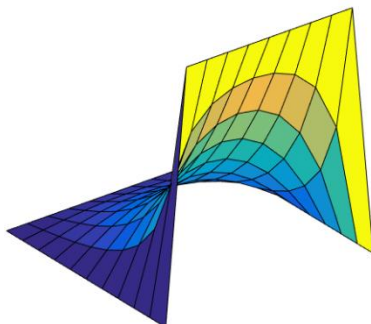


UNIT 1: DESIGN OF HEAT FINS: HEAT CONDUCTION, FOURIER SERIES, AND FINITE DIFFERENCE APPROXIMATION

Heat conduction is a wonderland for mathematical analysis, numerical computation, and experiment. It's also highly practical: engineers have to make sure engines don't melt and computer chips don't overheat.



Lecture 1: Fourier's law

(Lienhard and Lienhard pp. 10-17, 19-20)

Heat is the **flow of thermal energy** from a **warmer place** to a **cooler place**. There are three kinds of heat transfer:

- **Conduction:** Heat flow through motionless materials.
- **Convection:** Heat flow through moving fluids.
- **Radiation:** Heat flow through electromagnetic waves.

This unit is primarily concerned with heat conduction. We will give attention to convection only because **convective heat flow at the surface** of a solid affects the **conductive heat flow within** the solid.

We use the symbol **Q** to represent **heat transfer rate**. Q has units of watts.

What happens when heat flows into or out of a mass? The temperature T changes:

$$Q_{\text{in}} - Q_{\text{out}} = mc \, dT/dt,$$

where m is mass and c is called "**specific heat**." This formula is the time derivative of a calorimetry equation you may have used in chemistry (energy transferred = mcΔT).

If Q is the rate at which heat is flowing through a solid with cross-sectional area A, **q = Q/A** is the **heat flux**.

Fourier's law states that heat flux is proportional to thermal gradient:

$$q = -k \, dT/dx,$$

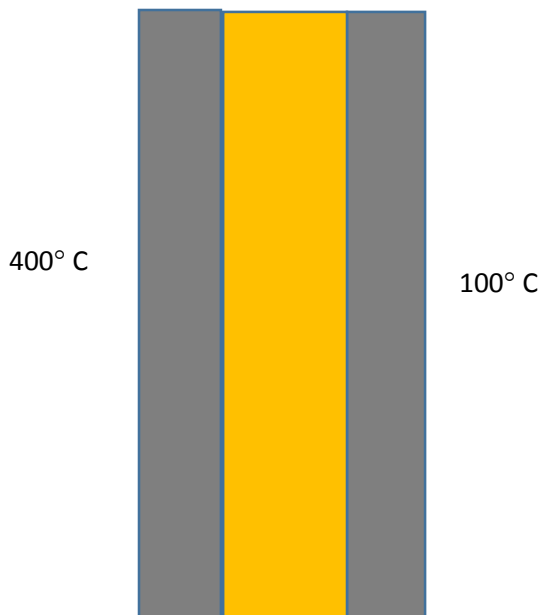
where **k** is **thermal conductivity**. Sometimes we care only about the magnitude of q, but what does the sign mean? Positive q represents heat flux in the direction of positive x. Negative q represents heat flux in the direction of negative x.

If heat flux is constant throughout a solid, then dT/dx can be replaced by ΔT/Δx:

$$|q| = k\Delta T/\Delta x = \text{constant.}$$

When is heat flux **constant**? In **one-dimensional, steady-state** heat flow. For example, if the two sides of a wall are held at two fixed temperatures, or the two ends of a laterally insulated wire are held at two fixed temperatures, then the heat flow is approximately one-dimensional and constant.

Example (from Lienhard and Lienhard)



A copper plate that is 3 mm thick is sandwiched between stainless steel plates that are each 2 mm thick. The left surface of the sandwich is held at 400° C, and the right surface is held at 100° C. The thermal conductivity of stainless steel is 17 W/(Km), and the thermal conductivity of copper is 373 W/(Km). What are the temperatures at the left and right surfaces of the copper plate?

Solution

This is a case of one-dimensional steady state heat flow, so q is constant.

$$(q \text{ through steel plate on left}) = (q \text{ through the copper plate}) = (q \text{ through steel plate on right})$$

$$k_{\text{Steel}}\Delta T_{\text{steel}}/(2 \text{ mm}) = k_{\text{Cu}}\Delta T_{\text{Cu}}/(3 \text{ mm}) = k_{\text{Steel}}\Delta T_{\text{steel}}/(2 \text{ mm})$$

(Why must ΔT_{steel} be the same for both steel plates?)

We also know that $\Delta T_{\text{steel}} + \Delta T_{\text{Cu}} + \Delta T_{\text{steel}} = 300^\circ \text{ C}$, the total temperature difference across the sandwich.

We now have two equations in two unknowns and can find $\Delta T_{\text{steel}} = 145 \text{ K}$ and $\Delta T_{\text{Cu}} = 10 \text{ K}$. From this, we see that the temperatures on the two sides of the copper plate are 255° C and 245° C.

Convective boundary conditions

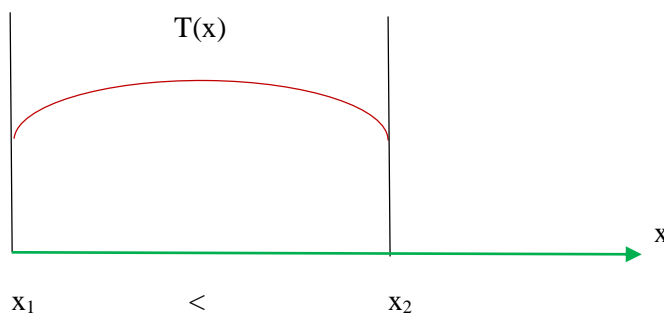
It would be nice if boundary conditions were always specified surface temperatures. Sometimes, instead, we have convection at surfaces. Convective heat flow is proportional to the difference between the surface temperature and the surrounding temperature (Newton's law of cooling):

$$q = \bar{h}(T_{surface} - T_{\infty})$$

where T_{∞} is the temperature of the surrounding fluid (possibly air), and \bar{h} is called the heat transfer coefficient. A convective boundary condition requires convective heat flow at the surface to equal the conductive heat flow at the surface:

$$\bar{h}[T_{surface} - T_{\infty}] = k \left. \frac{dT}{dx} \right|_{surface}$$

Let's try to get rid of the absolute value bars. Consider a slab between x_1 and x_2 , where $x_1 < x_2$. Suppose both surfaces are warmer than their surroundings, so heat flows out through the surfaces (heat flows from hot to cold):



We see that dT/dx is positive at x_1 and negative at x_2 . Therefore the convective boundary condition at x_1 is

$$\bar{h}[T(x_1) - T_{\infty}] = k \left. \frac{dT}{dx} \right|_{x_1}$$

and at x_2 it is (notice the minus sign!)

$$\bar{h}[T(x_2) - T_{\infty}] = -k \left. \frac{dT}{dx} \right|_{x_2}$$

These equations also work if the slab is colder than its surroundings. Why? Draw $T(x)$ and examine the slope at the surfaces. What if the temperature throughout the slab is hotter than the surrounding fluid on one side but colder than the surrounding fluid on the other side?

Lecture 2: The One-Dimensional Heat Equation

(Lienhard and Lienhard pp. 17-19, 56-61)

Now we're going to use Fourier's law to derive the one-dimensional heat equation. Please pay attention to the "tiny volume analysis" that we're about to do because we'll use this technique throughout the semester.

Consider a very thin slice of a solid with thickness δx and cross-sectional area A (perpendicular to the diagram). Heat $Q(x)$ flows in from the left, and $Q(x+\delta x)$ flows out through the right. The temperature changes according to

$$Q_{\text{in}} - Q_{\text{out}} = mc \frac{\partial T}{\partial t}$$

which is

$$Q(x) - Q(x+\delta x) = mc \frac{\partial T}{\partial t}.$$

(I'm using “ ∂ ” now because T in the tiny volume will depend on both position and time.)

