OPTO Prelim Notes

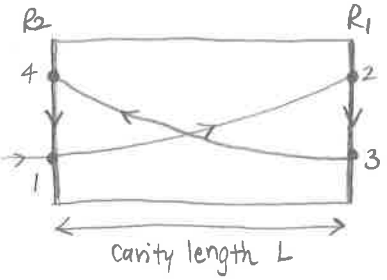
Patrick Xiao

Table of Contents

The primary section numbers are marked by boxes in the following pages. Sections 1-11 mostly cover EE232 material, sections 12&13 are a few selected topics from EE230A and EE218.

Prelim Review

① $\boxed{\text{Gain condition in a laser}}$



A mode is sustained in a cavity if the round-trip gain is balanced exactly with loss:

$$\underbrace{e^{(\Gamma g L - \alpha_i L)} R_1}_{\text{first pass \& reflection}} \; \underbrace{e^{(\Gamma g L - \alpha_i L)} R_2}_{\text{second pass \& reflection}} = 1$$

g = gain
α_i = intrinsic loss
Γ = confinement factor

What is the minimum gain required?

$$e^{2(\Gamma g L - \alpha_i L)} = \frac{1}{R_1 R_2} \Rightarrow \boxed{g_{th} = \frac{1}{\Gamma}\left(\alpha_i + \frac{1}{2L}\ln\frac{1}{R_1 R_2}\right)}$$

gain at threshold; α_i → intrinsic loss; $\frac{1}{2L}\ln\frac{1}{R_1R_2}$ → mirror loss (useful; gives output beam)

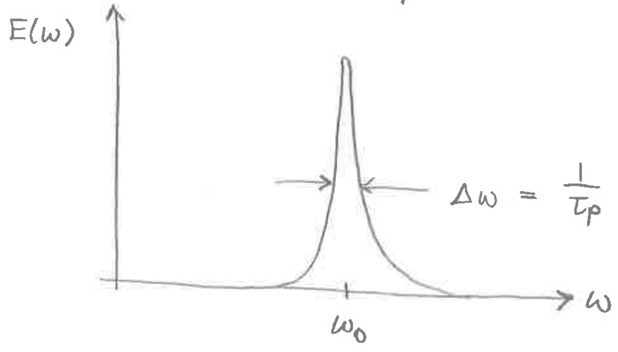
- Photon lifetime: consider a photon that is traveling inside a cavity. It is lost from the cavity after a characteristic time τ_p

$$I(t) = I_0 e^{-t/\tau_p}$$

$$E(t) = E_0 e^{-i\omega_0 t} e^{-t/2\tau_p}$$

$$FT(E(t)) = \frac{1}{i(\omega - \omega_0) + 1/2\tau_p}$$

(Lorentzian) ↖ $\Delta\omega/2$



- The Q-factor is:

$$Q = \frac{\omega}{\Delta\omega} \Rightarrow \boxed{Q = \omega\tau_p} \quad \text{and also note} \quad \frac{1}{\tau_p} = \alpha\frac{c}{n}$$

total absorption/loss coefficient ↙ α

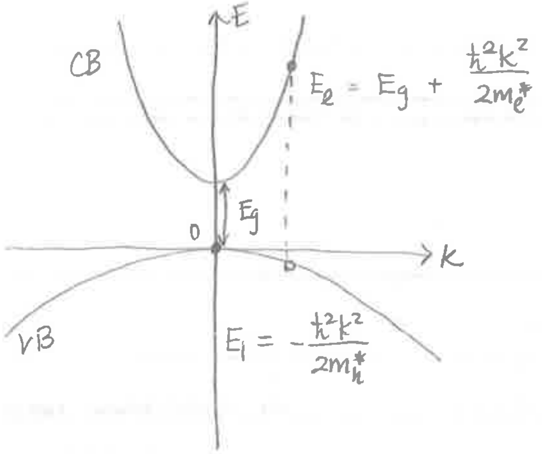
- Express the threshold gain in terms of the photon lifetime:

$$\underbrace{\Gamma g_{th} \frac{c}{n}}_{\text{gain rate}} = \underbrace{\frac{1}{\tau_p}}_{\text{loss rate}} \Rightarrow g_{th} = \frac{1}{v_g \Gamma \tau_p} = \frac{1}{v_g\Gamma}\frac{\omega}{Q}$$

$\frac{c}{n}$: group velocity v_g

The threshold gain is independent of diode physics, it is simply equating gain with loss

② Basic solid-state physics



In an optical absorption/emission process,

1) Energy conservation $\Rightarrow$ $E_2 - E_1 = \hbar\omega$

2) Momentum conservation:

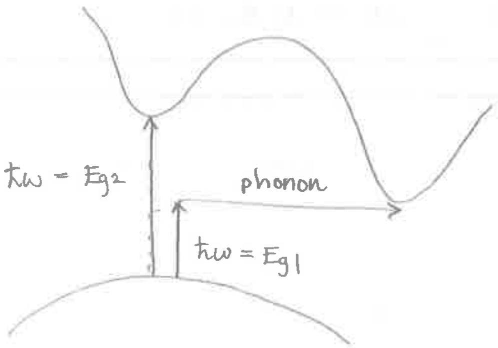
$$k_2 - k_1 = \frac{2\pi}{\lambda}$$

- Electronic wave-number $k_1, k_2 \sim \frac{2\pi}{a}$, $a \sim 0.5\text{nm}$
- Photon wavenumber $\frac{2\pi}{\lambda}$, $\lambda \sim 1000\text{nm}$

Thus, the photon momentum is negligible to the size of the Brillouin zone in k-space, and we have the condition $k_1 \approx k_2$

$\Rightarrow$ optical transitions are vertical in E-K

Indirect bandgap:



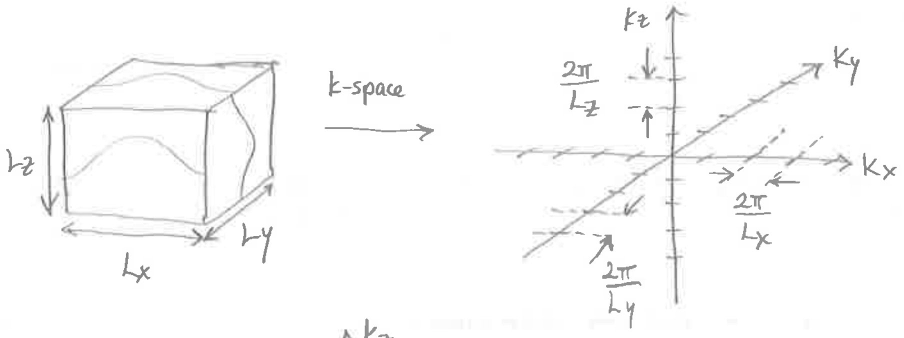
- Indirect transitions are not vertical, and require the participation of a phonon to conserve momentum

  $\Rightarrow$ much less likely
  lower absorption coefficient

- Ge has a small indirect bandgap but can be alloyed with Sn to bring the direct bandgap below the indirect bandgap

- Electronic density of states (3D)

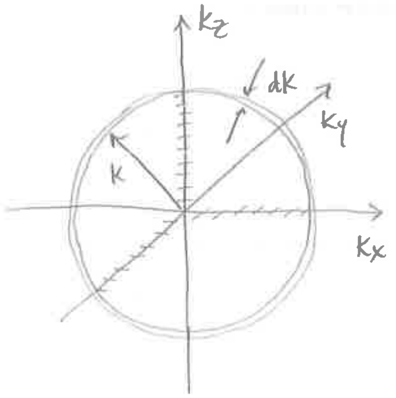
Consider a box of volume $V = L_x L_y L_z$ in which electronic wavefunctions exist:



- The DOS in k-space is:

$$dN = \frac{d^3k}{\left(\frac{2\pi}{L_x}\right)\left(\frac{2\pi}{L_y}\right)\left(\frac{2\pi}{L_z}\right)} \times \overset{\text{* spins}}{2}$$

- Integrate over a spherical shell representing a wavenumber $k = |\vec{k}|$



$$dN = \frac{4\pi k^2\, dk}{\left(\frac{2\pi}{L_x}\right)\left(\frac{2\pi}{L_y}\right)\left(\frac{2\pi}{L_z}\right)} \times 2 = \frac{k^2 dk}{\pi^2} V$$

- Per volume, $n \equiv N/V$, we have

$$\boxed{\frac{dn}{dk} = \frac{k^2}{\pi^2}}$$

- Convert k to energy

- kinetic energy $E' = \frac{\hbar^2 k^2}{2m}$ $\Rightarrow$ $\frac{dE'}{dk} = \frac{\hbar^2}{m} k$, $k = \frac{\sqrt{2mE'}}{\hbar}$, $dk = \frac{m}{\hbar^2 k} dE'$

carrier mass

$$\Rightarrow \quad dn = \frac{k^2}{\pi^2} dk = \frac{k^2}{\pi^2} \frac{m}{\hbar^2 k} dE' = \frac{m}{\pi^2 \hbar^2} k \, dE'$$

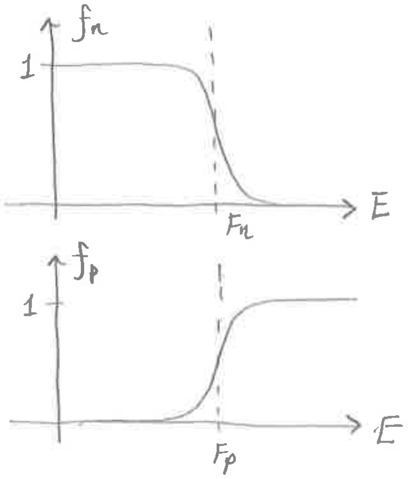
$$\frac{dn}{dE'} = \frac{m}{\pi^2 \hbar^2} \frac{\sqrt{2mE'}}{\hbar} = \frac{1}{2\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \sqrt{E'}$$

- The kinetic energy expressed in terms of absolute energy is:

$E' = E - E_c$ for electrons, $E' = E_v - E$ for holes

$$\Rightarrow \quad \boxed{\rho_e(E) = \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2} \right)^{3/2} \sqrt{E - E_c}} \qquad \boxed{\rho_h(E) = \frac{1}{2\pi^2} \left(\frac{2m_h^*}{\hbar^2} \right)^{3/2} \sqrt{E_v - E}}$$

- Fermi-Dirac distribution: the occupancy of an electronic energy state is ≤ 1 because of the Pauli exclusion principle



$$f_n(E) = \frac{1}{e^{(E - F_n)/kT} + 1} \quad \leftarrow \text{electron occupancy}$$

$$f_p(E) = \frac{1}{e^{(F_p - E)/kT} + 1} \quad \leftarrow \text{hole occupancy}$$

The hole occupancy is $f_p = 1 - f_n(E)$ with F_p as the Fermi level, since a state that is occupied by a hole is a valence band state that is not occupied by an electron

- Carrier densities

$$n = \int_{E_c}^{\infty} \rho_e(E) f_n(E) \, dE = \underbrace{2 \left(\frac{m_e^* kT}{2\pi\hbar^2} \right)^{3/2}}_{N_c} F_{1/2}\left(\frac{F_n - E_c}{kT} \right)$$

$$p = \int_{-\infty}^{E_v} \rho_h(E) f_p(E) \, dE = \underbrace{2 \left(\frac{m_h^* kT}{2\pi\hbar^2} \right)^{3/2}}_{N_v} F_{1/2}\left(\frac{E_v - F_p}{kT} \right)$$

same expressions for holes

$$\text{where } F_{1/2}(\eta) \approx \begin{cases} e^{\eta} & \text{when } \eta < 1 \quad \longrightarrow \quad \boxed{n = N_c e^{-\frac{E_c - F_n}{kT}}} \quad \text{Boltzmann} \\ \frac{4}{3} \left(\frac{\eta^3}{\pi} \right)^{1/2} & \eta > 1 \quad \longrightarrow \quad \boxed{n = N_c \frac{4}{3\sqrt{\pi}} \left(\frac{F_n - E_c}{kT} \right)^{3/2}} \quad \text{degenerate} \end{cases}$$

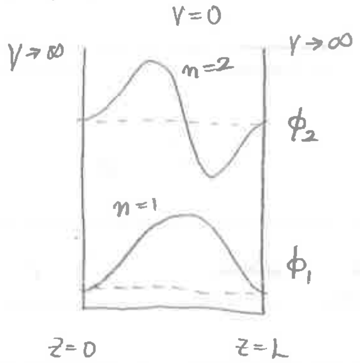
Electronic density of states in 2D

- Quantum well: structure in which electron is confined along one dimension, z

$$\psi = \frac{1}{\sqrt{A}} e^{ik_x x} e^{ik_y y} \phi(z) e^{-i\omega t}$$

$$\swarrow \qquad \searrow$$

$$E_x' = \frac{\hbar^2 k_x^2}{2m} \qquad E_y' = \frac{\hbar^2 k_y^2}{2m}$$



- To find the energy along z, solve Schrodinger's equation: $E\phi(z) = -\frac{\hbar^2}{2m}\frac{d^2\phi}{dz^2} + V(z)\phi(z)$

Let the well be infinite, so $V(z) = 0$ for $0 < z < L$, $V(z) = \infty$ elsewhere

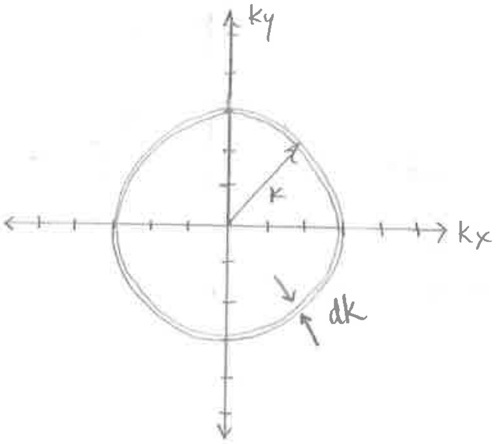
The solution is:

$$\phi_n(z) = \sqrt{\frac{2}{L_z}} \sin\left(\frac{n\pi}{L_z} z\right) \qquad n = 1, 2, 3, \ldots$$

$$E_{nz}' = \frac{\hbar^2}{2m}\left(\frac{n\pi}{L_z}\right)^2$$

- The total kinetic energy is: $E_n' = \frac{\hbar^2}{2m}\left[\underbrace{k_x^2 + k_y^2}_{\text{continuous}} + \underbrace{\left(\frac{n\pi}{L_z}\right)^2}_{\text{quantized}}\right]$

Within a subband (specific n) the DOS is found by integrating over the k-space of the unconfined directions k_x, k_y



$$dn = \frac{1}{V}\frac{2\pi k\, dk}{\left(\frac{2\pi}{L_x}\right)\left(\frac{2\pi}{L_y}\right)} \times 2 = \frac{k}{\pi L_z} dk$$

- From dispersion relation, $dk = \frac{m}{\hbar^2 k} dE'$, $k = \frac{\sqrt{2mE'}}{\hbar}$

$$dn = \frac{1}{\pi L_z} \cancel{k} \frac{m}{\hbar^2 \cancel{k}} dE' \Rightarrow \frac{dn}{dE'} = \frac{m}{\pi\hbar^2 L_z}$$

- Therefore, the density of states of each subband is:

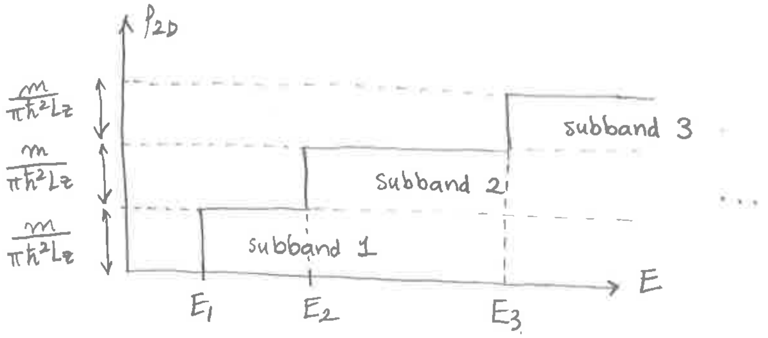
$$\rho_n^{2D}(E) = \frac{m}{\pi\hbar^2 L_z} \times \underset{\text{Step function}}{H}(E > E_n)$$

$$E_n = \frac{\hbar^2}{2m}\left(\frac{n\pi}{L_z}\right)^2$$

ground state energy of subband n: $E_n'(k_x = k_y = 0)$

- Sum over subbands: $$\boxed{\rho^{2D} = \sum_n \frac{m}{\pi\hbar^2 L_z} \times H(E > E_n)}$$

The quantum well DOS consists of a series of steps:



- Electron concentration in a quantum well

First, integrate only the first subband:

$$N_1 = \int_{E_1}^{\infty} \frac{m}{\pi\hbar^2 L_z} \cdot \frac{1}{e^{(E-F_c)/kT}+1}\, dE$$

$$N_1 = \frac{mkT}{\pi\hbar^2 L_z} \ln\left(1 + e^{(F_c - E_1)/kT}\right)$$

Then sum over all subbands:

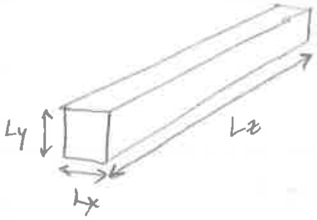
$$\boxed{N = \sum_n N_n = \frac{m_e^* kT}{\pi\hbar^2 L_z} \sum_n \ln\left(1 + e^{(F_c - E_n)/kT}\right)}$$

$$\boxed{P = \sum_m P_m = \frac{m_h^* kT}{\pi\hbar^2 L_z} \sum_m \ln\left(1 + e^{(E_m - F_v)/kT}\right)}$$

If $\frac{F_c - E_n}{kT} \gg 1$ $\left(\text{or } \frac{E_m - F_v}{kT} \gg 1\right)$, the exponential term dominates inside the logarithm and the carrier density in the subband is linear with $F_c - E_{en}$ (or $E_m - F_v$)

- Quantum wire : confinement in two dimensions

$$\psi \sim e^{ik_z z}\,\phi(x)\,\varphi(y)$$



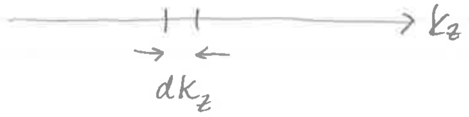
The wire can be approximately decomposed into orthogonal quantum wells with energies

$$E_{nx} = \frac{\hbar^2}{2m}\left(\frac{n\pi}{L_x}\right)^2 , \quad E_{my} = \frac{\hbar^2}{2m}\left(\frac{m\pi}{L_y}\right)^2$$

$$\Rightarrow \quad E' = \frac{\hbar^2}{2m}\left[\left(\frac{n\pi}{L_x}\right)^2 + \left(\frac{m\pi}{L_y}\right)^2 + k_z^2\right]$$

- There is one free dimension with dispersion relation $E_z' = \frac{\hbar^2}{2m} k_z^2$

Treat a 1D k-space



$$dn = \frac{1}{V}\,\frac{dk_z}{\left(\frac{2\pi}{L_z}\right)} \times 2 = \frac{2}{\pi L_x L_y}\,dk_z$$

$$= \frac{2}{\pi L_x L_y} \cdot \frac{m}{\hbar^2}\,\frac{\hbar}{\sqrt{2mE_z'}}\,dE_z$$

$$\Rightarrow \quad \frac{dn}{dE_z} = \left(\frac{2m}{\hbar^2}\right)^{1/2} \frac{1}{\pi L_x L_y}\,\frac{1}{\sqrt{E' - E_{nx} - E_{my}}}$$

Therefore,

$$\boxed{\rho_{mn}^{1D} = \frac{1}{\pi L_x L_y}\left(\frac{2m}{\hbar^2}\right)^{1/2} \frac{1}{\sqrt{E' - E_{nx} - E_{my}}}}$$

③ Optical absorption and gain in bulk semiconductors

- Time-dependent perturbation theory : Hamiltonian

  - Conventional Hamiltonian. $H_0 = \frac{p^2}{2m} + V(\vec r)$

  - Upon interaction with an electromagnetic wave, $\vec p \rightarrow \vec p - q\vec A$ ← vector potential of EM wave

$$\Rightarrow H = \frac{(\vec p - q\vec A)\cdot(\vec p - q\vec A)}{2m} + V(\vec r)$$

$$\mathcal{H} = \frac{p^2}{2m} \underbrace{- \frac{q}{2m}(\vec p\cdot\vec A + \vec A\cdot\vec p) + \frac{q^2A^2}{2m}}_{\mathcal{H}'} + V(\vec r) = \mathcal{H}_0 + \mathcal{H}'$$

  - Perturbed Hamiltonian is:

$$\mathcal{H}' = -\frac{q}{2m}(\vec p\cdot\vec A + \vec A\cdot\vec p) + \underbrace{\cancel{\frac{q^2A^2}{2m}}}_{0}$$

since A is a small perturbation

Simplify:

$$(\vec p\cdot\vec A)\psi = -i\hbar\vec\nabla\cdot(\vec A\psi) = -i\hbar\left[\overset{\text{Coulomb gauge}}{\cancel{(\vec\nabla\cdot\vec A)}}\psi + \vec A\cdot\nabla\psi\right]$$

$$= \vec A\cdot(-i\hbar\vec\nabla\psi) = (\vec A\cdot\vec p)\psi$$

$$\Rightarrow \mathcal{H}' = -\frac{q}{m}\vec A\cdot\vec p$$

  - For an electromagnetic wave of a single frequency, set $\vec A$ to be a plane wave:

$$\vec A = \hat e\frac{A_0}{2}e^{i(\vec k\cdot\vec r - \omega t)} + \hat e\frac{A_0}{2}e^{-i(\vec k\cdot\vec r - \omega t)}$$

$\uparrow$ polarization vector

$$\Rightarrow \mathcal{H}' = -\frac{q}{m}\frac{A_0}{2}\left[\hat e\cdot\vec p\, e^{i\vec k\cdot\vec r}e^{-i\omega t} + \hat e\cdot\vec p\, e^{-i\vec k\cdot\vec r}e^{i\omega t}\right]$$

$$= \mathcal{H}'(\vec r)e^{-i\omega t} + \mathcal{H}'^{+}(\vec r)e^{i\omega t}$$

- Apply the Hamiltonian: use Schrodinger's equation.

$$H\psi(\vec{r},t) = i\hbar\frac{\partial}{\partial t}\psi(\vec{r},t) \qquad \text{where} \qquad \psi(\vec{r},t) = \sum_n \underset{\substack{\uparrow \\ \text{eigenstates of } \mathcal{H}_0}}{a_n\,\phi_n(\vec{r})}\, e^{-i\frac{E_n}{\hbar}t}$$

$$(H_0 + H')\sum_n a_n \phi_n e^{-i\frac{E_n}{\hbar}t} = i\hbar\sum_n \frac{da_n}{dt}\phi_n e^{-i\frac{E_n}{\hbar}t} + i\hbar\sum_n a_n\phi_n\left(-i\frac{E_n}{\hbar}\right)e^{-i\frac{E_n}{\hbar}t}$$

We know that $H_0\sum_n a_n\phi_n e^{-iE_nt/\hbar} = i\hbar\sum_n a_n\phi_n\left(-i\frac{E_n}{\hbar}\right)e^{-iE_nt/\hbar}$ since we have used the eigenstates of $\mathcal{H}_0$

$$\Rightarrow \quad \mathcal{H}'\sum_n a_n\phi_n e^{-i\frac{E_n}{\hbar}t} = i\hbar\sum_n\frac{da_n}{dt}\phi_n e^{-i\frac{E_n}{\hbar}t}$$

Let $|n\rangle = \phi_n(\vec{r})$ and multiply both sides by $\langle m|$ to turn this into an overlap integral:

$$\sum_n a_n\langle m|H'|n\rangle e^{-i\frac{E_n}{\hbar}t} = i\hbar\sum_n\frac{da_n}{dt}\underset{=\delta_{mn}}{\langle m|n\rangle}e^{-i\frac{E_n}{\hbar}t} = i\hbar\frac{da_m}{dt}e^{-i\frac{E_m}{\hbar}t}$$

$$\frac{da_m(t)}{dt} = \frac{1}{i\hbar}\sum_n a_n\langle m|H'|n\rangle e^{-i\frac{E_n}{\hbar}t}e^{+i\frac{E_m}{\hbar}t}$$

① $$\frac{da_m(t)}{dt} = \frac{1}{i\hbar}\sum_n a_n(t)H'_{mn}e^{-i\omega_{mn}t}, \qquad \omega_{mn} = \frac{E_m - E_n}{\hbar}$$

- Use perturbation theory: $\mathcal{H} = \mathcal{H}_0 + \underset{\substack{\uparrow \\ \text{perturbation parameter}}}{\lambda}\mathcal{H}'$

$$a_n(t) = a_n^{(0)}(t) + \lambda a_n^{(1)}(t) + \lambda^2 a_n^{(2)}(t) + \dots$$

In applying eq. ①, we use the eigenstate amplitudes $a_n(t)$ of the lower order to get the $da_m(t)/dt$ of the higher order on the left side. So:

$$\frac{da_m^{(1)}}{dt} = \frac{1}{i\hbar}\sum_n a_n^{(2)}(t)\mathcal{H}'_{mn}e^{-i\omega_{mn}t}$$

$$\frac{da_m^{(2)}}{dt} = \frac{1}{i\hbar}\sum_n a_n^{(1)}(t)\mathcal{H}'_{mn}e^{-i\omega_{mn}t}$$

- Matrix element $\mathcal{H}_{fi}$

$$\mathcal{H}'_{fi} = \langle f | \mathcal{H}' | i \rangle$$

$$= \langle f | -\frac{qA_0}{2m} \hat{e}\cdot\vec{p}\, e^{i\vec{k}\cdot\vec{r}} | i \rangle = -\frac{qA_0}{2m} \langle f | \hat{e}\cdot\vec{p}\, e^{i\vec{k}\cdot\vec{r}} | i \rangle$$

. Invoke the dipole approximation: over the size of an electric dipole (or atom), the EM wave does not change much, again since $\lambda \gg a$ (lattice period) So we can say the wave has a constant phase across the dipole and eliminate the $e^{i\vec{k}\cdot\vec{r}}$ term:

$$\mathcal{H}'_{ba} = -\frac{qA_0}{2m} \langle b | \hat{e}\cdot\vec{p} | a \rangle = -\frac{qA_0}{2m} \hat{e}\cdot\vec{p}_{ba}$$

An alternative way to write this is in terms of the dipole moment:

$$\vec{p} = m\frac{d}{dt}\vec{r} \underset{\text{Ehrenfest's theorem}}{=} \frac{m}{i\hbar}[\vec{r}, H_0] = \frac{m}{i\hbar}(\vec{r}H_0 - H_0\vec{r})$$

$$\langle b | \vec{p} | a \rangle = \frac{m}{i\hbar} \langle b | \vec{r}H_0 - H_0\vec{r} | a \rangle = \frac{m}{i\hbar}\left[\langle b | \vec{r}\overbrace{H_0 | a\rangle}^{E_a|a\rangle} - \overbrace{\langle b | H_0}^{\langle b|E_b} \vec{r} | a \rangle\right]$$

$$= \frac{m}{i\hbar}(E_a - E_b)\langle b | \vec{r} | a \rangle = \frac{m_0\omega}{i}\vec{r}_{ba}$$

$$\mathcal{H}'_{ba} = -\frac{q}{m_0}\vec{A}\cdot(-im_0\omega)\vec{r}_{ba} = q(i\omega\vec{A})\cdot\vec{r}_{ba}$$

Since $\vec{E} = -\frac{\partial\vec{A}}{\partial t} = -i\omega\vec{A}$, we have

$$\boxed{\mathcal{H}'_{ba} = -\vec{E}\cdot\underbrace{q\vec{r}_{ba}}_{\text{dipole moment}} = -\frac{qA_0}{2m}\hat{e}\cdot\vec{p}_{ba}}$$

In a bulk semiconductor, the matrix element is isotropic (does not depend on polarization). An estimate of the magnitude from Kane's model gives $|\vec{r}_{ba}| \approx 0.5\,\text{nm}$

- Use first-order perturbation theory: let the initial state be

$$a_m^{(0)}(t) = \begin{cases} 1 & \text{if } m = i \\ 0 & \text{if } m \neq i \end{cases}$$ (i.e. initial wavefunction is an eigenstate of $\mathcal{H}_0$)

$$\frac{da_f^{(1)}(t)}{dt} = \frac{1}{i\hbar} \mathcal{H}'_{fi}(t)\, e^{i\omega_{fi} t} = \frac{1}{i\hbar}\left[\mathcal{H}'_{fi} e^{-i\omega t} + \mathcal{H}'^{\dagger}_{fi} e^{i\omega t} \right] e^{i\omega_{fi} t}$$

$$= \frac{1}{i\hbar}\left[\mathcal{H}_{fi}\, e^{i(\omega_{fi}-\omega)t} + \mathcal{H}^{\dagger}_{fi}\, e^{i(\omega_{fi}+\omega)t} \right]$$

$$a_f^{(1)}(t) = -\frac{1}{\hbar}\left[\mathcal{H}_{fi} \frac{e^{i(\omega_{fi}-\omega)t} - 1}{\omega_{fi} - \omega} + \mathcal{H}^{\dagger}_{fi} \frac{e^{i(\omega_{fi}+\omega)t} - 1}{\omega_{fi} + \omega} \right]$$

$\swarrow$ trig

$$|a_f^{(1)}(t)|^2 = \frac{4|\mathcal{H}_{fi}|^2}{\hbar} \frac{\sin^2(\frac{\omega_{fi}-\omega}{2} t)}{(\omega_{fi}-\omega)^2} + \frac{4|\mathcal{H}^{\dagger}_{fi}|^2}{\hbar} \frac{\sin^2(\frac{\omega_{fi}+\omega}{2} t)}{(\omega_{fi}+\omega)^2}$$

Ignore the cross terms: they cancel when time-averaged

Consider the function $$\frac{\sin^2(\frac{\omega_{fi}-\omega}{2} t)}{(\omega_{fi}-\omega)^2} = \frac{t^2}{4} \operatorname{sinc}^2\left(\frac{\omega_{fi}+\omega}{2} t\right)$$

As $t \to 0$, sinc $\to 0$ unless $\omega_{fi} \to \omega_i$. In exact terms, the limit is

$$\lim_{t\to\infty} \frac{t^2}{4} \operatorname{sinc}^2\left(\frac{\omega_{fi}-\omega}{2} t\right) = \frac{\pi t}{2} \delta(\omega_{fi} - \omega)$$

$$|a_f^{(1)}(t)|^2 = \frac{4|\mathcal{H}_{fi}|^2}{\hbar} \frac{\pi t}{2} \delta(\omega_{fi} - \omega) + \frac{4|\mathcal{H}^{\dagger}_{fi}|^2}{\hbar} \frac{\pi t}{2} \delta(\omega_{fi} + \omega)$$

- Transition rate is:

$$\frac{d|a_f^{(1)}(t)|^2}{dt} = \frac{2\pi}{\hbar} |\mathcal{H}_{fi}|^2 \delta(\omega_{fi} - \omega) + \frac{2\pi}{\hbar} |\mathcal{H}'_{fi}|^2 \delta(\omega_{fi} + \omega)$$

$$\boxed{W_{i\to f} = \frac{2\pi}{\hbar} |\mathcal{H}_{fi}|^2 \delta(E_f - E_i - \hbar\omega) + \frac{2\pi}{\hbar} |\mathcal{H}'_{fi}|^2 \delta(E_f - E_i + \hbar\omega)}$$

Fermi's golden rule

$\downarrow$ $E_f = E_i + \hbar\omega$ optical absorption

$\downarrow$ $E_f = E_i - \hbar\omega$ photon emission

- Sum Fermi's Golden Rule over all states

- Consider only the absorption component of the transition rate, and use the density of final states

$$W_{ab} = \int \frac{2\pi}{\hbar} |\mathcal{H}_{ba}|^2 \delta(E_b - E_a - \hbar\omega)\, \rho(E_b)\, dE_b$$

(i.e. $T = 0K$)

For simplicity, first consider a completely full valence band and a completely empty conduction band:

$$E_b = E_c + \frac{\hbar^2 k^2}{2m_e^*} \quad \Rightarrow \quad E_b - E_a = E_c - E_v + \frac{\hbar^2 k^2}{2}\left(\frac{1}{m_e^*} + \frac{1}{m_h^*}\right)$$

$$E_a = E_v - \frac{\hbar^2 k^2}{2m_h^*} \qquad\qquad = E_g + \frac{\hbar^2 k^2}{2m_r^*} \quad \text{where} \quad m_r^* = \left(\frac{1}{m_e^*} + \frac{1}{m_h^*}\right)^{-1}$$

- Furthermore, due to the conservation of momentum ($k_a = k_b = k$), we can also define a joint density of states that gives the effective number of possible electronic transitions when a photon of energy $\hbar\omega$ is absorbed

$$\rho_r(\hbar\omega - E_g) = \frac{1}{2\pi^2}\left(\frac{2m_r^*}{\hbar^2}\right)^{3/2}\sqrt{\hbar\omega - E_g}$$

only one value of k is allowable

$$W_{ab}(\hbar\omega) = \frac{2\pi}{\hbar}|\mathcal{H}_{ba}|^2 \int \delta\left(E_g + \frac{\hbar^2 k^2}{2m_r^*} - \hbar\omega\right)\frac{dn}{dk}\, dk$$

Apply δ, and convert to energy integral

$$W_{ab}(\hbar\omega) = \frac{2\pi}{\hbar}|\mathcal{H}_{ba}|^2 \rho_r(\hbar\omega - E_g)$$

- To get the absorption coefficient, we normalize this to the incident photon flux

$$N_{ph} = \underbrace{\tfrac{1}{2}\varepsilon_0\varepsilon_r\mathcal{E}^2}_{\text{electromagnetic energy density}} \cdot \underset{\substack{\uparrow \\ \text{group velocity}}}{\frac{c}{n_r}} \cdot \underset{\substack{\uparrow \\ \text{photon energy}}}{\frac{1}{\hbar\omega}} = \tfrac{1}{2}\varepsilon_0 n_r \frac{c}{\hbar\omega}(-i\omega A_0)^2$$

$$= \tfrac{1}{2}\varepsilon_0 n_r \frac{c}{\hbar\omega}\omega^2 A_0^2 \quad \Rightarrow \quad N_{ph} = \frac{\varepsilon_0 n_r c A_0^2}{2\hbar}\omega$$

- Finally

$$\alpha_0(\hbar\omega) = \frac{W_{ab}}{N_{ph}} = \overbrace{\frac{2\pi}{\hbar}\left|-\frac{qA_0}{2m}\hat{e}\cdot\vec{p}_{cv}\right|^2}^{|\mathcal{H}_{ba}|^2} \cdot \frac{2\hbar}{\varepsilon_0 n_r c A_0^2 \omega}\rho_r(\hbar\omega - E_g)$$

$$= \frac{\cancel{4}\pi}{\varepsilon_0 n_r c \cancel{A_0^2}\omega} \cdot \frac{q^2 \cancel{A_0^2}}{4m^2}|\hat{e}\cdot\vec{p}_{cv}|^2 \rho_r(\hbar\omega - E_g)$$

$$\alpha_0(\hbar\omega) = \underbrace{\frac{\pi q^2}{\varepsilon_0 n_r c m^2 \omega}}_{C_0}|\hat{e}\cdot\vec{p}_{cv}|^2 \rho_r(\hbar\omega - E_g) \quad \text{at} \quad \underline{T = 0K}$$

- Include the occupancy probabilities of the upper and lower electronic states to obtain accurate results at finite temperature

$$\underset{\text{absoption}}{W_{ab}} = \int dk\, \frac{2\pi}{\hbar}|\mathcal{H}_{ba}|^2\, \delta\left(E_g + \frac{\hbar^2k^2}{2m_r^*} - \hbar\omega\right)\cdot f_v(E_a)\cdot(1 - f_c(E_b))\cdot\frac{dn}{dk}$$

electron is occupied at VB state

unoccupied electron state in CB

$$\underset{\text{emission}}{W_{ba}} = \int dk\, \frac{2\pi}{\hbar}|\mathcal{H}_{ba}|^2\, \delta\left(E_g + \frac{\hbar^2k^2}{2m_r^*} - \hbar\omega\right)\cdot f_c(E_b)\cdot(1 - f_v(E_a))\cdot\frac{dn}{dk}$$

occupied CB electron

unoccupied electron in VB

* Note that both f_c, f_v refer to electron occupancies here!

