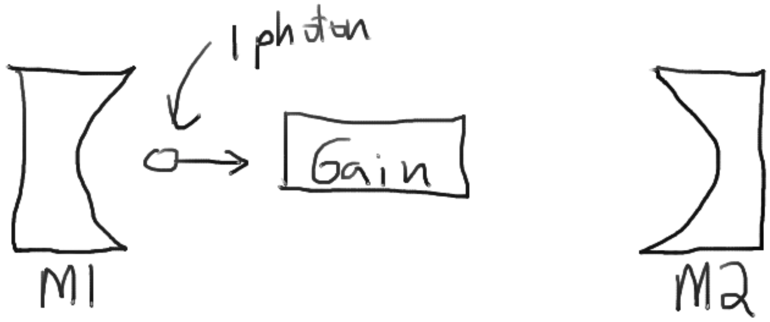
Laser Lecture
ERH

LASER: **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation

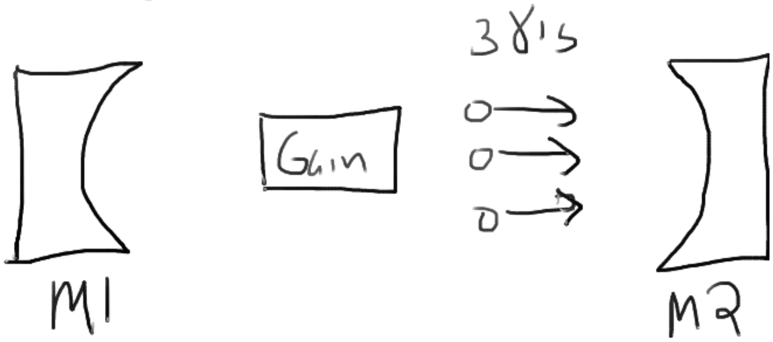
Last week we learned about optical cavities, where resonant photons bounce back and forth between high-reflectivity mirrors for many times before they either leak out of the cavity due to transmission or are absorbed due to losses at mirrors.

Now, imagine we could prepare a magical material that would amplify photons – i.e. one photon goes in and more than 1 identical photons come out. If we placed this magical material – let's call it a gain medium – inside of an optical cavity, what would happen?

Suppose one photon goes in:

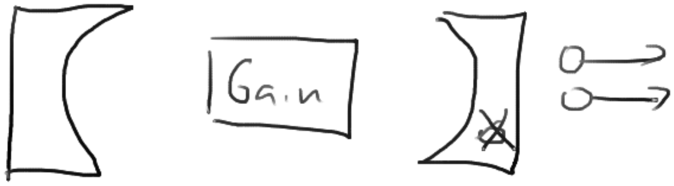


And three photons come out:



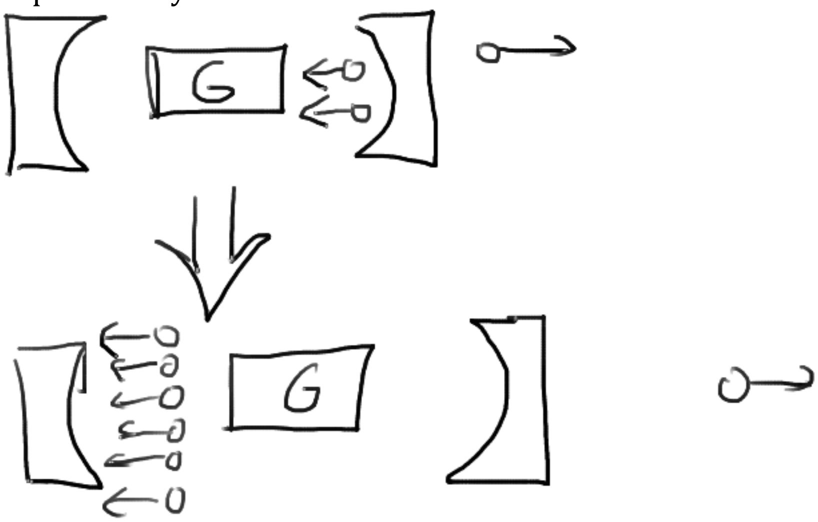
What happens next? Well three things can happen actually depending on the reflectivity and loss on the mirrors. Defining the total loss as $\mathcal{L} = T + L$ and the total gain as G, we have (technically speaking $\mathcal{L}$ and G are defined for a cavity round trip, so includes both mirrors and two passes through the gain medium, but below we're using them in a "half-round trip" manner for clarity):

1. $\mathcal{L} > G$. The photons either leave the cavity, are absorbed, or both. The process stops.



2. $\mathcal{L} = G$. Here the losses are just balanced by the gain and the process can continue indefinitely! (This is a laser! More specifically, this is known as a laser at threshold, because the output light intensity is right on the edge of growing tremendously – see next item.)

3. $\mathcal{L} < G$. Here the gain outpaces the losses and the photon number grows exponentially! This is an even better laser!



Of course, this process cannot continue forever, otherwise the photon number would go to infinity. So what happens? Well it turns out that you cannot make a gain medium that amplifies the photon by the same factor regardless of the input photon number. Eventually, the gain saturates. For example, the G is about 3 for small photon numbers in our example, but at large photons numbers it will be much smaller – we'll see why later in this lecture. Thus, eventually, the gain will equal the losses and the laser will reach a steady state where this is a large number of photons in the cavity and hopefully a large number in the output beam.

Already though we can see several key features of ALL lasers:

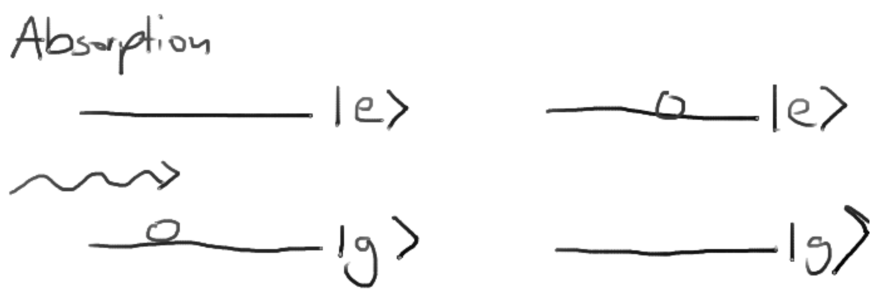
1. The laser works at the resonant frequency of the cavity. Otherwise the photons will not bounce back and forth for many round trips.
2. Steady state is reached when gain = loss.
3. Because these photons are travelling in the cavity, they take on the cavity spatial mode, e.g. TEM_{00}. Thus, when they leak out of the cavity they propagate as TEM_{00}. This is why laser beams have nice mode structure (hopefully!).