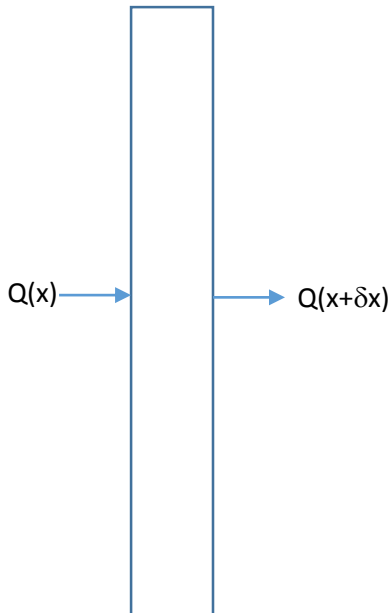
Now, using $q = Q/A$,

$$A q(x) - A q(x+\delta x) = mc \frac{\partial T}{\partial t}.$$

Next, we use $q = -k \frac{\partial T}{\partial x}$:

$$-kA \left. \frac{\partial T}{\partial x} \right|_x - \left(kA \left. \frac{\partial T}{\partial x} \right|_{x+\delta x} \right) = mc \frac{\partial T}{\partial t}$$

$$kA \left(\left. \frac{\partial T}{\partial x} \right|_{x+\delta x} - \left. \frac{\partial T}{\partial x} \right|_x \right) = mc \frac{\partial T}{\partial t}$$



Now let's write m in terms of density ρ , so $m = \rho A \delta x$:

$$kA \left(\left. \frac{\partial T}{\partial x} \right|_{x+\delta x} - \left. \frac{\partial T}{\partial x} \right|_x \right) = \rho A (\delta x) c \frac{\partial T}{\partial t}.$$

Now let's think about the approximation $f(x+dx) \approx f(x) + f'(x)dx$. This just means that any function is approximately linear if you zoom in close enough. Since $f'(x)$ is also a function, it's also true that $f'(x+dx) \approx f'(x) + f''(x)dx$. Therefore

$$\left. \frac{\partial T}{\partial x} \right|_{x+\delta x} \approx \left. \frac{\partial T}{\partial x} \right|_x + \frac{\partial^2 T}{\partial x^2} \delta x$$

Plugging this in,

$$kA \frac{\partial^2 T}{\partial x^2} \delta x = \rho A (\delta x) c \frac{\partial T}{\partial t}$$

Now we divide both sides by $A \delta x$:

$$k \frac{\partial^2 T}{\partial x^2} = \rho c \frac{\partial T}{\partial t}.$$

Finally, defining $\alpha = k/\rho c$ as the thermal diffusivity, we have the heat equation in one dimension:

$$\alpha \frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t}.$$

First, let's consider steady-state heat flow, so the time derivative is 0:

$$d^2T/dx^2 = 0$$

Example: Steady-state slab

A slab has one surface at temperature T_1 at $x = 0$, and the other surface is at temperature T_2 at $x = L$. Determine the steady-state temperature $T(x)$ throughout the slab.

Solution

We can confirm that the solution to $d^2T/dx^2 = 0$ is $T = C_1x + C_2$, where C_1 and C_2 are constants that we have to determine from the boundary conditions: $T(0) = T_1$ and $T(L) = T_2$. Doing the algebra, we obtain $T = T_1 + (T_2 - T_1)x/L$

Example: Steady-state slab with convection on one side

A slab has one surface at temperature T_1 at $x = 0$, and the other surface at $x = L$ is subject to convective heat transfer with the surroundings at temperature T_∞ . Determine the steady-state temperature $T(x)$ throughout the slab.

Solution

As before, the solution to the steady-state heat equation is $T = C_1x + C_2$, so we only have to determine the constants. The boundary conditions are $T(0) = T_1$ and $\bar{h} [T(L) - T_\infty] = -k dT/dx$ (at $x = L$), where $T(L) = C_1L + C_2$, and $dT/dx = C_1$. Doing the algebra results in $T = T_1 + (T_\infty - T_1) \bar{h} x / (k + \bar{h}L)$.

Internal energy generation

We're going to add one term to the heat equation to account for internal energy generation. For example, the hardening of concrete is exothermic: thermal energy is generated through the substance. Internal energy generation is represented by \dot{q} , which has units of W/m^3 : energy generated per unit volume per unit time.

To derive a more general form of the heat equation, again we'll begin with

$$Q_{in} - Q_{out} = mc \partial T / \partial t$$

which now is

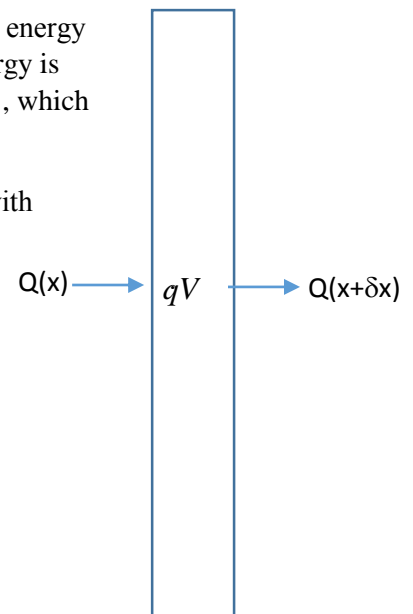
$$Q(x) + \dot{q}V - Q(x+\delta x) = mc \partial T / \partial t.$$

Now, using $q = Q/A$ and $V = A\delta x$,

$$A q(x) + \dot{q} A \delta x - A q(x+\delta x) = mc \partial T / \partial t.$$

Next, we use $q = -k \partial T / \partial x$:

$$-kA \left. \frac{\partial T}{\partial x} \right|_x + \dot{q} A \delta x - \left(kA \left. \frac{\partial T}{\partial x} \right|_{x+\delta x} \right) = mc \frac{\partial T}{\partial t}$$



$$kA \left(\left. \frac{\partial T}{\partial x} \right|_{x+\delta x} - \left. \frac{\partial T}{\partial x} \right|_x \right) + \dot{q}A \delta x = mc \frac{\partial T}{\partial t}$$

Just as before, we use $\left. \frac{\partial T}{\partial x} \right|_{x+\delta x} \approx \left. \frac{\partial T}{\partial x} \right|_x + \frac{\partial^2 T}{\partial x^2} \delta x$:

$$kA \frac{\partial^2 T}{\partial x^2} \delta x + \dot{q}A \delta x = mc \frac{\partial T}{\partial t}$$

And as before, we write $m = \rho A \delta x$ and divide by $A \delta x$:

$$k \frac{\partial^2 T}{\partial x^2} + \dot{q} = \rho c \frac{\partial T}{\partial t}$$

So we arrive at

$$\frac{1}{\alpha} \frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} + \frac{\dot{q}}{k}$$

Example: An exothermic slab

A slab of concrete is hardening such that its internal energy generation is \dot{q} . One side of the wall is at $x = 0$, and the other is at $x = L$. Both surfaces of the wall are held at T_∞ . Determine $T(x)$.

Solution

In steady state, the time derivative is 0, so

$$\frac{d^2 T}{dx^2} = -\frac{\dot{q}}{k}$$

If we integrate twice, we find the general solution

$$T = (-\dot{q}/2k)x^2 + C_1 x + C_2$$

To determine C_1 and C_2 , we use the boundary conditions, $T(0) = T(L) = T_\infty$. Plugging the boundary conditions into the general solution yields:

$$T(0) = T_\infty: C_2 = T_\infty$$

$$T(L) = T_\infty: (-\dot{q}/2k)L^2 + C_1 L + C_2 = T_\infty$$

Solving for C_1 : $C_1 = \dot{q}L/2k$. So the solution is

$$T = (-\dot{q}/2k)x^2 + \dot{q}Lx/2k + T_\infty$$

The temperature distribution is parabolic with a peak at the center of the slab. Why doesn't the temperature change with time? Why isn't temperature increasing since an exothermic process is occurring throughout?