- Net transition rate:

$$W(\hbar\omega) = \int dk\, \frac{2\pi}{\hbar}|\mathcal{H}_{ba}|^2\, \delta\left(E_g + \frac{\hbar^2k^2}{2m_r^*} - \hbar\omega\right)\frac{dn}{dk}\underbrace{\left[f_v(1-f_c) - f_c(1-f_v)\right]}_{= f_v(E_a) - f_c(E_b)}$$

- Convert to an energy integral, then apply the delta function:

$$W(\hbar\omega) = \frac{2\pi}{\hbar}|\mathcal{H}_{ba}|^2\rho_r(\hbar\omega - E_g)\cdot\left[-f_g(\hbar\omega - E_g)\right]$$

Let $E_c = E_g$, $E_v = 0$

where $$f_g(\hbar\omega - E_g) = f_c(E_b) - f_v(E_a)$$
$$= f_c\left(E_g + \frac{\hbar^2k^2}{2m_e^*}\right) - f_v\left(-\frac{\hbar^2k^2}{2m_h^*}\right)$$

From the delta function, we must have $$E_g + \frac{\hbar^2k^2}{2m_r^*} = \hbar\omega$$
$$\frac{\hbar^2k^2}{2m_r^*} = \hbar\omega - E_g$$

So: $$\frac{\hbar^2k^2}{2m_e^*} = (\hbar\omega - E_g)\frac{m_r^*}{m_e^*}, \quad \frac{\hbar^2k^2}{2m_h^*} = (\hbar\omega - E_g)\frac{m_r^*}{m_h^*}$$

So:

$$\boxed{f_g(\hbar\omega - E_g) = f_c\left(E_g + \frac{m_r^*}{m_e^*}(\hbar\omega - E_g)\right) - f_v\left(-(\hbar\omega - E_g)\frac{m_r^*}{m_h^*}\right)}$$

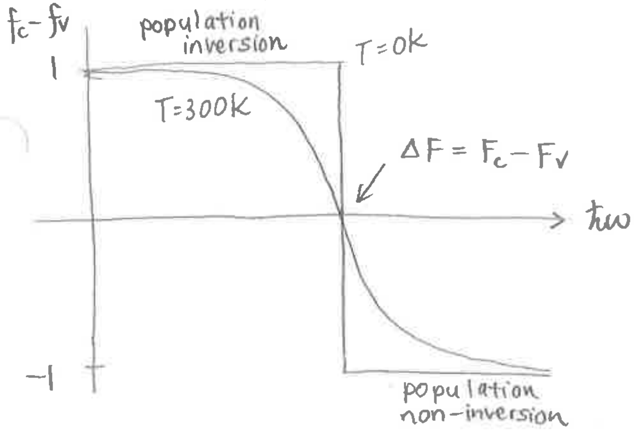
Fermi inversion factor

where $$f_c(E) = \frac{1}{e^{(E-F_c)/kT} + 1}$$

$$f_v(E) = \frac{1}{e^{(E-F_v)/kT} + 1}$$

(Make sure F_c, F_v are referenced to $E_v = 0$)

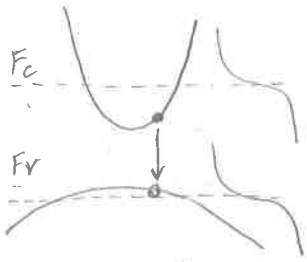
- The shape of the Fermi inversion factor is:



- If $\hbar\omega \sim E_g < \Delta F$
  - $f_c \approx 1$ and $f_v \approx 0$

    $\Rightarrow f_g \approx 1$

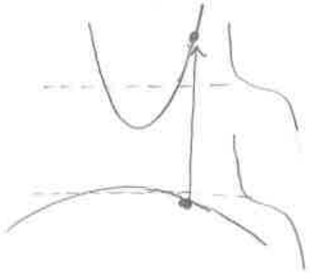
    emission



- If $\hbar\omega \sim E_g > \Delta F$
  - $f_c \approx 0$ and $f_v \approx 1$

    $\Rightarrow f_g \approx -1$

    absorption

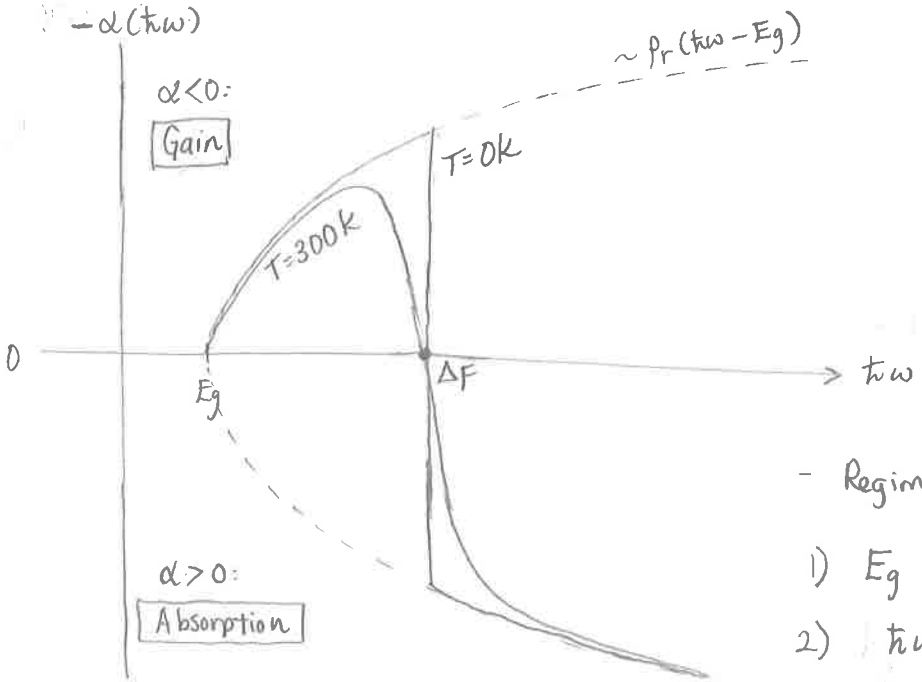


- Finally, the expression for the absorption coefficient becomes modified to:

$$\boxed{\alpha(\hbar\omega) = C_0 |\hat{e}\cdot\vec{p}_{cv}|^2 \rho_r(\hbar\omega - E_g)\cdot\left[-f_g(\hbar\omega - E_g)\right]}$$

be wary! + for gain

Let's plot $-\alpha(\hbar\omega)$



For $\hbar\omega < \Delta F$, $f_g > 0 \Rightarrow \alpha < 0$

$-\alpha > 0$

For $\hbar\omega > \Delta F$, $f_g < 0 \Rightarrow \alpha > 0$

$-\alpha < 0$

- Regimes of bias:

1) $E_g < \hbar\omega < \Delta F \Rightarrow$ Gain
2) $\hbar\omega = \Delta F \Rightarrow$ Transparency

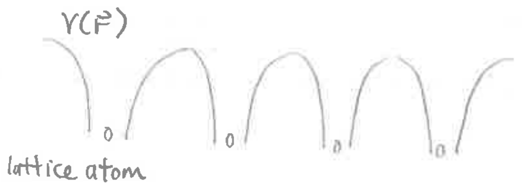
   $\alpha = 0$: no gain or absorption
3) $\hbar\omega > \Delta F \Rightarrow$ Absorption

- ΔF arises from an applied voltage V. Unless the applied voltage is large ($\sim E_g$), the response will always be absorption.
- Once $\Delta F = E_g$, the transparency condition (Bernard-Duraffourg condition) is reached: this is the onset of population inversion in the semiconductor

- Momentum conservation

In our derivations, it was important that $\vec{k}_a = \vec{k}_b = \vec{k}$. Let us revisit this, and justify it more rigorously

- In a periodic crystal, the electron wavefunction has two components:
  (1) A rapidly varying function with periodicity given by the atomic spacing
  (2) A slowly varying plane wave envelope



$$|a\rangle = u_v(\vec r)\,\frac{e^{i\vec{k}_v\cdot\vec r}}{\sqrt{V}}$$

$$|b\rangle = u_c(\vec r)\,\frac{e^{i\vec{k}_c\cdot\vec r}}{\sqrt{V}}$$

- Evaluate the optical matrix element for a transition from $|a\rangle$ to $|b\rangle$

$$\mathcal{H}'_{ba} = -\frac{qA_0}{2m}\,\hat{e}\cdot\langle b|\vec{p}\,e^{i\vec{k}_{op}\cdot\vec r}|a\rangle \leftarrow$$

before applying dipole approximation
$\vec{k}_{op}$ = optical wave-vector

$$= -\frac{qA_0}{2m}\,\hat{e}\cdot\int u_c^*(\vec r)\,\frac{e^{-i\vec{k}_c\cdot\vec r}}{\sqrt{V}}\,e^{i\vec{k}_{op}\cdot\vec r}\,(-i\hbar\vec\nabla)\,u_v(\vec r)\,\frac{e^{i\vec{k}_v\cdot\vec r}}{\sqrt{V}}\,d^3\vec r$$

$$= -\frac{qA_0}{2m}\,\hat{e}\cdot\int u_c^*(\vec r)\,e^{-i\vec{k}_c\cdot\vec r}\,e^{i\vec{k}_{op}\cdot\vec r}\left[(-i\hbar\vec\nabla u_v(\vec r)) + (\hbar k_v\,u_v(\vec r))\right]e^{i\vec{k}_v\cdot\vec r}\,\frac{d^3\vec r}{V}$$

Since u_v varies much more quickly than $e^{i\vec{k}_v\cdot\vec r}$, we can approximate $\vec\nabla u_v \gg k_v$ (slowly varying envelope approximation)

$$\mathcal{H}'_{ba} = -\frac{qA_0}{2m}\,\hat{e}\cdot\int u_c^*(\vec r)\,e^{-i\vec{k}_c\cdot\vec r}\,e^{i\vec{k}_{op}\cdot\vec r}\,(-i\hbar\vec\nabla u_v(\vec r))\,e^{i\vec{k}_v\cdot\vec r}\,\frac{d^3\vec r}{V}$$

Since u_c^* and ∇u_v vary on the scale of the unit cell, we can separate the above integral into two parts: one that is over the unit cell (volume Ω) and one over the whole volume V

$$\mathcal{H}'_{ba} = -\frac{qA_0}{2m}\,\hat{e}\cdot\underbrace{\left(\int_\Omega u_c^*(\vec r)\,\frac{\hbar}{i}\vec\nabla\,u_v(\vec r)\,\frac{d^3\vec r}{\Omega}\right)}\left(\int_V e^{i(\vec{k}_v+\vec{k}_{op}-\vec{k}_c)\cdot\vec r}\,\frac{d^3\vec r}{V}\right)$$

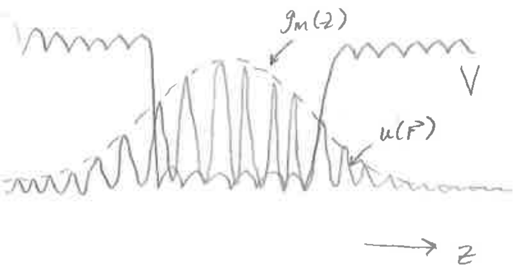
$$= -\frac{qA_0}{2m}\,\hat{e}\cdot\vec{p}_{cv}\cdot\delta(\vec{k}_v+\vec{k}_{op}-\vec{k}_c)$$

$\downarrow$
momentum matrix element over one unit cell

$\searrow$
Momentum conservation:
$\hbar\vec{k}_c = \hbar\vec{k}_v + \hbar\vec{k}_{op}$

④ Optical absorption in a quantum well

- Interband optical transitions: the quantum well wavefunction again has a component that follows the atomic-scale potential, and an envelope function that is a plane wave in (x, y) but is confined in z:



$$|a\rangle = u_v(\vec{r})\,\frac{e^{i\vec{k}_t\cdot\vec{\rho}}}{\sqrt{A}}\,g_m(z)$$

$$|b\rangle = u_c(\vec{r})\,\frac{e^{i\vec{k}_t'\cdot\vec{\rho}}}{\sqrt{A}}\,\phi_n(z)$$

$\vec{k}_t$, $\vec{\rho}$: transverse wave-vector and position vector

$$\mathcal{H}_{ba}' = \langle b|\mathcal{H}|a\rangle = -\frac{qA_0}{2m}\langle b|\,\hat{e}\cdot\vec{p}\,e^{i\vec{k}_{op}\cdot\vec{r}}\,|a\rangle$$

$$\cong -\frac{qA_0}{2m}\int d^3\vec{r}\; u_c^*(\vec{r})\,\frac{e^{-i\vec{k}_t'\cdot\vec{\rho}}}{\sqrt{A}}\,\phi_n^*(z)\left[\hat{e}\cdot\vec{p}\,e^{i\vec{k}_{op}\cdot\vec{r}}\right]u_v(\vec{r})\,\frac{e^{i\vec{k}_t\cdot\vec{\rho}}}{\sqrt{A}}\,g_m(z)$$

$$= -\frac{qA_0}{2m}\int d^3\vec{r}\; u_c^*(\vec{r})\,\frac{e^{-i\vec{k}_t'\cdot\vec{\rho}}}{\sqrt{A}}\,\phi_n^*(z)\,e^{i\vec{k}_{op}\cdot\vec{r}}\,\frac{e^{i\vec{k}_t\cdot\vec{\rho}}}{\sqrt{A}}\left[\hat{e}\cdot\vec{p}\right]u_v(\vec{r})\,g_m(z)$$

- where we have used the dipole approximation again to move $e^{i\vec{k}_t\cdot\vec{\rho}}$ outside of the operator, since we know $\vec{\nabla}u_v \gg \vec{\nabla}(e^{i\vec{k}_t\cdot\vec{\rho}})$
- Now apply the product rule:

$$[\hat{e}\cdot\vec{p}]\,u_v(\vec{r})\,g_m(z) = g_m(z)\,(\hat{e}\cdot\vec{p})\,u_v(\vec{r}) + u_v(\vec{r})\,(\hat{e}\cdot\vec{p})\,g_m(z)$$

g_m is only a function of z, so $\vec{p}\,g_m(z) = \frac{\hbar}{i}\vec{\nabla}g_m(z) = p_z\,\hat{z}\,g_m(z)$

$$\mathcal{H}_{ba}' = -\frac{qA_0}{2m}\int d^3\vec{r}\; u_c^*(\vec{r})\,\frac{e^{-i\vec{k}_t\cdot\vec{\rho}}}{\sqrt{A}}\,\phi_n^*(z)\,\underbrace{\cancel{e^{i\vec{k}_{op}\cdot\vec{r}}}}_{\text{constant phase}}\,\frac{e^{i\vec{k}_t\cdot\vec{\rho}}}{\sqrt{A}}\left[g_m(z)\,(\hat{e}\cdot\vec{p})\,u_v(\vec{r}) + e_z p_z\,g_m(z)\right]$$

$$= -\frac{qA_0}{2m}\Bigg[\left(\int d^3\vec{r}\,\frac{e^{i(\vec{k}_t-\vec{k}_t')\cdot\vec{\rho}}}{A}\,\phi_n^*(z)\,g_m(z)\right)\left(\int\frac{d^3\vec{r}}{\Omega}\,u_c^*(\vec{r})\,(\hat{e}\cdot\vec{p})\,u_v(\vec{r})\right)$$

$$+\left(\int d^3\vec{r}\,\frac{e^{i(\vec{k}_t-\vec{k}_t')\cdot\vec{\rho}}}{A}\,\phi_n^*(z)\,e_z p_z\,g_m(z)\right)\left(\int\frac{d^3\vec{r}}{\Omega}\,\underbrace{\cancel{u_c^*(\vec{r})\,u_v(\vec{r})}}_{0}\right)\Bigg]$$

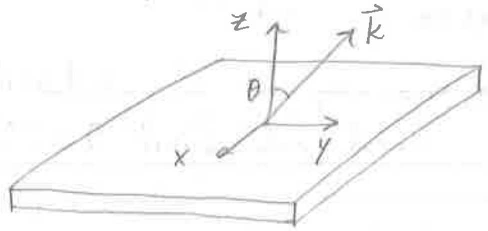
$$= -\frac{qA_0}{2m}\left(\int d^2\vec{\rho}\,\frac{e^{i(\vec{k}_t-\vec{k}_t')\cdot\vec{\rho}}}{A}\right)\left(\int dz\,\phi_n^*(z)\,g_m(z)\right)\left(\int\frac{d^3\vec{r}}{\Omega}\,u_c^*(\vec{r})\,(\hat{e}\cdot\vec{p})\,u_v(\vec{r})\right)$$

$$\boxed{\mathcal{H}_{ba}' = -\frac{qA_0}{2m}\,\underset{\substack{\uparrow\\ \text{conservation of}\\ \text{transverse momentum}}}{\delta(\vec{k}_t-\vec{k}_t')}\,\underset{\substack{\uparrow\\ \text{momentum}\\ \text{matrix element}\\ \text{for periodic parts of }\psi}}{(\hat{e}\cdot\vec{p}_{cv})}\cdot\underset{\substack{\uparrow\\ \text{confinement direction}\\ \text{selection rules}}}{\int dz\,\phi_n^*(z)\,g_m(z)}}$$

- Polarization dependence of interband transition in quantum wells

  + To analyze this, a detailed $\vec{k}\cdot\vec{p}$ methodology is needed to first obtain the dispersion relations for the conduction band, HH band, and LH band
    - This is done first by defining $\vec{k} = k\hat{z}$
    - Then a coordinate rotation is applied to the resulting wavefunctions to get a general expression for u_c, u_{hh}, u_{lh}

- The matrix elements $|\hat{e}\cdot\vec{p}_{cv}|^2$ are polarization-independent in a bulk Semiconductor
- In a quantum well, define $\hat{z}$ to be the confinement direction, and let θ be the angle of $\vec{k}$ from $\hat{z}$:



($\vec{k}$ refers to the electron, not the photon)

- The matrix element $|\hat{e}\cdot\vec{p}_{cv}|^2$ is:

	TE: $\hat{e}=\hat{x}$ or $\hat{y}$	TM: $\hat{e}=\hat{z}$
C-HH transition	$\frac{3}{4}(1+\cos^2\theta)M_b^2$	$\frac{3}{2}\sin^2\theta\ M_b^2$
C-LH transition	$(\frac{5}{4}-\frac{3}{4}\cos^2\theta)M_b^2$	$(\frac{1}{2}+\frac{3}{2}\cos^2\theta)M_b^2$
Sum of C-HH, C-LH	$2M_b^2$	$2M_b^2$

- Notes:
  - The sum of the C-HH, C-LH matrix elements is polarization independent
  - If $\theta = 0$ ($\vec{k}$ in plane of QW)
    - No C-HH transition for TM
    - Weak C-LH transition for TE

- Band edge:

HH → C : $M_{TE}^{HH} = \frac{3}{2}M_b^2$, $M_{TM}^{HH} = 0$ ⇒ HH → C responds strongly to TE, weakly to TM.

LH → C : $M_{TE}^{LH} = \frac{1}{2}M_b^2$, $M_{TM}^{LH} = 2M_b^2$ ⇒ LH → C responds strongly to TM, weakly to TE

For optical amplifiers, polarization independence is desired. To get this in a quantum well, tensile strain is used, to make the HH and LH transitions roughly balanced

- Interband transition selection rules

The interband selection rules are determined by the term,

$$\int_{-\infty}^{\infty} \phi_n^*(z)\, g_m(z)\, dz \sim \int_{-\infty}^{\infty} \sin\left(\frac{n\pi}{L_z} z\right) \sin\left(\frac{m\pi}{L_z} z\right) dz = \delta_{mn}$$

↑ conduction subband

↖ valence subband

⇒ Interband transitions are only allowed between the n^{th} subbands of the conduction and valence bands.

- Quantum well interband absorption and gain

The general expression is the same as in the bulk case,

$$g(\hbar\omega) = -\alpha(\hbar\omega) = C_0 |\hat{e}\cdot\vec{p}_{cv}|^2 \rho_r^{2D}(\hbar\omega - E_g)\, f_g(\hbar\omega - E_g)$$

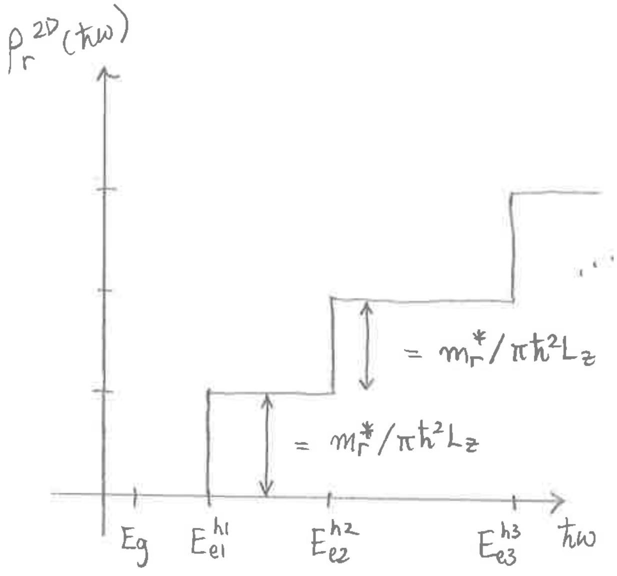
where the differences are in:

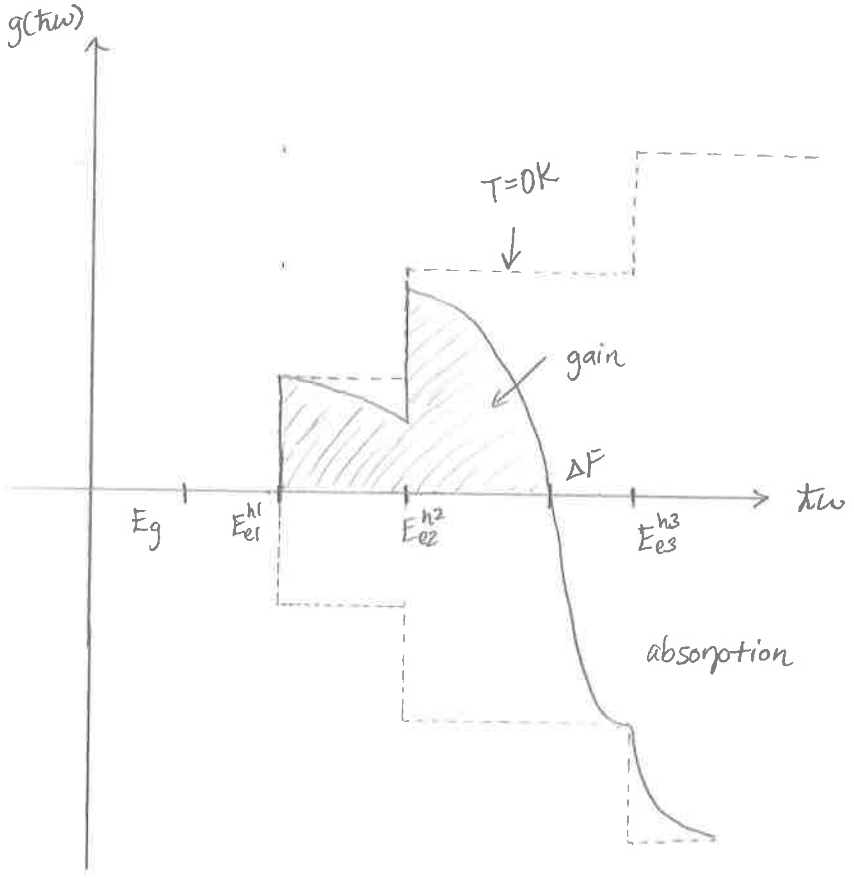
(1) The momentum matrix element $|\hat{e}\cdot\vec{p}_{cv}|^2$ has a new polarization dependence and selection rules between subbands

(2) The joint DOS is now given by:

$$\rho_r^{2D}(\hbar\omega - E_g) = \sum_m \sum_n \frac{m_r^*}{\pi\hbar^2 L_z} H(\hbar\omega - E_{en}^{hm}) = \sum_n \frac{m_r^*}{\pi\hbar^2 L_z} H(\hbar\omega - E_{en}^{hn})$$

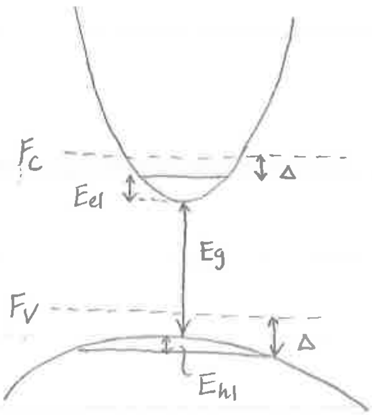
where $E_{en}^{hm} = (E_c + E_{en}) - (E_v - E_{hm}) = E_g + E_{en} + E_{hm}$





- Transparency condition in a quantum well

  - Consider an undoped, unstrained quantum well, and use the bulk values of electron and hole effective mass (which is not exactly valid)
  - The transparency condition occurs when $\Delta F = \hbar\omega$. What is the carrier density when this occurs?



$$N = \frac{m_e^* kT}{\pi\hbar^2 L_z} \ln\left(1 + e^{(F_c - (E_g + E_{e1}))/kT}\right)$$

$$p = \frac{m_h^* kT}{\pi\hbar^2 L_z} \ln\left(1 + e^{(-E_{h1} - F_V)/kT}\right)$$

where the energy is referenced to the bulk valence level $E_V = 0$. Now notice that the transparency condition implies

$$\Delta F = \hbar\omega = E_g + E_{e1} + E_{h1}$$

$$F_c - F_V = E_g + E_{e1} + E_{h1}$$

$$F_c - (E_g + E_{e1}) = -(-E_{h1} - F_V) \equiv \Delta \cdot kT$$

Therefore, $N = \frac{m_e^* kT}{\pi\hbar^2 L_z} \ln(1 + e^{\Delta})$

$$p = \frac{m_h^* kT}{\pi\hbar^2 L_z} \ln(1 + e^{-\Delta})$$

Net neutrality in an undoped semiconductor implies $N = P$

$$\Rightarrow \quad m_e^* \ln(1 + e^{\Delta}) = m_h^* \ln(1 + e^{-\Delta})$$

Numerically solving this using $m_e^*/m_h^* = \frac{1}{7.5}$, which may be accurate for an unstrained GaAs quantum well, with $m_e^* = 0.067 m_0$:

$\Delta = 1.41$; therefore:

$$N_{tr} = \frac{m_e^* kT}{\pi\hbar^2 L_z} \ln(1 + e^{\Delta}) = \frac{1}{L_z}(1.2 \times 10^{12}\,\text{cm}^{-2})$$

$$P_{tr} = N_{tr} \approx 1.2 \times 10^{18}\,\text{cm}^{-3}$$

$\Delta = 1.41$ implies that F_c lies inside the conduction band and outside the valence band at transparency, due to effective mass asymmetry

This is the necessary current density for gain to occur. Note again that this analysis holds only for an undoped, unstrained material; for the effects of strain and doping, see §⑥!

- Quantum well intersubband transitions

- Intersubband transitions respond to energies of $E_{e2} - E_{e1} = \frac{\hbar^2}{2m_e^*}\left(\frac{2\pi}{L_z} - \frac{\pi}{L_z}\right)$

  $\approx 168\text{mV}$ for $L_z = 10\text{nm}$ GaAs

- Transition matrix element:

  The wavefunctions are

$$|a\rangle = u_c(\vec{r}) \frac{1}{\sqrt{A}} e^{i\vec{k}_t\cdot\vec{\rho}} \phi_1(z) \quad \leftarrow 1^{st}\ e^-\ \text{subband}$$

$$|b\rangle = u_c(\vec{r}) \frac{1}{\sqrt{A}} e^{i\vec{k}_t'\cdot\vec{\rho}} \phi_2(z) \quad \longleftarrow 2^{nd}\ e^-\ \text{subband}$$

To evaluate $\mathcal{H}'_{ba}$, it is easier to use the alternate form:

$$\mathcal{H}'_{ba} = -\vec{E}\cdot q\vec{r}_{ba} = -E_0\,\hat{e}\cdot\langle b|q\vec{r}|a\rangle\cdot \cancel{e^{i\vec{k}_{op}\cdot\vec{r}}} \quad \text{slow}$$

$$= -qE_0\,\hat{e}\cdot\int u_c^*(\vec{r})\frac{1}{\sqrt{A}}e^{-i\vec{k}_t'\cdot\vec{\rho}}\phi_2^*(z)\,\vec{r}\,u_c(\vec{r})\frac{1}{\sqrt{A}}e^{i\vec{k}_t\cdot\vec{\rho}}\phi_1(z)\,d^3\vec{r}$$

$$\cong -qE_0\,\hat{e}\cdot\left[\int u_c^*(\vec{r})\,u_c(\vec{r})\frac{d^3\vec{r}}{\Omega}\right]\cdot\left[\int\frac{1}{A}e^{i(\vec{k}_t-\vec{k}_t')\cdot\vec{\rho}}\,d^2\vec{\rho}\right]\cdot\int\phi_2^*(z)\,\phi_1(z)\,\vec{z}\,dz$$

- In the above step, the x and y components of $\vec{r}$ vanish because integrating $(x\hat{x}+y\hat{y})\,e^{i(\vec{k}_t-\vec{k}_t')\cdot\vec{\rho}}$ returns 0
- Note that:

$$\int u_c^*(\vec{r})\,u_c(\vec{r})\frac{d^3\vec{r}}{\Omega} = 1$$

$$\int\frac{1}{A}e^{i(\vec{k}_t-\vec{k}_t')\cdot\vec{\rho}}\,d\vec{\rho} = \delta(\vec{k}_t-\vec{k}_t')$$

$$\Rightarrow \quad \mathcal{H}'_{ba} = -qE_0\,\hat{e}\cdot\int\phi_2^*(z)\,\phi_1(z)\,z\,\hat{z}\,dz\cdot\delta(\vec{k}_t-\vec{k}_t')$$

$$\boxed{\mathcal{H}'_{ba} = -E_0\,(\hat{e}\cdot\hat{z})\,\mu_{21}\,\delta(\vec{k}_t-\vec{k}_t')}$$

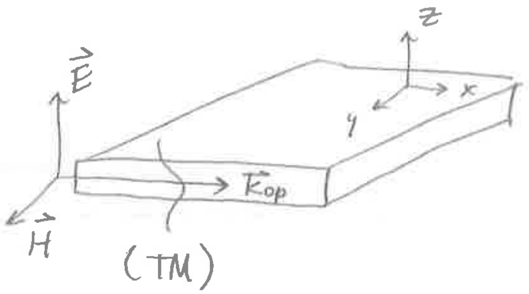
where $\mu_{21} = q\int\phi_2^*(z)\,\phi_1(z)\,z\,dz$

dipole moment between subband envelope functions

- Polarization dependence of intersubband transitions

Note that the dipole moment is always polarized along the confinement direction, in this case $\hat{z}$.

$$\mathcal{H}'_{ba} = -E_0\,(\underset{\swarrow}{\hat{e}\cdot\hat{z}})\,\mu_{21}\cdot\delta(\vec{k}_t-\vec{k}_t')$$



- If the incident field does not have a component along $\hat{z}$, there will be no intersubband matrix element; i.e. responds only to TM polarization

- If $\vec{E}$ is polarized at an angle θ away from $\hat{z}$, the matrix element diminishes with θ:

$$\mathcal{H}'_{ba} = -E_0\,\mu_{21}\cos\theta\cdot\delta(\vec{k}_t-\vec{k}_t')$$

Even more generally:

* An electric field can induce an intersubband transition if it has a component along one of the confinement directions

* If there are multiple confinement directions (e.g. quantum wire), intersubband transitions can occur along any confinement direction that is excited by a nonzero component of $\vec{E}$

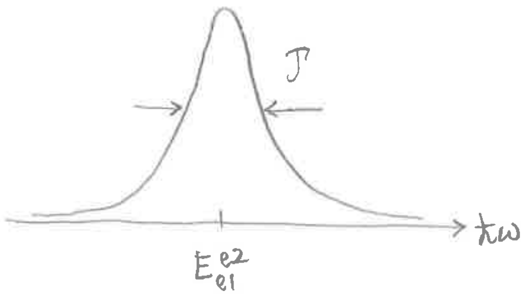
  - If $\vec{E}$ is polarized only along one of the confinement directions, it can only induce transitions between two subbands of that one confinement direction

- Absorption: inter-subband

$$\alpha(\hbar\omega) = \underset{\substack{\uparrow\\ \text{incident flux}\\ \text{normalization}}}{\frac{1}{N_{ph}}}\,\frac{2\pi}{\hbar}\,|-E_0\underset{\substack{\uparrow\\ \text{matrix}\\ \text{element}}}{\hat{e}\cdot\vec{\mu}_{21}}|^2\,\underset{\substack{\uparrow\\ \text{energy conservation}}}{g(E_{e1}^{e2}-\hbar\omega)}\cdot\frac{2}{V}\underbrace{\sum_k[f_c(E_{e1})-f_c(E_{e2})]}_{\substack{\text{counting all initial and}\\ \text{final states}}}$$

- Energy conservation: without accounting for energy broadening,

$g(E_{e1}^{e2}-\hbar\omega) = \delta(E_{e1}^{e2}-\hbar\omega)$ which has an infinite peak



- Instead, use a Lorentzian

$$g(E_{e1}^{e2}-\hbar\omega) = \frac{1}{\pi}\,\frac{\Gamma/2}{(E_{e1}^{e2}-\hbar\omega)^2+(\Gamma/2)^2}$$

This gives the shape of the absorption spectrum $\alpha(\hbar\omega)$

- The sum over states gives:

$$\frac{2}{V}\sum_{K_1} f_c(E_{e1}) = N_1 \quad \leftarrow \text{\# } e^- \text{ in subband } 1$$

$$\frac{2}{V}\sum_{k_2} f_v(E_{e2}) = N_2 \quad \leftarrow \text{\# } e^- \text{ in subband } 2$$

- Then the expression becomes:

$$\alpha(\hbar\omega) = \frac{1}{N_{ph}} \frac{2\pi}{\hbar} \left| -E_0 \hat{e}\cdot\vec{\mu}_{21} \right|^2 g(E_{e1}^{e2} - \hbar\omega)(N_1 - N_2)$$

$$= \frac{\hbar\omega}{\frac{1}{2}\varepsilon_0 n_r E^2 c} \frac{2\pi}{\hbar} E_0^2 \left|\hat{e}\cdot\mu_{21}\right|^2 g(E_{e1}^{e2} - \hbar\omega)(N_1 - N_2)$$

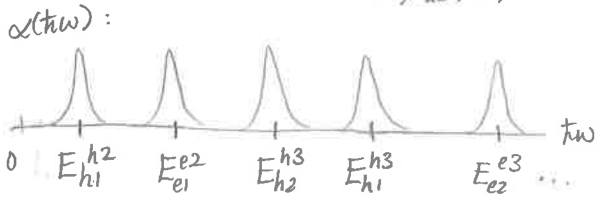
↖ There is probably a factor of ½ missing in front of E inside $\mathcal{H}'_{ba}$ which selects the forward-going wave

$$\boxed{\alpha(\hbar\omega) = \frac{\pi\omega}{\varepsilon_0 n_r c} \left|\hat{e}\cdot\vec{\mu}_{21}\right|^2 g(E_{e1}^{e2} - \hbar\omega)(N_1 - N_2)}$$

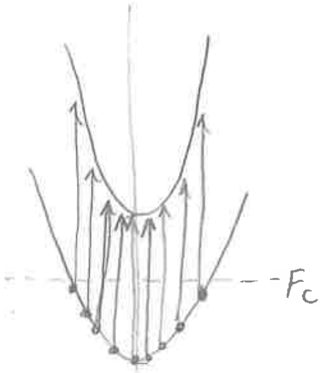
When integrated, $\int g \, d\omega \rightarrow 1$ and we can assume it is narrow around E_{e1}^{21} so the ω out front can be pulled out:

$$\int \alpha(\hbar\omega)\, d(\hbar\omega) = \frac{\pi E_{e1}^{e2}}{\varepsilon_0 n_r c \hbar} \left|\hat{e}\cdot\vec{\mu}_{21}\right|^2 (N_1 - N_2)$$

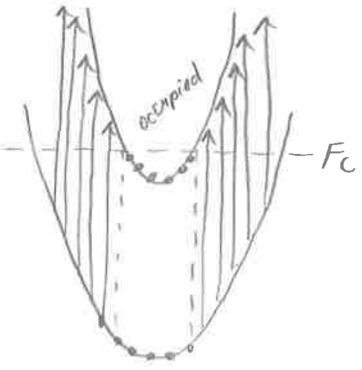
individual peak values depend on ω, $\vec{\mu}_{ab}$, ΔN



- Absorption as a function of carrier concentration

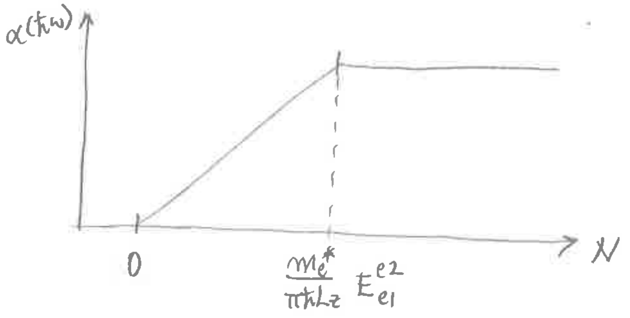


Only the first subband is filled



Both subbands filled

Once both subbands are filled, the number of possible transitions does not change! So the absorption coefficient tends to saturate



- More rigorously,

$$N_1 - N_2 = N_1 \qquad \text{if} \quad N_2 = 0$$

$$N_1 - N_2 \cong \frac{m_e^*}{\pi\hbar^2 L_z}(F_c - E_{e1}) - \frac{m_e^*}{\pi\hbar^2 L_z}(F_c - E_{e2}) = \frac{m_e^*}{\pi\hbar^2 L_z}(E_{e2} - E_{e1})$$

fixed!

- Intersubband selection rules

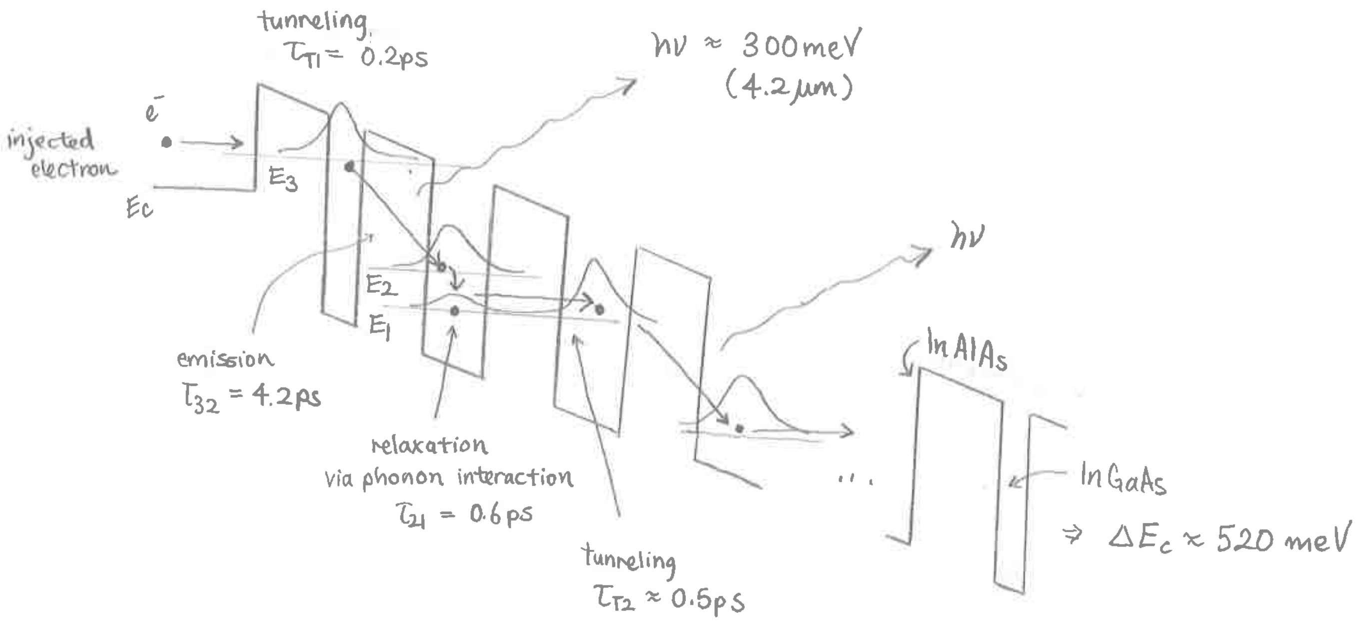
This is governed by

$$\int_0^L z\, \phi_n^*(z)\, \phi_m(z)\, dz \quad \text{is zero if the integrand is odd}$$

- So as long as ϕ_n and ϕ_m have different parity (i.e. one is even and one is odd), this integral will be nonzero
- Thus, the allowable transitions are

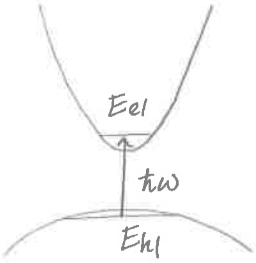
1 → 2	2 → 1	
1 → 4	2 → 3	...
1 → 6	2 → 5	

- Quantum cascade lasers : gain in inter-subband transitions
  - Form a heterojunction superlattice, in which the energy bands in very closely spaced quantum wells are strongly coupled
  - The QCL is based on a 3-level system:



  - Due to the slow intersubband radiation process ($\tau_{32} = 4.2ps$) relative to the speed of filling E_3 and emptying E_2, the population inversion is very efficient
  - One injected electron can produce several infrared photons as it travels through the superlattice (cascading, or carrier recycling)
  - Very useful emitters in the mid-IR and THz range

- Differential gain : $\frac{dg}{dN}$, the rate at which gain increases with the number of injected carriers (important for determining modulation bandwidth)



$$g(\hbar\omega) = C_0 |\hat{e}\cdot\vec{p}_{cv}|^2 \rho_r(\hbar\omega)(f_c(E_e) - f_v(E_h))$$

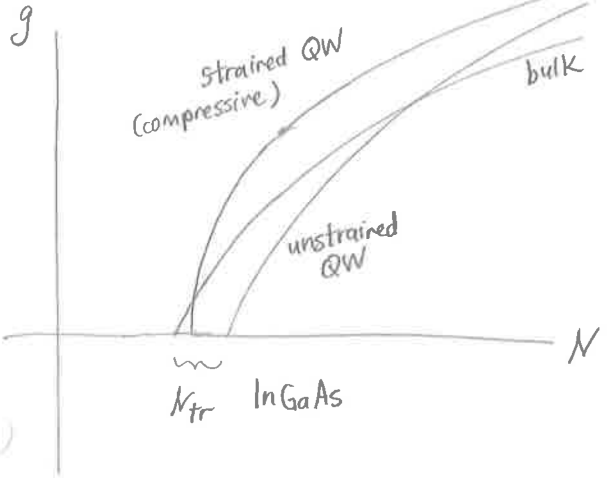
$$= C_0 |\hat{e}\cdot\vec{p}_{cv}|^2 \left[\rho_r(\hbar\omega) f_c(E_e) - \rho_r(\hbar\omega) f_v(E_h)\right]$$

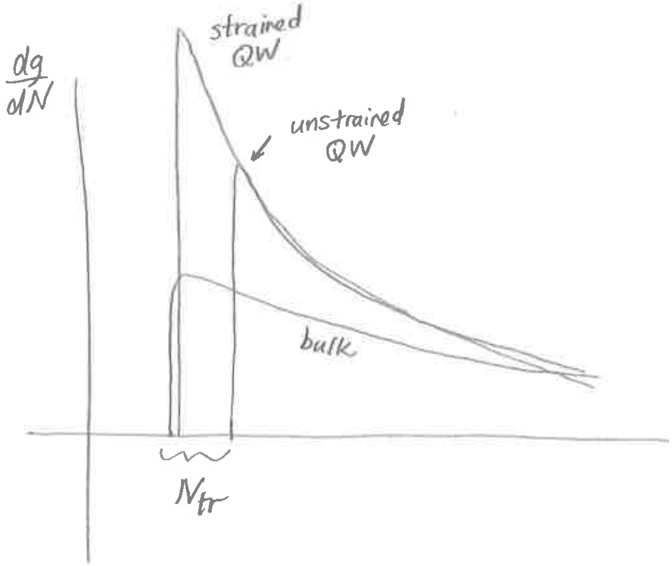
$$= C_0 |\hat{e}\cdot\vec{p}_{cv}|^2 \left[\underbrace{\rho_r(\hbar\omega) f_c(E_e)}_{\approx \frac{dN}{dE}(E_e)} + \underbrace{\rho_r(\hbar\omega)(1 - f_v(E_h))}_{\approx \frac{dP}{dE}(E_h)} - \rho_r(\hbar\omega)\right]$$

- Since ρ_r is proportional to the density of states ρ_e, ρ_h in either band, the two terms above correspond approximately to the carrier density (per unit energy) at the two band edges E_e and E_h
- Thus, if the band edge carrier density is a sensitive function of the Fermi level positions, the differential gain becomes larger. This is achieved when two conditions are met:

1) ρ is steep at the band edges → use a quantum well
2) The Fermi levels lie close to their respective band edges, since the occupancy functions f_c, f_v change most rapidly at the Fermi levels
→ use strain to get symmetric electron and hole bands

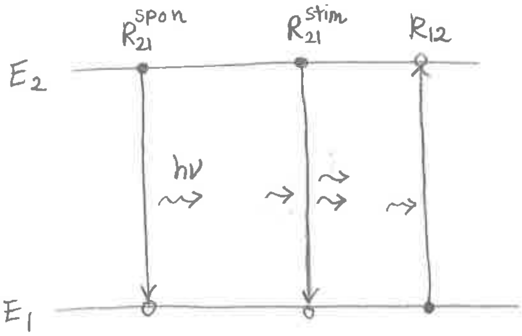
- Curves from Coldren:





⑤ Spontaneous emission

- The most famous derivation of the spontaneous emission rate invoke's Einstein's A&B coefficients:



$$R_{21}^{spon} = A_{21} f_2 (1 - f_1)$$
$$R_{21}^{stim} = B_{21} f_2 (1 - f_1) \cdot P(E_{21})$$
$$R_{12} = B_{12} (1 - f_2) f_1 \cdot P(E_{21})$$

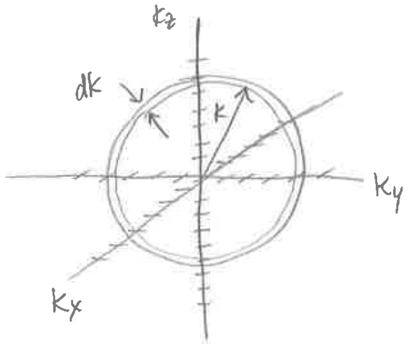
where $P(E_{21}) = \frac{\#\text{ photons}}{\text{volume} \cdot \text{energy}} = \langle n_{ph} \rangle \cdot \rho_{ph}(E_{21})$

$\uparrow$ Bose-Einstein distribution

$\uparrow$ Photon DOS at $h\nu = E_{21}$

- Photon density of states

First find DOS in k-space, then use the photon dispersion relation



- Consider a spherical shell in k-space w/ radius k

$$dN = \frac{4\pi k^2\, dk}{(2\pi/L)^3} \times 2 \leftarrow \#\text{ polarizations}$$

$$\underset{\text{(per volume)}}{dn} = \frac{k^2}{\pi^2}\, dk$$

- Now use $E = \hbar\omega = \frac{\hbar c k}{n}$

$$k = n \cdot \frac{E}{\hbar c}\,, \quad dk = \frac{n}{\hbar c}\, dE$$

$$\Rightarrow \quad dn = \frac{1}{\pi^2} \frac{n^2 E^2}{\hbar^2 c^2} \cdot \frac{n}{\hbar c}\, dE = \frac{1}{\pi^2} \frac{n^3 E^2}{\hbar^3 c^3}\, dE$$

$$\boxed{\rho_{ph}(E) = \frac{8\pi n^3 E^2}{h^3 c^3}}$$ photon DOS

Meanwhile the Bose-Einstein distribution is $\langle n_{ph} \rangle = \frac{1}{e^{E/kT} - 1}$

This represents the expectation value of the number of photons that occupy a state at energy E, derived from statistical mechanics.