Okay, so we see how lasing could work. The only problem is how do we make this magical medium that takes one photon in and gives us more identical photons out?!

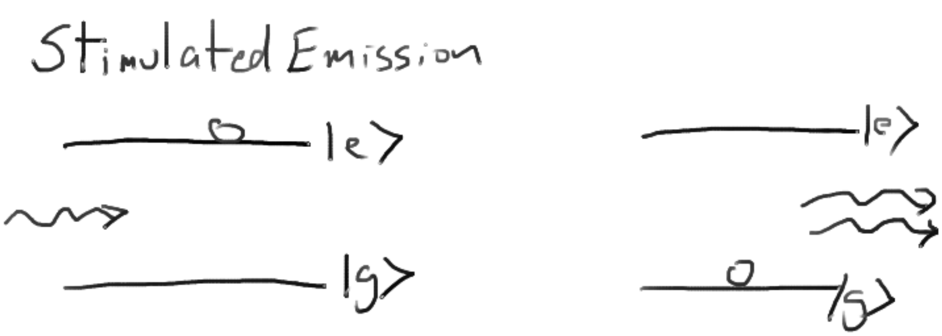
Think about the acronym?

Stimulated Emission.

Consider the lowest two quantum states of an atom, subject to an electromagnetic field (photons) that oscillates (has the same energy as) the splitting between the two quantum states. What happens if the atom was initially in the ground state?



What if the atom had initially been in the excited state?



Even without detailed quantum mechanics calculations, we can see that stimulated emission should exist by time reversal symmetry. If absorption happens, then so should stimulated emission. And, as you know, the stimulated photon is identical to the original one -- to prove this requires quantum electrodynamics.

More serious derivation of Stimulated Emission (Skip in lecture)
The transition between the two states happens because the electric field of the electromagnetic field perturbs the electron orbit. If the electric field is just:
$\vec{E} = E_o \cos(\omega t)\,\hat{x}$.

The energy of the electron in the electric field is given as:
$V = e \int \vec{E} \cdot d\vec{r} = e\, E_o z \cos(\omega t)$

Thus, the Hamiltonian of the system is:

$H = H_o + e\, E_o z \cos(\omega t)$

where H_o is the Hamiltonian of the atom in the absence of the field. If we only look at the lowest two levels of H_o (say the S and P state of the H atom), we can write the effective Hamiltonian as:

$$H = \begin{pmatrix} E_g + V_{gg} & V_{eg} \\ V_{ge} & E_e + V_{ee} \end{pmatrix}$$

where $E_i = \langle i|H_o|j\rangle$, $V_{ij} = e\, E_o\langle i|z|\mathrm{j}\rangle \cos(\omega t)$, and a general solution is just $\psi = a(t)|g\rangle + b(t)|e\rangle$.

Since z is parity odd and $|g\rangle$ and $|e\rangle$ are eigenfunctions of parity, the diagonal contributions of V vanish. (You can try this with the hydrogen wavefunctions if you don't believe me.) And defining the ground state energy as 0 and remembering that $E_e = \hbar\omega$ (i.e. we are on resonance), we have:

$$H = \begin{pmatrix} 0 & dE\cos(\omega t) \\ dE\cos(\omega t) & \hbar\omega \end{pmatrix}$$

where we have defined $d = \langle e|z|g\rangle = \langle g|z|e\rangle$. With one final definition of $\hbar\Omega = dE$ (Ω is called the Rabi frequency), the Schrodinger equation gives us two coupled differential equations:

$$i\dot{a} = \Omega\cos(\omega t)\, b$$
$$i\dot{b} = \Omega\cos(\omega t)\, a + \omega\, b$$

Looking at this equations, we see that time derivative of a is equal to just one thing, but the time derivative of b is equal to two things. Thus, b must be two time dependent things multiplied together. Also, we know that in the absence of the perturbation $b(t) = e^{-iE_e t/\hbar} = e^{-i\omega t}$ and $a(t) = e^{-i\,0\,t/\hbar} = 1$, therefore we guess that $b(t) = c(t)e^{i\omega t}$ and our equations becomes:

$$i\dot{a} = \Omega\cos(\omega t)\, c\, e^{i\omega t} = \frac{\Omega}{2}\left(e^{i\omega t} + e^{-i\omega t}\right)ce^{i\omega t} \approx \frac{\Omega}{2}c$$

and

$$\omega ce^{-i\omega t} + i\dot{c}e^{-i\omega t} = \frac{\Omega}{2}\left(e^{i\omega t} + e^{-i\omega t}\right)a + \omega ce^{-i\omega t}$$

$$i\dot{c} = \frac{\Omega}{2}\left(1 + e^{-2i\omega t}\right)a \approx \frac{\Omega}{2}a$$

Where we have thrown away terms with $e^{\pm 2i\omega t}$, as these oscillate very quickly and there effect "averages" out compared to the other term. (This is called the rotating wave approximation, since our guess for $b(t)$ can be tied with a transformation to the interaction picture, which when you are on resonance is the same as transforming to a frame that rotates at ω.)

Now, we just have two trivially coupled equations:

$$i\dot{a} = \frac{\Omega}{2}c$$

$$i\dot{c} = \frac{\Omega}{2}a$$

I've seen these solved a bunch of ways, but the coolest is in they Feynman lectures, which he probably stole from somebody. It goes like this. Add the two together and subtract them to make two new equations:

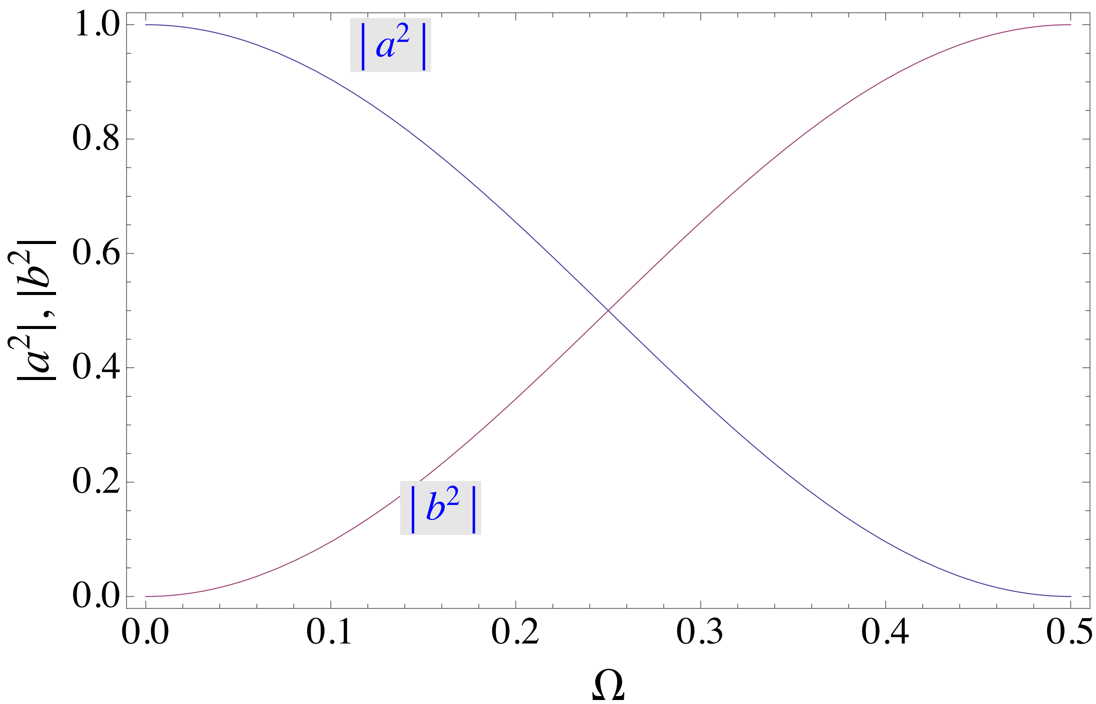
$$i(\dot{a} + \dot{c}) = \frac{\Omega}{2}(a + c)$$
$$i(\dot{a} - \dot{c}) = -\frac{\Omega}{2}(a - c)$$

Thus,:

$$(a + c) = e^{-i\left(\frac{\Omega}{2}t + \phi\right)}$$
$$(a - c) = e^{i\left(\frac{\Omega}{2}t + \phi\right)}$$

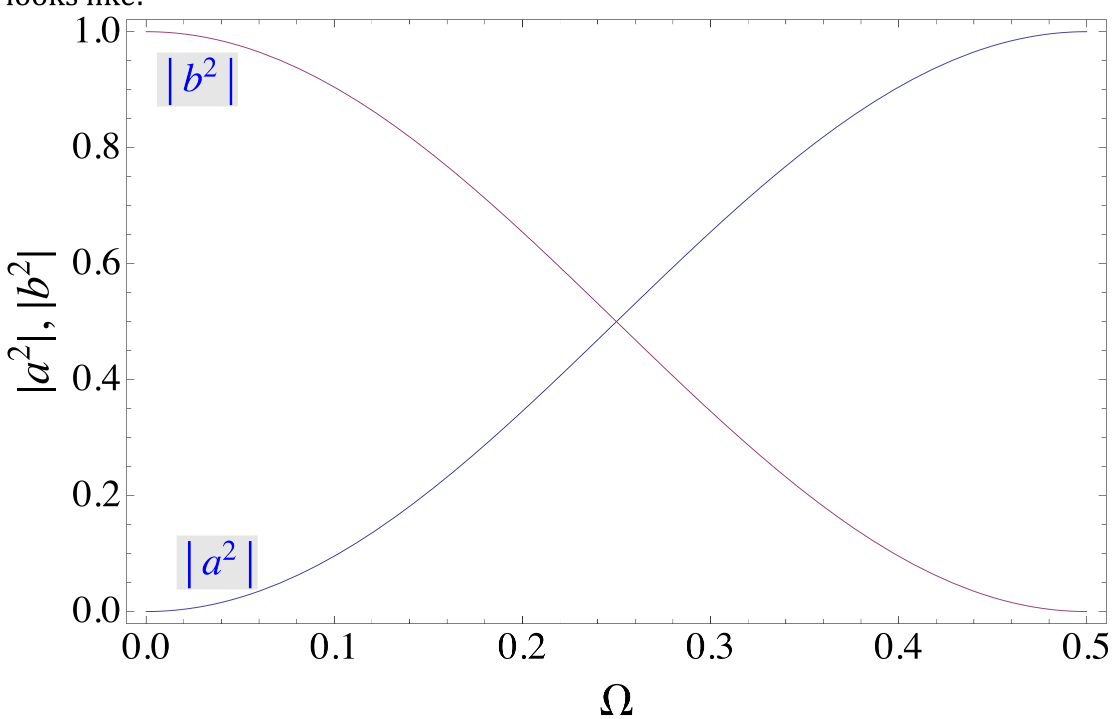
Therefore, $a = \cos\left(\frac{\Omega}{2}t + \phi\right)$ and $b = \sin\left(\frac{\Omega}{2}t + \phi\right)$.

So, if the population started in the ground state: $|a^2| = 1$ at *t = 0* and thus $\phi = 0$. The time evolution looks like:



This is absorption. The population starts in the ground state and moves to the excited state.

If the population started in the excited state $|a^2| = 0$ at $t = 0$ and thus $\phi = \frac{\pi}{2}$. This looks like:



This is stimulated emission. The population starts in the excited state and goes to the ground state!

A few words on this: You might be wondering exactly "where are the photons" in this treatment. Well, we didn't really put them in. We used a "classical" field amplitude – i.e. we let E be a continuous variable. To really do this problem correctly requires quantum electrodynamics, where the fields are also quantized. This is beyond the scope of this class, but the main feature is very similar in both treatments. The field drives transitions between the two states in both directions.

(END: Skip in lecture)

Thus, we can now see how to make our magical gain medium. We need to prepare a sample of atoms with more electrons in the excited state than in the ground state:

Then when photons pass through this medium they will cause spontaneous emission to happen and more photons will be generated.

Vocabulary word: When there is more population in an excited state than the ground state it is called an inversion

So, now, the only question is how do we create such a situation where all (or at least most!) of the population is in the excited state? Can we do it by say turning up the temperature? Well the Boltzmann factor for the two states are $P_g = 1$ and $P_e = e^{-\frac{\hbar\omega}{k_b T}}$, so the only way the excited state could have more population than the ground state is if the medium had a negative temperature! So, we probably need another way!

Perhaps we could use another light source like a lamp, to try to excite all of the ground state atoms into the excited state before we try to start the lasing. Well, since we just saw that absorption and stimulated emission both depend on the number of photons, then our lamp would stimulate some of the excited state atoms to emit a photon. And when you go through the math you find that no matter how bright your lamp is the best you can do is a 50/50 split between the ground and the excited states. And to make matters even worse, we have left out one very important physical phenomenon from our description: Spontaneous Emission!

