, has slowed the application of fiber lasers in this area, but the situation is likely to change as time division multiplexed optical fiber communication systems become more common.

2. Analog Telecommunications

A major emerging application of optical fibers is in the distribution of radio signals for broadband services and wireless communication systems; fiber can provide a low loss backbone with large bandwidth. Such systems require optical sources with narrow linewidth, low noise, and moderate output power. Externally modulated single mode fiber lasers are therefore well suited to applications in these systems. Dual wavelength fiber lasers (e.g., dual polarization mode DFB lasers) are also potentially useful for the generation of millimeter wave radio carriers.

C. Medical Applications

Optical fiber endoscopes are minimally invasive tools used in medical diagnosis and therapy. Optical fibers facilitate the combination of low power diagnostic functions (imaging and sensing) with high power for therapeutic functions (surgery) in the one device. Optical sources with very different characteristics in both wavelength and peak and average powers are required to implement the different functions.

1. Imaging and Sensing

Biological tissue is relatively transparent in the near-infrared, but highly scattering. Various techniques may be used to overcome or alleviate the problem of scattering, such as optical coherence tomography (OCT) and multiphoton excitation. The latter techniques, among others, enable the powerful diagnostic capabilities of light to be extended at least a small distance into biological tissue, e.g., to provide high resolution imaging of tissue morphology and localized optical spectroscopy. High resolution OCT requires broadband optical sources with low coherence and high brightness. Superfluorescent rare-earth-doped fiber sources are superior to light-emitting and superluminescent diodes in this application, providing significantly higher power over a broader bandwidth. Ultrashort optical pulses with high peak power and moderate average power are required for multiphoton excitation and imaging. Mode locked fiber lasers are likely to prove preferable to large, expensive, and inefficient bulk lasers in these applications.

2. Laser Surgery

High power diode-pumped rare-earth-doped fiber lasers with about 1 W average power around the 2 and/or 3-μm absorption bands of water are capable of providing high quality cutting and ablation of tissue for surgical procedures. Cladding pumped erbium-, holmium-, or thulium-doped ZBLAN fibers are the key elements in these systems, resulting in compact sources with much higher efficiency than comparable bulk lasers. The absorption of water at 3 μm is two to three orders of magnitude larger than at 2 μm, and consequently 3-μm sources are preferred for precise cutting without thermal damage, while the larger penetration of 2-μm sources results in coagulation of the incised tissue. Holmium fiber lasers generating both wavelengths simultaneously have been demonstrated, and may provide the optimum combination of both sources.

VII. FUTURE DEVELOPMENTS

Developments in fiber laser technology have mostly followed innovations in fiber fabrication and processing technology. For example, the development of rare-earth-doped single mode optical fibers with low loss was a crucial step, which enabled rapid and significant advances in fiber

laser technology, as in other systems using fiber amplifiers. The latter advances were then sustained and extended by the development of UV-written in-fiber Bragg grating fabrication technology, and aided by developments in related fields such as semiconductor and integrated optics. The development of high power pump laser arrays and improvements in semiconductor saturable absorbers and integrated waveguide modulators have also made noticeable impacts.

Two emerging optical fiber fabrication technologies will potentially have a significant impact on the development of fiber lasers. The first is polymer optical fibers. While low-loss single mode polymer fibers are still in the relatively early stages of development, they could be doped with a wide range of new materials, such as laser dyes, which would result in fiber lasers at a wide range of new wavelengths. Furthermore the large optical nonlinearity of many polymers may allow the use of novel mode locking and frequency conversion techniques within polymer fiber lasers. Secondly, it is fairly certain that another recent innovation in optical fiber fabrication, known as "holey" or "photonic crystal" optical fibers, will have a major impact on future optical fiber systems, including fiber lasers. These fibers are fabricated with a periodic array of holes running parallel to the fiber axis. Photonic crystal fibers have some very interesting and unusual properties, with the potential to take optical fiber lasers into entirely new operating regimes and novel applications.

SEE ALSO THE FOLLOWING ARTICLES

FIBER-OPTIC CHEMICAL SENSORS • LASERS, NUCLEAR PUMPED • LASERS, SOLID-STATE • LASERS, ULTRAFAST PULSE TECHNOLOGY • NONLINEAR OPTICAL PROCESSES • OPTICAL AMPLIFIERS • OPTICAL FIBERS, FABRICATION AND APPLICATIONS • OPTICAL FIBER TECHNIQUES FOR MEDICAL APPLICATIONS • OPTICAL WAVEGUIDES AND WAVEGUIDE DEVICES • TELECOMMUNICATIONS

BIBLIOGRAPHY

Agrawal, G. P. (1995). "Nonlinear Fiber Optics," Academic Press, San Diego.

Desurvire, E. (1994). "Erbium-doped Fiber Amplifiers: Principles and Applications," Wiley, New York.

Digonnet, M. J. F., ed. (1993). "Rare Earth Doped Fiber Lasers and Amplifiers," Marcel Dekker, New York.

Duling, I. N., ed. (1995). "Compact Sources of Ultrashort Pulses," Cambridge University Press, U.K.

France, P. W., ed. (1991). "Optical Fibre Lasers and Amplifiers," Blackie, Glasgow/London.

Kashyap, R. (1999). "Fiber Bragg Gratings," Academic Press, San Diego.

Siegman, A. E. (1986). "Lasers," Oxford University Press, London.

Silfvast, W. T. (1996). "Laser Fundamentals," Cambridge University Press, U.K.

Weber, M. J., ed. (1995). "CRC Handbook of Laser Science and Technology," CRC Press, Boca Raton, Florida.

Lasers, Semiconductor

Peter J. Delfyett
University of Central Florida

I. Brief Historical Overview
II. Semiconductor Basics
III. Laser Structures
IV. Applications
V. Directions and Trends

GLOSSARY

Auger recombination A three-body collision involving two electrons and a hole (or two holes and an electron). The energy released by the recombination of an electron and a hole is immediately absorbed by another electron (or hole), which then dissipates the energy by emitting phonons.

Band gap energy The energy difference between the conduction and valence bands in a semiconductor material, which also corresponds to the energy required to promote an electron from the valence band to the conduction band.

Conduction band A band of energy levels above the valence band that can support the conduction of electrons.

Degenerate doping The amount of doping required to bring the Fermi energy level to a level comparable to the conduction band energy for electrons and the valence band energy for holes.

Direct band gap semiconductor A semiconductor in which the valence band maximum and the conduction band minimum occur at the same position in momentum space.

Dopant A material which is incorporated into a semiconductor material that adds excess electrons (n-type dopant) or excess holes (p-type dopant).

Electron–hole recombination A radiative (or nonradiative) process which an electron in the conduction band recombines with a hole in the valence band. In the radiative recombination process, photons are generated, while in the nonradiative recombination process, phonons are generated.

Fermi level The energy level that represents a 50% chance of finding a state occupied by an electron.

Indirect band gap semiconductor A semiconductor in which the conduction band minimum and the valence band maximum do not occur at the same place in momentum space. In these semiconductors, electrons in the conduction band minimum require a change in momentum in order to recombine with a hole at the valence band maximum. The recombination process in this case usually requires a phonon.

Nonlinear gain The nonlinear part of the gain in semiconductor lasers, which manifests itself in the gain versus photon density curve. In this case, the gain decreases with an increasing photon density due to finite intraband scattering and dynamic carrier heating.

***p–n* junction** A region which joins two materials of opposite doping. This occurs when *n*-type and *p*-type materials are joined to form a continuous crystal.

Population inversion A necessary condition which must occur in order for the stimulated emission process to occur. This condition occurs when the population of the upper lasing level exceeds the population of the lower lasing level.

Q-switching A method for producing a high-power pulse from a laser system. This is usually accomplished by inserting an optical shutter in the cavity to control resonant Q or quality factor.

Quantum well A material structure that can spatially confine electronic charges to spatial dimensions on the order of tens of angstroms (10^{-10} m). These dimensions are comparable to the de Broglie wavelength of the electron/hole.

Schawlow–Townes equation The equation which predicts theoretically the linewidth of the lasing transition of a single-mode laser. The linewidth predicted from this equation is inversely proportional to the output power and the cavity quality factor (Q-factor).

Valence band The energy band corresponding to the valence electrons of a semiconductor crystal. This energy band is normally filled, and as a result does not allow the conduction of electrical current.

Window structure A modification to the standard laser facets of a semiconductor injection laser. The modification involves processing the laser facets so that the energy gap of the laser facet material is larger than the emitted photon energy. This reduces optical absorption at the laser facets and prevents irreversible facet damage.

SEMICONDUCTOR LASERS are highly efficient emitters of coherent laser radiation. A key characteristic is that these devices are extremely small, with typical linear dimensions being on the order of a few hundred microns ($1\ \mu m = 10^{-6}$ m) on a side. These lasers belong to a specific class of solid state laser, which are constructed from semiconductor materials, as opposed to conventional solid state lasers, which are made from insulating crystals doped with active ions. The semiconductor injection laser derives its input power from an electrical current that is directly passed or injected through the device. The typical threshold currents for initiating lasing in these devices is on the order of a few milliamperes, with conversion efficiencies between the injected electrons and the generated photons exceeding 90%. Other methods of excitation are possible, such as optical pumping and electron beam pumping. However, these methods of pumping are less attractive because they do not take advantage of the compactness of the diode laser, due to the relatively large size of the pumping sources, and the pumping scheme is normally less efficient.

Semiconductor lasers are normally constructed from several semiconductor material systems, most notably the gallium arsenide/aluminum gallium arsenide system and the indium phosphide/indium gallium arsenide phosphide system. The specific material system that is used is generally defined by the desired emission wavelength of the laser. The field of semiconductor lasers began in 1962, just 3 years after the first laser was invented. These early devices were very crude; however, tremendous advances have been made in the development of these devices, allowing many scientific, industrial, and commercial applications. This article first reviews some of the fundamental concepts in semiconductor material physics to lay the groundwork for subsequent discussions. The basic principles of light generation in semiconductor injection lasers are then outlined, which continues with a survey of several important semiconductor laser structures and their applications in the scientific and commercial communities. Finally, the article concludes by contemplating future directions and trends in this rapidly growing and exciting field of research.

I. BRIEF HISTORICAL OVERVIEW

The first semiconductor laser devices were made from chips of gallium arsenide (see Fig. 1). The gallium arsenide was grown such that a *p–n* junction, or diode, was formed inside the crystal. The chip had a metallic base with a wire contact attached to the top to allow the injection of the electrical current. Smooth end faces were formed on the diode, which acted as mirrors to provide the optical feedback necessary to attain laser oscillation, while the sidewalls of the laser chip were roughened to prevent laser oscillation in the direction perpendicular to

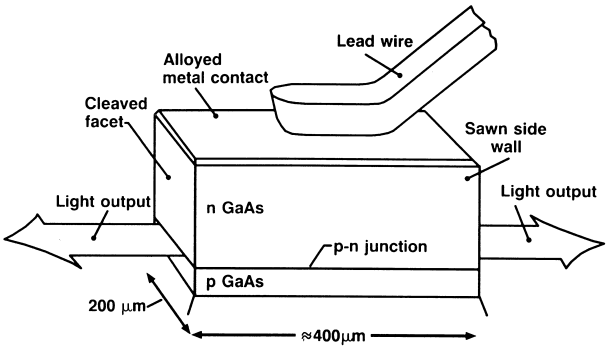


FIGURE 1 Broad-contact semiconductor injection laser mounted on a heat sink with a wire contact. [After Thompson (1980).]

the desired direction of lasing. These devices had very high threshold currents and could only be operated at very low temperatures. The light output characteristics from these laser chips were far from what their solid state and gas laser counterparts could produce. The laser emission from the chips suffered from a lack of coherence due to the wide spectral bandwidth inherent in all semiconductor light-emitting devices, and the output beam emission was contained in a very broad far-field pattern. Despite these initial drawbacks, it was apparent that semiconductor lasers would have a very promising future. Semiconductor lasers have been developed to the point where they are one of the key components in commercial home entertainment systems, such as video and compact disk players.

The techniques for developing semiconductor lasers are identical to the technology used for manufacturing electronic devices. This means that semiconductor lasers can be mass-produced with a comparable reliability to standard electronic components. As a result of this link in processing technology, electronic devices can be integrated with semiconductor lasers on the same wafer. This feature has had a tremendous impact on the fields of integrated optoelectronics, optical communications, and optical data storage, and is what makes semiconductor lasers very attractive for technological and commercial applications.

II. SEMICONDUCTOR BASICS

A. The Energy Band Theory of Crystals and the Origin of the Band Gap

In order to provide a complete picture of how a semiconductor laser operates, it is important to understand the fundamental concepts that form the basis for the modern theory of how electrons interact in crystalline solids. The salient features of the theory arise from the natural assumption that semiconductors are solids where the atoms are arranged in a periodic crystal structure or lattice. The periodicity of the lattice leads to the formation of energy bands that the electrons can occupy. One can obtain substantial insight into the structure of these energy bands owing to the periodicity of atoms, if the periodicity of the atoms generates a periodic potential that is weak. The idea of the structured arrangement of atoms generating a weak potential for electrons to interact with turns out to be an excellent approximation. This can be understood if one considers that metal atoms, especially in groups I, II, and III–IV of the periodic table, are atoms with *s* and *p* electrons outside of a closed-shell noble gas configuration. These *s* and *p* electrons are then viewed as conduction electrons (nearly free electrons) that move about in a potential with near-perfect uniformity. In other words, the description of the movement of these electrons follows closely that of a perfect free electron gas, which then is modified by the presence of a weak periodic potential. Figure 2 shows a simple picture of a periodic chain of atoms, the weak potential owing to the atoms, and the existence of a conduction electron wavefunction along the atomic chain.

One might initially think that these conduction electrons would interact strongly with each other and with the positively charged atomic nucleus. However, it is owing to the presence of these free electrons that the potential is weak. This is because the electron–ion interaction is strongest at small separations and the Pauli exclusion principle forbids the conduction electrons from entering the immediate neighborhood of the ions because the core electrons already occupy this region. In addition, in the region where the conduction electrons are allowed, their high mobility further reduces the net potential any single electron experiences, for they can block, or screen out, the fields of positively charged ions, diminishing the total effective potential.

Owing to the weak potential, we can determine the band structure of the semiconductor by first considering the

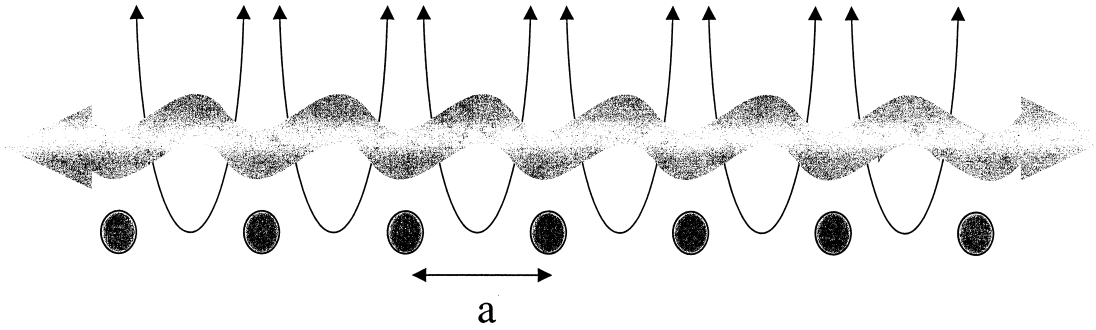


FIGURE 2 Representation of an electron wave propagating along a periodic chain of atoms. The periodic potentials due to the atoms are also shown.

solution to Schrödinger's equation for an electron in a constant potential. The equation that describes the motion of the electron is given by $\exp(i\mathbf{k}\cdot\mathbf{r})$, where $\mathbf{k}$ is the wavevector of the electron and $\mathbf{r}$ is the direction taken by the electron. The energies and k values of the free electrons in a cube of side L are given by

$$E = \frac{\hbar^2 k^2}{2m} \qquad \text{and} \qquad k = \frac{2\pi}{L}n, \tag{1}$$

where m is the effective mass of the electron, $\hbar$ is Planck's constant divided by 2π, and $n = 0, \pm 1, \pm 2, \ldots$, resulting from the periodic boundary conditions that allow these discrete values of k. For the case of a free electron (no periodic potential), the E versus k diagram is a parabola (see Fig. 3).

The addition of a weak periodic potential modifies the solution of Schrödinger's equation such that the resulting E versus k diagram also becomes periodic. To show schematically the effect of the periodic potential on the E versus k diagram, we can first consider the effect in one dimension, with the results for three dimensions following by analogy. First, it should be recognized that electrons, while being primarily described as discrete particles, also exhibit wavelike behavior. In one dimension, as an electron travels in the direction of the atomic chain, the electrons will be reflected, or Bragg scattered, in the reverse direction if the energy or $\mathbf{k}$ vector is large enough (or the electron wavelength is short enough). For the one-dimensional case, Bragg reflection occurs for wavevector $k = n\pi/a$, where a is the distance between atoms. For electrons with an energy or wavevector of $k = n\pi/a$, a bandgap occurs due to these reflections, and a steplike discontinuity is created in the E versus k diagram (see Fig. 3a).

To see how the band gap is created graphically, one can plot the E versus k diagram for a free electron and superimpose another free electron diagram shifted from the original by an amount of $k = 2\pi/a$, as show in Fig. 3b. At the point of intersection of these two $E-k$ diagrams, i.e., $k = \pi/a$, a gap is formed owing to the interference of the two forward- and backward-propagating electron wave functions. Since the crystal lattice is periodic with a lattice spacing of a, a multiplicity of E–k diagrams are drawn, each separated by $k = n\pi/a$. The effective total composite E–k diagram is represented by the extended zone representation. It should be noted that we can specify all the levels of a wavevector k within the first Brillouin zone. As a result, to plot the resultant E–k diagram, the individual segments in the extended zone scheme must be shifted into the first Brillouin zone, giving the reduced zone scheme. From this diagram in the reduced zone scheme, the production of an energy gap at $k = 0$ between two energy bands is clearly evident. This energy gap is the familiar gap between the conduction band and energy band in semiconductors.

B. Energy States and the Density of States

In order to achieve lasing action in semiconductor medium, the step of achieving a population inversion is required, as in any conventional laser. To determine whether a population inversion can be achieved, one needs to know the concentrations of electrons in the conduction band and holes in the valence band as a function of energy. In order to know these relative concentrations, it is first required to know the density of allowed energy levels, or density of states, and the probability that each of these levels is occupied.

As noted earlier, an electron in a semiconductor material is characterized by its energy E, its wavevector $\mathbf{k}$, and its spin. Near the conduction band edge, an electron can be described as a particle of mass m_C confined to a three-dimensional cubic box of dimension d with perfectly reflecting walls, i.e., a three-dimensional infinite rectangular potential well. The standing wave solutions require that the components of the wavevector $\mathbf{k} = (k_X, k_Y, k_Z)$ assume the discrete values $\mathbf{k} = (q_1\pi/d, q_1\pi/d, q_1\pi/d)$, where the q's are positive mode integers. The tip of the wavevector $\mathbf{k}$ must lie on the points of a lattice whose cubic unit cell has dimension π/d (see Fig. 3c). As a result, there are $(\pi/d)^3$ points per unit volume in $\mathbf{k}$ space. The number of states whose wavevectors $\mathbf{k}$ have magnitudes between 0 and k is determined by counting the number of points lying within the positive octant of a sphere of radius k, $V = \frac{1}{8}(4\pi k^3/3) = \pi k^3/6$. It should be noted that each point in $\mathbf{k}$ space has two values, one each for the two spins of an electron. As a result, there are approximately $2(\pi k^3/6)/(\pi/d)^3 = (k^3/3\pi^2)d^3$ such points in the volume d^3 and $(k^3/3\pi^2)$ points per unit volume. It follows that the number of states with electron wavenumbers between k and $k + \Delta k$ per unit volume is $\rho(k)\Delta k = [(d/dk)(k^3/3\pi^2)]\Delta k = (k^2/\pi^2)\Delta k$, so that the density of states is $\rho(k) = k^2/\pi^2$.

If $\rho_C(E)\Delta E$ represents the number of conduction band energy levels per unit volume lying between E and $E + \Delta E$, then owing to the one-to-one correspondence between E and k determined by $E = E_C + (\hbar^2 k^2/2m_C)$, the densities $\rho_C(E)$ and $\rho(k)$ must be related by $\rho_C(E)dE = \rho(k)dk$. Thus the allowed energy in the conduction band is $\rho_C(E) = \rho(k)/(dE/dk)$, with an analogous result for electrons in the valance band. The expression for dE/dk is obtained by using the approximate quadratic $E-k$ relation that is valid near the conduction and valance band edges. The final, resulting expression

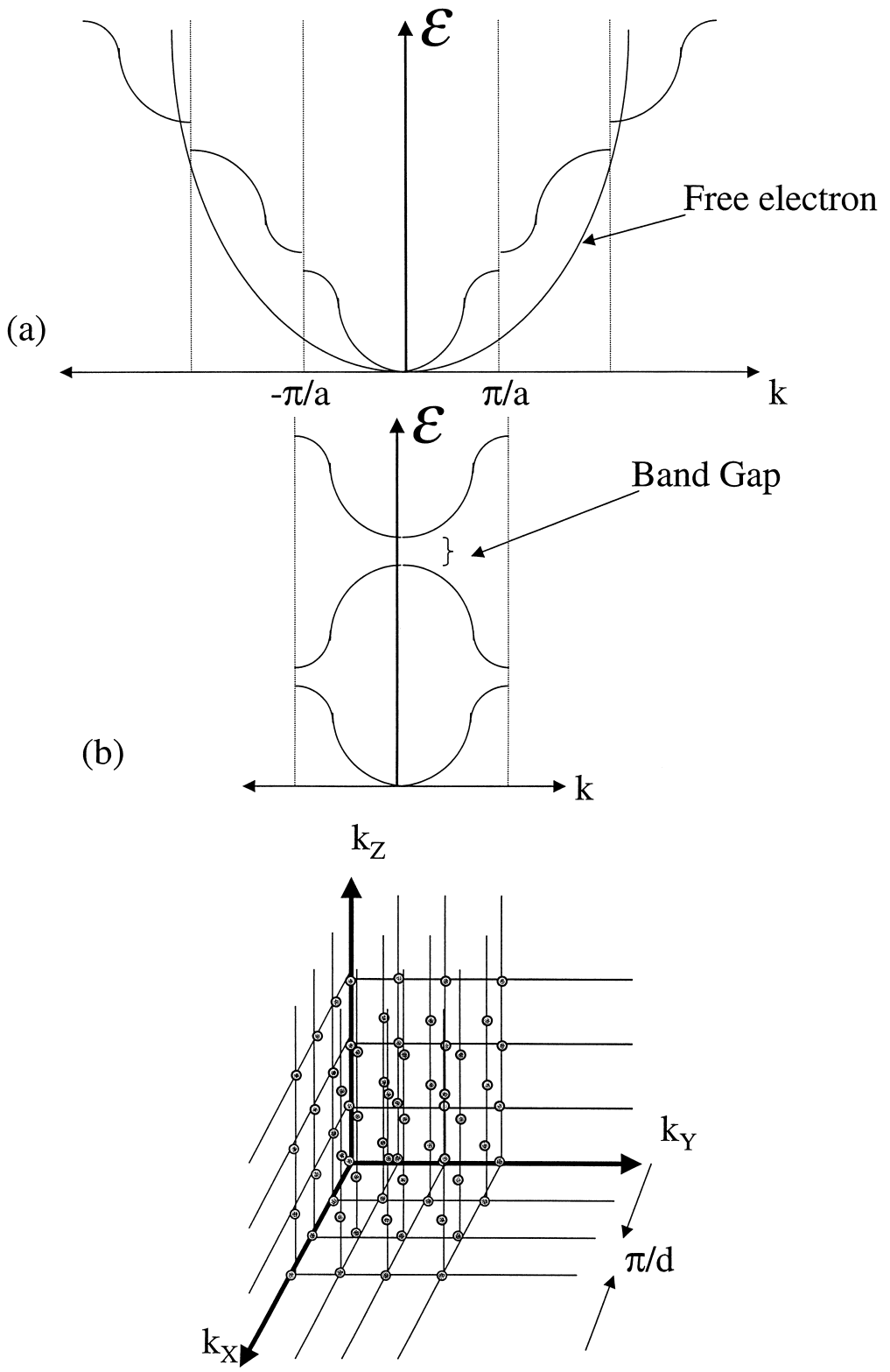


FIGURE 3 The energy versus momentum curve of an electron traveling in a periodic potential. (a) Extended zone representation, (b) reduced zone representation, and (c) the endpoints of the wavevector for a three-dimensional cubic box of dimension *d*.

for the density of states near the conduction and valance band edges is then given by

$$\rho_C(E) = \frac{2m_C^{3/2}}{2\pi^2\hbar^3}(E - E_C)^{1/2}, \qquad E \geq E_C$$
$$\rho_V(E) = \frac{2m_V^{3/2}}{2\pi^2\hbar^3}(E - E_V)^{1/2}, \qquad E \geq E_V \tag{2}$$

for the conduction band and valence band, respectively, where E_C and E_V are the conduction and valence band energies, and m_C and m_V are the effective masses of the electron in the conduction band and valence bands, respectively. In the absence of thermal excitation ($T = 0$), all electrons must occupy the lowest possible energy levels, owing to the Pauli exclusion principle. When the temperature is raised, the thermal excitations raise some electrons from the valence band to the conduction band, leaving behind empty states, or holes, in the valence band. The probability that a given state of energy E is occupied by an electron is determined by the Fermi function.

$$f(E) = \frac{1}{\exp[(E - E_f)/k_B T] + 1}, \tag{3}$$

where k_B is Boltzmann's constant and E_f is a constant known as the Fermi energy. Given the number density of available states and the relative probability that a state is occupied, the number of occupied states is simply given by

$$n(E) = \rho_C(E) f(E), \qquad p(E) = \rho_V(E)[1 - f(E)]. \tag{4}$$

Figure 4 shows a plot of the Fermi distribution, density of states, and the number of occupied states for a typical intrinsic direct gap semiconductor.

For a semiconductor laser, a key requirement is the possibility of obtaining a population inversion between the conduction and valence bands. In this case, one must have materials with an excess density of electrons in the conduction band and holes in the valence band. One can achieve this excess in semiconductor materials by doping the material with donor or acceptor atoms, for *n*- and *p*-type materials, respectively. By doping the semiconductor, the effect is to modify the probability statistics by shifting the Fermi function to a new Fermi level. Figure 5 shows the carrier concentration for *n*- and *p*-type doped semiconductors.

C. Quantum Well Physics

As will be shown later, certain types of semiconductor lasers can be made of semiconductor quantum structures,

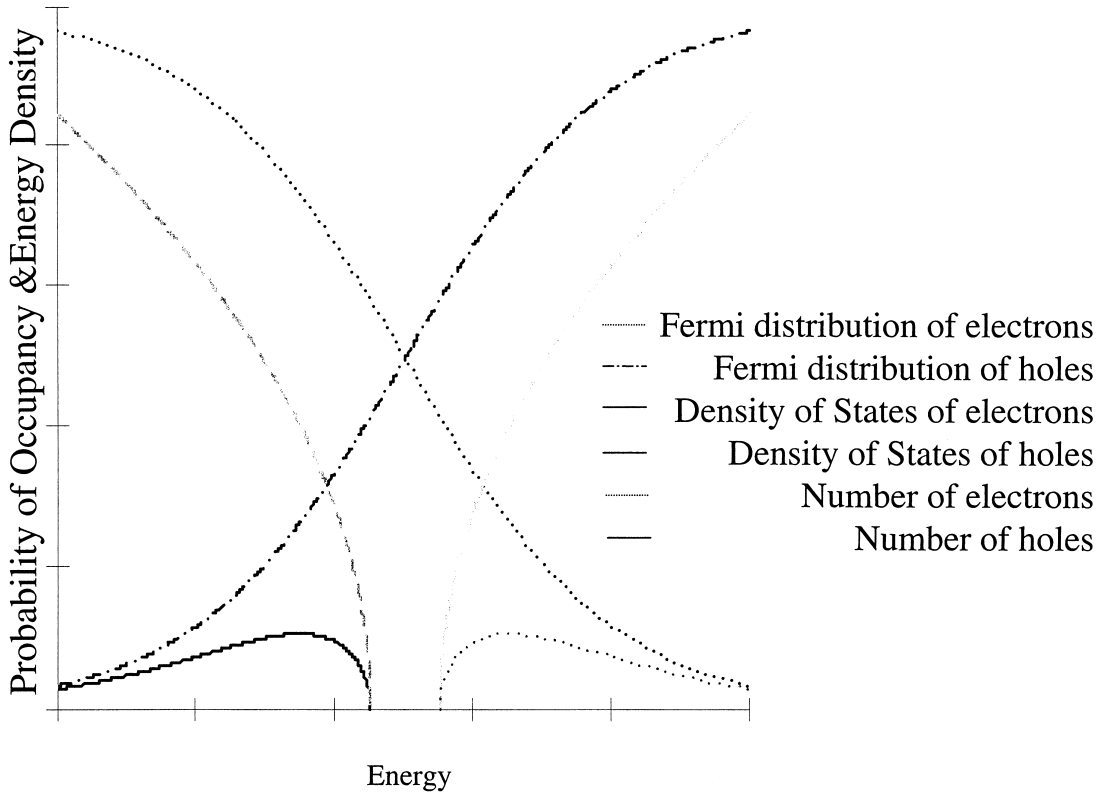


FIGURE 4 Plot of the Fermi distribution, density of states, and number of occupied states for a typical intrinsic direct gap semiconductor.

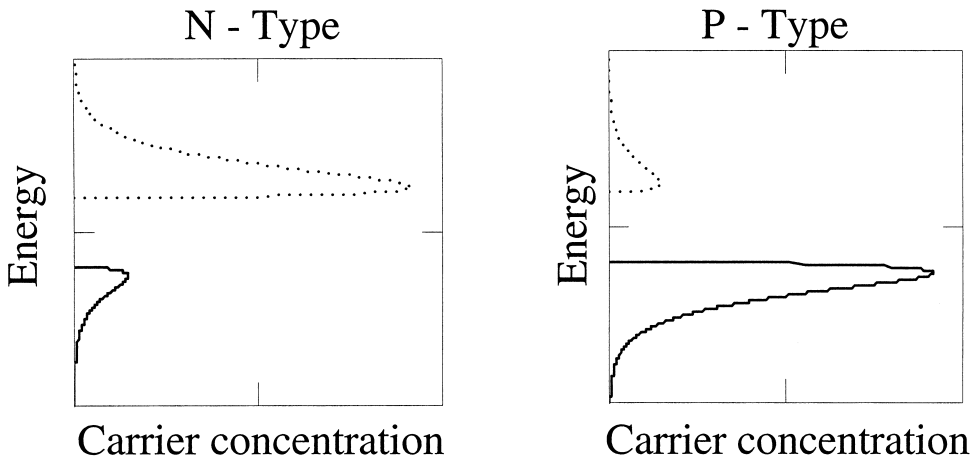


FIGURE 5 Plots of the carrier concentration for *n*- and *p*-type doped semiconductors.

such as quantum wells, quantum wires, and quantum dots. The use of these specially designed semiconductor materials leads to unique operating characteristics. In order to understand the implications of these materials, a short description of their fundamental physical characteristics is provided. Quantum well materials are made of thin layers of differing semiconductor materials, and can be grown such that the movement of the free electron is confined to a two-dimensional plane. It should also be noted that this confinement can also be achieved in one dimension (quantum wires) as well as zero dimensions (quantum dots). The confining of free electrons in two dimensions (or holes) can easily be accomplished by epitaxially growing a double heterostructure consisting of a very thin, e.g., 100 Å, layer of semiconductor material surrounded by different semiconductor material whose bandgap is larger than the thin layer. For example, a typical quantum well structure could comprise a thin layer of GaAs surrounded by AlGaAs. This process of surrounding an ultrathin layer of semiconductor material forms potential wells for the conduction band electrons and the valence band holes, such that these conduction band electrons and valence band holes are confined within the conduction band well and the valence band well, respectively.

The energy levels E_q of a particle of mass $m_{c,v}$ that is confined to a sufficiently deep one-dimensional rectangular potential well of width d, typically approximated as an infinitely deep well, can easily be solved by using Schrödinger's equation. The resulting energy levels are

$$E_q = \frac{\hbar^2(q\pi/d)^2}{2m}, \qquad q = 1, 2, 3, \ldots, \tag{5}$$

and E_q represents the qth energy level.

As an example of a quantum well heterostucture, Fig. 6 shows the geometry of the structure, energy level diagram, the $E-k$ relation, and the density of states of electrons and holes. The electrons and holes are confined in the x direction to within a distance d_1, the thickness of the well. On the other hand, the electrons and holes can move freely in the plane of the confining layer. As a result, in the y–z plane, the electrons and holes behave as if they were in a conventional bulk semiconductor. The energy–momentum relation is

$$E = E_C + E_{q1} + \frac{\hbar^2 k^2}{2m_C}, \qquad q_1 = 1, 2, 3, \ldots, \tag{6}$$

where E_C is the conduction band energy of the bulk semiconductor, E_{q1} is the energy of the first allowed energy level associated with the quantum well layer, and the third term is the dispersion relation of the quantum well layer, where k is the magnitude of the two-dimensional **k** vector that lies in the y–z plane.

It should be noted that the salient difference between the energy–momentum relation between bulk semiconductor and quantum well material is that the **k** vector associated with E_{q1} takes on discrete, well-separated values. In the quantum well device, the density of states is obtained from the magnitude of the two-dimensional **k** vector associated with the y–z plane, as compared to the three-dimensional wavevector for the bulk semiconductor. As a result, the final density of states for the quantum well structure is given by

$$\rho_C(E) = \begin{cases} m_C/\pi\hbar^2 d_1, & E > E_C + E_{q1} \\ 0 & E < E_C + E_{q1} \end{cases} \tag{7}$$

for q_1 values of 1, 2, ... This result shows that the density of states per unit volume of a quantum well is constant for energies $E > E_C + E_{q1}$. The total density of states is then just the sum of the density of states for all values of q_1. The resultant density of states distribution takes on the form of a sum of constants, where each constant is

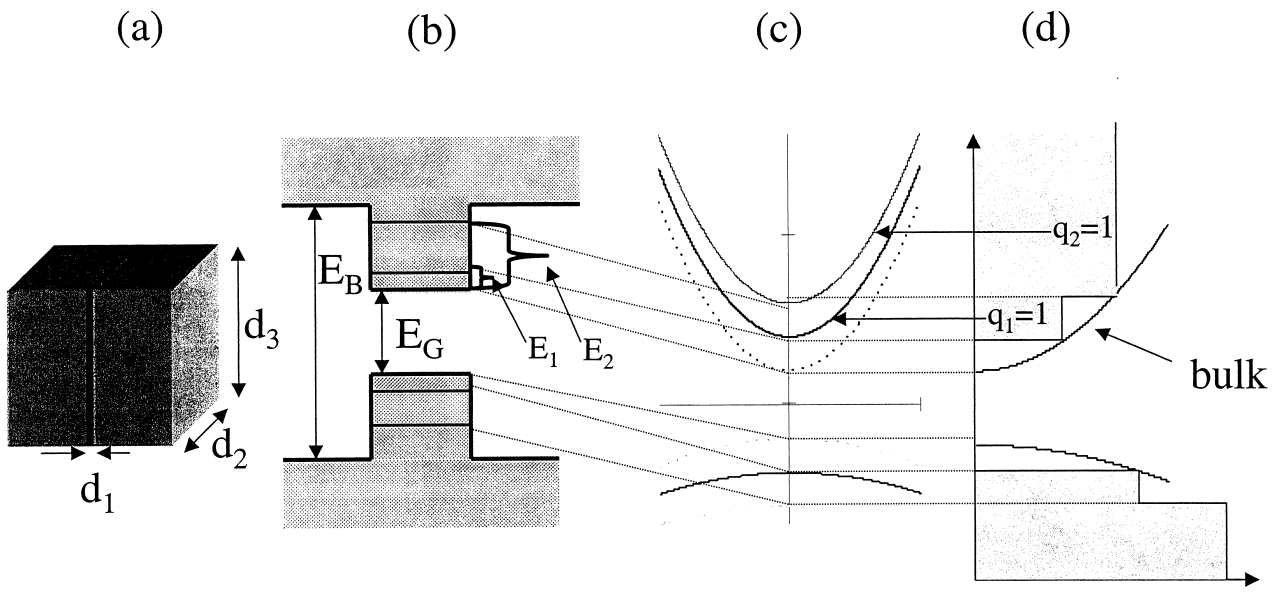


FIGURE 6 Quantum well semiconductor structure. (a) Geometry of the quantum well. (b) Energy level diagram for the conduction and valence bands. (c) Cross-sectional view of the energy versus wavevector relation for electrons and hole traveling in the plane of the quantum well. (d) Density of states of a quantum well structure as compared to a bulk semiconductor.

"turned on" for each value of q_1. The distribution is thus in the shape of a steplike distribution, where each step of the distribution corresponds to a different quantum number q_1 and can be considered as a subband within the conduction band. It should be noted that while the above discussion focused on the conduction band, a similar relation results for the valence band as well. An important difference contrasting the density of states in bulk media as compared to quantum well structures is that the quantum well structure possesses a substantial density of states at its lowest allowed conduction band energy and at its highest allowed valence band energy level. This resulting dramatic effect lowers the required amount of current applied to achieving lasing in quantum well lasers.