- In thermal equilibrium,

$$R_{21}^{spon} + R_{21}^{stim} = R_{12}$$

$$A_{21} f_2(1-f_1) + B_{21} f_2(1-f_1) P(E_{21}) = B_{12} f_1(1-f_2) P(E_{21})$$

$$P(E_{21})\left[B_{12} f_1(1-f_2) - B_{21} f_2(1-f_1)\right] = A_{21} f_2 (1-f_1)$$

$$P(E_{21}) = \frac{A_{21} f_2(1-f_1)}{B_{12} f_1(1-f_2) - B_{21} f_2(1-f_1)}$$

Recall that the Fermi-Dirac occupancy is

$$f_1 = f(E_1) = \frac{1}{e^{(E_1-E_F)/kT}+1}, \quad 1-f_1 = \frac{e^{(E_1-E_F)/kT}}{e^{(E_1-E_F)/kT}+1}$$

$$f_2 = f(E_2) = \frac{1}{e^{(E_2-E_F)/kT}+1}, \quad 1-f_2 = \frac{e^{(E_2-E_F)/kT}}{e^{(E_2-E_F)/kT}+1}$$

All the Fermi products have the same denominator, so:

$$P(E_{21}) = \frac{A_{21} e^{(E_1-E_F)/kT}}{B_{12} e^{(E_2-E_F)/kT} - B_{21} e^{(E_1-E_F)/kT}} = \frac{A_{21}}{B_{12} e^{(E_2-E_1)/kT} - B_{21}}$$

Equate this to the known photon density $P(E_{21})$:

$$P(E_{21}) = \frac{A_{21}}{B_{12} e^{E_{21}/kT} - B_{21}} = \frac{8\pi n^3 E_{21}^2}{c^3 h^3} \cdot \frac{1}{e^{E_{21}/kT} - 1}$$

$$\left(\frac{A_{21}}{B_{12}}\right) \frac{1}{e^{E_{21}/kT} - \frac{B_{21}}{B_{12}}} = \left(\frac{8\pi n^3 E_{21}^2}{c^3 h^3}\right) \frac{1}{e^{E_{21}/kT} - 1}$$

This equation immediately implies:

1) $B_{12} = B_{21}$: the stimulated emission rate equals the absorption rate

2) $A_{21} = \left(\frac{8\pi n^3 E_{21}^2}{c^3 h^3}\right) B_{21}$

↳ proportionality between spontaneous emission and absorption pre-factors

- Recall that the absorption coefficient is

Fermi inversion factor x -1

$$\alpha(E_{21})\,dE = -g(E_{21})\,dE = \frac{n}{c}B_{21}\left[f_1 - f_2\right]$$

Meanwhile the spontaneous emission rate is

$$r_{21}(E_{21})\,dE = A_{21}\,f_2(1-f_1)$$

$$= \left(\frac{8\pi n^3 E_{21}^2}{c^3 h^3}\right) B_{21}\; f_2(1-f_1)$$

$$= \left(\frac{8\pi n^3 E_{21}^2}{c^3 h^3}\right) \frac{c}{n}\;\frac{\alpha(E_{21})\,dE}{f_1 - f_2}\cdot f_2(1-f_1)$$

$$r_{21}(E_{21}) = \frac{8\pi n^2 E_{21}^2}{c^2 h^3}\;\frac{f_2(1-f_1)}{f_1 - f_2}\;\alpha(E_{21})$$

where

$$\frac{f_2(1-f_1)}{f_1-f_2} = \frac{e^{(E_1-F_v)/kT}}{\left[e^{(E_2-F_c)/kT}+1\right]\left[e^{(E_1-F_v)/kT}+1\right]}\cdot\left(\frac{e^{(E_2-F_c)/kT} - e^{(E_1-F_v)/kT}}{\left[e^{(E_2-F_c)/kT}+1\right]\left[e^{(E_1-F_v)/kT}+1\right]}\right)^{-1}$$

$$= \frac{e^{(E_1-F_v)/kT}}{e^{(E_2-F_c)/kT} - e^{(E_1-F_v)/kT}} = \frac{1}{e^{(E_{21}-\Delta F)/kT} - 1}$$

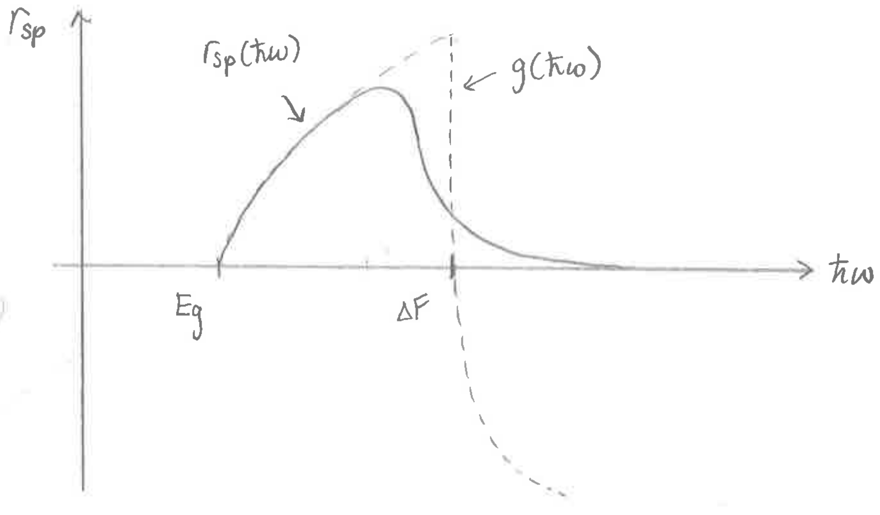
Therefore:

Note that this also depends on ΔF, near or above transparency

$$\boxed{r_{sp}(E) = \frac{8\pi n^2 E^2}{c^2 h^3}\;\alpha(E)\;\frac{1}{e^{(E-\Delta F)/kT} - 1}}$$

spontaneous emission rate
$(\#/s\cdot m^3\cdot ev)$

This is the Roosbroeck-Shockley relation!



- For $E \ll \Delta F$, last term → -1
- For $E \gg \Delta F$, last term → 0
- On this plot, designate emission as positive, absorption as negative
- At transparency, $E = \Delta F$, the Bose-Einstein distribution blows up but this is balanced by α going to 0 due to the Fermi inversion factor

- Spontaneous emission lifetime : an alternative expression for r_{sp} is

$$r_{sp}(\hbar\omega) = \frac{1}{\tau_{sp}} \rho_r(\hbar\omega - E_g) \left[\underbrace{f_c(E_2)\cdot(1 - f_v(E_1)}_{f_e(\hbar\omega)\ :\ \text{probability of emission}} \right]$$

Equate this to the previously derived expression:

$$r_{sp}(\hbar\omega) = \frac{8\pi n^2 E^2}{c^2 h^3}\,\alpha(\hbar\omega)\,\frac{1}{e^{(E-\Delta F)/kT}-1} = \frac{1}{\tau_{sp}}\,\rho_r(\hbar\omega - E_g)\,f_e(\hbar\omega)$$

$$\frac{8\pi n^2(\hbar\omega)^2}{c^2h^3}\,g(\hbar\omega)\,\frac{1}{1-e^{(E-\Delta F)/kT}} = \frac{1}{\tau_{sp}}\,\rho_r(\hbar\omega - E_g)\,f_e(\hbar\omega)$$

$$\frac{8\pi n^2(\hbar\omega)^2}{c^2h^3}\left[C_0|\hat{e}\cdot\vec{p}_{cv}|^2\rho_r(\hbar\omega-E_g)\,f_g(\hbar\omega)\right]\frac{1}{1-e^{(E-\Delta F)/kT}} = \frac{\rho_r(\hbar\omega-E_g)\,f_e(\hbar\omega)}{\tau_{sp}}$$

Recall that

$$\frac{f_2(1-f_1)}{f_1-f_2} = -\frac{f_c(1-f_v)}{f_c-f_v} = \frac{-f_e}{f_g} = \frac{1}{e^{(E-\Delta F)/kT}-1}$$

$$\Rightarrow \quad \frac{8\pi n^2(\hbar\omega)^2}{c^2h^3}\cdot C_0|\hat{e}\cdot\vec{p}_{cv}|^2\cdot\left(\frac{f_g}{f_e}\;\overset{1}{\cancel{\frac{1}{1-e^{(E-\Delta F)/kT}}}}\right) = \frac{1}{\tau_{sp}}$$

So:

$$\tau_{sp} = \frac{c^2h^3}{8\pi n^2(\hbar\omega)^2}\cdot\frac{1}{C_0|\hat{e}\cdot\vec{p}_{cv}|^2}$$

multiply constants

$$\boxed{\tau_{sp} = \frac{hc^3\varepsilon_0 m_0^2}{2q^2 n\omega}\cdot\frac{1}{|\hat{e}\cdot\vec{p}_{cv}|^2}} \sim 1\,ns$$

for semiconductors
(interband)

⑥ Effects of doping and strain

- Threshold current : the threshold current is best derived from the rate equations in steady state (see § 9)

  - At transparency or at threshold, $S \approx 0$

    $$\Rightarrow \frac{dN}{dt} = \frac{\eta_i I_{th}}{qV} - \frac{N_{th}}{\tau} = 0$$

    The current must exactly balance the rate of carrier loss in steady state:

    $$\boxed{I_{th} = \frac{qV}{\eta_i}\left[AP_{th} + BN_{th}P_{th} + CN_{th}^2 P_{th}\right]} \quad \text{(n-type active region)}$$

    At threshold in a III-V optoelectronic material, the radiative rate usually dominates, so

    $$I_{th} \approx \frac{qV}{\eta_i} BN_{th}P_{th} = \frac{qlwd}{\eta_i} BN_{th}P_{th}$$

    Let $d = L_z$ be the confinement direction and the transverse area of the quantum well is $A_t = lw$

    $$\Rightarrow J_{th} = \frac{1}{\eta_i} qL_z BN_{th}P_{th}$$

  - If the quantum well is undoped, $N_{th} = P_{th}$ so

    $$J_{th} = \frac{1}{\eta_i} qL_z (BN_{th}) N_{th} = \frac{1}{\eta_i} \frac{qL_z}{\tau_e} N_{th}$$

    where BN_{th} is in units of s^{-1} and $\tau_e = 1/BN_{th}$ can be considered the carrier life time

  - The transparency current follows the same relationship with N:

    $$J_{tr} = \frac{1}{\eta_i} qL_z \left(AP_{tr} + BN_{tr}P_{tr} + CN_{tr}^2 P_{tr}\right) \approx \frac{1}{\eta_i} qL_z BN_{tr}P_{tr}$$

    In order to reduce J_{tr}, and therefore J_{th}, we must reduce the product $N_{tr} \times P_{tr}$.

- Undoped quantum well :

  - In § 4, we found that $N_{tr} = P_{tr} = 1.2 \times 10^{18} \text{cm}^{-3}$

    Let our figure of merit be $\bar{N}_{tr} = \sqrt{N_{tr}P_{tr}} = 1.2 \times 10^{18} \text{ cm}^{-3}$

    with $m_e^*/m_h^* = 1/7.5$, $m_e^* = 0.067 m_0$

    $L_z = 10 \text{ nm}$

- Effect of doping

- n-doping:

  Charge neutrality becomes $N = p + N_D^+$ where we assume $N_D^+ \cong N_D$

  This becomes:

$$\frac{m_e^* kT}{\pi\hbar^2 L_z} \ln(1+e^{\Delta}) = \frac{m_h^* kT}{\pi\hbar^2 L_z} \ln(1+e^{-\Delta}) + N_D$$

  Let $N_D = 1\times10^{18}\text{cm}^{-3}$ and numerically solve the above

  $\Rightarrow \Delta = 2.12,\quad N_{tr} = 1.62\times10^{18}\text{cm}^{-3}$

  $P_{tr} = 6.2\times10^{17}\text{cm}^{-3} \Rightarrow \bar{N}_{tr} = 1.0\times10^{18}\text{cm}^{-3}$

  (~17% smaller than undoped!)

- p-doping:

  Charge neutrality becomes $N + N_A^- = p$ where we assume $N_A^- = N_A$

  This becomes:

$$\frac{m_e^* kT}{\pi\hbar^2 L_z} \ln(1+e^{\Delta}) = \frac{m_h^* kT}{\pi\hbar^2 L_z} \ln(1+e^{-\Delta}) - N_A$$

  Let $N_A = 1\times10^{18}\text{cm}^{-3}$ and numerically solve:

  $\Rightarrow \Delta = 0.878,\quad N_{tr} = 8.9\times10^{17}\text{cm}^{-3}$

  $P_{tr} = 1.9\times10^{18}\text{cm}^{-3} \Rightarrow \bar{N}_{tr} = 1.3\times10^{18}\text{cm}^{-3}$

  (~8% larger than undoped)

- Conclusions:
  - To reduce the threshold/transparency current density, n-doping is desirable
    - This is due to the lower effective mass of the conduction band, which populates slower than the valence band for a given shift in Fermi levels
  - To get higher differential gain, p-doping is desirable
    - This is because in the p-doped case, the Fermi levels both shift down in energy and become closer to their respective band edges
  - The effect on transparency current is relatively small in both cases for doping ~10^{18}cm^{-3} (~10-20% difference)
  - Both types of doping will significantly increase the majority carrier concentration, and can increase the Auger rate at transparency if the minority carrier concentration does not substantially decrease

- Overall, how does a quantum well laser improve upon a bulk laser?

1) Lower threshold current, since $J_{th} = \frac{1}{\eta_i} q d \frac{N_{th}}{\tau_e}$
   - In bulk, $d \sim 100\,\text{nm}$ or more. In a quantum well, $d = L_z \sim 10\,\text{nm}$
     ⇒ 10x reduction
   - Possibility of strain to further reduce N_{tr}
2) Higher differential gain : $\frac{dg}{dN}$ is enhanced by the steepness of the DOS function. This is desirable for extending modulation bandwidth
3) Reduced frequency chirping under direct modulation

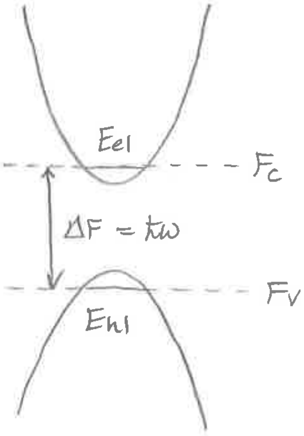
- Benefit of strain

- Consider the hypothetical case where $m_e^* = m_h^* = 0.067 m_0$

We still have:

$$N = \frac{m_e^* kT}{\pi \hbar^2 L_z} \ln(1 + e^{\Delta})$$

$$P = \frac{m_e^* kT}{\pi \hbar^2 L_z} \ln(1 + e^{-\Delta})$$



Assume undoped:

$$N = P \quad \Rightarrow \quad \ln(1 + e^{\Delta}) = \ln(1 + e^{-\Delta})$$

So: $\underline{\Delta = 0}$

Therefore,

$$P_{tr} = \boxed{N_{tr} = \frac{m_e^* kT}{\pi \hbar^2 L_z} \ln 2} = 5.1 \times 10^{17}\,\text{cm}^{-3}$$

There is a 2x reduction in N_{tr}!

- Effect on threshold current: J_{th} goes down as J_{tr} goes down

$$J_{tr} = \frac{1}{\eta_i} q L_z (A N_{tr} + B N_{tr}^2 + C N_{tr}^3) < \frac{1}{2} J_{tr} \text{ of undoped, unstrained QW}$$

2x reduction

4x reduction

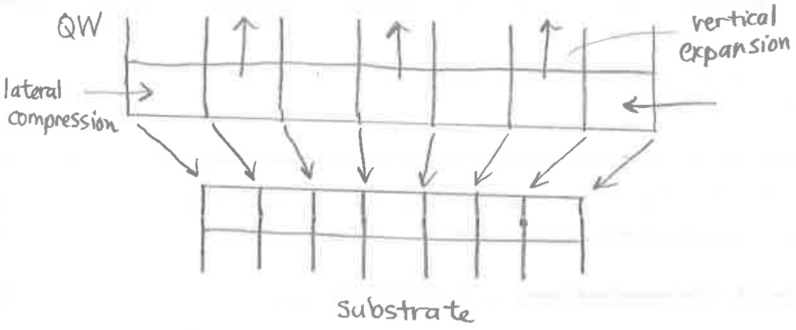
8x reduction!

C is also reduced by strain

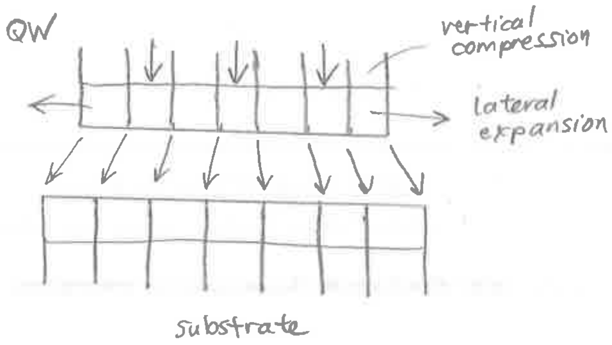
- Note: the L_z dependence in N_{tr} cancels out in J_{tr}.

- Effects of strain on band structure

- Compressive strain: the QW has a larger lattice constant than the substrate



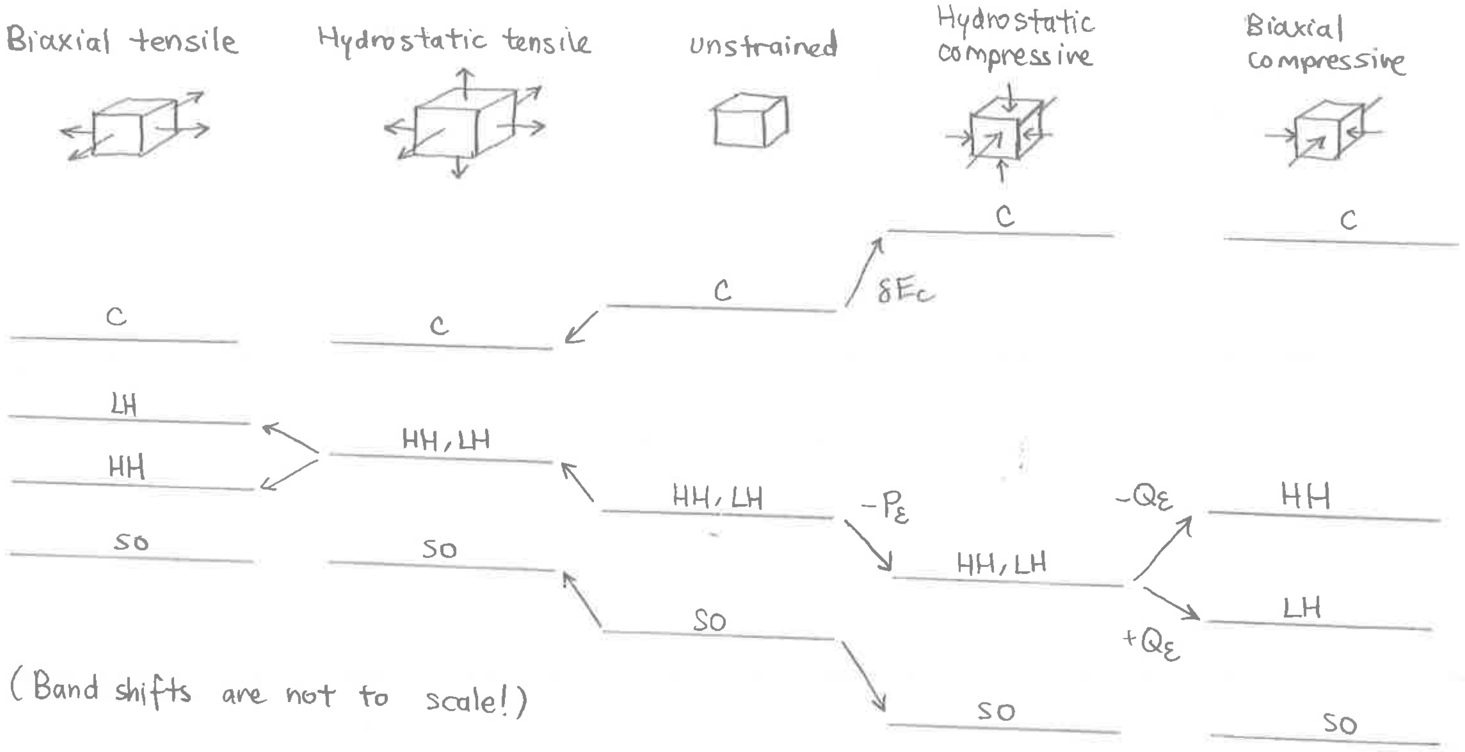
- Tensile strain: the QW has a smaller lattice constant than the substrate



- Directions of strain:
  - Hydrostatic: equal strain in all 3 directions
    - ⇒ shifts energy levels (changes bandgap) but affects all bands equally
  - Biaxial: strain along 2 of 3 directions (useful for QWs)
  - Uniaxial: strain only along 1 direction

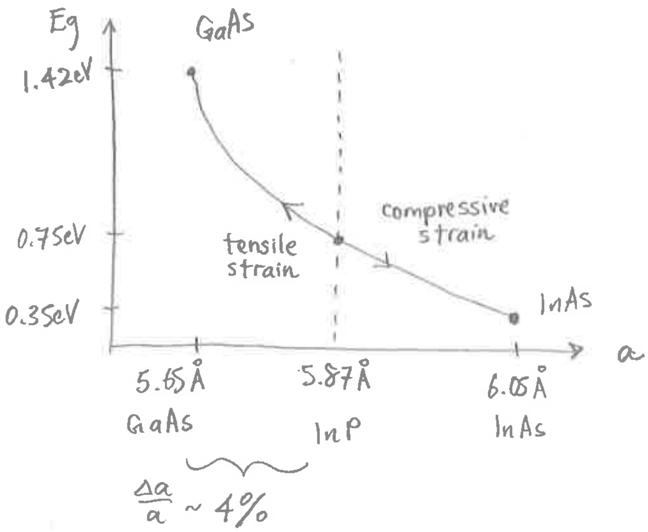
- Strain is only possible if the quantum well is below a critical thickness; above this thickness, the lattice relaxes to its native state, and generates many structural defects in the process

- Qualitative band energy shifts under strain:



(Band shifts are not to scale!)

- To reduce the valence band effective mass in the interband transition, it seems preferable to use tensile strain to bring the LH band higher in energy. But in reality, compressive strain is preferable (for InGaAs) because the effective masses separate into transverse and longitudinal components in a QW.
- Strain & bandgap: we introduce strain commonly by adjusting the Ga/In ratio of InGaAsP grown on InP



- Notice that by applying tensile strain by increasing the Ga content,
  - Eg decreases due to tensile strain effect
  - Eg increases due to larger Ga content

$\downarrow$

Net effect: Eg increases

w/ application of stress

$$\varepsilon \equiv \frac{a_0 - a(x)}{a_0} = \frac{\Delta a}{a_0}$$

- Stress/strain analysis
  - Stress: deforming force per unit area applied in a particular direction $\rightarrow \sigma$
  - Strain: relative change in length caused by deforming force $\rightarrow \varepsilon$
- In general, both σ and ε are 3×3 tensors. We will assume $\sigma_{ij} = 0$ if $i \neq j$ meaning only normal forces $\sigma_x, \sigma_y, \sigma_z$ act to deform the lattice; no shear or rotation occurs. This means we also only consider ε_{ij}, $i=j$.
- The stress and strain are related by the stiffness tensor C:

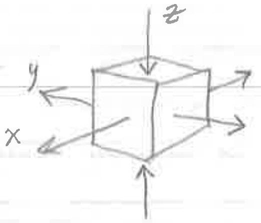
$$\begin{bmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{bmatrix} \cong \begin{bmatrix} C_{11} & C_{12} & C_{12} \\ C_{12} & C_{11} & C_{12} \\ C_{12} & C_{12} & C_{11} \end{bmatrix} \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{bmatrix}$$

Cubic symmetry:
- All diagonal elements equal $\rightarrow C_{11}$
- All off-diagonal elements equal $\rightarrow C_{12}$

Biaxial strain: $\sigma_x = \sigma_y > 0$ (tensile) $\quad \sigma_x = \sigma_y < 0$ (compressive)

$\sigma_z = 0$

- Consider $\sigma_x = \sigma_y$, $\sigma_z = 0$ so we also have $\varepsilon_x = \varepsilon_y$



$$\sigma_{xy} = C_{11}\,\varepsilon_{xy} + C_{12}\,\varepsilon_{xy} + C_{12}\,\varepsilon_z$$

$$0 = C_{12}\,\varepsilon_{xy} + C_{12}\,\varepsilon_{xy} + C_{11}\,\varepsilon_z$$

$$\Rightarrow \quad \varepsilon_z = \frac{-2C_{12}}{C_{11}}\,\varepsilon_{xy}$$

- Under biaxial strain, the deformation along z will be opposite in sign from the deformation along x, y
(like squeezing effect)

- Combining this with band structure models, we get:

$$\delta E_c = a_c(\varepsilon_x + \varepsilon_y + \varepsilon_z) = a_c\left(2\varepsilon_{xy} - \frac{2C_{12}}{C_{11}}\varepsilon_{xy}\right)$$

$$\delta E_c = 2a_c\left(1 - \frac{C_{12}}{C_{11}}\right)\varepsilon_{xy}$$

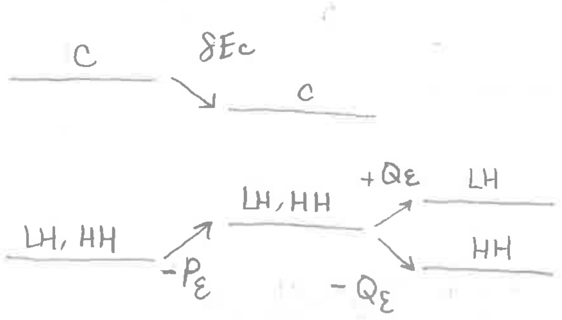
$$P_\varepsilon = -2a_v\left(1 - \frac{C_{12}}{C_{11}}\right)\varepsilon_{xy}$$

$$Q_\varepsilon = -b\left(1 + 2\frac{C_{12}}{C_{11}}\right)\varepsilon_{xy}$$

$a = a_c - a_v$
= hydrostatic potential

b = shear potential

where the definitions are:



- Assuming a direct bandgap, the band edges (at the minimum $k=0$) are:

$$E_C = E_g(x) + \delta E_c(x)$$

$$E_{HH} = -P_\varepsilon(x) - Q_\varepsilon(x)$$

$$E_{LH} = -P_\varepsilon(x) + Q_\varepsilon(x)$$

where energy is referenced to E_{HH} at $k=0$ without strain.

- Tensile vs. compressive strain
  - Tensile: $\delta E_c < 0$, $P_\varepsilon < 0$, $Q_\varepsilon > 0$
  - Compressive: $\delta E_c > 0$, $P_\varepsilon > 0$, $Q_\varepsilon < 0$

- The dispersion relations away from the center of k-space are given by:

$$E_{HH}(k) = -P_\varepsilon(x) - Q_\varepsilon(x) - \frac{\hbar^2}{2m^t_{hh}}k_t^2 - \frac{\hbar^2}{2m^z_{hh}}k_z^2$$

$$E_{LH}(k) = \underbrace{-P_\varepsilon(x) + Q_\varepsilon(x)}_{k=0} - \underbrace{\frac{\hbar^2}{2m^t_{lh}}k_t^2}_{\text{transverse dispersion}} - \underbrace{\frac{\hbar^2}{2m^z_{lh}}k_z^2}_{\text{longitudinal dispersion}}$$

(without valence band mixing)

transpareny carrier density

- Use m^t to calculate the 2D DOS (which goes into the gain expression)
- Use m^z to calculate the energy levels of the QW subbands

- Effective masses: HH vs. LH

$$m^t_{hh} = \frac{m_0}{\gamma_1 + \gamma_2}, \quad m^t_{lh} = \frac{m_0}{\gamma_1 - \gamma_2}$$

$$m^z_{hh} = \frac{m_0}{\gamma_1 - 2\gamma_2}, \quad m^z_{lh} = \frac{m_0}{\gamma_1 + 2\gamma_2}$$

where $\gamma_1, \gamma_2 > 0$ are band structure parameters

In a quantum well,

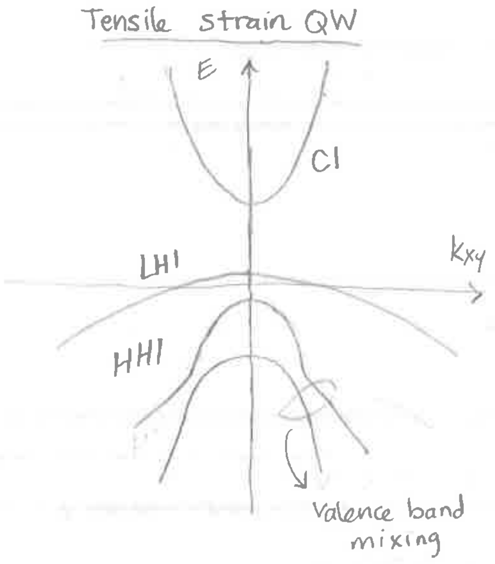
- The heavy hole has a lighter transverse effective mass!
  - ⇒ HH → C transition is preferred to reduce N_{tr}
- The light hole has a lighter longitudinal effective mass
- The values of the effective mass change with strain, however

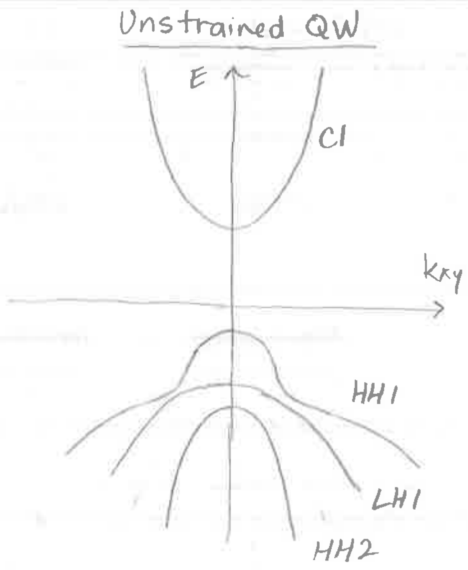
↙

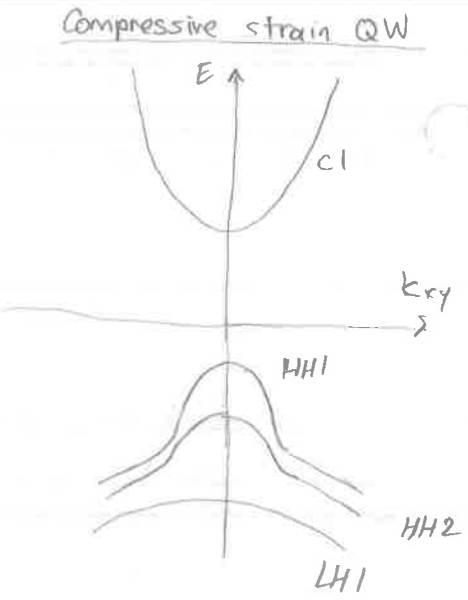
* An unstrained quantum well seems to have a DOS advantage over a bulk laser:

1) HH has a lower transverse DOS than LH in a QW
2) HH has a heavier m^z, so it is closer to the conduction band than LH ⇒ HH → C transitions are already preferred over LH → C transitions

- The benefit of strain:







- LH1 → C1 preferred (not desirable)
- HH1 → C1 preferred
- HH1 → C1 further preferred (desirable)

- Compressive strain moves the LH bands further from the HH bands, leading to a lower valence band effective mass due to a stronger preference for the HH → C transition

  ⇒ lower N_{tr}, J_{th}

- A lower m_h^* (more comparable to m_e^*) also brings the Fermi levels closer to the band edges on both sides. Since the Fermi occupancy function changes most rapidly at the Fermi level, this means that the carrier density becomes more sensitive to Fermi level shifts

  ⇒ larger differential gain dg/dN

  (this can be achieved with p-doping as well, but this increases N_{tr} due to the extra holes that need to be supplied)

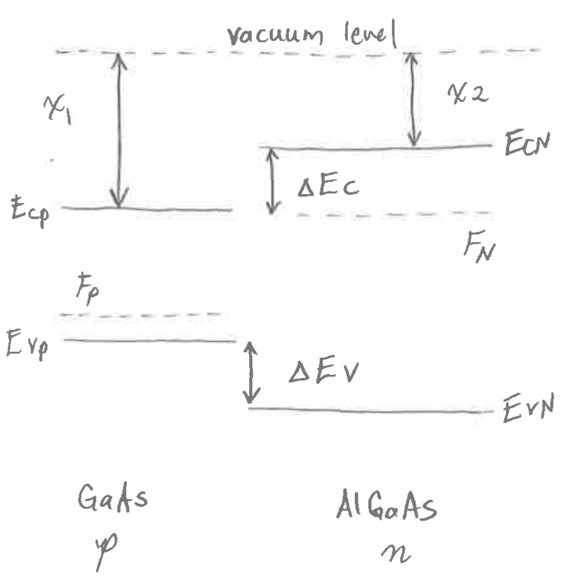
- Polarization dependence
  - Compressive strain: most transitions are HH → C, so the gain is more responsive to TE polarization (also the case with unstrained, but less so)
  - Tensile strain: most transitions are LH → C, so the gain is more responsive to TM polarization; at moderate tensile strain, the gain is independent of polarization (desirable for optical amplifiers)
  - $\boxed{\text{Compressive} \rightarrow \text{HH transitions} \rightarrow \text{TE}}$ $\boxed{\text{Tensile} \rightarrow \text{LH transitions} \rightarrow \text{TM}}$

⑦ Heterojunctions and the Double Heterostructure

- Attaining the transparency condition $\Delta F = E_g$ is difficult to achieve in practice in a normal pn-diode, because the current becomes too large. Today, most lasers use the double heterostructure

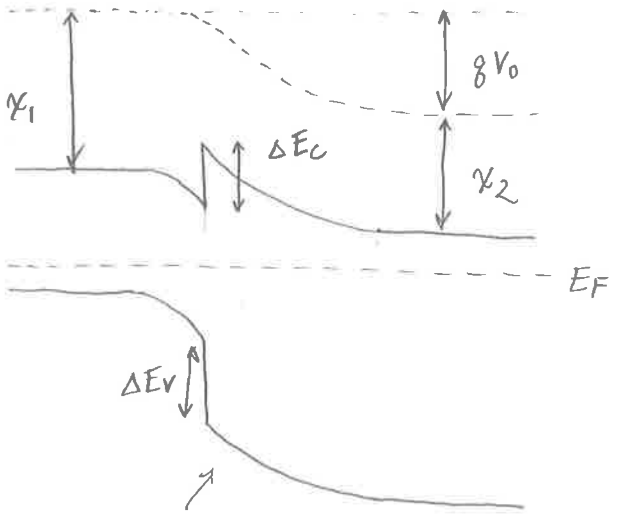
- Band alignment

No bias, before forming junction:



→

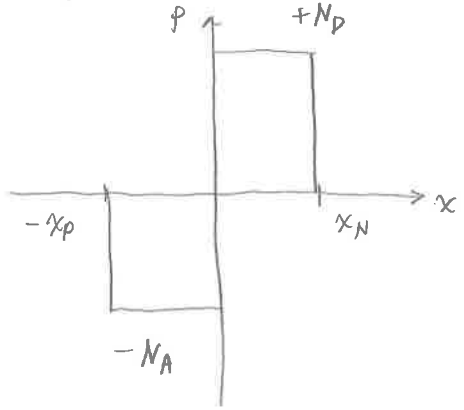
After contact, line up the Fermi levels:



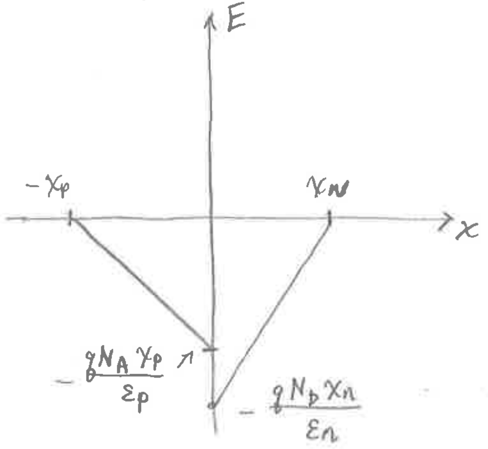
Band bending: e^- transfers out of N-type in this part, so it bends upward

- Electrostatics

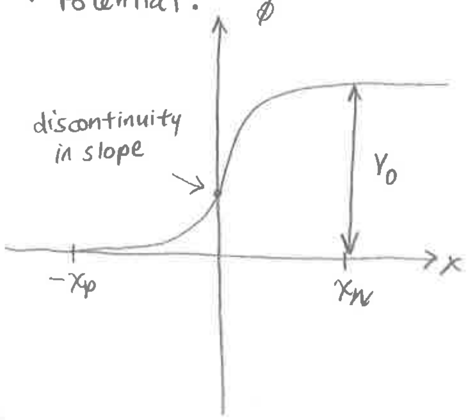
- Charge density:



- Electric field:



- Potential:



Gauss's law: $\frac{d(\varepsilon E)}{dx} = \rho$

Poisson's eq: $\frac{dV}{dx} = -E(x)$

- The built-in potential is:

from Fermi levels before contact

$$qV_0 = F_N - F_p = E_{gp} + \Delta E_c - (E_{CN} - F_N) - (F_p - E_{Vp})$$

$$= E_{gp} + \Delta E_c - kT \ln \frac{N_{CN}}{N_D} - kT \ln \frac{N_{Vp}}{N_A}$$

$$qV_0 = E_{gp} + \Delta E_c + kT \ln \frac{N_A N_D}{N_{CN} N_{Vp}}$$

- Built-in potential

$$\boxed{qV_0 = E_{gp} + \Delta E_c + \underbrace{kT \ln \frac{N_A N_D}{N_{CN} N_{vp}}}_{<0 \text{ if non-degenerately doped}}}$$

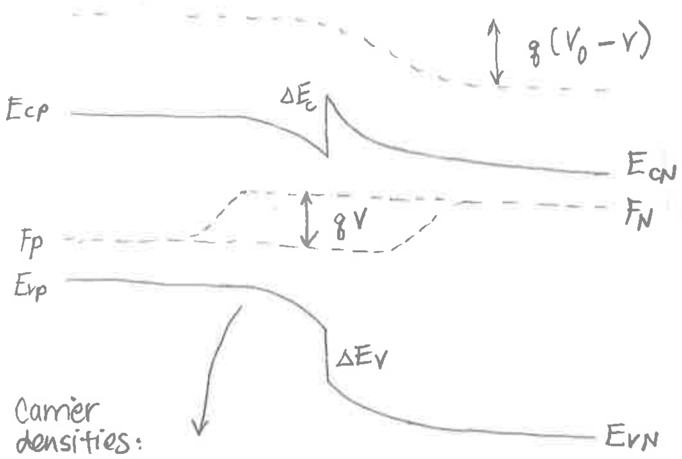
(if degenerately doped, this expression is invalid)

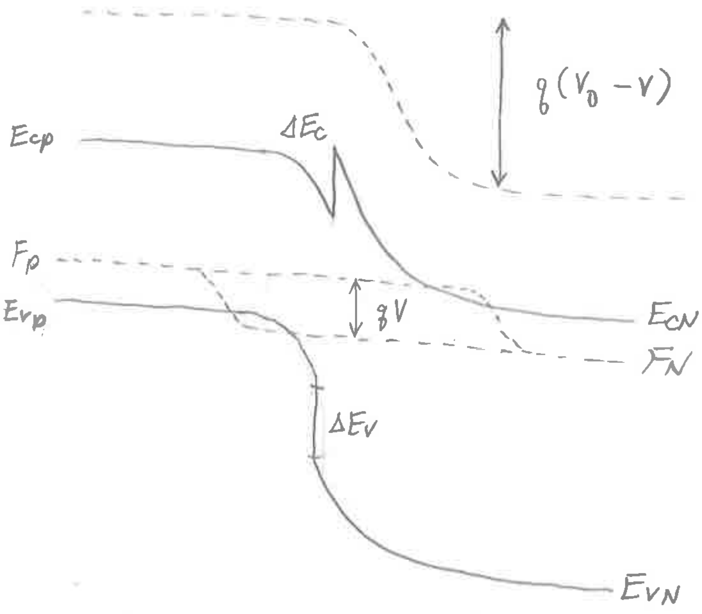
- The valence band $E_V(x)$ follows $-q\phi(x)$, but with a discontinuity of ΔE_V at the junction
- The conduction band $E_c(x) = E_V(x) + E_g(x)$; again, there is a discontinuity at the junction, due to both ΔE_V and the change in E_g

- Biased heterojunction

Forward bias: $V > 0$

Reverse bias: $V < 0$





Carrier densities:

$$n_p(x) = N_{cp}\, e^{-(E_c(x) - F_N(x))/kT}$$
$$p_p(x) = N_{vp}\, e^{-(F_p(x) - E_V(x))/kT}$$
$$N_N(x) = N_{CN}\, e^{-(E_c(x) - F_N(x))/kT}$$
$$p_N(x) = N_{VN}\, e^{-(F_p(x) - E_V(x))/kT}$$