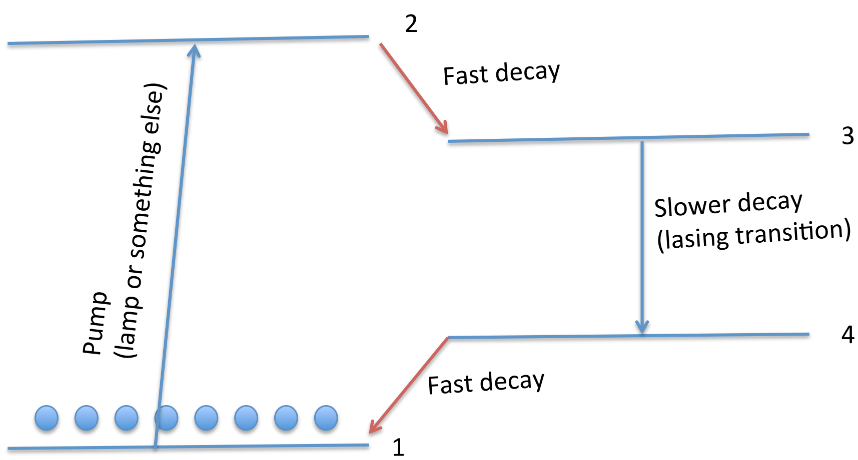
You have probably heard of spontaneous emission of a photon. In this phenomenon, an excited atom can spontaneously emit a photon and relax to the ground state. It turns out that this cannot be described in quantum mechanics if you use a classical field amplitude as we have above. To see spontaneous emission actually requires the use of QED. Though we cannot derive the spontaneous emission rate here we can state a few facts about it that are fairly apparent.

1. Unlike stimulated emission, which clearly happens faster for more photons (i.e. Ω gets bigger and therefore the rate of transition increases), spontaneous emission does not depend on any applied field (photons) and therefore happens at a rate given by the details of the atomic structure.
2. Spontaneous emission will be a "problem" for our laser, as it will fight our efforts to make an inversion.

So, the situation is really hopeless for our two level atom. Even without spontaneous emission we can never get more than a 50/50 split and so we have no hopes of making the necessary population distribution.

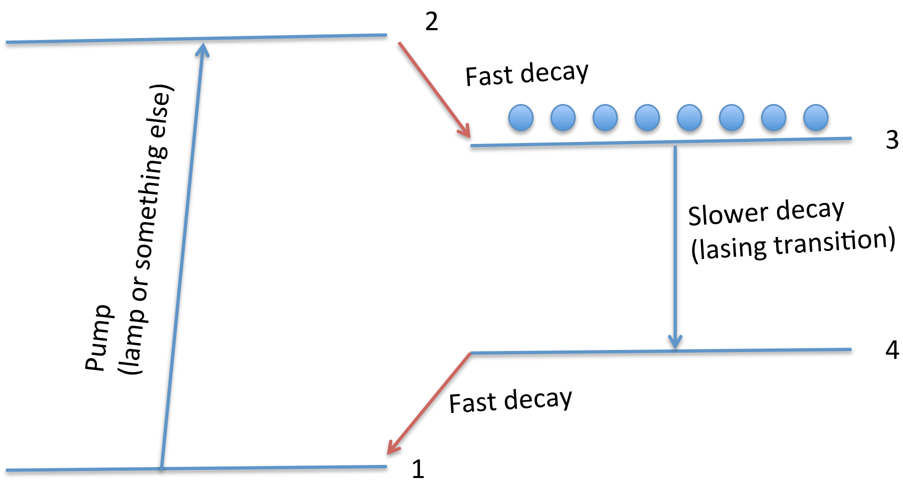
Well, we know lasers actually work so how do they do it? How do they make an inversion? The answer is that there are more than 2 levels in atoms!

The **classic** (there are other implementations) lasing scheme is a four-level system:



Obviously, to analyze this system quantum mechanically would be a nightmare and even then unless we did QED we wouldn't even get the right answer. So the best route is to use rate equations, which calculate the behavior of the population based on average rates. If necessary these average rates can be calculated from QM, but in practice you usually just look them up! And then use them as inputs into our rate equations, which are just ordinary differential equations.

We'll see this come out of the maths, but just looking at this you can probably already see how this works: because then slowest step in the process is the 3→ 4 decay, population can pile up in the 3 state and we can create an inversion:



Now, we have all of the basic ingredients to build a laser. We know that putting a gain medium inside of a cavity can lead to lasing. And we know how to make that gain medium. The only remaining question is really how much power will come out of our laser? As you can intuit from the initial discussion, the number of photons coming out of the laser is given by the balance of round trip gain and round trip loss. But in our simple model of the gain so far, it's hard to see how this would happen. Since it looks like once you get the gain> loss the photon number will grow without bound. Evidently, for high photon number the gain must decrease.

Now that we understand stimulated emission we see how this is. Once there are a lot of photons in the lasing mode, then they "deplete" the population in state 3 and that lessens the gain. We can treat this quantitatively with a rate equation model.

For simplicity we imagine the two fast decays occur infinitely fast (for a "good" laser this is good approximation). Then the number of atoms in the upper laser state (#3) is given as:

$$\dot{N}_3 = R - BIN_3 - \Gamma_{34}N_3$$

That is, the rate of change of the number in the excited state is given by the rate at which you pump atoms in minus the amount lost to stimulated and spontaneous emission. (The constant B comes from a quantum mechanics calculation in principle, but in practice you can just look it up. It is called the Einstein B coefficient and gives the probability for a stimulated emission or absorption event per unit intensity per unit population.) Note that, absorption of laser photons from N_4 normally gives a term $+BIN_4$, however, we assume that population in #4 decays infinitely fast, thus $N_4 = 0$ and this term disappears. This is a very good approximation for a **good** laser.

And the rate of change of intensity of the light in the cavity is given as:

$$\dot{I} = -\Gamma_c I + BIN_3$$

where Γ_c is the decay rate out of the cavity (which is just given as the loss per round trip*c/2L). Strictly speaking we should also include a spontaneous emission term into the cavity, since this is what seeds the lasing process. However, this term has no real effect on the laser intensity as soon as there is ~1 photon in the cavity, so we ignore it here to make the equations simpler.