Lecture 3: The lumped-capacity (uniform temperature) approximation

(Lienhard and Lienhard pp. 22-26)

If an object is small and conductive, its temperature may be approximately uniform, even if it's heating up or cooling off. Suppose a hot copper bead at temperature T is cooling off in air at temperature T_∞ . We begin with

$$Q_{\text{in}} - Q_{\text{out}} = mc \, dT/dt.$$

$Q_{\text{in}} = 0$, and Q_{out} , the convective heat loss to the air, is

$$Q_{\text{out}} = qA = \bar{h}A(T - T_\infty).$$

So,

$$-\bar{h}A(T - T_\infty) = mc \frac{dT}{dt}.$$

Defining $\theta = T - T_\infty$,

$$-\bar{h}A\theta = mc \frac{d\theta}{dt}.$$

Now we can separate variables: put dt on one side, and everything with θ in it on the other:

$$\frac{-\bar{h}A}{mc} dt = \frac{d\theta}{\theta}$$

Integrating both sides gives

$$\frac{-\bar{h}A}{mc} t + C = \ln \theta,$$

where C is a constant of integration. Solving for θ :

$$\theta = C' \exp\left(\frac{-\bar{h}A}{mc} t\right),$$

where $C' = e^C$ is a new constant defined for convenience. Replacing θ with $T - T_\infty$,

$$T = C' \exp\left(\frac{-\bar{h}A}{mc} t\right) + T_\infty.$$

If initial temperature $T(0) = T_i$, then we find $C' = T_i - T_\infty$, so

$$T = (T_i - T_\infty) \exp\left(\frac{-\bar{h}A}{mc} t\right) + T_\infty.$$

Now we'd like to know when this formula is valid. Our assumption was that T was uniform throughout the object. When is this approximately true? What's really happening is that **heat is flowing convectively out of the object at the surfaces**, as **heat is flowing conductively to the surfaces from the interior** of the object. Let's suppose that the temperature in the center of the object is ΔT larger than T at the surface. Then the conductive heat flow is approximately

$$q \approx k\Delta T/L,$$

where L is a "characteristic length" of the object (its radius, if it's a sphere).

Conductive heat flow to the surfaces equals convective flow from the surfaces:

$$q = \bar{h}(T - T_\infty).$$

Setting conductive and convective heat flow equal, and solving for ΔT :

$$\Delta T = \frac{L\bar{h}}{k}(T - T_\infty).$$

If this is small, temperature is approximately uniform. Specifically, if $\frac{L\bar{h}}{k} \ll 1$, temperature in the object is approximately uniform. $\frac{L\bar{h}}{k}$ is called the Biot number: $Bi = \frac{L\bar{h}}{k}$. If $Bi > 1$, temperature is not uniform, but as Bi increases, the surface temperature approaches T_∞ .

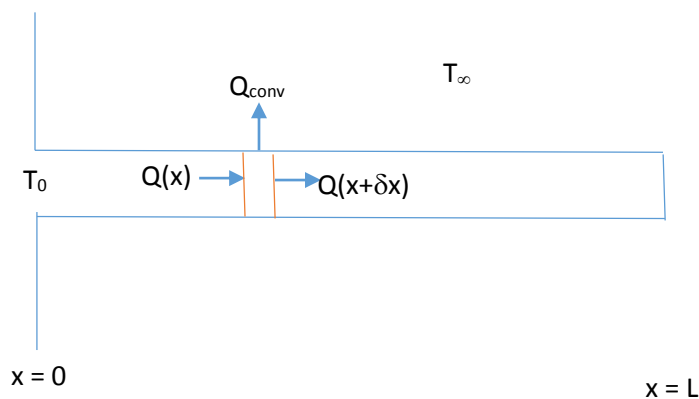
Why is it impossible for the temperature of the bead to be exactly uniform if it's not at the same temperature as its surroundings?

Lecture 4: Heat fins

(Lienhard and Lienhard pp. 163-173, 176-177)

A very practical application of heat conduction is in the design of heat fins. Heat fins are used to increase the rate of convective heat transfer, to cool off engines and electronics, for example.

To analyze a heat fin, let's assume it has constant cross-sectional area A with perimeter P . At $x = 0$, the fin is in contact with a solid at temperature T_0 . The fin has length L and is surrounded by a fluid at temperature T_∞ . Let the heat transfer coefficient of the fin be \bar{h} , except at the tip ($x = L$), where it is \bar{h}_L .



To derive a fundamental equation, we once again analyze a tiny volume with length δx . In steady state, the heat flowing in equals the heat flowing out:

$$Q(x) = Q_{\text{conv}} + Q(x+\delta x)$$

Using $Q(x) = qA$ and $Q_{\text{conv}} = \bar{h}(P\delta x)(T - T_\infty)$, where $P\delta x$ (perimeter \times length) is the external surface area of the tiny volume, we find

$$-kA \frac{dT}{dx} \Big|_x = \bar{h}(P\delta x)(T - T_\infty) - kA \frac{dT}{dx} \Big|_{x+\delta x}$$

Rearranging and dividing by δx :

$$kA \frac{\frac{dT}{dx} \Big|_{x+\delta x} - \frac{dT}{dx} \Big|_x}{\delta x} = \bar{h}P(T - T_\infty),$$

once again using $\frac{dT}{dx} \Big|_{x+\delta x} \approx \frac{dT}{dx} \Big|_x + \frac{d^2T}{dx^2} \delta x$, and solving for d^2T/dx^2 :

$$\frac{d^2T}{dx^2} = \frac{\bar{h}P(T - T_\infty)}{kA} = m^2(T - T_\infty),$$

where $m \equiv \sqrt{\frac{\bar{h}P}{kA}}$ is defined for convenience. This fin parameter can be thought of as “convectiveness” divided by “conductiveness.”

To solve the differential equation, we need the boundary conditions

$$T(0) = T_0$$

and, at $x = L$, conduction equals convection:

$$-kA \frac{dT}{dx} \Big|_L = \bar{h}_L A [T(L) - T_\infty].$$

To simplify the differential equation, we again define $\theta = T - T_\infty$, so

$$\frac{d^2\theta}{dx^2} = m^2\theta.$$

The general solution is

$$\theta = C_1 e^{mx} + C_2 e^{-mx},$$

and we must use the boundary conditions to solve for the constants.

The boundary conditions become

$$\theta(0) = T_0 - T_\infty, \text{ so } C_1 + C_2 = T_0 - T_\infty$$

and

$$-k \left. \frac{d\theta}{dx} \right|_L = \bar{h}_L \theta(L), \text{ so } -km(C_1 e^{mL} - C_2 e^{-mL}) = \bar{h}_L (C_1 e^{mL} + C_2 e^{-mL}).$$

Now we have two equations in two unknowns, C_1 and C_2 . It's straightforward but tedious to do the algebra to solve for them. Using the first boundary condition, we solve for $C_2 = T_0 - T_\infty - C_1$. Then we can plug this into the second boundary condition:

$$-km(C_1 e^{mL} - (T_0 - T_\infty - C_1)e^{-mL}) = \bar{h}_L (C_1 e^{mL} + (T_0 - T_\infty - C_1)e^{-mL})$$

Taking terms with C_1 to one side and all other terms to the other:

$$C_1 \bar{h}_L (e^{mL} - e^{-mL}) + C_1 km (e^{mL} + e^{-mL}) = (km - \bar{h}_L)(T_0 - T_\infty) e^{-mL}$$

It will be convenient to use the hyperbolic trig functions

$$\sinh x = \frac{e^x - e^{-x}}{2}, \quad \cosh x = \frac{e^x + e^{-x}}{2}, \quad \text{and} \quad \tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

Solving for C_1 ,

$$C_1 = \frac{(km - \bar{h}_L)(T_0 - T_\infty) e^{-mL}}{2\bar{h}_L \sinh(mL) + 2km \cosh(mL)}.$$

It takes some algebra (but just straightforward algebra) to find

$$C_2 = \frac{(km + \bar{h}_L)(T_0 - T_\infty) e^{mL}}{2\bar{h}_L \sinh(mL) + 2km \cosh(mL)}.$$

Now we simply plug these into the general solution:

$$\theta = \frac{(T_0 - T_\infty) \left((km - \bar{h}_L) e^{m(x-L)} + (km + \bar{h}_L) e^{m(L-x)} \right)}{2\bar{h}_L \sinh(mL) + 2km \cosh(mL)}$$

Doing some algebra puts this in a slightly nicer form:

$$\theta = \frac{(T_0 - T_\infty) \left(km \cosh[(m(L-x))] + \bar{h}_L \sinh[(m(L-x))] \right)}{\bar{h}_L \sinh(mL) + km \cosh(mL)}$$