Quantum Wires and Quantum Dots

In a manner analogous to confining the movement of an electron to a two-dimensional plane as in a quantum well structure, additional confinement conditions can also be imposed to realize confinement in both one and zero dimensions, to generate quantum wire and quantum dot structures (see Fig. 7). These structures can be realized physically by having a semiconductor material take on the form of an ultrasmall rectangular-shaped wire or an ultrasmall box, where the surrounding material is another semiconductor material of wider bandgap. For the quantum wire structure, the wire acts as a potential well that narrowly confine electrons and holes in two dimensions, say x and y. Assuming that the cross-sectional area is $d_1 d_2$, we find for the energy relation in the conduction band

$$E = E_C + E_{q1} + E_{q2} + \frac{\hbar^2 k^2}{2m_C}, \tag{8}$$

where

$$E_{q1} = \frac{\hbar^2 (q_1 \pi / d_1)^2}{2m_C},$$

$$E_{q2} = \frac{\hbar^2 (q_2 \pi / d_1)^2}{2m_C}, \qquad q_1, q_2 = 1, 2, \ldots,$$

and k is the wavevector component in the z direction, which in this case is the direction along the length of the wire. For each pair of quantum numbers q_1 and q_2, there is an energy subband with a density of states $\rho(k) = 1/\pi$ per unit length of the wire and therefore $1/\pi d_1 d_2$ per unit volume. The corresponding density of states per unit volume is

$$\rho_C(E) = \begin{cases} \dfrac{(1/d_1 d_2)\left(m_C^{1/2}/\sqrt{2}\pi\hbar\right)}{(E - E_C - E_{q1} - E_{q2})^{1/2}}, & E > E_C + E_{q1} + E_{q2} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

for $q_1, q_2 = 1, 2, \ldots$.

In a quantum dot structure, the electrons are confined in all three dimensions within an ultrasmall box

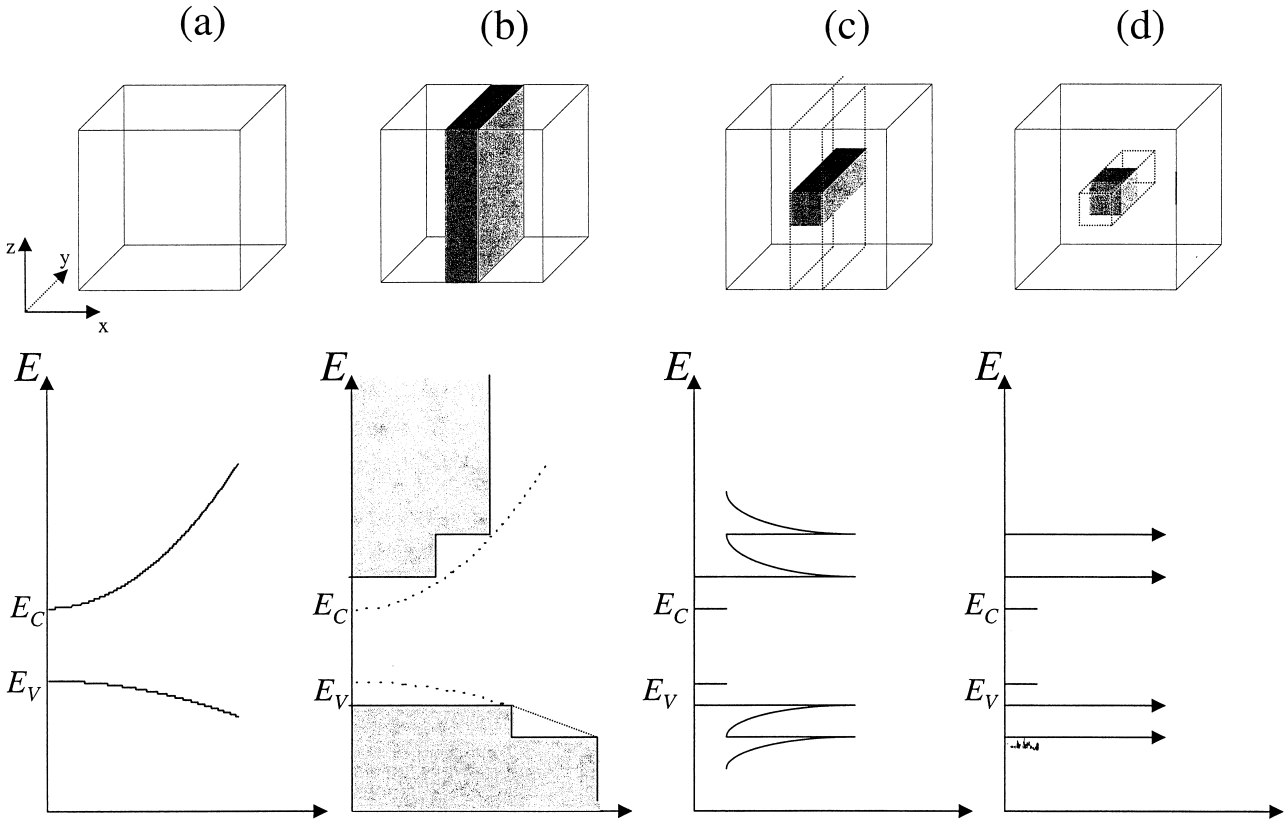


FIGURE 7 Density of states for (a) bulk semiconductor, (b) quantum well, (c) quantum wire, and (d) quantum dot.

of volume $d_1 d_2 d_3$. The energy is therefore quantized to $E = E_C + E_{q1} + E_{q2} + E_{q3}$, where

$$E_{q1} = \frac{\hbar^2 (q_1 \pi / d_1)^2}{2m_C}, \qquad E_{q1} = \frac{\hbar^2 (q_1 \pi / d_1)^2}{2m_C}, \qquad E_{q1} = \frac{\hbar^2 (q_1 \pi / d_1)^2}{2m_C} \tag{10}$$

for $q_1, q_2, q_3 = 1, 2, \ldots$. The allowed energy levels associated with the quantum box structure are discrete and well separated so that the density of states is represented by a sequence of impulse or delta functions at the allowed energy levels. An important feature of the quantum dot structure is that the discrete energy levels can be chosen by appropriately choosing the semiconductor material and designing the size and dimensions of the quantum box.

D. Basic Principles

The basic principles involved in the light generation process of semiconductor injection lasers can most easily be described by examining the energy level diagram of a *p–n* junction made from a direct band gap semiconductor. Indirect band gap semiconductors are not used for semiconductor lasers due to the nonradiative processes, which are the dominant mechanism for electron–hole recombination. These processes generate lattice vibrations, or phonons, which are mechanisms for generating sound and heat, and are not useful for light generation. Figure 8 shows an energy level diagram of a *p–n* junction. Two semiconductor layers with opposite doping are grown in contact with each other. The *n*-doped side has an excess number of electrons; the *p*-doped side has an excess number of holes. The doping levels are typically 10^{18} cm^{-3}, which assures that the Fermi level, shown as the dashed line in Fig. 8a, lies within the upper level or conduction band on the *n*-doped side and the lower level or valence band on the *p*-doped side. When this occurs, the doping is said to be degenerate.

In thermal equilibrium, electrons and holes cannot recombine with each other due to the potential barrier that exists between the *p–n* junction. As a forward bias voltage is applied across the junction, the potential barrier is lowered. If the forward bias voltage is increased to a level which is nearly equal to the band gap energy as in Fig. 8b, i.e.,

$$V_{fb} = \frac{E_{bg}}{e}, \tag{11}$$

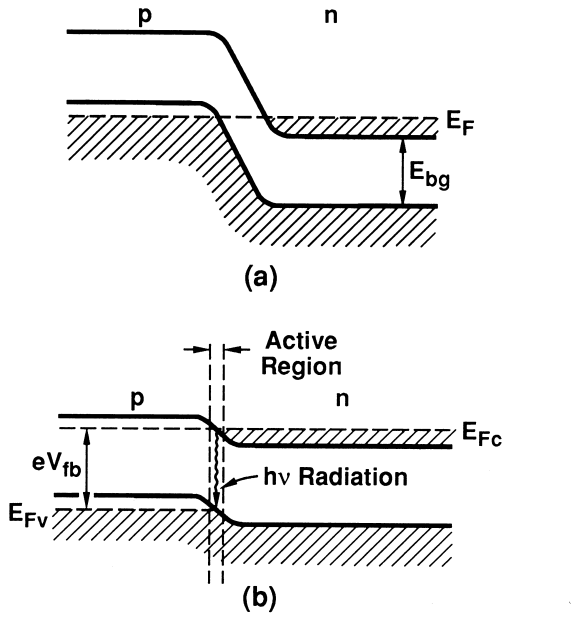


FIGURE 8 (a) Degenerate *p–n* junction at zero bias. (b) Junction with a forward bias voltage comparable to the band gap energy. [After Yariv (1975).]

then both electrons and holes are injected into the active region of the *p–n* junction. In the above equation, V_{fb} is the forward bias voltage, E_{bg} is the band gap energy of the host semiconductor in units of electron volts, and e is the electron charge. This biasing condition produces a population inversion in the active region, which is a necessary condition for the lasing process. The electrons and holes are now allowed to recombine with each other directly, emitting photons that experience a gain that satisfies the relation

$$E_{bg} < h\nu < E_{FC} - E_{FV}. \tag{12}$$

In the above equation, E_{FC} and E_{FV} are the quasi-Fermi levels of the conduction and valence bands, respectively, h is Planck's constant, and ν is the optical frequency of the laser emission. Once the light has been generated and amplified, the cleaved end facets of the semiconductor provide the optical feedback. The cleaved end facets act as mirrors which reflect the light back and forth inside the semiconductor initiating laser oscillation. The gain inside the semiconductor laser can approach 1000 times in a single pass. Due to this large gain factor, relatively low reflectivities from the cleaved facets are sufficient to initiate lasing. The typical reflectivity of the cleaved facets is approximately 30%, as compared to 60–99% for solid-state and dye laser mirrors.

Optical and Electronic Confinement: Transverse Confinement

The first semiconductor injection lasers developed were called homostructures because the *p–n* junction was made from one type of semiconductor material, e.g., GaAs. In these lasers, there is no mechanism to confine the injected carriers or created photons. As a result, these lasers suffered from high threshold currents and a poor-quality output beam. These problems were overcome with the development of the heterostructure laser, i.e., a semiconductor laser made from different semiconductor materials, such as GaAs and aluminum gallium arsenide, $Al_xGa_{1-x}As$. In this notation x is the fractional content of aluminum, typically 30%.

The most common type of heterostructure laser is the double heterostructure laser, illustrated in Fig. 9. This laser is made from a combination of semiconductor materials having different band gap energies for current confinement and different optical indices of refraction for optical confinement. The double heterostructure laser is made of three types of semiconductor materials; a layer of *p*-type, low-bandgap material, such as GaAs, sandwiched between a *p*-type and an *n*-type material with higher band gap, such as $Al_xGa_{1-x}As$. The aluminum content is chosen such that the AlGaAs regions will have a higher band gap energy and a lower optical index of refraction than the GaAs active region.

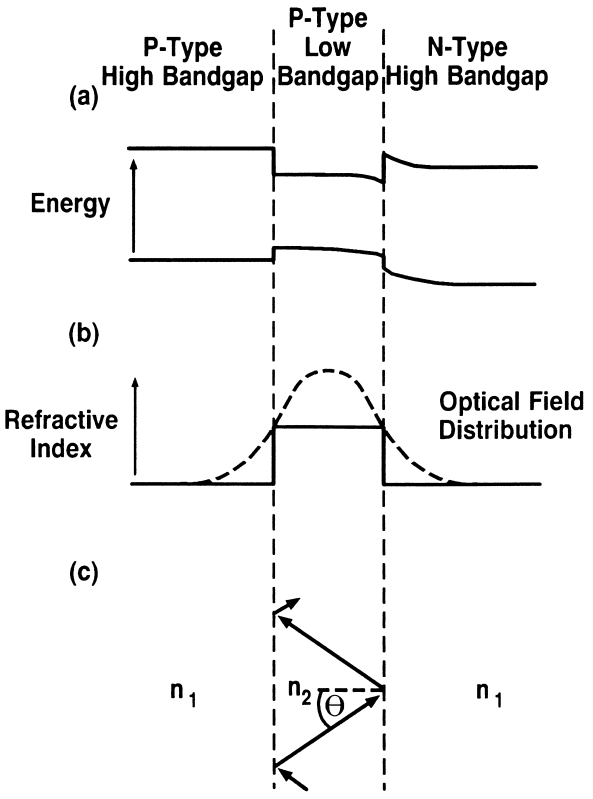


FIGURE 9 Schematic of a double heterostructure laser. (a) Band diagram under forward bias. (b) Refractive index profile and optical field distribution. (c) Waveguiding effect produced by total internal reflection and the index profile shown in (b). [After Kapon (1989).]

With the design of the band gaps of the double heterostructure laser as depicted in Fig. 9a it is possible to understand the current confinement by considering the injected carriers being trapped inside a potential well. Electrons and holes are injected into the active region from the high-bandgap material AlGaAs. The electrons and holes become trapped in the potential well created by the low-bandgap material GaAs. This potential well acts to confine the carriers in the active region of the laser. As a result, the electrons and holes cannot diffuse out of the active layer and are forced to recombine with each other in the GaAs material, contributing to the light generation.

The optical beam confinement is provided by the waveguiding properties of the AlGaAs/GaAs/AlGaAs material structure, depicted in Figs. 9b and 9c. The AlGaAs layers have a lower optical index of refraction than the GaAs (Fig. 9b). This type of spatial index profile leads to the confinement of an optical beam inside the higher index material because of the total internal reflection experienced by an optical beam inside the GaAs material (Fig. 9c). The minimum angle of incidence which an optical beam can have and still undergo total internal reflection is given by

$$\sin(\theta) = \frac{n_1}{n_2}, \tag{13}$$

where θ is the critical angle of incidence and n_1 and n_2 are the indices of refraction of the AlGaAs and GaAs materials, respectively.

E. Growth Techniques

The process of growing differing semiconductor materials such that the lattice structure of each material is matched to the other is referred to as epitaxial growth. The differing materials comprise structures called heterostructures and are epitaxially grown using techniques such as liquid phase epitaxy (LPE), molecular beam epitaxy (MBE), and vapor phase epitaxy (VPE), of which a common variant is metal organic chemical vapor deposition (MOCVD). LPE uses the cooling of a saturated solution containing the constituents in contact with the substrate. MBE makes use of molecular beams of the constituent elements that are caused to impinge on an appropriately prepared crystalline substrate held at high temperature in an ultrahigh vacuum. In the VPE technique, also referred to as chemical vapor deposition (CVD), depending on the constituents of the reactants, the epitaxial layer is grown by the reaction of gaseous elements or compounds at the surface of a heated substrate. MOCVD is similar in its growth process with the deposited elements being derived from metal alkyls. In all of these growth techniques, the composition and doping of the individual layers are determined by adjusting the arrival rates of the atoms or molecules and the temperature of the substrate surface, and can be made as thin as a single atomic monolayer.

1. LPE

The liquid phase epitaxy technique was first developed in 1963. It has been successfully used to fabricate various types of III–IV compound semiconductor devices including injection lasers, light-emitting diodes (LEDs), photodetectors, solar cells, bipolar transistors, and field effect transistors (FETs). In LPE, a supersaturated solution of the material to be grown is brought into contact with the substrate for a desired period of time. If the substrate is single crystalline and the material has nearly the same lattice constant as the substrate, some of the material precipitates on the substrate while maintaining the crystalline quality. The precipitated material forms a lattice matched epitaxial layer on the surface of the substrate.

Three basic types of growth apparatus have been used for LPE: (1) The tipping furnace, in which the substrate is brought into contact with the solution by tipping the furnace, (2) the vertical furnace, in which the substrate is dipped into the solution, and (3) the multibin furnace, in which the substrate can be brought into contact with different solutions kept in successive bins. The multibin furnace type of growth apparatus is extensively used for the fabrication of laser structures that require the successive growth of several epitaxial layers, and is generally used for growing double heterostructure lasers (see Fig. 10). The graphite boat has a number of reservoirs, each of which contains a saturated solution corresponding to the epitaxial layers to be grown. The substrate is placed in a graphite slider that has a groove to hold the substrate. The slider is attached to a long rod that allows the grower outside the furnace to position the substrate under each of the different reservoirs. As a result, several layers of different semiconductor materials, each with a different thickness if desired, can be successfully grown on the substrate.

2. VPE and MOCVD

In vapor phase epitaxy, the source chemicals from which the epitaxial layers are grown are gaseous. The technique has been widely used for the growth of several III–V compound semiconductors. VPE is often classified into two different methods, the chloride and the hydride methods. As an example, in the chloride method, $AsCl_3$ or PCl_3 is passed over elemental Ga or In, or pieces of GaAs or InP, to form metal chlorides. These metal chlorides then react with AsH_3 or PH_3 near the substrate to form epitaxial layers. In the hydride method, passing HCl gas over hot In or Ga metal forms metal chlorides. VPE of InGaAsP material

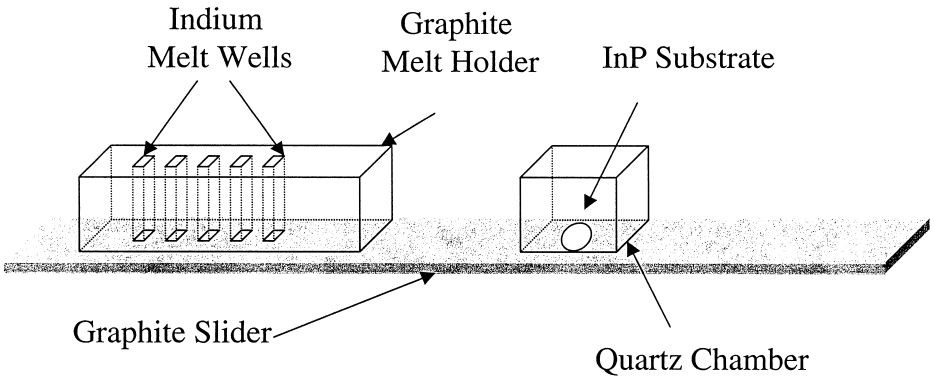


FIGURE 10 Schematic diagram of a liquid phase epitaxial growth system.

for laser devices has often used the hydride method. The chloride method is more useful for fabrication of field effect transistors because it produces layers with lower background doping levels.

Figure 11 is a schematic diagram of a vapor phase epitaxial growth system. The deposition of epitaxial layers, say, InGaAsP, is initiated by passing HCl gas over hot In or Ga metal held at ∼850–900°C, resulting in the formation of metal chlorides. Arsine and phosphine are mixed with the metal chlorides in the mixing zone. The InP substrate is held in a chamber at room temperature that is flushed with H_2. Prior to growth, the substrate is heated to a temperature of ∼700°C in an atmosphere of AsH_3 and PH_3 to prevent surface decomposition. The substrate is then inserted into the growth chamber for the desired period of time. The growth rate generally lies in the range of 0.1–1.0 μm/min. Accurate mass-flow controllers control all reactant flow rates. *P*-type doping with zinc can be accomplished by passing H_2 gas over hot zinc metal, which carries zinc vapor to the mixing chamber. Similarly, *n*-type doping can be accomplished using 100-ppm H_2S along with AsH_3 and PH_3 in the mixing chamber.

Altering the flow rates of the reactants in a VPE reactor allows the composition of the layers to be changed during growth. It takes a certain amount of time for gaseous equilibrium to be established each time that the flow rates are changed in the reactor. Owing to this delay, it is difficult to grow successive epitaxial layers with different compositions and abrupt interfaces in a single-chamber reactor. As a result, the single-chamber reactor has evolved to include multichamber reactors. In the multichamber reactor, each chamber, which is specified for growing layers

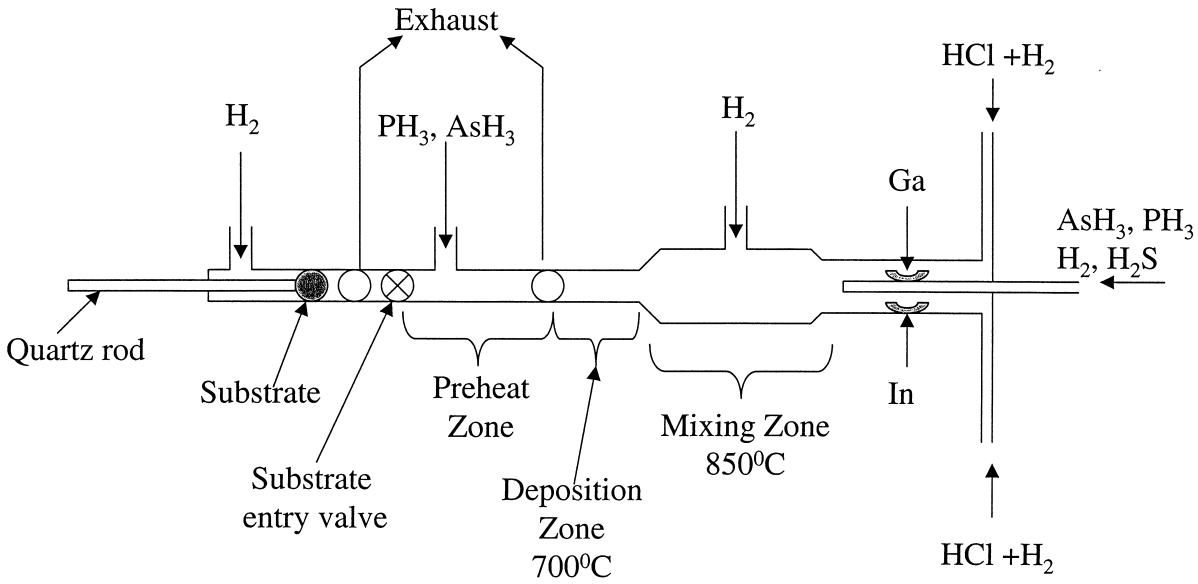


FIGURE 11 Schematic diagram of a vapor phase epitaxial growth system.

of a well-defined composition, is initially stabilized with the appropriate gas flow rate. The substrate is then introduced into each growth chamber separately, with the transfer time of the substrate between chambers on the order of 1 sec.

3. MBE

In the molecular beam epitaxy technique, layers are grown by directing an atomic or molecular beam onto a heated substrate located in an ultrahigh-vacuum environment. The particles in the molecular beam adhere to the substrate, resulting in a lattice-matched layer. The strength or flow rate of the atomic/molecular beams can be adjusted to compensate for differences in the adhesion properties of the constituents making up the epitaxial layers. Figure 12 shows a schematic of an MBE system for the growth of heterostructure devices. The substrate is heated by a molybdenum heating block inside a vacuum chamber held at a pressure of 10^{-7}–10^{-10} torr. Diagnostics, such as electron diffraction, mass spectrometer, and Auger analysis, can be incorporated onto the MBE chamber to characterize the epitaxial layers during growth. The atomic and molecular sources (e.g., Ga, GaAs, In, As_2, P_2, etc.) as well as the *n*- and *p*-type dopants are contained in independently heated effusion ovens, surrounded by liquid nitrogen shrouds. The effusion ovens are heated to a sufficiently high temperature to generate an adequate beam flux on the substrate surface. The flux rate falls in the range of 10^{12}–10^{14} atoms/(cm-sec) for the growth of heterostructures, for example, AlGaAs on GaAs. The required flux rate is dependent upon the adhesion coefficient, i.e., the fraction of atoms in the beam adhering to the substrate surface. For many of the group III elements, specifically Al or Ga, the adhesion coefficient is nearly unity. On the other hand, the adhesion coefficient differs considerably for group V elements. As a result, the adhesion coefficient must be determined empirically for the range of operating substrate temperatures and beam intensities.

III. LASER STRUCTURES

The optical and electrical confinement just discussed provide confinement of the injected carriers and the optical beam in a direction that is parallel to the growth direction of the layered structure. Optical and electrical confinement is also needed in the lateral direction, i.e., in a direction

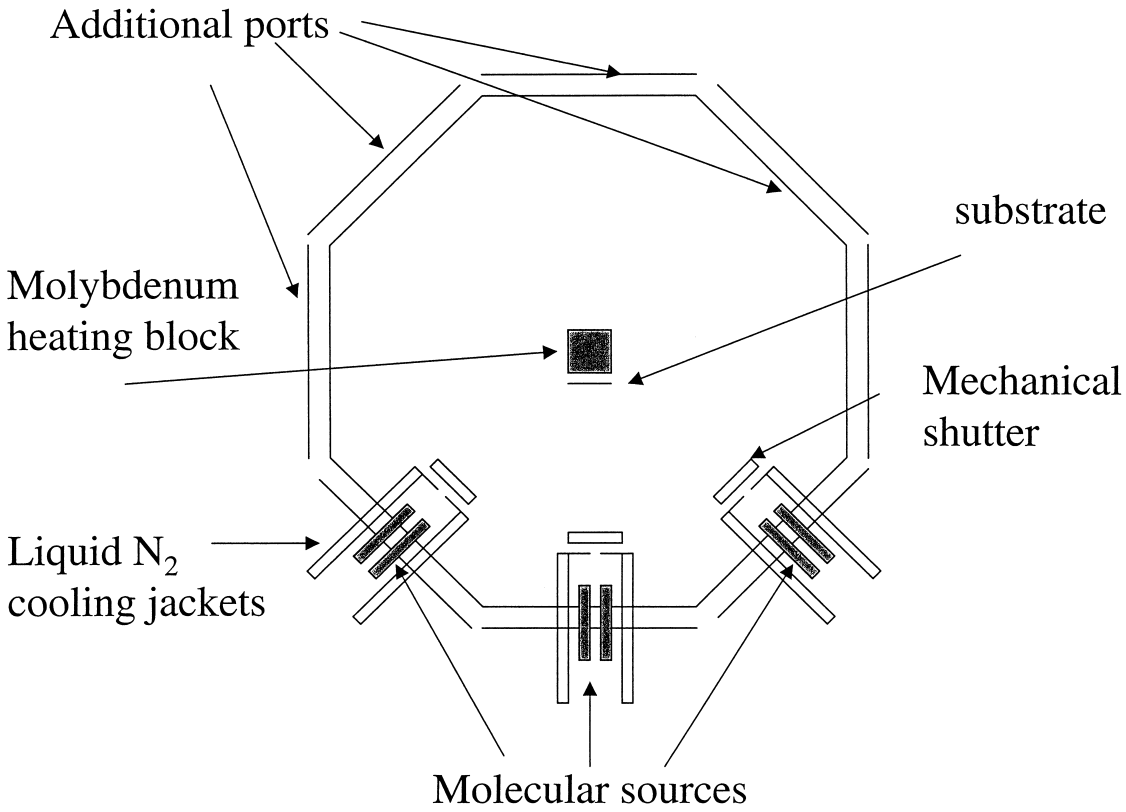


FIGURE 12 Schematic diagram of a molecular beam epitaxial growth system.

parallel to the layers. This can be provided by two means: (i) gain guiding and (ii) index guiding. The physics describing the operation of these two general classes of semiconductor injection lasers can best be understood by considering some specific laser structures and is described in the next subsection.

A. Gain-Guided and Index-Guided Laser Structures: Lateral Confinement

Gain guiding is one method of providing electrical and optical confinement of the injected carriers and the generated photons in a direction perpendicular to the direction of growth of the layered structure. It is most easily accomplished by fabricating a narrow electrical opening in an otherwise normally insulating region, forming a stripe on the top of the laser structure. The carriers are injected and confined into the active layer directly under the contact stripe. This provides optical gain only under the stripe contact, thus giving the name gain-guided laser. Specific laser structures that utilize the gain guiding principle include the oxide stripe laser and the proton-implanted laser. These lasers are depicted schematically in Fig. 13.

The oxide stripe laser in Fig. 13a has a narrow opening on the order of several microns, which is created in an electrically isolating oxide layer. A metalization layer is then deposited on top of the oxide layer to create the stripe contact. The layers located above the active region are made sufficiently thin so that the injected carriers are confined in a narrow stripe in the active region.

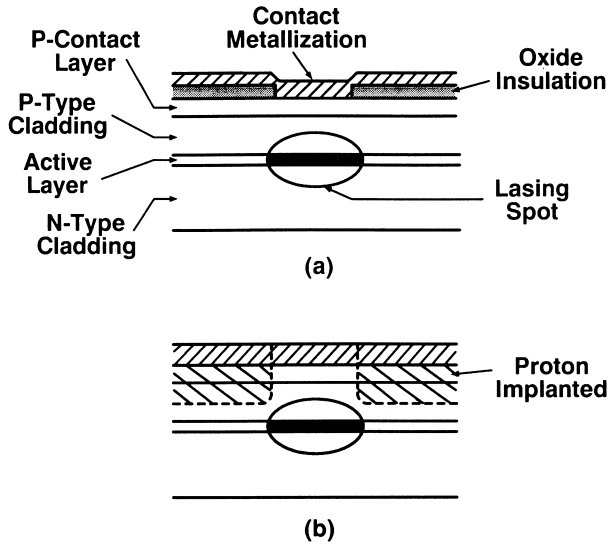


FIGURE 13 Schematic cross sections of two types of gain-guided laser structures. (a) The oxide-stripe laser. (b) The proton-implanted laser. [After Kapon (1989).]

The proton-implanted laser in Fig. 13b has a contact stripe which is created by implanting protons into the metalization layer, leaving only a small stripe which has not been implanted. The implanted regions become highly resistive to current injection, while the unimplanted region has a low resistivity, which serves as the electrical stripe contact. Using this method, it is possible to control the amount of current confinement by varying the depth of proton implantation.

In gain-guided structures, the unpumped regions on both sides of the stripe are lossy at the lasing wavelength. This leads to a quasi-Gaussian lateral gain profile. The peak of the gain profile is located directly under the stripe, while the wings of the gain profile extend into the lossy regions. This leads to guidance of the optical field in a direction along the laser stripe. The features of the gain-guided optical field are not only determined by the gain distribution. Additional guiding is naturally provided by the index change due to local lateral temperature gradients and carrier density gradients in the active layer.

Alternative methods of providing lateral electrical and optical beam confinement can be seen by examining index-guided structures. Index-guided structures provide optical lateral confinement by fabricating a refractive index distribution which is parallel to the laser's active layer. This, in combination with the double heterostructure configuration, gives a two-dimensional confinement of the optical field. The electrical confinement can be provided by a small opening in an electrically isolating layer as in the gain-guided structures or by employing reversed-biased p–n junctions which sandwich the lasing region in the lateral direction. Index-guided structures include the ridge waveguide laser and the buried heterostructure (BH) laser. These laser structures are depicted schematically in Fig. 14.

In the ridge waveguide structure (Fig. 14a) the top of the laser device is etched down to be very close to the active layer. Only a small part of the diode is not etched. This produces a plateau above the active layer which ultimately becomes the lasing region. The evanescent field of the propagation wave extends into the plateau region and is efficiently guided in this process. In this structure, the electrical confinement is provided by the opening in the insulating oxide layer.

The buried heterostructure laser in Fig. 14b is designed so that the active region is completely surrounded by material which has a lower optical index and also a higher energy band gap. As a result of this two-dimensional index profile, the optical beam is tightly confined in both transverse and lateral directions to the direction of propagation.

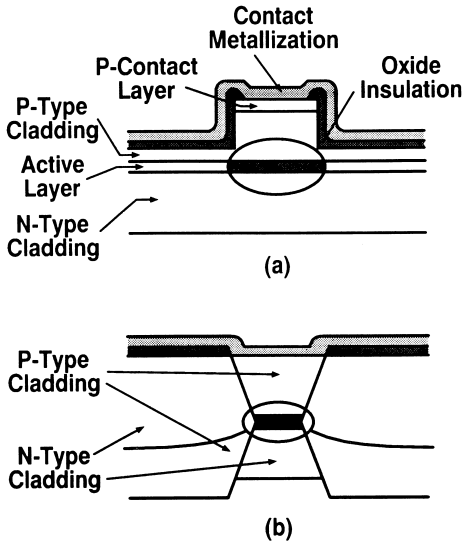


FIGURE 14 Schematic cross sections of two types of index-guided lasers. (a) The ridge waveguide laser. (b) The buried heterostructure laser. [After Kapon (1989).]

This structure employs two reversed-biased *p–n* junctions so that the current also becomes laterally confined to the lasing region. With this type of optical and current confinement, threshold currents of less than 1 mA have been achieved.

As we have seen, there are several methods for confining both the injected electrical current and the generated optical beam. The gain-guided structures are easier to fabricate than the index-guided structures; however, gain-guided structures have higher threshold currents and an inferior output beam than index-guided lasers. As a general rule, by properly designing the optical and electronic properties of the semiconductor laser, the injected carriers and generated photons can be confined into a well-defined region, yielding the maximum interaction, which leads to lower threshold currents and a high-quality output beam.

B. High-Power Semiconductor Lasers

Typical optical power output levels in standard single-stripe diode lasers are on the order of several milliwatts. These power levels are adequate for many commercial applications that use diode lasers, such as fiber communications, optical discs, etc. However, for many other applications, such as optical pumping, optical time-domain reflectometry, laser radar, and nonlinear optics, these power levels are not sufficient.

The most obvious way of increasing the output power of a semiconductor laser is to increase the driving current of the laser. This method works up to the point where the output power damages the facets of the laser or until the heat generated in the active region starts to degrade the performance of the laser diode.

Other methods of increasing the output power of the laser diode rely on increasing the volume of the active region. This is typically done by increasing either the width or the thickness of the active area. Increasing the length of the laser only reduces the threshold current. The output power in this case is still limited by the catastrophic facet damage.

By increasing the active-region stripe width, one produces what is termed a broad-area laser. These devices have very wide active areas typically ranging from 10 to over 250 μm. These lasers are capable of providing large output powers on the order of a few watts with several amperes of driving current. These lasers can be used as an optical pump in solid-state laser systems or in other applications that require a compact, high-output-power laser. The problems associated with these lasers are that they usually require large biasing currents. This leads to potentially damaging thermal effects in the active region, ultimately leading to device failure. Another problem associated with broad-area lasers is that they tend to operate in a multispatial mode pattern. The poor beam quality that is produced makes it difficult to focus the beam to a small spot, making the device unsuitable for fiber optic applications. Another serious limitation to broad-area lasers is due to beam filamentation. The filamentation is caused by a nonlinear optical effect called self-focusing. Self-focusing is caused by an increase of the optical index of refraction which is proportional to the optical intensity of a laser beam. An optical beam that has spatial nonuniformities in the intensity of the transverse beam profile causes a spatial distribution of varying refractive index. This causes the beam to collapse or self-focus into a filament, which causes damage to the laser and ultimately into device failure.

To overcome these difficulties, new device structures were developed. The problem of high driving currents was addressed by the development of quantum well lasers. These lasers rely on properties of the electron which occur when the electron is confined to a small region in space. The active-area thickness of typical quantum well lasers is approximately 100 Å. Broad-area lasers which utilize a quantum well as the active area can have threshold currents as low as 80 mA with a stripe width of 100 μm, and have output powers as high as 1 W with 1 A of driving current. Most recently, using device structures that employ a broadened or graded index waveguide region, output powers of 5 W have been achieved in a single transverse output mode, employing a 100-μm emitter.

The problem of beam filamentation has been addressed by the development of laser arrays. A typical laser array

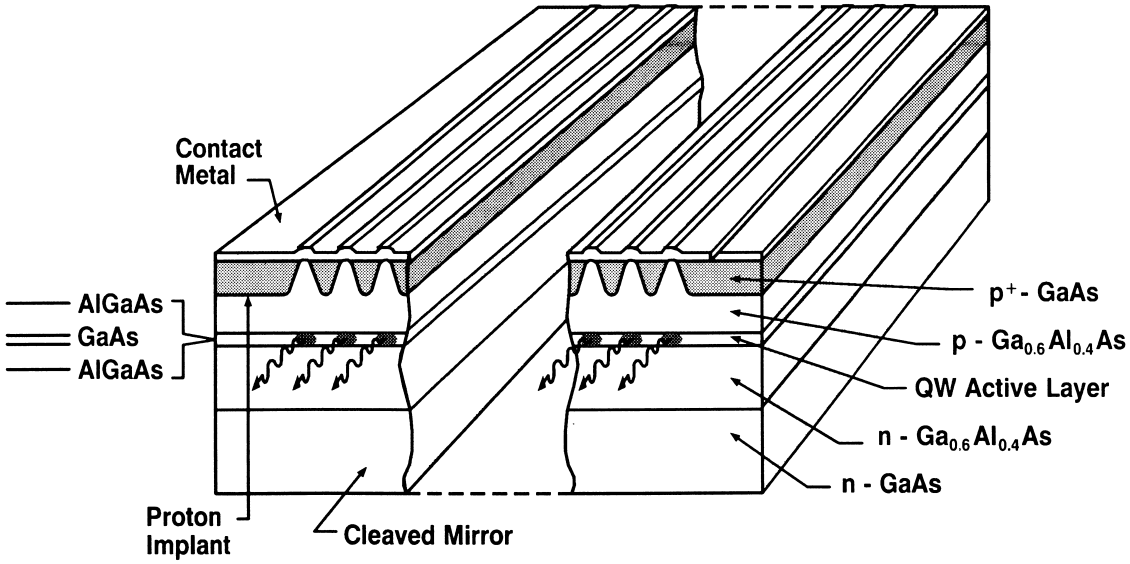


FIGURE 15 Schematic cross section of a multistripe laser array. [After Streifer *et al.* (1988).]

structure is shown in Fig. 15. Several laser stripes are formed by proton implantation of an initially wide stripe. This is shown by the shaded areas of the p^+ GaAs region. The disordered regions become highly resistive to current flow, while the unimplanted regions become the lasing stripes. The stripes which define the width of the active region are kept small, typically a few micrometers. This forces each individual laser to operate in the single fundamental transverse mode, which eliminates the problem of the beam filamentation. The individual stripes are positioned close enough to the neighboring stripes to allow the evanescent fields of the individual stripes to overlap and interact. This interaction allows coupling of the spatial modes of the individual stripes, forcing the laser array structure to operate coherently. Utilizing a structure with 10 single emitters with 10-μm-wide stripes in an external cavity configuration, output powers in excess of 700 mW have been produced in a single-lobed beam.