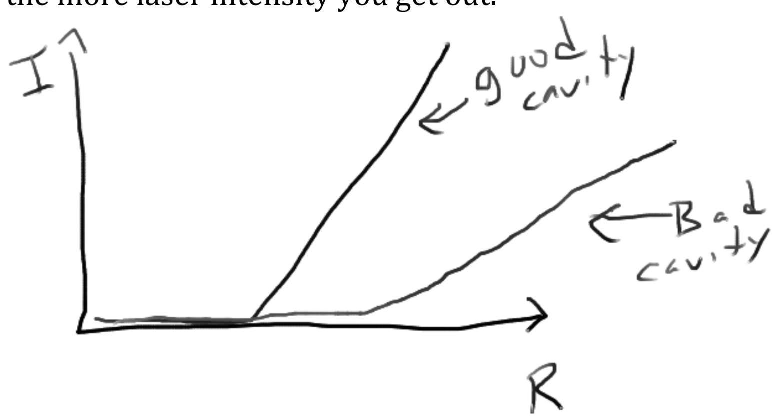
Now at equilibrium, $\dot{N} = 0$ and $\dot{I} = 0$, thus the laser intensity in the cavity is given as:

$$I = \frac{R - \Gamma_{34}N_3}{BN_3} = \frac{R - \frac{\Gamma_{34}\Gamma_c}{B}}{\Gamma_c} = \frac{R}{\Gamma_c} - \frac{\Gamma_{34}}{B}$$

Therefore, unless the pumping rate is greater than the loss rate to spontaneous emission, you get zero photons out.

Note that because $\dot{I} = 0$, the stimulated emission rate and the cavity loss rate are balanced at equilibrium. This makes sense because in solving the second equation we've implicitly assumed I is not zero, and if I is not zero, as the intensity increases the stimulated emission rate increases so that it always matches the cavity loss rate.

Once the pumping rate increases beyond the spontaneous loss rate, the laser intensity increases linearly with R. And the better cavity you make (small loss rate), the more laser intensity you get out.



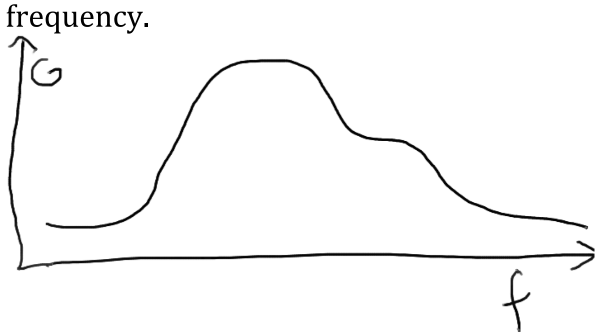
Vocabulary word: Laser threshold. The threshold is the pumping rate, where the lasing action "starts." (i.e. G=L).

Lasing frequency

So far we've seen how the lasing operation is established, but we have not said too much about the frequency of the laser, other than stating that it works at the cavity resonance frequency. Why is that? And can we be more precise?

The key to understanding the lasing frequency comes from analyzing the gain and losses as a function of frequency. Let's first look at the frequency dependence of the gain – sometimes called the gain profile.

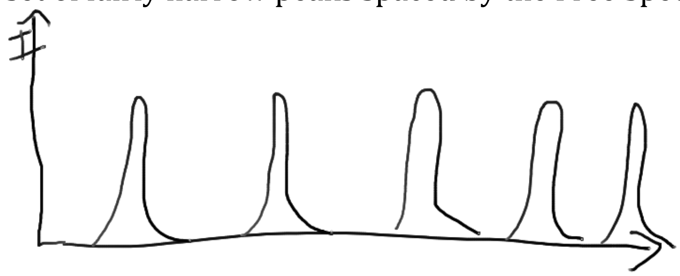
Well, in many laser systems the laser levels are between bands in a solid state device (for example, Ti:Sapphire). And even in lasers using a gaseous gain medium, things like Doppler and pressure broadening, not to mention the natural lifetime of the transition, work to broaden the transition energy so that it can occur at a spread of photon energies. Thus, in **every** laser gain can happen for more than one single frequency.



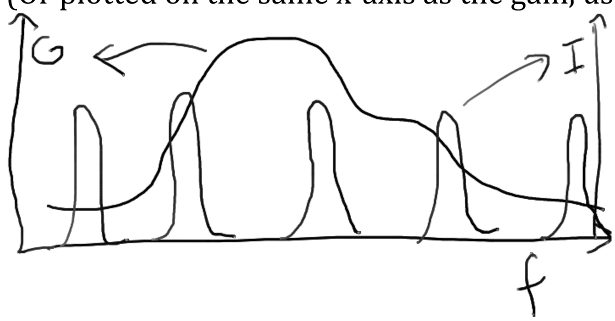
So, how does the laser "choose" which frequencies to operate at? To understand this requires analyzing the losses in the laser cavity.

Well what are the losses? Of course the mirror losses matter, but those tend to be fairly broadband. The biggest loss is the effect of the cavity itself, which will only allow certain wavelengths to build-up inside of it!

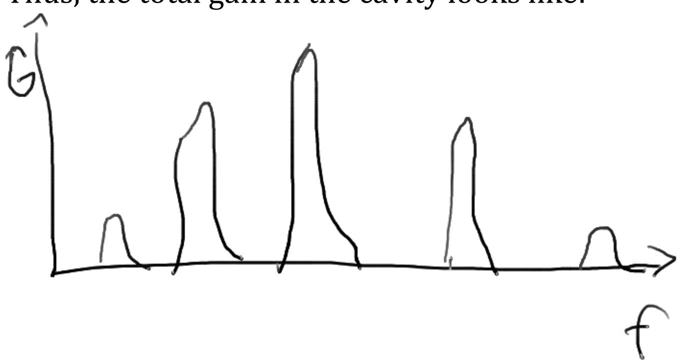
Recall that for a constant input power, the intensity inside of an optical cavity is a set of fairly narrow peaks spaced by the Free Spectral Range:



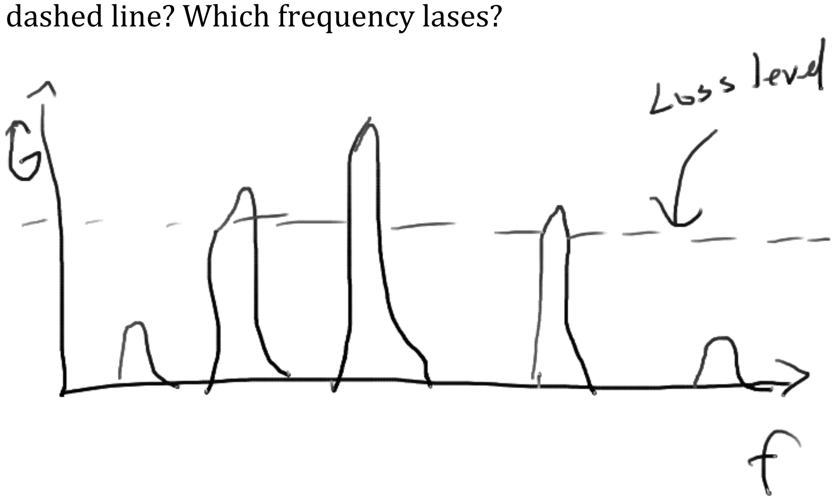
(Or plotted on the same x-axis as the gain, using a right y-axis:)



Thus, the total gain in the cavity looks like:



Now, suppose the loss due to things like transmission, absorption in the mirror coatings, or reflections from the gain medium are at the value shown below by the dashed line? Which frequency lases?