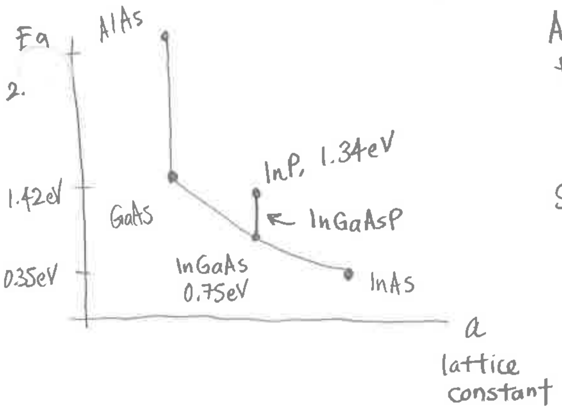
$$\Rightarrow \quad n_p(x) p_p(x) = N_{cp} N_{vp}\, e^{(E_V(x) - E_c(x))/kT}\, e^{(F_N(x) - F_p(x))/kT}$$
$$= N_{cp} N_{vp}\, e^{-E_g(x)/kT}\, e^{qV/kT}$$

$$n_p(x) p_p(x) = n_{ip}^2\, e^{qV/kT}$$

The law of mass action still holds

- Double heterostructure laser basics:

- Communication wavelength : $\lambda = 1.55\mu m$

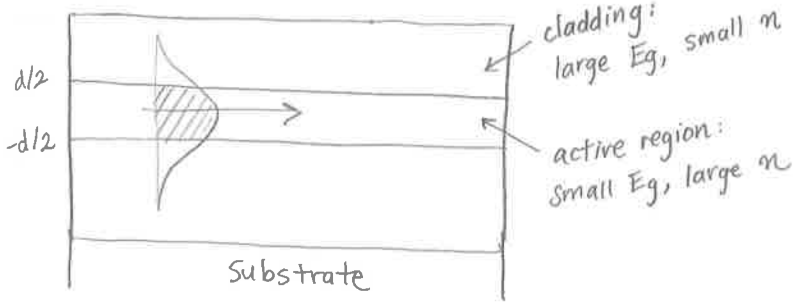


Absorption minimum in silica, optimal for transmission through optical fiber → InGaAsP is commonly used

$\lambda = 1.30\mu m$

Smallest $dn/d\lambda$: least material dispersion

- Edge-emitting laser:



- At a given wavelength, approximate the gain as a linear function of the carrier concentration:

$$g(N) = a \cdot (N - N_{tr})$$

Threshold gain (see §①)

$$g_{th} = \frac{1}{\Gamma}(\alpha_i + \alpha_m) = a(N_{th} - N_{tr})$$

$$\Rightarrow N_{th} = N_{tr} + \frac{1}{\Gamma a}(\alpha_i + \alpha_m)$$

- Threshold current

use confinement to reduce Jth

$$\boxed{J_{th} \approx \frac{qN_{th}}{\tau_e} d} = \frac{qd}{\tau_e}\left(N_{tr} + \underbrace{\frac{1}{\Gamma a}(\alpha_i + \alpha_m)}_{\text{loss increases threshold current}}\right)$$

(undoped approximation)

reduce N_{tr} to lower threshold current (e.g. by using a QW)

- The carrier lifetime τ_e is determined by how long it takes a carrier to recombine:

$$\text{recombination rate per volume } (cm^{-3}s^{-1}) = \frac{N}{\tau_e} \approx AN + BN^2 + CN^3$$

↑ Shockley-Reed-Hall recombination

Radiative recombination

Auger recombination

- Current in a heterojunction

Although the derivation is lengthy, we can show that the current across a heterojunction is given by:

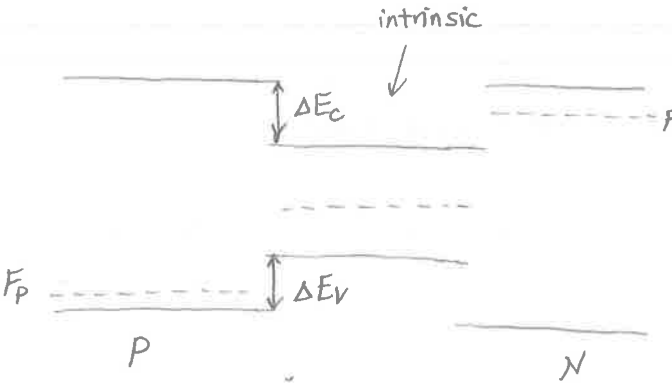
$$J = q\left(\frac{D_n}{L_n} n_{p0} + \frac{D_p}{L_p} p_{N0}\right)\left(e^{qV/kT} - 1\right)$$

electrons in p $\quad$ holes in N

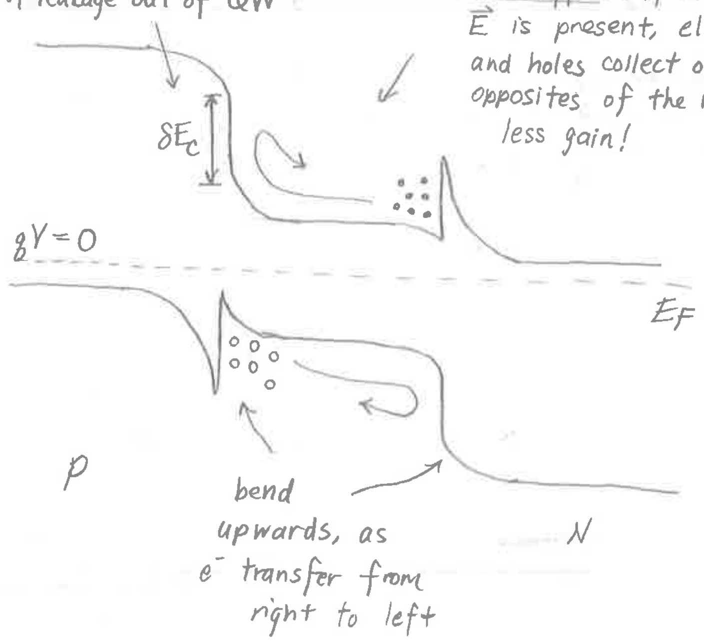
$n_{p0} = n_{ip}^2 / N_A$ $\quad$ $p_{N0} = n_{iN}^2 / N_D$

- Double heterostructure

- Carrier confinement: high concentration of electrons and holes in a small-bandgap region



⇒

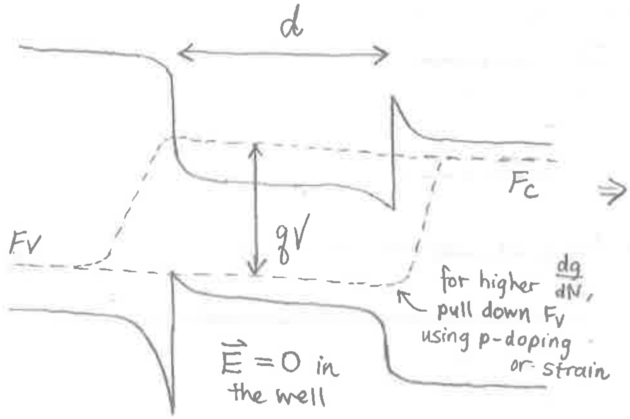


Band offset determines the amount of leakage out of QW

Quantum confined Stark effect: if an internal $\vec{E}$ is present, electrons and holes collect on opposites of the well: less gain!

bend upwards, as e^- transfer from right to left

Under forward bias:

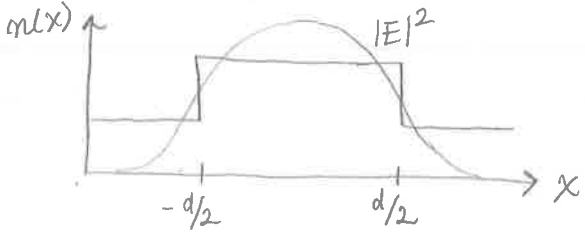


for higher $\frac{dg}{dN}$, pull down F_V using p-doping or strain

$\vec{E} = 0$ in the well

⇒ large ΔF in small-bandgap region, away from either heterojunction (this is like the heterostructure equivalent of a p-i-n diode)

- Optical confinement: structure is a slab waveguide along x



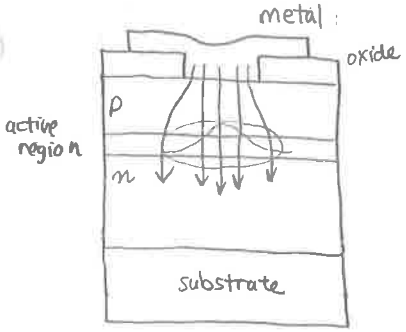
- Confinement factor

$$\Gamma = \frac{\int_{-d/2}^{d/2} |\vec{E}|^2 \, dx}{\int_{-\infty}^{\infty} |\vec{E}|^2 \, dx}$$

* To increase Γ while using small QWs for current confinement, use multiple quantum wells (MQW)

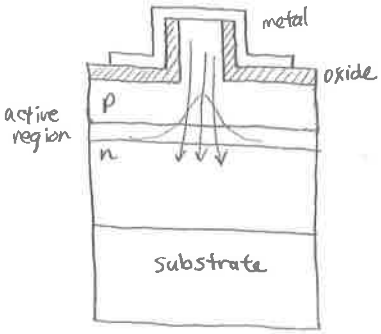
- Common DH laser structures

  - Gain-guided laser



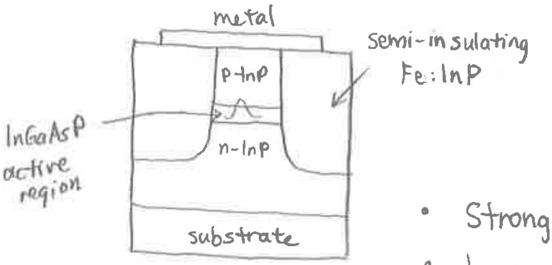
    - Gain guiding: the oxide layer ensures that only a small part of the active region provides gain
      ⇒ weak guiding of the optical mode
    - Current can spread laterally from oxide to reach the active region (weak current confinement)
      ⇒ larger I_{th}
    - Higher order side modes are easily generated (a bad thing!)

  - Ridge waveguide laser



    - Reduces lateral current spreading
      ⇒ reduces I_{th} compared to gain guiding structure
    - Provides weak lateral index guiding of the optical mode due to the ridge above the active region
      - Relatively small confinement factor Γ
        ⇒ relatively high threshold gain needed ⇒ $I_{th}\uparrow$

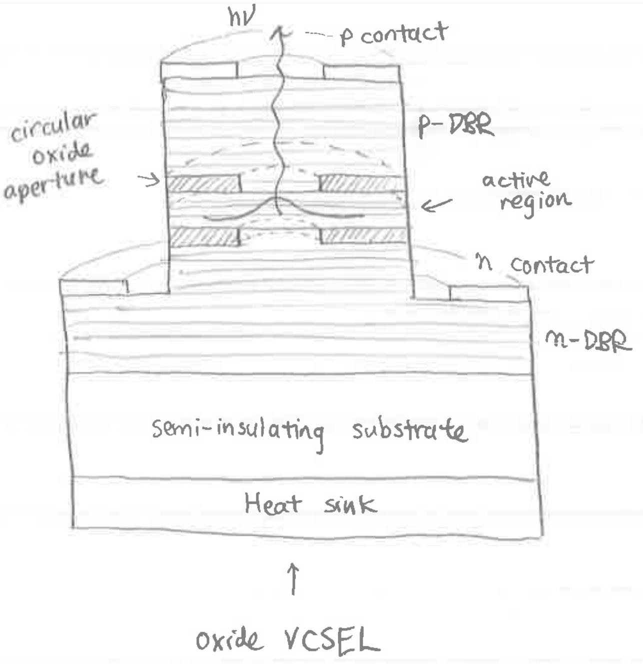
  - Buried heterostructure laser

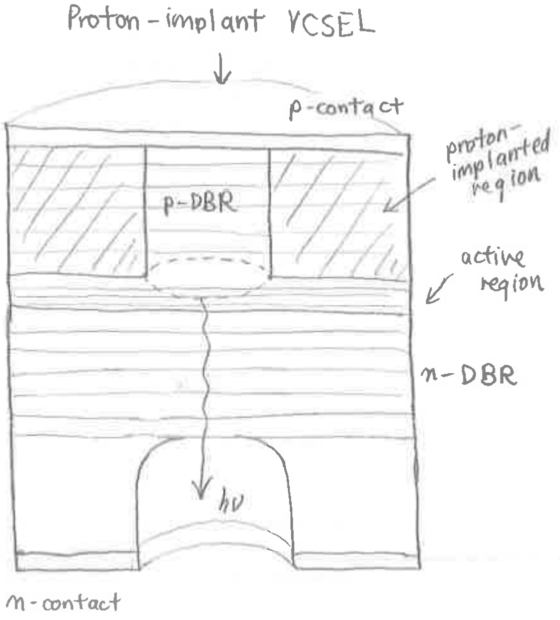


    - Improves current confinement by surrounding the active region with heterojunctions on all sides, including semi-insulating layers ⇒ $I_{th}\downarrow$
    - Stronger optical confinement with larger Δn
    - Lower parasitic capacitance : improves bandwidth

- Optical cavities for edge-emitting lasers
  - Fabry-Perot: two-mirror arrangement (see § ⑦)
    - Many longitudinal modes within gain spectrum ⇒ limits bandwidth
  - Distributed feedback (DFB) laser: mode selectivity by placing a diffractive grating directly above active region
    - Widely used for 1550 nm long-haul optical communications
  - Distributed Bragg reflector (DBR) laser: mode selectivity by introducing periodic index variation at the two ends of laser cavity
    - Like Fabry-Perot with wavelength selective dielectric mirrors

- Vertical cavity surface-emitting laser (VCSEL)

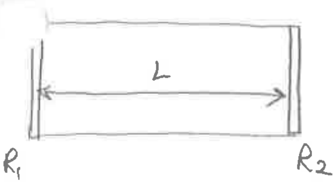




- Emits out of the top or bottom surface, rather than the edges
  - No need for optically flat facets
  - No need for cleaving/dicing ⇒ batch array processing
    ⇒ much cheaper!
  - Easily coupled to optical with a circular aperture
  - Small cavity volume
    - Low threshold current I_{th}
    - Single mode operation, but large linewidth $\Delta\lambda$
- Cavity consists of an active region surrounded by DBR stacks with $R > 99\%$
  - Not practical for 1.3-1.55 μm (long haul), as too many DBR pairs would be needed
- Current confinement can be provided in several ways
  - Oxide VCSEL: oxidized AlGaAs apertures provide both lateral current confinement and index confinement of optical mode
  - Proton implantation: parts of the structure that are implanted with protons become insulating
    - Weaker optical confinement, relies on gain guiding
- The cavity length is usually $\sim \lambda / n$ (one wavelength) which results in a peak of the electric field at the center of the cavity between the DBRs
- The DBR layers can be made to have low resistivity by using graded, highly doped heterojunctions

⑧ Optical modes in a laser cavity

- **Longitudinal modes**: modes along propagation direction



The electric field has the form $\vec{E}(x,y,z) = \vec{E}_0(x,y)\, e^{ik_z z}$

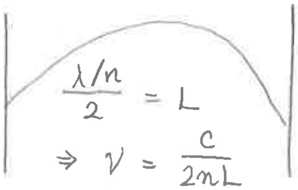
transverse field profile ↓ ($\vec{E}_0(x,y)$)

Round trip condition gives the allowed k_z:

$$e^{ik_z(2L)} = 1 \quad \Rightarrow \quad 2k_z L = 2\pi m$$

$$k_z = \frac{m\pi}{L}$$

$m = 1$:



$\frac{\lambda/n}{2} = L$

$\Rightarrow \nu = \frac{c}{2nL}$

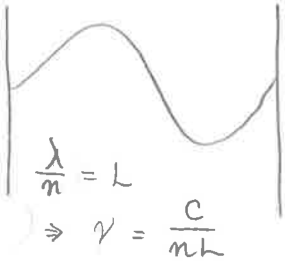
- Use the dispersion relation for light to find the mode frequencies:

$$k_z = \frac{m\pi}{L} = \frac{n\omega}{c} = \frac{2\pi n \nu}{c}$$

$$\boxed{\nu_m = m\frac{c}{2nL}}$$

This is obvious simply from inspecting the cavity, as shown on left

$m = 2$:



$\frac{\lambda}{n} = L$

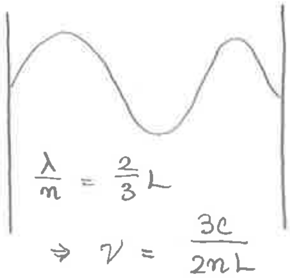
$\Rightarrow \nu = \frac{c}{nL}$

- The mode spacing is

$$\Delta\nu = \frac{c}{2nL} \quad \text{or since} \quad d\nu = -\frac{c}{\lambda^2}d\lambda$$

$$\Delta\lambda = -\frac{\lambda^2}{2nL}$$

$m = 3$:



$\frac{\lambda}{n} = \frac{2}{3}L$

$\Rightarrow \nu = \frac{3c}{2nL}$

- If we include material dispersion,

$$\Delta\nu = \nu_m - \nu_{m-1} = \frac{mc}{2n_m L} - \frac{(m-1)c}{2n_{m-1}L}$$

$$= \frac{c}{2n_m L} + \frac{(m-1)c}{2L}\left[\frac{1}{n_m} - \frac{1}{n_{m+1}}\right]$$

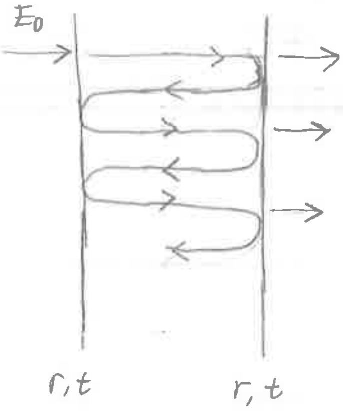
$$= \frac{c}{2nL} + \nu_{m-1} n_{m-1}\left[\frac{-\Delta n}{n^2}\frac{1}{\Delta\nu}\Delta\nu\right]$$

$$= \frac{c}{2nL} - \frac{\nu}{n}\frac{dn}{df}\Delta\nu$$

$$\Rightarrow \boxed{\Delta\nu = \frac{c}{2nL}\left(1 + \frac{\nu}{n}\frac{dn}{d\nu}\right)^{-1}} \quad \text{or} \quad \Delta\lambda = \frac{-\lambda^2}{2nL\left(1 - \frac{\lambda}{n}\frac{dn}{d\lambda}\right)}$$

↑ without dispersion

Transmission characteristics of longitudinal modes



$$\rightarrow E_0 t_1 t_2 e^{ik_zL}$$

$$\rightarrow E_0 t_1 r_2 r_1 t_2 e^{ik_z3L}$$

$$\rightarrow E_0 t_1 (r_2r_1)^2 t_2 e^{ik_z5L}$$

Add coherently:

$$\Rightarrow E_T = E_0 t_1 t_2 e^{ik_zL}\left(1 + r_1r_2 e^{ik_z2L} + (r_1r_2e^{ik_z2L})^2 + \dots\right)$$

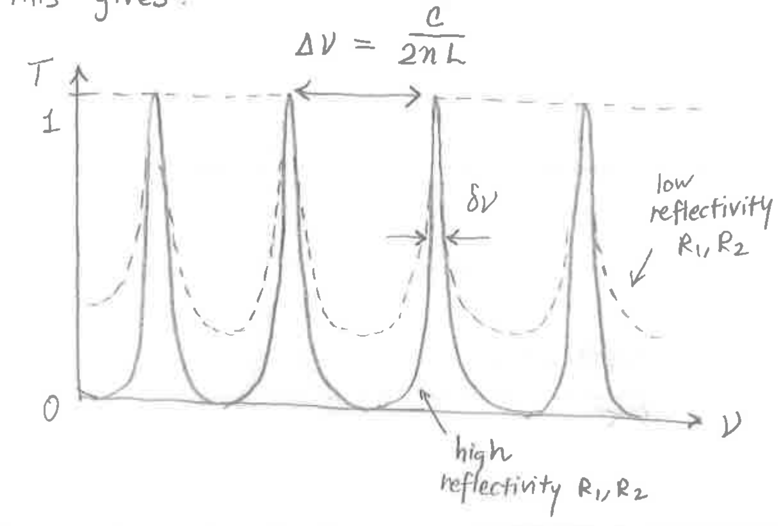
$$t = \frac{t_1 t_2 e^{ik_zL}}{1 - r_1r_2e^{ik_z2L}}$$

If internal and external media have the same index,

$$T = |t|^2 = \frac{|t_1t_2|^2}{|1 - r_1r_2e^{ik_z2L}|^2} = \underbrace{\frac{|t_1t_2|^2}{(1-|r_1r_2|)^2}}_{\substack{T_{max}\\=100\%}} \frac{1}{1+\left(\frac{2F}{\pi}\right)^2\sin^2(k_zL)}$$

$$F = \frac{\pi\sqrt{|r_1r_2|}}{1-|r_1r_2|} = \frac{\pi(R_1R_2)^{1/4}}{1-(R_1R_2)^{1/2}}$$

This gives:



- The finesse F describes the sharpness of each peak:

$$\text{FWHM: } \delta\nu = \frac{\Delta\nu}{F}$$

- The quality factor is therefore:

$$Q_{FP} = \frac{\nu}{\delta\nu} = \frac{\nu}{\Delta\nu}F$$

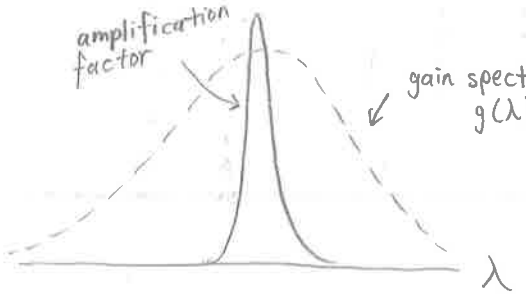
$$\boxed{Q_{FP} = \frac{2nL}{\lambda}\cdot\frac{\pi(R_1R_2)^{1/4}}{1-(R_1R_2)^{1/2}}}$$

$n=1$
$\lambda = 1550\,nm$
$L = 1\,\mu m$
$R_1 = 0.95$
$R_2 = 0.995$
↓
$Q \approx 144$
$T_p \approx 1.2ps$

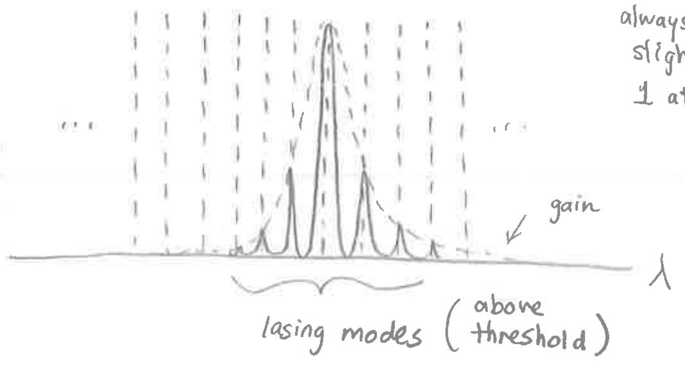
- Laser modes: overlap the gain spectrum onto Fabry Perot modes. The amplification factor due to gain is:

$$Amp = 1 + R_1R_2e^{2gL} + (R_1R_2e^{2gL})^2 + \dots = \frac{1}{1-R_1R_2e^{2gL}}$$

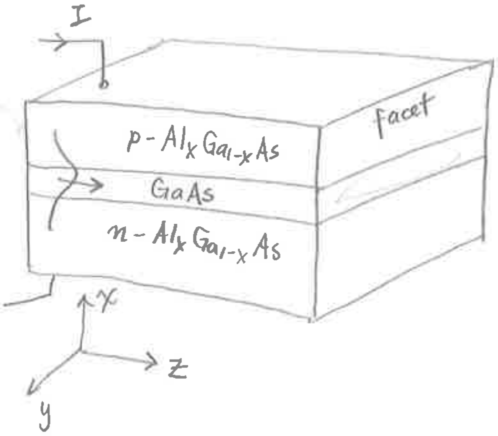
this value will always be clamped slightly below 1 at gain peak

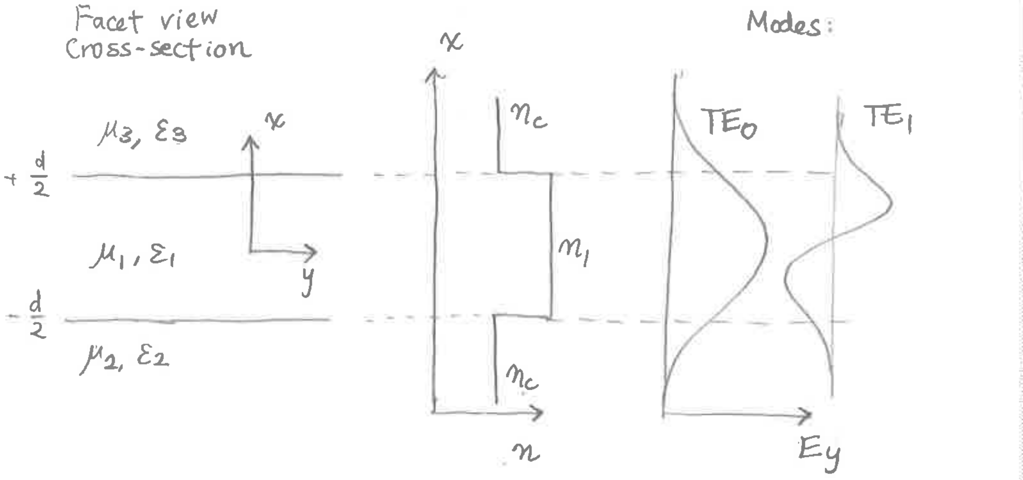


$\Rightarrow$

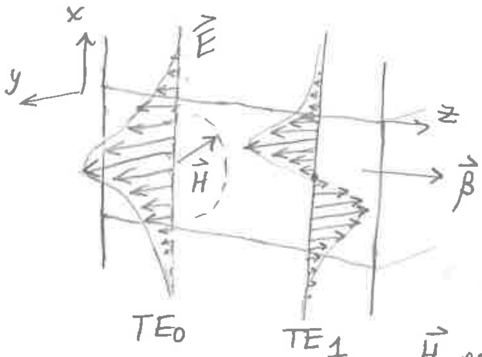


- Dielectric slab waveguide

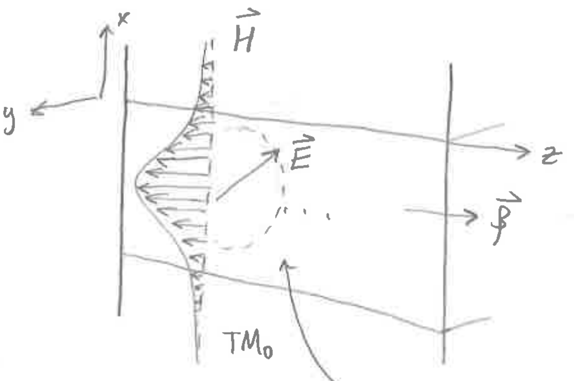




- TE modes: $\vec{E} = E_y \cdot \hat{y}$ - TM modes: $\vec{H} = H_y \hat{y}$

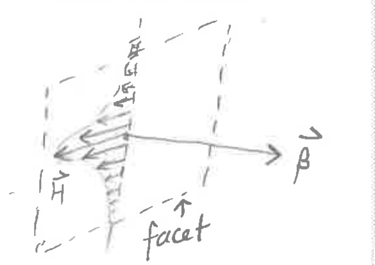


$\vec{H}$ points in the xz plane, $H_y = 0$



$\vec{E}$ points in the xz plane, $E_y = 0$

Emerges as TM polarized Gaussian beam



Assume structure is extended infinitely along y: All $\frac{\partial}{\partial y} \to 0$.

- Apply the electromagnetic wave equation

$$\vec{\nabla}^2 \vec{E} = -\omega^2 \mu \varepsilon \vec{E} \quad \Rightarrow \quad (\vec{\nabla}^2 + \omega^2 \mu \varepsilon) E_y = 0$$

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + \omega^2 \mu \varepsilon \right) E_y = 0$$

Guess that the solution will be of the form $E_y = \phi(x) e^{i k_z z} e^{-i\omega t}$ and that furthermore,

$$\phi(x) = \begin{cases} C_3 e^{-\alpha_3 (x - d/2)} & x > \frac{d}{2} \\ C_1 \cos(k_x x) + D_1 \sin(k_x x) & -\frac{d}{2} < x < \frac{d}{2} \\ C_2 e^{\alpha_2 (x + d/2)} & x < -\frac{d}{2} \end{cases}$$

C_1 corresponds to even modes: TE_0, TE_2, ...
D_1 corresponds to odd modes: TE_1, TE_3, ...

These can be written as plane waves, so E_y satisfies the wave equation

- Now apply boundary conditions at $x = \pm \frac{d}{2}$

1) E_y must be continuous (transverse to interface)

2) H_x and H_z must be continuous

Use Faraday's law to get $\vec{H}$

$$\vec{\nabla} \times \vec{E} = -\mu \frac{\partial \vec{H}}{\partial t}$$

$$\begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ 0 & E_y & 0 \end{vmatrix} = -\frac{\partial E_y}{\partial z}\hat{x} + \frac{\partial E_y}{\partial x}\hat{z} = +\mu i\omega \vec{H}$$

$$\vec{H} = \frac{1}{i\omega\mu}\left(-ik_z E_y \hat{x} + \frac{\partial E_y}{\partial x}\hat{z}\right)$$

$$H_x = -\frac{k_z}{\omega\mu}E_y \quad , \quad H_z = \frac{1}{i\omega\mu}\frac{\partial E_y}{\partial x}$$

- H_x is automatically continuous if E_y is continuous, assuming $\mu = \mu_0$ for all materials
- H_z is continuous if $\frac{\partial E_y}{\partial x}$ is continuous

Thus:

At $x = -\frac{d}{2}$:

$$C_2 = C_1 \cos\left(-\frac{k_x d}{2}\right) + D_1 \sin\left(-\frac{k_x d}{2}\right) \quad ①$$

$$\alpha_2 C_2 = -k_x C_1 \sin\left(-\frac{k_x d}{2}\right) + k_x D_1 \cos\left(-\frac{k_x d}{2}\right) \quad ②$$

At $x = +\frac{d}{2}$:

$$C_3 = C_1 \cos\left(\frac{k_x d}{2}\right) + D_1 \sin\left(\frac{k_x d}{2}\right) \quad ③$$

$$-\alpha_3 C_3 = -k_x C_1 \sin\left(\frac{k_x d}{2}\right) + k_x D_1 \cos\left(\frac{k_x d}{2}\right) \quad ④$$

We have also 3 additional equations from the wave equation: $\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + \omega^2\mu\varepsilon\right)E_y = 0$

Region 3: $-\alpha_3^2 + k_z^2 = \omega^2\mu_0\varepsilon_3$

Region 1: $k_x^2 + k_z^2 = \omega^2\mu_0\varepsilon_1$ ⑤

Region 2: $-\alpha_2^2 + k_z^2 = \omega^2\mu_0\varepsilon_2$

* Let's assume symmetry in x:

$\varepsilon_2 = \varepsilon_3 = \varepsilon_c \Rightarrow \alpha_2 = \alpha_3 = \alpha$

$\Rightarrow -\alpha^2 + k_z^2 = \omega^2\mu_0\varepsilon_c$ ⑥

- Even modes: symmetric, $C_2 = C_3$

① + ③ : $C_2 + C_3 = 2C_2 = 2C_1 \cos\left(\frac{k_x d}{2}\right) \Rightarrow \quad C_2 = C_1 \cos\left(\frac{k_x d}{2}\right)$

② - ④ : $\alpha_2 C_2 + \alpha_3 C_3 = 2\alpha C_2 = 2k_x C_1 \sin\left(\frac{k_x d}{2}\right) \Rightarrow \quad C_2 = \frac{k_x}{\alpha} \sin\left(\frac{k_x d}{2}\right)$

Combining these, we get: $\boxed{\alpha = k_x \tan\left(k_x \frac{d}{2}\right)}$ Even modes

- Odd modes: assymetric : $C_2 = -C_3$

① - ③ : $C_2 - C_3 = 2C_2 = -2D_1 \sin\left(\frac{k_x d}{2}\right) \Rightarrow \quad C_2 = -D_1 \sin\left(\frac{k_x d}{2}\right)$

② + ④ : $\alpha_2 C_2 - \alpha_3 C_3 = 2\alpha C_2 = 2k_x D_1 \cos\left(\frac{k_x d}{2}\right) \Rightarrow \quad C_2 = \frac{k_x}{\alpha} D_1 \cos\left(\frac{k_x d}{2}\right)$

$\Rightarrow \boxed{\alpha = -k_x \cot\left(k_x \frac{d}{2}\right)}$ Odd modes

- How do we interpret these characteristic equations?
If a solution exists, a mode exists. Otherwise, the mode does not exist. If multiple solutions exist, multiple modes can propagate.

Let $X = k_x \frac{d}{2}$, $Y = \alpha \frac{d}{2}$

⑤ - ⑥ : $k_x^2 + \alpha^2 = \omega^2 \mu_0 (\varepsilon_1 - \varepsilon_c)$

$$\left(\frac{k_x d}{2}\right)^2 + \left(\frac{\alpha d}{2}\right)^2 = \left(\frac{d}{2}\right)^2 \omega^2 \mu_0 (\varepsilon_1 - \varepsilon_c)$$

$$\boxed{X^2 + Y^2 = \left(\frac{d}{2}\right)^2 \omega^2 \mu_0 \varepsilon_0 (n_1^2 - n_c^2)}$$

Since RHS is constant (not dependent on k_x or α), this is an equation for a circle in the X-Y plane

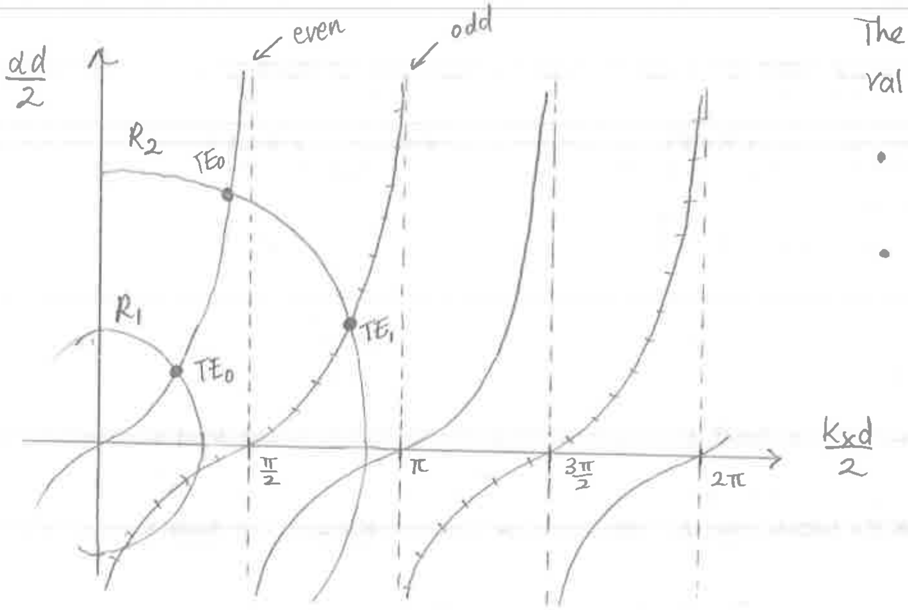
Meanwhile, the characteristic equations can be written as:

$$Y = X \tan X, \qquad Y = -X \cot X$$

(even) (odd)

Plotting the circle together with these expressions gives us a way to find the modes

- Graphical solution:



The two circles correspond to different values of $R^2 = (\frac{d}{2})^2\omega^2\mu_0\varepsilon_0(n_1^2 - n_c^2)$

- As R increases, more modes can exist
- To increase R,
  - Increase the size, d
  - Increase the index contrast $n_1^2 - n_c^2$

- The first even mode always exists. The condition for having a single waveguide mode is

$$R < \frac{\pi}{2} \Rightarrow \frac{d}{2}\omega\sqrt{\mu_0\varepsilon_0}\sqrt{n_1^2 - n_c^2} = \frac{d}{2}\frac{\omega}{c}\sqrt{n_1^2 - n_c^2} = \frac{d}{2}k_0\sqrt{n_1^2 - n_c^2} < \frac{\pi}{2}$$

Single-mode condition: $\boxed{k_0\frac{d}{2}\sqrt{n_1^2 - n_c^2} < \frac{\pi}{2}}$

Another way to write this is: $d\sqrt{n_1^2 - n_c^2} < \frac{\lambda_0}{2}$

$$d^2(n_1^2 - n_c^2) < \left(\frac{\lambda_0}{2}\right)^2$$

small index contrast

$$\approx d^2(2n_1)\Delta n < \frac{\lambda_0^2}{4}$$

$$\Rightarrow \boxed{d^2\Delta n < \frac{\lambda_0^2}{8n_1}}$$

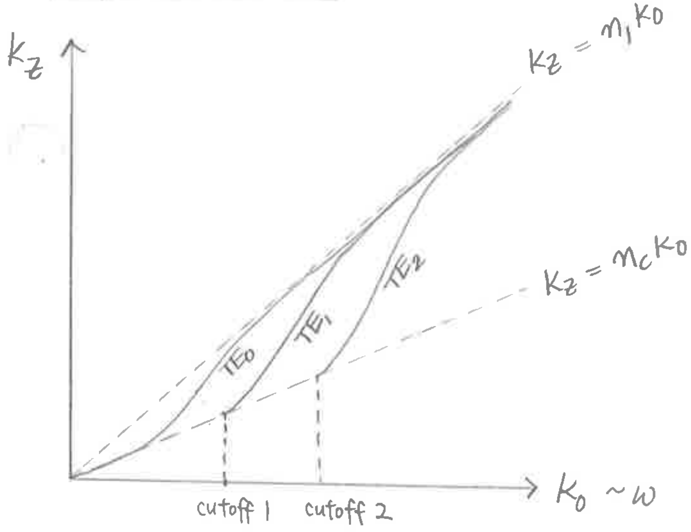
For $\lambda = 873\text{nm}$ (GaAs)
$n_1 \approx 3.59$ (GaAs)
$n_2 \approx 3.385$ ($Al_{0.3}Ga_{0.7}As$)
$\rightarrow d < 0.36\mu m$

- Solve for k_z

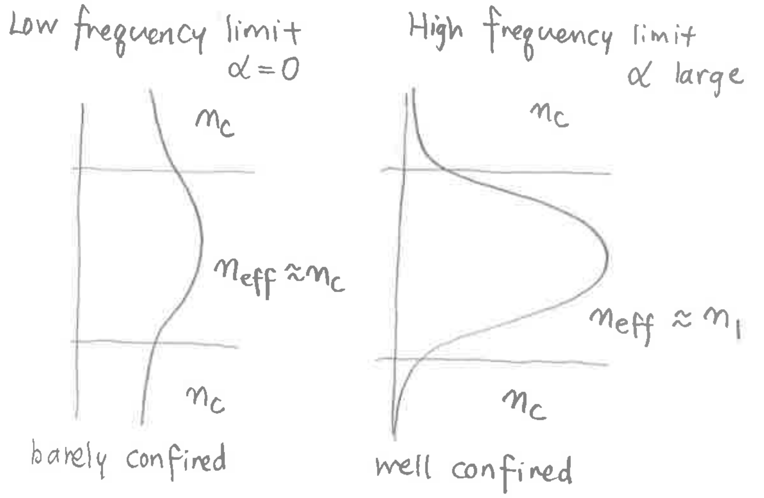
⑤ $\rightarrow k_z^2 = \omega^2\mu_0\varepsilon_1 - k_x^2$, ⑥ $\rightarrow k_z^2 = \omega^2\mu_0\varepsilon_c + \alpha^2$

- When a mode is at the cutoff of being excited, $\alpha = 0 \Rightarrow k_z = n_c k_0$
  In the limit of high frequency, k_x asymptotes, so $\omega^2\mu_0\varepsilon_1 >> k_x^2 \Rightarrow k_z = n_1 k_0$
  At intermediate frequencies, $n_c k_0 < k_z < n_1 k_0$
- At the mode's cutoff $\alpha = 0$ (not confined) and α increases with frequency

- Dispersion curves



Mode profile : (TE$_0$)



- The effective index of a mode is defined as $n_{eff} = \dfrac{k_z}{k_0}$
For the reasons stated above, every mode starts at n_c and ends at n_1 as a function of ω

- It is customary to define the following 3 parameters:

Normalized frequency $\quad V = k_0 d\sqrt{n_1^2 - n_2^2}$

Propagation parameter $\quad b = \dfrac{n_{eff}^2 - n_2^2}{n_1^2 - n_2^2} \longrightarrow$ 0 at low V ($\overset{\text{TE}_1}{\text{cutoff}} = \pi$), 1 at high V

Asymmetry parameter $\quad a = \dfrac{n_2^2 - n_3^2}{n_1^2 - n_2^2} \longrightarrow$ $a = 0$ for symmetric slab

Re-express single-mode condition : $V < \pi$

- Power normalization

$$P = \frac{1}{2}\,\mathrm{Re}\int_{-\infty}^{\infty} (\vec{E}\times\vec{H}^*)\cdot\hat{z}\,dx = 1 \;\Rightarrow\; -\frac{1}{2}\,\mathrm{Re}\int_{-\infty}^{\infty} E_y H_x^*\,dx = \underbrace{\frac{1}{2}\int_{-\infty}^{\infty}\frac{k_z}{\omega\mu}|E_y|^2\,dx = 1}$$

This last condition can be used to find the remaining unknown in specifying $E_y(x,z)$

- Optical confinement factor

$$\Gamma = \frac{\int_{-d/2}^{d/2} |\vec{E}(x)|^2\,dx}{\int_{-\infty}^{\infty} |\vec{E}(x)|^2\,dx}$$

* Γ is small at mode cutoff, and becomes larger with increasing k_0 or d

$\Rightarrow$ Therefore, for single mode operation, it is desirable to be just below the cutoff of the first excited mode (i.e. near the single mode condition)

- TM modes: to solve, use the duality principle and make the following replacements,

$$\begin{array}{ccc} \vec{E} & \rightarrow & \vec{H} \\ \vec{H} & \rightarrow & -\vec{E} \\ \text{TE} & & \text{TM} \end{array} \quad \text{and} \quad \begin{array}{c} \mu \rightarrow \varepsilon \\ \varepsilon \rightarrow \mu \end{array}$$

- The characteristic equations are

$$\alpha \frac{d}{2} = \frac{\varepsilon_c}{\varepsilon_1}\left(\frac{k_x d}{2}\right) \tan\left(\frac{k_x d}{2}\right) \leftarrow \text{even}$$

$$\alpha \frac{d}{2} = -\frac{\varepsilon}{\varepsilon_1}\left(\frac{k_x d}{2}\right) \cot\left(\frac{k_x d}{2}\right) \leftarrow \text{odd}$$

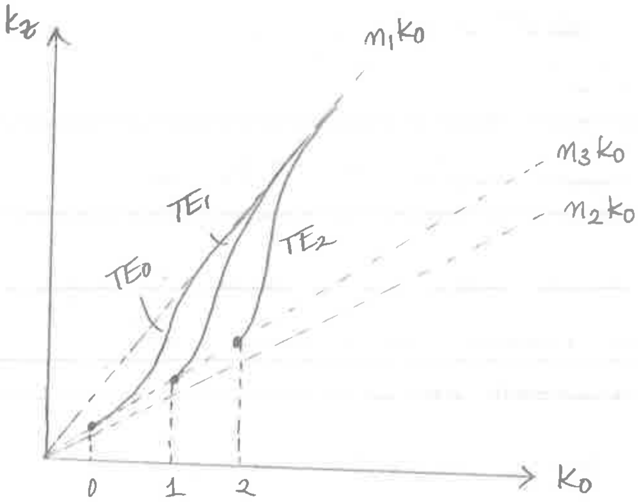
- Asymmetric waveguide : $n_2 > n_3$
  - The cutoff condition changes to

$$V = k_0 d \sqrt{n_1^2 - n_2^2} = \tan^{-1}\sqrt{a} + m\pi$$

where a is the asymmetry parameter

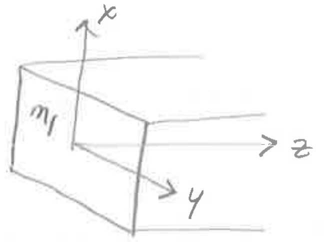
  - Notice that now even the TE_0 mode has a finite cutoff frequency! The single mode condition changes to:

$$\tan^{-1}\sqrt{a} \leq V \leq \tan^{-1}\sqrt{a} + \pi$$

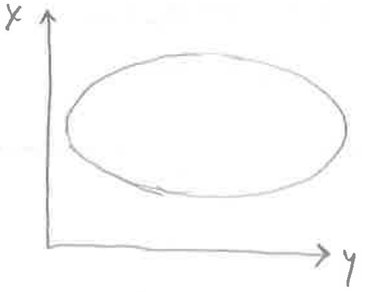


  - Now the low frequency limit of n_{eff} is the index of the higher-index cladding

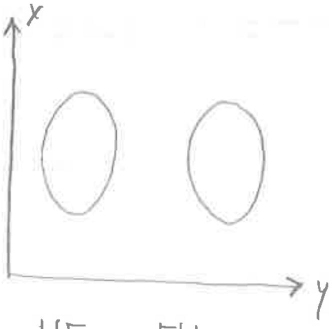
- Rectangular waveguide modes



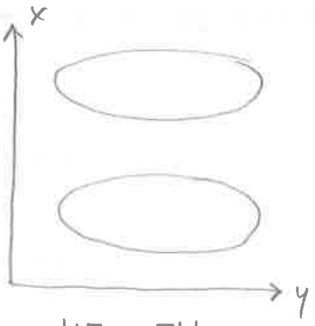
  - The modes are described as:
    - HE_{pq} : H_x and E_y dominant (like TE for infinite slab)
    - EH_{pq} : E_x and H_y dominant (like TM for infinite slab)
    - p, q denote the mode frequency along x and y : intensity contours shown below



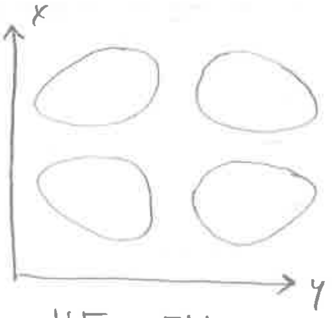
HE_{00}, EH_{00}



HE_{01}, EH_{01}



HE_{10}, EH_{10}



HE_{11}, EH_{11}

Effective index method for ridge waveguides

① Solve the asymmetric slab waveguide $n_3 - n_1 - n_5$, treating it as infinite along y. Find $n_{eff,1}$ for this waveguide

② If $n_2 > n_7, n_6$, use the same method to find $n_{2,eff}$. Otherwise, approximate weak fields in 6, 7 and let $n_{2,eff} = n_2$

③ Follow the same procedure as ② for the $n_8 - n_4 - n_8$ waveguide to find $n_{4,eff}$.