These one-dimensional array structures can be stacked on top of each other to produce a two-dimensional laser array. However, at the present stage of development, it is difficult to achieve coupling of all of the individual stripes without the use of external optics. The highest power produced from this type of two-dimensional laser array structure has exceeded 100 W. Although the emitted energy is not totally coherent due to the inability to couple all of the individual emitters, the output power is sufficient to allow the device to be used as an optical pump to excite a solid-state laser material such as Nd:YAG. This type of application is very promising for semiconductor laser array structures and is discussed in more detail in the section on diode-pumped solid-state lasers.

C. High-Modulation-Frequency Lasers

The direct modulation property of semiconductor injection lasers is one of their unique characteristics and this feature provides the potential for many scientific and commercial applications. Unlike other solid state and liquid lasers, modulating the injection current can directly control the optical output of the semiconductor laser.

The modulation dynamics of semiconductor injection lasers can be described by the following rate equations for the laser, which are derived from the Maxwell–Bloch equations after adiabatically eliminating the atomic polarization:

$$\begin{aligned} \frac{dN}{dt} &= \frac{I}{eV} - \frac{N}{\tau_S} - g(1-\varepsilon S)(N-N_a)S \\ \frac{dS}{dt} &= g(1-\varepsilon S)(N-N_a)S - \frac{S}{\tau_P} + \beta\frac{N}{\tau_S}, \end{aligned} \tag{14}$$

where N is the electron density in the active region, S is the photon density, V is the active region volume, τ_S and τ_P are the electron lifetime and photon lifetime, respectively, N_a is the carrier density at transparency, β is the spontaneous emission coupling coefficient, g is the differential gain coefficient, ε is the gain saturation coefficient, and e is the electron charge.

By employing a small-signal analysis of the laser rate equations above the lasing threshold, one can derive the resonance frequency f_r and the damping constant α_d, for

the modulation response of the photon density to the injection current:

$$f_r = \frac{1}{2\pi}\sqrt{\frac{gS_0}{\tau_P}} \tag{15}$$

for the resonance frequency, and

$$\alpha_d = \frac{1}{\tau_S}(1+gS_0\tau_S) + \frac{\tau_S}{\tau_P}\left[\beta\frac{1+g\tau_P N_a}{g\tau_S S_0}\frac{1}{\tau_S} + \varepsilon g S_0\right] \tag{16}$$

for the damping constant, where S_0 is the steady-state photon density.

Figure 16 shows typical modulation response curves experimentally measured for several values of output power. The circuit diagram shows how the measurement is performed. The modulating signal $i_1(\omega)$ is applied to the laser diode and varied from a low modulation frequency of ~200 MHz to a high modulation frequency of ~10 GHz. In this notation, ω denotes the applied frequency. The current $i_2(\omega)$ is measured from a wide-bandwidth photodiode, and the ratio of the output current to the input current is plotted on a logarithmic scale. These curve shows three salient features characteristic of the modulation response of semiconductor laser diodes. For quasistatic modulation, the modulation frequency applied to the laser is much less than the laser resonance frequency. In this region, the optical output of the laser follows the modulating current, and the modulation efficiency is equal to the slope efficiency of the steady state light output versus dc current input curve of the laser diode. At the high-frequency limit, the coupled electron–photon system cannot follow the modulation current, and the modulation efficiency decreases rapidly. The interesting feature, the resonant enhancement of the modulation response, is observed when the modulating current is at or near the resonance frequency of the laser. This resonance peak in the modulation response in indicative of a damped oscillation of the photon density after being perturbed from the steady state.

The peak in the modulation response can be described by a strong two-way interaction between the populations of injected carriers and photons. The stored energy of the system can swing between the two populations with a natural resonance frequency which depends on particular circumstances but is normally in the vicinity of a few gigahertz. Little damping is supplied by the optical resonator since under lasing conditions its Q is very large and the main contribution to damping comes from the spontaneous recombination time of the carriers, giving a decay time of the order of 5 nsec. The resonance manifests itself as a transient oscillation during laser switching and also as an enhancement of the modulation response to a small sinusoidal current in the relevant frequency range.

The modulation bandwidth of the semiconductor injection laser given by Eq. (15), is determined by the gain coefficient, photon lifetime, and the steady-state photon density. These parameters are influenced by the structure of the laser, the operating temperature, and the laser diode material system. For a given laser, increasing the injection current, as shown in Fig. 16, can increase the bandwidth. However the maximum current is limited by the

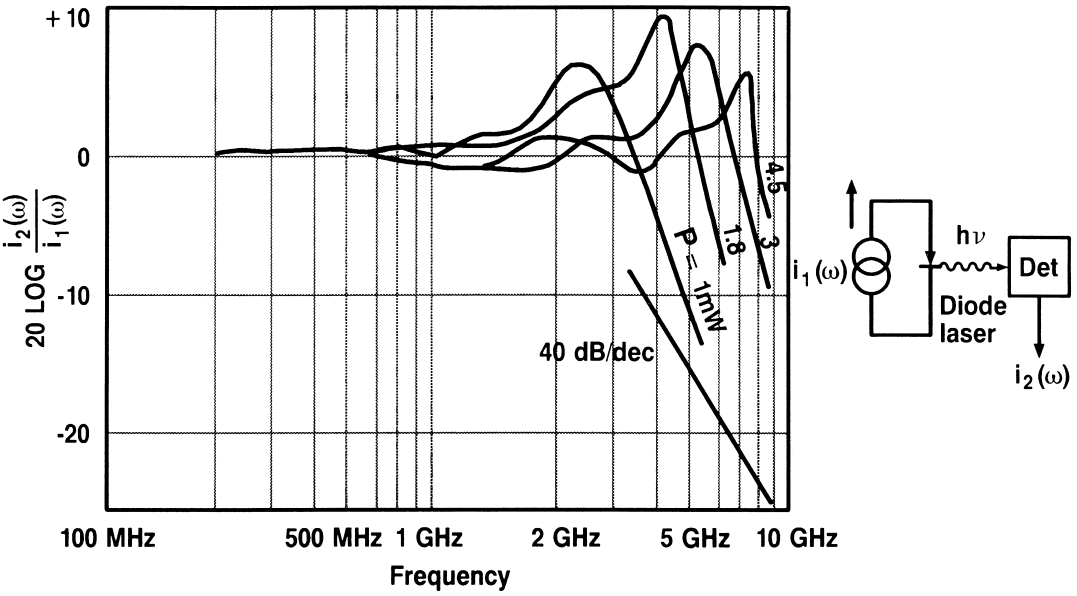


FIGURE 16 Experimentally measured continuous wave modulated output power versus current characteristic of a laser of length 120 μm. The inset shows the method of measurement. [After Lau *et al.* (1983).]

catastrophic facet damage for AlGaAs lasers and by nonlinear gain, Auger recombination, and current leakage in InGaAsP lasers. By fabricating a window structure near the facets in AlGaAs lasers, the damage threshold can be increased due to the decrease in the optical absorption occurring at the facets. These lasers have been demonstrated to have modulation bandwidths in excess of 11 GHz at room temperature.

Another method of increasing the modulation bandwidth is to decrease the photon lifetime. This is most easily accomplished by decreasing the laser diode cavity length. This, however, increases the laser threshold current level, and as a result, lasers with extremely low threshold current levels are required for this method. Utilizing a 40-μm-long AlGaAs multiple quantum well laser, a modulation bandwidth of 50 GHz has been achieved.

In addition to excessive current pumping and shortening the laser diode cavity length as a means to increase the modulation bandwidth, the differential gain coefficient can be increased. This can be achieved by lowering the operation temperature of the laser diode. For example, the bandwidth of a GaAs buried heterostructure laser which is 7 GHz at room temperature (25°C) can be extended to 9.5 GHz at −50°C for a 175-μm-long cavity. Similarly, the bandwidth of a InP constricted mesa laser can be increased from 16 GHz at 20°C to 26.5 GHz at −60°C. The drawback with this method is that there is a decrease in the modulation efficiency due to an increase in the series resistance of the diode which occurs at low temperatures.

Higher differential gain coefficients can also be obtained by utilizing semiconductor quantum well structures. Due to the confinement of carriers in a direction of the quantum well growth, the density of states become stepwise and increases the differential gain coefficient. This staircase density of states distribution increases the modulation bandwidth by a factor of two. It is expected that another factor of two increase in the modulation bandwidth will be obtainable with quantum wire structures, which have a quasi-discrete density of states distribution.

Other phenomena that influence the modulation dynamics of semiconductor lasers include the spontaneous emission, gain saturation, and the external electrical connections to the device. These effects normally manifest themselves as a suppression of the relaxation oscillations, which reduces the peak of the modulation response and also reduces the modulation bandwidth.

D. Narrow-Linewidth Lasers

The spectral linewidth of standard Fabry–Perot semiconductor lasers is typically on the order of 100 MHz. For many applications, such as holography, spectroscopy, and coherent optical communication systems, these broad linewidths are unsatisfactory. The main difficulty of obtaining narrower linewidths in semiconductor lasers is that there is a modulation of the index of refraction due to fluctuations of the excited population caused by spontaneous transitions. This has an effect of broadening the spectral linewidth of the lasing transition that is predicted by the Schawlow–Townes formula by a factor of $\left(1+\alpha_{\mathrm{L}}^2\right)$. The parameter α_{L} is called the linewidth enhancement factor and is defined as

$$\alpha_{\mathrm{L}} = \frac{\chi_{\mathrm{r}}^3}{\chi_{\mathrm{i}}^3} = \frac{\partial \chi_{\mathrm{r}}/\partial N}{\partial \chi_{\mathrm{i}}/\partial N}, \tag{17}$$

where $\chi_{\mathrm{r/i}}^3$ is the real/imaginary part of the third- order nonlinear susceptibility, χ is the linear plus nonlinear components of the susceptibility, and N is the population density. In most gas and solid state lasers, the lasing gain spectrum is symmetric due to a single atomic transition. This implies that $\alpha_{\mathrm{L}} = 0$ for these laser systems, and the linewidth is correctly predicted by the Schawlow–Townes equation. In semiconductor injection lasers, the gain spectrum is not symmetric due to the density of states and the Fermi–Dirac statistics of electrons and holes. This leads to a nonzero value for α_{L} in semiconductor lasers and thus a broader spectral linewidth of the lasing transition. The typical value of α_{L} is $\sim$5 for bulk semiconductor injection lasers and $\sim$2.5 for quantum well semiconductor injection lasers.

Because of this difficulty in obtaining inherently narrow spectral linewidths, there has been great interest in developing new laser structures and techniques that can yield linewidths less than a few megahertz. The most promising method for reducing the linewidth of a semiconductor laser incorporates a grating structure inside the semiconductor chip itself or utilizes a grating in an external cavity configuration.

The two most popular laser structures which utilize gratings as an integral part of the laser structure are (1) the distributed feedback (DFB) laser structure and (2) the distributed Bragg reflector (DBR) laser structure. In these laser structures, a region of the semiconductor waveguide is fabricated to include a corrugated region. This region acts as a diffraction grating which couples a narrow spectral portion of the total spontaneous emission back into the laser diode. The corrugation period and the laser emission wavelength must satisfy the relation

$$m\Lambda \cong \frac{\lambda}{2n}, \tag{18}$$

where Λ is the grating period, λ is the laser wavelength, n is the index of refraction, and m is an integer.

In the DFB laser, the grating structure is distributed along the entire length of the laser diode, just slightly

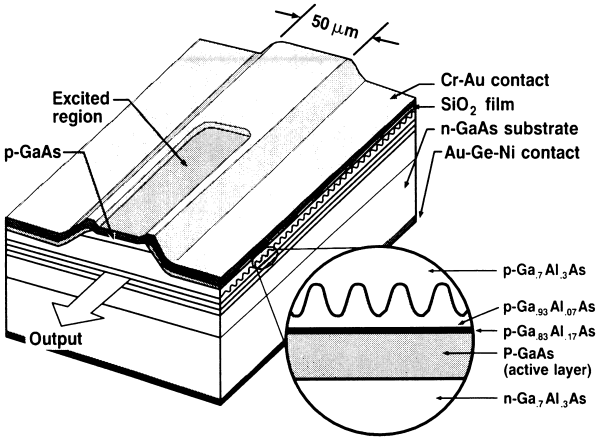


FIGURE 17 Schematic of a GaAs–GaAlAs distributed feedback laser. The inset gives a detailed illustration of the laser layers and the corrugated structure. [After Yariv (1988).]

above the gain region. This is shown schematically in Fig. 17. Due to the frequency-selective feedback provided by the grating, lasing occurs at a single longitudinal frequency defined by the laser cavity. The power contained in the longitudinal modes that have frequencies that do not correspond to the Bragg condition (18) is lower by 1000 times than that in the lasing mode. Another attractive feature of the DFB laser is that the cavity is defined by the grating structure, not by cleaved facets. This allows the DFB laser to be incorporated directly into an integrated optoelectronic device.

The distributed Bragg reflector laser is similar to the DFB laser. Both lasers incorporate grating structures directly into the waveguiding path of the laser mode. The difference between these two structures in that the DBR laser utilizes a passive waveguiding region to contain the corrugated structure. The main advantage of this type of structure is that it does not require epitaxial growth over the grating structure and hence is easier to fabricate. The disadvantage is that the coupling efficiency between the grating structure and the active region is lower due to the waveguide discontinuity between the active and passive regions. Both of these laser structures provide lasing emission linewidths on the order of 1 MHz, which is ~100 times narrower than the conventional Fabry–Perot laser structure.

Narrower linewidths can be obtained by utilizing an external cavity. Typical cavities include an objective lens for collecting and collimating the laser emission and a reflecting diffraction grating or étalon for frequency selection. When the optical feedback from the external cavity is strong, an antireflection coating is required on the laser facet that is inside the external cavity and faces the frequency-selective element. The antireflection coating is necessary because it eliminates the longitudinal modes of the Fabry–Perot structures. If the antireflection coating in not employed, there will be competition between the longitudinal modes defined by the external cavity and the cleaved facets. With external cavity semiconductor lasers employing a diffraction grating, linewidths as narrow as 1 kHz have been achieved. The disadvantage of external cavity lasers relates to the relatively large dimensions, typically on the order of several centimeters.

E. New Devices

1. Quantum Well Lasers

Quantum well semiconductor lasers with both single and multiple active layers have been fabricated. Quantum well lasers with one active are called single-quantum-well (SQW) lasers and lasers with multiple quantum well active regions are called multiquantum-well (MQW) lasers. The layers separating the active layers in a multiquantum well structure are called barrier layers. Typical examples of the energy band diagram of both SQW and MQW are schematically represented in Fig. 18.

A key feature associated with quantum well laser structures as compared to their bulk counterparts is that the laser threshold current is greatly reduced by the effect of the modified density of states. If the density of states of the bulk and quantum well structures is given as above, one can express the optical gain coefficient as

$$\Gamma(\nu_0) = \frac{\lambda^2}{8\pi\tau_r}\rho(\nu)\, f_g(\nu) \tag{19}$$

where f_g is the Fermi factor associated with optical gain, i.e., $f_c(E_2) - f_v(E_1)$, $\rho(\nu)$ is the density of states, λ is the wavelength in the medium with index of refraction n, τ_r is the radiative recombination time, and E_2 and E_1 are the quasi-Fermi levels of the conduction and valence bands, respectively. The gain coefficient for both bulk and quantum well lasers is shown in Fig. 19. The salient feature in this figure is that the quantum well structure has a smaller peak gain and a narrower, smaller spectral width of the gain spectrum than the bulk material.

By increasing the pumping current density, the concentration of excess electrons and holes is increased and therefore so is the separation between the quasi-Fermi levels. The effect of this increase on the gain coefficient is borne out by considering Fig. 19. For sufficiently small current there is no gain, i.e., the difference in quasi-Fermi levels is not larger than the energy gap of the quantum well E_{GQW}. When the injection current is sufficient such that the difference in energy between the quasi-Fermi levels $E_2 - E_1$ exceeds E_{GQW}, the quantum well material provides optical gain. The gain increases sharply to the level of Γ_{MAX} and

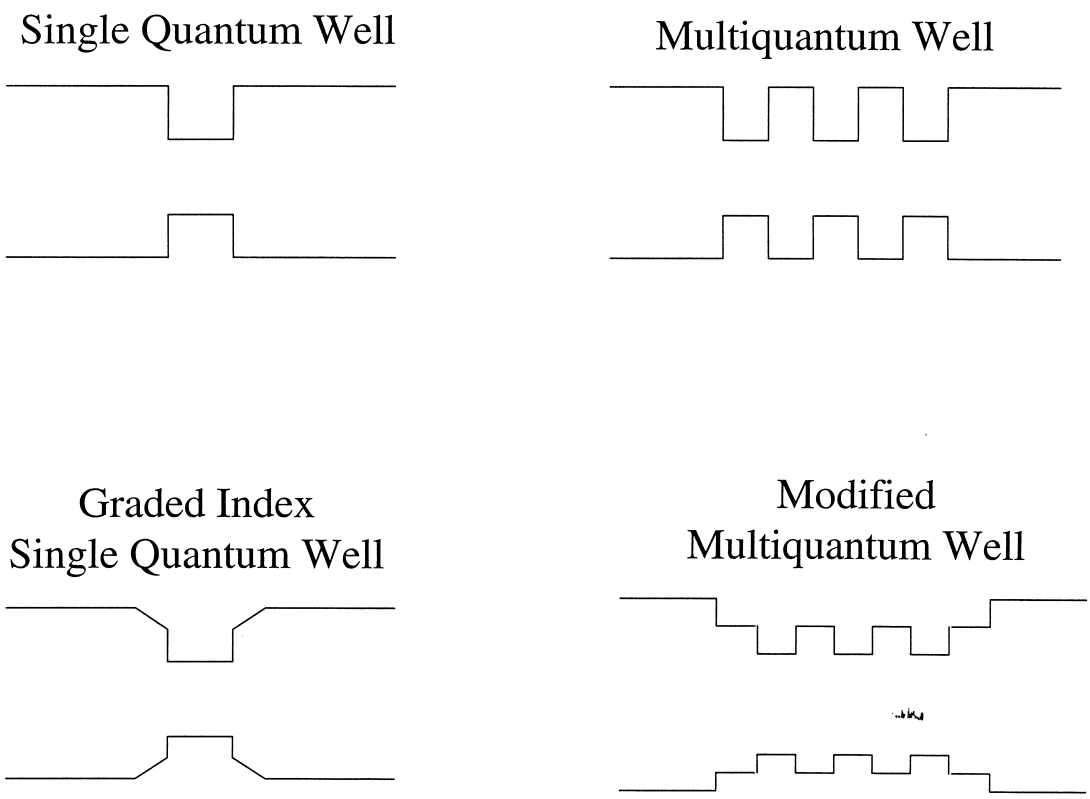


FIGURE 18 Energy band diagram of four types of semiconductor quantum well structures.

saturates. By continuing to increase the pumping current, the spectral width over which optical gain can be realized is increased; however, its peak value is kept constant. A continued increase in current can allow the quasi-Fermi level difference to exceed the second quantum well level ($q = 2$), allowing another steplike increase in the gain coefficient. An interesting result of this steplike increase in gain coefficient is the accompanying increase in spectral width, i.e., the gain profile can potentially be quite broad, allowing the possibility of a wide gain bandwidth or wide tuning range for these structures.

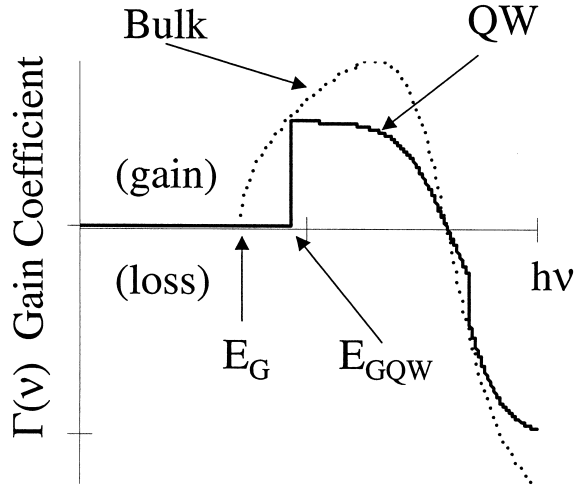


FIGURE 19 Plot of the optical gain coefficient in inverted semiconductor media, showing the difference between bulk and quantum well material.

Owing to the additional confinement of carriers to within the quantum well structure, the resultant threshold current density becomes considerably less than for bulk double heterostructure devices. Additionally, quantum well lasers generally have a narrower gain spectrum (for similar bias currents to DH structures), a smaller lasing linewidth of the lasing modes, a reduced temperature dependence, and the potential for achieving higher modulation frequencies.

The active-layer thickness of an SQW laser is typically less than 10 nm, which is to be compared to 0.1 μm for a DH laser. The threshold current for a typical SQW laser is typically less than 1 mA, while DH structures have threshold currents of several tens of milliamperes. The spectral width of the lasing emission from an SQW laser is usually less than 10 MHz as compared to 100 MHz for a typical DH structure. The output power from single quantum well laser is on the order of 100 mW, although

arrays, in both linear and two-dimensional stacking, can be constructed to give output powers in excess of 100 W. The overall quantum efficiency of such arrays can be on the order of 50%.

To increase the gain coefficient of an SQW structure further, a parallel stack of quantum wells can be epitaxially grown to yield multiquantum well structures. The gain of a multiquantum well laser structure with N layers is N times the gain of each of its wells. However, a more realistic comparison of the performance of SQW and MQW lasers requires that both structures be injected with the same current. Assume that a single quantum well is injected with an excess carrier density and has a peak gain coefficient Γ_{MAX}. Each of the N wells in the MQW structure would then be injected with only $1/N$ of the current. Owing to the nonlinear dependence of the gain on carrier concentration, the gain coefficient of each well is $(\Gamma_{MAX}/N)\varepsilon$, where ε is a factor on the order of 1, depending on the operation conditions. The total gain provided by the MQW laser is $N(\Gamma_{MAX}/N)\varepsilon$. In general, it is found that at low current densities, the SQW is superior and has a higher gain coefficient, while at high current densities, the MQW is superior, but by a factor of less than N.

2. Vertical Cavity Lasers

The semiconductor lasers that have been described thus far employ cleaved facets that serve as reflecting mirrors and form the optical cavity required for feedback. The facets are perpendicular to the surface of the wafer and light is emitted from the edge of the device, i.e., parallel to the surface of the wafer, and are sometimes referred to as edge emitters. For many applications, such as with optical interconnects, that require a two-dimensional laser array or monolithic integration of lasers with electronic components, it is desirable to have the laser output normal to the surface of the wafer. Such lasers are referred to as surface emitting lasers. A specific class of surface emitting lasers that have their optical cavity normal to the surface of the wafer are known as vertical cavity surface emitting lasers (VCSELs).

An importance feature of a typical VCSEL structure is that it incorporates Bragg reflecting mirrors at the top and bottom of the VCSEL structure to form the optical cavity for feedback. The active region is sandwiched between both n- and p-type cladding layers that are also sandwiched between a top p-type Bragg mirror and a bottom n-type Bragg mirror. The Bragg mirrors are made from a multilayer stack of alternating layers of semiconductor materials, where the semiconductor materials are similar to those used to make both the cladding and active regions. A schematic diagram of a typical VCSEL structure using distributed Bragg reflecting mirrors made of multiple semiconductor layers is shown in Fig. 20. Owing to the vertical emission of the laser radiation, the electrical contacts can be fabricated such that the emitted optical beam is circular. This is in contrast to conventional edge emitters, which typically emit a beam that is elliptical. This difference in the emitted optical beam mode profile allows the light from VCSEL devices to be more easily focused and coupled into an optical fiber, and has important implications in cost-effective optical data and communication links.

The key technical advance that led to the development of VCSELs was the ability to fabricate high-quality, high-reflective mirrors. These Bragg mirrors consist of alternating layers of low-index and high-index semiconductor materials. The thickness of each layer is equal to one-quarter of the wavelength of light in that material. These periodic Bragg reflecting mirrors can achieve very high reflectivities. For normal incidence the mirror reflectivity is given by

$$R = \left[\frac{1-(N_H/N_L)^{2N}}{1+(N_H/N_L)^{2N}}\right]^2, \tag{20}$$

where N_H, N_L, and N are the high index, low index, and number of layer pairs, respectively. In general, the effective reflectivity of the mirror stack increases with the number of pairs, while the index difference controls the wavelength range where the reflectivity is large, for example, a large index difference generates a spectrally broad reflector.

In a VCSEL, the effective single-pass gain is approximately 1% since the thickness of the gain layer is thin in the direction of the lasing emission. As a result, in order for VCSELs to have a threshold current comparable to that of edge emitters, the mirror reflectivity needs to be 99% to assure that the round-trip losses do not exceed the gain.

Most of the initial research and development of VCSELs has been carried out using the GaAs/AlGaAs material system, where the active region is 0.1-μm-thick bulk active layer or a quantum well active region and provides an output lasing wavelength of ~850 nm. After the initial development of short wavelength VCSELs, long-wavelength VCSELs with an emission wavelength of 1.3–1.55 μm have been actively pursued owing to their potential impact in data and communication links and networks. The main technical difference in the development of short-wavelength VCSELs and long-wavelength VCSELs is in the techniques employed for mirror fabrication. For short-wavelength devices the mirrors are easily achieved by epitaxial growth. One the other hand, long-wavelength VCSELs can realized by fusing two GaAs/AlGaAs Bragg mirrors designed for 1.54 μm on each side of a separate, thinned active layer operating near 1.54 μm. Alternately, dielectric mirrors of SiO_2–TiO_2

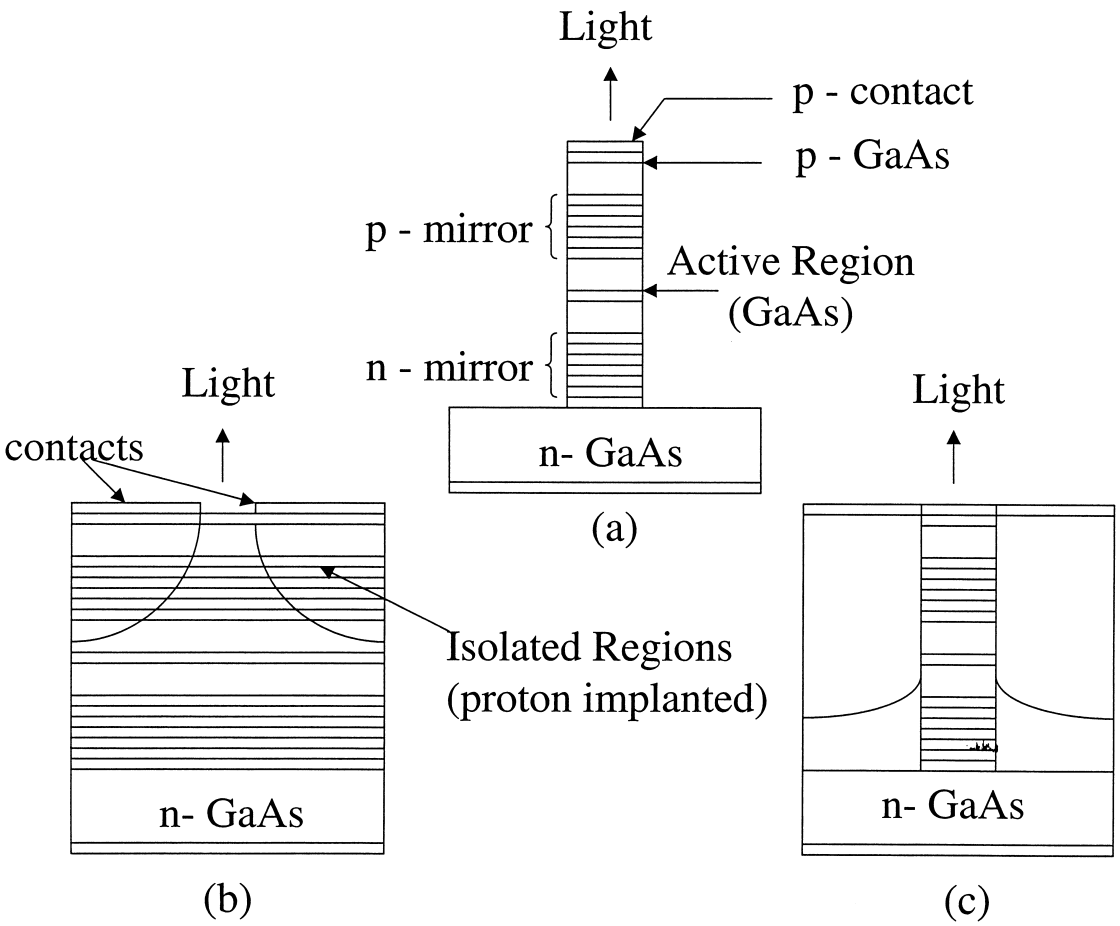


FIGURE 20 Schematic representation of three types of vertical cavity surface emitting laser structures. (a) Etched mesa, (b) ion-implanted structure, (c) buried heterostructure.

or Si–SiO_2 can also be used for creating the Bragg mirrors.

A key advantage of VCSELs is the high packing density of individual devices into a two-dimensional array on a single wafer. The diameter of such devices can be on the order of $\sim 2\ \mu$m, so that ~ 1 million VCSELs could be packed into a 1-cm^2 chip. In addition, owing to the very small active-region volume, the threshold current for achieving lasing can theoretically be as small as 10 μA. One potential drawback of VCSELs is the relatively low output power of these devices, which is typically less than 1 mW; however, the overall advantages of low cost, excellent output beam quality, and high packing density will undoubtedly allow VCSELs to be incorporated into many commercial applications.

3. Quantum Cascade Lasers

Bandgap engineering techniques of controlling the composition and thicknesses of the layers in semiconductor lasers have allowed for the unprecedented control of the energy levels and thus control of the emission wavelength and other unique properties of semiconductor lasers. The quantum cascade laser is an excellent example of how bandgap-engineered quantum well structures can be used to develop new laser materials and the devices designed from these materials.

The quantum cascade laser is very different from the conventional semiconductor laser that has been described in this article, and is based on the intersubband transitions between the excited states of coupled quantum wells, or superlattice structures, and on the resonant tunneling between the wells as the pumping mechanism. This means that lasing action takes place between energy levels within the conduction band (not between the conduction and valance bands). More importantly, since the electron is still in the conduction band, novel bandgap engineering can provide for a transport mechanism that allows for this electron to be reinjected into another set of coupled quantum wells, and is therefore reused. As a result, one injected

electron can be used to generate many photons. This type of semiconductor lasers is also referred to as a unipolar device since laser action occurs with one type of carrier, e.g., the electron, as compared to the required recombination of an electron and a hole in a conventional semiconductor diode laser.

A key feature of quantum cascade laser is that it uses semiconductor superlattice structures, which are periodic stacks of nanometer-thick quantum well and barrier pairs. The periodicity created by the stack of quantum well/barrier pairs is typically several nanometers, as compared the atomic lattice spacing of a few angstroms. The periodicity of the barriers creates a splitting of the conduction (and valance bands) so that a series of minibands is created. The separation, or gaps, and widths of these minibands are on the order of a few hundred millielectronvolts and can allow for transport of carriers in a direction normal to the layers. For a given choice of materials, the miniband gaps and widths can be engineered by a suitable choice of the layer thicknesses. In addition, this control of the layer thickness also provides a degree of freedom that allows the tunneling or scattering times between wells to be designed, and adds a unique dimension to quantum-engineered semiconductor laser devices. Owing to the dimensions of the quantum wells and the subsequent energy levels, cascade lasers are typically designed to operate in the mid-infrared wavelength region, for example, 3–20 μm.

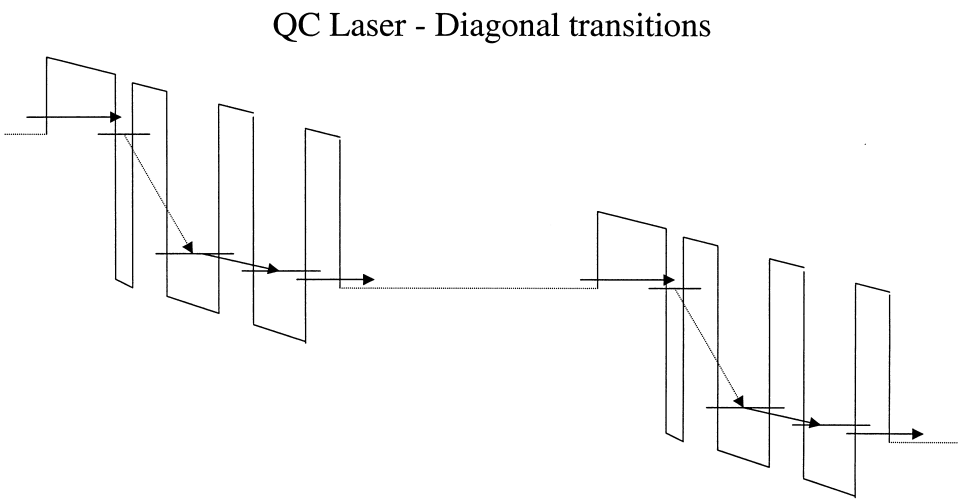
Figure 21 shows two types of quantum cascade intersubband energy band diagrams, one that employs diagonal transitions and one that employs vertical transitions. In the quantum cascade laser employing diagonal transitions, the operation is as follows. Electrons are injected into the active region, which consists of, for example, a pair of GaInAs quantum wells (0.8 and 3.0 nm) separated by a 3.5-nm AlInAs barrier. The thin, 0.8-nm well serves as the upper lasing level, while the 3.0-nm well serves as the lower lasing level. It should be recalled that the thickness of each quantum well determines the allowed energy levels, i.e., thin layers generate high-lying energy levels, while thicker layers allow for lower energy level states. Once an electron makes a photon-generating, lasing transition from level 3 to level 2, the electron quickly tunnels through a 3.0-nm AlInAs barrier to a 2.8-nm GaInAs well. This strong coupling of the 3.0- and 2.8-nm wells provides for a fast tunneling of the electron out of the lower lasing level, to prevent electron buildup. The electron in level 1 then undergoes a fast tunneling through a AlInAs barrier into a digitally graded semiconductor alloy, or superlattice injector section. The salient feature of the injector section is that the superlattice structure generates a miniband that allows electrons to tunnel from the lower lasing level of one active-layer region to the upper lasing level of the adjoining active-layer region on the other side of the injector. This electron can now undergo the processes as describe above and continue to contribute additional lasing photons. The operation of the quantum cascade laser that employs vertical transitions is similar. Injected electron from the injector section undergo a photon-generated, lasing transition where the electron can quickly tunnel out of the lower lasing level into the adjoining quantum well. Subsequent tunneling into the next injector section continues the process.

There are many variations on these standard models of quantum cascading laser action. Cascade lasing action can also be achieved by using conventional interband transitions between the conduction and valance bands. In this case, the injector is designed to allow for tunneling from the lower lasing level in the valence band to the upper lasing level in the conduction band at the end of the injector. From these examples of quantum cascade lasers, it should be clear that the ability to control, with atomic accuracy, the thicknesses of semiconductor layers and, as a result, the energy levels and their relaxation rates brings a unique capability to the design and operation of semiconductor lasers.

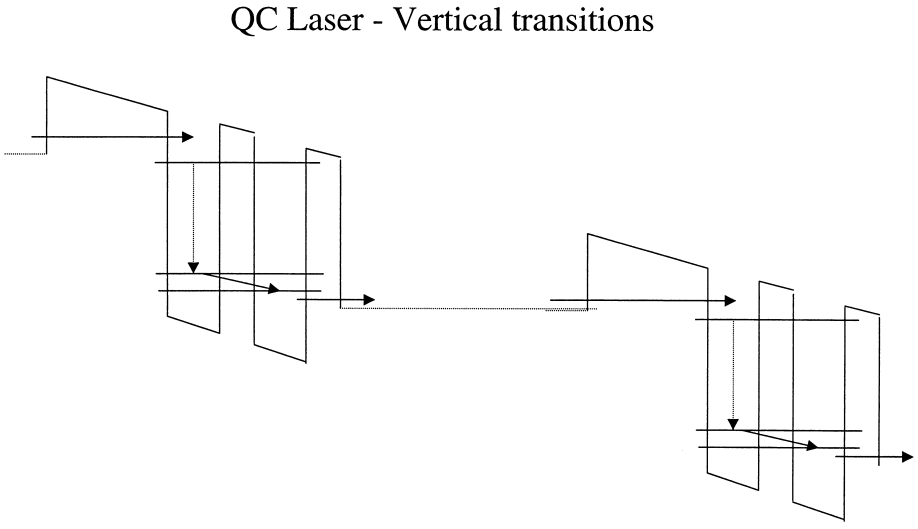
4. Semiconductor Optical Amplifiers

Semiconductor optical amplifiers (SOA), also referred to as semiconductor laser amplifiers (SLA), are devices very similar to semiconductor lasers, which amplify light that is injected into the device. The principle behind the operation of a semiconductor optical amplifier is identical to that of other semiconductor lasers. It is the creation of a population inversion that allows stimulated emission and optical gain. As in a conventional semiconductor laser, the population inversion is achieved by injecting carriers by an electrical current into the active region, which subsequently recombine through spontaneous and stimulated emission.