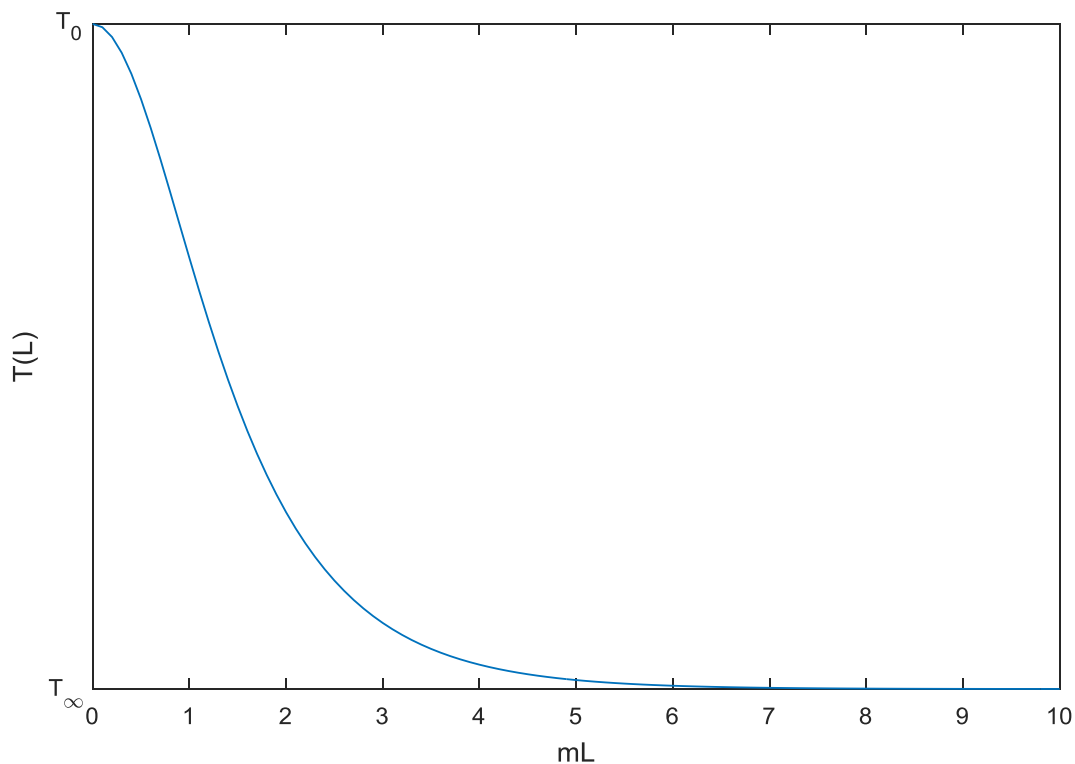
This is a very versatile equation. The fun really begins when we start investigating special cases. Let's consider an insulated tip ($\bar{h}_L = 0$):

$$\theta = \frac{(T_0 - T_\infty) \cosh[m(L - x)]}{\cosh(mL)}$$

Let's further specialize by setting $x = L$ to find $\theta(L)$, the temperature (relative to ambient) of an insulated tip:

$$\theta(L) = \frac{(T_0 - T_\infty)}{\cosh(mL)}, \text{ or } T(L) = \frac{(T_0 - T_\infty)}{\cosh(mL)} + T_\infty$$

What does this function look like? As mL gets big, $1/\cosh(mL) \approx e^{-mL}/2$. This means that as L increases, the temperature at the tip decreases exponentially.



We see that for $L > 5/m$, the temperature at the tip is approximately T_∞ .

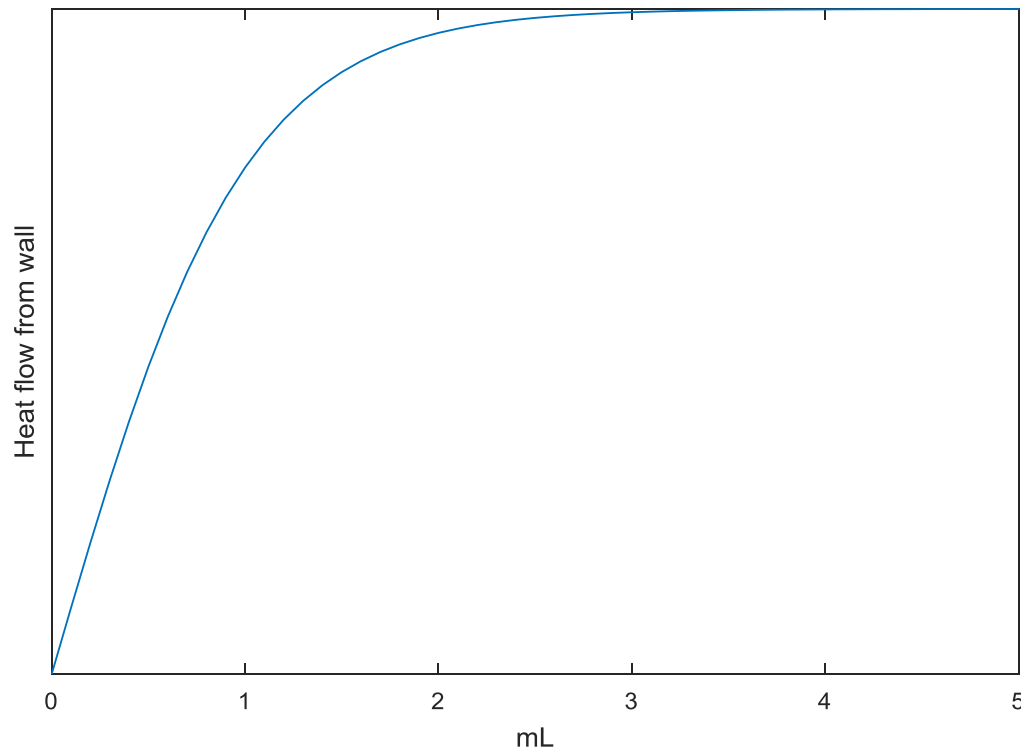
What's most important is the heat flow from the object. This is what we're trying to maximize. Heat flow from the object into the fin at $x = 0$ is

$$Q = -kA \frac{dT}{dx} = -kA \frac{d\theta}{dx}, \text{ where the derivative is evaluated at } x = 0.$$

For an insulated tip,

$$Q = \frac{kAm(T_0 - T_\infty) \sinh(mL)}{\cosh(mL)} = kAm(T_0 - T_\infty) \tanh(mL)$$

Plotting Q as a function of L is just a plot of \tanh , assuming fixed m :



We see that **increasing L beyond $3/m$ doesn't improve heat flow**. Conceptually, why is this the case?

Lecture 5: Fourier series—invented to solve the heat equation!

(Lienhard and Lienhard pp. 203-207)

We've looked at steady state heat flow (no time dependence). We also looked at the lumped-sum approximation (negligible spatial dependence). What if temperature depends on both position and time? Then we have to solve the full heat equation, which has (partial) derivatives with respect to both time and position:

$$\alpha \frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t}$$

Example: A lucky initial condition

The sides of a slab are held at 0°C : $T(0,t) = 0$ and $T(L,t) = 0$. The initial temperature distribution in the slab is $T(x,0) = T_0 \sin(\pi x/L)$. Determine the temperature distribution $T(x,t)$ at all times.

Solution

The way to solve this partial differential equation is to assume that the temperature $T(x,t)$ is a product of a function of x and a function of t : $T(x,t) = X(x)\tau(t)$. (When is this assumption true? It's true as long as there are no constant terms in the boundary conditions. If the temperature was T_0 instead of 0 at the boundaries, we'd have to define $\theta = T - T_0$ to make $\theta = 0$ at the boundaries.)

Now we plug $T(x,t) = X(x)\tau(t)$ into the heat equation:

$$\alpha\tau \frac{d^2 X}{dx^2} = X \frac{d\tau}{dt}$$

Now I'm going to divide both sides by τX to get all the x dependence on one side and all the t dependence on the other. (I'll divide by α too.)

$$\frac{1}{X} \frac{d^2 X}{dx^2} = \frac{1}{\alpha\tau} \frac{d\tau}{dt}$$

The left hand side is a function of x (which can't depend on t). The right hand side is a function of t (which can't depend on x). Then how can the two sides possibly equal each other? Only if they're both equal to the same constant! I could call the constant C or k , but for later convenience, I'll call the constant $-\lambda^2$.

$$\frac{1}{X} \frac{d^2 X}{dx^2} = \frac{1}{\alpha\tau} \frac{d\tau}{dt} = -\lambda^2$$

Now I can separately determine $X(x)$ and $\tau(t)$:

$$\frac{1}{X} \frac{d^2 X}{dx^2} = -\lambda^2 \quad \text{and} \quad \frac{1}{\alpha\tau} \frac{d\tau}{dt} = -\lambda^2$$

Let's find X first:

$$\frac{d^2 X}{dx^2} = -\lambda^2 X$$

What functions are proportional to the negative of their second derivatives? Sines and cosines! By plugging the following expression into the preceding equation, you can confirm that

$$X = A\sin(\lambda x) + B\cos(\lambda x).$$

A and B are constants that we'll determine from the boundary conditions and initial condition. If you didn't know in advance to write $A\sin(\lambda x) + B\cos(\lambda x)$, you could have written $A\sin(kx) + B\cos(kx)$ and plugged into the differential equation to determine $k = \lambda$.