Well, there are three cavity modes that experience a gain greater than the total losses so in principle all three can lase! Whether or not they all lase depends on some details of the gain medium.

If the gain profile is due to homogeneous broadening (i.e. all the active atoms exhibit the same gain profile) then the three lasing modes are in competition for the same gain! And since the lasing steady state is reached when the stimulated emission balances the pumping rate, the mode with the highest gain can "steal" the gain from all of the other modes. Thus, the laser lases at the frequency of the cavity mode with the highest gain. This is sometimes called mode-competition or gain narrowing – caution though: these words can sometimes mean other things too. Now, does the laser really just operate at one single frequency? No, it will lase at a band of frequencies centered on the cavity mode, but with a linewidth given by the cavity linewidth.

If the gain profile is due to inhomogenous broadening (i.e. the active atoms exhibit different gain profiles that add up to total gain profile), then the three lasing modes are addressing different atoms and therefore they are not in competition. In this case, all three modes can lase at the same time.

Diode lasers and ECDL

In this week's lab we will be building a laser, where the active medium is a solid-state device, known as a diode laser. Diode lasers typically have enormous gain compared to other laser systems and are much simpler as the inversion can be created by simply passing an electrical current through the device! For these reasons, they are ubiquitous in both science and technology (CD players, etc.)

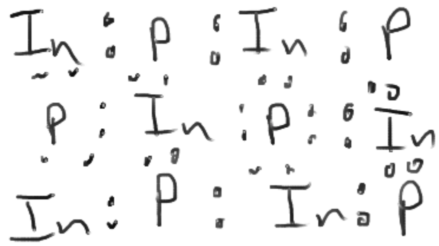
There are many classes of laser diodes (Fabry Perot, DBR, DFB) and these classes have many variation (ridge waveguide, submerged structures, etc.), but they all work on one basic principle: Inject electrons into a n-doped material and holes into a p-doped material and allow them to recombine in a junction. When they recombine you get light. The earliest laser diodes were homojunction devices – just a single p-n junction – but these suffer from two main difficulties:

1. The light is not guided inside the material.
2. The current density is difficult to contain so the gain region varies between devices.

Both of these problems were solved by the heterojunction diode, sometimes called p-i-n or pin diodes. These diodes have an intrinsic layer (no-doping) between the p and n layers where the recombination (gain happens).

But first, in case you don't remember what p and n doping means, let's look at InP semiconductor:

InP undoped

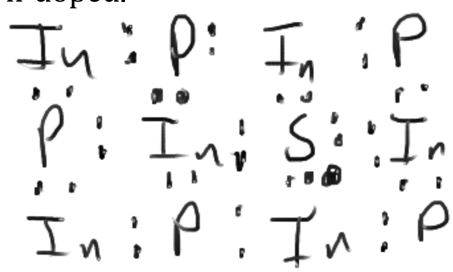


In the undoped case all the electrons are bound to an ion core, so charge can't move (no current can flow = insulator or at least a semiconductor)

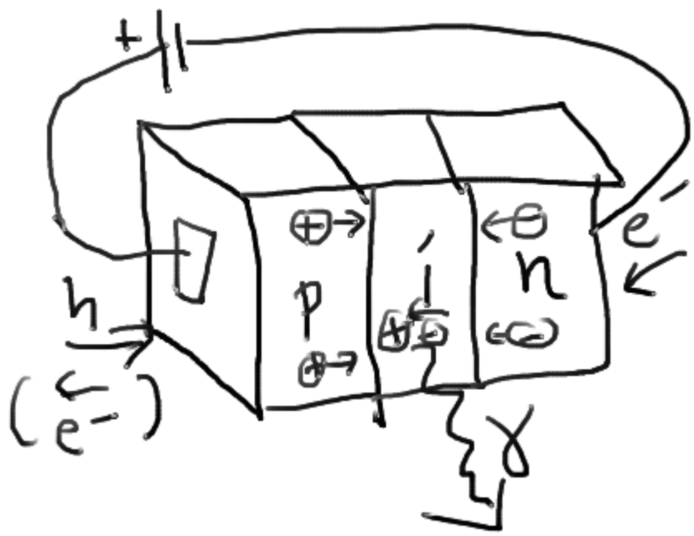
With doping, either holes or electrons can conduct current:

p-doped

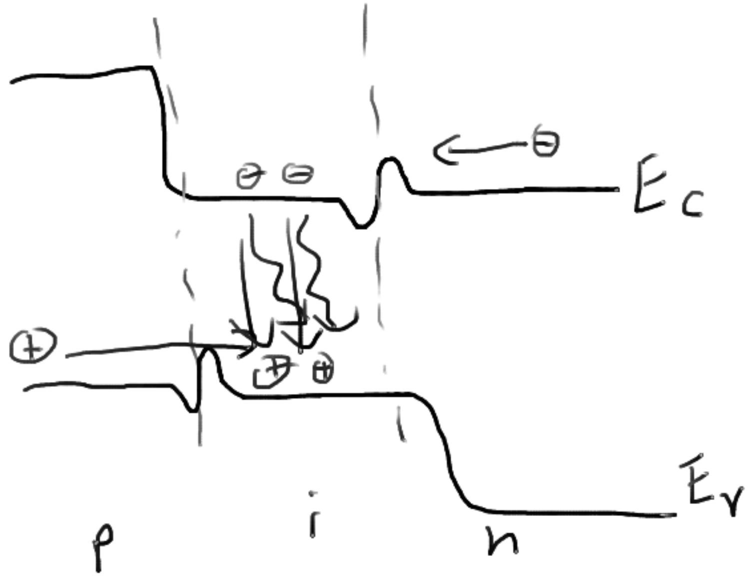
n-doped.



With that out of the way, here's what a pin laser diode looks like.