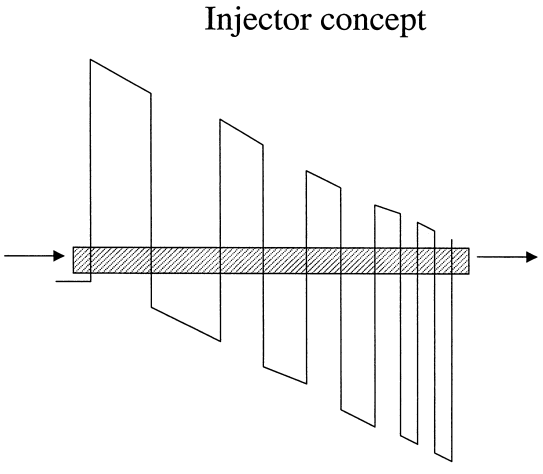
Semiconductor optical amplifiers can be realized by employing a conventional Fabry–Perot laser or by fabricating a traveling-wave optical amplifier. The Fabry–Perot amplifier has large reflection coefficients at both end facets of the device, which lead to a resonant optical amplification of light signals that have wavelengths that correspond to the resonant peaks, or longitudinal modes, of the FP cavity. On the other hand, the traveling-wave optical amplifier is designed to have close to zero reflectivity on the ends of the cleaved facets. Since the effects of the reflectivity of the facets is removed this allows signal amplification over a continuous broad wavelength range that corresponds to the gain of the semiconductor medium. It should be noted, however, that in the FP amplifiers, owing to the resonant amplification, the overall



(a)



(b)



(c)

FIGURE 21 Energy level diagram of the quantum cascade structure. (a) Diagonal transitions, (b) vertical transitions, (c) detailed energy diagram of the injector used in both vertical and diagonal cascade laser structures.

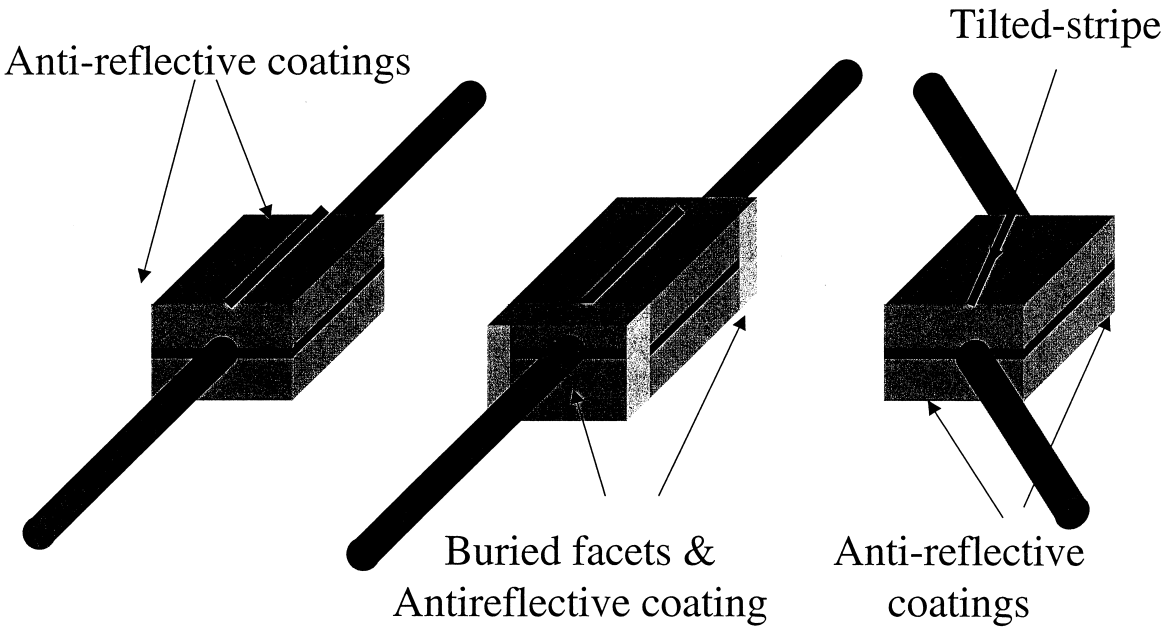


FIGURE 22 Semiconductor optical amplifier structures: Antireflection coated, buried facets, and angle striped.

gain can be large, but since the gain is resonant, the injected signal must experience multiple reflections in the FP cavity. For TWA structures, the gain experienced by the injected optical signal occurs in a single pass, and thus has very useful properties for amplifying picosecond optical pulses that will be used in high-capacity optical communications applications. Owing to these features of broadband optical gain and ability to amplify short optical pulses, considerable effort has gone into developing traveling-wave optical amplifiers that minimize the facet reflectivities.

For good traveling-wave optical amplifier performance, a salient feature of the device design is the ability to achieve a low facet reflectivity. In general, there are three methods to achieve low facet reflectivity: (1) high-quality antireflective dielectric facet coatings, (2) buried-facet amplifier design, and (3) tilted-stripe amplifiers (see Fig. 22). Below, a brief discussion is given for each approach.

a. High-quality antireflective coatings. In order to assure good performance in traveling-wave semiconductor optical amplifiers, the effect of the facet reflections must be minimized. This can be achieved by depositing a layer of a dielectric material on the facets such that the combined effect of the facet reflection and the dielectric coating effectively reduces the reflection of the facet. The key property of the dielectric coating is that the thickness of the layer should be one-quarter of the wavelength of light that the amplifier is designed for, and the index of refraction of the dielectric coating should be $(n_o n_s)^{1/2}$, where n_s is the index of refraction of the semiconductor medium and n_o is the index of the medium that the light will enter, e.g., n_o is generally taken as $\sim$1 since the light from the SOA enters into free space. It should be noted, however, that a coating could be designed such that $n_o \cong 1.5$, in order to directly attach an optical fiber to the facet of the SOA. It should also be noted that the effective reduction in reflectivity should have a residual effective reflectance of 10^{-4} to supress the FP modes of the device and to prevent lasing since the typical single-pass gain of a semiconductor optical amplifier can be in excess of 30 dB. The lowest reflectance occurs over a narrow range of wavelengths, but is still broader than the separation of the Fabry–Perot modes of the FP optical amplifier. To achieve a reduction over a broader wavelength range, alternative coating approaches have been considered that use multiple coatings instead of a single coating.

b. Buried-facet amplifier design. An alternate approach to achieve a low effective facet reflection is to fabricate a buried facet amplifier. The key concept of the buried facet design, sometimes referred to as a buried window, is to remove portions of the optical waveguide where gain occurs and replace the removed material with semiconductor material that has the same index as the semiconductor laser but is transparent to the amplified optical signal. In this design the amplified optical signal exits the optical waveguide and starts to diffract. Owing to the beam spreading during diffraction, the optical beam does ultimately reflect from the cleaved facet, but the diverging amplified optical beam is not efficiently coupled back into the waveguide. This provides a low effective facet reflectivity, even though

the optical reflection coefficient of the semiconductor material has not been altered. The effective reflectivity of a buried facet decreases with increasing distance between the facet and the end of the active region. While increasing the length of the buried facet region decreases the reflectivity, a limit is imposed by the diffraction of the amplified beam. If the length is too long, the beam emerging from the active region will spread and reflect from the top metalized surface, leading to lateral interference fringes in the optical beam, which decreases the coupling efficiency into an optical fiber. In practice, a simple antireflective coating is also used to assist in reducing any residual facet reflectivity. The nominal index of refraction of semiconductor laser media is ∼3.3–3.5 and yields a reflection coefficient of the facet of ∼30%. The additional antireflection coating used in conjunction with the buried facet helps to reduce this facet reflectivity, which allows more of the injected light to enter the optical amplifier and also allows more of the amplified light to be extracted.

c. Tilted-stripe amplifiers. A third method for reducing the facet reflectivity of a conventional Fabry–Perot semiconductor laser is to fabricate the gain stripe so that it is tilted with respect to the cleaved facets. The effect of tilting the gain stripe is that the amplified light that is internally incident on the cleaved facet reflects from the facet and is directed away from the gain stripe so the reflected beam is prevented from experiencing any additional gain. This is in direct contrast to conventional FP devices, where the gain stripe is perpendicular to the cleaved facets. Any residual reflection from the cleaved facet in the FP case is easily coupled back into the gain region, thus requiring an antireflection coating of high quality. As in the buried facet approach, the tilted-stripe approach provides a reduction in the facet reflectivity without a need to modify the cleaved facet surfaces. The amount of residual reflectivity is controlled by the tilt angle of the gain stripe, and for most devices the tilt angle ranges from 5 to 7 deg. Also, as in the buried facet approach, simple antireflection coatings are applied to the facets to allow for increased optical injection and better optical extraction of the amplified light. Owing to the simplicity of fabrication in the tilted-stripe approach, this approach is becoming the preferred design of semiconductor optical amplifiers. One drawback of the tilted-stripe approach is that the injected and extracted beams do not exit the device in a direction perpendicular to the cleaved facets, but are refracted in a direction away from the normal direction. The relatively large refractive index of the semiconductor gain medium implies that the injected and emerging beams must enter and exit the optical amplifier at an angle from 15 to 20 deg. This refractive effect makes it more difficult to couple light in and out of the optical amplifier with optical fibers. It should be noted that optical amplifiers serve as optical gain media as in any other conventional laser amplifier. As such, these devices can then be employed as gain devices in external cavity lasers, i.e., lasers that employ conventional collection optics and mirrors. The reason for employing these devices as gain media in external cavity lasers is that complex cavity designs can be employed which would not otherwise be possible with conventional semiconductor fabrication techniques. As a final note on semiconductor optical amplifiers, these devices can be made to realize a broad range of functionalities. For example, since the optical amplifier is powered by an electric current, the current of the device can be modulated to impress information or data onto the injected optical beam. In this case the device is operating as an optical modulator; however, the modulator device in this case possesses optical gain. Also, SOA devices have been used as devices in external cavities for ultranarrow-linewidth lasers, high-power cw lasers, and modelocked lasers. The nonlinear optical properties of SOAs also allow them to be used for wavelength converters, i.e., devices that can transfer information impressed on one optical beam and convert it into another wavelength, and also as all-optical switches, i.e., devices that can perform switching or routing/demultiplexing of a data stream that is controlled by another optical signal.

5. Modelocked Semiconductor Laser

Modelocking is a technique for obtaining very short bursts of light from semiconductor lasers, and can be easily achieved employing both integrated semiconductor devices and external cavity designs (see Fig. 23). A salient feature of modelocking is that the temporal duration of the optical pulse produced by modelocking is much shorter than the period of the driving signal. To contrast this, consider a DFB laser whose light is externally modulated. In this case, the temporal duration of the optical signals will be proportional to the temporal duration of the electrical signal driving the external modulator. As a result, in a communication application, the maximum possible data transmission rate achievable from a DFB will be limited to the speed of the electronic driving signal. With modelocking, however, a low-frequency electrical drive signal can be used to generate ultrashort optical pulses. By following the light production with external modulation and optical bit interleaving, one can realize the ultimate in optical time division multiplexing (OTDM) transmission rates. To show the difference between a modelocked pulse train and its drive, Fig. 23 plots a sinusoid and a modelocked pulse train consisting of five locked optical modes.

To understand the process of modelocking, it should be recalled that a laser can oscillate on many longitudinal modes that are equally spaced by the longitudinal mode spacing $c/(2nL)$. Normally, these modes oscillate independently; however, techniques can be employed to

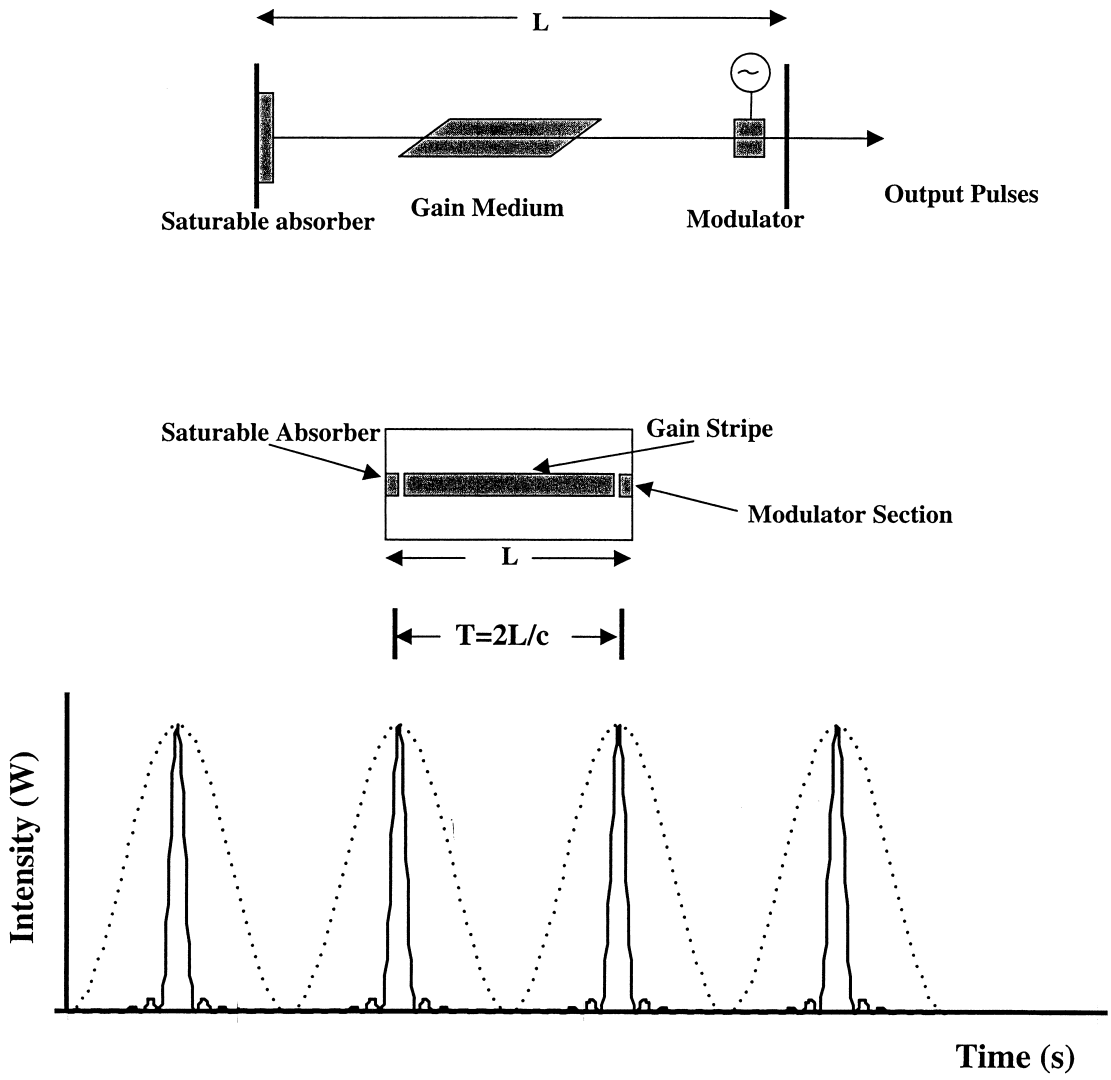


FIGURE 23 Modelocked semiconductor laser. External cavity, using a semiconductor optical amplifier as the gain element, and a monolithic integrated structure.

couple and lock their relative phases together. The modes can then be regarded as the components of a Fourier series expansion of a periodic function of time of period $T = 2nL/c$ that represents a periodic train of optical pulses. Consider, for example, a laser with multiple longitudinal modes separated by $c/(2n\mathrm{L})$; the output intensity of a perfectly modelocked laser as a function of time t and axial position z, with M locked longitudinal modes each with equal intensity, is given by

$$I(t, z) = M^2|A|^2 \frac{\sin c^2[M(t - z/c)/T]}{\sin c^2[(t - z/c)T]}, \qquad (21)$$

where T is the periodicity of the optical pulses and $\mathrm{sinc}(x)$ is $\sin(x)/x$. In practice, methods to generate optical pulse trains by modelocking generally fall into two categories: (1) active modelocking and (2) passive modelocking. In both cases, to lock the longitudinal modes in phase, the gain of the laser is allowed to increase above its threshold for a short duration by opening and closing a shutter placed within the optical cavity. This allows a pulse of light to form. By allowing the light to propagate around the cavity and continually reopening and closing the shutter at a rate inversely proportional to the round-trip time, a stable, well-defined optical pulse is formed. If the shutter is realized by using an external modulator, the technique is referred to as active modelocking, whereas if the shutter is realized by a device or material that is activated by the light intensity itself, the process is called passive modelocking. Both techniques can be used simultaneously, referred to as hybrid modelocking (see Fig. 23b).

From the above equation, it is observed that the pulse duration is determined by the number of modes M, which

in practice is generally limited by the gain bandwidth of the medium. Since the gain bandwidth of semiconductor lasers can be very broad, the resultant pulse width can be very short. In addition, since the modes are added coherently, the peak intensity is M times the average power, making these optical pulses sufficiently intense for inducing nonlinear optical effects. Generally, large optical power in many applications is useful for large signal-to-noise ratios in the detection process; however, other effects, such as nonlinear optical effects, can be detrimental. While nonlinear optical effects are typically avoided in certain applications, such as data transmission, the peak intensity may exploit novel forms of optical propagation, as in optical soliton propagation. In addition, ultrafast all-optical switching and demultiplexing only become possible with such high-peak-intensity pulses. As a result, modelocked semiconductor lasers may ultimately become the preferred laser transmitters for telecommunications.

IV. APPLICATIONS

Since the development of semiconductor lasers, there has been an increasing desire for their utilization in many applications due to their compact size, efficient electrical to optical conversion, and extremely low cost. In this section, several applications of semiconductor injection lasers are reviewed for both scientific and commercial applications.

A. Diode-Pumped Solid State Lasers

The advances of semiconductor laser arrays have recently caused a resurgence of research activity in the solid state laser community. This is mainly because diode laser arrays have the potential to be used as an efficient optical pump for these laser systems. The advantages of a semiconductor laser array pump as compared to a conventional flashlamp-pumped system are numerous. The semiconductor laser arrays are compact, efficient, robust, and potentially inexpensive. The emission wavelength of the laser arrays can be controlled by varying the material composition during the wafer growth. As a result, the emission wavelength can be made precisely to coincide with the peak absorption lines of the solid state laser crystal, thus making the laser array a much more efficient optical pump than to a broadband flashlamp. In addition, diode pumps produce less undesirable heating of the solid state laser crystal and less potential damage that can occur due to the high-energy photons present in a flashlamp source. Another attractive feature of a semiconductor laser array pump is that the light emitted is partially coherent, enabling the light to be efficiently focused or mode matched to the solid state laser system.

Specific examples of semiconductor laser array pumps include devices made from the AlGaAs material system. The emission wavelength can be varied from 700 to 900 nm by varying the Ga/Al composition. These emission wavelengths can be made to coincide with absorption bands of several important solid state lasing ions, such as neodymium, chromium, holmium, erbium, and promithium.

Solid state lasers can be made to operate in various manners and geometric configurations, all of which can utilize a diode array as the pump source. The solid state laser can be an optical fiber, a slab, or a standard cylindrical rod. The solid state laser may operate continuous wave (cw), quasi-cw or pulsed, or Q-switched, and the emission may be single or multi spatial and longitudinal mode. Each operating condition and geometry places different requirements on the pump source. Longitudinal or end pumping is most efficient for optical fiber and rod active media, whereas transverse or side pumping is particularly suitable for slab active media.

The diode laser arrays used as the pump source in a longitudinally pumped solid state laser system typically employ an array with 10–20 emitters, each with a stripe width of 100 μm, separated by 250–500 μm. The arrays are rated to operate $\sim$5 years with up to $\sim$60 W of cw output. These single-bar arrays can be stacked into a two-dimensional structure yielding output powers approaching 1 kW. The pumping is normally coupled by optical fibers or bulk optics so that there is mode matching between the diode pump and the solid state laser. The advantages of this geometry are the long absorption path of the pump light and the efficient utilization of the pump light resulting from a mode-matched system.

Figure 24 illustrates a typical longitudinally diode-pumped solid state laser system. The diode array is made from the AlGaAs material system and is mounted on a thermoelectric cooler that controls the operating temperature and the emission wavelength of the laser array. The coupling optics collect the light emission from the laser array and spatially manipulates the beam shape so that the transverse spatial mode patterns of the diode pump and the Nd:YAG laser cavity are matched. The laser rod is an Nd:YAG crystal that has high-reflecting (HR) and antireflecting (AR) coatings to allow the reflection and transmission of the desired wavelengths. The 95% reflecting curved mirror serves as the output coupler of the complete Nd:YAG laser. The longitudinally pumped scheme is limited in output power by the diode output power per unit width. Higher power can be obtained by overdriving the laser arrays, which ultimately reduces the operating lifetime of the array. One simple solution is to use several arrays, each coupled into an optical fiber. The fiber is then combined into a bundle and then used as the pumping

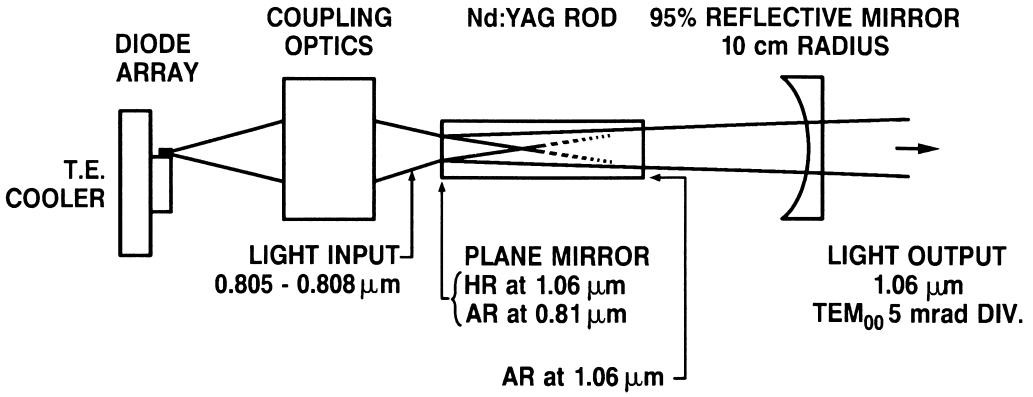


FIGURE 24 Schematic of a longitudinally diode-pumped Nd:YAG laser. [After Streifer *et al.* (1998).]

source. Systems employing this approach are commercially available, and may generally employ two 25-W diode array bars operating with an output power of 35 W to increase the diode pump's useful lifetime. Utilizing this method, ~25 W of TEM_{00} power was produced from an Nd:YAG laser emitting at 1.06 μm. Intracavity frequency doubling can also be employed, which can produce ~10 W of light at 530 nm from this system.

The field of semiconductor laser arrays is still a new and developing technology. With future advancement in achieving higher output power and higher efficiencies, their use in increasingly larger solid state laser systems will be unavoidable. In addition, as manufacturing volumes and yields increase, the cost of these pump sources will decrease with an improvement of diode quality. Even with the present state of technology, it is apparent that semiconductor laser arrays will be an important player in the field of compact solid state laser systems.

B. Optical Communications

Semiconductor lasers have a tremendous potential for use in optical communication systems due to their high efficiency, small size, and direct modulation capability. Because of their small size, they become the ideal source for optical radiation to be utilized in an optical fiber transmission system. The reason for developing an optical communications system is that these systems can take advantage of the large bandwidth that optics has to offer. For example, in a conventional copper wire communication system, the bandwidth of the transmission channel, i.e., the coaxial cable, is limited to approximately 300 MHz. In a normal telephone conversation, frequency components up to 3 kHz are required for the listener to understand the conversation. By multiplexing many telephone conversations on different carrier frequencies, approximately 100,000 conversations can be sent through a single conventional coaxial cable without interference between conversations. However, in an optical fiber communication system, the bandwidth of the fiber is typically several terahertz. Thus, by multiplexing the conversations with a suitable optical carrier signal, more than one billion conversations can be sent over a single optical fiber. This example clearly shows the potential advantage of utilizing an optically based communication system.

Communication systems have three main components: the transmitter, the receiver, and the transmission path. Each component has certain characteristics which ultimately define how the communication system is to be designed. An example of this is clearly seen in the early designs of an optical communication system. These systems were based on the GaAs semiconductor laser wavelengths. The main reason for this was only due to the availability of these devices. However, due to the physical properties of the transmission channel, i.e., the optical fiber, this wavelength region was considered only to be appropriate for short-distance communication networks due to dispersive and absorption losses. Long-distance communication networks that were based on this wavelength would be costly due to the losses of the transmission channel and the necessity of a short repeater spacing. With the development of longer wavelength semiconductor injection lasers from the InP/InGaAsP material systems, communication systems can take advantage of special properties characteristic of glass optical fibers.

In most optical communication systems, the information is encoded onto a beam of light and transmitted through an optical fiber. In an ideal system, the information is undistorted by the transmission path. However, in a real optical fiber, the power level of the signal is decreased as it is transmitted, due to absorption and scattering losses in the fiber. As a result, the signal needs to be amplified or regenerated along the transmission path in order to be detected at the receiver side. Another detrimental effect

caused by the fiber is due to the chromatic or group velocity dispersion. This has an effect of broadening bits of information, causing them to overlap in time. These detrimental effects are minimized by utilizing specific wavelengths that take advantage of special physical properties of optical fibers. For example, optical fibers have a minimum loss window at 1.55 μm, typically 0.2 dB/km, and a minimum dispersion window at 1.3 μm, typically 3 psec/km/nm. By utilizing semiconductor lasers that emit in these wavelength regions, these problems can be minimized. The InP/InGaAsP semiconductor lasers emit in these wavelength regions and have been developed sufficiently so that long-distance communication systems are being developed so that designers can take advantage of these optical windows. Most recently, optical communication systems are being deployed using a multiplexing format referred to as wavelength division multiplexing (WDM). In this multiplexing scheme, different wavelengths are used to carry information, analogous to conventional radio communication systems. An active area of semiconductor laser research is in the development of semiconductor lasers that can simultaneously emit multiple wavelengths compatible with these new communication multiplexing formats. Owing to the use of closely spaced multiple wavelengths, other nonlinear optical properties of optical fibers manifest themselves, such as four-wave mixing, which exchanges energy (and information) between these wavelengths. This effect, similar to dispersion, causes a degradation of the fidelity of the communication channel and is referred to as cross talk.

In general, present-day optical communication systems utilize an intensity modulation scheme to encode information onto an optical beam. In these schemes, the coherent properties of the laser are not employed. For detection, the detector simply counts the number of incident photons during a given amount of time in order to recover the modulated signal. Although the intensity modulation scheme does not utilize the coherent properties of the semiconductor laser output, many applications still require single longitudinal mode lasers in order to reduce the effects of chromatic dispersion in an optical fiber. These effects are most important for long-distance fiber communications.

It should be noted, however, that many additional advantages can be obtained in an optical communication system if the system exploits the coherent properties of the laser output. As in a conventional radio communication system, coherent optical communication systems have a local oscillator at the receiver side. By employing a local oscillator, the detected signal sensitivity is increased by an order of magnitude. This ultimately translates into a longer repeater spacing. This feature is very important for ultralong-distance communications, i.e., transoceanic optical communication links. Figure 25 shows a typical coherent optical communication system. The transmitter consists of a semiconductor injection laser, an external modulator, and an amplifier which amplifies the modulated signal before it is launched into the transmission channel. The transmission channel consists of an optical fiber and an optical repeater, which regenerates the transmitted signal. On the receiver end is a photodetector which detects the transmitted signal along with the local oscillator signal. The demodulation circuit demodulates the signal from the photodetector to produce the desired received information. The demodulation circuit also provides a feedback signal to the automatic frequency control circuit. This circuit controls the frequency of the local oscillator so that it can track the fluctuation of the input frequency.

There are numerous modulation schemes that depend upon which parameter of the electric field is modulated, i.e., amplitude shift keying (ASK), frequency shift keying

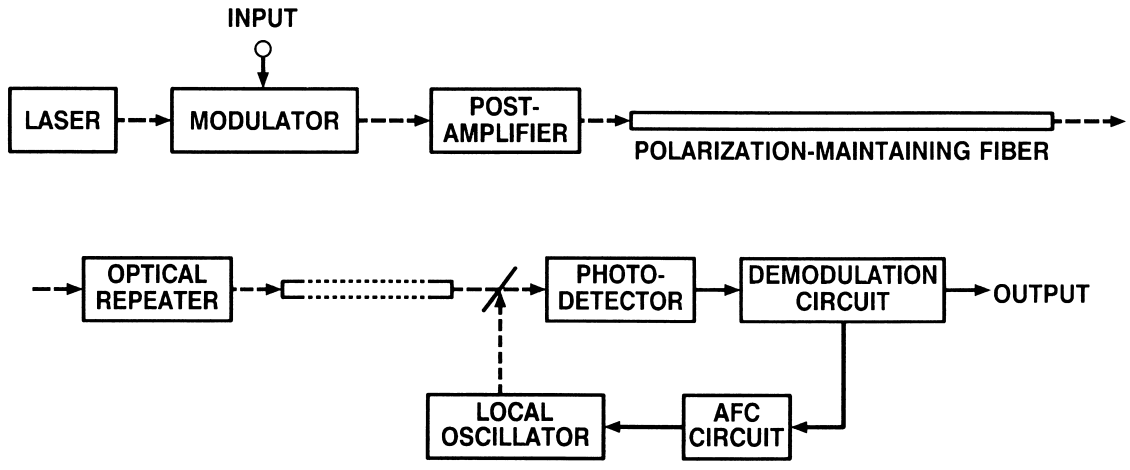


FIGURE 25 Schematic of a coherent optical communication system.

(FSK), and phase shift keying (PSK). Since the coherent properties of the laser output are utilized in these systems, the bandwidth of the transmitted signal can be reduced to the bandwidth of the modulation signal. Thus, by appropriately multiplexing many signals together, whether by time division multiplexing (TDM) or frequency division multiplexing (FDM), the communication system can take full advantage of the channel capacity that the optical fiber has to offer.

To show how semiconductor laser technology can realize state-of-the-art performance in an optical time division multiplexed (OTDM) telecommunications system, an example is shown here that incorporates a variety of the device technologies discussed in this article. While there have been many OTDM links operating in excess of 100 Gb/sec using different device technologies for pulse generation and subsequent demultiplexing, the general concepts are consistent between each demonstration, and are reproduced here to bring together the concepts of OTDM. Figure 26 shows an ultrahigh-speed optical data transmitter and demultiplexing system. To demonstrate the system performance, the transmitter can realized by employing a 10-GHz modelocked semiconductor laser to generate ~3-psec optical pulses. The data pulses are then modulated using a pseudo-random non-return-to-zero electrical signal that drives a 10-Gb/sec $LiNbO_2$ modulator. The optical pulses are then reduced in temporal duration using optical nonlinear effects in fiber, e.g., adiabatic soliton pulse compression, resulting in optical pulses of 250 fsec in duration. The pulses are then temporally interleaved by a factor of 64 using a planar light-wave multiplexer. This device is simply a cascade of integrated Mach–Zehnder (MZ) interferometers that employ a fixed delay equal to half of the input repetition rate for each MZ stage. The resulting output from the PLC MUX is a pseudo-random modulated pulse train at 640 Gb/sec. In a real system, the data from 64 individual users would need to be interleaved accurately for this scheme to work. This, however, can be achieved by distributing or broadcasting the initial 10-Gb/sec pulse train to each user for modulation. Since the propagation distance to each user is fixed and deterministic, the return signals can be interleaved easily.

After the data has been generated and amplified in an optical amplifier, such as a semiconductor optical amplifier, to compensate for losses, the pulses are launched into

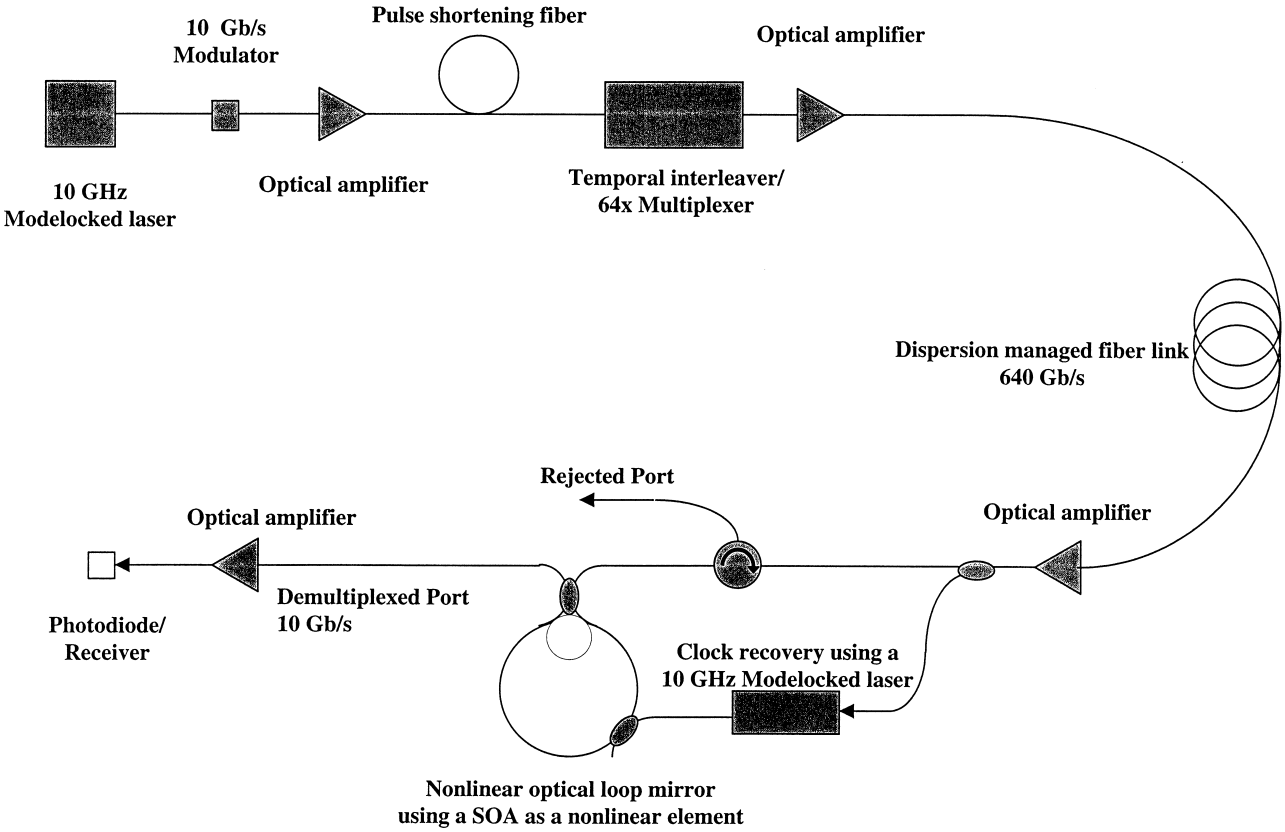


FIGURE 26 High-speed optical communication link.

a fiber span of approximately 60 km. It should be noted that the typical distance between central switching stations or central offices in conventional telecommunication networks is 40–80 km. The fiber span is composed of fiber with varying dispersion, to reduce the effects of pulse broadening. The pulses at the demultiplexing end are initially amplified using an optical amplifier to again compensate for losses encountered in transmission. A portion of the data is extracted from the data line and is used to generate an optical clock signal that will be used as a control pulse in an all-optical switch/demultiplexer. The clock recovery unit comprises a modelocked laser that operates at the nominal data rate of 10 GHz. This clock signal could also be generated by injection locking a passively modelocked semiconductor laser. Once the clock pulses are generated, they are injected into a nonlinear loop mirror along with the data to provide all-optical demultiplexing at 10 Gb/sec. The clock and data pulses generally are distinguishable either by wavelength or polarization, so that the clock and data can easily be separated or filtered after demultiplexing. Finally, after the data pulses are separated, detection is performed with a photodetector, with the subsequent electrical signal analyzed by a bit-error-rate measurement system. The resulting performance of this system showed a bit error rate of 10^{-10}, i.e., less than one error for every 10 billion bits received, with a received power of -23 dBm or 5 mW of average power.

V. DIRECTIONS AND TRENDS

The field of semiconductor lasers has spanned the past four decades. Since its development, it has revolutionized the science of laser optics and given birth to the field of integrated optoelectronics. The field is continuing to grow at a tremendous rate, with new laser structures yielding lower threshold currents and higher output power being reported in the technical journals each month. With advances such as these, it is apparent that the future will be even more exciting than the development up to the present. It is difficult to predict the future; however, recent trends point to areas in which tremendous progress will be made. Below, a few areas of semiconductor research that will play an important part in the future development of the field are highlighted.

New semiconductor laser structures will rely on the physics of quantum wire and quantum dot effects. This will enable the devices to operate with only a few microamperes of injection current. In addition, these devices will have tunable bandwidths spanning several tens of nanometers.

Two-dimensional surface emitting lasers represent another area that will contribute to future semiconductor laser development. These devices have the potential to impact high-power semiconductor laser arrays, optical storage, optical computing, and multiwavelength lasers.