Now we'll solve the equation for τ :

$$\frac{d\tau}{dt} = -\alpha\lambda^2 \tau$$

What function is proportional to its own first derivative? An exponential function:

$$\tau = \exp(-\alpha\lambda^2 t)$$

If we multiplied $\exp(-\alpha\lambda^2 t)$ by an unknown constant, it would still satisfy the differential equation. However, since we're going to multiply τ and X together, we can safely ignore the multiplicative constant for τ ; equivalently, we can pretend that A and B were already multiplied by the multiplicative constant for τ .

$$\text{So } T(x,t) = X(x)\tau(t) = [A\sin(\lambda x) + B\cos(\lambda x)]\exp(-\alpha\lambda^2 t).$$

To determine A, B, and λ , we apply the boundary conditions and initial condition.

The boundary condition on the left is $T(0,t) = 0$. If we plug this into $T(x,t)$, we find $B \exp(-\alpha\lambda^2 t) = 0$, which can only be true at all t if $B = 0$. Now we only have A and λ left to solve for:

$$T(x,t) = A\sin(\lambda x)\exp(-\alpha\lambda^2 t)$$

The boundary condition on the right is $T(L,t) = 0$. If we plug this into $T(x,t)$, we find

$$A\sin(\lambda L)\exp(-\alpha\lambda^2 t) = 0$$

We'd better not satisfy this by setting $A = 0$ because then $T(x,t) = 0$ and we can't satisfy the initial condition. So we must satisfy this boundary condition by setting $\sin(\lambda L) = 0$, so $\lambda = n\pi/L$, where n is an unknown integer. Now our expression for $T(x,t)$ is

$$T(x,t) = A\sin(n\pi x/L)\exp[-\alpha(n\pi/L)^2 t]$$

Finally, we look at the initial condition: $T(x,0) = T_0\sin(\pi x/L)$. Plugging this into $T(x,t)$ yields

$$A\sin(n\pi x/L) = T_0\sin(\pi x/L)$$

Evidently, the initial condition is satisfied if $A = T_0$ and $n = 1$. Therefore, the final solution is

$$T(x,t) = T_0\sin(\pi x/L)\exp[-\alpha(\pi/L)^2 t]$$

We were **lucky**. What if the initial condition isn't a sinusoid that equals 0 at $x = 0$ and $x = L$?

Example: Uniform initial temperature

The sides of a slab are held at 0°C : $T(0,t) = 0$ and $T(L,t) = 0$. The initial temperature distribution in the slab is uniform: $T(x,0) = T_0$. Determine the temperature distribution $T(x,t)$ at all times.

Solution

Everything except the initial condition is the same as in the previous problem. And the initial condition was the last fact used. So it's still true that we need to solve the heat equation

$$\alpha \frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t}.$$

It's still true that we define $T(x,t) = X(x)\tau(t)$ and find

$$\frac{1}{X} \frac{d^2 X}{dx^2} = \frac{1}{\alpha\tau} \frac{d\tau}{dt} = -\lambda^2.$$

It's still true that a solution is

$$T(x,t) = X(x)\tau(t) = [A\sin(\lambda x) + B\cos(\lambda x)]\exp(-\alpha\lambda^2 t).$$

It's still true that the boundary condition on the left requires $B = 0$, so

$$T(x,t) = A\sin(\lambda x)\exp(-\alpha\lambda^2 t).$$

It's still true that the boundary condition on the right requires $\lambda = n\pi/L$, so

$$T(x,t) = A \sin(n\pi x/L) \exp[-\alpha(n\pi/L)^2 t].$$

Now, there's no way this expression can satisfy the initial condition $T(x,0) = T_0$. So we must use the fact that **the sum of solutions to a linear differential equation is also a solution**. So the general solution to the differential equation is actually

$$T(x,t) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{L}\right) \exp\left[-\alpha\left(\frac{n\pi}{L}\right)^2 t\right],$$

where each term in the sum can have a different coefficient A_n . Now our job is to use the initial condition to solve for A_n for each term in the infinite sum! The initial condition $T(x,0) = T_0$ gives

$$\sum_n A_n \sin\left(\frac{n\pi x}{L}\right) = T_0.$$

This is called a "Fourier series expansion" of T_0 . In fact, any function can be written as a sum of sinusoids (a Fourier series). In this case, the function happens to be a constant.

We have a single equation with an infinite number of unknowns, A_n ($n = 1, 2, 3, \dots$)! How can we possibly find an infinite number of unknowns with a single equation? We'll use Fourier's ingenious idea: we'll multiply both sides by $\sin(m\pi x/L)$, where m is an integer, and integrate from 0 to L .

$$\sum_n A_n \int_0^L \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{m\pi x}{L}\right) dx = T_0 \int_0^L \sin\left(\frac{m\pi x}{L}\right) dx$$

This isn't really new math; this is just a **clever application of Calculus 1**. Since this isn't a calculus class, we don't need to discuss the method of solving these integrals; we can even just use an online integral calculator (wolfram alpha or integral-calculator.com). However we do the integrals, we obtain an amazing result: all the terms on the left hand side integrate to 0, except for $n = m$! So we can set $n = m$ and drop the terms that become 0:

$$A_n \int_0^L \sin^2\left(\frac{n\pi x}{L}\right) dx = T_0 \int_0^L \sin\left(\frac{n\pi x}{L}\right) dx$$

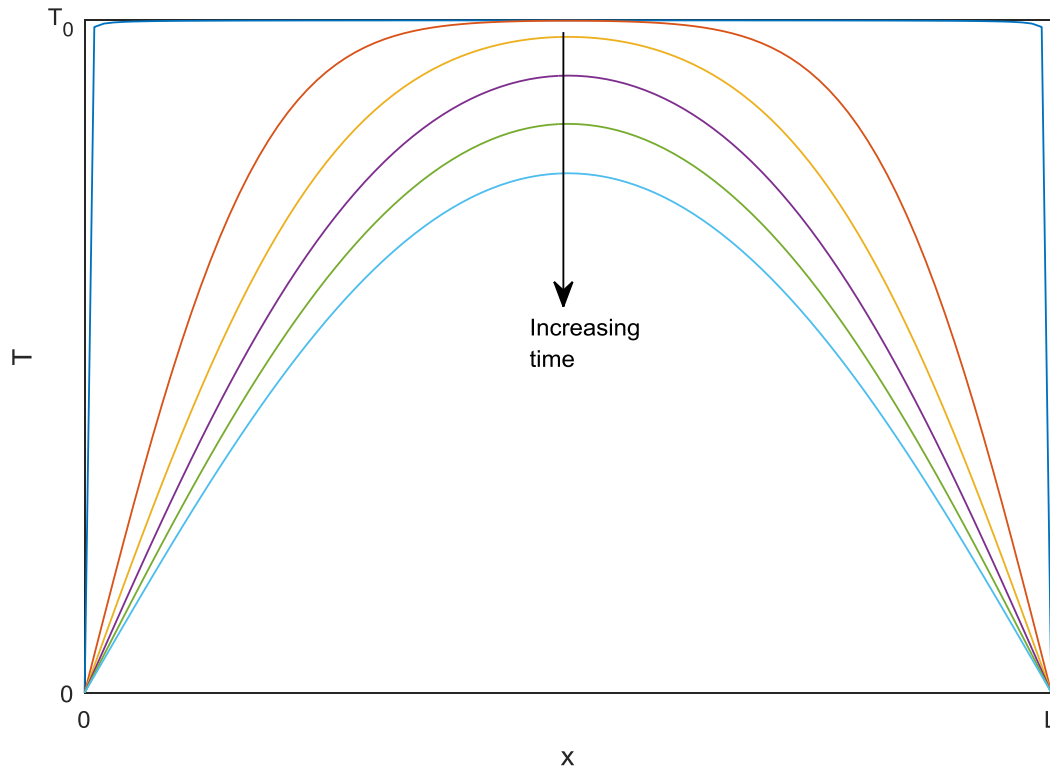
These are elementary integrals that we can do by hand, by using a table, or by using an online calculator. The result is

$$\begin{aligned} A_n L/2 &= 2T_0 L/n\pi \text{ for odd } n, \\ &= 0 \text{ for even } n \end{aligned}$$

So $A_n = 4T_0/n\pi$ for odd n , and 0 for even n . We've now determined A_n for all positive integers n ! The final solution is thus

$$T(x,t) = \frac{4T_0}{\pi} \sum_{n=1,3,5,\dots} \frac{1}{n} \sin\left(\frac{n\pi x}{L}\right) \exp\left[-\alpha\left(\frac{n\pi}{L}\right)^2 t\right]$$

Notice that the “higher order terms” (bigger n) have two characteristics: the spatial dependence is wigglier (smaller distance between peaks), and the exponential decrease with time occurs much more rapidly. The lowest order term ($n = 1$) has the least wiggly spatial dependence (it’s a single sinusoidal “hump”), and the slowest decay. So over time, the temperature distribution evolves toward a single sinusoidal hump (the $n = 1$ term); the higher order terms decay away.