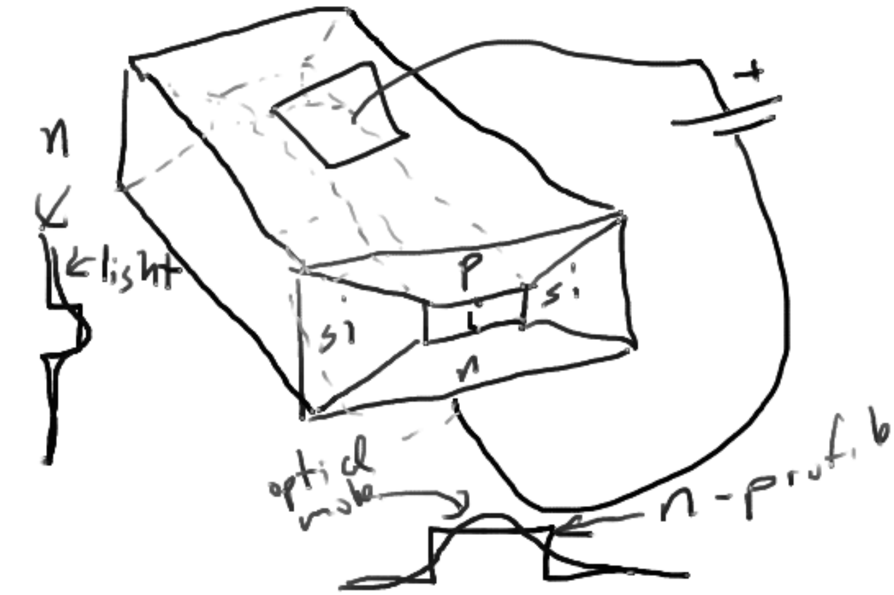


And the band structure throughout the device looks roughly like:



Because the recombination happens in the intrinsic layer the region of gain is now well defined. And because doping (p or n reduces) usually reduces the index of refraction, the device has a larger index of a refraction in the middle and thus, just like a fiber, we have now made a waveguide. This keeps the light together inside of the diode. (The fact that index of refraction decreases with doping can be seen from the behavior of the bandgap. The bandgap grows with doping generally and thus the wavelength of the emitted light in the intrinsic region is farther from “resonance” in the doped region. And by Kramers-Kronig, the index is closer to 1.)

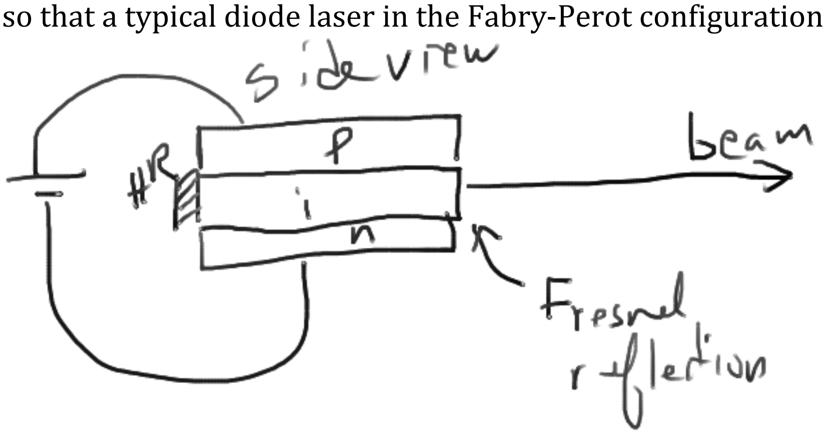
In the simplest implementation, these a laser is constructed from a pin device in the following way:



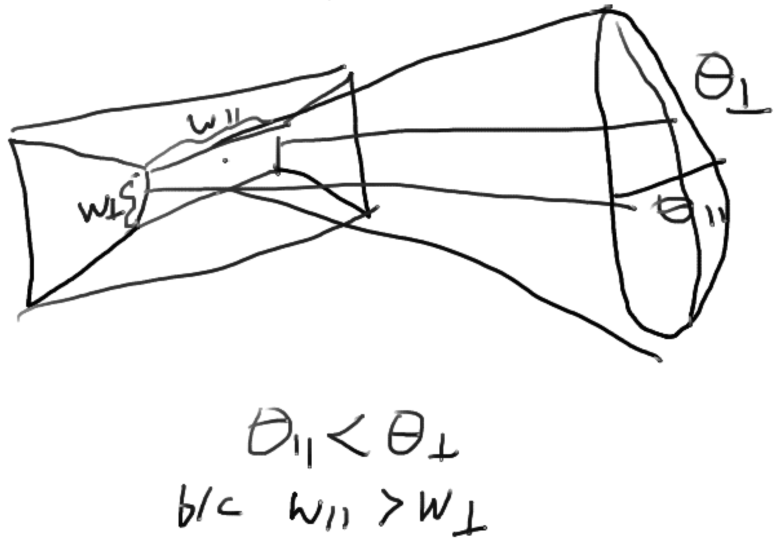
The si layer (semi-insulating) has lower index of refraction than the i layer and makes the waveguide work in both the vertical and horizontal directions.

And because the i layer has such a large index of refraction (n = 3-4), there is enough reflection at the edge of the device that you don't even need to put a mirror there to make a laser cavity. Recall the Fresnel reflection coefficient for perpendicular incidence is: $R = \frac{(n-1)^2}{(n+1)^2}$, which for n ~3.5 is close to 30%. Of course, this makes a bad cavity with a lot of loss, but the pin diode can give an enormous amount of gain.

Typically, though, the backside is coated with a mirror coating to improve efficiency, so that a typical diode laser in the Fabry-Perot configuration looks like:

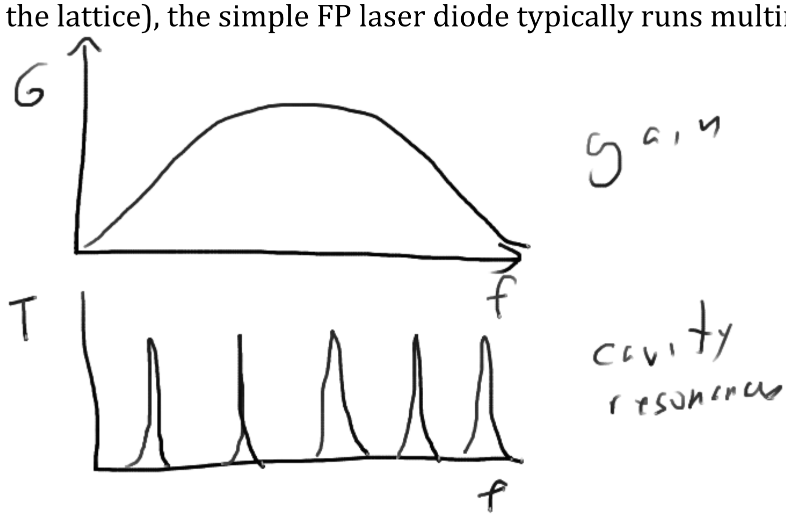


And because of the pin geometry the output beam has an elliptical shape:



The laser oscillation is usually polarized along the || direction so the polarization comes out parallel to the minor axis of the beam ellipse.