Semiconductor lasers have certainly advanced to the stage where in the near future they will replace the more common solid state and gas lasers that have been the workhorses in both the scientific and industrial arenas. Semiconductor lasers will be successfully modelocked and the resultant ultrashort optical pulses will be amplified to peak power levels approaching the kilowatt region. This will have a tremendous impact on the ultrafast laser community, by providing an inexpensive, efficient, and compact source for ultrafast nonlinear optical studies. In addition, real-time optical signal processing and optical computing will take one step closer to reality with this advancement.

New wavelengths will become available by utilizing strained layer semiconductor material and expanding on the idea of the quantum cascade laser. These new material structures will not only provide wavelengths previously unattainable from semiconductor injection lasers, but will undoubtedly uncover new physics resulting from the combined interaction of quantum structures and strained layer materials.

Visible diode lasers are now commercially available with wavelengths in the 600-nm regime and can be purchased for a few dollars. Utilizing wide-bandgap III–V material systems such as AlGaN; wavelengths as short as 400 nm are now becoming available. The major hurdle in this field is that p–n junctions of sufficiently high quality are difficult to fabricate due to the outdiffusion process of the p-type dopant. However, present research indicates that this problem will soon be overcome.

Another area in which semiconductor lasers are having a major impact is in the area of optical amplification. As noted earlier, by simply modifying the device structure of a Fabry–Perot laser, a semiconductor laser is transformed into an amplifying device. This type of device is of paramount importance in integrated optical receivers, optical repeaters, wavelength converters, all-optical switches, and broadband optical amplification for WDM communication systems.

Multi-terabit communication systems are now becoming a part of current communication networks. These links and networks will continue to advance with the continued development of high levels of integrated semiconductor optoelectronics, incorporating lasers, modulators, amplifiers, detectors, wavelength converters, and drive electronics on a single microchip. With continued advances in high-speed electronics, these devices will directly emit pulses on the order of a trillionth of a second, with repetition rates exceeding 100 GHz.

From this brief look into the future, it is apparent that semiconductor lasers will tremendously impact the telecommunications, medical, scientific, and industrial

communities. The advances in device technology have far exceeded the expectations of the early researchers. We can only assume that the past advances will continue into the future, bringing devices and technologies which at this time can only be dreamed of. How these new technologies will shape the development of the world and what impact they will have on society remain to be seen.

SEE ALSO THE FOLLOWING ARTICLES

EPITAXIAL TECHNOLOGY FOR INTEGRATED CIRCUIT MANUFACTURING • EXCITONS, SEMICONDUCTOR • LASERS • LASERS, OPTICAL FIBER • LASERS, SOLID-STATE • LIGHT EMITTING DIODES • METALORGANIC CHEMICAL VAPOR DEPOSITION (MOCVD) • MOLECULAR BEAM EPITAXY, SEMICONDUCTORS • OPTICAL AMPLIFIERS (SEMICONDUCTOR) • SEMICONDUCTOR ALLOYS

BIBLIOGRAPHY

Agrawal, G. A., and Dutta, N. K. (1993). "Semiconductor Lasers," 2nd ed., Van Nostrand Reinhold, New York.

Aschroft, N. W., and Mermin, N. D. (1976). "Solid State Physics," Holt Reinholt Winston, New York.

Kapon, E. (1989). *In* "Handbook of Solid State Lasers" (P. K. Cheo, ed.), Marcel Dekker, New York.

Lau, K., Bar-Chaim, N., Ury, I., and Yariv, A. (1983). "Direct amplitude modulation of semiconductor GaAs lasers up to X-band frequencies," *Appl. Phys. Lett.* **43,** 11.

Saleh, B. E. A., and Teich, M. C. (1991). "Fundamentals of Photonics," Wiley, New York.

Streifer, W., Scifres, D., Harnagel, G., Welch, D., Berger, J., and Sakamoto, M. (1988). "Advances in diode laser pumps," *IEEE J. Quant. Electron*. **QE-24,** 883.

Thompson, G. H. B. (1980). "Physics of Semiconductor Laser Devices," Wiley, New York.

Yariv, A. (1975). "Quantum Electronics," 2nd ed., Wiley, New York.

Yariv, A. (1988). "Quantum Electronics," 3rd ed., Wiley, New York.

Lasers, X-Ray

Raymond C. Elton
Naval Research Laboratory

I. Introduction
II. Pumping Modes and Experimental Results
III. Characteristics of Plasma X-Ray Lasers
IV. Alternative (to Plasmas) and Complimentary Approaches
V. Applications
VI. Summary and Prognosis

GLOSSARY

Amplified spontaneous emission (ASE) Amplification of internal spontaneous emission from a transition throughout the medium on a single pass.
Coherence Correlation between instantaneous amplitudes and phase angles of two or more waves.
Holography Formation of three-dimensional images by recording and then reconstructing amplitude and phase distributions of a wave disturbance.
***jj* coupling** Coupling of the total angular momenta j for multiple bound electrons in an atom or ion, where $j = l + s$ for coupled orbital l and spin s electron momenta.
Photolithography Photographic formation of relief structures in semiconducting materials for electronic microcircuitry.
Metastable state Excited energy state in an atom or nucleus whose lifetime is unusually long.
Plasma Ionized gaseous medium in which the net charge is zero because of a balance between the total charges of ions and electrons present.
Self-terminating laser Laser in which the population inversion terminates naturally when the lower state becomes sufficiently populated from the upper state to destroy the inversion.

X-RAY LASERS are devices that amplify very short wavelength radiation along a well-defined direction in order to obtain an intense, collimated, and coherent beam of monochromatic and penetrating emission. Such a radiation source can be used for photolithography of microcircuitry, holographic microscopy of cellular structure in biology, plasma diagnostics, and many other scientific and technical purposes.

I. INTRODUCTION

In principle, a *laser* (light amplification by stimulated emission of radiation) is not limited to wavelengths normally associated with "light," that is, the visible wavelength range of approximately 380–780 nm (see Fig. 1).

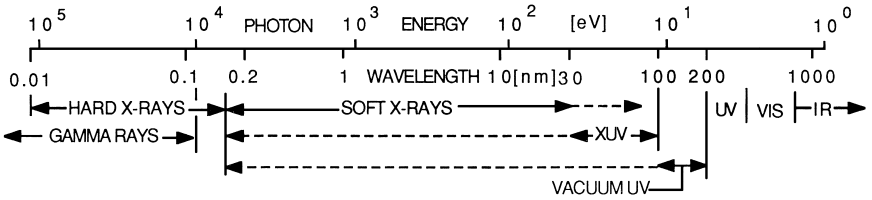


FIGURE 1 Wavelength and photon-energy ranges and designations. [From Elton, R. C. (1990). "X-ray Lasers," Academic Press, San Diego.]

In fact, the first device demonstrating such amplification operated in the microwave region (millimeters to centimeters in wavelength) as a "maser." However, the availability of high-quality reflecting mirrors for resonant cavities has resulted in the development of powerful and useful coherent beams of collimated radiation in the infrared, visible, and near ultraviolet spectral regions. From the very earliest experiments and publications on lasers, there has been anticipation of the eventual further extension into the extreme ultraviolet (xuv) and X-ray spectral regions (see Fig. 1). This might occur either by a giant leap or by a gradual advance through the vacuum ultraviolet (vacuum-uv or vuv) region, in which material penetration and optical properties are generally poor compared to the visible region. In either case, it was recognized that the challenges would be formidable. This was basically because of what were recognized to be four somewhat-associated difficulties, discussed in the following paragraphs.

(a) The first difficulty is a generally unfavorable scaling of the cross section σ_{stim} (and hence the probability) for stimulated emission with the square of the lasing wavelength λ, i.e.,

$$\sigma_{\text{stim}} = \frac{nr_0 cf\lambda^2}{\Delta\lambda} \cdot \frac{g_l}{g_u} \tag{1}$$

where $r_0 = e^2/mc^2 = 2.8 \times 10^{13}$cm is the classical electron radius, $\Delta\lambda$ is the lasing-line width, f is the oscillator strength of the lasing transition, and g_u, g_l are the statistical weights of the upper and lower energy levels, respectively, for the laser transition. From the product of this cross section and the upper state density N_u is obtained the net gain coefficient $G = N_u\sigma_{\text{stim}} - N_l\sigma_{\text{abs}}$, where the latter product represents absorption on the lower laser state "l." Then, in the simplest case of a photon transversing an amplifying medium of length L, the net amplification achieved is given by $\exp(GL)$.

(b) The second difficulty is an inefficiency of reflecting mirrors for cavities, because vacuum-uv normal-incidence reflectivities decrease with shorter wavelengths and are generally less than 50%, compared to 99 + % for the visible and/or infrared spectral regions. This generally has meant that amplification must be achieved on a single pass through the inverted medium ("lasant") by the amplification of spontaneous emission (ASE), rather than through ~100 multiple passes in an efficient cavity. To compensate for the lack of a cavity, directionality (collimation) and some coherence are achieved by forming the medium as a thin filament. Furthermore, the density of amplifying atoms with inverted level populations must be scaled up accordingly in order to achieve a similar overall amplification.

(c) The third difficulty is the large power required to generate (pump) and maintain the required large population density of inverted atomic states. At a minimum it is necessary to compensate for the spontaneous radiation power P_r per unit volume V losses, i.e.,

$$\frac{P_r}{V} = N_u \sum_j A_{uj} \cdot \frac{hc}{\lambda_{uj}}, \tag{2}$$

where A_{uj} and λ_{uj} are the transition probabilities and wavelengths, respectively, for a transition from the upper-laser state u to an arbitrary lower state j over which the summation is performed. For single-pass gain conditions this loss alone can be as high as 10^{11} W/cm^3 for a laser wavelength of 10 nm. Furthermore, for fixed gain it can be shown that this scales approximately as λ^{-4} and hence reaches 10^{15} W/cm^3 at a wavelength of 1 nm, still in the "soft" X-ray region (see Fig. 1). Such high power densities in the tenuous medium required to transmit the X radiation automatically generates a high temperature and degree of multiple ionization and, hence, an overall-neutral ionized gas or "plasma" lasing medium.

(d) The final difficulty is the short pulse required. There were early concerns that X-ray lasing would necessarily have to operate on traditional deep-innershell transitions after the creation of, for example, K-shell vacancies. Even if population inversions were so created, these vacancies would be filled and the lasing action would "self-terminate" on a femtosecond (10^{-15} sec) time scale. This seemed impossible in the 1960s, even though lasers now exist in this range.

Hence, a considerable degree of early pessimism existed. However, as known in astrophysics and plasma physics, there are associated with high temperature plasmas multiply (but not completely) ionized atoms which closely mimic neutral atoms with the same number of

bound electrons. Such a group of ions is called an "isoelectronic sequence" and labeled H-like, He-like, Li-like, etc., for 1-, 2-, and 3-bound electrons, respectively. This in principle makes it possible to transfer atomic and ionic lasing at long wavelengths to vacuum-uv and shorter wavelengths with ions of increasing nuclear charge Z. The wavelengths of all transitions become increasingly shorter as Z increases along the sequence. For example, in the simplest case of the one-electron H-like sequence the wavelength scales as Z^{-2}, so that a transition from principal quantum number 3 to 2 at 656.2 nm in neutral hydrogen translates to 18.2 nm in five-times ionized C^{5+} ions. Conversely, a laser wavelength of 0.2 nm in the penetrating hard X-ray region could in principle be obtained on the same H-like 3-2 transition in a Ba ($Z = 56$) plasma. The scaling along other isoelectronic sequences is similar but more complicated because of the screening of additional bound electrons.

More importantly for X-ray lasers, it was realized that quasi-continuous-wave (cw) laser operation ("quasi" in the sense that lasing is limited only by the duration of the pumping impulse) could be achieved at short wavelengths using outershell "optical" transitions in such multiply ionized atoms in a high-temperature plasma medium. In this cw mode, typically lasting for tens of nanoseconds, the population inversion is maintained by rapid depletion of the lower-state density, usually accomplished through radiative decay to the ground state. The price paid is in efficiency lost in generating the highly ionized ions at the high densities required for short wavelength gain.

Indeed, success to date in generating ultrashort wavelength ASE has come about by using highly ionized, high-density plasmas as the lasing medium. A typical geometry is a thin (~100–200 μm diameter) filament of several centimeters in length. Such a plasma is created by line-focusing a high power infrared or visible laser onto a target of the material desired for the plasma. X-ray lasing is then observed along the plasma axis, transverse to the driving-laser beam direction. An absolute value for the gain is determined either by varying the length and measuring an exponential increase in emission or by comparing axial amplified emission with transverse spontaneous emission. These methods are particularly convenient for such a transverse-driven geometry, because both ends of the plasma are readily accessible.

An alternative method of lasant creation is to use a linear electrical discharge, either free or confined (e.g., a capillary), to create the plasma; and there are preliminary reports now arriving of some success in this method (see Table II and footnote *a* to Table II). Instabilities in such discharge plasmas have led to axial inhomogeneities and a less than satisfactory X-ray laser medium. Hence, most of the experimental gain results collected in Tables I and II are based on more uniform laser-produced plasmas as lasants.

II. PUMPING MODES AND EXPERIMENTAL RESULTS

As with any laser, it is necessary to "pump" electrons into an upper bound state of the lasing transition in order to establish therein a net population density exceeding that for the lower bound state, that is, a "population density inversion." Then stimulated emission from the former will exceed absorption from the latter as photons traverse the length of the plasma.

Self-contained in a high-temperature high-density plasma is the electron thermal energy per unit volume $N_e k T_e$ (N_e and T_e are the electron density and temperature, respectively). This intrinsic energy has been tapped very successfully for pumping X-ray laser transitions through inelastic collisions of free electrons with ions. Excitation, recombination, and (innershell) ionization are all possible electron-induced processes for effective pumping. The first two have proven to be most successful to date.

Electron-collisional excitation pumping of X-ray lasers, perhaps more than any other method, has evolved as an isoelectronic extrapolation of ion lasers from the ultraviolet spectral region (where efficient cavities exist). The pumping arises from the inelastic collision of a free electron with a ground-state ion in the plasma, resulting in a loss of free-electron energy and an excitation of a bound electron. The rate of pumping increases with electron temperature and hence is naturally compatable with the high temperature plasmas in which the ions exist. Indeed, $2p^{n-1}3p - 2p^{n-1}3s$ $\Delta n = 0$ lasing transitions ($n = 1$–6), pumped from a $2p^n$ ground state are abundant in the ultraviolet region and have proven to be widely successful for the soft X-ray region. This is shown both in Fig. 3 and Table I for $n = 6$ neonlike ions (and for $3d^{10}$ ground state nickel-like ions), where the levels are indicated in jj-coupling notation which is the most accurate for the medium atomic weight elements involved. Amplified spontaneous emission measurements have been published for selected ions extending from Ti^{12+} through Mo^{32+} in the soft X-ray spectral region (Table I). Agreement between theory and experiments for the wavelengths of the lasing transitions shown is very good. However, calculated gains for the $3p$ ($J = 0$) to $3s$ ($J = 1$) "A" lasing transitions continue to exceed those from experiments by factors of 2–3. In this regard, it has been suggested that the pumping may not be pure collisional excitation from the $2p^6$ ground state and perhaps includes some innershell ionization and recombination contributions.

TABLE I Wavelengths and Gain Coefficients for e-Collisional Pumping[x]

Ne-like ions $3p$-$3s$ transition:	A[a]	B[b]	C[c]	D[d]	E[e]	F[f]	Source[s]
Wavelength (nm):							
Ti^{12+}	32.65				508		LLE,RIKEN
Cu^{19+}	22.111	27.931	28.467				NRL
Zn^{20+}	21.217	26.232	26.723				NRL
Ga^{21+}		24.670	25.111				NRL
Ge^{22+}	19.606	23.224	23.626	24.732	28.646		NRL
As^{23+}		21.884	22.256				NRL
Se^{24+}	18.243	20.638	20.978	22.028	26.294	16.92	LLNL
	18.244	20.635	20.973				NRL
Sr^{28+}	15.98	16.41	16.65	17.51	22.49		LLNL
Y^{29+}		15.50	15.71	16.5	21.8		LLNL
Mo^{32+}	14.16	13.10	13.27	13.94		10.64	LLNL
Gain Coefficients (cm^{-1}):							
Ti^{12+}	2.7[r]						LLE,LLNL
Cu^{19+}	2.0	1.7	1.7				NRL
Zn^{20+}	2.3	2.0	2.0				NRL
Ge^{22+}	3.1	4.1	4.1	2.7	4.1		NRL
	1.0	2.1	2.2		2.8		LLNL
	1.5	2.5	2.5	1.0	1.8		LLE
	2.7	3.7	3.8	2.2	3.0		RAL
	2.3	3.3	3.4	2.3	3.2		SLL
As^{23+}			5.4				NRL
Se^{24+}	<1–2.7	4.0	3.8	2.3	3.5	≤1	LLNL
	2.6	4.9	4.9				NRL
Sr^{28+}		4.4	4.0				LLNL
Y^{29+}		4	4				LLNL
Mo^{32+}		4.1	4.2	2.9		2.2	LLNL

Ni-like ions: Transition:	G[g]	H[h]	I[i]	J[j]	
Wavelength (nm):					
Eu^{35+}	10.039	10.456	7.100	6.583	LLNL
Yb^{42+}	8.10[y]	8.440	5.609	5.026	LLNL
Ta^{45+}	7.442	7.747	5.097	4.483	LLNL
W^{46+}	7.240	7.535	4.924[y]	4.318	LLNL
Re^{47+}				4.157[y]	LLNL
Au^{51+}				3.56	LLNL
Gain Coefficients (cm^{-1}):					
Eu^{35+}	0.08	(−0.07)	1.11	1.1	LLNL
Yb^{42+}		(−1)	(−0.4)	2.2	LLNL
Ta^{45+}	(−1)	1.3	(−1)	3.0	LLNL
W^{46+}	(−0.6)	0.8		2.6	LLNL
Au^{5+}				2.0	LLNL

[a] $(2p^5_{1/2},3p_{1/2})_0$-$(2p^5_{1/2},3s_{1/2})_1$ [b] $(2p^5_{3/2},3p_{3/2})_2$-$(2p^5_{3/2},3s_{1/2})_1$
[c] $(2p^5_{1/2},3p_{3/2})_2$-$(2p^5_{1/2},3s_{1/2})_1$ [d] $(2p^5_{3/2},3p_{3/2})_1$-$(2p^5_{3/2},3s_{1/2})_1$
[e] $(2p^5_{3/2},3p_{1/2})_2$-$(2p^5_{3/2},3s_{1/2})_1$ [f] $(2p^5_{3/2},3p_{3/2})_0$-$(2p^5_{3/2},3s_{1/2})_1$
[g] $(3d^9_{5/2},4d_{5/2})_2$-$(3d^9_{5/2},4p_{3/2})_1$ [h] $(3d^9_{5/2},4d_{5/2})_1$-$(3d^9_{5/2},4p_{3/2})_1$
[i] $(3d^9_{3/2},4d_{3/2})_0$-$(3d^9_{5/2},4p_{3/2})_1$ [j] $(3d^9_{3/2},4d_{3/2})_0$-$(3d^9_{3/2},4p_{1/2})_1$
[r] Pumping may be enhanced by resonance absorption (see text).
[s] LLE: U. Rochester LLNL: Lawrence Livermore Nat' l Lab.
NRL: Naval Research Laboratory RAL: Rutherford Laboratory, U.K.
RIKEN: Inst. Phys. Chem. Res., Japan. SLL: Shanghai Laser Lab., PRC.
[x] All values are measured except: [y] Calculated wavelengths.

TABLE II Measured Wavelengths and Gains for Recombination Pumping

Species	Transition	Target[a]	Wavelength(nm)	Gain Coef(cm^{-1})	Source[b]
Hydrogenic:					
C^{5+}	3–2	Fi	18.2	25	HULL
		M		6.5	PPL
		Fo		3	NRL, SLL
		Fi		4.1	RAL
		Fo		2.0	ILE
		S		8	PPL
		C		1.2–1.7	RUHR
		Fo/C		2.0	ILE
O^{7+}	3–2	Fo	10.2	~0.5	LLNL
	3–2	Fo	10.2	2.3	ILE
F^{8+}	3–2	Fi	8.1	5.5	RAL
		Fi	8.1	4.4	RAL
		Fo		2.0	ILE
Na^{10+}	3–2	Fi,Fo,R	5.4	5.3	ILE
Mg^{11+}	3–2	R	4.5	1.5	ILE
Al^{12}_{+}	3–2	R	3.9	0.3	ILE
He-like:					
Al^{11+}	3d–2p	S	4.5	9.8	UTJ
	3s–2p	S	4.6	7	UTJ
	3p–3s	S	4.2	4.4	UTJ
Li-like:					
O^{5+}	4f–3d	M	5.20	1.8	PPL
Al^{10+}	5d–3p	Fo	10.4	1.0	ILE
	5d–3p	Fi	10.4	1.5	LSAI
	5f–3d	S	10.6	2–2.5	LSAI
		S		3.4	RIKEN
		S		3.1	SLL
		Fi		3	RAL
		Fo		1.5	ILE
		Fo		3.5	UMD
		Fi		1.5	LSAI
		S		2.2	LLE
	4d–3p	Fo	15.1	4.5	UMD
		Fi	15.1	1.4	LSAI
	4f–3d	M	15.5	3–4	PPL
		Fo		4.1	UMD
		Fi		2–3	RAL,LSAI
		S		4.5	RIKEN
		S		2.0	LLE
Si^{11+}	4f–3d	M,S	13.0	0.9–2	PPL,RIKEN
	5f–3d	S	8.9	1.5–1.7	SLL,RIKEN
	5d–3p	S	8.7	1.4	SLL
	6f–3d	S	7.6	1.7	SLL
	6d–3p	S	7.5	1.3	SLL
S^{13+}	5f–3d	S	6.5	1.3	LSAI
Cl^{14+}	4f–3d	Fi	8.0	3.5	RAL
Ti^{19+}	4f–3d	S	4.7	2.7	LLE
Be-like:					
Al^{9+}	5d–3p	S	12.3	3.4	RIKEN
	4f–3d	S	17.8	3.5	RIKEN
Si^{10+}	5f–3d	S	10.2	0.9	RIKEN
	4f–3d	S	14.8	0.6	RIKEN
Na-like:					
Cu^{18+}	6g–4f	S	7.2	2	SSL

[a] Targets: S, slab; M, magn. conf.; Fi, fiber; Fo, foil; R, ribbon; C, capillary.
[b] See Table I footnote s, and: HULL: Hull Univ., U.K.; ILE: Inst. Laser Engr., Osaka, Japan; LSAI: Lab Spect. Atom. Ion., France; PPL: Princeton Plasma Lab.; RUHR: Ruhr Univ., Bochum, Germany; UMD: Univ. of Maryland; UTJ: Univ. Tokyo, Japan.

Collisional-recombination pumping is also an inelastic free-electron collisional-induced process that, in contrast to collisional excitation, requires considerable "unnatural" plasma cooling to be effective. Sometimes referred to as "three-body recombination," the process involves a free electron interacting both with an ion and simultaneously with another free electron such that one electron is captured into a highly excited bound state in the atom. (This may be thought of as the inverse process of electron collisional ionization of a bound electron from an excited state.) From there it cascades downward in energy. For certain pairs of levels the lower bound state depopulates more rapidly than the higher state, so that there is a net population density inversion and gain. The rate for this pumping process is largest at high electron density ($\propto N_e^k k \approx 2$) and low electron temperature ($\propto T_e^{-2}$), so that again sudden and significant cooling is desired once the plasma is created at a high temperature and density. Such conditions often exist in very limited regions of parameter space for an expanding plasma. There have been numerous reports of measured gain associated with recombination pumping (see Table II and Fig. 3). In fact, the most "efficient" X-ray lasers produced to date are associated with recombination pumping, as discussed below in relation to Fig. 3. Numerical modeling tends to underestimate the magnitude of the gain compared to measurements, sometimes by as much as 100 times for certain transitions. Hence, there is a strong suggestion that the pumping is much less restrictive and more effective than explained by present recombination theory.

Pumping by photon collisions is also a possibility. Broadband flashlamps (most familiar from long-wavelength lasers) of sufficient power for pumping lasers in the X-ray spectral range do not exist at present. However, photoionization has been used for initiating the laser process for long wavelengths in the vacuum ultraviolet region (~100 nm, see "Auger" in Fig. 3). Here the removal of a tightly bound innershell electron by X-ray photoionization creates an ion sufficiently excited that Auger (or autoionization) processes can occur at a very rapid rate. When such processes more rapidly populate an upper laser state than lower, a population inversion and amplification are achieved. With the development of higher power drivers, photoionization pumping for the X-ray region may become a reality in the future. Also, very narrow-band photoexcitation pumping using the emission from an intense single spectral line of one ion of almost precisely the same photon energy as that of an absorbing excitation transition in a second lasant ion has been proposed as a form of flashlamp pumping. At least 112 such wavelength coincidences between spectral transitions have been identified for pumping X-ray lasers. To date, however, there is only one experiment in which there is evidence of such pumping of gain in the soft X-ray region, as noted in Table I, footnote *r*.

III. CHARACTERISTICS OF PLASMA X-RAY LASERS

The quality of the beam from a present ASE-type X-ray laser is considered poor by usual laser standards, again because of the lack of high quality cavities. It is hoped that this will improve with further development. Vital operating parameters, present values, and crystal-ball projections for the near future are discussed in the following paragraphs.

Wavelengths extending from the present soft X-ray range of 4–30 nm to, say, 0.1–2.5 nm will require pumping powers continuing to scale-up approximately as λ^{-4}, as previously discussed. Also, concerning *size*, much smaller diameters will be required at such shorter wavelengths in order to reduce radiative trapping on the transition connecting the lower laser level with the ground state. Smaller diameters also serve to reduce the number of transverse modes and hence increase the spatial coherence, as discussed below. The angular beam spread θ (the opposite of *collimation*) is determined for a single-pass ASE device by the length L and diameter d of the lasant plasma, i.e., $\theta = d/L \approx 10$ mrad at present. This is much larger than that for a diffraction limited perfectly coherent laser beam, where $\theta_d \approx 0.1$ mrad. The collimation can be expected to improve dramatically when efficient cavities are developed.

The *output power* will increase exponentially with length until saturation occurs, that is, until the stimulated emission rate competes with other rates and depletes the population inversion density. This occurs for $GL \gtrsim 14$. Largely because of expected monochromaticity, the *brightness* of the X-ray laser in units of photons $\cdot\,\mathrm{cm}^{-2} \cdot \mathrm{s}^{-1} \cdot \Delta\nu^{-1}$ is extremely high and advantageous for appropriate applications (ν is the photon frequency). This is illustrated in Fig. 2, where the a comparison is made to other popular pulsed X-ray sources. Decreasing the *pulse length* leads to higher power for a given energy output. For applications requiring maximum power, this is most desirable. However, for pulse durations less than $L/c \approx 300$ psec (or even longer for multiple passes through a cavity), traveling wave pumping synchronized with the travel of the gain packet through the medium has been demonstrated, with additional complexity.

An important measure of the quality of any laser beam is the degree of *coherence*. Of primary interest for lasers are the transverse and longitudinal spatial coherence parameters, which are more-or-less independent. These characteristics are related to the focusing and holographic

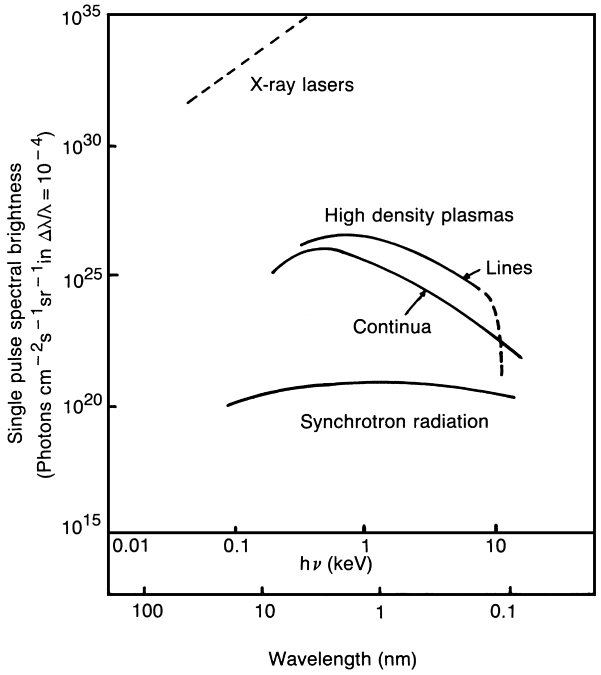


FIGURE 2 Comparison of absolute spectral brilliance of various xuv and X-ray sources. [Adapted from Key, M. H. (1988). *J. de Physique* **49,** C1–135.]

capabilities of the beam, as well as the spectral purity, or monochromaticity. The *transvers spatial coherence* across the X-ray laser beam is the extent D_{coh} over which a particular mode is coherent. At long wavelengths in more conventional lasers, a high degree of coherence and single transverse mode operation is achieved with high-quality cavities and other optics not yet available for the vacuum ultraviolet and X-ray regions. Hence, values for present ASE systems would be considered quite poor by normal laser standards. This coherence is given by $D_{\text{coh}} = L_{\text{v}}\theta_{\text{d}} = L_{\text{v}}\lambda/d$, where L_{v} is the viewing distance from the laser and $\theta_{\text{d}} = \lambda/d$ is the diffraction limited beam spread for a perfectly coherent beam of diameter d or for a particular mode in a multimode beam. Hence, if a beam diameter of $d = 30\ \mu$m can be created, the transverse coherence size at a viewing distance of 1 m is $D_{\text{coh}} = 300\ \mu$m, so that holograms of objects as large as 1 mm might be possible. However, with ~1000 transverse modes present, the intensity in a single mode will be greatly reduced, thereby puttin a great demand on the brightness delivered by the X-ray laser.

The Doppler broadened spectral *line width* in a high-temperature plasma is typically 3×10^{-4} times the wavelength. Narrower lines are an advantage in producing higher gain [Eq. (1)] as well as improved monochromaticity and temporal (and longitudinal-spatial) coherence (discussed in Section IV). Such a decreased bandwidth might be obtained in very cold laser media or by so-called gain narrowing near saturation. Directly related to the line width is the *temporal coherence*. This is the interval t_{coh} during the passage of a wavefront over which coherence is maintained and is given simply by the inverse of the spectral bandwidth $\Delta\nu$ in frequency units, i.e., $t_{\text{coh}} = (\Delta\nu)^{-1}$. It follows that the more monochoromatic the beam the longer the coherence time, as desired. As a plasma X-ray laser example, for a typical Doppler profile with $\Delta\nu/\nu \approx 3 \times 10^{-4}$ and for a laser wavelength of $\lambda = c/\nu = 10$ nm, the coherence interval is $t_{\text{coh}} \approx 0.1$ psec. This increases to ~0.3 psec for already reported narrowed lines of $\Delta\nu/\nu \approx 1 \times 10^{-4}$ and proiected to ~1 psec for even further narrowing. A temporal coherence of 0.3 psec translates into a *longitudinal spatial coherence* of axial distance $L_{\text{coh}} = ct_{\text{coh}} \approx 100\ \mu$m for the same conditions. This already is reasonable for holographic imaging of biological objects and microlithography, two specific applications discussed in Section V.

The overall *efficiency* is poor for present X-ray lasers, largely because of the power required to produce and sustain a significant population density in a high-temperature, high-density plasma. Unfortunately, there have been insufficient measurements of the output power from X-ray lasers to make meaningful comparisons of the output–input energy ratios and thereby overall efficiency. However, a figure-of-merit ratio GL/P achieved is obtainable from published data (see Table I), where P represents the pumping power used to create the plasma. This parameter is plotted in Fig. 3 and labeled for various dominant pumping processes. Many of the data are seen to fall near a somewhat arbitrary λ^2 line. Some data associated with recombination pumping show phenomenally large figures-of-merit. It must be said that none of the experiments giving these unusually high values has been duplicated to date, whereas many of the collisional excitation results have been reproduced independently in a number of laboratories. Nevertheless, these apparently anomolously high values for GL/P are very encouraging for future compact and efficient X-ray lasers.

IV. ALTERNATIVE (TO PLASMAS) AND COMPLIMENTARY APPROACHES

Besides direct amplification in a plasma medium, other novel X-ray laser approaches have been proposed. Most are still in the conceptual stage of development; some are complimentary to plasma amplifiers as master oscillators.

A prominent example of the latter is *harmonic generation* or frequency up-conversion. Here, a high-quality long wavelength laser beam is converted to the soft X-ray region through the nonlinear response of certain transmitting

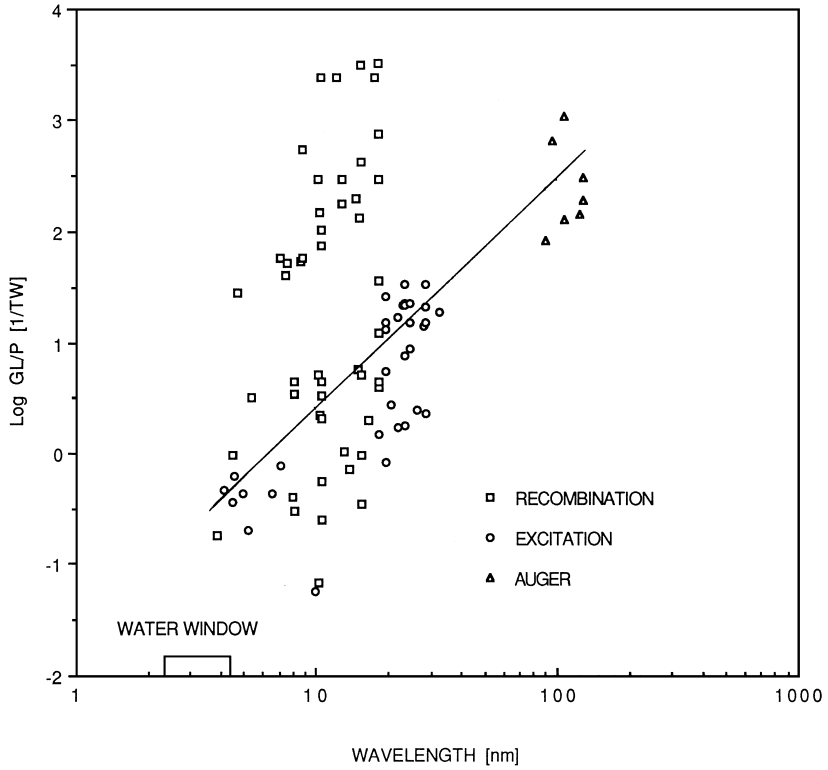


FIGURE 3 Survey of measured gain products (*GL*) divided by input power *P* versus wavelength for various pumping mechanisms. The biological water window is indicated. The λ^2 line is somewhat arbitrary. [Expanded from Key, M. H. (1988). *J. de Physique* **49,** C1–135.]

materials to electromagnetic waves. This is analogous to the nonlinear response of audio and microwave electronic circuit elements, which results in harmonic distortion, parametric amplification, modulation, and rectification. For vacuum-uv conversions, gases, vapors, and even plasmas are possible nondestructive media of high transmission. Typical higher order nonlinear processes used include odd-harmonic generation and the associated frequency mixing interactions, sometimes with resonant enhancement. Wavelengths throughout the vacuum-uv region reaching the 20–30 nm range have been generated. Anti-Stokes Raman scattering has also been successful at the longer vacuum-uv wavelengths. An advantage of this approach is that coherence and tunability at long wavelengths is preserved in the conversion. A disadvantage is the generally low efficiency for conversion ($\lesssim 10^{-5}$). Hence the proposed coupling to a plasma amplifier to increase the output power.

Free-electron lasers (FELs) are essentially an outgrowth and extension of modern synchrotron light sources. The amplification and radiation are achieved by a beam of free electrons forced into transverse oscillations by a spatially periodic wiggler magnetostatic field, thereby emitting magnetic bremsstrahlung radiation in the forward direction. They are sometimes advocated to combine the best of both broadband synchrotron sources and monochromatic, coherent lasers. Free-electron lasers share with synchrotron sources the technologies of low-divergence electron beams and periodic magnetic structures. The major advantages of FELs over conventional lasers is the broad and continuous (over a band) tunability offered, even with diffraction-limited beam quality.

The wavelength in the laboratory frame of reference is approximately proportional to the spatial period of the wiggler and inversely to the electron beam energy squared. Hence, for operation in the vacuum-uv spectral region a very short wiggler period and/or high beam energy is required. For the former, even substitution of an optical wiggler (perhaps laser) is proposed. This is projected to provide amplification at a wavelength near 0.5 nm with a 5 MeV electron beam. Plasma and solid-state wigglers are also under discussion. Once again, efficient cavities are required to achieve the level of operation expected from success with FELs in the infrared and visible spectral regions. In general, near-uv FELs are now a reality and vacuum-uv FELs are on the horizon, but operation in the X-ray region is still quite remote and faces many

similar problems associated with present plasma X-ray lasers.

For wavelengths shorter than 1 nm, nuclear *gamma-ray* emitting transitions in solids are usually considered. Gamma-ray laser ("graser") design revolves around self-terminating transitions between the ground and first excited isomeric states. It has been projected that output powers as large as 10^{21} W might be possible. Certain aspects of the energy storage, inversion, and gain mechanisms are considerably more challenging than for X-ray lasers.