↓

Field profiles $F(x, y)$ where the field in y is stitched together from the 3 parts

④ Solve the infinite slab waveguide above to find the field profile $G(y)$ along y.

⑤ Combine the results to get the overall field

$$E(x,y) = F(x,y)\,G(y)$$

- **Gaussian beam**: the most spatially localized type of wave that can exist in free space

- Origin of the Gaussian beam: consider a paraxial wave $U(\vec{r}) = A(\vec{r})\,e^{-ikz}$ where A is an envelope that varies much less slowly than k. For U to satisfy the Helmholtz equation, A must satisfy:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) A - i2k\frac{\partial A}{\partial z} = 0$$

The Gaussian beam is a solution to this equation, and is given by

$$A(\vec{r}) = \frac{A_1}{q(z)}\, e^{-ik\left(\frac{x^2+y^2}{2q(z)}\right)}$$

Gaussian beam envelope

where $q(z) = z + iz_0$

also written as

$$\frac{1}{q(z)} = \frac{1}{R(z)} - i\,\frac{\lambda}{\pi W^2(z)}$$

The overall equation for the Gaussian beam is

wavefront curvature

$$\boxed{U(\vec{r}) = A_0 \frac{W_0}{W(z)} e^{-\frac{x^2+y^2}{W^2(z)}} e^{-ikz - ik\frac{x^2+y^2}{2R(z)} + i\zeta(z)}}$$

This is the excess delay of a Gaussian beam in comparison with a plane wave

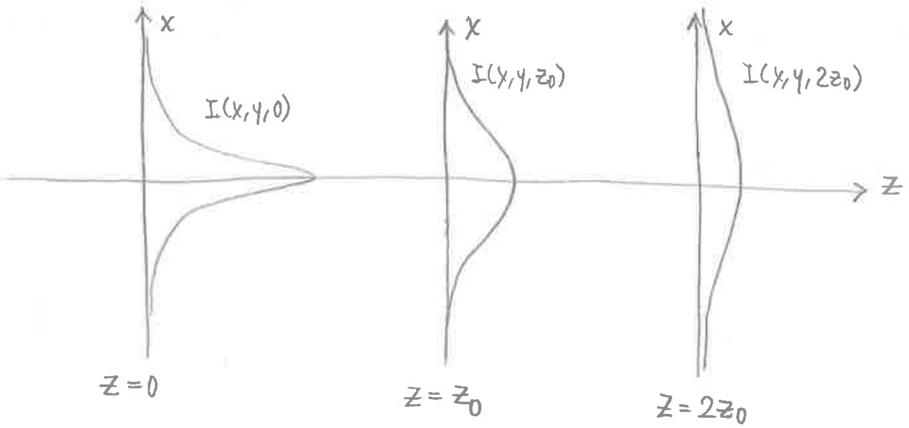
where
$$W(z) = W_0\left[1+\left(\frac{z}{z_0}\right)^2\right]^{1/2} \qquad W_0 = \sqrt{\frac{\lambda z_0}{\pi}}$$
$$R(z) = z\left[1+\left(\frac{z_0}{z}\right)^2\right]$$
$$\zeta(z) = \tan^{-1}\frac{z}{z_0}$$

A_0 and z_0 are determined by boundary conditions; everything can be found

- Properties of the Gaussian beam

1) Intensity

$$I(x,y,z) = |U(\vec{r})|^2 = I_0\left(\frac{W_0}{W(z)}\right)^2 e^{-2\frac{x^2+y^2}{W^2(z)}}$$



- The transverse field profile always falls off as a Gaussian of the radial distance.
- The width $W(z)$ of the beam increases with z
- The on-axis intensity is

$$I(0,0,z) = \frac{I_0}{1+(z/z_0)^2}$$

- At $z = z_0$, $I = \frac{1}{2}I_0$
- As $z \to \infty$, $I = I_0 z_0^2/z^2$ (inverse square law, like a spherical wave)

2) The power is

$$P = \int_0^\infty I(x,y,z)\,dx\,dy = \frac{1}{2}\underset{\substack{\uparrow \\ \text{peak} \\ \text{intensity}}}{I_0}\underset{\substack{\uparrow \\ \text{beam} \\ \text{area}}}{\pi W_0^2}$$

We can rewrite the intensity as

$$\boxed{I(x,y,z) = \frac{2P}{\pi W^2(z)} e^{-2\frac{x^2+y^2}{W^2(z)}}}$$

- About 86% of the power is within a circle of radius $\rho = \sqrt{x^2+y^2} = W(z)$
- About 99% is within $1.5W(z)$

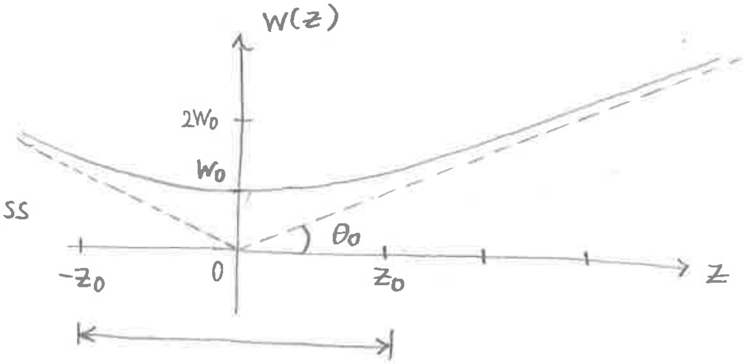
3) Beam radius: $W(z) = W_0\left[1+\left(\frac{z}{z_0}\right)^2\right]^{1/2}$

The field drops to e^{-2} at this radius (≈ 0.135)

- At $z=0$, W is at its minimum and is equal to W_0. This plane is called the waist and W_0 is the waist radius.
- As $z \to \infty$, $W(z) \approx \frac{W_0}{z_0} z = \theta_0 z$ where the beam divergence is:

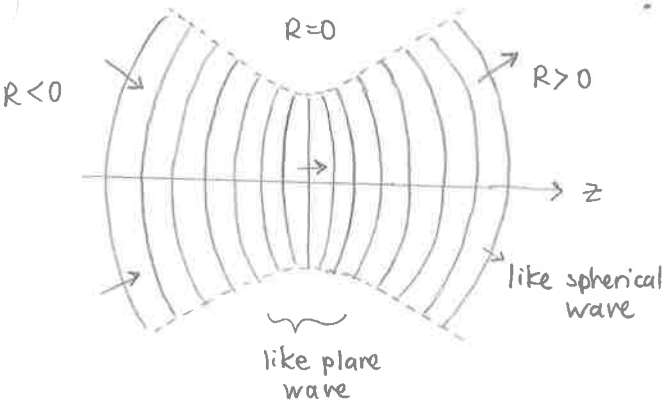
$$\boxed{\theta_0 = \frac{2}{\pi}\frac{\lambda}{2W_0}}$$

  - Shorter wavelengths diverge less
  - A wider beam diverges less
  - Beams with a greater depth of focus diverge less



Depth of focus $= 2z_0 = \frac{2\pi W_0^2}{\lambda}$

4) Wavefronts: due to the phase term $e^{-ik(x^2+y^2)/2R(z)}$ the wavefronts have some curvature

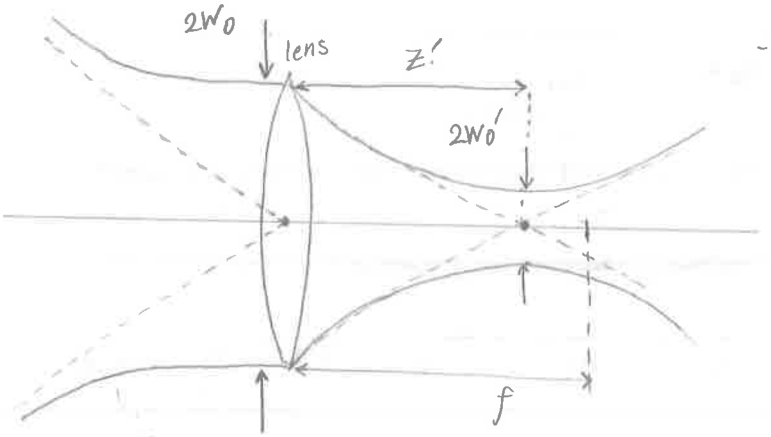


- $R(z)$ represents the radius of curvature of the wavefront at z
  - At $z=0$, $R \to \infty$ ⇒ planar wavefronts
  - As $z \to \infty$, $R \to z$ ⇒ spherical wavefronts

- The Gaussian beam needs to be specified by 4 parameters:
  1) Amplitude A_0
  2) Direction, or beam axis
  3) Location of waist ($z=0$)
  4) Waist radius W_0 or Rayleigh range z_0

- Manipulation of Gaussian beams by passive optics

- A thin lens retains the shape of the Gaussian, but magnifies its waist radius: $W_0' = MW_0$, $\theta_0' = \frac{1}{M}\theta_0$
- Focusing a Gaussian beam: what is the smallest possible spot size?



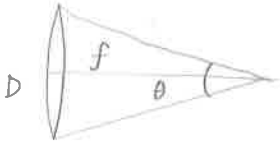
- Put a lens directly at the waist of a Gaussian. The new Gaussian has its waist at z' and has a new diameter $2W_0'$

$$W_0' = \frac{W_0}{\sqrt{1+(z_0/f)^2}}$$

$$z' = \frac{f}{1+(f/z_0)^2}$$

To make the smallest possible spot, use a beam with a large waist W_0 and therefore large depth of focus $2z_0$. Since this is approximately a plane wave, the lens focuses to a distance f. The waist diameter is:

$$2W_0' \approx \underset{\substack{\uparrow \\ \text{lens } f\text{-number}}}{\frac{4}{\pi}\lambda F_\#} = \frac{4}{\pi}\lambda\frac{f}{D} \approx 1.27\frac{\lambda}{2\sin\theta}$$



$\tan\theta \approx \sin\theta = \frac{D}{2f}$

$\frac{f}{D} \approx \frac{1}{2\sin\theta}$

This roughly corresponds to the diffraction limit,

- Passive optics always preserves the quantity $W_0 \cdot \theta_0$, which depends only on the wavelength λ. This is essentially a statement of étendue conservation.

- A Bessel beam is given by

$$U(\vec{r}) = A_m J_m(k_T\rho)\, e^{im\phi}\, e^{-ik_z z}$$

where $m = 0, \pm1, \pm2, \ldots$

$\rho = \sqrt{x^2+y^2}$, $x = \rho\cos\phi$, $y = \rho\sin\phi$

J_m = Bessel function of 1st kind, mth order

A Bessel beam has zero beam divergence
but infinite rms beam width

⑨ Dynamic response of semiconductor lasers

The rate equations express the rate of change of N = carrier density
and S = photon density

$$\boxed{\frac{dN}{dt} = \overset{\substack{\text{injection rate of}\\\text{carriers into the}\\\text{active region}}}{\frac{\eta_i I}{qV}} - \overset{\substack{\text{carrier loss}\\\text{rate by}\\\text{recombination}}}{\frac{N}{\tau(N)}} - \overset{\substack{\text{carrier loss rate}\\\text{by stimulated emission}}}{v_g\, g(N) S}}$$

$$\boxed{\frac{dS}{dt} = \underset{\substack{\text{stimulated emission}\\\text{rate of photons}}}{\Gamma v_g\, g(N) S} - \underset{\substack{\text{loss rate}\\\text{of photons}}}{\frac{S}{\tau_p}} + \underset{\substack{\text{spontaneous emission}\\\text{rate of photons into}\\\text{the lasing mode}}}{\Gamma \beta R_{sp}}}$$

η_i = IQE = $\dfrac{\#\text{ photons produced}}{\#\text{ carriers injected}}$

V = active region volume

Γ = optical confinement factor = $\dfrac{V}{V_p}$

where V_p = photon mode volume

R_{sp} = spontaneous emission rate/vol.

v_g = group velocity c/n_r

τ_p = photon lifetime

τ = carrier lifetime

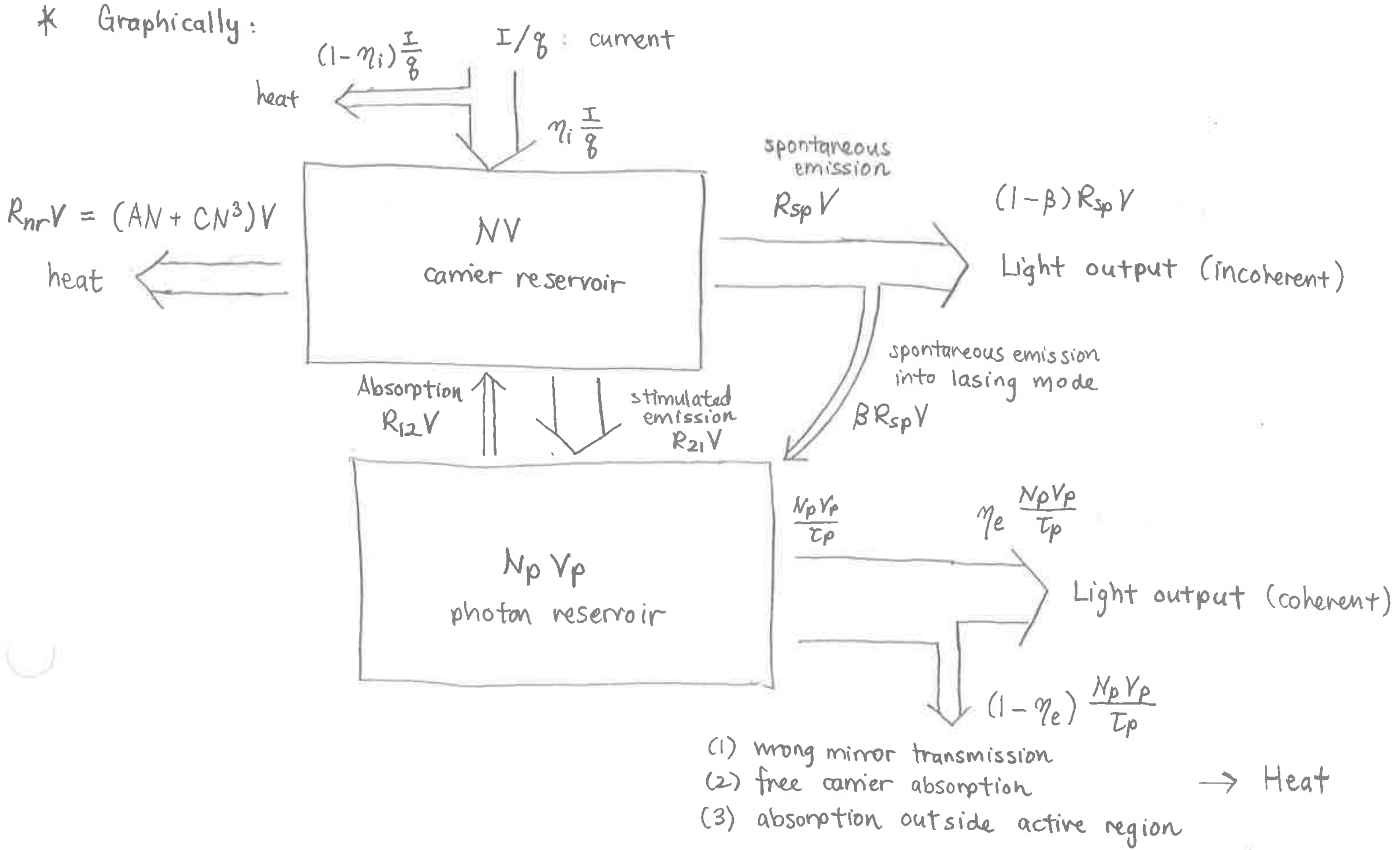
η_e = photon extraction efficiency

where $\dfrac{N}{\tau(N)} = AN + BN^2 + CN^3$, $R_{sp} = BN^2$

β = fraction of spontaneously emitted photons in the same direction as stimulated emission

$= \Omega/4\pi \approx 10^{-3} - 10^{-4}$

* Graphically:



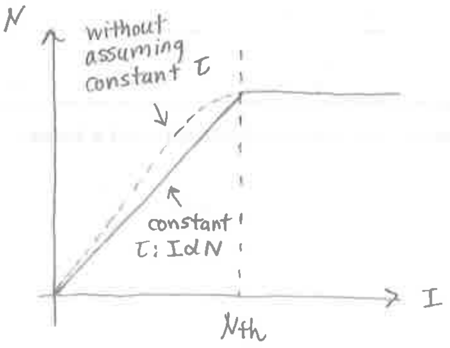
- Steady state solutions : $\frac{d}{dt} = 0$

1) Simplified analysis: let $R_{sp} \approx 0$

$$\frac{dS}{dt} = \left(\Gamma v_g\, g(N) - \frac{1}{\tau_p}\right) S = 0$$

$$\Rightarrow \quad g(N) = \frac{1}{\Gamma v_g \tau_p} = \frac{\alpha_{total}}{\Gamma} = g_{th} = g(N_{th})$$

- Due to photon loss, the gain will be clamped after the laser reaches threshold: otherwise, the field in the cavity increases without bound → no steady state!
- As a consequence, the carrier density is clamped at N_{th} above threshold, since the gain is not allowed to increase



- If the current is increased, N will exceed N_{th} briefly, but all of the excess carrier density will quickly be used up via increased stimulated emission, until a new steady state is reached

- Below threshold: $S \approx 0$

$$\frac{dN}{dt} = \frac{\eta_i I}{qV} - \frac{N}{\tau} = 0 \qquad \Rightarrow \qquad I = \frac{qV}{\eta_i \tau} N$$

- Above threshold, $N = N_{th}$, $g = g_{th}$

$$\frac{dN}{dt} = \frac{\eta_i I}{qV} - \frac{N_{th}}{\tau} - v_g g_{th} S = 0$$

$$\Rightarrow \quad S = \frac{1}{v_g g_{th}}\left(\frac{\eta_i I}{qV} - \frac{N_{th}}{\tau}\right) = \frac{1}{v_g g_{th}}\left(\frac{\eta_i I}{qV} - \frac{\eta_i I_{th}}{qV}\right)$$

$$S = \frac{1}{v_g g_{th}} \frac{\eta_i}{qV} (I - I_{th})$$

The power out is: then use $\frac{1}{\tau_p} = \Gamma g v_g$

$$P_{out} = V_p S \underbrace{\frac{\alpha_m}{\alpha_m + \alpha_i}}_{\eta_e} \frac{1}{\tau_p} \hbar\omega = \frac{SV}{\Gamma} \frac{\eta_e}{\tau_p} \hbar\omega = \frac{1}{v_g g_{th}} \frac{\eta_i}{q\cancel{V}} (I - I_{th}) \frac{\cancel{V}}{\Gamma} \frac{\eta_e}{\tau_p} \hbar\omega$$

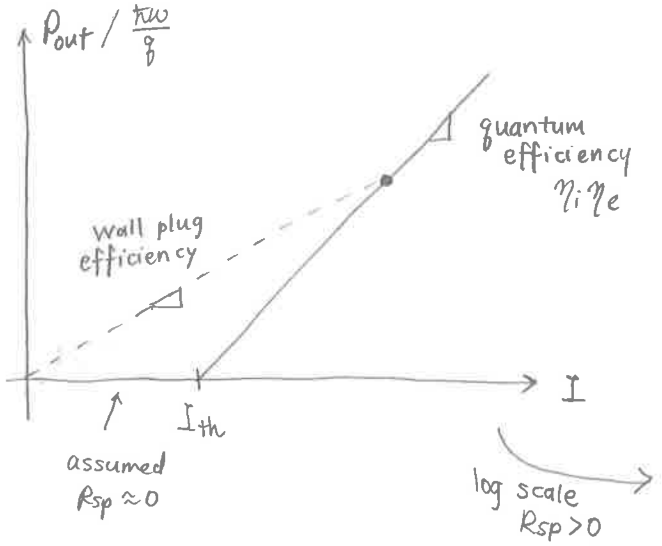
extraction efficiency

$$\Rightarrow \quad \boxed{P_{out} = \frac{\hbar\omega}{q} \eta_i \eta_e (I - I_{th})}$$

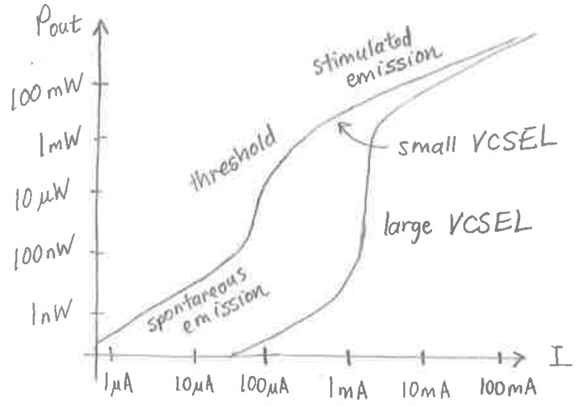
photon energy to charge conversion (W/A)

quantum efficiency

$$P_{out} = \frac{\hbar\omega}{q}\eta_i\eta_e(I - I_{th})$$



- The light output increases linearly with current above threshold under the approximation that $R_{sp} \approx 0$



- Detailed steady-state analysis : $R_{sp} \neq 0$

- Continue to assume τ is constant (this implies that the SRH term $= AN$ dominates the recombination rate)
- Apply the rate equations:

$$\frac{dS}{dt} = \Gamma v_g g(N) S - \frac{S}{\tau_p} + \Gamma\beta R_{sp} = 0$$

$$S(N)\left[\frac{1}{\tau_p} - \Gamma v_g g(N)\right] = \Gamma\beta R_{sp}$$

$$① \quad S(N) = \frac{\Gamma\beta R_{sp}(N)}{\frac{1}{\tau_p} - \Gamma v_g g(N)}$$

$$\frac{dN}{dt} = \frac{\eta_i I}{qV} - \frac{N}{\tau} - v_g g(N) S(N) = 0$$

$$② \quad I(N) = \frac{qV}{\eta_i}\left[\frac{N}{\tau} + v_g g(N) S(N)\right]$$

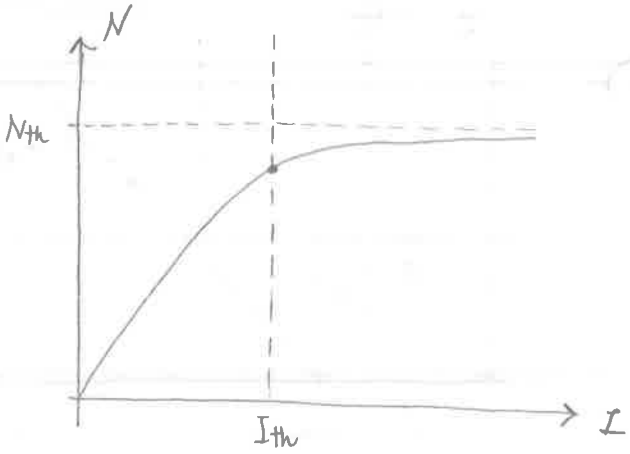
- Below threshold: $S \approx 0$, so from ② we again have $I(N) = \frac{qV}{\eta_i}\frac{N}{\tau(N)}$
- Above threshold: recall $\frac{1}{\tau_p} = \Gamma v_g g_{th}$

$$① \Rightarrow S(N) = \frac{\Gamma\beta R_{sp}(N)}{\Gamma v_g (g_{th} - g(N))} \Rightarrow \boxed{S(N) = \frac{\beta R_{sp}(N) / v_g}{g_{th} - g(N)}}$$

- If the photon density stays finite, the threshold gain cannot actually be reached!
- g and N stay clamped just slightly below g_{th} and N_{th}

$$S(N) = \frac{\beta R_{sp}(N)/v_g}{g_{th} - g(N)} \quad \leftarrow$$ the spontaneous emission βR_{sp} is clearly being amplified!

Use ② to solve for current:



$$I(N) = \underbrace{\frac{qV}{\eta_i}\frac{N}{\tau(N)}}_{\approx I_{th}} + \frac{qV}{\eta_i} v_g\, g(N)\, S(N)$$

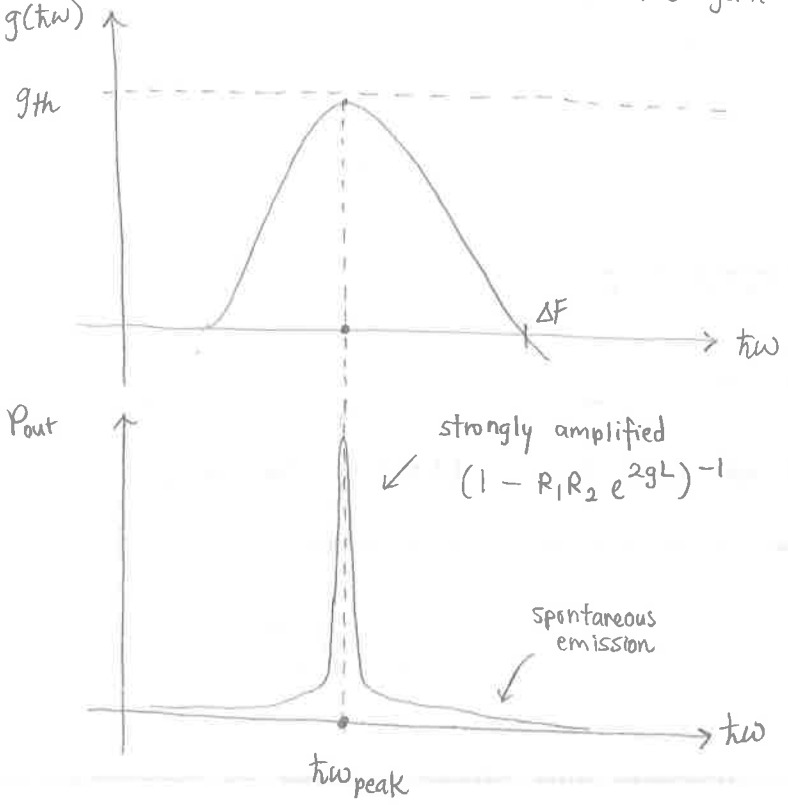
$$I(N) = I_{th} + \frac{qV}{\eta_i} v_g\, g(N) \frac{\beta R_{sp}(N)/v_g}{g_{th} - g(N)}$$

$$\frac{qV}{\eta_i} \beta R_{sp}(N) \frac{g(N)}{g_{th} - g(N)} = I(N) - I_{th}$$

$$\Rightarrow \quad g(N) = g_{th} \cdot \left(\frac{1}{\frac{qV\beta R_{sp}(N)/\eta_i}{I(N) - I_{th}} + 1} \right) < g_{th}$$

$$\Rightarrow \quad N < N_{th}$$

- This shows again that the gain never reaches g_{th} (and same with N) but asymptotically approaches it with increasing current.
- This means that the gain spectrum appears as below



Independent of the gain physics, g is clamped at g_{th} due to the loss rate of photons from the cavity

→ Therefore, even without considering the allowable longitudinal lasing modes, it is a good approximation to assume a narrow emission spectrum with

$$P_{out} = \frac{\hbar\omega}{q} \eta_i \eta_e (I - I_{th})$$

- Differential analysis

  - To analyze the modulation characteristics of semiconductor lasers, we need to linearize the rate equations around a steady state solution and treat the modulation as a small signal
  - Take the differential of the rate equations:

$$d\left(\frac{dN}{dt}\right) = d\left(\frac{\eta_i I}{qV} - \frac{N}{\tau(N)} - v_g\, g(N) S\right)$$

③ $$d\left(\frac{dN}{dt}\right) = \frac{\eta_i}{qV}\, dI - \frac{dN}{\tau_{\Delta N}} - v_g\, g\, dS - v_g\, S\, dg$$

④ $$d\left(\frac{dS}{dt}\right) = \Gamma v_g\, g\, dS + \Gamma v_g\, S\, dg - \frac{dS}{\tau_p} + \Gamma \frac{dN}{\tau'_{\Delta N}}$$

where the differential carrier lifetimes are found from

$$d\left(\frac{N}{\tau}\right) = dN \frac{d}{dN}\left(\frac{N}{\tau}\right) = dN \underbrace{\left(A + 2BN + 3CN^2\right)}_{= \frac{1}{\tau_{\Delta N}}}$$

likewise, $$\frac{1}{\tau'_{\Delta N}} = \frac{d}{dN}\left(\beta R_{sp}\right) = \frac{d}{dN}\left(\beta B N^2\right)$$

$$\Rightarrow \quad \frac{1}{\tau'_{\Delta N}} = 2\beta B N + \frac{d\beta}{dN} B N^2$$

- The differential gain dg is found by differentiating $g(N,S)$:

$$dg = \frac{\partial g}{dN}\, dN + \frac{\partial g}{dS}\, dS = a\, dN - a_p\, dS$$

where the differential gain values are defined as

$$a = \frac{\partial g}{dN}, \quad a_p = -\frac{\partial g}{dS}$$

- Although a and a_p are not constant with N and S, we can set them to their values under a steady state bias in a differential analysis. To know their actual dependences, we need a model for $g(N,S)$. From laser physics, we can approximate the gain as:

$$g(N,S) = \frac{g_0}{1+\varepsilon S} \ln\left(\frac{N}{N_{tr}}\right) \quad \Rightarrow \quad a = \frac{\partial g}{dN} = \frac{g_0}{1+\varepsilon S} \cdot \frac{1}{N}, \quad a_p = -\frac{\partial g}{\partial S} = \frac{\varepsilon}{1+\varepsilon S} \cdot g$$

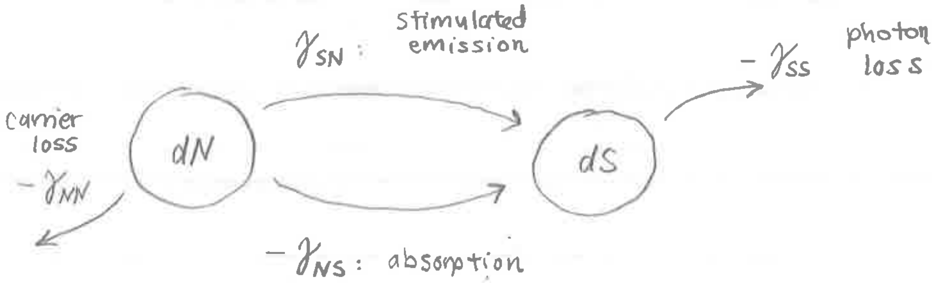
gain saturation

- If we cast ③ and ④ into matrix form and do some algebra, we can write them as:

$$\frac{d}{dt}\begin{bmatrix} dN \\ dS \end{bmatrix} = \begin{bmatrix} -\gamma_{NN} & -\gamma_{NS} \\ \gamma_{SN} & -\gamma_{SS} \end{bmatrix}\begin{bmatrix} dN \\ dS \end{bmatrix} + \frac{\eta_i}{qV}\begin{bmatrix} dI \\ 0 \end{bmatrix}$$

where
$$\gamma_{NN} = \frac{1}{\tau_{\Delta N}} + v_g a S$$
$$\gamma_{NS} = v_g g - a_p v_g S$$
$$\gamma_{SN} = \Gamma v_g a S$$
$$\gamma_{SS} = \Gamma v_g a_p S$$

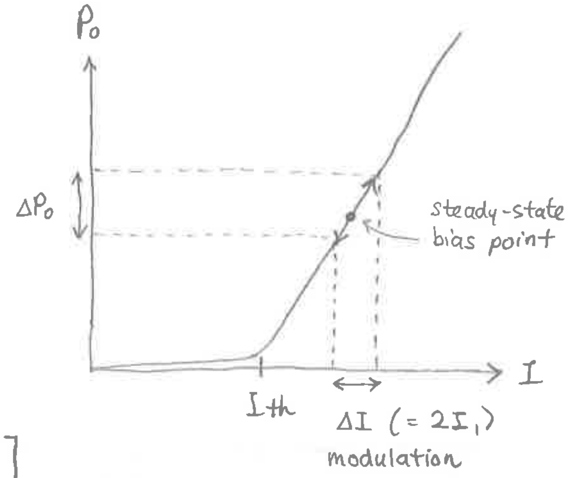
Interpretation:



where we can approximate the steady state gain as

$$g \approx \frac{1}{\Gamma v_g \tau_p}$$

- Small signal intensity modulation response



Let
$$dI(t) = I_1 e^{i\omega t}$$
$$dN(t) = N_1 e^{i\omega t}$$
$$dS(t) = S_1 e^{i\omega t}$$

so $\frac{d}{dt} \to i\omega$

The differential rate equations become

$$\begin{bmatrix} \gamma_{NN} + i\omega & \gamma_{NS} \\ -\gamma_{SN} & \gamma_{SS} + i\omega \end{bmatrix}\begin{bmatrix} N_1 \\ S_1 \end{bmatrix} = \frac{\eta_i I_1}{qV}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

We solve this by first determining the determinant of the matrix on the LHS, then applying Cramer's rule to solve for N_1, S_1.

determinant $\Delta = (\gamma_{NS}\gamma_{SN} + \gamma_{NN}\gamma_{SS}) - \omega^2 + i\omega(\gamma_{NN} + \gamma_{SS})$
$= \omega_R^2 - \omega^2 + i\omega\gamma$

where $\omega_R^2 = \gamma_{NS}\gamma_{SN} + \gamma_{NN}\gamma_{SS} \approx \frac{v_g a S}{\tau_p}$ ← relaxation resonance frequency

and $\gamma = \gamma_{NN} + \gamma_{SS} \approx K\omega_R^2 + \gamma_0$ ← damping

where $K = \tau_p\left[1 + \Gamma\frac{a_p}{a}\right]$

- The relaxation resonance

steady-state expression for S

$$\omega_R^2 = \frac{v_g a}{\tau_p} S = \frac{v_g a}{\tau_p}\left(\frac{1}{v_g g_{th}}\,\frac{\eta_i}{qV}(I - I_{th})\right) = \frac{\Gamma v_g a}{V}\,\frac{\eta_i}{q}(I - I_{th})$$

$$\Rightarrow \boxed{\omega_R^2 = \frac{\Gamma v_g a}{qV}\eta_i (I - I_{th})}$$

ω_R, the relaxation resonance, represents the natural resonance in the laser cavity when the photons interact with the carriers. Consider the response to a sudden change in input current:

- N increases : $\frac{\eta_i I}{qV}$, gain increases
- S then increases, due to increased gain : $\Gamma v_g g(N) S$
- N decreases, through increased stimulated emission : $-v_g g S$
- S decreases, due to decreased gain
- N increases, due to decreased stimulated emission

ω_R describes the natural frequency of this oscillation

- Frequency response

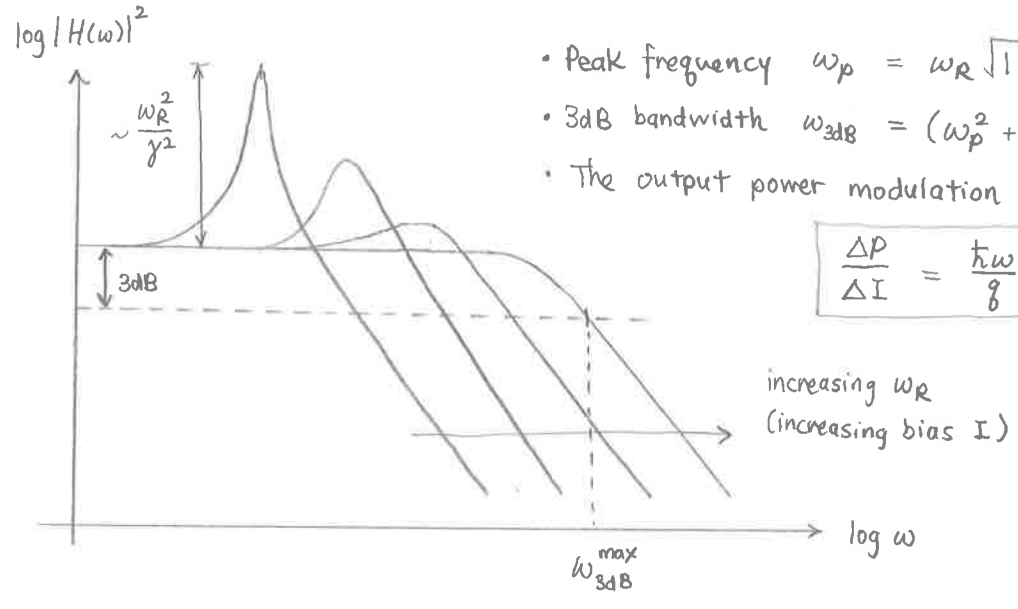
Now if we use Cramer's rule to solve the rate equations, we find:

$$N_1 = \frac{\eta_i I_1}{qV}\,\frac{\gamma_{SS} + i\omega}{\omega_R^2}\,H(\omega)$$

$$S_1 = \frac{\eta_i I_1}{qV}\,\frac{\gamma_{SN}}{\omega_R^2}\,H(\omega)$$

where the modulation transfer function is given by:

$$\boxed{H(\omega) = \frac{\omega_R^2}{\omega_R^2 - \omega^2 + i\omega\gamma}}$$



- Peak frequency $\omega_p = \omega_R\sqrt{1 - \frac{1}{2}(\gamma/\omega_R)^2} < \omega_R$
- 3dB bandwidth $\omega_{3dB} = \left(\omega_p^2 + \sqrt{\omega_p^4 + \omega_R^4}\right)^{1/2} > \omega_R$
- The output power modulation is:

$$\boxed{\frac{\Delta P}{\Delta I} = \frac{\hbar\omega}{q}\eta_i\eta_e H(\omega)}$$

- Some notes on frequency response

- Beyond the resonance peak, which occurs at $\omega_p \approx \omega_R$, the response falls off rapidly, at 40 dB/decade

- The damping $\gamma \approx K\omega_R^2 + \gamma_0$, increases with the bandwidth ω_R
  - At low power (small ω_R), the damping is small and the resonance peak is strong; at high power the resonance flattens out due to damping

- The modulation bandwidth goes with the relaxation resonance, given by:

$$\omega_R^2 = \frac{\Gamma v_g a}{qV} \eta_i (I - I_{th})$$

  - The modulation bandwidth increases as we drive the laser with more power $I > I_{th}$
  - The bandwidth benefits directly from:
    1) Large differential gain $a = \frac{dg}{dN}$ ⇒ use a strained QW laser!
    2) Large mode confinement factor Γ ⇒ use multiple QWs
    3) Small active region volume V.
    4) Injection quantum efficiency $\eta_i \rightarrow 100\%$
  - The limit to ω_R is ultimately set by $a = \frac{dg}{dN} = \frac{a_0}{1 + \varepsilon S}$

* As the bias I is increased, ω_R increases until the photon density S is $\sim 1/\varepsilon$. At that point a begins to decrease quickly as a result of gain compression

- However, in practice such high photon densities are not usually reached. The limit is then set by the damping γ, which increases as ω_R increases. If γ increases too much, ω_{3dB} will fall below ω_R. The maximum possible 3dB bandwidth occurs at an optimal damping, and is found to be:

high power limit →
$$\omega_{3dB}^{max} = \frac{\sqrt{2}}{2\pi K} = \boxed{\frac{\sqrt{2}}{2\pi} \frac{1}{1 + \Gamma a_p / a} \times \frac{1}{\tau_p} = \omega_{3dB}^{max}}$$

This value represents a limit to the intrinsic modulation capabilities of the laser. It is not surprising that its ultimate limit is $\sim 1/\tau_p$

- Thermal management, mirror damage, etc. are also potential risks associated with high power operation

- Transient intensity modulation response
  - In response to a sharp step increase in input current, the system response depends on the amount of damping
    - If underdamped ($\gamma \ll \omega_R$), the carrier and photon densities will oscillate at the relaxation resonance frequency ω_R
      - The photon density (and power output) will oscillate and settle to the new, higher value
      - The carrier density will oscillate and settle back to N_{th}
    - If the damping is large ($\gamma \sim \omega_R$), the carrier and photon densities will exponentially rise/fall to their steady-state values
      - This regime is not useful in practice, since ω_{3dB}^{max} is at $\frac{1}{\sqrt{2}}$

- Frequency modulation / chirping

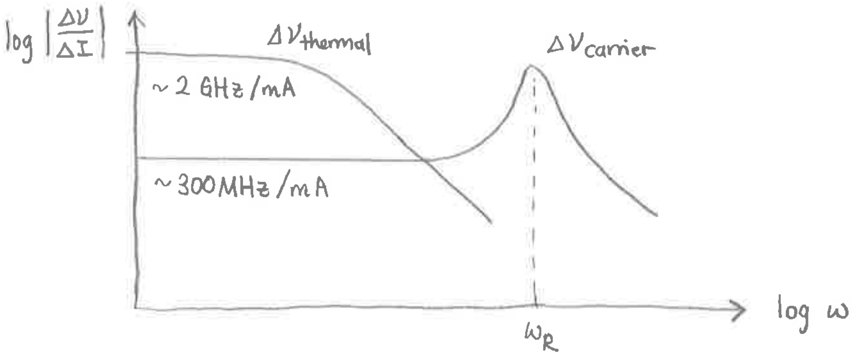
As the current I is modulated,
the carrier density N is modulated, and thus
the active region refractive index n_a is modulated, and
the cavity length length is modulated, and finally
the resonant mode frequency is modulated.