In my solution for $T(x,t)$, why didn't I allow the index n to be negative?

Lecture 6: Nonzero boundary conditions: $T_{\text{steady}}(\mathbf{x}) + T_{\text{transient}}(\mathbf{x},t)$

If we look at a single solution to the heat equation:

$$[A\sin(\lambda x) + B\cos(\lambda x)]\exp(-\alpha\lambda^2 t),$$

we see that it **always contains a term that exponentially decays to 0 over time**. So, unless the final, steady state of a system is 0° throughout, **we need a steady-state term $T_s(\mathbf{x})$ to add to the transient term $T_t(\mathbf{x},t)$, which always decays away**.

Example: Reversing the temperatures of the walls of a slab

The surfaces of a slab are set to 0° at $x = 0$ and T_0 at $x = L$. We know the steady state solution is $T_s(x) = T_0x/L$. But suppose the initial temperature distribution is $T(x,0) = T_0(L-x)/L$, the mirror image of the final steady state distribution. What is $T(x,t)$: how does the temperature distribution evolve from $T_0(L-x)/L$ to T_0x/L ?

Solution

The problem statement gives the steady state solution:

$$T_s(x) = T_0x/L.$$

(If this wasn't given, we'd obtain it quickly from the steady state heat equation, $d^2T_s/dx^2 = 0$, combined with the boundary conditions.)

The full solution has a steady-state and a transient term:

$$T(x,t) = T_s(x) + T_t(x,t).$$

Now we need the boundary conditions and initial condition on $T_t(x,t)$. Since $T_t(x,t) = T(x,t) - T_s(x)$,

$$T_t(0,t) = T(0,t) - T_s(0) = 0 - 0 = 0$$

and

$$T_t(L,t) = T(L,t) - T_s(L) = T_0 - T_0 = 0.$$

Good! The transient term always decays to 0, so we'll have no trouble satisfying the boundary conditions at all t . Now, what's the initial condition on $T_t(x,t)$?

$$T_t(x,0) = T(x,0) - T_s(x) = T_0(L-x)/L - T_0x/L = T_0(L-2x)/L.$$

Now we can plug $T(x,t) = T_s(x) + T_t(x,t)$ into the heat equation:

$$\alpha \left(\frac{d^2T_s}{dx^2} + \frac{\partial^2T_t}{\partial x^2} \right) = \frac{\partial T_t}{\partial t}.$$

But since $d^2T_s/dx^2 = 0$,

$$\alpha \frac{\partial^2T_t}{\partial x^2} = \frac{\partial T_t}{\partial t}.$$

T_t must independently satisfy the heat equation.

The method we use to solve for this function, whose boundary conditions are 0, is identical to the method used in the previous lecture. We set $T_t(x,t) = X(x)\tau(t)$ and just as before find

$$\frac{1}{X} \frac{d^2X}{dx^2} = \frac{1}{\alpha\tau} \frac{d\tau}{dt} = -\lambda^2.$$

It's still true that a solution is

$$T_t(x,t) = X(x)\tau(t) = [A\sin(\lambda x) + B\cos(\lambda x)]\exp(-\alpha\lambda^2t).$$

It's still true that the boundary condition on the left requires $B = 0$, so

$$T_t(x,t) = A\sin(\lambda x)\exp(-\alpha\lambda^2t).$$

It's still true that the boundary condition on the right requires $\lambda = n\pi/L$, so

$$T_t(x,t) = A\sin(n\pi x/L)\exp[-\alpha(n\pi/L)^2t].$$

It's still true that the general solution is

$$T_t(x,t) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{L}\right) \exp\left[-\alpha\left(\frac{n\pi}{L}\right)^2 t\right].$$

To solve for A_n , we apply the initial condition $T_t(x,0) = T_0(L-2x)/L$:

$$\sum_n A_n \sin\left(\frac{n\pi x}{L}\right) = \frac{T_0(L-2x)}{L}.$$

As before, we multiply by $\sin(m\pi x)$ and integrate from 0 to L . As before, all the terms on the left integrate to 0 except $m = n$. So we're left with:

$$A_n \int_0^L \sin^2\left(\frac{n\pi x}{L}\right) dx = T_0 \int_0^L \sin\left(\frac{n\pi x}{L}\right) dx - \frac{2T_0}{L} \int_0^L x \sin\left(\frac{n\pi x}{L}\right) dx$$

Doing the integrals (using an online calculator, for example) gives

$$\begin{aligned} A_n L/2 &= 2T_0 L/n\pi - (2T_0/L)(L^2/n\pi) = 0 \text{ for odd } n, \\ &= 0 - (2T_0/L)(-L^2/n\pi) = 2T_0 L/n\pi \text{ for even } n \end{aligned}$$

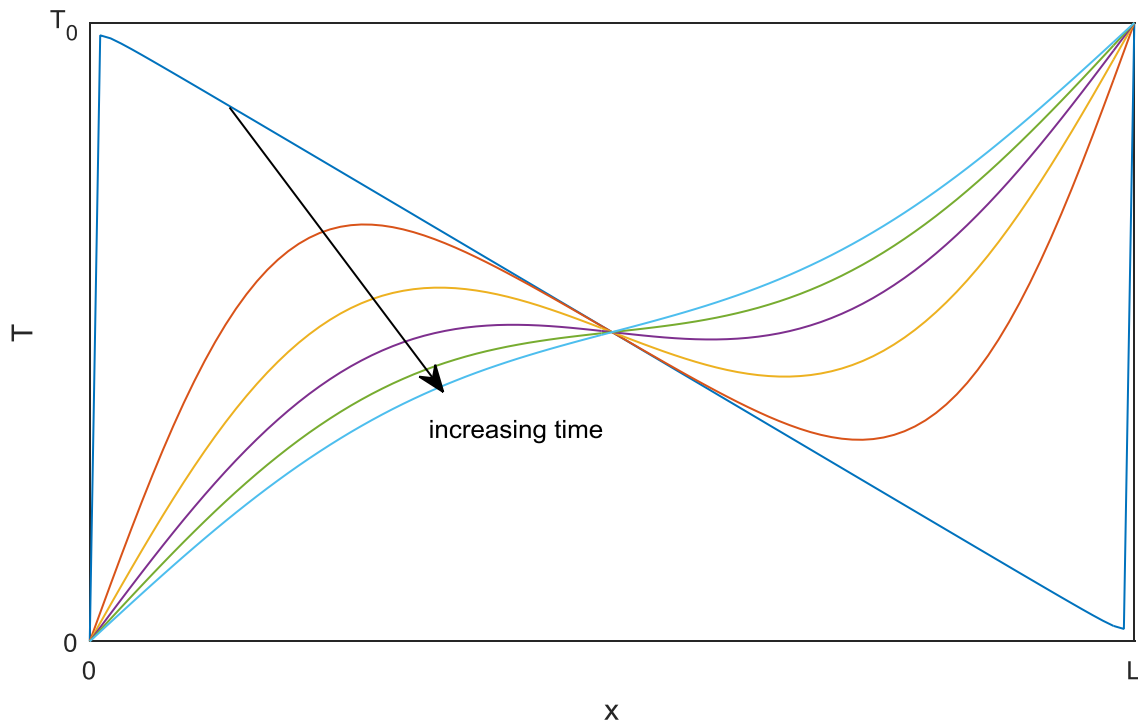
So

$$T_t(x,t) = \frac{4T_0}{\pi} \sum_{n=2,4,6,\dots} \frac{1}{n} \sin\left(\frac{n\pi x}{L}\right) \exp\left[-\alpha\left(\frac{n\pi}{L}\right)^2 t\right]$$

and finally

$$T(x,t) = T_s(x) + T_t(x,t) = \frac{T_0 x}{L} + \frac{4T_0}{\pi} \sum_{n=2,4,6,\dots} \frac{1}{n} \sin\left(\frac{n\pi x}{L}\right) \exp\left[-\alpha\left(\frac{n\pi}{L}\right)^2 t\right]$$

How do I know if I'm right? I plot the equation in MATLAB and see if the temperature changes as expected: gradually evolving from the initial condition $T(x,0) = T_0(L-x)/L$ to the final steady state solution $T_s = T_0 x/L$.