The threshold density for a graser is proportional to the product of the line width and the lifetime of the excited isomeric state. Hence, if the line width can be reduced, the density and overall pumping power can be kept reasonable with the long lifetimes typical of forbidden isomeric transitions. Indeed, operation on very narrow lines may be possible for certain Mössbauer recoilless transitions. In this mode, pumping must be "gentle" in order not to overheat or vibrate the crystal. Because of these problems, excited isomers must first be prepared slowly and gently into extremely metastable storage states. The next step in pumping involves the actual segregation of the excited isomeric nuclei from others, followed by crystallization in a recoilless lattice of a cooled slender thread or whisker. From there lasing would be initiated by various schemes, depending on the upper state lifetimes. For example, in the very slow mode, isomers are pumped to a metastable state and lasing is initiated by a change in environment, for example, by rapid cooling to reduce the line width and increase the stimulated emission above the losses in the medium. In a faster mode, the broadening limits are within the range of Mössbauer narrowing. One pumping approach uses X-ray line emission to pump a short-lived excited intermediate state in a three-level system, either from a ground state or a nearby long-lived preformed excited isomeric state. This storage state then rapidly decays to fill the upper-laser level.

The challenge to developing grasers continues to be the contradictory requirements of an intense pump and a relatively undisturbed solid host material. Efforts are currently aimed at laying the groundwork for future experiments. The current consensus seems to be that basic studies of the structure and properties of nuclear excitation must precede a ture estimate of the feasibility of a gamma-ray laser, with a reasonable time scale for operation a couple of decades.

There is a continual need for new approaches for X-ray lasers. Several somewhat speculative ideas have been suggested. One category involves solid lasants in the form of crystals, with which both high densities and low temperatures are available for high gain. One major problem is absorption of the X-rays in the crystal. It has been suggested that certain diffraction effects may effectively offset some of this absorption. One proposed approach for lasing in a crystal involves photon pumping by the absorption of localized bremsstrahlung fluorescent radiation by a doped lasant substance. It is also proposed that the crystal lattice serve as the resonant cavity. Another crystal approach involves channeled radiation in a specific direction, which results from the interaction between the field in an aligned crystal and charged particles. The radiation from the oscillatory motion of the particles may be Doppler shifted from a few electron volts in photon energy to the kiloelectron volt or even megaelectron volt regions. It has been suggested also that X-ray lasing may be induced for channel radiation under certain circumstances. One approach considers the field as a very short-period undulator for a solid-state free-electron laser; the second involves more conventional transitions between channeling states.

Stimulating transitions of free electrons in the Coulomb field of a nucleus to make an X-ray laser would in principle result in tunability, because of the spread in energy for electrons in nonbound states. Transitions into other free-electron states results in *stimulated bremsstrahlung* emission. *Stimulated recombination* into bound states has also been proposed for pumping an X-ray laser. The central problem for achieving high gain in either case is a lack of sufficient free-electron density in a specific energy interval. These concepts are likely to become practical candidates for X-ray lasing only if a nonequilibrium situation can be generated at high density.

Another approach involves *Compton scattering*, in which photons injected into an oncoming relativistic electron beam are scattered back at greatly increased energies. The additional energy derives from the electron beam. Again, it may be possible to reach the X-ray spectral region starting with a long-wavelength laser. In addition, the output wavelength would be tunable by varying the acceleration of the electron beam. For 5-MeV electrons, 200-nm photons then emerge as 0.5-nm X-rays. Shorter wavelengths are conceivable with gigaelectron volt accelerators. It also has been proposed that the relativistic electron beam used in Compton scattering be replaced by a beam of multiply ionized atoms. The population inversion created by optical pumping could lase with tunability in the X-ray region because of the Doppler shift.

V. APPLICATIONS

While there has been continual "crystal ball" reflection on potential applications for X-ray lasers, the general attitude in the scientific community is that probably the most

important applications will develop after a useful device has arrived, depending upon the characteristics. With a brilliance of $\sim 10^8$-times that of state-of-the-art pulsed X-ray sources now known (see Fig. 2), such a device could not possibly remain unappreciated for any significant period. The most exciting applications most likely will be as revolutionary in the X-ray region as they have been with visible and infrared lasers. Numerous (nonprioritized) potential X-ray laser applications are discussed in the following paragraphs.

In the scientific arena, photoexcitation and photoionization are applicable to pumping other laser transitions, both for other wavelengths and for investigating atomic kinetics. Photoionization may be used also to generate multicharged ions for accelerator sources. Photoexcitation also can produce resonance-fluorescence pumping of atoms and ions into excited states for quantitative studies of further excitation and ionization by either photons or electrons. Resonance absorption and fluorescence can also be used for measuring ground-state population densities as well as excitation temperatures in atoms and plasma ions. In the surface science electron spectroscopy for chemical analysis (ESCA) technique, one or more core electrons is photoionized by X-rays or released by Auger emission, and the kinetic energy distribution is analyzed to determine the elemental compostion and chemical state of the surface of a specimen. Also, nuclear processes can be influenced by X-ray lasers through the surrounding electrons. The responsible processes are internal conversion and K-electron capture. In addtion, the impact of high-energy radiation on matter eventually leads to chemical changes. This field involves in part photolysis as well as radiolysis on very short time scales. Finally, miniature electron accelerators have been conceived as a possible centimeter-long version of the two-mile Stanford linear accelerator.

Moving from pure science to technical applications, the formation of diffractive gratings with sub-100 nm spacings may be possible using X-ray lasers. Also, high-resolution ($<1\ \mu$m) photolithography is one of the more promising applications. X-rays are preferred to uv radiation to reduce diffraction, scattering, and reflection effects. Line-widths on the order of 20 nm have been reported using conventional X-ray sources with exposures of up to hours. Microelectronic devices appropriate for X-ray lithography include integrated circuits, diffraction gratings, magnetic bubbles, transistors, surface acoustic wave devices, gratings for distributed-feedback lasers, and Josephson junction cryogenic devices.

In surface science, photoelectron spectroscopy is a versatile probing technique for studying intrinsic and absorbate surface states in solids. Photons are used to release electrons from the surface, and their energy distribution is measured with an electron energy analyzer. The intensity and monochromaticity of an X-ray laser are decided advantages here. X-ray lasers in the 1 nm range also will be useful in metallurgy for ion implantation, surface alloying, laser annealing, wear, friction and corrosion detection, mirror surface evaluation, and phase transformation.

For biology, it is desirable to study cellular organelles with high spatial and millisecond-to-picosecond temporal resolutions in a natural and dynamic fluid state. Some vital questions that can be addressed are the locations of soluble enzymes in the cytoplasm and the presence of protein-associated aggregated structures. Since proteins in a live biological sample are more absorbing than the water present for wavelengths shorter than 4.4 nm (the K-absorption edge for carbon) and longer than the corresponding K-edge for oxygen at 2.3 nm wavelength, this particularly useful range has been dubbed the "water window" of high contrast (noted in Fig. 3). This has served to set one of the most important goals in terms of short wavelengths. Some gain in this wavelength region has already been achieved (see Tables I and II and Fig. 3), but the output power is still quite low for applications. With a high brightness, single-mode spatially coherent X-ray laser, biological holography becomes possible, thereby providing time-resolved three-dimensional images of the live cell. Also, three dimensional images of macromolecules and atoms may become a reality. In the field of biomolecular crystallography, micrometer scale imaging with an X-ray laser could immensely simplify the task of growing crystals free of irregularities. With added coherence it also may be possible to use X-ray scattering to determine the electron distribution in a crystal.

In the radiological medical field, some angiographic applications using X-ray and gamma-ray lasers are suggested. Differential absorption can be used to ascertain the distribution of heavy elements in organs *in vivo* with improved contrast, in a situation in which a pathological condition has led to a change in elemental composition. Some familiar examples are iodine in the thyroid gland and osteoporosity in bones. Holographic X-ray imaging could have tremendous clinical potential for the diagnosis of abnormal anatomical situations caused by congenital abnormalities or disease, such as the location and analysis of tumor masses.

VI. SUMMARY AND PROGNOSIS

X-ray laser research and development has advanced to the success that it enjoys today through a series of spurts of interest and progress, with a periodicity of $\sim$3–5 years, beginning around 1970. Some of the periodically renewed interests have arisen with the availability of new and more powerful drivers; some have followed a breakthrough in

thinking and progressive experiments. Mostly, however, they have resulted from changes in policies and emphasis at various levels.

Now that lasers exist in the xuv and soft X-ray spectral regions of useful output power, the next logical step is to "institutionalize" this area of research, rather than continuing it as an adjunct to laser fusion and pulsed power programs. Fully dedicated, worldwide cooperative facilities are needed. Likely research areas of fruitful pursuit include shorter wavelengths, increased output, cavity capability, improved beam quality, reduced size, enhanced efficiency, and improved convenience and accessibility. We are on the threshold of acceptance of X-ray lasers as practical and useful tools. Clearly the promise is exciting for X-ray laser research, development, and application in the coming years. The truly revolutionary capabilities and applications are limited only by our imaginations and resources.

SEE ALSO THE FOLLOWING ARTICLES

HOLOGRAPHY • LASERS • LASERS, FREE-ELECTRON • X-RAY ANALYSIS • X-RAY PHOTOELECTRON SPECTROSCOPY

BIBLIOGRAPHY

Baldwin, G. C., Solem, J. C., and Gol'danski, V. I. (1981). *Rev. Mod. Phys*. **53,** 687.

Brau, C. (1990). "Free Electron Lasers," Academic Press, San Diego.

Elton, R. C. (1990). "X-Ray Lasers," Academic Press, San Diego.

J. Quant. Spectros. Rad. Transfer (1988) **40,** n. 40.

Marshall, T. C. (1985). "Free-Electron Lasers," Macmillan, New York.

Reintjes, J. F. (1985). Coherent ultraviolet and vacuum ultraviolet sources. In "Laser Handbook," (M. Bass and M. L. Stitch, eds.). North-Holland, New York.

Skinner, C. H. (1991). *Phys. Fluids* **B3,** 2420.

Lasers, Solid-State

Stephen A. Payne
Georg F. Albrecht
Lawrence Livermore National Laboratory

I. Introduction
II. Solid-State Laser Materials
III. Solid-State Laser Architectures
IV. Future Directions

GLOSSARY

Active ion The ion whose electronic transitions enable storing of the pump energy as an inversion and subsequently converting this inversion to laser light.
Active medium The material, that contains the active ion as a dopant ion. Dopant ions typically constitute from 10^{-1} to 10^{-4} of the material atoms.
Energy storage Capability of a doped laser material to store energy as inversion until it is converted to laser light.
Flash lamps High-current plasma discharges contained in a fused silica envelope whose light output pumps a laser and thus creates the inversion in the active ions.
Fusion driver Laser system of sufficient peak power to initiate nuclear fusion in the laboratory.
Host material The crystal or glass, that is host to the ion whose energy transitions make the lasing possible.
Inversion The ion's "readiness" to produce laser light by keeping the active ion in a state that is de-excited through amplification of laser light. Only few ions have the necessary structure of electronic transitions to produce this effect in practice.
Laser transition Electronic structure change in the state of an active ion that gives rise to the emission of a photon at the lasing wavelength. It is that same transition that is inverted upon pumping of the active medium.
Optical switch Device used to control the pulse format generated by a laser by allowing electronic control of the passage or blockage of light.
Optical fiber Optical medium, usually made of SiO_2, which guides light in the core region of 3–100 μm.
Pumping Process by which the active ions are storing energy in an inversion. The goal is to pump hard enough to drive a laser medium to, or above, the threshold inversion density at which lasing will begin.
Sensitizer ion Impurity ion used to enhance the absorption of pump light by the laser medium, which then transfers this energy to the active ions.
Tunable laser Laser for which the output wavelength may be tuned to different values.

A SOLID-STATE LASER is a device in which the active medium is based on a solid material. While this material can be either an insulator or a semiconductor, semiconductor lasers will be covered elsewhere in this volume. Solid-state lasers based on insulators include both materials doped with, or stoichiometric in, the laser ions and materials that contain intrinsic defect laser species,

known as F-centers. F-center lasers, however, will be considered elsewhere in this volume as well. Thus, the designation, solid-state lasers, is intended to denote those laser systems that are based on crystalline or glassy insulating media, in which a stoichiometric component, or extrinsic dopants, incorporated into the material serve as the laser species.

The physics and engineering of solid-state lasers is both a mature field and an area burgeoning with new activity. While there are many concepts and laser designs that have been established, each year continues to bring remarkable discoveries that open new avenues of research. In this article, it is not only intended to provide an accounting of the known physics of solid-state laser sources, but also to convey a sense of the enormity of the field and the likelihood that many new laser materials and architectures will be discovered during the next decade.

I. INTRODUCTION

The principle of laser action was first demonstrated in 1960 by Maiman. This first system was a solid-state laser; a ruby crystal served as the active element, and it was pumped with a flash lamp. With this report of laser action, the main concepts upon which solid-state lasers are based became established (see Fig. 1). The idea of optically pumping the laser rod was realized, as well as the use of an impurity-doped solid as the laser medium. Lastly, the concept of a laser resonator, as adapted from the work of Townes and Schalow, was experimentally demonstrated. Much of this article is essentially an exposition of the extensive technical progress that has occurred in each of these three areas. Optical pumping has evolved considerably, by way of optimization of flash lamps, and through the additional use of laser-pumping techniques. The number of impurity-doped solids that have now been lased stands at over 200. Optical resonators have also become remarkably sophisticated in terms of the manipulation of the spatial, temporal, and spectral properties of the laser beam.

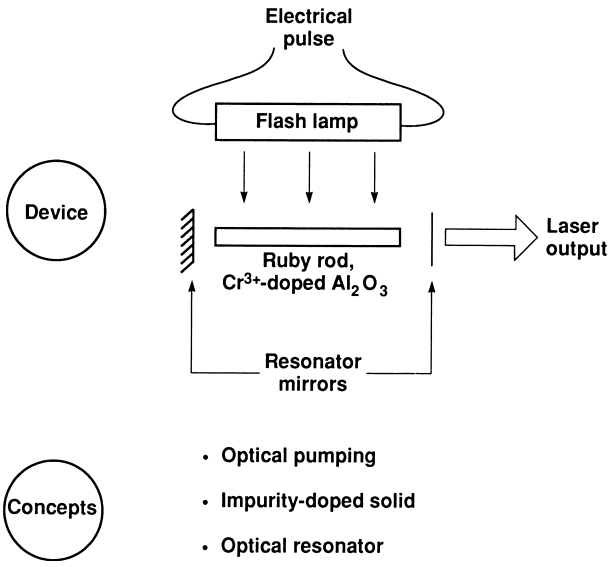


FIGURE 1 Schematic description of the ruby system for which laser action was first reported, and a listing of the fundamental concepts that were introduced.

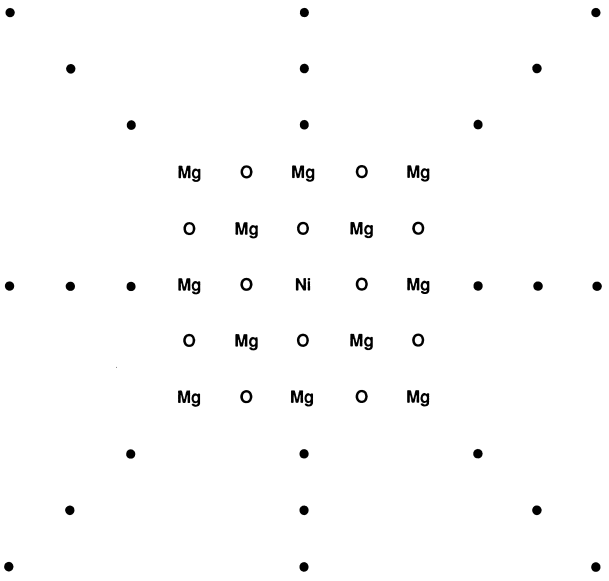


FIGURE 2 Example of an impurity-doped host, MgO:Ni^{2+}.

II. SOLID-STATE LASER MATERIALS

A. Laser-Active Ions in Solids

Solid-state lasers are based on a wide variety of materials. All of these materials are conceptually similar, however, in that a laser-active impurity ion is incorporated into the solid material, referred to as the host. In nearly all cases of interest to us, the host is an ionic solid (e.g., MgO), and the impurity carries a positive charge (e.g., Ni^{2+}). As a simple illustration of this situation, a two-dimensional view of the MgO:Ni_2^+ system is pictured in Fig. 2. Here, a small fraction of the Mg^{2+} sites are substituted by Ni^{2+} ions. While the pure MgO crystal is clear, the NiO doping leads to green coloration. It is the impurity ions that are responsible for the laser action. The host medium nevertheless profoundly affects the electronic structure of the impurity and is, of course, responsible for the bulk optical, thermal, and mechanical properties of the laser medium.

1. Ions in Crystals and Glasses

On the basis of the introduction, and a cursory glance at the periodic table of the elements, it may seem that

TABLE I Laser Ions and Abbreviated Listing of Host Materials

Laser ions
Transition metal ions: Ti^{3+}, V^{2+}, Cr^{2+}, Cr^{3+}, Cr^{4+}, Fe^{2+}, Mn^{5+}, Co^{2+}, Ni^{2+}
Trivalent rare earth ions: Ce^{3+}, Pr^{3+}, Nd^{3+}, Pm^{3+}, Sm^{3+}, Eu^{3+}, Gd^{3+}, Tb^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, Tm^{3+}, Yb^{3+}
Divalent rare earth ions: Sm^{2+}, Dy^{2+}, Tm^{2+}
Actinide ion: U^{3+}
Examples of laser hosts
Fluoride crystals: MgF_2, CaF_2, LaF_3, $LiYF_4$, $LiSrAlF_6$, BaY_2F_8
Oxide crystals: MgO, Al_2O_3, $BeAl_2O_4$, Mg_2SiO_4, $YAlO_3$, $CaWO_4$, YVO_4, $Y_3Al_5O_{12}$, $Gd_3Sc_2Ga_3O_{12}$, $LiNdP_4O_{12}$, $Sr_5(PO_4)_3F$
Chalcogenide crystals: ZnS, ZnSe
Glasses: ZrF_4-BaF_2-LaF_3-AlF_3, SiO_2-Li_2O-CaO-Al_2O_3, P_2O_5-K_2O-BaO-Al_2O_3, SiO_2, La-Ga-S

the number of potential impurity-host systems is virtually limitless. The reported laser ions are indicated in Table I, along with a very abbreviated list of host crystals. While 22 laser ions are reported in Table I, it should be noted that these ions lase with varying degrees of proficiency and generally possess both advantages and disadvantages. For example, Ti^{3+} has a wide tuning range but lacks the ability to store energy; Pm^{3+} is predicted to lase quite efficiently but is radioactive; Sm^{2+} lacks good chemical stability, and so on.

Each of the laser host materials possesses certain attributes and handicaps. The ease of crystal growth is important, as is the meltability of glassy materials. It is also noted that only certain impurity ions are compatible with a particular host. Generally, the size and charge of the substitutional host metal ion must be similar to that of the impurity ion. For example, Ni^{2+} can be incorporated into the Mg^{3+} sites of MgF_2 and MgO, and Nd^{3+} onto the Y^{3+} sites of $LiYF_4$, YVO_4, $YAlO_3$ and $Y_3Al_5O_{12}$.

2. Energy Levels

The nature of the energy levels and dynamics of the impurity-host system determine the character and effectiveness of the laser. A generic representation of the impurity energy levels and the energy flow appears in Fig. 3. An ideal laser crystal or glass would efficiently absorb the light from the pump source (Step 1). The energy then relaxes to the lowest excited state (Step 2). This level typically has a lifetime that is long enough to "store" energy. Gain occurs in the next step, as the impurity undergoes a transition between the upper and lower laser levels. While there are many potential loss mechanisms in a laser material, two that are fundamental in nature are noted in Fig. 3. One is excited-state absorption, where the upper level "sees" some amount of loss, rather than gain, due to the excitation to a higher lying excited state. A second loss involves the degradation of the energy of the upper laser level into heat, resulting in nonemissive loss. Step 4 shows the lower laser level relaxing back to the ground state. This last step is critically important since a system in which this process is absent (known as three-level lasers), requires considerably more energy in order to lase. The advantage of the four-level scheme depicted in Fig. 3 is that the lower laser level is unoccupied and therefore cannot absorb the laser light (thereby introducing ground state absorption loss into the system).

B. Host Materials

The host materials that are utilized in laser systems must exhibit adequate transparency, mechanical strength, and thermal properties. In addition, the material must be able to sustain a precise optical polish, and be cast or grown adequately within reasonable economic and time constraints. The host must afford the impurity ions the type of spectroscopic properties that are appropriate for good laser performance. As a result of the numerous requirements, not many materials turn out to be useful in practical circumstances. Below, the nature of glasses and crystals is discussed, and the important physical properties are briefly outlined.

1. Glasses

Most laser glasses fall into one of several categories: silicates, phosphates, fluorides, and chalcogenides. These glasses may also be mixed, yielding fluorophosphates, silicophosphates, and so on. In all cases, the glass is imagined to consist of two components: the network former and the modifiers. The network is a covalently bonded

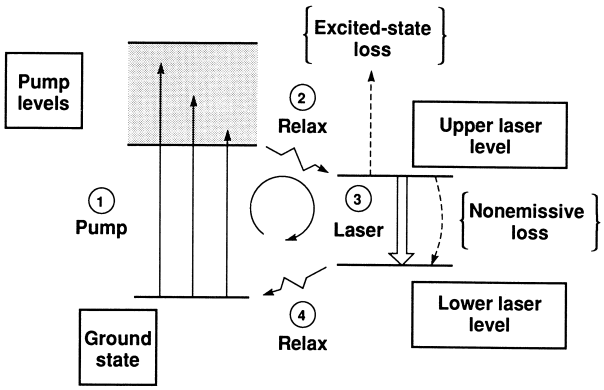


FIGURE 3 Generic representation of the energy levels and energy flow (steps 1–4) for an idealized impurity laser ion. The dashed arrows illustrate two fundamental loss mechanisms.

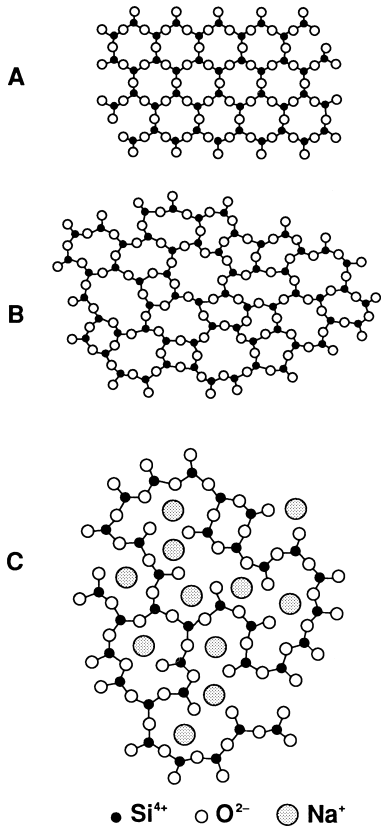


FIGURE 4 Two-dimensional view of the structure of SiO_4 tetrahedrons in (a) quartz crystal, (b) fused silica, and (c) sodium silicate glass.

three-dimensional system, whereas the modifiers are ionically bonded and are imagined to disrupt the network structure. The silicate glasses provide the simplest description of the interplay between the network and the modifiers. First, consider crystalline quartz, or SiO_2, as illustrated in Fig. 4. Here, every oxygen bridges between two silicons. Fused silica is similar, although it is glassy, meaning that the highly ordered nature of the system has been eliminated. If modifiers such as Na_2O are added, some of the oxygens become "nonbridging." There are several favorable features afforded to the glass by the modifier ions: the melt acquires a much lower viscosity and may be easily poured and cast, and the glass is able to dissolve rare earth ions much more effectively compared to fused silica.

A similar situation exists for other types of glasses, as well. For example, the P_2O_5 in phosphate glasses forms the network, and alkali and alkaline earth oxide compounds are added as modifiers. For the case of fluoride glasses, ZrF_4, ThF_4, or BeF_2 may serve as the network former, as does Ga-S for the chalcogenide La-Ga-S mixture.

2. Crystals

The growth of most crystals turns out to be considerably more difficult than is the case for melting and casting glassy materials. Crystals provide important advantages, however, since a precisely defined site is available to the laser ion, rather than the broad distribution of sites that characterize a glass. Crystals often have more favorable thermal and mechanical properties as well. For example, the thermal conductivity tends to be much higher, and oxide crystals tend to be very strong mechanically compared to glasses. As a result, it is often advantageous to produce the crystalline media.

Crystals may be grown many different ways, two examples of which are shown in Fig 5. The Bridgman method typically involves slowly lowering a crucible through a zone in which the temperature abruptly drops from above to below the melting point of the crystal. A seed crystal is sometimes placed at the bottom of the crucible to initiate the growth. Also shown is the Czochralski method,

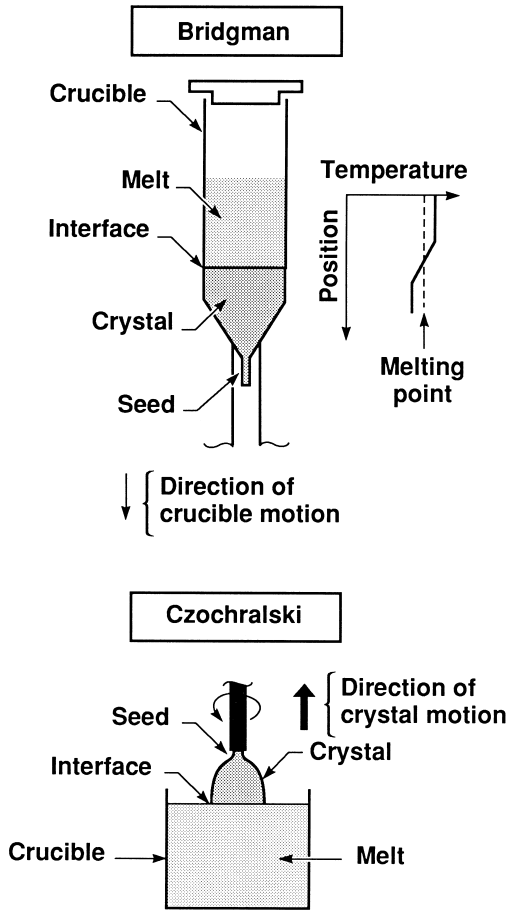


FIGURE 5 Schematic drawing of two common crystal growth methods.

TABLE II Summary of Some Thermal, Mechanical, and Optical Properties for Several Laser Hosts

Material	Thermal		Mechanical		Optical	
	Melting (softening) points, T (°C)	Thermal conductivity, κ (W/m°C)	Young's modulus E (GPa)	Fracture toughness, K_{IC} (MPa/m$^{1/2}$)	Index change, dn/dT (10^{-6}/°C)	Nonlinear index, n_2 (10^{-13} esu)
Glasses: ED-2 (Silicate)	(590)	1.36	92	0.83	+3.8	1.4
LG-750 (Phosphate)	(545)	0.62	52	0.45	−5.1	1.1
SiO_2 (fused silica)	(1585)	1.38	72	0.75	+11.8	0.9
Crystals: Al_2O_3 (sapphire)	2040	34	405	2.2	+1.6	1.3
$Y_3Al_5O_{12}$ (YAG)	1970	13	282	1.4	+8.9	2.7
CaF_2 (fluorite)	1360	9.7	110	0.27	−11.5	0.4
$LiYF_4$ (YLF)	820	5.8, 7.2	75	0.31	−2.0, −4.3	0.6
ZnSe	1525	18	67	0.2	+70	170

in which a seed is dipped into the melt, and then slowly raised as it is rotated. Most crystals are grown at the rate of 0.1–10 mm/hr. It is important to note that there are many other methods of crystal growth that have not been discussed here (solution and flux growth, flame-fusion, etc.). All methods are based on the concept of slowly enlarging the seed crystal.

3. Physical Properties

The physical properties of the host material, in part, stipulate the architecture of the laser system. The properties of several types of host materials are outlined in Table II, including glasses and crystals. The qualitative trends evidenced in the table can be enumerated as follows: (a) The high-melting oxides tend to provide the highest mechanical strengths, as noted by the magnitudes of Young's modulus (the "stiffness") and the fracture toughness (a measure of the material's resistance to breaking). (b) The thermal conductivity of crystals is much higher than glasses. This permits the rapid cooling of the laser material when deployed in a system. (c) The temperature difference between the surface and center of a laser material gives rise to a thermal lens. The magnitude and sign of the change of refractive index with temperature, dn/dT, partially determines the extent of thermal lensing, which is typically smallest for fluorides, larger for oxides, and largest for ZnSe in Table II. (d) The nonlinear refractive index, n_2, is smallest for fluorides. The value of n_2 determines the power at which self-focusing occurs. In this process the power of the beam itself creates a lens in the material, which potentially may result in catastrophic self-focusing of the light.

C. Rare-Earth Ion Lasers

As mentioned above, thirteen of the rare-earth (RE) ions have been lased. In passing from Ce^{3+} to Y^{3+}, the 4f shell becomes filled with electrons, $4f^1$ to $4f^{13}$. It is the states that arise from the $4f^n$ shell that give rise to nearly all of the RE laser transitions. In this section a few specific systems will be discussed in detail, in order to provide representative examples of the nature of RE lasers.

1. Nd^{3+}-Lasers

All of the relevant absorption and emission features of Nd^{3+} are due to $4f \rightarrow 4f$ transitions. Dozens of electronic states arise from the $^4f^3$ electronic configuration of Nd^{3+}, many of which are indicated in Fig. 6. The ground states

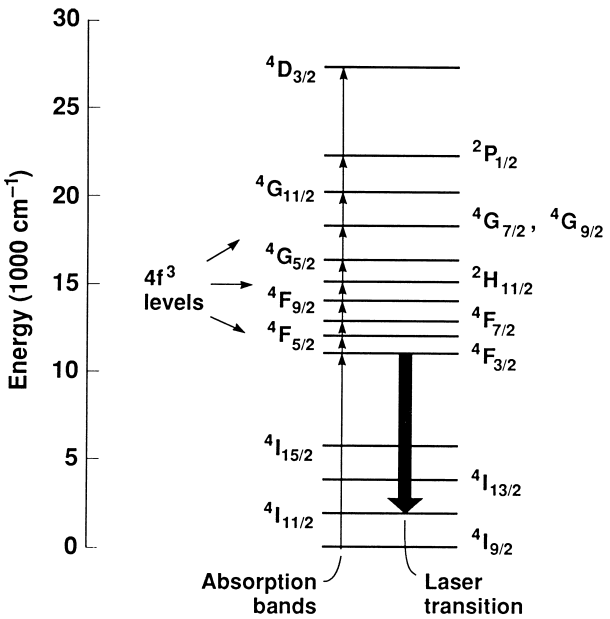


FIGURE 6 Energy levels of Nd^{3+} showing the transitions responsible for the absorption bands and the laser action. Note that only the major states involved in the transitions have been indicated for clarity of presentation.

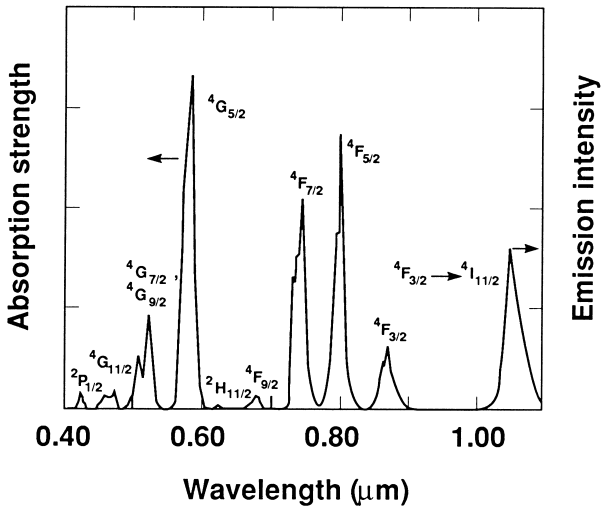


FIGURE 7 Absorption and emission spectra of Nd^{3+} in phosphate glass (LG-750). The final states of the absorption transitions are indicated.

consist of the ^4I_J manifold, where the ^4I designation describes the spin ($S = 3/2$) and orbital ($L = 6$) angular momenta, and the $J = 9/2, 11/2, 13/2, 15/2$ indicate the coupling between these two momenta. Absorption transitions occur from the $^4I_{9/2}$ ground state to the indicated energy levels, and each of these transitions are manifested in the absorption spectrum. Figure 7 contains the absorption spectrum of Nd^{3+} in a phosphate glass, known as LG-750. The appearance of the spectrum for Nd^{3+} is similar in any host, since it is primarily due to the free-ion properties of Nd^{3+}. The main effect of the host glass (crystal) is to split and broaden the atomic transitions into bands (lines).

Following excitation into any of the absorption bands shown in Figs. 6 and 7, the energy rapidly relaxes to the metastable $^4F_{3/2}$ excited state, and generates significant heat during this process. The $^4F_{3/2}$ state, on the other hand, decays in a radiative manner to all the states of the ^4I manifold. The $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition produces an emission band that peaks at 1054 nm (see Figs. 6 and 7). It is this transition that provides the strongest gain for Nd^{3+}-doped laser glasses and crystals. Nd^{3+} lasers are by far the most technologically important materials. The three main materials that are routinely utilized with Nd^{3+} are $Y_3Al_5O_{12}$, $LiYF_4$ (known as Nd:YAG and Nd:YLF), and several kinds of glass.

2. Er^{3+}-Lasers

The Er^{3+} ion has been lased in many crystals and glasses, and this ion has, in fact, been reported to lase on 13 different transitions. The relevant energy levels and the wavelengths of the laser transitions are shown in Fig. 8. Of all of the laser lines, however, only the $^4I_{13/2} \rightarrow {}^4I_{15/2}$ transition is utilized in commercial systems. The most significant technology here is the use of diode-pumped Er:SiO_2 fibers as the means of amplifying the 1.55-μm signal within the telecommunications network. The Er^{3+} laser ion illustrates the wide gap between number of demonstrated laser transitions and those that are suitable for technological application. This disparity may be taken as an indication of the progress in laser physics and design that is likely to occur in the coming years.

One question that may come to the reader's mind is why all of the energy relaxes to the $^4F_{3/2}$ metastable level for Nd^{3+} irrespective of which state is excited, while so many of the higher-lying states of Er^{3+} are found to emit sufficiently to allow for laser action. The answer to this question lies in the "energy gap law." Simply put, it states that the rate of *non* emissive decay between two energy levels of a RE ion is dependent on the energy gap between them and the highest phonon frequency of the host medium. Since the energy gaps are generally larger for Er^{3+} than for Nd^{3+} (compare Figs. 8 and 6), more of the energy levels are able to emit in the case of Er^{3+}, and therefore may potentially be able to lase.

3. Yb^{3+} -Lasers

With the advent of high-power laser diodes operating in the 0.9–1.0-μm region, Yb^{3+}-lasers have become a

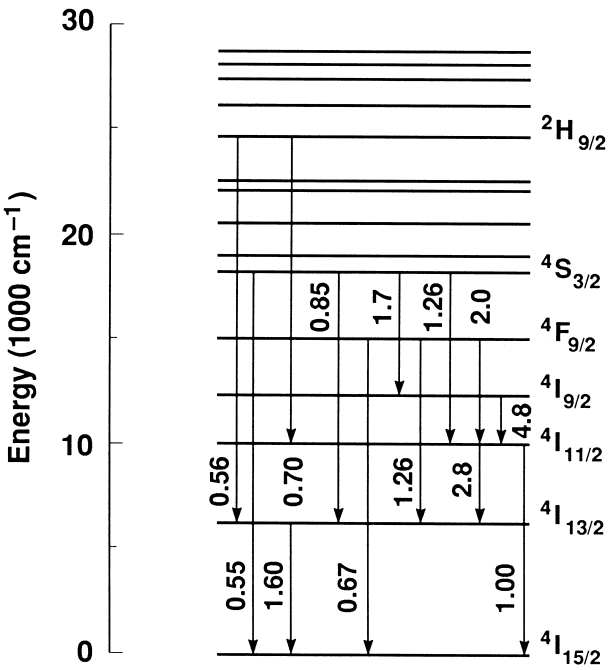


FIGURE 8 Energy levels and demonstrated laser transitions of Er^{3+}-doped materials. The wavelengths of the laser action are indicated in micrometers.

practical reality. Diode-pumped Yb:YAG lasers have been reported to yield up to a kilowatt of power, partly owing to the low heat load of the Yb^{3+} ion in YAG (pumped at 0.94 μm, lases at 1.03 μm, 9% of pump light is converted to heat). Yb:SiO_2 in the form of fiber lasers is having a tremendous impact in the marketplace, especially in the so-called cladding-pumped configuration, where separate pump and lasing cores are defined and overlapped. Diode-pumped Yb:fiber lasers yielding several tens of watts are being used for marking and materials processing. Another rapidly emerging Yb-doped material, Yb:S-FAP or Yb-doped $Sr_5(PO)_3F$, is being developed as a high-energy diode-pumped solid-state laser at Lawrence Livermore National Laboratory; the nominal specifications of the laser are 100 J–10 Hz–1 kW–2 nsec.

4. Other Rare-Earth Lasers

As noted above, there is a great disparity between the number of RE ions that have been lased and the relatively small number of materials that are actually used in practical applications. Nevertheless, great potential exists for the future. A summary of all the wavelengths that have been generated by RE ions appears in Fig. 9. (Note that the particular transitions involved have not been indicated on the figure.) The wavelengths span the range from 0.18 μm to >7 μm, and both divalent and trivalent RE ions are represented. The efficiencies of many of these systems are extremely low, however, and may never prove to be useful in practical circumstances.