- We can write this relationship as: $\Delta\nu = -\frac{\Gamma v_g}{\lambda}\frac{dn_a}{dN}\Delta N$

  We then find the modulation to be given by:

$$\frac{\Delta\nu}{\Delta I} = \frac{\alpha}{4\pi}\Gamma v_g a \frac{\eta_i}{qV} \cdot \frac{\gamma_{ss} + i\omega}{\omega_R^2} H(\omega) \qquad \text{where} \quad \alpha \equiv -\frac{4\pi}{\lambda a}\frac{dn}{dN}$$

- Additionally, there is a further frequency modulation arising from the temperature modulation of the laser $\Rightarrow$ $\Delta\nu_{thermal}$. This tends to dominate at low frequency
- The two effects are illustrated below:



These are important effects to keep in mind when designing a directly modulated laser system!

- Noise in lasers:

  - Relative intensity noise: time variations in the photon and carrier densities due to random recombination and generation events in the absence of external modulation

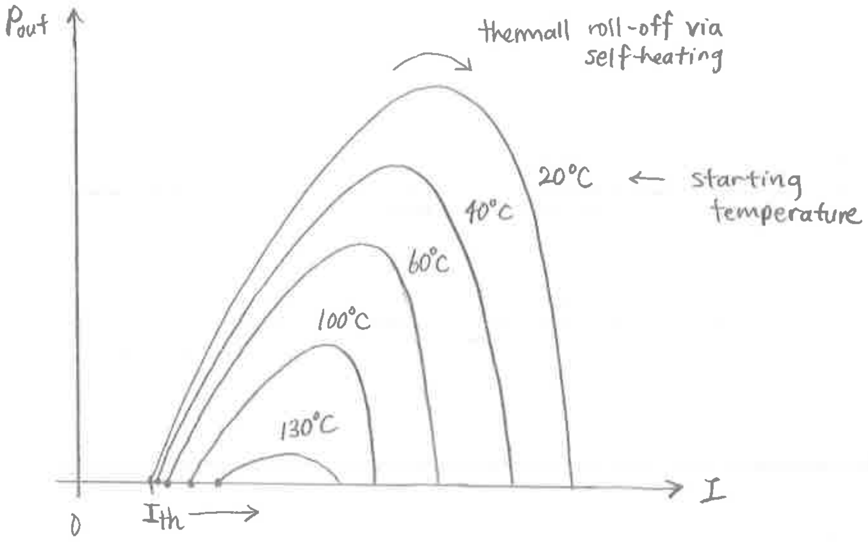
    ⇒ Noise floor on the output power magnitude

  - Frequency noise: random fluctuations in output frequency due to
    (1) spontaneous emission
    (2) carrier density fluctuations, through the frequency chirping effect

- Temperature-dependent effects

  - Any input power that does not leave the laser as light output is dissipated in the laser: $P_D = P_{in} - P_{out}$. The resulting temperature rise of the laser is $P_D Z_T$ where Z_T is the thermal impedance

  - The laser's L-I characteristic changes: $N_{tr} \propto T$ and $I_{th} \propto e^{T/T_0}$ (approximately)
    - Above threshold, η_i decreases and α_i increases with T, to give a lower quantum efficiency
    - Eventually the output power decreases with current due to heating effects
      - Auger recombination
      - Electron leakage from active region by thermal excitation

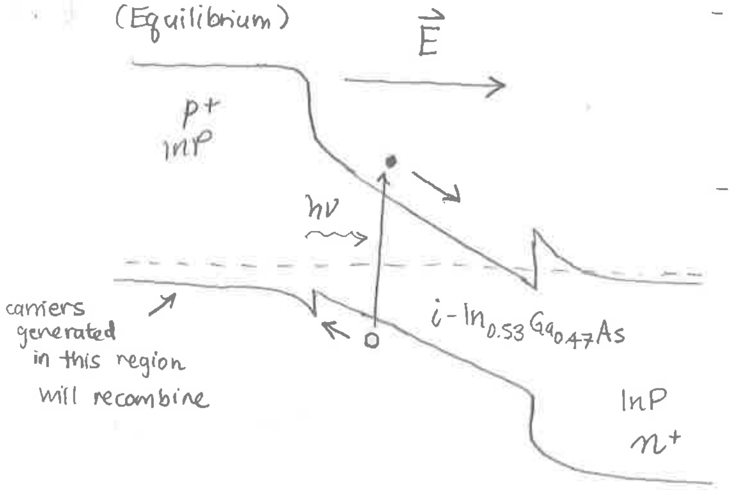


  - Wavelength matching: a big issue in VCSELs
    (1) the gain peak shifts with temperature
    (2) the wavelength of the cavity mode shifts with temperature, due to thermal expansion, etc.

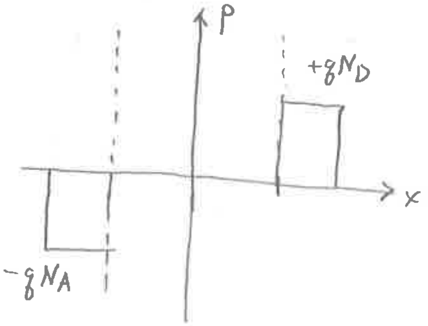
    ⇒ the wavelength mismatch lowers the gain of the cavity mode, so I_{th} increases away from the optimal temperature (usually 300 K)

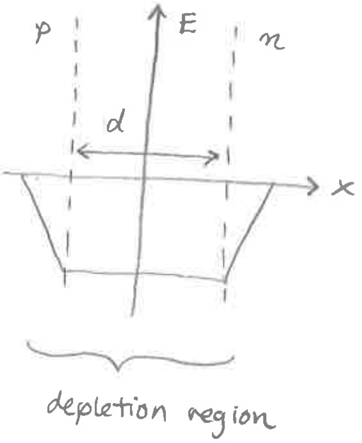
⑩ Photodetectors and photoconductors

- p-i-n photodiode



- The light-absorbing region is an intrinsic semiconductor sandwiched between heavily doped, preferably larger bandgap p- and n- regions
- The i-region becomes almost fully depleted of carriers and has a constant electric field:





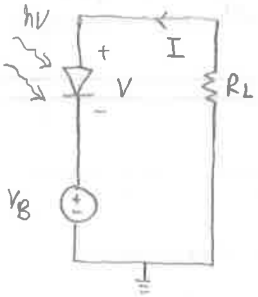
- The photodiode typically is reverse-biased so that a high electric field exists in the i-region, which helps sweep the carriers to their respective contacts.

- Wavelength sensitivity: by making the i-region width $d \sim 1/\alpha(\lambda)$, the p-i-n will absorb most of the incident light at the wavelength λ
  - This is only practical for λ shorter than the bandgap
  - Photodiodes of different thicknesses can be stacked so that each layer is sensitive to a different wavelength (shortest λ goes on top)

- Speed: $f_{max} \approx \dfrac{1}{\text{carrier transit time}} = \dfrac{1}{d/v}$ where v = carrier velocity
  - A thinner absorption region provides a larger f_{max}, but less optical absorption ⇒ a fundamental trade-off between speed and signal strength!
  - A thinner absorption region also increases the diode capacitance $C = \dfrac{\varepsilon A}{W_I}$ This will end up limiting the speed, but there may be structures that are able to circumevent this

- The p-i-n provides no gain, but has low noise.

- Photodiode I-V curve



- The photodiode current in the dark is:

$$I = I_0(e^{qV/kT} - 1)$$

which is that of a normal diode

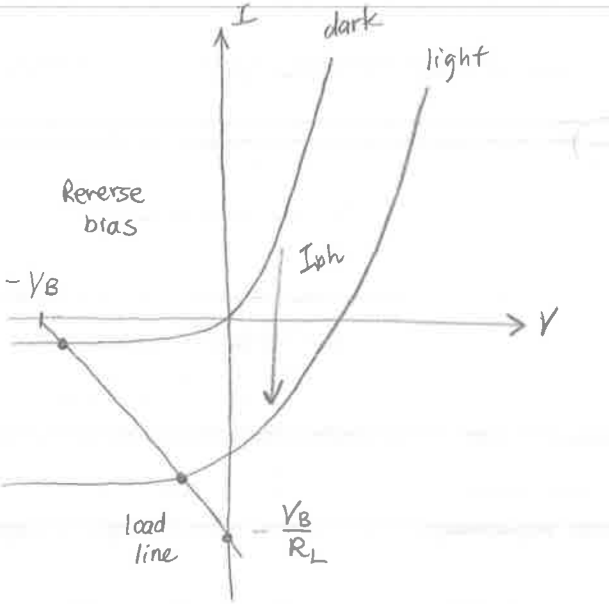
- In the presence of optical absorption, this is modified to

$$I = I_0(e^{qV/kT} - 1) + I_{ph}$$

where

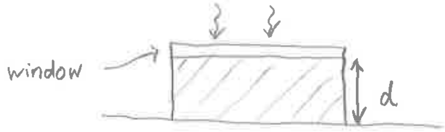
$$I_{ph} = -\eta \frac{q}{\hbar\omega} P_{opt}$$

photocurrent is a reverse current (−); quantum efficiency $\eta \leq 1$ (η); incident optical power (P_{opt})



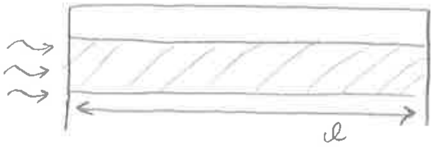
- 3rd quadrant: "photoconductive" mode
- 4th quadrant: "photovoltaic" mode

- The quantum efficiency depends on structure
  - Surface-illuminated p-i-n
  - Waveguide p-i-n



$$\eta = \eta_i(1-R)(1-e^{-\alpha d})$$

IQE (η_i); window reflectivity (R); absorption fraction ($1-e^{-\alpha d}$)

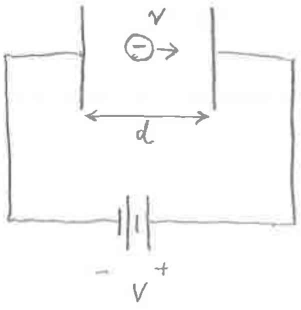


$$\eta = \eta_i(1-R)(1-e^{-\Gamma\alpha l})$$

facet reflectivity (R); confinement factor (Γ)

- Paradox: a photon generates an electron and a hole, does it produce 2q of charge?

- First, introduce Ramo's theorem



- The current produced in the external circuit by a charge moving with velocity v(t) between the plates is:

$$\boxed{i(t) = \frac{qv(t)}{d}}$$

- Proof: work done by the field to move the charge dx is

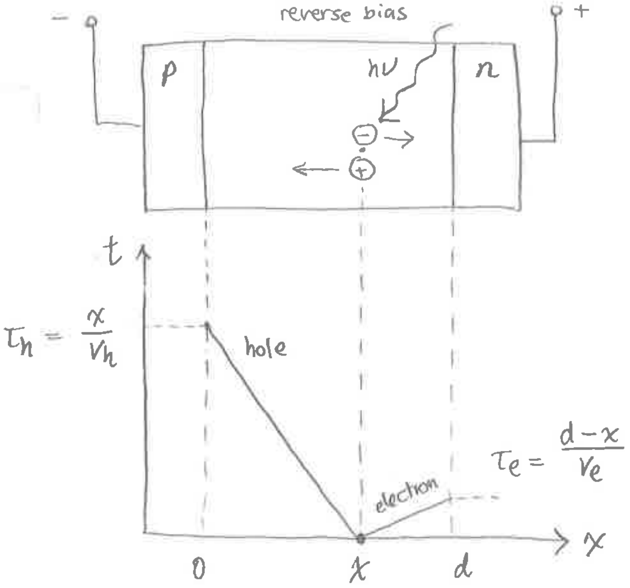
$$dW = qE\,dx = q\frac{V}{d}\,dx$$

- Equate this to the work done by the power supply

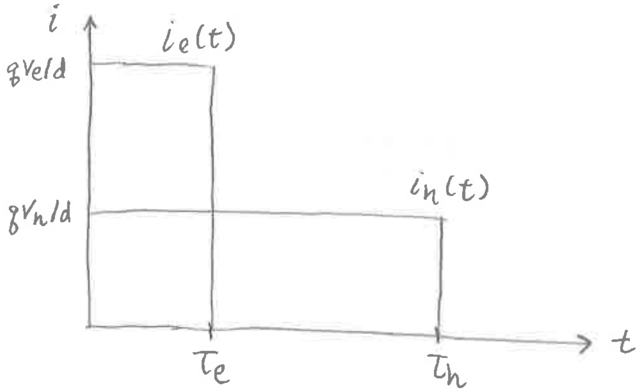
$$dW = \frac{qV}{d}\,dx = i(t)V\,dt \Rightarrow i(t) = \frac{q}{d}\frac{dx}{dt} = \frac{q}{d}v(t)$$

- The current stops when $v(t) \to 0$; when e^- reaches contact

- Find the response of a photo-generated electron-hole pair



- Find how long it takes each carrier to reach the contact
- Use Ramo's theorem to find the photocurrent over time



- Integrate the current to find the total collected charge

$$Q = \int i_e(t)\,dt + \int i_h(t)\,dt$$

$$= \frac{qv_e}{d}\tau_e + \frac{qv_h}{d}\tau_h = \frac{qv_e}{d}\frac{d-x}{v_e} + \frac{qv_h}{d}\frac{x}{v_h}$$

$\boxed{Q = q}$ → 1 charge is detected!

- Now suppose that a pulse of many photons produces electron-hole pairs through the volume of the absorption region. The resulting current $i(t)$ goes to 0 once the last carrier reaches its contact
  - Since in the worst case the carrier must travel the whole distance d and holes are slower, the transit time $\tau_{tr} \approx d/v_h$

- The response time is limited both by the transit time and the RC time

$$\tau = \tau_{tr} + \tau_{RC} = \frac{d}{v_h} + R\frac{\varepsilon A}{d}$$

Optimize: $\frac{\partial \tau}{\partial d} = \frac{1}{v_h} - \frac{R\varepsilon A}{d^2} = 0 \Rightarrow d = \sqrt{R\varepsilon A v_h}$

$$\tau_{opt} = 2\sqrt{\frac{R\varepsilon A}{v_h}} \Rightarrow f_{3dB}^{opt} \approx \frac{1}{2\pi\tau_{opt}}$$

$$f_{3dB}^{opt} = \frac{1}{4\pi}\sqrt{\frac{v_h}{R\varepsilon A}}$$

Typical values

$\tau_{RC} \approx 14ps$, $\tau_{tr} \approx 20ps$

$\Rightarrow f_{3dB} = 9.7\,GHz$

- Bandwidth - efficiency product
  - Surface-illuminated p-i-n: assume $R = 0\%$, assume transit time limited

$$\eta \times f_{3dB} = \eta_i (1 - e^{-\alpha d}) \frac{1}{2\pi} \frac{v_h}{d} \approx \eta_i (1 - (1 - \alpha d)) \frac{1}{2\pi} \frac{v_h}{d}$$

weakly absorbing limit

$$\Rightarrow \eta \times f_{3dB} = \frac{\eta_i \alpha v_h}{2\pi}$$

The product does not depend on dimensions: a trade-off between efficiency and speed exists! (or rephrased: signal strength vs. speed)

  - Waveguide p-i-n: assume $R = 0\%$, assume RC-limited

$$\eta \times f_{3dB} = \eta_i (1 - e^{-\Gamma \alpha \ell}) \frac{1}{2\pi RC} \quad , \quad C = \frac{\varepsilon \ell w}{d}$$

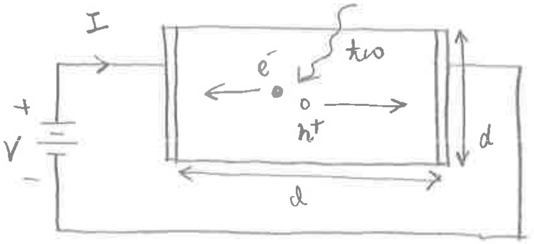
$$\approx \eta_i \Gamma \alpha \ell \frac{1}{2\pi R} \frac{d}{\varepsilon \ell w}$$

$$\eta \times f_{3dB} = \frac{\eta_i \Gamma \alpha d}{2\pi R \varepsilon w} \qquad (\text{independent of } \ell)$$

* In either type of detector, there is a tradeoff between efficiency and bandwidth (through d or ℓ)

- Photoconductors

- A photoconductor's resistivity changes with the absorption of incident light



- In the dark,

$$J_0 = \sigma_0 E = (q \mu_n n_0 + q \mu_p p_0) E$$

Under illumination,

$$J = [q \mu_n (n_0 + \delta n) + q \mu_p (p_0 + \delta p)] E$$

$$\Delta J = J - J_0 = (q \mu_n \delta n + q \mu_p \delta p) E$$

- Since $\delta n = \delta p$,

$$\Delta J = q (\mu_n + \mu_p) \delta n E$$

$$\text{or } \boxed{\Delta \sigma = q (\mu_n + \mu_p) \delta n}$$

- Both contacts on the photoconductor must be Ohmic
- The optically injected carrier density is given by the steady-state condition:

$$\frac{\partial}{\partial t} (\delta n) = \underset{\substack{\uparrow \\ \text{photogeneration} \\ \text{rate}}}{G_0} - \underset{\substack{\uparrow \\ \text{carrier} \\ \text{lifetime}}}{\frac{\delta n}{\tau_n}} = 0 \quad \Rightarrow \quad \delta n = G_0 \tau_n$$

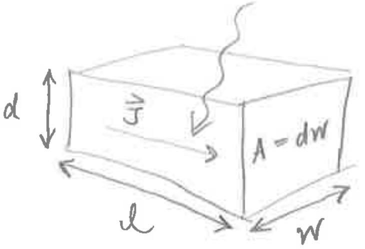
- The photogeneration rate is $G_0 = \underbrace{\frac{P_{opt}}{\hbar\omega}}_{\text{\# photons incident per second (assuming monochromatic)}} \cdot \frac{1}{lwd} \cdot \eta$

volume

quantum efficiency same as in photodiode

$\eta = \eta_i(1-R)(1-e^{-\alpha d})$

$\Rightarrow \delta n = G_0 \tau_n = \eta \frac{P_{opt}}{\hbar\omega}\frac{1}{lwd}\tau_n$



- The generated current is:

$$\Delta I = dw\Delta J = lwq(\mu_n + \mu_p)\delta n E$$

Assume $\mu_n \gg \mu_p$:

$$\Delta I = dwq\mu_n E \cdot \eta \frac{P_{opt}}{\hbar\omega}\frac{1}{lwd}\tau_n = q v_n \eta \frac{P_{opt}}{\hbar\omega}\frac{1}{l}\tau_n$$

Recall the transit time expression: $\tau_{tr} = l / v_n$

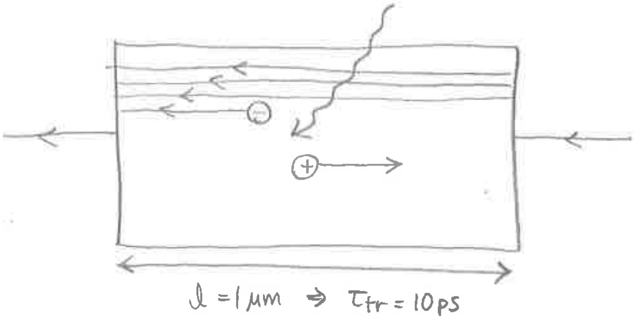
$$\Rightarrow \boxed{\Delta I = \underbrace{\left(\eta P_{opt}\frac{q}{\hbar\omega}\right)}_{\text{primary photocurrent } I_{ph}} \cdot \frac{\tau_n}{\tau_{tr}}}$$

photoconductive gain

\# absorbed photons × q

- The photoconductive gain $\frac{\Delta I}{I_{ph}} = \frac{\tau_n}{\tau_{tr}}$ ← recombination lifetime ← transit time

Interpretation: if many electrons travel across the electrodes before recombining with a hole, there is net gain.



$l = 1\,\mu m \Rightarrow \tau_{tr} = 10ps$

- The generated electron reaches the electrode before the hole
- Current continuity forces the external circuit to supply anther electron, which also crosses the device before the hole reaches its electrode
- This continues until the electron recombines with the hole

$\Rightarrow$ Many electrons pass through the circuit for every photon absorbed, leading to gain

- The mechanism is the same as that in a bipolar transistor:

$$\beta = \frac{i_c}{i_b} = \frac{\tau_t}{\tau_{rb}}$$

- In III-V, gain $\approx \frac{1\,ns}{10\,ps} = 100$
- In Si, gain $\approx \frac{1\,\mu s}{10\,ps} = 10^6$

- The photoconductor's responsivity is

$$R_\lambda = \frac{\Delta I}{P_{opt}} = \eta \frac{q}{\hbar\omega} \cdot \frac{\tau_n}{\tau_{tr}}$$ which has units of $\frac{A}{W}$

- Photoconductor frequency response (small signal)

Consider again the equation

$$\frac{d}{dt}(\delta n) = G_0 - \frac{\delta n}{\tau_n}$$

If $\delta n = N_1 e^{i\omega t}$, $\delta P = P_1 e^{i\omega t}$, this becomes

$$i\omega N_1 = \eta \frac{P_1}{\hbar\omega} \frac{1}{lwd} - \frac{N_1}{\tau_n}$$

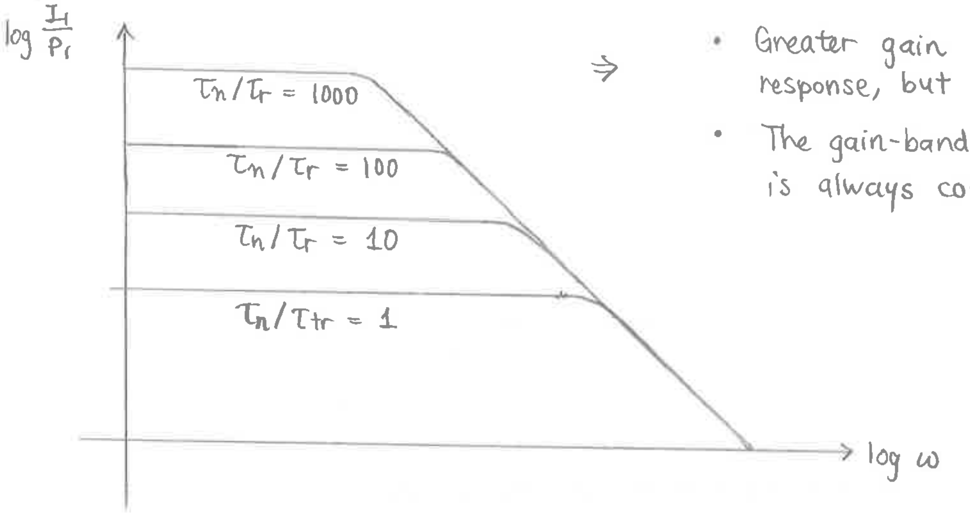
$$\left(i\omega + \frac{1}{\tau_n}\right) N_1 = \eta \frac{P_1}{\hbar\omega} \frac{1}{lwd}$$

$$N_1 = \eta \frac{P_1}{\hbar\omega} \frac{1}{lwd} \frac{1}{i\omega + 1/\tau_n}$$

Recall $I_1 = dw J_1 = dw\, q v_e N_1$

$$I_1 = P_1 \eta \frac{q}{\hbar\omega} \frac{v_e}{l} \frac{1}{i\omega + 1/\tau_n}$$

$$\Rightarrow \frac{I_1}{P_1} = \frac{\eta q}{\hbar\omega}\left(\frac{\tau_n}{\tau_{tr}}\right)\frac{1}{1 + i\omega\tau_n}$$



⇒
- Greater gain increases the low frequency response, but reduces the 3dB bandwidth
- The gain-bandwidth product of the device is always conserved

- Photoconductor structure
  - A conventional photoconductor has an n-i-n structure, in which electrons can be injected from both sides
  - To increase τ_n, n-i-p-i superlattice structures are sometimes used to create spatially separated potential wells for electrons and holes
  - To further improve τ_n, heterojunction superlattices with $\Delta E_c \approx 0$ but large ΔE_V can be used to create potential wells for holes but allow electrons to flow freely

Noise in photodetection

- Noise can arise due to:
  1) Random arrival of photons
  2) Randomness in the generation of photoelectron-holes
  3) Randomness in the gain process (for avalanche photodiodes)
  4) Circuit noise: shot noise and thermal noise

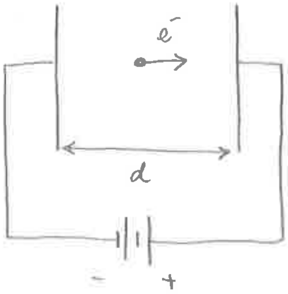
- The number of photons arriving on a detector follows a Poisson distribution (i.e. the probability that an event occurs within a given time interval is distributed uniformly over that interval)
  - If $\bar{n}$ = # average photons arriving within an interval T, the probability that exactly n photons arrive within that interval is

$$p(n) = \frac{\bar{n}^n \cdot e^{-\bar{n}}}{n!}$$

Of course, the mean = $\bar{n}$
also, the variance $\sigma^2 = \bar{n}$

$$\Rightarrow \quad SNR = \frac{\bar{n}^2}{\sigma^2} = \bar{n}$$

- Shot noise: the random arrival times of photons upon the detector translantes into random flow of charge, which produces a noisy current signal. Consistent with the properties above, the noise current is derived as follows:



- We model the generation of charge in a photodetector as a current pulse associated with a moving charge between two capacitor plates. From Ramo's theorem,

$$i_e(t) = \frac{q}{d} v(t) \qquad \text{for} \quad 0 \le t \le \tau_t$$

- Now add together a sequence of charges generated at random times:

$$i_T(t) = \sum_{i=1}^{N_T} i_e(t - t_i) \qquad \text{for} \quad 0 \le t \le T$$

N_T = # e^- within interval T

Take the Fourier transform:

$$i_T(f) = \sum_i^{N_T} \underset{\substack{\uparrow \\ \text{F.T. of } i_e(t)}}{i_e(f)} \; \underset{\substack{\uparrow \\ \text{shift property of F.T.}}}{e^{-i2\pi f t_i}}$$

Noise power is proportional to $|i|^2$:

$$i_T(f)\, i_T^*(f) = |i_e(f)|^2 \cdot \sum_i^{N_T} \sum_j^{N_T} e^{-i2\pi f(t_i - t_j)}$$

$$|i_T(f)|^2 = |i_e(f)|^2 \cdot \left[\underset{\substack{\uparrow \\ i=j \text{ terms}}}{N_T} + \sum_{i\neq j}^{N_T} \sum_j^{N_T} e^{-i2\pi f(t_i - t_j)} \right]$$

$$|i_T(f)|^2 = |i_e(f)|^2 \left[N_T + \sum_{i\neq j}\sum_j e^{-i2\pi f(t_i - t_j)} \right]$$

- Now take the ensemble average: since the times t_i are randomly correlated, all the terms $e^{-i2\pi f(t_i - t_j)}$ with $i \neq j$ average to zero

$$\Rightarrow \langle |i_T(f)|^2 \rangle = |i_e(f)|^2 N_T = |i_e(f)|^2 \, T \cdot \bar{N}$$

average # electrons / time

- Then find the FT $i_e(f)$:

$$\begin{aligned} i_e(f) &= \int_0^{\tau_t} i_e(t)\, e^{i2\pi f t}\, dt \\ &\approx \frac{q}{d}\int_0^{\tau_t} \frac{dx}{dt}\, dt = \frac{q}{d}\int_0^{d} dx = q \end{aligned}$$

Assumed low frequency $f \ll \frac{1}{\tau_t}$

$$\Rightarrow \langle |i_T(f)|^2 \rangle = q^2 T \bar{N}$$

- The final step is to take out the dependence on the interval T by finding the spectral density of the noise:

$$\begin{aligned} P_T &= \frac{1}{T}\int_{-T/2}^{T/2} \langle | i_T^2(t) | \rangle dt = \frac{1}{T}\int_{-\infty}^{\infty} \langle | i_T(f)|^2 \rangle df = \frac{2}{T}\int_0^{\infty} \langle | i_T(f)|^2 \rangle df \\ &= \int_0^{\infty} S(f)\, df \end{aligned}$$

where $S(f) = \frac{2}{T}\langle |i_T(f)|^2 \rangle = 2q^2 \bar{N} = 2q\,(q\bar{N}) = 2q\bar{I}$ DC bias current

- Therefore, the noise current is given by:

$$\boxed{\langle i_N^2 \rangle = 2q \langle I \rangle \Delta f}$$

mean current (DC)

bandwidth of noise

Unlike Johnson noise, shot noise is only present when the mean photocurrent is nonzero, i.e. when photons are actually incident on the detector.

- Thermal noise : noise associated with random thermal motion of charge

Although it will not be derived here, this noise is :

$$\underset{\text{thermal current noise}}{\langle i_N^2 \rangle} = \frac{4kT}{R}\Delta f \qquad \underset{\text{thermal voltage noise}}{\langle v_N^2 \rangle} = 4kTR\,\Delta f$$

Unlike shot noise, the thermal noise in the photodetector does not depend on the number of photons incident

- Total noise :

$$\langle i_N^2 \rangle_{total} = 2q\langle I \rangle \Delta f + \frac{4kT}{R}\Delta f$$

The ratio of signal to noise power in a detector without gain (such as a p-i-n photodiode) is:

$$\boxed{SNR_{pin} = \frac{(RP)^2}{2q(RP + I_D)\Delta f + \frac{4kT}{R}\Delta f}}$$

where the photocurrent is

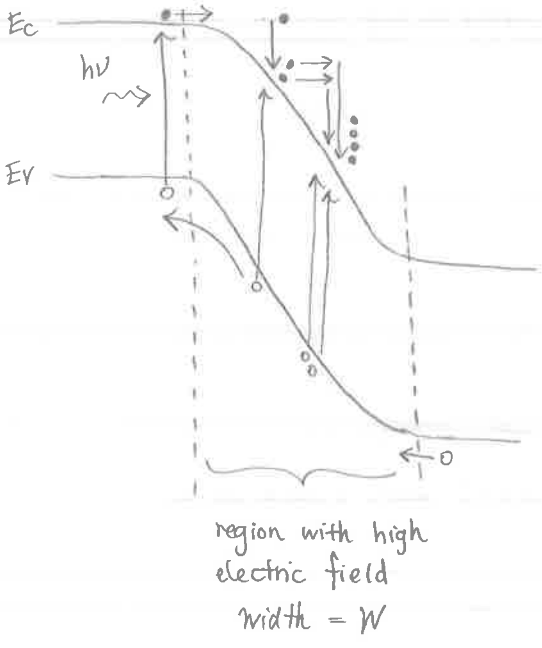
$$I_{ph} = \underset{\text{responsivity}}{R}\,\underset{\text{incident optical power}}{P}$$

and I_D = dark current

In a detector with gain, the shot noise term needs to be modified (see the next section on avalanche photodiodes)

- ## Avalanche Photodiode (APD)

A photodiode with an intrinsic carrier multiplication mechanism



- In the absorption region, the large electric field induces avalanche breakdown; carriers moving in this region to quickly gain enough kinetic energy to cause impact ionization, creating a new electron-hole pair
- In the ideal case, only one type of carrier experiences a multiplication effect. The current enhancement in this case is:

$$\frac{dJ_n}{dx} = \alpha_n J_n(x) \Rightarrow J_n(x) = e^{\alpha_n x} J_{no}$$

at the edge of the absorption region,

$$J_n(W) = e^{\alpha_n W} J_{no}$$

where $\alpha_n = e^-$ impact ionization coefficient (cm^{-1})

multiplication factor

current without multiplication

- In practice, since each impact ionization process generates an electron-hole pair, the generated holes can also cause impact ionization while moving in the opposite direction. In this case, coupled differential equations must be solved:

$$\frac{dJ_n}{dx} = \alpha_n J_n + \alpha_p J_p$$

$$-\frac{dJ_p}{dx} = \alpha_n J_n + \alpha_p J_p$$

where α_n and α_p are functions of the electric field in the depletion region

Note that

$$\frac{d}{dx}(J_n + J_p) = 0 \Rightarrow J = J_n + J_p \text{ is constant, due to current continuity}$$

$$J_p = J - J_n$$

- Solve:

$$\frac{dJ_n}{dx} = \alpha_n J_n + \alpha_p (J - J_n) = (\alpha_n - \alpha_p) J_n + \alpha_p J$$

If we assume $J_p(W) = 0$, then we can find that the multiplication factor is:

$$\boxed{M_n = \frac{1-k}{e^{-(1-k)\alpha_n W} - k}} \qquad \text{where } k = \frac{\alpha_p}{\alpha_n}$$

($k = 0$ recovers the single-carrier case)

- APD response time

$$\tau = \underset{\substack{\uparrow \\ \text{transit time}}}{\tau_t} + \underset{\substack{\nwarrow \\ \text{time of} \\ \text{multiplication} \\ \text{process}}}{\tau_m} \qquad \text{where} \quad \tau_m \approx \underbrace{\frac{M_n k W}{v_e} + \frac{W}{v_h}}_{\substack{\text{approximate expressing} \\ \text{for the duration of the} \\ \text{entire process, assuming} \\ \text{large } M_n}} , \qquad \tau_t = \frac{W_{abs}}{v_h}$$

W = multiplication region width

W_{abs} = absorption region width

$$\Rightarrow \tau \cong \frac{M_n k W}{v_e} + \frac{W + W_{abs}}{v_h}$$

- The gain-bandwidth product ignoring quantum efficiency is:

$$G \times BW = M_n \cdot \frac{1}{2\pi\tau}$$

$$G \times BW = \underset{\substack{\uparrow \\ \text{Generally} \\ \text{dominates}}}{\frac{v_e}{2\pi k W}} \cdot \frac{1}{1 + \frac{1}{M_n k} \frac{v_e}{v_h}\left(1 + \frac{W_{abs}}{W}\right)}$$

- Conclusions: the factor $k = \alpha_p / \alpha_n$ describes the extent to which feedback (due to impact ionization caused by hot holes) is present in the device

(1) A larger k gives larger gain

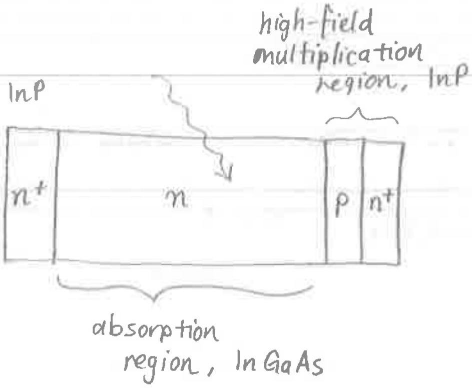
(2) A larger k increases response time, reduces bandwidth

↳ Overall, the second effect is stronger: the gain-bandwidth product drops with larger k, so this feedback is generally undesirable unless a large gain is the primary goal

(3) A larger k also increases noise

(4) The feedback process can cause instabilities, leading to uncontrolled avalanche: generally want to avoid this by making the multiplication region thin, $W \ll W_{abs}$

- Separate-absorption-multiplication (SAM) APD
  - Absorb photons in a large intrinsic region
  - Allow generated carriers to drift under a moderate electric field
  - Avalanching occurs in a thin multiplication region



- Noise in APDs:

- Since the gain mechanism involves random processes, noise is generated and then amplified. What is the effect on SNR?
  - Signal power is multiplied by $\langle M\rangle^2$
  - Shot noise power is multiplied by $\langle M^2\rangle$
- The quantity $\langle M^2\rangle$ is re-expressed using the noise factor F

$$\langle M^2\rangle = \langle M\rangle^2 F \qquad \text{where} \quad F = \frac{\langle M^2\rangle}{\langle M\rangle^2} = k\langle M_n\rangle + (1-k)\left(2 - \frac{1}{\langle M_n\rangle}\right)$$

* $F > 1$ and quantifies the excess noise caused by the gain process

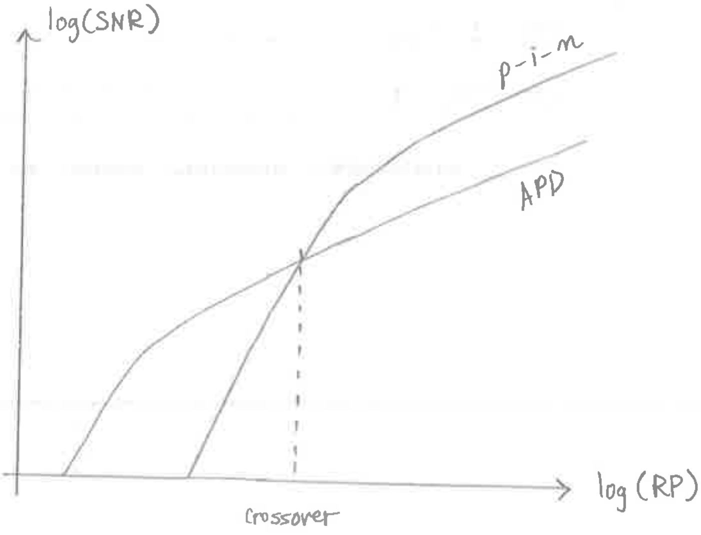
- Note: F increases with k due to the feedback process, which produces even more noise
- The noise figure is defined as $NF = 10 \log F$ (dB)

- Finally, the APD SNR is:

$$\boxed{SNR_{APD} = \frac{(RP)^2\langle M\rangle^2}{2q(RP + I_D)F\langle M\rangle^2\Delta f + \frac{4kT}{R}\Delta f}}$$

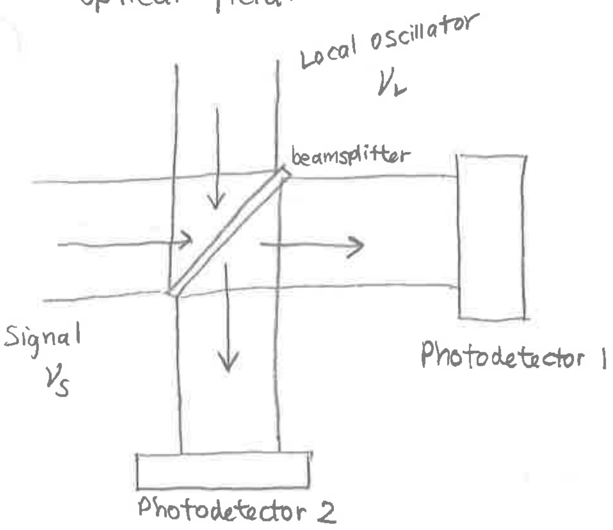
Since the noise factor F has a greater effect on SNR when the current is large, the APD is superior to the PIN only when the photocurrent is sufficiently small

The change in slope occurs once the shot noise becomes larger than the thermal noise →



- Coherent detection

It is possible to measure both the magnitude and phase of an incident optical field.