Which of the curves implies the greatest heat flux at the center of the slab? Which implies the least?

Lecture 7: Finite difference approximation

(Yener and Kakac, Heat Conduction, 4th ed., Ch. 12)

We've seen that solving the heat equation can take a lot of work. Depending on the boundary conditions, **it may even be impossible to find an exact solution!** Therefore, we need to develop a method of finding an approximate solution, using a computer. (This is called a **numerical** solution, as opposed to an exact **analytical** solution.)

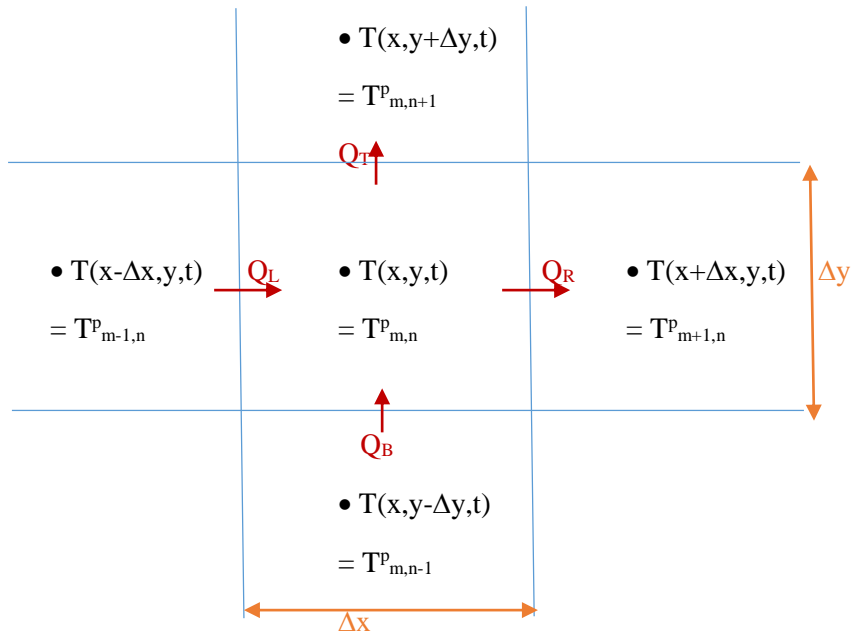
Let's develop an approximate equation for two-dimensional heat flow. Instead of a slab (through which heat flows in only one direction, approximately), we might think of a column. We can impose different thermal boundary conditions on the four sides of a square column.

We divide our solid into tiny volumes $\Delta x \Delta y \Delta z$, where Δz is perpendicular to the diagram. We assume the temperature is approximately uniform throughout each tiny volume. We use the subscripts m and n to represent the x and y coordinates, as shown in the diagram, and the superscript p represents time. We want to determine $T(x, y, t + \Delta t) = T^{p+1}_{m,n}$.

We do the energy balance

$$Q_{\text{in}} - Q_{\text{out}} = mc \frac{\partial T}{\partial t}$$

$$Q_L + Q_B - Q_R - Q_T = mc \frac{\partial T}{\partial t}$$



Now use $Q = qA = -kA \times (\text{spatial derivative of } T)$, with $A = \Delta y \Delta z$ or $\Delta x \Delta z$, the surface through which heat flows into or out of the central volume.

$$-k\Delta y \Delta z (\partial T / \partial x \text{ on the left}) - k\Delta x \Delta z (\partial T / \partial y \text{ on the bottom}) + k\Delta y \Delta z (\partial T / \partial x \text{ on the right}) + k\Delta x \Delta z (\partial T / \partial y \text{ on the top}) = mc \partial T / \partial t$$

The next (key) step is to **replace the derivatives with finite differences**:

$$-k\Delta y \Delta z (T^p_{m,n} - T^p_{m-1,n}) / \Delta x - k\Delta x \Delta z (T^p_{m,n} - T^p_{m,n-1}) / \Delta y + k\Delta y \Delta z (T^p_{m+1,n} - T^p_{m,n}) / \Delta x + k\Delta x \Delta z (T^p_{m,n+1} - T^p_{m,n}) / \Delta y = mc (T^{p+1}_{m,n} - T^p_{m,n}) / \Delta t$$

Divide by mc and use $m = \rho \Delta x \Delta y \Delta z$ and $\alpha = k / \rho c$:

$$-\alpha (T^p_{m,n} - T^p_{m-1,n}) / (\Delta x)^2 - \alpha (T^p_{m,n} - T^p_{m,n-1}) / (\Delta y)^2 + \alpha (T^p_{m+1,n} - T^p_{m,n}) / (\Delta x)^2 + \alpha (T^p_{m,n+1} - T^p_{m,n}) / (\Delta y)^2 = (T^{p+1}_{m,n} - T^p_{m,n}) / \Delta t$$

Finally, collect terms:

$$\alpha (T^p_{m+1,n} + T^p_{m-1,n} - 2T^p_{m,n}) / (\Delta x)^2 + \alpha (T^p_{m,n+1} + T^p_{m,n-1} - 2T^p_{m,n}) / (\Delta y)^2 = (T^{p+1}_{m,n} - T^p_{m,n}) / \Delta t$$

We see that this is the **approximation to the two-dimensional heat equation**, where

$$\partial^2 T / \partial x^2 \approx (T^p_{m+1,n} + T^p_{m-1,n} - 2T^p_{m,n}) / (\Delta x)^2,$$

$$\partial^2 T / \partial y^2 \approx (T^p_{m,n+1} + T^p_{m,n-1} - 2T^p_{m,n}) / (\Delta y)^2,$$

and

$$\partial T / \partial t \approx (T^{p+1}_{m,n} - T^p_{m,n}) / \Delta t.$$

For one-dimensional heat flow, simply drop the y term.

How do we use the finite difference approximation to solve the heat equation? There are separate strategies for time-dependent and steady-state heat flow. **The time-dependent case is actually easier**, assuming we know initial conditions. We simply solve for $T^{p+1}_{m,n}$:

$$T^{p+1}_{m,n} = T^p_{m,n} + \alpha\Delta t(T^p_{m+1,n} + T^p_{m-1,n} - 2T^p_{m,n})/(\Delta x)^2 + \alpha\Delta t(T^p_{m,n+1} + T^p_{m,n-1} - 2T^p_{m,n})/(\Delta y)^2$$

and step forward in time. Specifically:

- Choose small values of Δx , Δy , and Δt . Requirement: the total coefficient of $T^p_{m,n}$ in the previous equation must be positive, so $1 - 2\alpha\Delta t/(\Delta x)^2 - 2\alpha\Delta t/(\Delta y)^2 > 0$.
- Given initial conditions, we know $T^1_{m,n}$ for all points. (I'm using $p = 1$ to represent $t = 0$, so $p = 2$ represents $t = \Delta t$, $p = 3$ represents $t = 2\Delta t$, etc.)
- Determine $T^2_{m,n}$ for all interior points using the equation above. Apply boundary conditions for all points on the boundaries. (If the temperature is specified at the boundaries, this is easy. We'll look at convective boundary conditions soon.)
- Iterate the previous step: Knowing $T^2_{m,n}$, we can use the equation to determine $T^3_{m,n}$. Continue for as long as desired.

The steady-state case is harder because there's no starting point. We solve for $T_{m,n}$ and drop all p 's because there's no time dependence:

$$T_{m,n} = [(T_{m+1,n} + T_{m-1,n})/(\Delta x)^2 + (T_{m,n+1} + T_{m,n-1})/(\Delta y)^2]/[2/(\Delta x)^2 + 2/(\Delta y)^2]$$

This equation is valid for every interior point in the solid. (Assume for now that the temperatures are specified on the boundaries.) If we have a grid of 100×100 interior points, we have 10,000 unknown temperatures $T_{m,n}$, and 10,000 equations. How do we solve 10,000 equations in 10,000 unknowns? The simplest method is the Gauss-Seidel method:

- Guess the temperature $T_{m,n}$ at all points.
- Now calculate $T_{m,n}$ at all points using the above equation. (The temperature at each point depends on its neighbors.) This is an improvement over the initial guess.
- Iterate until the temperature changes negligibly from one iteration to the next.