Because the gain profile of a semiconductor device is quite large (think band structure) and is inhomogenously broadened (recombination at different parts in the lattice), the simple FP laser diode typically runs multimode:



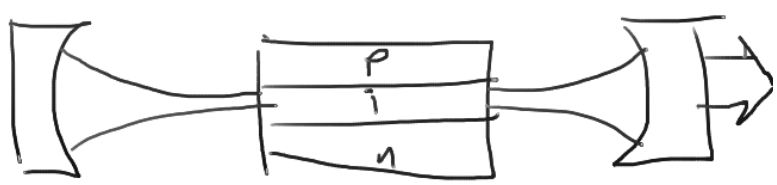
This is called multi-mode operation. In practice, as you turn up the current there is a bit of mode-competition and you will see some gain narrowing.

Finally, because the cavity length and index both depend on the temperature of the structure and the current passing through it (current causes a change in index through changing the carrier densities and by changing the temperature), you can tune the wavelength of these lasers according to the FP resonance condition: $\frac{n(T,I)\lambda}{2} = L(T)$.

You will play with this in lab.

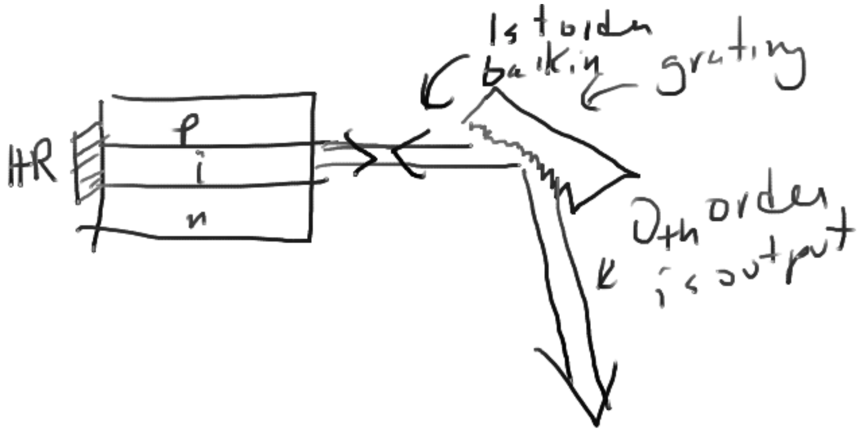
External Cavity diode laser

A slightly more complicated, but much better performing laser can be made from these pin diodes if they are placed inside an external cavity:



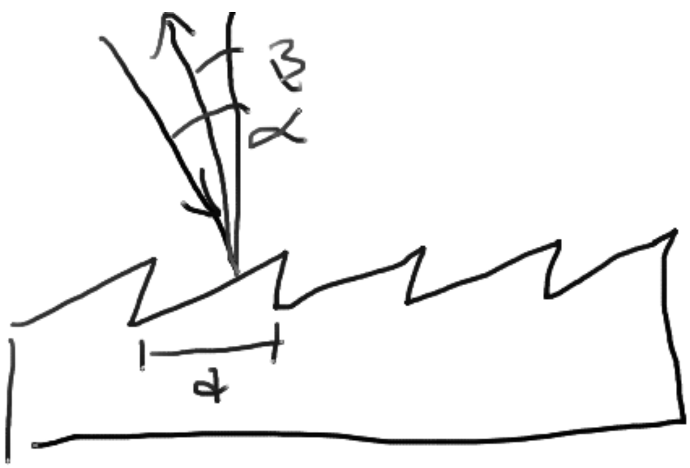
In this way, single mode operation can be achieved and the linewidth will be much reduced as long as the external cavity is a "good" cavity – i.e. high Finesse.

Though the above picture looks quite simple, in practice it requires quite careful alignment. Another way to achieve effectively the same end, but with a much simpler system is to use a grating to provide wavelength selection. There are several different implementations, but the simplest (and best for most things) is the Littrow configuration:

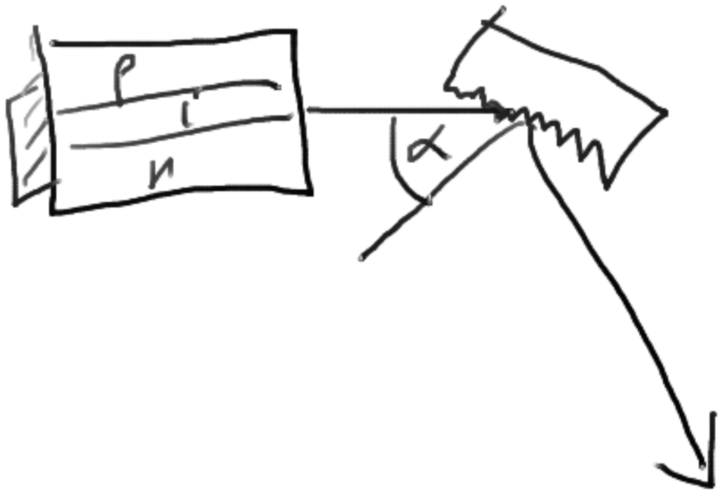


In this configuration, a diffraction grating is used to refract light of a chosen wavelength back into the laser diode. In this way, the cavity is formed by the back high reflectivity mirror (HR) and the grating for only one wavelength and the multimode spectrum from before becomes single mode.

Recall the grating refracts a beam at angle according to:
$d(\sin\alpha + \sin\beta) = m\,\lambda$



For the Littrow configuration $\alpha = \beta$. Thus, first order beam is refracted back into the diode when the angle between the grating normal and the incoming beam is $2d\sin\alpha = m\lambda$. And the zeroth order (i.e. just regular reflection from the grating) can be used as the output beam!



With this, the net gain can be inferred as follows:

Thus, only the middle mode sees enough gain to lase!

Technically, we also need to consider the mode spacing of the cavity formed by HR mirror and the grating. This is a much smaller FSR than of the FP cavity and depending on how you rotate the grating you can get hops between the different modes of the longer cavity. These hops are discontinuous jumps of the laser frequency as you try to tune it. Using special geometries, which we will not consider, these can be mitigated.