- Signal field

$$\mathcal{E}_s = \mathcal{E}_{so}\, e^{i2\pi\nu_s t}\, e^{i\phi_s}$$

where the amplitude and phase $\mathcal{E}_{so}$ and ϕ_s are modulated much more slowly than ν_s

- Local oscillator field

$$\mathcal{E}_L = \mathcal{E}_{LO}\, e^{i2\pi\nu_L t}\, e^{i\phi_L}$$

- Each photodetector will measure the intensity of the optical field.

- An ideal double balanced mixer is one in which:

(1) The beams are perfect plane waves and are perfectly aligned

(2) The beamsplitter equally splits the optical power into the 2 paths

(3) The two paths have relative lengths such that E_s and E_L will have a π phase lag in one path, compared to their relative path in the other

Therefore the intensities on the two detectors will be:

$$|\mathcal{E}_1|^2 = \tfrac{1}{2}|\mathcal{E}_s - \mathcal{E}_L|^2 = \tfrac{1}{2}\left[|\mathcal{E}_{so}|^2 + |\mathcal{E}_{LO}|^2 - 2\mathcal{E}_{so}\mathcal{E}_{LO}\cos\left(2\pi(\nu_s-\nu_L)t + (\phi_s-\phi_L)\right)\right]$$

$$\Rightarrow |\mathcal{E}_1|^2 = \tfrac{1}{2}\left(|\mathcal{E}_{so}|^2 + |\mathcal{E}_{LO}|^2\right) - \mathcal{E}_{so}\mathcal{E}_{LO}\cos\left(2\pi\Delta\nu\cdot t + (\phi_s-\phi_L)\right)$$

$$|\mathcal{E}_2|^2 = \tfrac{1}{2}\left(|\mathcal{E}_{so}|^2 + |\mathcal{E}_{LO}|^2\right) + \mathcal{E}_{so}\mathcal{E}_{LO}\cos\left(2\pi\Delta\nu\cdot t + (\phi_s-\phi_L)\right)$$

If $\Delta\nu$ is small, that is $\nu_s \approx \nu_L$, the incident light on the detector will be nearly monochromatic and the current is proportional to the optical intensity by: $I = \eta P_{opt}/h\nu = \eta\left(\tfrac{1}{2}\varepsilon|\mathcal{E}|^2 A\right)/h\nu \propto |\mathcal{E}|^2$

$$\Rightarrow I_1 = \tfrac{1}{2}I_s + \tfrac{1}{2}I_L + \sqrt{I_s I_L}\cos\left(2\pi\Delta\nu\cdot t + (\phi_s(t) - \phi_L(t))\right)$$

$$I_2 = \tfrac{1}{2}I_s + \tfrac{1}{2}I_L - \sqrt{I_s I_L}\cos\left(2\pi\Delta\nu\cdot t + (\phi_s(t) - \phi_L(t))\right)$$

current due to direct detection of signal power

current due to direct detection of LO power

interference term

- Notice that if we subtract the two detected signals,

$$I_2 - I_1 = 2\sqrt{I_S I_L}\,\cos\left[2\pi \Delta\nu \cdot t + \phi_s(t) - \phi_L(t)\right] \qquad \text{(heterodyne)}$$

$$\approx 2\sqrt{I_S I_L}\,\cos(\phi_s(t) - \phi_L(t)) \qquad (\text{homodyne}, \Delta\nu = 0)$$

Knowing all the parameters of the local oscillator ($\mathcal{E}_{L0}, \nu_L, \phi_L$), the magnitude of the signal $\mathcal{E}_{so}$ can be estimated from $I_1(t)$ or $I_2(t)$, while the phase $\phi_s(t)$ can easily be estimated from $I_2(t) - I_1(t)$.

- Advantages of coherent detection:
  - Measure both the modulated intensity and modulated phase of the optical field
  - The measured field is large even if the signal is weak, due to the presence of a strong local oscillator; this improves the SNR,

$$SNR_{\text{coherent}} = \frac{\langle |2\sqrt{I_S I_L}\,\cos(\phi_s - \phi_L)|^2\rangle}{2q\langle I_L\rangle \Delta f} = \frac{2\langle I_S\rangle\cancel{\langle I_L\rangle}}{2q\cancel{\langle I_L\rangle}\Delta f}$$

$$SNR_{\text{coherent}} = \frac{\langle I_S\rangle}{q\Delta f}$$

where we assumed the local oscillator is much stronger than the signal: $\langle I_L\rangle \gg \langle I_S\rangle$ and is also large enough to overcome thermal noise: $2q\langle I_L\rangle \gg 4kT/R$

- Meanwhile, the SNR of direct detection without gain is:

$$SNR_{\text{direct}} = \frac{\langle I_S\rangle^2}{2q\langle I_S\rangle\Delta f + (4kT/R)\Delta f} \rightarrow \frac{\langle I_S\rangle}{2q\Delta f}$$

So the SNR of coherent detection is ~2x for relatively strong signals, and even more for weak signals

- Compare to an APD:

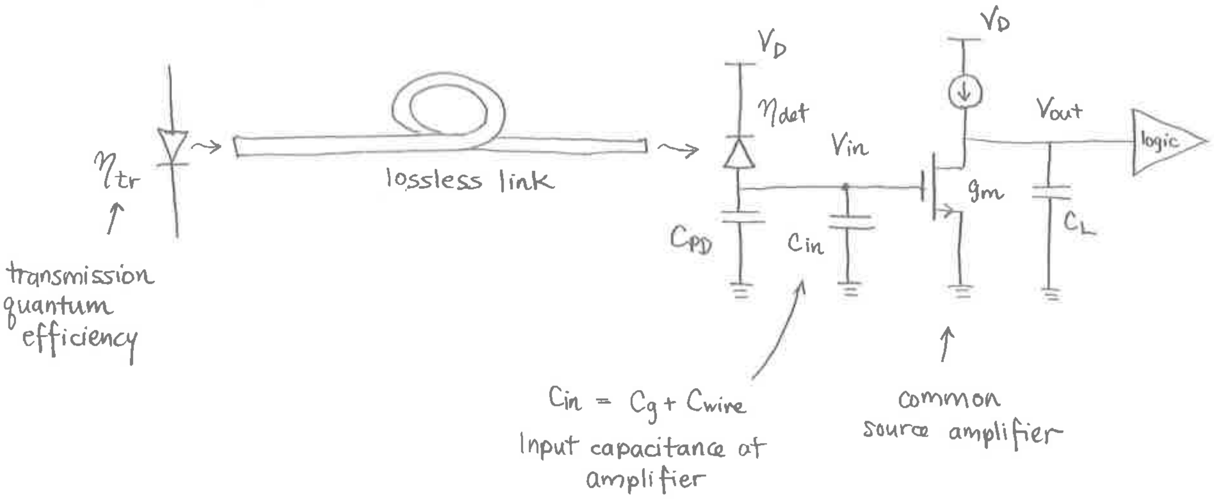
- Note: An APD is not desirable with coherent detection, because it possesses an SNR benefit only when it is not shot noise limited!

$$SNR_{\text{APD}} = \frac{\langle I_S\rangle^2 M^2}{2q\langle I_S\rangle F M^2 \Delta f} = \frac{\langle I_S\rangle}{2qF\Delta f}$$

So coherent detection is still 2x better than even a noiseless APD ($F = 1$); the local oscillator provides a type of noiseless gain!

- Receiver circuit analysis (from Ryan Going's thesis, Ch 2)

Consider the optical link below:



- The power expended emitter is:

$$P_{tx} = \hbar\omega \cdot \eta_{tr} \cdot N_{ph} \cdot f_{bit}$$

where N_{ph} = # photons/bit
f_{bit} = bit rate (s^{-1})

- The power expended at the receiver is:

$$P_{rx} = I_D V_D$$

  - It can be shown that the transconductance of the MOS amplifier is:

$$g_m = \frac{I_D}{2(V_{GS} - V_{th})} = \frac{I_D}{2V_{ov}} \quad \text{where} \quad V_{ov} \equiv V_{GS} - V_{th}$$

  - We assume that the transistor is designed so that its 3dB bandwidth coincides with the bit rate f_{bit}. Assuming some other ideal amplifier properties, we can obtain

$$2\pi f_{bit} = \frac{g_m}{A_v C_L} \quad \text{where} \quad A_v = \frac{V_{out}}{V_{in}} \text{ is the gain}$$

  - The input voltage is the voltage produced on the collection of capacitors C_{PD}, C_{wire}, and C_g when charge is generated by the absorption of a photon (assume a photodiode with no gain)

$$\Rightarrow V_{in} = \frac{q N_{ph} \eta_{det}}{C_{PD} + C_{wire} + C_g}$$

  - Putting this together, we get:

$$P_{rx} = V_D \cdot 2V_{ov} g_m = 2V_D V_{ov} 2\pi f_{bit} C_L \frac{V_{out}}{V_{in}}$$

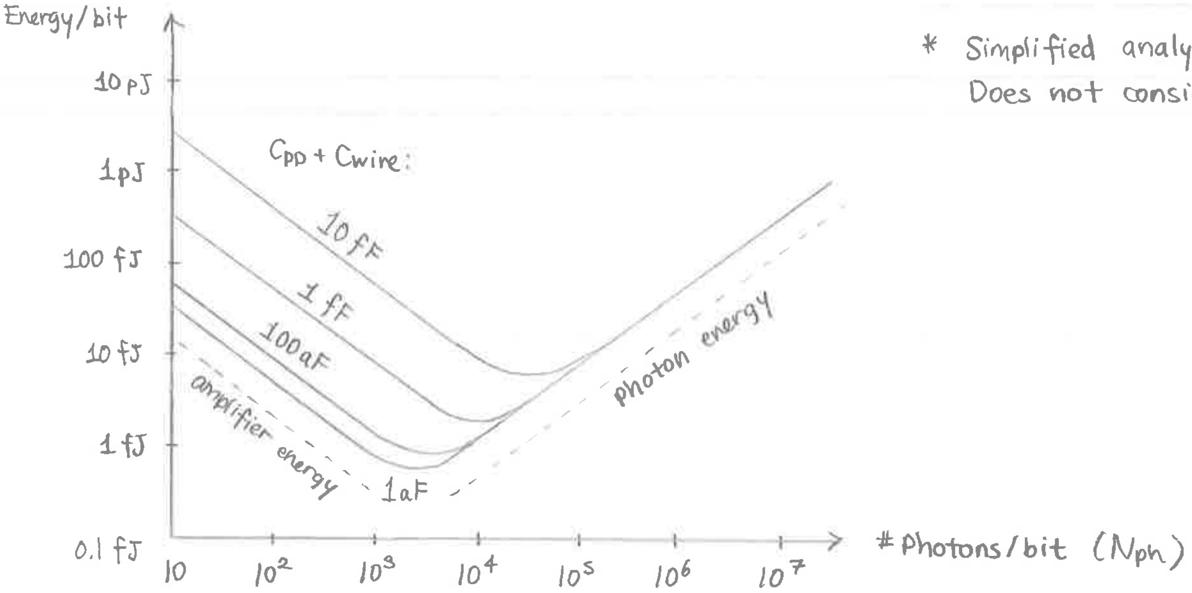
$$= \frac{4\pi V_{ov} V_{out} V_D C_L (C_{PD} + C_{wire} + C_g)}{q N_{ph} \eta_{det}} \cdot f_{bit}$$

- Find the total power in the transmitter and receiver, then divide by the bit rate to get the energy/bit of the optical link

$$\frac{\text{Energy}}{\text{bit}} = \frac{P_{tx} + P_{rx}}{f_{bit}} = \underbrace{\hbar\omega\,\eta_{tr} N_{ph}}_{\text{transmitter energy}} + \underbrace{\frac{4\pi V_{ov} V_{out} V_D C_L (C_{PD} + C_g + C_{wire})}{q N_{ph} \eta_{det}}}_{\text{receiver energy}}$$

- Using the values: $\eta_{tr} = \eta_{det} = 1$
  $V_D = V_{out} = 1V$, $V_{ov} = 100\,mV$ → realistic values compared to ITRS, 2013
  $C_g = C_L = 200\,aF$

- This leads to the following plot:



* Simplified analysis!
Does not consider SNR

- Conclusions:
  - Reducing the number of photons/bit does not always improve the energy/bit of communication:
    - For large N_{ph}, the transmitter dominates the energy cost, so reducing N_{ph} helps
    - For small N_{ph}, the amplifier energy dominates: the photodiode signal becomes very low, and more electrical power will be needed to amplify the signal
  - The photodetector capacitance (including C_{wire}) must be reduced to decrease the lowest achievable energy/bit
    - Low C_{PD} → more voltage per photon
    - C_{wire} can be reduced by tight integration of the photodiode and amplifier

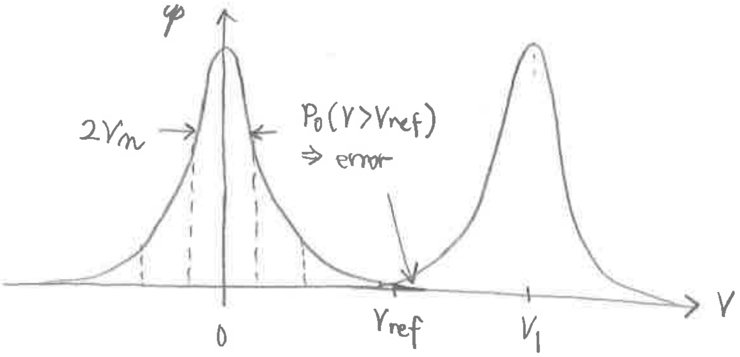
- ## Bit error rate and SNR

  - The bit error rate is broadly defined as the fraction of detected bits that contains an error; a tolerable value is no more than 10^{-9}, a specification of 10^{-15} is common

  - Suppose that if a '1' signal is detected, the expected voltage signal is V_1. If a '0' signal is detected, the expected signal voltage is 0. Suppose also that the noise is Gaussian, with a probability distribution functions given below:

$$\varphi_1(V) = \frac{1}{V_n\sqrt{2\pi}} e^{-\frac{(V-V_1)^2}{2V_n^2}}$$

$$\varphi_0(V) = \frac{1}{V_n\sqrt{2\pi}} e^{-\frac{V^2}{2V_n^2}}$$



where V_n^2 is the noise variance.

  - In a receiver system, a digital comparator inspects the signal voltage V and decides that a bit is '1' if $V > V_{ref}$, and '0' if $V < V_{ref}$, and usually $V_{ref} = \frac{1}{2}V_1$ is chosen.

    → An error occurs if $V > V_{ref}$ even though the bit is '0'. We can find this probability by integrating φ_0:

$$\text{BER} = \int_{V_{ref}}^{\infty} \varphi_0(V)\, dV = \int_{V_{ref}}^{\infty} \frac{1}{V_n\sqrt{2\pi}} e^{-\frac{V^2}{2V_n^2}} dV$$

Let $z = \frac{V}{V_n}$, $dz = \frac{dV}{V_n}$, so $z_{ref} = \frac{V_{ref}}{V_n}$

$$\Rightarrow \text{BER} = \frac{1}{\sqrt{2\pi}} \int_{z_{ref}}^{\infty} e^{-z^2/2} dz \equiv Q(z_{ref}) = Q\left(\frac{V_{ref}}{V_n}\right)$$

  - Notice that if V_1 represents the expected voltage when a voltage is present, then $V_{ref}/V_n = V_1/2V_n$ represents half the SNR

$$\Rightarrow \boxed{\text{BER} = Q\left(\tfrac{1}{2}\text{SNR}\right) = \frac{1}{\sqrt{2\pi}} \int_{\frac{1}{2}\text{SNR}}^{\infty} e^{-z^2/2} dz} \quad (\text{Gaussian noise})$$

For BER $= 10^{-9}$, we find SNR $= 6$

- Quantum noise limit

  - If we consider only shot noise, what is the minimum number of photons we can use and still transmit with a specified BER?

$$SNR = \frac{I_{ph}}{\sqrt{2qI_{ph}\Delta f}} = \frac{I_{ph}}{\sqrt{qI_{ph}f_{bit}}} = \sqrt{\frac{I_{ph}/q}{f_{bit}}} = \underbrace{\sqrt{N_{ph}}}_{\frac{\#\text{ photons}}{\text{bit}}}$$

$\Delta f = \frac{1}{2} f_{bit}$

(Nyquist limit)

We know that if the noise is Gaussian, the required $SNR = 6$ to get $BER = 10^{-9}$
This means that:

$$N_{ph} = SNR^2 = \underline{36 \text{ photons / bit}} \quad (\text{for '1'})$$

  - The shot noise actually follows a Poisson distribution, which is well approximated as Gaussian noise only for large photon numbers. For small photon numbers, the error is significant, and the above is modified to 20 photons / bit (for '1')

⑪ Optical modulators

- There are various ways to modulate an optical signal other than directly modulating the laser diode itself; these methods are called external modulation
- Motivations for external modulation
  - Potentially achieve a greater modulation bandwidth (not limited by ω_R)
  - Avoid the frequency chirping effect in direct modulation
  - Makes possible phase modulation → coherent detection
- The cost is higher power consumption and greater complexity
- Most high-speed optical links > 10 Gbps rely on external modulators

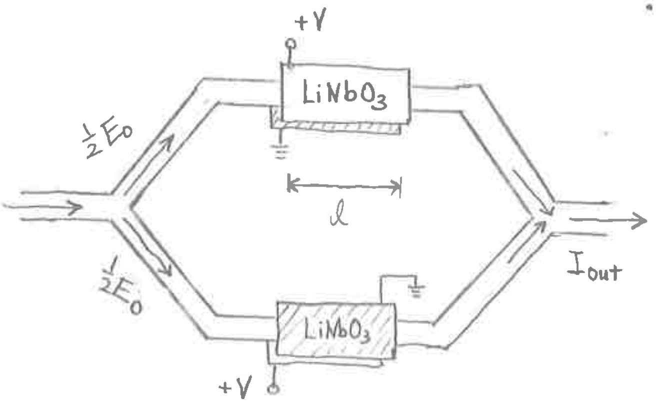
---

- Electro-optic modulators
  - In general, the refractive index of a material can be a function of the field, due to non-linearities:

$$n(E) = \underbrace{n_0 + a_1 E}_{\text{linear electro-optic (Pockels) effect}} + \underbrace{a_2 E^2}_{\text{quadratic electro-optic (Kerr) effect}} + \ldots$$

  - Crystals with centro-symmetry have $a_1 = 0$, so the quadratic effect dominates (though it still tends to be weak)
  - Careful engineering is necessary to produce a desired Δn along the direction of propagation of the field
    - e.g. A field of 10^5 V/cm on GaAs can produce $\Delta n \approx 3 \times 10^{-4}$

- Integrated electro-optic intensity modulator:



The two paths experience the same Δn but opposite in sign

- The waveguide mode is split equally at the first fork, then they are combined at the second fork:

$$I_{out} = \left| \tfrac{1}{2} E_0 e^{i\frac{2\pi}{\lambda}(n_0+\Delta n)l} + \tfrac{1}{2} E_0 e^{i\frac{2\pi}{\lambda}(n_0-\Delta n)l} \right|^2$$

$$= |E_0|^2 \left| \tfrac{1}{2} + \tfrac{1}{4} e^{i\frac{2\pi}{\lambda}2\Delta n l} + \tfrac{1}{4} e^{-i\frac{2\pi}{\lambda}2\Delta n l} \right|$$

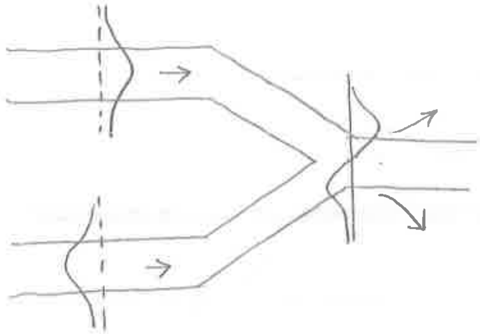
$$= |E_0|^2 \left(\tfrac{1}{2} + \tfrac{1}{2}\cos\left(\tfrac{4\pi}{\lambda}\Delta n l\right)\right)$$

$$\boxed{I_{out} = |E_0|^2 \cos^2\left(\tfrac{2\pi}{\lambda}\Delta n \cdot l\right)}$$

intensity modulation through $\Delta n(V)$

$$I_{out} = \frac{1}{2}|E_0|^2 \left[1 + \cos\left(\frac{4\pi}{\lambda}\Delta n(V) l\right)\right] = |E_0|^2 \cos^2\left(\frac{2\pi}{\lambda}\Delta n(V) l\right)$$

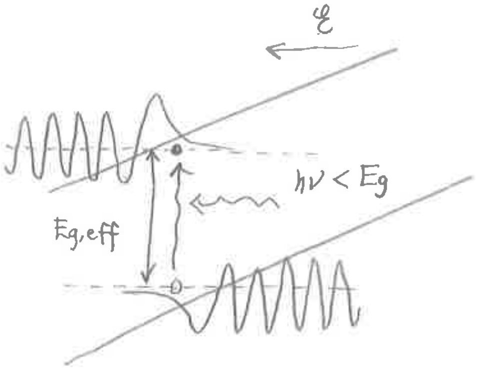
- When $V = 0$, $\Delta n = 0$ and $I_{out} = |E_0|^2$
- When V is such that $\frac{4\pi}{\lambda}\Delta n \cdot l = \pi$, $I_{out} = 0$. What is the meaning of this destructive interference?



The interfering modes add up to a first-order mode, which is not supported by the output fiber, so it quickly leaks out

- $LiNbO_3$ electro-optic modulators have been demonstrated with speeds in the 10 - 40 Gbps range

- Franz-Keldysh effect
  - When a voltage is applied across a semiconductor absorption region, the electron wavefunctions become Airy functions (triangular potential well solutions to Schrödinger's equation):



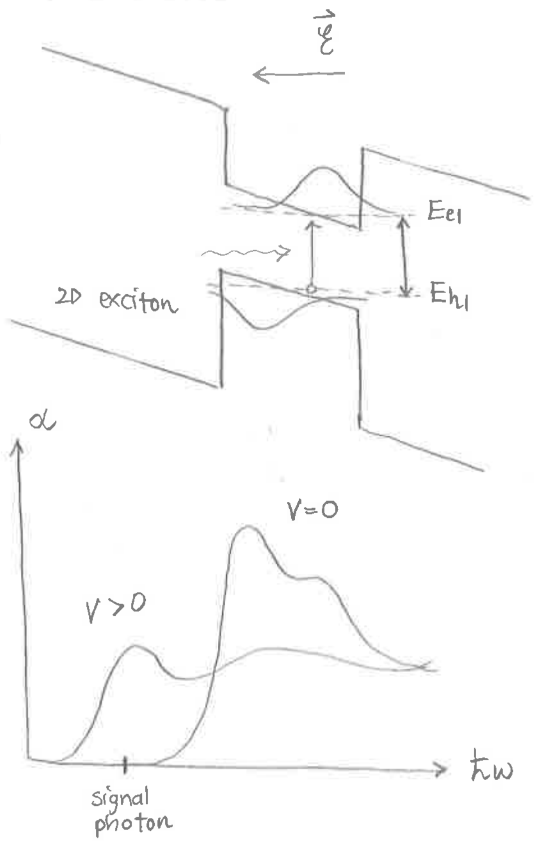
  - As a result, due to wavefunction penetration into the gap, it becomes possible for photons below the bandgap energy to be absorbed. The effective bandgap is:

$$\Delta E_g = \left(\frac{q}{m^*}\right)^{1/3} (q\hbar\mathscr{E})^{2/3}$$

  - This can be considered a photon-assisted tunneling process

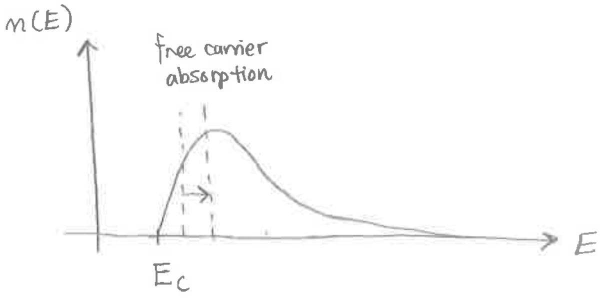
  - The absorption can be modulated with voltage
    - Choose a semiconductor with $E_g > h\nu$, where $h\nu$ is the energy of the signal photon
    - If no voltage is applied, the material is transparent to the photon → '1'
    - If a sufficient voltage is applied, the material can absorb the photon via the Franz-Keldysh effect → '0'
      - Up to $\alpha \approx 1000\ cm^{-1}$ is achievable

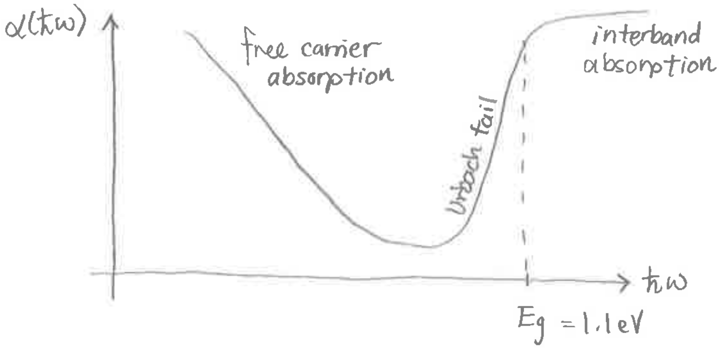
- Quantum-confined Stark effect



- Under an electric field, the effective bandgap E^{e1}_{h1} of a quantum well also decreases, due to the separation of the electron and hole wavefunctions to opposite sides of the quantum well
- The absorption edge effectively shifts to a lower energy, much like the Franz-Keldysh effect in a buk material under an applied voltage
- The generated electron and hole stay separated, so this can be considered a form of exciton absorption; the electron and hole interact through the Coulomb potential

- Free carrier effect in Si
  - Free carrier absorption: absorption of photons by carriers, with initial and final states in the same band ⇒ photon energy is converted to kinetic energy





- Due to the Fermi-Dirac distribution, most of the electrons occupy states within a few kT of the conduction band, with a decaying thermal tail at high energies
  - ⇒ number of available free carrier energy transitions decrease with larger photon energy
  - ⇒ $\alpha(\hbar\omega)$ decreases with photon energy, until the onset of interband absorption
- Carrier density dependence: the above physics can be captured in a free carrier absorption cross-section:

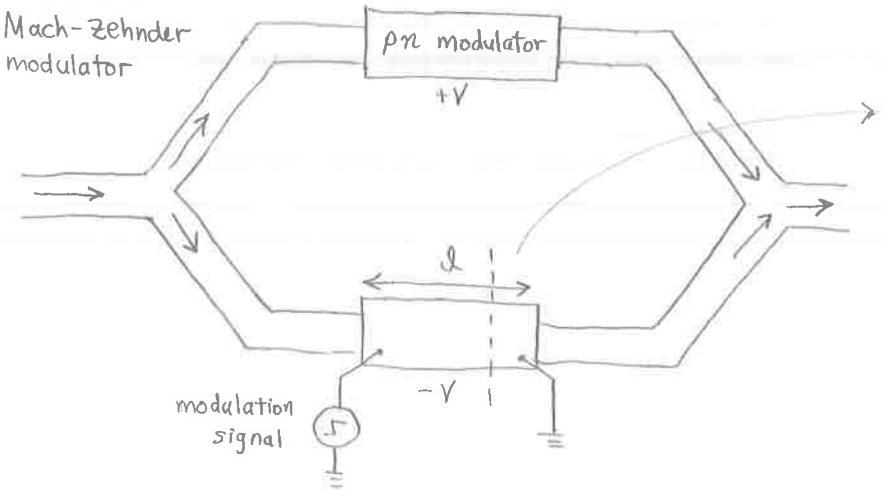
$$\underset{(cm^{-1})}{\alpha(\hbar\omega)} = \underset{(cm^{2})}{\sigma(\hbar\omega)} \cdot \underset{(cm^{-3})}{n}$$

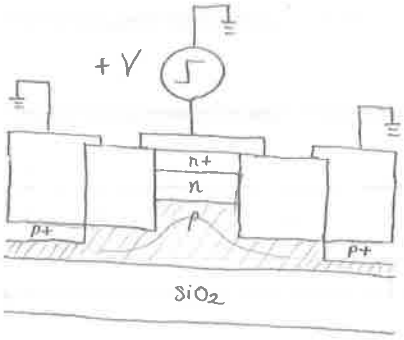
- Free carrier index modulation: the Kramers-Kronig relations require that a change in the absorption coefficient of the material be accompanied by a change in the real permittivity, and therefore the refractive index Δn
  - In Si, $\Delta n \approx (-8.8 \times 10^{-22}\,\text{cm}^3) \cdot \Delta N_e + (-8.8 \times 10^{-18}\,\text{cm}^{3 \cdot 0.8}) \cdot \Delta N_h^{0.8}$

    Thus, if $\Delta N_e = \Delta N_h = 10^{18}\,\text{cm}^{-3}$, $\Delta n \approx -0.003$

    This index change is capable of producing a large index change over $l \sim 100\mu m - 1mm$

- Free carrier effect modulator for Si photonics





- By applying a reverse bias to the p-n junction, the depletion region which coincides with the optical mode can be further deprived of carriers (or supplied with carriers if biased in the forward direction)

  - There is a fundamental trade-off between index change and loss!
    - A longer device produces a larger phase shift, but also more loss
    - A greater carrier density produces a greater index change, but also gives rise to more absorption
  - Carrier density can be modulated at ~ 50 Gbps

- Thermo-optic effect
  - The refractive index is a function of temperature, due to thermal expansion and the bandgap dependence on temperature
    - In Si, $\frac{dn}{dT} \approx 2 \times 10^{-4}\,K^{-1} \Rightarrow \Delta n = 0.05$ for $\Delta T = 300K - 270K$

      Larger than the free carrier effect!
  - Slow response: time to heat/cool is ~1ms or ~1μs for very small structures
    Suitable for tunable filters, but not modulators
  - Large power consumption

⑫ $\boxed{\text{Semiconductor Devices}}$: selected concepts

- Current flow:
  - **Drift**: $J_{dr} = qnv = \overbrace{qn\mu}^{\sigma} E$ where v = drift velocity

    where μ is the mobility: $\mu = q\tau/m^*$ with τ = mean free time
    - μ decreases with dopant concentration due to ionized impurity scattering
    - temperature dependence:
      - lattice scattering increases with temperature $\Rightarrow \mu\downarrow$
      - ionized impurity scattering decreases with temperature because ions have a stronger effect on carriers with smaller thermal velocity: $v_{th} = \sqrt{\frac{3kT}{m^*}}$ $\Rightarrow \mu\uparrow$ with temperature
      - In heavily doped samples, the mobility peaks at an intermediate temperature due to these effects
    - At high fields, the drift velocity saturates due to strong scattering from optical phonons
  - **Diffusion**: $J^n_{dif} = qD_N\nabla n$ (electrons) where D is the diffusivity

    $J^p_{dif} = -qD_p\nabla p$ (holes)
  - In equilibrium, the total current is zero for each carrier type: for electrons, in 1D we have

$$J^n_{dr} + J^n_{dif} = q\mu_n nE + qD_N\frac{dn}{dx} = 0$$

Note:

$$E = \frac{1}{q}\frac{dE_c}{dx} = \frac{1}{q}\frac{dE_i}{dx}, \qquad n = n_i e^{(E_F - E_i)/kT}$$

0: equilibrium

$$\Rightarrow \frac{dn}{dx} = \frac{n_i}{kT}e^{(E_F - E_i)/kT}\left(\cancelto{0}{\frac{dE_F}{dx}} - \frac{dE_i}{dx}\right)$$

$$\frac{dn}{dx} = -\frac{qE}{kT}\cdot n$$

Therefore we have

$$q\mu_n nE + qD_N\left(-\frac{qE}{kT}\right)n = 0 \quad\Rightarrow\quad \mu_n = \frac{qD_N}{kT}$$

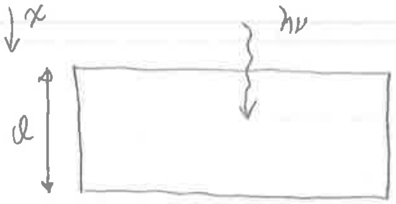
$$\text{or} \quad \boxed{\frac{D_N}{\mu_N} = \frac{kT}{q}}$$

<u>Einstein relation</u>

D_N : cm^2/s , μ_N : cm^2/V·s

- Recombination / generation

  - Optical generation



- Assume monochromatic incident light with $h\nu > E_g$. The optical density decays into the semiconductor as:

$$I = I_0 \, e^{-\alpha(h\nu)\cdot x}$$

- Photons generate electron-hole pairs: thus the generation rate is the same for both carriers and has the same spatial dependence:

$$\left.\frac{\partial n}{\partial t}\right|_{opt} = \left.\frac{\partial p}{\partial t}\right|_{opt} = G_{LO} \, e^{-\alpha x}$$

  - Shockley - Reed - Hall recombination/generation

  - Let n_0, p_0 = equilibrium carrier concentrations, and $\Delta n = n - n_0$, $\Delta p = p - p_0$

    - low-level injection: the excess minority carrier concentration is still much smaller than the majority carrier concentration. This constitutes a small perturbation: $\Delta n \ll p_0$ (in p-type), $\Delta p \ll n_0$ (in n-type)
      - This will be assumed through the analysis

  - The net thermal SRH recombination rate is given by (in an n-type material):

$$\left.\frac{\partial p}{\partial t}\right|_{SRH} = \left.\frac{\partial p}{\partial t}\right|_{SRH,R} + \left.\frac{\partial p}{\partial t}\right|_{SRH,G}$$

$$= -C_p N_T p + C_p N_T p_0 = -C_p N_T (p - p_0)$$

hole recombination rate:
$E_F > E_T$, so # electrons that
occupy the travel simply
equals the trap density N_T

hole generation rate

- The constant C_p depends on the trap cross section σ_{Tp}, the carrier thermal velocity $v_{th,p}$, etc.

$$\left.\frac{\partial p}{\partial t}\right|_{SRH} = -\frac{\Delta p}{\tau_p} \quad \text{where} \quad \tau_p = \frac{1}{C_p N_T}$$

- In equilibrium, $\left.\frac{\partial p}{\partial t}\right|_{SRH} = 0$, as expected. In nonequilibrium conditions, this quantity is nonzero, and is of course balanced in steady state by the excitation that drives the system out of equilibrium

- τ_p is the minority carrier lifetime. In a p-type material, τ_n is the minority carrier lifetime.

- Continuity equations:

- Start with Gauss's law: $\vec{\nabla}\cdot\vec{E} = \frac{1}{\varepsilon}\rho$

  Take time derivative: $\vec{\nabla}\cdot\frac{\partial\vec{E}}{\partial t} = \vec{\nabla}\cdot\left(\frac{1}{\varepsilon\mu_0}\vec{\nabla}\times\vec{B} - \frac{1}{\varepsilon}\vec{J}\right) = \frac{1}{\varepsilon}\frac{\partial\rho}{\partial t}$

$$\vec{\nabla}\cdot\vec{J} = -\frac{\partial\rho}{\partial t}$$

If we apply this separately to electrons and holes, we have:

$$\frac{\partial n}{\partial t} = \frac{1}{q}\vec{\nabla}\cdot\vec{J}_N + \left.\frac{\partial n}{\partial t}\right|_{\text{other}}$$

$$\frac{\partial p}{\partial t} = \underbrace{-\frac{1}{q}\vec{\nabla}\cdot\vec{J}_P}_{\substack{\text{drift and diffusion}\\ \text{in/out of a volume}}} + \underbrace{\left.\frac{\partial p}{\partial t}\right|_{\text{other}}}_{\substack{\text{recombination}\\ \text{and generation}\\ \text{within a volume}}}$$

$\rightarrow$ These are the continuity equations for carriers in a semiconductor

- Carrier lifetime measurement via optical excitation

  - Consider an n-type material in 1D, and that the intensity of radiation is uniform within its volume (i.e. weak absorption), and that there is no applied field:

$$\frac{\partial\Delta p}{\partial t} = -\frac{1}{q}\frac{\partial}{\partial x}J_{p,\text{drift}} + \left.\frac{\partial p}{\partial t}\right|_{\text{SRH}} + G_L$$

$$= -\frac{1}{q}\frac{\partial}{\partial x}\left(-qD_p\frac{\partial}{\partial x}\Delta p\right) - \frac{\Delta p}{\tau_p} + G_L$$

$$\frac{\partial\Delta p}{\partial t} = \underset{\substack{\uparrow\\ \text{diffusion}}}{D_p\frac{\partial^2\Delta p}{\partial x^2}} - \underset{\substack{\uparrow\\ \text{SRH}\\ \text{recombination}}}{\frac{\Delta p}{\tau_p}} + \underset{\substack{\uparrow\\ \text{optical}\\ \text{generation}}}{G_L}$$

  - Under the assumption of uniform generation, Δp is the same everywhere

$$\Rightarrow \frac{\partial\Delta p}{\partial t} = G_L - \frac{\Delta p}{\tau_p}$$

Applying the boundary condition that $\Delta p(t=0) = 0$, the solution is:

$$\Delta p(t) = \underset{\substack{\uparrow\\ \text{excess hole}\\ \text{density in steady}\\ \text{state}}}{G_L\tau_p}\left(1 - e^{-t/\tau_p}\right)$$

  - After the light shuts off, Δp decays from its steady-state value:

$$\Delta p(t) = G_L\tau_p e^{-t/\tau_p}$$

- A photoconductivity measurement can be used to extract τ_p:

$$\Delta\sigma(t) = q(\mu_n + \mu_p)\,\Delta p(t)$$

The decay constant of conductivity is easily manifested in the decay of a voltage across a load; this gives the lifetime τ_p

- Diffusion length

  - Suppose that light is fully absorbed within an infinitesimal sheet at the semiconductor's surface: what is the distribution of minority carriers?

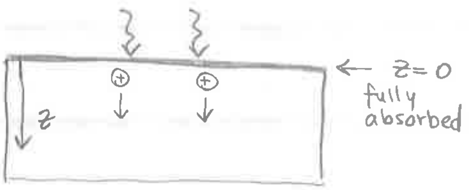
    - Away from the surface ($z>0$), $G_L = 0$ and in steady-state $\partial\Delta p/\partial t = 0$:

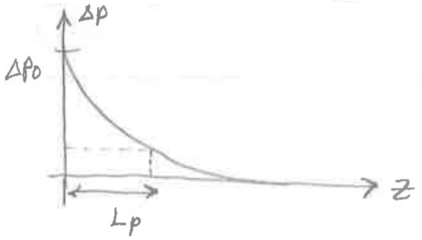
$$\frac{\partial\Delta p}{\partial t} = 0 = D_p\frac{\partial^2\Delta p}{\partial x^2} - \frac{\Delta p}{\tau_p}$$

Solve this subject to $\Delta p(z=0) = \Delta p_0$, $\Delta p(z\to\infty) = 0$

$$\Rightarrow \quad \Delta p(z) = \Delta p_0\, e^{-z/L_p}$$

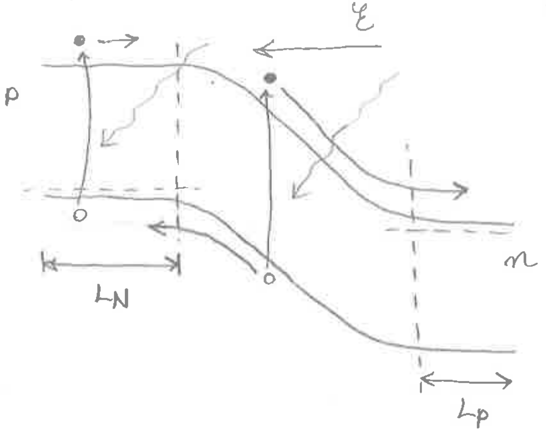
where $\boxed{L_p = \sqrt{D_p\tau_p}}$ is the diffusion length





- This length refers to the average distance that a minority carrier can travel before recombining

  - In photodetectors:



- If a photon is absorbed in the depletion region, it is swept by the large electric field to the region where it is a majority carrier
  - ↳ the external quantum efficiency (# absorbed photons to the # collected carriers) is ≈ 1
- If a photon is absorbed within a diffusion length outside the depletion region, the generated minority carrier has a good chance of reaching the depletion region, then collected
  - ⇒ EQE ≈ 50%
- If a photon is absorbed more than a diffusion length from the depletion region, the generated minority carrier will almost certainly recombine before reaching the depletion region (via SRH)

  - In solar cells:

- To fully absorb sunlight, the cell must be sufficiently thick; the diffusion length must be at least this thickness, as otherwise the collection of minority carriers becomes very inefficient (low EQE)

Diode ideality factor

- An ideal diode has the following I-V characteristic:

$$I = I_0 \left(e^{qV/kT} - 1\right)$$

In getting to this equation, several ideal assumptions were made:

(1) Low injection
(2) No recombination / generation in the depletion region
(3) Excess minority concentration → 0 at Ohmic contacts
(i.e. quasi-neutral regions are several diffusion lengths thick)

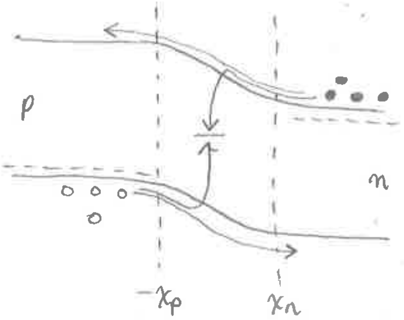
- Deviating from any of these assumptions changes the ideal characteristic. The "ideality factor" measures these deviations

$$I(V) \propto e^{qV/nkT} \qquad n = 1 \text{ for ideal diode}$$

This concept is relevant under forward bias when the I-V curve is exponential. The value of n can vary with voltage.