D. Transition Metal Ion Lasers

1. Ti^{3+}-Lasers

The optical properties of transition metal ions are fundamentally different from the rare-earth species. This is the case primarily because the 3d $\rightarrow$3d electronic transitions that are responsible for the absorption and emission features of transition metal ions interact strongly with the host, in contrast to the relative insensitivity of the rare-earth 4f $\rightarrow$ 4f transitions. The type of situation that arises is depicted in Fig. 10 in terms of a "configuration coordinate" model. The Ti^{3+} ion in Al_2O_3 has been selected for illustrative purposes because its valence shell contains only a single d electron, 3d. The 3d electron is split into two states, the ^2E and $^2T_2$, by the six nearest-neighbor oxygen anions surrounding the Ti^{3+} ion. (The splitting of 3d orbitals is much larger than that experienced by the 4f electrons of rare earth ions.) As depicted in Fig. 10, the average Ti-O distance is slightly larger in the ^2E state, compared to the $^2T_2$ state. This point is particularly important, because it produces wide absorption and emission features.

The configuration coordinate diagram of Fig. 10 explains how the different Ti-O distances in the ground and excited states give rise to broad spectral features. The Gaussian curve drawn on the ground state potential energy surface indicates the probabilistic distribution of Ti-O distances that occurs. Since the electronic transition to the excited state occurs rapidly compared to the motion of the Ti-O atoms, this ground state Ti-O distance distribution is simply "reflected" off the rising side of the upper state energy surface, thereby producing a broad absorption feature. A similar argument applies to the emission process.

The actual absorption and emission spectra of Al_2O_3:Ti^{3+} are shown in Fig. 11. This material (known as Ti:sapphire) may be optically pumped with a doubled Nd:YAG laser, an Ar^+ laser, or a flash lamp. The output of the laser can be tuned from 0.66 to 1.2 μm. Ti:sapphire operates efficiently and is not adversely impacted by the detrimental loss mechanisms indicated in Fig. 3. Ti:sapphire is the material of choice for generating ultrashort pulses of light in the femtosecond regime.

2. Cr^{3+}-Lasers

Cr^{3+}-lasers are similar to Ti^{3+}-lasers, in that these crystals exhibit broad spectral features. Cr^{3+}-doped systems possess an important advantage, however, in that they have three absorption bands, rather than one, and therefore absorb flash lamp light more efficiently. Furthermore, because the trivalent oxidation state is very stable, Cr^{3+} may be incorporated into a wide variety of hosts. A summary of the tuning ranges achieved by Cr^{3+}-doped materials appears in Fig. 12, where it is seen that wavelengths from .70 μm to 1.25 μm can be covered with different host materials. It is crucial to emphasize, however, that many of the reported Cr^{3+}-lasers are flawed in various ways, such as by having low efficiency, or perhaps by permanently coloring under the influence of ultraviolet flash lamp light. The two most common lasers are Cr^{3+}-doped $BeAl_2O_4$ and $LiSrAlF_6$ (known as alexandrite and Cr:LiSAF). Diode-pumped Cr:LiSAF lasers are often utilized in the short-pulse mode, providing a compact architecture in the form of a mode-locked oscillator.

Another important distinction between Ti^{3+} and Cr^{3+} lasers pertains to the lifetime of the metastable excited state, which is typically near 1–10 μs for Ti^{3+}, and 100–300 μs for Cr^{3+}. As a result, Cr^{3+} lasers can be arranged to "store" more energy than Ti^{3+} lasers.

3. Other Transition Metal Ion Lasers

There are, in total, 30 transition metal ions, including the first, second, and third row ions. In addition, several oxidation states exist for each of these ions. In spite of this diversity, only four other ions have been lased in addition

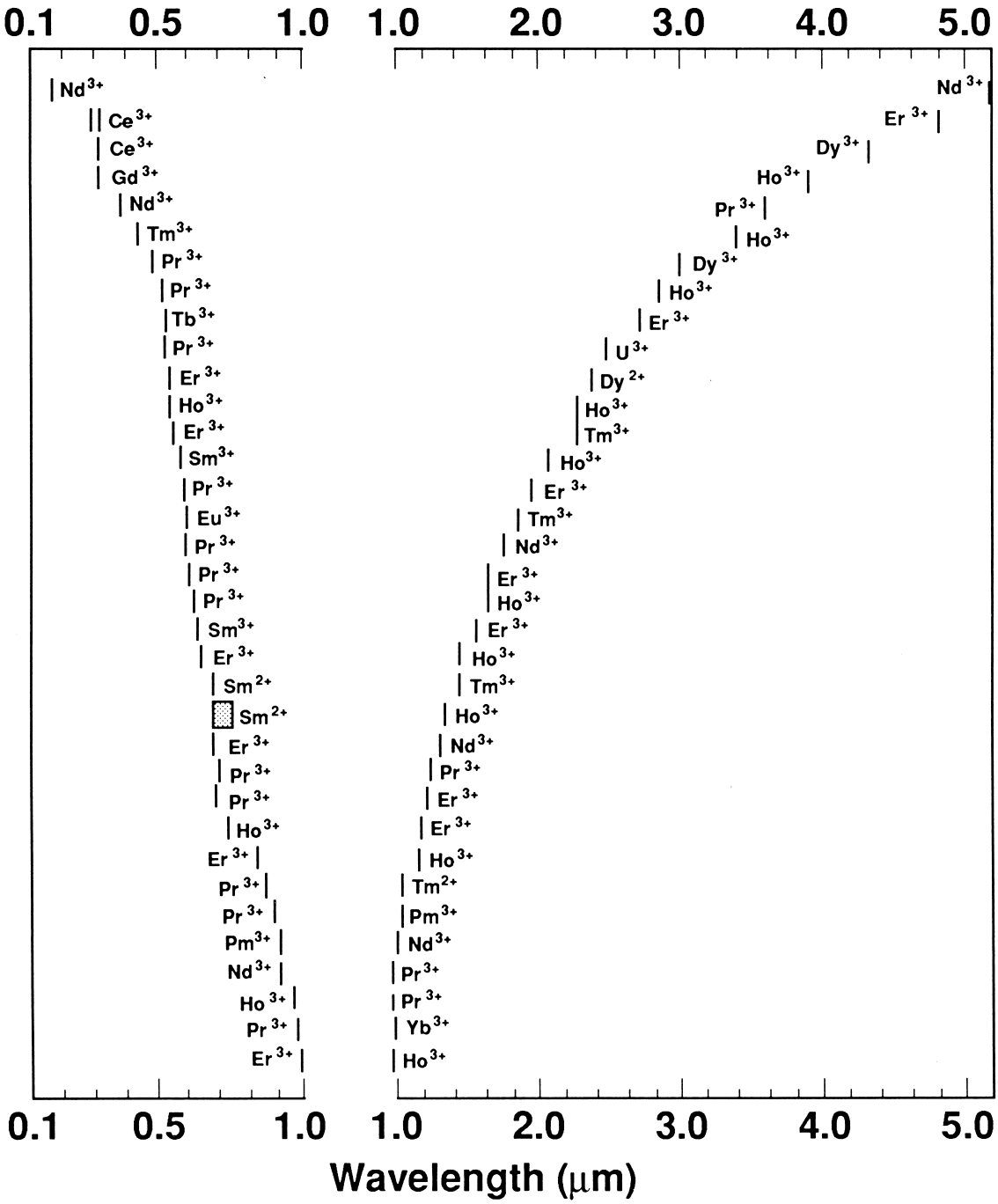


FIGURE 9 Summary of all the wavelengths that may be generated by RE ions. Each line indicates a particular transition for an RE ion, that may have been lased in numerous host materials. Recently laing action has been extended in the midinfrared to beyond 7 μm with Pr:$LaCl_3$.

to Ti^{3+} and Cr^{3+}; they are V^{2+}, Cr^{2+}, Cr^{4+}, Fe^{2+}, Mn^{4+}, Co^{2+}, and Ni^{2+}. These ions exhibit laser output in the 1–4.5 μm region.

The limited number of transition metal ion lasers is related to several factors. First, many of the ions do not have stable oxidation states and tend to vaporize rather than dissolve in the host material. Second, many ions turn out to be either nonemissive, or to have serious excited state absorption losses, rendering them useless as laser ions.

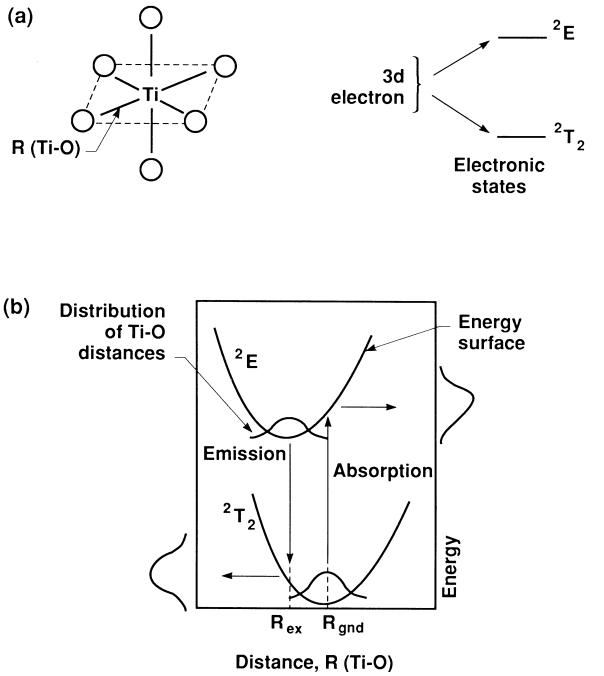


FIGURE 10 (a) Splitting of the 3*d* electron of Ti^{3+} into the ^2E and $^2T_2$ states due to the interaction with the six nearest neighbor oxygen anions of Al_2O_3. (b) Configuration coordinate model of the Ti^{3+} impurity depicting how the displacement between the $^2T_2$ and ^2E states results in broad absorption and emission features.

E. Energy Transfer in Laser Materials

1. Background

The performance of a laser material may be enhanced by the presence of additional impurity ions. Figure 13 illustrates two possible roles for these extra ions. A sensitizer increases the level of pump-light absorption by the laser material, by absorbing in spectral regions where the laser ion normally does not absorb. The sensitizer then efficiently transfers energy to the upper level of the laser ion.

Another role that the additional ions may play is that of "deactivating" the lower laser level. As discussed above, a four-level laser has important advantages since the lower laser level remains unpopulated. If this level does not rapidly drain following the laser action, it will become populated. The deactivator serves to funnel this energy away from the laser ion.

2. Examples of Sensitized Laser Materials

There are several sensitized laser materials that have proved to be useful. As an example, consider the case of Cr^{3+} and Nd^{3+} co doped into $Gd_3Sc_2Ga_3O_{12}$ (GSGG), where Nd^{3+} serves as the laser ion and Cr^{3+} as the sensitizer. The absorption spectrum of GSGG:Cr^{3+},Nd^{3+} is shown in Fig. 14. Since the sharp features are due to the Nd^{3+} ion and the broad bands to Cr^{3+}, it is clear that the Cr^{3+} ion provides greatly enhanced absorption for the flash-lamp-pumped system. (The flash-lamp output is essentially a quasi-continuum throughout the ultraviolet-visible infrared regions.) It is also crucial to note that the $Cr \rightarrow Nd$ energy transfer is extremely efficient (>90%). GSGG:Cr^{3+},Nd^{3+} and related systems have provided the highest flash-lamp-pumped efficiencies measured to date.

The energy transfer of the Cr^{3+}, Tm^{3+}, Ho^{3+}:YAG crystal is an example of an elegant laser system. The Cr^{3+} ions efficiently absorb the flash lamp light. The energy is then transferred to the Tm^{3+} ions, as shown in Fig. 15.

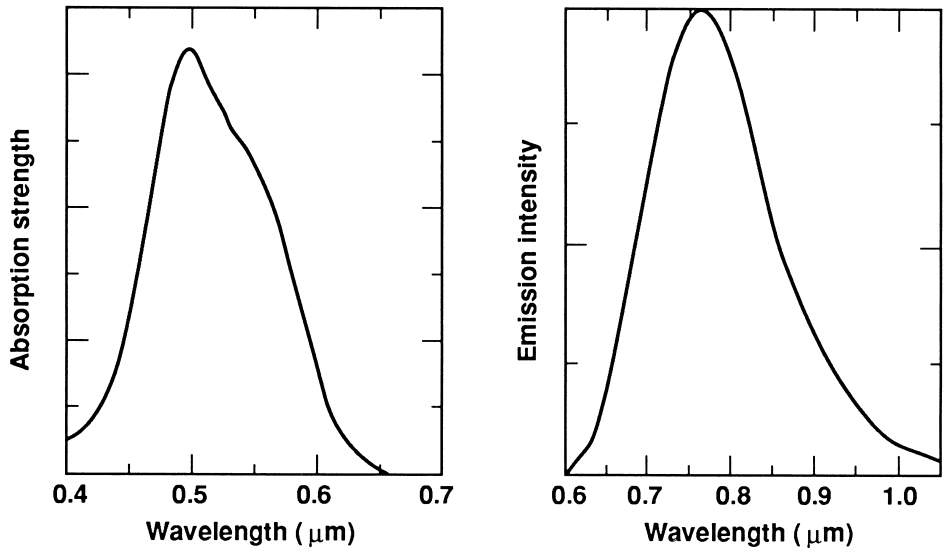


FIGURE 11 Absorption and emission spectra of Ti^{3+}-doped Al_2O_3.

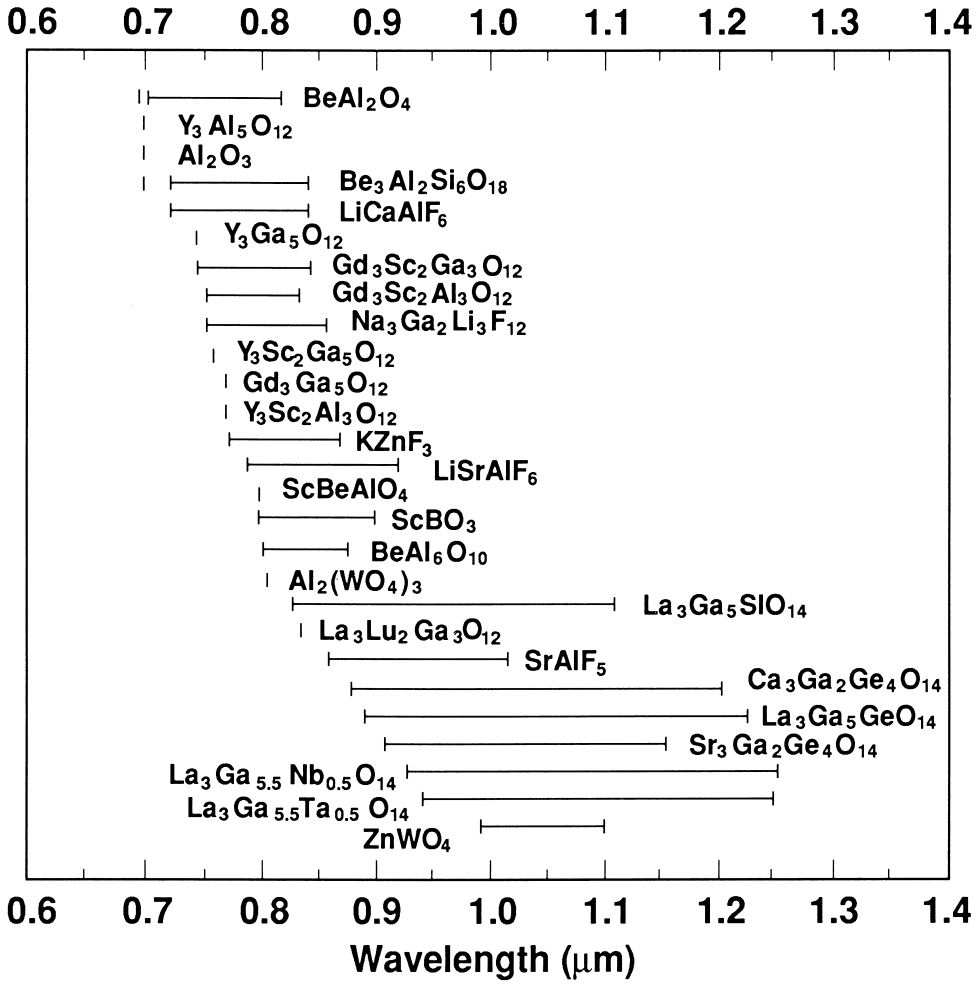


FIGURE 12 Reported tuning ranges of Cr^{3+}-doped crystals.

The Tm^{3+} ions are doped at high concentration to allow for efficient cross-relaxation in which two Tm^{3+} ions are generated in the $^3F_4$ excited state, for each Tm^{3+} ion initially in the $^3H_4$ state. Last, the energy is transferred to the Ho^{3+} ions, which exhibit gain near 2.1 μm. Each of the concentrations of the ions must be carefully chosen to optimize the energy transfer steps.

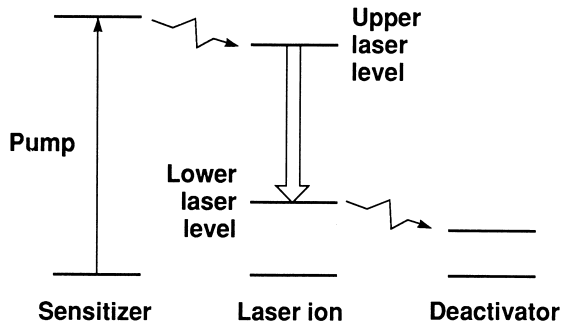


FIGURE 13 Illustration showing the role of a sensitizer to excite the upper laser level, or that of a deactivator, to depopulate the lower laser level.

F. Practical Laser Materials

A myriad of issues impact the effectiveness of a laser material. The issues range from the laser parameters to the thermal-optical-mechanical factors, the raw materials costs, and the ease of fabrication. A listing of some viable solid-state laser materials appears in Table III. Nd:YAG is by far the most widely used material. It offers high gain and has a long storage time of 240 μs. The R_T value, a measure of the material's resistance to thermal stress-induced fracture, shows that YAG is a strong crystal. Nd:YLF is mechanically weaker than YAG (low R_T value), by virtue of its being a fluoride rather than an oxide. The main advantages of Nd:YLF are that it endures significantly reduced thermal lensing, and it stores energy about twice as long.

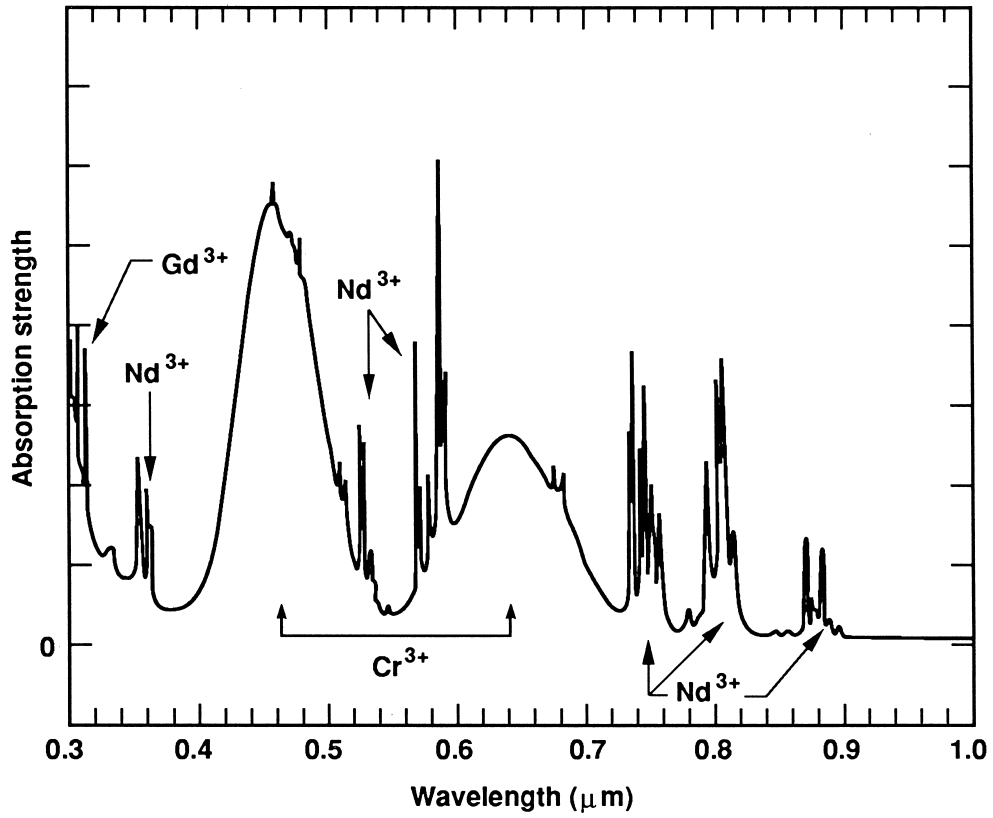


FIGURE 14 Absorption spectrum of Cr^{3+} and Nd^{3+} doped into $Gd_3Sc_2Ga_3O_{12}$. The Cr^{3+} sensitizers provide for more efficient flash lamp absorption, and then rapidly transfer their energy to the Nd^{3+} laser ions.

The main advantage of Nd:glass is that it can be fabricated in large sizes with excellent optical quality. By contrast, the size of crystals is usually much more limited. The sensitization of Nd^{3+} by Cr^{3+}, as has been accomplished into the GSGG host, results in a material that gives flash lamp-pumped efficiencies that are about twice as large as can be obtained for Nd:YAG.

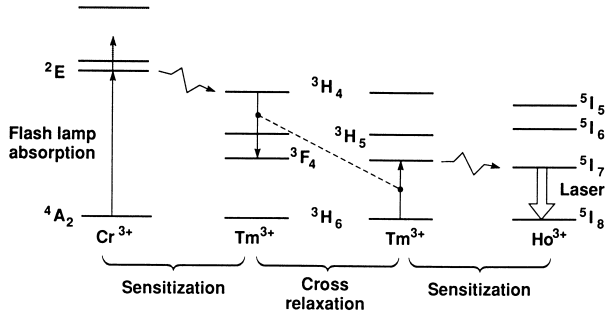


FIGURE 15 Energy-transfer dynamics of the Cr^{3+}, Tm^{3+}, Ho^{3+}:YAG laser crystal. The Cr^{3+} impurities initially absorb the light and transfer the energy to Tm^{3+}. The Tm^{3+} ions then cross-relax to produce two excited states for each ion initially excited. Last, the energy is transferred to the Ho^{3+} laser ions.

While Nd^{3+} is probably the most important laser ion, the Ti^{3+}, Cr^{3+}, Ho^{3+}, and Er^{3+} ions are also useful dopants for laser materials (see Table III). The Ti:sapphire and alexandrite crystals allow for broadly tunable output from the laser, whereas Ho^{3+} and Er^{3+} are capable of generating infrared light. Yb:SiO_2 fibers are now offering high-power performance at 1.1 μm for materials-processsing, while Er:SiO_2 fiber amplifiers have become omnipresent in the telecommunications industry by replacing electronic repeaters with these "transparent" high bandwidth optical amplifiers.

III. SOLID-STATE LASER ARCHITECTURES

The many different laser materials described in the previous section are complemented by a similarly large variety of oscillator and amplifier configurations. On one end of the scale are the very large lasers built to initiate fusion, whose light output pulse delivers petawatts (that is 10^{15} W) of peak power, much more than the total at the size of a pinhead and deliver milliwatts of power. On

TABLE III Properties of Some Common Laser Materials

Name	Material	Laser wavelength λ (μm)	Upper level lifetime, τ (μs)	Thermal stress resistance R_T (W/m$^{1/2}$)	Type of pumping[a]
Nd:YAG	$Y_3AI_5O_{12}$:Nd^{3+}	1.064	240	4.6	FL, DL
Nd:YLF	$LiYF_4$:Nd^{3+}	1.047 1.053	480	1.1	FL, DL
LG-750	Phosphate Glass: Nd^{3+}	1.054	350	0.35	FL
Cr,Nd:GSGG	Gd_3 $Sc_2Ga_3O_{12}$:Cr^{3+},Nd^{3+}	1.061	250	3.3	FL
Ti:Sapphire	Al_2O_3:Ti^{3+}	0.66–1.2	3.2	22	ArL, D-YAG
Alexandrite	$BeAl_2O_4$:Cr^{3+}	0.70–0.82	250	14	FL
Cr:LiSAF	$LiSrAlF_6$:Cr^{3+}	0.78–1.01	67	0.4	FL, DL
Cr,Tm,Ho:YAG	$Y_3Al_5O_{12}$:Cr^{3+} Tm^{3+}, Ho^{3+}	2.1	8000	4.6	FL
Er:glass	Er^{3+}-doped silicates	1.54	10,000	1.2	FL
Yb:SiO_2	Yb^{3+}-doped SiO_2 fiber	1.03	800	22	DL
Cr:ZnSe	ZnSe:Cr^{2+}	2.1–2.9	8	5.3	DL

[a] FL, flashlamps; DL, diode laser; ArL, argon ion laser; D-YAG, doubled neodymium YAG laser.

power consumption of the entire United States at any one instant. Solid-state lasers are used in a wide variety of applications, and the following examples are only an indication of the possibilities. The smallest lasers are used for a multitude of alignment tasks, and the units contained in CD readers and laser pointers are popular representatives of this smallest class. Though physically tiny, the market behind them is huge and measured in billions of dollars. Somewhat larger lasers enable activities involving marking, medical, and military applications. Such units are built by the hundreds to thousands each year. As the power of the laser increases further, applications such as ranging, remote sensing, and wind velocity measurements become important. These systems are quite complicated, since detection and processing play a major part. Some of the most powerful systems are designed for cutting, drilling, and welding at high rates of throughput. A modern automotive production line now includes many robot-controlled solid-state lasers. The very largest lasers are quite unique and serve special research purposes. On the top end of the list is the fusion driver. Nova, as the name suggests, generates enough energy and power to initiate the same process of thermonuclear fusion that powers the stars, but it will soon be considered a smaller representative of this class of lasers.

Regardless of the size of the laser, one always has to start with an active medium, which is energized, or pumped, by either flash lamps or another laser. This creates an inversion, which is then extracted by amplification of an externally applied laser signal or spontaneous emission. This process takes place in a resonator cavity that consists of two or more mirrors, and which may contain optical switches as well. Such an assembly is referred to as an oscillator. There is a multitude of ways to arrange the optical switches in an oscillator to change the temporal format of the laser output; several of these will be discussed in this section. The significant difference between solid-state lasers and other types is that they are storage lasers. The word *storage* refers to the fact that the typical fluorescent lifetime of an inversion is very much longer (several microseconds to milliseconds) than the time it takes for the lasing process to extract the inversion and convert it to laser output (microsecond to nanoseconds). Hence, as the active medium is excited by the pump, the resulting inversion can be accumulated over time (it can be *stored*) and, by use of an electro-optic switch, be extracted at a chosen instant. To get increased amounts of power or energy one can, up to a point, make larger oscillators. Eventually, however, one has to build additional amplifiers to further increase the output of a system. We will discuss amplification towards the end of Section III.

A. Methods of Pumping

1. Flash Lamp Pumping

There are two basic methods with which an inversion can be created in a solid-state laser: flash lamp pumping and laser diode array pumping. The cheapest and most common method uses flash lamps (Fig. 16). A flash lamp essentially consists of a fused silica tube of suitable diameter and length with an electrode at each end. Flash lamps for a laser like that in Fig. 23 are nominally four to five inches long, and around a centimeter in diameter. Those for large fusion lasers can be as long as a meter, and be several centimeters in diameter. Once triggered with a short

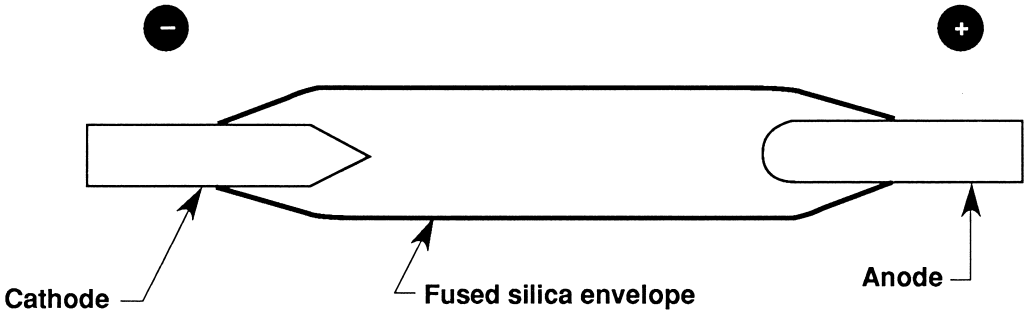


FIGURE 16 The basic elements of a flash lamp. The envelope is typically made from fused silica, the electrodes from a tungsten alloy. The shape of the cathode helps to increase lamp life.

high-voltage spike, a plasma discharge occurs between the electrodes that converts the supplied electrical power to power radiated as light in the infrared, visible, and ultraviolet. Some of this light is absorbed by the active ion with which the host medium is doped and, by virtue of its energy levels and decay dynamics (e.g., Fig. 3), an inversion is created. Pulsed lasers operated in a storage-type mode use different kinds of flash lamps than continuous wave (CW) or quasi-CW systems (pulse duration is long compared to the fluorescent lifetime of the laser ion). The flash lamp used for a CW Nd:YAG laser is filled with several atmospheres of Kr gas. In this case the plasma radiates around 800 nm, which is a wavelength readily absorbed by the neodymium ion. If one wants to pump a pulsed laser, the gas of choice tends to be Xe. In that case the electrical pulse applied to the lamp is of the order or shorter than the lifetime of the upper laser level. The energy delivered to the lamp electrodes can range from joules for smaller lamps, to many kilojoules for large lamps. Pulse durations typically are several hundred microseconds, commensurate with ion fluorescent lifetimes. The plasma radiation pumping the laser is approximately described by a black body spectrum, which extends from the ultraviolet to the infrared spectral regions. As a result, only a small fraction of it is actually absorbed by the active medium. Nevertheless, the peak power densities possible with such pulsed Xe flash lamps are sufficient to construct intense high-power pulsed lasers in this way. Excessive loading of the flash lamp will eventually result in explosive failure of the component, and the best trade-off between lamp size, efficiency, and lamp wall loading is the goal of specifying the correct flash lamp.

2. Diode Pumping

The pumping of solid-state lasers with arrays of laser diodes has attracted great interest within the last few years. This method allows for very efficient solid-state lasers, because the diode lasers themselves efficiently (∼40–50%) convert electrical power to radiated pump power, and the light output by the diode pump array can be accurately tuned to the absorption line of the active medium of the solid-state laser. In other words, unlike a flash lamp, the diode pump array only emits light at a wavelength that is actually used by the active ion to build up inversion. Although still far more expensive than flash lamp pumping, it is the method of choice where efficiency is a premium, as is the case for most military applications. Diode pumped lasers are, nevertheless, likely to become cheaper in the future through increased volume of production. Since laser diodes are limited by their peak power, the total energy output necessarily decreases with a shorter pump pulse duration. The single-shot aerial power density of a two-dimensional diode array is on the order of one kW/cm^2. As a consequence, laser ions with a long fluorescent lifetime are easier to diode pump efficiently than ions with a short fluorescent lifetime. Presently, the most common device involves the use of AℓGaAs diode lasers to pump the Nd^{3+} absorption band near 810 nm in various hosts, although InGaAs at 943 nm for pumping the Yb^{3+} ion are receiving increasingly greater interest. Since the last edition of this chapter, the prediction of cheaper diode arrays has come true to the tune of roughly a factor ten cost decrease. Diode arrays pumping has now moved routinely into the kW laser domain, and very much larger diode-pumped systems, and concomitant cost reductions are on the drawing boards.

The previous edition of this chapter added a section on laser pumping of lasers. This method has remained a laboratory specialty that has never matured outside the scientific marketplace due to the very small overall efficiency and system complexity it entails. A possible exception in the future might be "bleach pumping," where not only a very small fraction of the active ions present in the active medium are being inverted, as is usually the case, but essentially most of them. The result is that there are basically few uninverted ions left in the active medium. This is a useful effect for ions whose lasing transition terminates in, or

close to, the ground state. Bleach pumping then makes the active medium transparent to the pump radiation, as well as the lasing wavelength for this particular transition. In ions like Yb^{3+}, where the lower lasing level is indeed close to the ground state, this is an enabling effect especially at somewhat higher active medium temperatures. The practical drawback is that the pump power densities (kW/cm^2) necessary to achieve this bleach pumping effect are not easy to implement.

B. Optical Switches

The aligned mirrors that surround the active medium and permit repeated passes of laser light through the inversion, form the resonator cavity. Aside from the active medium one can also place a variety of electro-optic switches inside the cavity, including Pockels cells, acousto-optic switches, and saturable dyes. It is these switches that contribute to the great versatility of solid-state lasers and the different output pulse formats that they can generate. Therefore, we shall briefly describe three of them: the Pockels cell, the acousto-optic switch, and the saturable dye.

1. Pockels Cell

Figure 17 explains the basic functioning of a very commonly used device known as the Pockels cell. Such a cell consists of an electro-optic crystal, often a crystal called KD*P [Potassium (chemical symbol K), Dihydrogen (deuteration is indicated by the star), Phosphate] fitted with electrodes to create an electric field within the crystal. The presence of a strong electric field (typically about 3 kV) changes the refractive index of the KD*P crystal for a particular direction of polarization. Since voltages can

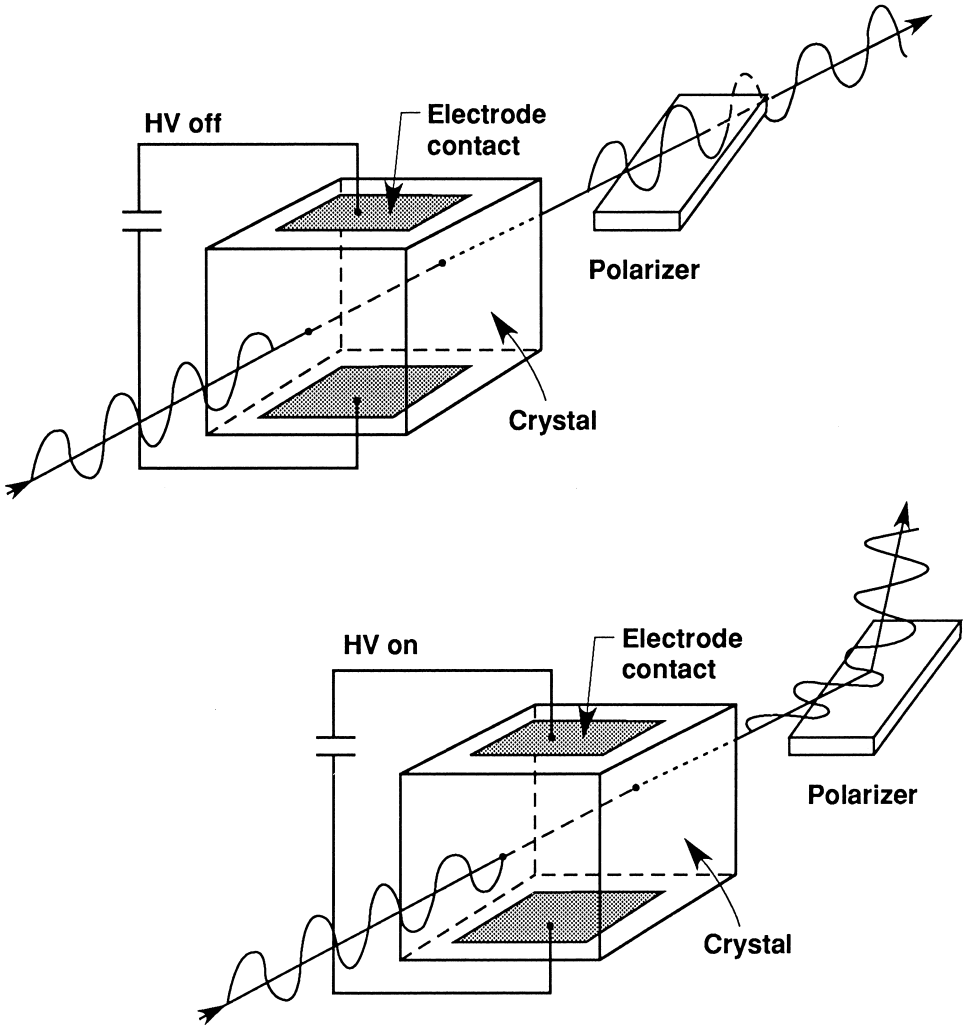


FIGURE 17 The basic operation of the Pockels cell is to rotate the polarization of a transmitted beam upon application of a high-voltage signal. In conjunction with one or two polarizers, Pockels cells are widely used inside of resonator cavities as Q-switches or cavity dumpers, and outside of resonators as fast switchout systems and pulse carvers.

be switched rapidly, one can now also abruptly switch the polarizing properties of the KD*P crystal. For the case of a setup with zero voltage applied to the crystal, the incident beam with vertical polarization passes through the KD*P switch unaffected; this beam will also pass through an appropriately oriented polarizer on the exit side of the crystal. With a voltage applied, however, the KD*P crystal will rotate the direction of polarization of the incident beam by 90°, so that the transmitted beam will now reflect off the polarizer (see Fig. 17). Although the intrinsic response time of the crystal is in the range of tens of femtoseconds, practical rise and fall times are limited by the HV circuitry to switching times on the order of nanoseconds. Hence a Pockels cell can be used as a very fast polarization switch with a good degree of synchronization between the applied high-voltage pulse and the switched light beam. A switching contrast of 1000 :1 with a single Pockels cell and good polarizers is readily achievable. This method is often used for Q-switching (see below), and pulse carving where, in the latter case, a desired pulse shape can be carved out of a longer laser pulse.