- Assumption ① is broken at high voltage, ② is broken at low voltage, and ③ is generally valid.

- Recombination / generation in the depletion region



- Under forward bias, an excess current flows due to SRH recombination in the depletion region

$$I_{SRH} = -qA \int_{-x_p}^{x_n} \left.\frac{\partial n}{\partial t}\right|_{SRH} dx$$

A rigorous treatment of SRH gives:

$$\left.\frac{\partial n}{\partial t}\right|_{SRH} = \frac{np - n_i^2}{\tau_p(n+n_1) + \tau_n(p+p_1)}$$

$$\text{where} \quad n_1 \equiv n_i e^{(E_T - E_i)/kT}, \quad p_1 \equiv n_i e^{(E_i - E_T)/kT}$$

- The result for forward bias is:

$$I_{SRH} = \frac{qAn_i}{2\tau_0} \underset{\uparrow}{W} \frac{e^{qV/kT} - 1}{1 + \frac{V_{bi} - V}{kT/q} \frac{\sqrt{\tau_n\tau_p}}{2\tau_0} e^{qV/2kT}} \approx \frac{qAn_i}{\sqrt{\tau_n\tau_p}} \frac{V_{bi} - V}{kT/q} W \cdot \underset{\uparrow}{e^{qV/2kT}}$$

depletion width

The SRH current has ideality factor $n = 2$

- Note: in reverse bias, an excess reverse current is seen that arises from SRH generation in the depletion region

- High current phenomena:

- Series resistance:

$$I = I_0 e^{qV/kT} \rightarrow I_0 e^{q(V - \overset{\text{series resistance}}{IR_s})/kT}$$

This causes a high-current roll-over of the diode I-V curve

- High-level injection: the injected minority carrier concentration becomes comparable to the majority carrier concentration on either side

⇒ in p-type: $n_p p_p = n_p(N_A + n_p) = n_i^2 e^{qV/kT}$ where it is no longer true that $n_p << N_A$

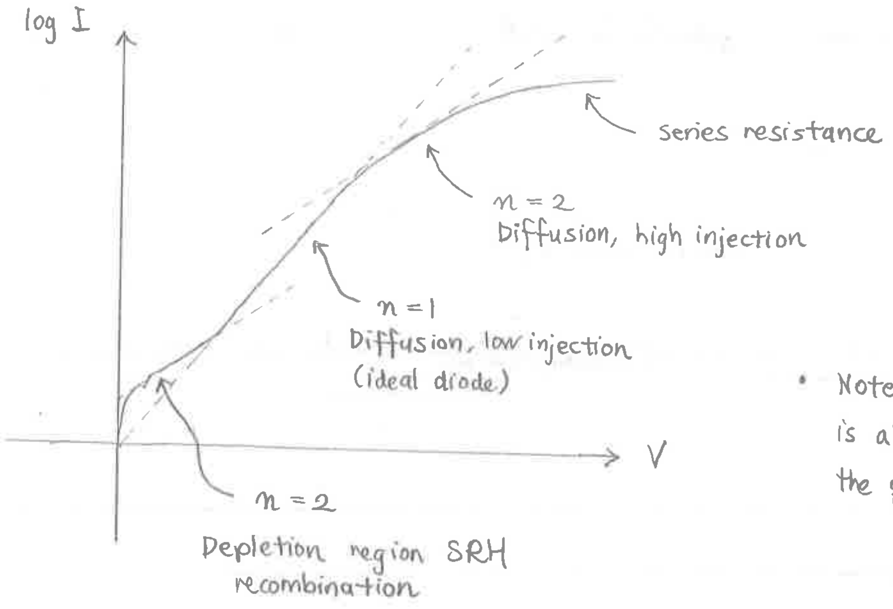
$$n_p = -\frac{N_A}{2} + \frac{1}{2}\sqrt{N_A^2 + 4n_i^2 e^{qV/kT}}$$

$$n_p = \frac{N_A}{2}\left[\sqrt{1 + \frac{4n_i^2}{N_A^2} e^{qV/kT}} - 1\right] \approx n_i e^{qV/2kT}$$

and in n-type: $$p_n = \frac{N_D}{2}\left[\sqrt{1 + \frac{4n_i^2}{N_D^2} e^{qV/kT}} - 1\right] \approx n_i e^{qV/2kT}$$

$$\Rightarrow J = \left(q\frac{D_N}{L_N} + q\frac{D_P}{L_P}\right) n_i e^{qV/2kT}$$ ⟶ Thus, $n = 2$ under high level injection

- Combining these effects, a typical I-V curve may look like:

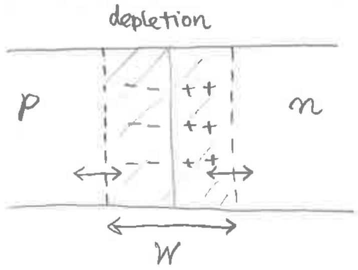


- Note: under high injection, there is also a voltage drop induced in the quasi-neutral region

- AC response of diodes

The response is significantly different for a reverse-biased diode (e.g. a photodiode) vs. a forward-biased diode (e.g. an LED / laser)

- Reverse bias: the response is dominated by the junction capacitance



- An AC voltage signal in reverse bias tends to modulate the depletion width W
- Since the conductance is small in reverse bias, the diode acts as a variable capacitor:

$$\boxed{C_J = \frac{\varepsilon A}{W}} \quad \text{where} \quad W \cong \sqrt{\frac{2\varepsilon}{q N_D}(V_{bi} - V)}$$

N_D ← lightly doped side

- If the modulation is small, C_J is constant and the speed of the diode circuit to communicate its signal to a load is: $\tau = R_L C_J$

- Forward bias:
  - The junction capacitance still exists, but there is also significant contribution from minority carrier charge oscillations; this is the origin of diffusion capacitance, which is frequency-dependent
  - We must solve the small-signal version of the minority carrier diffusion equation. Consider the n-side of a p^+-n junction:

$$\frac{\partial \Delta p}{\partial t} = D_p \frac{\partial^2 \Delta p}{\partial x^2} - \frac{\Delta p}{\tau_p}$$

small signal AC

$$i\omega \tilde{p} e^{i\omega t} = D_p \frac{\partial^2 \tilde{p}}{\partial x^2} e^{i\omega t} - \frac{\tilde{p}}{\tau_p} e^{i\omega t}$$

where $\tilde{p} e^{i\omega t}$ is the small AC modulation of the minority carrier concentration

$$0 = D_p \frac{\partial^2 \tilde{p}}{\partial x^2} - \tilde{p}\left(\frac{1}{\tau_p} + i\omega\right)$$

$$0 = D_p \frac{\partial^2 \tilde{p}}{\partial x^2} - \frac{\tilde{p}}{\tau_p}(1 + i\omega\tau_p)$$

- The boundary conditions are: $\tilde{p}(x \to \infty) = 0$

and $\tilde{p}(x = x_n) = \frac{n_i^2}{N_D} e^{qV/kT}\left(e^{q\tilde{v}/kT} - 1\right)$ where $\tilde{v}$ is the small AC voltage signal, V is the DC bias

- Solution for current:

$$\tilde{i} = \frac{q\tilde{v}}{kT} I_0 e^{qV/kT} \sqrt{1 + i\omega\tau_p} = \tilde{v} \cdot G_0 \sqrt{1 + i\omega\tau_p}$$

where $G_0 = \frac{dI}{dV} = \frac{q I_0}{kT} e^{qV/kT}$ is the low-frequency conductance

$$= \frac{I}{kT/q}$$

$$\tilde{i} = \underbrace{G_0\sqrt{1+i\omega\tau_p}}_{\text{small-signal admittance } Y_D \,=\, G_D + i\omega C_D} \cdot \tilde{v}$$

- The admittance can be separated into:
  - <u>Diffusion conductance</u> $G_D = \frac{1}{\sqrt{2}} G_0 \left(\sqrt{1+\omega^2\tau_p^2}+1\right)^{1/2}$
  - <u>Diffusion capacitance</u> $C_D = \frac{1}{\sqrt{2}} G_0 \cdot \frac{1}{\omega}\left(\sqrt{1+\omega^2\tau_p^2}-1\right)^{1/2}$
- Since $G_0 \sim e^{qV/kT}$, the diffusion capacitance becomes very large as the forward bias is increased
- Frequency dependence:
  - Low frequency $\omega \ll \frac{1}{\tau_p}$: $C_D \simeq G_0 \frac{\tau_p}{2} = \frac{I\tau_p}{2kT/q}$
  - Beyond this frequency, C_D decreases with increasing ω

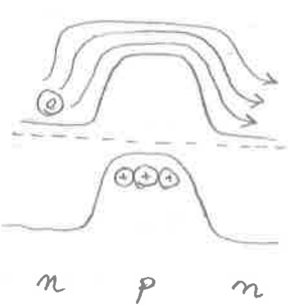
- <u>Implications of the diffusion capacitance</u>

How much signal voltage is produced by the absorption of N photons?

- Photodetector (reverse biased):

$$qN = C_J\Delta V \Rightarrow \Delta V = \frac{qN}{C_J}$$

- Phototransistor (forward biased):



- In a phototransistor (npn) the absorbed photons generate holes in the base. The holes attract electrons, and for each hole stored there many electrons can flow across the base without being lost; the ratio is the same as that of photoconductive gain, τ_p/τ_{tr}

$$\Rightarrow qN = C_J\Delta V + \underbrace{\Delta I\cdot\tau_{tr}}_{\substack{\#\, e^- \text{ that flow}\\ \text{across junction, each}\\ \text{taking 1 transit time}}}$$

= # holes stored in the base

$$= C_J\Delta V + \frac{I}{kT/q}\Delta V\,\tau_{tr}$$

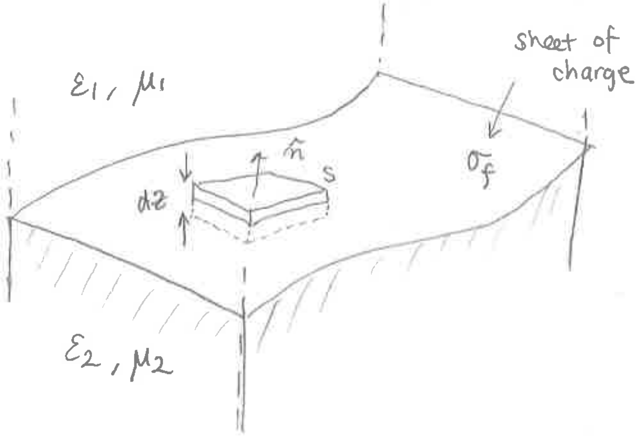
$$qN = \left(C_J + \frac{I\tau_{tr}}{kT/q}\right)\Delta V \Rightarrow \Delta V = \frac{qN}{C_D + C_J}$$

⇒ because of diffusion capacitance, more photons are needed to produce a voltage ΔV; less sensitive!

In the BJT base, $\tau_p \to \tau_{tr}$ in calculating C_D. This is because the base side of the pn junction is very narrow.

⑬ Electromagnetics and optics : selected concepts

- Electromagnetic boundary conditions



Apply Gauss's law to the pill box: the sides contribute nothing to flux since the boundary is infinitely thin

$$\int_S \vec{D} \cdot d\vec{A} = Q_{f,enc}$$

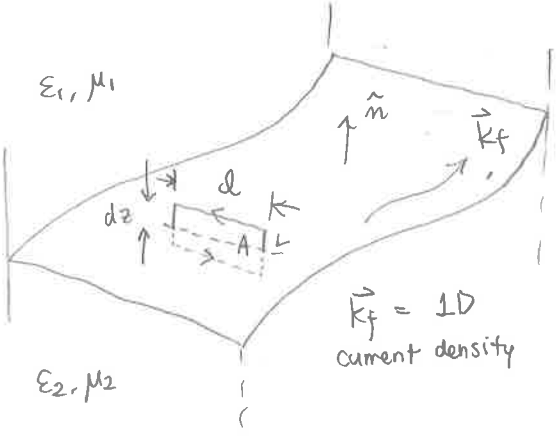
$$D_1^{\perp} A - D_2^{\perp} A = \sigma_f A$$

$$\boxed{\varepsilon_1 E_1^{\perp} - \varepsilon_2 E_2^{\perp} = \sigma_f}$$

- Use the same reasoning for magnetic field

$$\vec{\nabla} \cdot \vec{B} = 0 \Rightarrow \int_S \vec{B} \cdot d\vec{A} = 0$$

$$B_1^{\perp} A - B_2^{\perp} A = 0 \Rightarrow \boxed{B_1^{\perp} = B_2^{\perp}}$$



$\vec{K}_f$ = 1D current density

- Now consider an amperian loop along the surface:

$$\vec{\nabla} \times \vec{E} = - \frac{\partial \vec{B}}{\partial t}$$

$$\oint_L \vec{E} \cdot d\vec{l} = - \frac{d}{dt} \int_S \vec{B} \cdot d\vec{A}$$

$$\vec{E}_1 \cdot \vec{l} - \vec{E}_2 \cdot \vec{l} = 0 \leftarrow$$ flux vanishes as $dz \to 0$

$$\Rightarrow \boxed{\vec{E}_1^{\parallel} = \vec{E}_2^{\parallel}}$$

- Finally, use Ampere's law:

$$\vec{\nabla} \times \left(\frac{\vec{B}}{\mu}\right) = \vec{J}_f + \varepsilon \frac{\partial \vec{E}}{\partial t} \Rightarrow \oint \vec{H} \cdot d\vec{l} = I_f + \frac{d}{dt} \int_S \varepsilon \vec{E} \cdot d\vec{A} = I_f$$

$$\Rightarrow \vec{H}_1 \cdot \vec{l} - \vec{H}_2 \cdot \vec{l} = I_f = \vec{K}_f \cdot (\hat{n} \times \vec{l})$$

points in the direction normal to the amperian loop

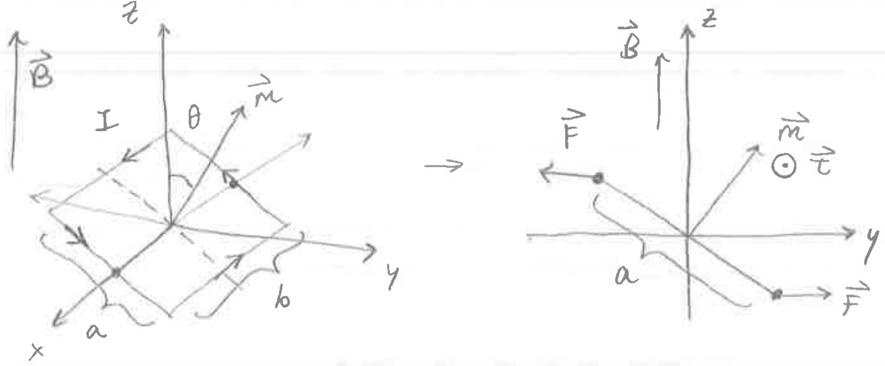
$$\vec{H}_1^{\parallel} - \vec{H}_2^{\parallel} = \vec{K}_f \times \vec{n}$$

$$\boxed{\frac{1}{\mu_1} \vec{B}_1^{\parallel} - \frac{1}{\mu_2} \vec{B}_2 = \vec{K}_f \times \vec{n}}$$

$\vec{K}_f \times \hat{n}$ is the free surface current density in the plane of the boundary. e.g. if $\hat{n} = \hat{z}$, $\vec{K}_f = K_{fy} \hat{x} - K_{fx} \hat{y}$

- Torques and forces on magnetic dipoles

Consider a dipole to be this current loop:



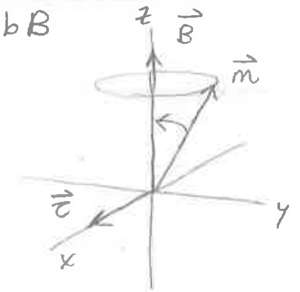
- Recall: $\vec{F} = q\vec{v} \times \vec{B}$

There is no net torque on the sides of length a, because the forces cancel. On the other sides with length b, the torque is:

$$\vec{\tau} = \vec{r} \times \vec{F}$$

$$\vec{\tau} = aF\sin\theta\,\hat{x}$$

$$F = IbB$$



$$\Rightarrow \vec{\tau} - abIB\sin\theta\,\hat{x} = mB\sin\theta\,\hat{x}$$

$$\Rightarrow \boxed{\vec{\tau} = \vec{m} \times \vec{B}} \rightarrow \text{preecession}$$

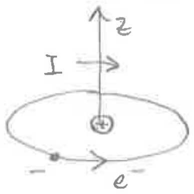
The magnetic field acts to rotate the magnetic dipole, until it aligns with the field

- The potential energy of the dipole in a magnetic field is given by: $\boxed{U = -\vec{m}\cdot\vec{B}}$

Thus, in order to minimize U, the preferred configuration is to have $\vec{m}$ parallel with $\vec{B}$. A precession dipole tends to dampen and relax to this state

- Effect on atomic orbits → Diamagnetism

- Imagine that in an atom, an electron orbits the nucleus at a radius R and completes a revolution in time $T = 2\pi R/v$



$$I = \frac{-q}{T} = \frac{-qv}{2\pi R} \Rightarrow \vec{m} = I\pi R^2$$

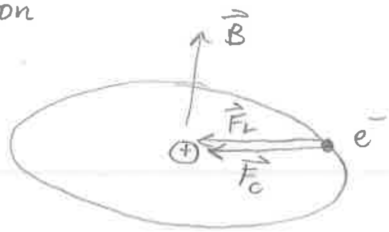
$$\vec{m} = -\tfrac{1}{2} q v R\,\hat{z}$$

I is opposite the motion of electrons

- If we introduce a magnetic field, we need to add a Lorentz force to the classical centripetal force of the orbiting electron

$$m_e \frac{v_B^2}{R} = \frac{1}{4\pi\varepsilon_0}\frac{q^2}{R^2} + q v_B B$$

centripetal force — Coulomb component — Lorentz component



where v_B = velocity w/ magnetic field

This leads to:

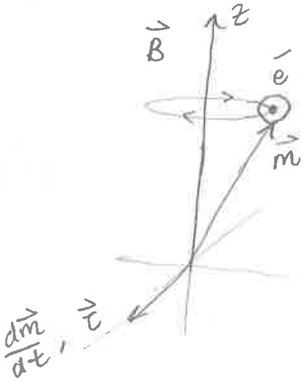
$$v_B = \frac{1}{qB}\left(m_e \frac{v_B^2}{R} - \frac{1}{4\pi\varepsilon_0}\frac{q^2}{R^2}\right) = \frac{m_e}{qB}\left(v_B^2 - v^2\right) \approx \frac{m_e}{qB} 2v\Delta v$$

$$\Rightarrow \Delta v = \frac{1}{2}\frac{qBR}{m_e}$$ (change in electron speed)

$$\Delta\vec{m} = -\frac{1}{2} q \Delta v R \hat{z} = -\frac{q^2 R^2}{4 m_e}\vec{B}$$

- If a magnetic field is applied, the magnetic dipole moment changes in the direction opposite the direction of the field. This is ultimately because the magnetic field speeds up the electrons, but the electrons' orbital motion has a negative dipole moment
- When $\vec{B} = 0$, $\vec{M} = 0$ in a material since the atoms are randomly oriented. But if $\vec{B} > 0$, $\vec{M} < 0$ due to this effect: this is diamagnetism

– Paramagnetism



- If we ignore atomic orbital motion, the electron is still a magnetic dipole due to its intrinsic orbital momentum, or spin:

$$\vec{m} = \gamma \vec{J}$$

electron magnetic moment

electron spin angular momentum

gyromagnetic ratio, $\gamma > 0$

- Thus, we still have the torque $\vec{\tau} = \vec{m} \times \vec{B}$

But note that the electron also has angular momentum $\vec{J}$, so there will be precession of the magnetic dipole:

$$\vec{\tau} = \frac{d\vec{J}}{dt} = \vec{m}\times\vec{B} \quad \Rightarrow \quad \frac{d\vec{m}}{dt} = \gamma \vec{m}\times\vec{B}$$

If we now include energy relaxation, which tends to minimize $U = -\vec{m}\cdot\vec{B}$, and phase relaxation, which reduces M_x and M_y:

$$\boxed{\frac{d\vec{m}}{dt} = \underbrace{\gamma\vec{m}\times\vec{B}}_{\text{Precession}} - \underbrace{\frac{m_z - m_0}{T_1}}_{\text{Energy relaxation}} - \underbrace{\frac{m_x + m_y}{T_2}}_{\text{Phase relaxation}}}$$ ← Bloch equation

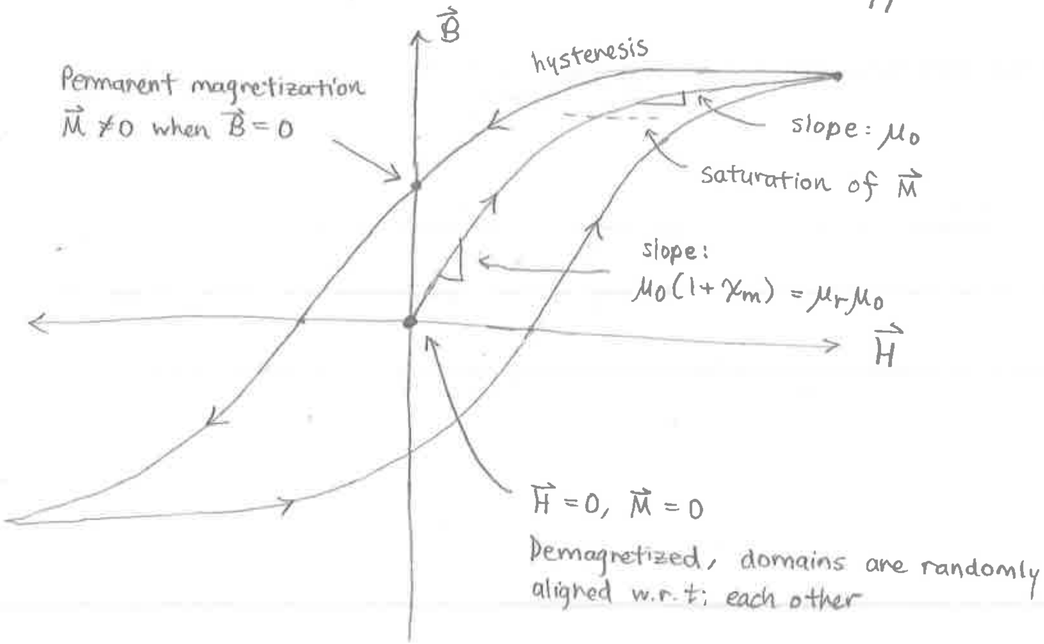
γB = precession frequency where $\vec{B} = B\hat{z}$

$m_z = m_0$ in steady state

- Note that $\vec{m}$ tends to align with $\vec{B}$, as expected classically. However, due to the Pauli exclusion principle, the electrons in atoms tend to come in pairs with opposite spin ⇒ no net torque!
- This effect is therefore only seen in some metals which have unpaired outer shell electrons, and in these materials is stronger than diamagnetism
  - As $\vec{B} = 0$, $\vec{M} = 0$, again due to random orientation of electrons
  - If $\vec{B} > 0$, $\vec{M} > 0$ ⇒ paramagnetism
    Remains weak, as random collisions disrupt the alignment
- Protons in the nucleus are also paramagnetic: this is the basis for nuclear magnetic resonance and magnetic resonance imaging

- Ferromagnetism
  - Exchange interaction: two unpaired valence electrons in adjacent atoms repel each other if they have the same spin, due to the Pauli exclusion principle
    - The reduced wavefunction overlap gives the parallel-spin state less electrostatic potential energy

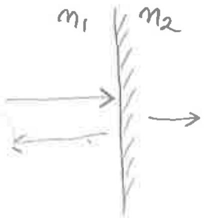
      ⇒ long-range alignment of magnetic dipole moments
  - In ferromagnets, the spins of all the unpaired electron within a domain (which can have macroscopic size) are all aligned
    - The dipoles within a domain respond in unison to a field $\vec{B}$, and can retain their magnetization after $\vec{B}$ switches off



Recall

$$\vec{B} = \mu_0\vec{H} + \mu_0\vec{M}$$
$$= (\mu_0 + \mu_0\chi_m)\vec{H}$$
$$= \mu_0\underbrace{(1+\chi_m)}_{\mu_r}\vec{H}$$

- Phase changes on reflection



- The normal incidence amplitude reflectivity is polarization-independent:

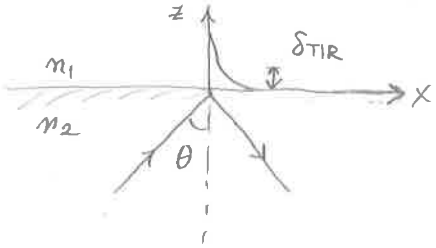
$$r = \frac{n_1 - n_2}{n_1 + n_2} \longrightarrow$$

π phase shift if $n_1 < n_2$ (low → high)

0 phase shift if $n_1 > n_2$ (high → low)

π phase shift for reflection from perfect conductor

- Total internal reflection



  - From Snell's law: $n_1 \sin\theta_1 = n_2 \sin\theta_2$

    Let the wave be incident from $n_2 > n_1$ at $\theta_2 = \theta$

$$\sin\theta_1 = \frac{n_2}{n_1}\sin\theta$$

At $\theta = \theta_c$, $\frac{n_2}{n_1}\sin\theta_c = 1$ so $\theta_1 = \frac{\pi}{2}$. If $\theta > \theta_c$, θ_1 is no longer a real number! But let's proceed to examine the wave in n_1

$$E_1 = e^{i(k_0 n_1 \sin\theta_1)x}\, e^{i(k_0 n_1 \cos\theta_1)z}$$

$$= e^{ik_0 n_2 \sin\theta \cdot x}\, e^{ik_0 n_1 \cos\theta_1 \cdot z}$$

$$\text{where} \quad \cos\theta_1 = \cos\left(\sin^{-1}\left(\frac{n_2}{n_1}\sin\theta\right)\right) = \sqrt{1 - \left(\overbrace{\frac{n_2}{n_1}\sin\theta}^{>1}\right)^2}$$

$$\cos\theta_1 = i\sqrt{\left(\frac{n_2}{n_1}\sin\theta\right)^2 - 1}$$

$$\Rightarrow \quad E_1 = e^{ik_0 n_2 \sin\theta x}\, e^{-k_0 z\sqrt{\left(\frac{n_2}{n_1}\sin\theta\right)^2 - 1}\, \cdot n_1}$$

The penetration depth is: $e^{-n_1 k_0 \delta\sqrt{\left(\frac{n_2}{n_1}\sin\theta\right)^2 - 1}} = e^{-1}$

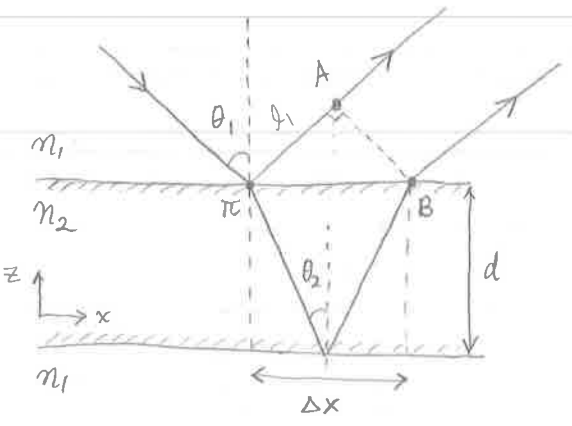
$$\delta_{TIR} = \frac{\lambda}{2\pi n_1\sqrt{\left(\frac{n_2}{n_1}\sin\theta\right)^2 - 1}} = \frac{\lambda}{2\pi\sqrt{n_2^2 \sin^2\theta - n_1^2}}$$

  - There is a phase shift on total internal reflection that is a continuous function of angle; the amount of the phase shift is tedious to calculate

    - Phase shift is 0 at $\theta = \theta_c$
    - Phase shift increases as θ increases beyond θ_c
    - Phase shift is π at $\theta = \frac{\pi}{2}$

    The phase shift might be interpreted as the penetration of the wavefront (NOT the energy) into the other medium before being reflected back.

- Thin film reflection



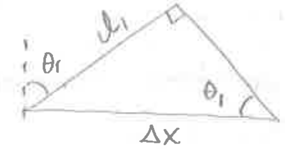
$n_2 > n_1$: this is important!

- In order for the two reflected waves to interfere constructively, their phase must match (to a factor of 2π) at points A and B
- Phase accumulated by reflected ray 1

$$\Delta x = 2d \tan \theta_2$$

where $\sin \theta_2 = \frac{n_1}{n_2} \sin \theta_1$

then:



$$\sin \theta_1 = \frac{\ell_1}{\Delta x}$$

$$\ell_1 = \Delta x \sin \theta_1$$

$$\ell_1 = 2d \tan \theta_2 \sin \theta_1$$

$$\phi_1 = k_0 n_1 \ell_1 + \pi$$

$$= k_0 \cdot 2d \tan \theta_2 \cdot n_1 \sin \theta_1 \underset{\text{from top surface reflection, since } n_1 < n_2}{+ \pi}$$

- Phase accumulated by reflected rate 2

$$\phi_2 = k_0 n_2 \cdot 2 \frac{d}{\cos \theta_2} = k_0 \frac{2dn_2}{\cos \theta_2}$$

- The phase difference is:

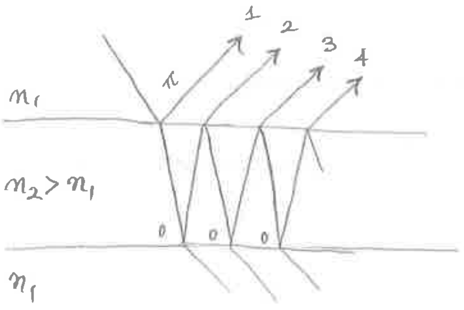
$$\Delta\phi = \phi_2 - \phi_1 = k_0 \frac{2dn_2}{\cos \theta_2} - k_0 2d \tan \theta_2 n_1 \sin \theta_1 - \pi$$

$$= 2k_0 d \left(\frac{n_2}{\cos \theta_2} - n_1 \tan \theta_2 \sin \theta_1 \right) - \pi$$

$$= 2k_0 d n_2 \left(\frac{1}{\cos \theta_2} - \tan \theta_2 \sin \theta_2 \right) - \pi$$

$$= 2n_2 k_0 d \left(\frac{1 - \sin^2 \theta_2}{\cos \theta_2} \right) - \pi \Rightarrow \Delta\phi = k_0 \, (\overbrace{2n_2 d \cos \theta_2}^{\Delta OPL}) \underset{\text{phase shift of first reflected ray}}{- \pi}$$

- Subsequent reflected rays have $\Delta\phi = k_0 (2n_2 d \cos \theta_2)$ since all of the internal reflections induce no phase shift in this geometry, and transmitted rays always have no phase shift at the boundary



  - Therefore, unless the initially reflected ray is very strong (unlikely), the film's reflectivity depends on whether rays 2, 3, 4, ... interfere constructively:

$$\Rightarrow \Delta\phi = k_0 2n_2 d \cos \theta_2 = 2\pi m$$

$$\lambda_m = \frac{1}{m} 2n_2 d \cos \theta_2$$

Reflectivity maxima: $\boxed{\frac{\lambda_m}{n_2} = \frac{2d}{m} \cos \theta_2} \approx \frac{2d}{m}$ near normal

- Reflectivity is maximized for $d = \frac{1}{2}(\lambda / n_2)$ (half wavelength)

– Metal optics

- Conductivity and the complex permittivity

Ampere's law → $$\vec{\nabla} \times \vec{H} = \vec{J} + \varepsilon \frac{\partial \vec{E}}{\partial t}$$ where $\varepsilon = \varepsilon_r \varepsilon_0$, ε_r is real

$$= \sigma \vec{E} - i\omega\varepsilon \vec{E}$$ sinusoidal field $\vec{E} \sim e^{-i\omega t}$

$$= \left(\varepsilon + \frac{\sigma}{-i\omega}\right)(-i\omega \vec{E})$$

$$= \left(\varepsilon_r \varepsilon_0 + i\frac{\sigma}{\omega}\right)\frac{\partial \vec{E}}{\partial t} = \tilde{\varepsilon}\frac{\partial \vec{E}}{\partial t}$$

where the complex permittivity is $$\boxed{\tilde{\varepsilon} = \varepsilon_r \varepsilon_0 + i\frac{\sigma}{\omega}}$$

In metals, imaginary permittivity = conductivity

- Skin depth

Consider a wave traveling through a metal $\tilde{\varepsilon}$, which varies as $e^{i\tilde{k}z}$:

$$\tilde{k} = k_0 \tilde{n} = \frac{\omega}{c}\sqrt{\frac{\tilde{\varepsilon}}{\varepsilon_0}\cdot\frac{\tilde{\mu}}{\mu_0}} = \frac{\omega}{c}\sqrt{\varepsilon_r\mu_r + i\frac{\sigma\mu_r}{\omega\varepsilon_0}}$$

where $\tilde{\mu} = \mu_r\mu_0$ is real

$$\Rightarrow \operatorname{Im}\tilde{k} = \frac{\omega}{c}\sqrt{\frac{1}{2}\left[\sqrt{(\varepsilon_r\mu_r)^2 + \left(\frac{\sigma\mu_r}{\omega\varepsilon_0}\right)^2} - \varepsilon_r\mu_r\right]}$$

$$= \frac{\omega}{c}\frac{1}{\sqrt{2}}\left[\varepsilon_r\mu_r\sqrt{1 + \left(\frac{\sigma}{\omega\varepsilon_r\varepsilon_0}\right)^2} - \varepsilon_r\mu_r\right]^{1/2}$$

$$= \frac{\omega}{c}\sqrt{\frac{\varepsilon_r\mu_r}{2}}\left[\sqrt{1 + \left(\frac{\sigma}{\omega\varepsilon_r\varepsilon_0}\right)^2} - 1\right]^{1/2}$$

The skin depth is defined where $e^{-\operatorname{Im}\tilde{k}\cdot\delta} = e^{-1}$

$$\Rightarrow \delta = \frac{1}{\operatorname{Im}\tilde{k}} = \sqrt{\frac{2}{\varepsilon_r\mu_r}}\,\frac{c}{\omega}\left[\sqrt{1 + \left(\frac{\sigma}{\omega\varepsilon_r\varepsilon_0}\right)^2} - 1\right]^{-1/2}$$

$$= \sqrt{\frac{2}{\omega\mu_r\mu_0}}\,\frac{1}{\sqrt{\varepsilon_r\varepsilon_0\omega}}\cdot\left[\frac{\sigma}{\omega\varepsilon_r\varepsilon_0}\sqrt{1 + \left(\frac{\omega\varepsilon_r\varepsilon_0}{\sigma}\right)^2} - 1\right]^{-1/2}$$

$$= \sqrt{\frac{2}{\omega\sigma\mu_r\mu_0}}\left[\sqrt{1 + \left(\frac{\omega\varepsilon_r\varepsilon_0}{\sigma}\right)^2} - \frac{\omega\varepsilon_r\varepsilon_0}{\sigma}\right]^{-1/2}$$

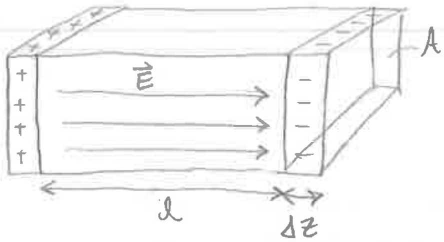
$$\boxed{\delta = \sqrt{\frac{2}{\omega\sigma\mu_r\mu_0}}\left[\sqrt{1 + \left(\frac{\omega\varepsilon_r\varepsilon_0}{\sigma}\right)^2} + \frac{\omega\varepsilon_r\varepsilon_0}{\sigma}\right]^{1/2}}$$ general formula for skin depth

low frequency skin depth

↳ At high frequencies or poor conductivity $\sigma \ll \omega\varepsilon_r\varepsilon_0$,

$$\delta \to \frac{2}{\sigma}\sqrt{\frac{\varepsilon_r\varepsilon_0}{\mu_r\mu_0}}$$

- Bulk plasmons : oscillations in the free electron gas inside a metal's volume
  - Suppose that a metal becomes polarized



- The charge separation creates an electric field:

$$\varepsilon_0 E \cdot A = -qnA\,\Delta z \quad \leftarrow$$

Gauss's law around box of negative charge

$$E = -\frac{qn}{\varepsilon_0}\Delta z$$

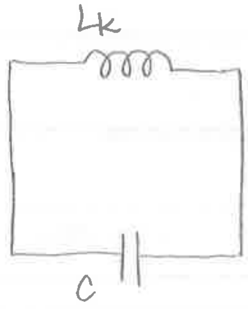
- What is the frequency of the oscillation?

$$qE = -\frac{q^2 n}{\varepsilon_0}\Delta z = m(\ddot{\Delta z}) \qquad \text{by } F = ma$$

$$\frac{nq^2}{m\varepsilon_0}\Delta z = \omega^2 \Delta z$$

$$\Rightarrow \quad \omega = \boxed{\omega_p = \sqrt{\frac{nq^2}{m\varepsilon_0}}}$$

bulk plasma frequency



- Consider the oscillation to that of an LC circuit

$$\omega = \frac{1}{\sqrt{L_K C}} \Rightarrow \frac{1}{\omega_p^2} = L_K C = L_K \frac{\varepsilon_0 A}{l}$$

$$\Rightarrow \quad L_K = \frac{l}{\varepsilon_0 A}\frac{1}{\omega_p^2} = \frac{l}{\cancel{\varepsilon_0} A}\frac{m\cancel{\varepsilon_0}}{nq^2}$$

Kinetic inductance
$$\boxed{L_K = \frac{m}{nq^2}\cdot\frac{l}{A}}$$

This is the effective inductance associated with the inertia of electrons: at very high frequencies, the response of the metal is delayed because the electrons need to move

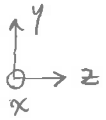
- This can also be derived from the AC conductivity of a metal, from the Drude model
  ↓

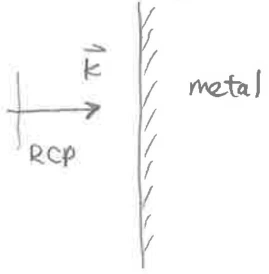
$$\sigma = \frac{nq^2\tau}{m}\left(\frac{1}{1+i\omega\tau}\right)$$

The impedance of a metal wire is

$$Z = \frac{1}{\sigma}\frac{l}{A} = \frac{m}{nq^2\tau}(1+i\omega\tau)\frac{l}{A} = \overbrace{\frac{m}{nq^2\tau}\frac{l}{A}}^{\text{ohmic resistance } R} + i\omega\overbrace{\frac{m}{nq^2}\frac{l}{A}}^{\text{kinetic inductance } L_K}$$

Reflection of circularly polarized light from a metal surface





- Consider the RCP incident light below:

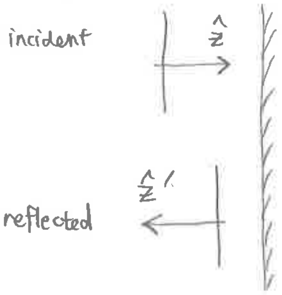
$$\vec{E}_I(z,t) = \frac{E_0}{\sqrt{2}}\cos(kz-\omega t)\,\hat{x} + \frac{E_0}{\sqrt{2}}\sin(kz-\omega t)\,\hat{y}$$

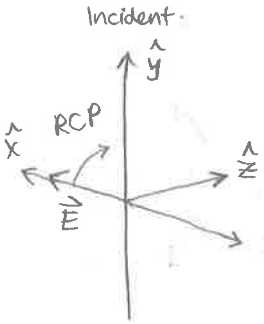
- The electric field must vanish at the surface of a perfect conductor, so $\vec{E}_R = -\vec{E}_I$:

$$\vec{E}_R(z,t) = -\frac{E_0}{\sqrt{2}}\cos(-kz-\omega t)\,\hat{x} - \frac{E_0}{\sqrt{2}}\sin(-kz-\omega t)\,\hat{y}$$

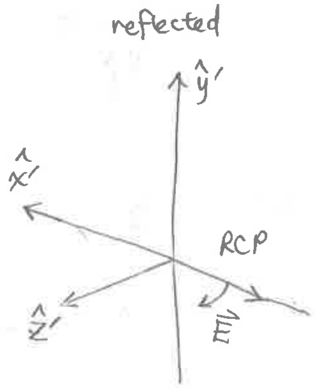
This ensures zero field at $z=0$ while $\vec{E}_R$ propagates along $-\hat{z}$.

- Now the subtle part: we need to do a coordinate transformation:



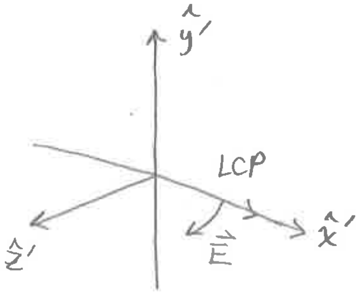


Incident coordinate system is right-handed
$\hat{x}\times\hat{y} = \hat{z}$



① First, flip propagation direction so $\hat{z}' = -\hat{z}$ while $\hat{x}' = \hat{x}$, $\hat{y}' = \hat{y}$
But now $\hat{x}'\times\hat{y}' = -\hat{z}'$.

② Next, flip the x-axis so
$\hat{x}' = -\hat{x}$, $\hat{y}' = \hat{y}$, $\hat{z}' = -\hat{z}$
Now we have $\hat{x}'\times\hat{y}' = \hat{z}'$



- With this transformation, the reflected wave is:

$$\vec{E}_R = \frac{E_0}{\sqrt{2}}\cos(kz'-\omega t)\,\hat{x}' - \frac{E_0}{\sqrt{2}}\sin(kz'-\omega t)\,\hat{y}'$$

This describes a left circularly polarized wave, so there is a handedness reversal of circularly polarized light when reflected from a conductor