2. Acousto-Optic Switch

Another common switch, based on the acousto-optic effect, is sketched in Fig. 18. A piezo-electric transducer driven by a radio frequency (typically tens of Mhz), low-voltage signal launches an acoustic wave into a block of fused silica, where the acoustic wave sets up a sinusoidal refractive index grating in accordance with the photo-elastic effect. This refractive index grating then scatters the incident beam with good efficiency out of its original direction, introducing a corresponding loss in the transmitted beam. There are two typical applications of this technique. For a traveling acoustic wave of a given duration, the transmitted beam becomes switched off while the acoustic wave is present in the beam aperture. Turn-on and turn-off speeds are equal to the time it takes the leading or falling edge of the radio frequency train to travel across the beam diameter (a few microseconds for typical intracavity applications). Where this switch is used for Q-switching, it is important that the acoustic wave be terminated without reflection from the surface of the quartz block in order to achieve fast switching speeds (see below). In comparing this technique to switching an optical beam with a Pockels cell, one finds that the Pockels cell method allows much faster rise times and provides better contrast, but since it also requires considerably higher voltages, acousto-optic switches are preferred for high repetition rate (kHz and higher) applications.

The other key application of acousto-optic switching is based on generating a standing (or stationary) wave in a

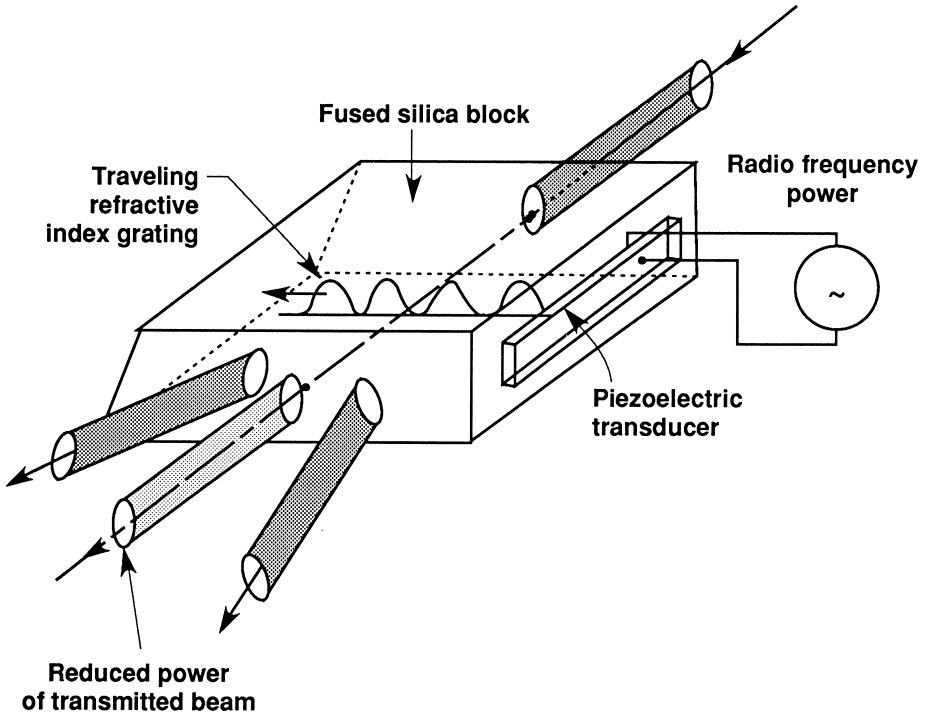


FIGURE 18 The basic operation of an acousto-optic device is to reduce the intensity of a transmitted beam by diffraction from a refractive index grating that is generated, via the photo-elastic effect, by the RF power supplied to the transducer. The device depicted is used for Q-switching and cavity dumping. If configured as a resonator for the acoustic wave, it serves as a mode-locker.

fused silica block, which is explicitly configured to form a resonant cavity for the acoustic wave. This resonant acoustic wave will then create a refractive index grating that oscillates at twice the radio-frequency drive frequency, thus periodically scattering, or switching, the beam from its original direction. An acousto-optic device utilized this way inside of a laser cavity is called a mode-locker (see below).

3. Saturable Absorber

The last switching method to be discussed involves the use of a saturable absorber. The optical beam enters a cell that contains a material that absorbs at the wavelength of the incident light pulse. If the beam enters with enough power per unit area to bleach through, or saturate, the absorbing molecules in the absorber (see bleach pumping, above), the beginning of the pulse will bleach through the cell, making it transparent for the rest of the pulse. Hence such a saturable absorber solution acts more like an intensity threshold filter than an actual switch. The cell can contain a solution of an appropriate dye, or it can be a crystal containing suitable F-centers. The switching of laser light by this method is very popular as a simple and cheap Q-switching and mode-locking technique in the laboratory, and it has even found its way into some products. But the statistics of the bleach process, the lack of timing control, and the chemical stability of the dyes make it unsuited for lasers, which need to be switched precisely and with good repeatability.

C. Modes of Operation

The different switches may be configured in various ways within the resonator cavity, such that many modes of operating a solid-state laser become possible. A wide variety of output pulse durations and formats can be achieved, ranging in duration from picoseconds to a continuous mode of operation, although some pulse durations are easier to produce than others. The availability of different pulse durations and pulse timing, in conjunction with variable output energies and wavelengths, contribute to the versatility of solid-state lasers. Other intracavity elements include wavelength-tuning or wavelength-narrowing devices such as prisms, birefringent tuners and etalons, and apertures to control the transverse mode behavior. In addition, frequency converters outside the resonator cavity can double, triple, quadruple, or even quintriple the optical frequency of output pulses having sufficient peak power. For example, a large fraction of the output of a 1.06-μm Nd^{3+} laser may be shifted to the visible wavelength of 0.53 μm. Diverse medical applications provide an example of the diversity of the wavelength and pulse duration requirements that arise. For example, a CW output at one particular wavelength may be used to cut tissue and another wavelength to coagulate blood, while long pulses are used to break apart kidney stones and short pulses to spot weld loose retinas in the eye.

1. CW Operation

The simplest mode of operating a laser is with no switches in the cavity, so that the resonator only contains the active medium. For a continuous pump, the laser operates in a CW mode. If no transverse or longitudinal mode control is implemented, the output will fluctuate due to the complex ways in which the longitudinal and transverse modes beat and couple to each other via the active medium. Other output fluctuations will come from fluctuations in the power supply. A laser with full mode control will operate only on a single longitudinal and transverse mode. After stabilizing the power supply such lasers can have output fluctuations of less than 1%. CW lasers have many applications, depending on the degree to which they are stabilized and their level of output power. These range from small alignment lasers (milliwatt range), to medical lasers used for surgery (watt range), to lasers used for cutting steel (kilowatt range).

2. Free Running

The term *free running* is generally used to describe a laser, that runs CW for longer times compared to the storage time of the laser, which is typically greater than several milliseconds. Figure 19 shows the pump pulse, the gain

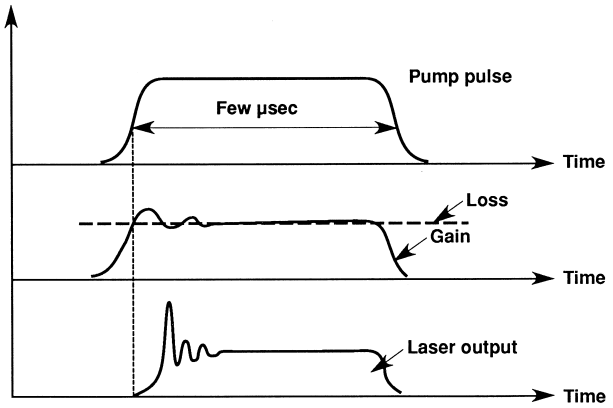


FIGURE 19 Timing diagram for a free-running, pulsed oscillator. Depending on how fast the pump pulse causes the gain to rise above the loss initially, the gain and laser output will react with overshoots that will damp out. The loss level includes that of the active medium and the transmission of the resonator mirror, through which the laser light exits the resonator cavity. In CW and quasi-CW lasers the steady-state gain equals the loss, and the excess pump power is converted to laser output.

and loss in the active medium, and the output energy. As the active medium is pumped sufficiently to exceed threshold, lasing starts with a few output spikes, after which it settles down to the CW output level. The output of lasers is often temporally shaped by tailoring the current pulse to the flash lamps of the diode pump array. The most frequent application of such lasers is for materials working, industrial drilling, and welding applications.

3. Q-switching

This mode of operation requires the addition of a Pockels cell and a polarizer, as sketched in Fig. 17, inside the resonator. Figure 20 shows the timing sequence of the pump pulse from the flash lamp, the impact of the Pockels cell voltage on the loss within the resonator cavity, and the laser output. At first the Pockels cell blocks the light from traveling back and forth between the mirrors. As the active medium is pumped with flash lamp light, the inversion in the upper laser level (the gain) accumulates, thereby storing the inversion until the pump pulse is nearly over. Note that far larger inversions are built up in this way than are possible for the free running case. Once the peak inversion density is reached, the Pockels cell is then switched to low loss. The net gain in the resonator cavity suddenly becomes very high so that energy of the output pulse rapidly builds up and the inversion is extracted efficiently. Note that the pulse duration depends on the amount of inversion just before switching and that the generation of short pulses requires that a large initial inversion be stored. Typical Q-switched oscillators produce 100-mJ pulses in tens of nanoseconds.

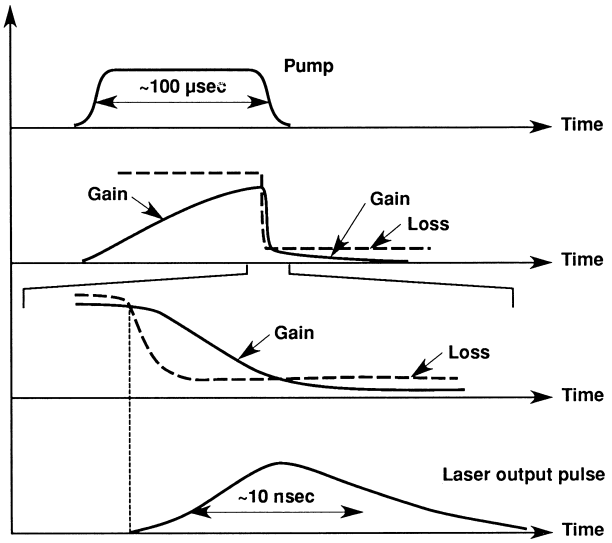


FIGURE 20 Timing diagram for a Q-switched oscillator. Initially, the Pockels cell causes a high loss in the cavity so that lasing is inhibited and the pump power is integrated as inversion in the upper laser level. When the pump pulse is over, the Pockels cell loss is switched off, lowering the loss in the resonator cavity to a value corresponding again to the transmission of the mirror through which the laser light leaves the cavity. The laser pulse will build up rapidly (note the much expanded time scale) and reach the peak at the time where the gain is equal to the loss.

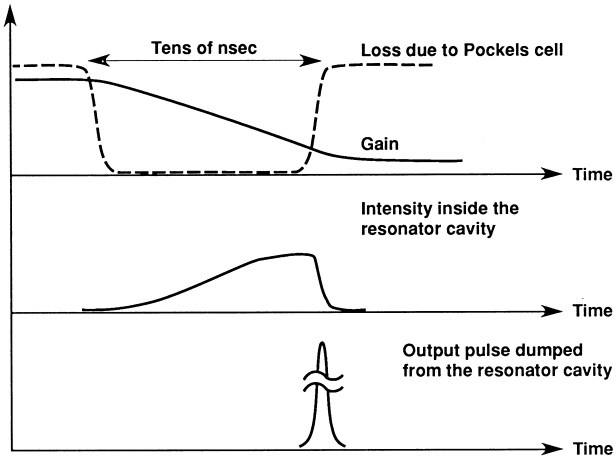


FIGURE 21 In the pump phase the timing diagram for the cavity-dumped case is identical to the Q-switched case. After switching the Pockels cell to enable lasing, the cavity loss now is very small since, for a cavity-dumped architecture, the reflectivity of both resonator mirrors is 100%. Once the power in the cavity has reached the maximum value, the Pockels cell is again switched and ejects the intracavity intensity from the laser in a pulse equal to the round-trip time of the resonator cavity.

4. Cavity Dumping

A mode of operation closely related to Q-switching is cavity dumping. The essential architecture is the same as that for Q-switching, and the timing diagram is given in Fig. 21. The first phase of cavity-dumped operation is similar to that of Q-switching in that energy is stored. Both of the resonator mirrors are selected to be 100% reflective, such that the amplified light remains trapped within the cavity. As the peak intracavity intensity is reached, the Pockels cell rapidly switches the cavity transmission off again. This ejects the light circulating in the cavity by reflection off the polarizer in a pulse whose duration is equal to two cavity passes. This technique is in essence the same as electrical engineers use in a cable discharge, and is used to produce pulses of a few nanoseconds duration, since the pulsewidth now depends on the length of the resonator cavity and not on the amount of inversion stored before switching. The peak power output is significantly limited in this method since the light circulating inside the resonator can become intense enough to destroy critical components of the laser.

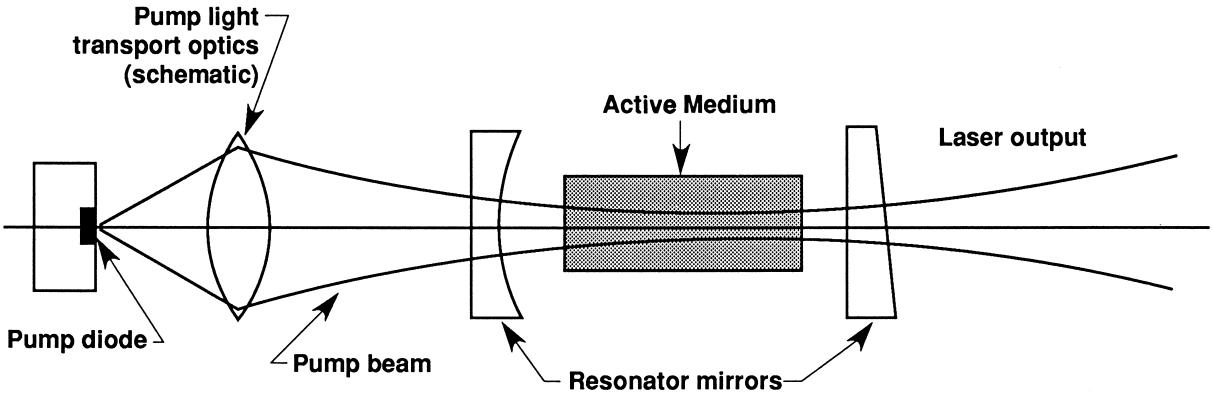


FIGURE 22 The essential elements of a diode-pumped minilaser as explained in the text. In practice, the pump light transport optics is a sophisticated array of lenses to change the spatial characteristics of the diode output to those optimally suited to pump the minilaser.

An architecture very closely related to the cavity-dumped oscillator is the regenerative amplifier. The principle difference is that the laser oscillation in a regenerative amplifier does not build up from spontaneous emission, but is initiated by a signal injected into the resonator from the outside, as the Pockels cell is switched to transmission. This injected signal is then trapped in the cavity and amplified until it has reached maximum intensity, at which point it is ejected (dumped) from the cavity.

5. Mode-Locking

By inserting an acousto-optic mode-locker in the cavity it is possible to produce very short pulses in an oscillator. With a mode-locker in the cavity, and the transverse modes suitably constrained, the oscillator can be operated quasi-CW to produce a steady stream of short pulses (i.e., 25-psec pulses every 10 ns). It is also possible to add a Pockels cell and a polarizer and produce an output that has the pulse envelope of a Q-switched pulse, but that is composed of individual short pulses from the mode-locking process. External to the cavity one can then pick out a single one of those pulses by placing an additional Pockels cell between two polarizers and applying a short high-voltage pulse at just the instant when the desired pulse is at the Pockels cell. Such an arrangement is commonly called a "single-pulse switchout." This is a standard way to produce individual pulses with durations from 100 psec to 1 nsec and energies on the order of 100 μJ. Most recently, the technique of Kerr lens mode-locking has been developed where the cavity is configured in such a way that only mode-locked operation is stable (by virtue of an intensity-dependent lens).

D. Types of Oscillators

In this section representative examples of frequently used solid-state lasers are described, starting again with the smallest lasers and proceeding to lasers of increasing power and energy. These lasers are practical embodiments of the pumping and operating schemes described above.

1. Mini-Lasers

Figure 22 is a sketch of a generic diode-pumped mini-laser. A single diode is focused with suitable optics through one of the resonator mirrors into the active medium (typically Nd:YAG, or Nd:YVO_4). This "end-pumping" method allows the precise matching of the pumping volume to the lasing volume of the lowest order transverse mode, thereby contributing to the excellent efficiency of mini-lasers. Output powers are typically in the milliwatt range. The mode of operation ranges from CW to repetitively Q-switched (requiring the addition of a small Pockels cell or saturable absorber to the resonator cavity). These lasers offer a stability of a few percent, and since they are physically very small, they are extremely rugged. The entire device, including power and cooling, comfortably fits in the palm of the hand.

2. Watt and Joule Level Laser

This heading encompasses lasers with outputs in the range of a few Watts, and with pulsed energies of 0.1–1 J. CW systems, as well as pulsed systems, are characterized by a great variety of resonator configurations and make up the majority of solid-state lasers sold today. These lasers typically contain a rod with a diameter on the order of 6 mm and a length of 10 cm, which is pumped with one or two flash lamps. The resonator cavity is generally 30 cm to 1.5 m long, depending on whether there is a mode-locker, a Pockels cell, or another combination of switches. The cavity may be made from low thermal expansion materials for more demanding applications. The power supply operates in the kilowatt range, and typically is a floor unit occupying a few cubic feet of space. A photograph of

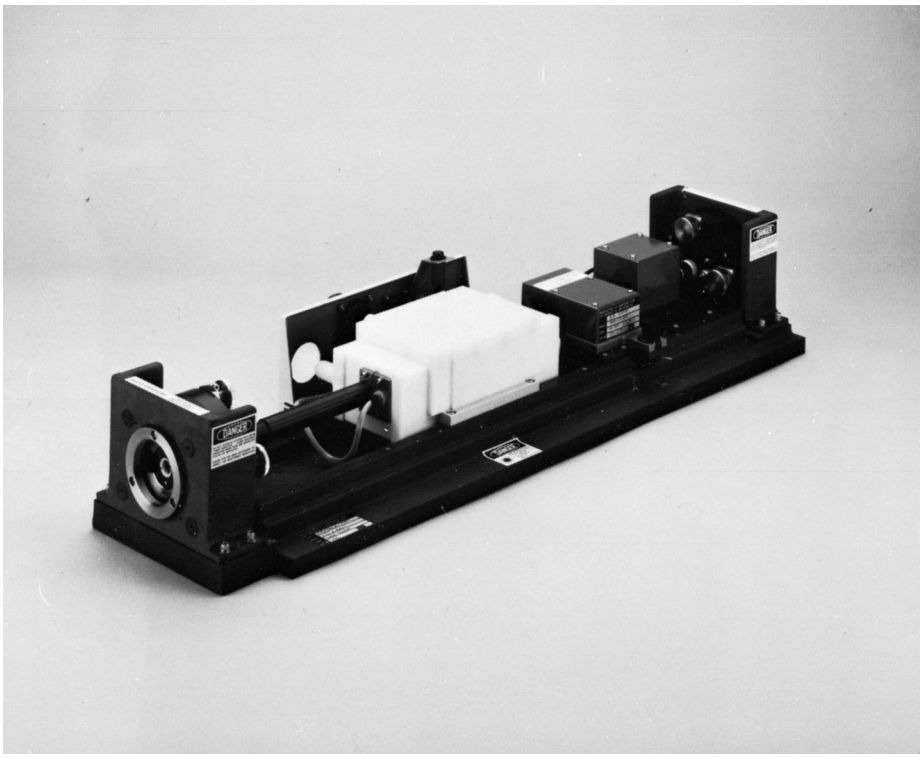


FIGURE 23 A 250-W Nd:YAG laser used in the semiconductor industry. (Courtesy of Quantronix Corp.) The complete unit includes a beam delivery system and the means to observe the work piece through a microscope.

a representative 200-W system appears in Fig. 23 (courtesy of Quantronix Corporation). Various CW versions are used for an ever increasing variety of medical applications and for marking and scribing in industrial settings. The CW mode-locked laser is an important component of many ultrafast pulse setups. Free-running, pulsed, and Q-switched systems have other uses in the medical field (e.g., breaking up of kidney stones without intrusive surgery). Industrial applications include semiconductor processing and working thin metal films. Many lasers for scientific applications operate in this range as well. Commercial laser companies offer a variety of resonators and modes of operation for the scientific market, and almost arbitrarily complex variations can be seen in the laboratory. Most recently, diode-pumped fiber lasers have been extended to operate at the watt level.

3. Material Working Lasers

These lasers operate mostly in the CW or free running pulsed mode and deliver output powers up to and exceeding 1 kW. They are configured as oscillators, but often have two or three laser rods in the resonator cavity. The resonator cavity is still approximately 1 m long, although the power supply is now significantly larger. Furthermore, these lasers, as well as some of the smaller lasers described above, often incorporate sophisticated beam delivery systems, which allow the use of one central laser system on several work stations. Larger systems are used for demanding welding and drilling application (e.g., components for jet engines) where throughput and reliability are essential.

A more sophisticated embodiment is the slab laser (Fig. 24). As opposed to the rod-shaped active media typically used in conventional solid-state lasers, the active medium of slab lasers is in the form of a thin plate or slab. Higher power slabs are around 5 mm thick or a bit thicker, from 10–20 cm long, and 1–2 cm wide. The beam enters and exits through tilted surfaces and zigzags through the active medium in such a way that the optical distortions across the slab thickness, which a beam experiences by passing through the material, can largely be eliminated. A slab laser can generate increased output power by employing a larger volume of the active medium than a rod, while maintaining the dimension across which the active medium must be cooled. This results in a higher quality output beam, which can be brought to a tighter focus at the work piece. Slab lasers involve a number of complex technological issues. The decision of the marketplace so far appears to be that, despite a considerable development effort, these complexities are not being outweighed by the more sophisticated advantages these lasers offer.

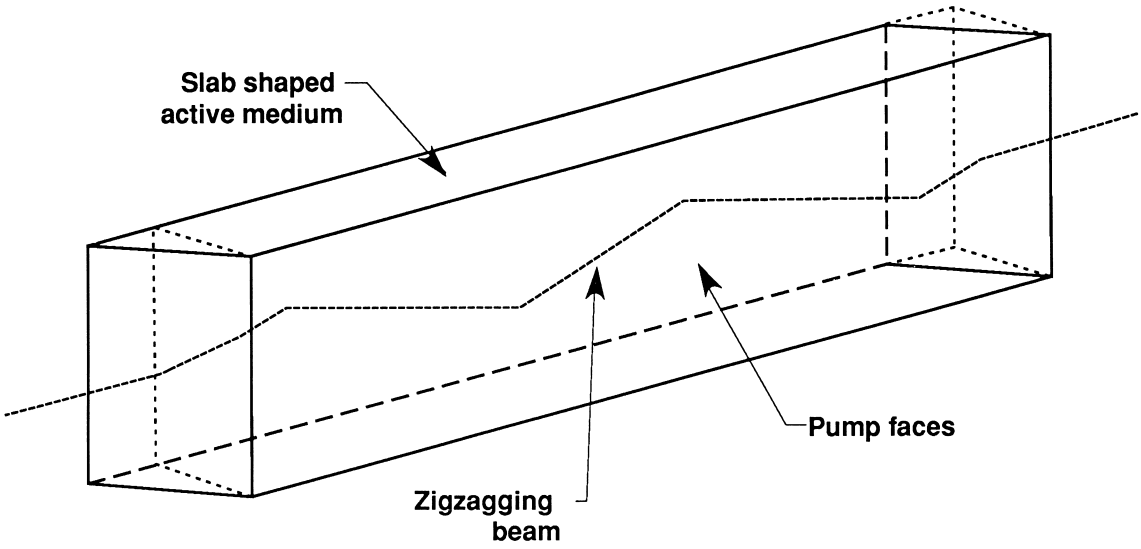


FIGURE 24 The basic idea behind a slab laser is to zigzag the beam through the active medium and thus greatly reduce the thermal distortions of the beam. The dotted lines are meant to indicate where perpendicular edges would lie. The actual slab laser has entrance and exit surfaces that are inclined with respect to the incident beam.

They have remained a specialty and have not found broad commercial acceptance.

E. Laser Scaling

We have thus far primarily described laser oscillators. To achieve higher energies, amplification of the pulse from the oscillator is required. The regenerative amplifier was already mentioned in the section on cavity dumping. In comparison to a single-pass amplifier, the regenerative amplifier is a far more complex setup, although it provides for very efficient extraction of energy from the medium and is favored for short pulse amplifiers in the laboratory. To get to higher energy, further amplification stages are required. One would, of course, like to make the most efficient use of the inversion stored in such an amplifier, while avoiding the potentially destructive consequences of propagating a high peak power pulse through the amplification medium. Consider a pulse making a single pass through an amplifier in which an inversion is stored. The key parameter to describe the amplification process is the saturation fluence, F_{sat}, which is given by

$$F_{sat} = h\nu/\sigma_{SE}, \tag{1}$$

where σ_{SE} is the stimulated emission cross section, h is Planck's constant, and ν is the frequency of the laser light. If the input fluence [J/cm^2] to the amplifier remains small compared to the saturation fluence of the transition (called the small signal gain regime), the amplification G proceeds in an exponential manner according to,

$$G = \exp(N_i \sigma_{SE} x), \tag{2}$$

where N_i is the inversion density and x is the pathlength in the amplifier. In this regime the amplification factor G is high, but the extraction of the inversion is consequently low, since saturation of the inversion does not occur. The other extreme is that the input fluence is much larger than the saturation fluence. In this case extraction of the inversion will approach 100%, and the energy stored in the amplifier is simply added to the energy in the input beam. Clearly, the best place to operate is in between the two extremes, where there is a desirable degree of signal amplification at an acceptable extraction efficiency of the energy originally stored in the amplifier. The basic equation that describes this amplification process is the Frantz–Nodvik equation.

There are a variety of effects that limit the energy that may be generated in a large-scale amplifier chain. First, the amplifier medium cannot be made arbitrarily large since it will lose its storage capability through the following process. Spontaneous emission, naturally generated by fluorescence decay of the inverted ion population inside the amplifier, travels through the active medium and will, of course, be amplified. As this occurs, this amplified spontaneous emission (ASE) will deplete some of the inversion originally created with the intent to amplify an external signal. Worse, this ASE can even get trapped inside the active medium in a path that leads back onto itself. This process is, understandably, called *parasitic oscillation*. Another amplification-limiting effect is that as the peak power in the amplified pulse increases its instantaneous power density [W/cm^2], it becomes "intense" enough that the electric field of the amplified light wave itself changes

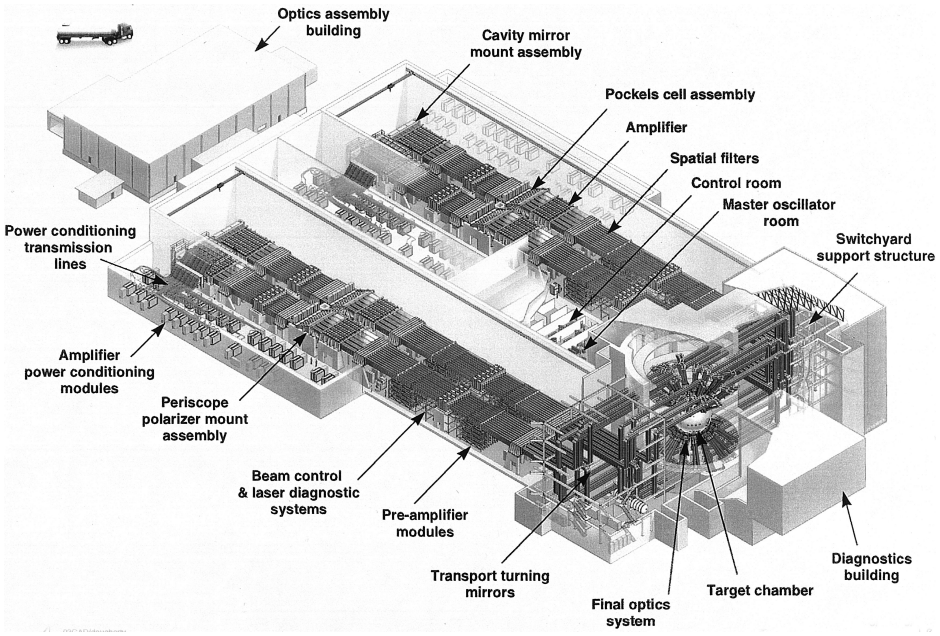


FIGURE 25 The National Ignition Facility (NIF) under construction at Lawrence Livermore National Laboratory, will be the biggest laser there ever was. It will have 192 individual beams. Note the delivery truck for scale.

the refractive index of the very medium through which it travels. At some point this induces a process called *self-focusing*, where the amplified beam destroys the active medium by creating tracks and bubbles inside the amplifier. Finally, amplifier surfaces and other optics are often covered with dielectric coatings to increase, or decrease, their reflectivities. These coatings can be destroyed by a light pulse that has too high a peak power, often at power levels below that at which selffocusing sets in. The resistance of such coatings to this effect is described by a damage threshold that characterizes the fluence the coating can withstand at a given pulse duration. Note that the last two of these effects have nothing to do with a laser system being "large" per se; it is the power densities, not the amount of absolute power, that determines if these effects occur and to what extent.

As the ultimate example of a laser that can only be built if all of these types of processes are well understood, Fig. 25 shows the National Ignition Facility (NIF) laser currently under construction at the Lawrence Livermore National Laboratory. It is designed to deliver precisely shaped pulses, several nanoseconds long, with several megajoules of energy.

IV. FUTURE DIRECTIONS

In the previous edition of this article, we predicted three clear areas where solid-state laser architectures will advance significantly in the next few years. The first is to take diode-pumped lasers from the minilaser scale to the tens of Watts level. This requires development of high-average power, two-dimensional diode arrays, which is now well underway. The second is to utilize slab laser technology to operate lasers at the 1–10-J/pulse level at repetition rates of about 100 Hz. Finally, large fusion lasers may ultimately operate at several Hertz, and produce average output powers in the megawatt regime. As for laser systems, all three of these expectations have been very nicely fulfilled. As for the next several years, we expect the development of solid-state lasers to tens and even hundreds of kW of power for burst durations of many seconds, and a further significant decrease in the cost of diode-pump arrays. In the longer term, these developments, if they are brought to a meaningful level of completion, will enable the further spread of diode-pumped solid-state lasers into more and more areas of medical, industrial, and military applications.

The new laser materials that are developed will be deployed in systems to generate new wavelengths, operate more efficiently, be produced at lower cost, and have optical properties that are tailored to meet specific technical objectives. Solid-state lasers that operate efficiently in the ultraviolet-blue region have been developed, and we may reach into the vacuum ultraviolet in the future. Many new transition metal lasers may be discovered in the next decade having useful optical properties. One change that may occur will involve the advent of tailoring some types

of laser materials for a specific application. (This has already occurred in the case of Nd-doped glasses for fusion lasers.) This process of designing, rather than discovering, laser materials will be enhanced by a better understanding of the physics and chemistry of solid-state media.

ACKNOWLEDGMENT

This research was performed under the auspices of the Division of Materials Sciences of the Office of Basic Energy Sciences, U.S. Department of Energy, and the Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48.

SEE ALSO THE FOLLOWING ARTICLES

ACOUSTO-OPTICS • LASERS • LASERS, SEMICONDUCTOR • LASERS, ULTRAFAST PULSE TECHNOLOGY • LIGHT EMITTING DIODES

BIBLIOGRAPHY

Adair, R., Chase, L. L., and Payne, S. A. (1989). "Nonlinear refractive index of optical crystals," *Phys. Rev. B.* **39,** 3337.

Albrecht, G. F. (1990). "Average power slab lasers with garnet crystals as the active medium," in YAG and Other Garnet Lasers, (De Shazer, ed.), Wiley.

Ballhausen, C. J. (1962). "The Theory of Transition-Metal Ions," University Press, Cambridge.

Belforte, D., and Levitt, M. (1989). "The Industrial Laser Annual Handbook," PennWell, Oklahoma.

Berger, J., Welch, D. F., Scifres, D. R., Streifer, W., and Cross, P. S. (1987). "High power, high efficiency Nd:YAG laser end pumped by a laser diode array," *Appl. Phys. Lett*. **51,** 1212.

Bernard, J. E., McCulkbyh, E., and Alcock, A. J. (1994). "High gain, diode-pumped Nd:YVO_4 slab amplifier," *Opt. Comm*. **109,** 109.

Bibeau, C., Beach, R. J., Mitchell, S. C., Emanuel, M. A., Skidmore, J., Ebbers, C. A., Sutton, S. B., and Jancaitis, K. S. (1998). "High-average power 1-μm performance and frequency conversion of a diode-and-pumped Yb:YAG lasers," *IEEE J. Quantum Electron* **34,** 2010.

Bowman, S. R., Shaw, L. B., Feldman, B. J., and Ganem, J. (1995). "A seven micror solid state laser," in Advanced Solid State Laser Topical Meeting, Memphis, TN, January 30–February 2.

Brown, D. C. (1981). "High Peak Power Nd:Glass Laser Systems," Series in Optical Sciences, Vol. 25. Springer-Verlag, New York.

DeLoach, L. D., Page, R. H., Wilke, G. D., Payne, S. A., and Krupke, W. F. (1996). "Transition metal-doped zinc chalcogenides: spectroscopy and laser demonstration of a new class of gain media," *IEEE J. Quantum Electronics* **32,** 885.

Desurvire, E. (1994). "Erbium-Doped Fiber Amplifiers," Wiley, New York.

Desurvire, E. (1989). "Analysis of Erbium-Doped Fiber Amplifiers Pumped in the $^4I_{15/2}$–$^4I_{13/2}$ Band," *IEEE Photonics Technol. Lett*. **1,** 293.

Dieke, G. M. (1968). "Spectra and Energy Levels of Rare Earth Ions in Crystals," Interscience, New York.

Izumitani, T. S. (1986). "Optical Glass," American Institute of Physics, U.S.A.

Kaminskii, A. A. (1981). "Laser Crystals," Springer-Verlag, New York.

Koechner, W. (1988). "Solid State Laser Engineering," Series in Optical Sciences, Vol. 1, Springer, New York.

Krupke, W. F., Shinn, M. D., Marion, J. E., Caird, J. A., and Stokowski, S. E. (1986). "Spectroscopic, optical and thermo-mechanical properties of neodymium- and chromium-doped gadolinium scandium gallium garnet," *J. Opt. Soc. Am. B* **3,** 102.

Maiman, T. H. (1960). *Nature* **187,** 493.

Moulton, P. F. (1986). "Spectroscopic and laser characteristics of Ti:Al_2O_3," *J. Opt. Soc. Am. B* **3,** 4.

Muendel, M., Engstrom, B., Kea, D., Lediberte, B., Mims, R., Robinson, R., Rockney, B., Y-Zhang, Collins, R., Gaurilovic, P., and Rawley, H. (1997). "35 cw single-mode ytterbium fiber laser," Conference on Lasers and Electrooptics.

Paschotta, R., Nilsson, J., Tropper, A. C., and Hanna, D. C. (1997). "Ytterbium-doped fiber amplifiers," *IEEE J. Quantum Electron* **33,** 1049.

Payne, S. A., Chase, L. L., Smith, L. K., Kway, W. L., and Newkirk H. W. (1989). "Laser Performance of $LiSrAlF_6:Cr^{3+}$," *J. Appl. Phys*. **66,** 1051.

Payne, S. A., Chase, L. L., Newkirk, H. W., Smith, L. K., and Krupke, W. F. (1988). "$LiCaAlF_6:Cr^{3+}$: A promising new solid-state laser material," *IEEE J. Quant. Electron*. **24,** 2243.

Petricevic, V., Gayen, S. K., and Alfano, R. R. (1989). "Continuous-Wave Laser Operation of Chromium-Doped Forsterite," *Opt. Lett*. **14,** 612.

Pfaender, H. G., and Schroeder, H. (1983). "Schott Guide to Glass," Van Nostrand Reinhold, New York.

Robertson, A., Knappe, R., and Wallenstein, R. (1997). "Kerr-lens mode-locked Cr:LiSAF femtosecond laser pumped by the diffraction-limited output of a 672-nm diode-laser master-oscillator power-amplifier system," *J. Opt. Soc. Am. B* **14,** 672.

Schawlow, A. L., and Townes, C. H. (1958). *Phys. Rev*. **112,** 1940.

Siegman, A. E. (1986). "Lasers," University Science Books, Mill Valley, CA.

Walling, J. C., Heller, D. F., Samelson, H., Harter, D. J., Pete, J. A., and Morris, R. C. (1985). "Tunable alexandrite lasers: development and performance," *IEEE J. Quantum Electron*. **QE-21,** 1568.

Weber, M. J., ed. (1982). "Handbook of Laser Science and Technology," CRC Press, Boca Raton, FL.

Yariv, A. (1975). "Quantum Electronics," John Wiley and Sons, New York.