Convective boundary conditions

Let's place a convective boundary condition on the right. We'll put our rightmost points right along the boundary. Once again we apply the energy balance:

$$Q_{in} - Q_{out} = mc \partial T/\partial t$$

$$Q_L + Q_B - Q_R - Q_T = mc \partial T/\partial t$$

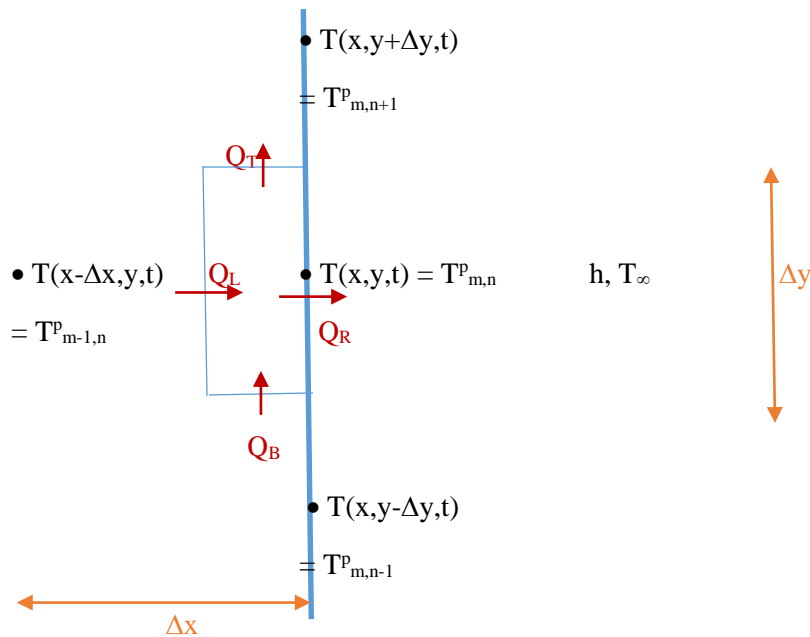
Q_L is the same as before, but Q_R is now $h\Delta y\Delta z(T^p_{m,n} - T_\infty)$, and the surface area through which Q_B and Q_T flow is only $(\Delta x/2)\Delta z$. ($T^p_{m,n}$ now represents the temperature in the little rectangle.)

$$-k\Delta y\Delta z(T^p_{m,n} - T^p_{m-1,n})/\Delta x - k\Delta x\Delta z(T^p_{m,n} - T^p_{m,n-1})/2\Delta y - h\Delta y\Delta z(T^p_{m,n} - T_\infty) + k\Delta x\Delta z(T^p_{m,n+1} - T^p_{m,n})/2\Delta y = mc(T^{p+1}_{m,n} - T^p_{m,n})/\Delta t$$

As before, divide by mc and use $m = \rho(\Delta x/2)\Delta y\Delta z$ and $\alpha = k/\rho c$:

$$-2\alpha(T_{m,n}^p - T_{m-1,n}^p)/(\Delta x)^2 - \alpha h(T_{m,n}^p - T_\infty)/k\Delta x + \alpha(T_{m+1,n}^p - T_{m,n}^p)/(\Delta x)^2 + \alpha(T_{m,n+1}^p - T_{m,n}^p)/(\Delta y)^2 = (T_{m,n}^{p+1} - T_{m,n}^p)/\Delta t$$

This result can be manipulated as necessary.

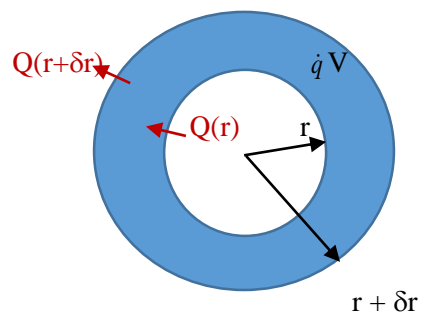


For an insulated boundary, set $h = 0$. For a corner with convective boundaries, the expression may be derived by drawing the appropriate diagram and following the procedure used above. Do we need a special expression for a corner where both walls have fixed temperatures?

Lecture 8: Radially symmetric temperature distribution

(Lienhard and Lienhard pp. 67-68)

Let's derive the heat equation in cylindrical coordinates, assuming the temperature is radially symmetric (no dependence on θ or z). Our tiny volume in this case will be a cylindrical shell, of infinitesimal thickness δr , and height L (perpendicular to the diagram):



I'll allow for the possibility of internal energy generation (\dot{q}). As always, we begin with the energy balance:

$$Q_{in} - Q_{out} = mc \partial T / \partial t$$

$$Q(r) + \dot{q}V - Q(r+\delta r) = mc \frac{\partial T}{\partial t},$$

where the volume of a cylindrical shell is $V = \pi(r+\delta r)^2L - \pi r^2L \approx 2\pi r\delta rL$, dropping the $(\delta r)^2$ term because $\delta r \ll r$. (The figure isn't to scale; δr approaches infinitesimal dr .) Using $Q = qA = (-k\partial T/\partial r)A$,

$$-k2\pi rL \left. \frac{\partial T}{\partial r} \right|_r + \dot{q}2\pi r\delta rL + k2\pi(r+\delta r)L \left. \frac{\partial T}{\partial r} \right|_{r+\delta r} = mc \frac{\partial T}{\partial t}$$

Now I'll use

$$\left. \frac{\partial T}{\partial r} \right|_{r+\delta r} \approx \left. \frac{\partial T}{\partial r} \right|_r + \frac{\partial^2 T}{\partial r^2} \delta r$$

I'll plug this in and drop the $(\delta r)^2$ term from the result:

$$k2\pi L \left(r\delta r \frac{\partial^2 T}{\partial r^2} + \delta r \frac{\partial T}{\partial r} \right) + \dot{q}2\pi r\delta rL = mc \frac{\partial T}{\partial t}$$

I'll use $m = \rho 2\pi r\delta rL$ and divide by $2\pi r\delta rLk$:

$$\frac{\partial^2 T}{\partial r^2} + \frac{1}{r} \frac{\partial T}{\partial r} + \frac{\dot{q}}{k} = \frac{1}{\alpha} \frac{\partial T}{\partial t}$$

This is usually written slightly more compactly as

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial T}{\partial r} \right) + \frac{\dot{q}}{k} = \frac{1}{\alpha} \frac{\partial T}{\partial t}$$

The same strategy allows us to derive the heat equation for a **spherically symmetric** temperature distribution. The only difference is that the volume of a spherical shell is $(4/3)\pi(r+\delta r)^3 - (4/3)\pi r^3 \approx 4\pi r^2\delta r$, and the final result is

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial T}{\partial r} \right) + \frac{\dot{q}}{k} = \frac{1}{\alpha} \frac{\partial T}{\partial t}.$$

If you've taken multivariable calculus, you recognize that we've derived the radial part of the Laplacian in cylindrical and spherical coordinates.

Example: Cylinder with fixed wall temperatures

A cylindrical shell has inner radius r_i and outer radius r_o . The temperature of the inner surface is T_i , and the temperature of the outer surface is T_o . Determine the temperature throughout the cylindrical shell.

Solution

Since there's no internal heat generation and no time dependence, the heat equation reduces to

$$\frac{d}{dr} \left(r \frac{dT}{dr} \right) = 0.$$

Integrate this once:

$$r(dT/dr) = C_1,$$

where C_1 is a constant of integration. Bring dr/r to the right side and integrate a second time:

$$T = C_1 \ln r + C_2.$$

Solve for C_1 and C_2 using the boundary conditions, $T(r_i) = T_i$ and $T(r_o) = T_o$:

$$T_i = C_1 \ln r_i + C_2$$

$$T_o = C_1 \ln r_o + C_2$$

Subtract one from the other to find

$$T_i - T_o = C_1 \ln r_i - C_1 \ln r_o,$$

so

$$C_1 = (T_i - T_o) / \ln(r_i/r_o).$$

Plugging this into the T_i equation:

$$T_i = (T_i - T_o) \ln r_i / \ln(r_i/r_o) + C_2,$$

so

$$C_2 = T_i - (T_i - T_o) \ln r_i / \ln(r_i/r_o).$$

The final result is

$$T = (T_i - T_o) \ln r / \ln(r_i/r_o) + T_i - (T_i - T_o) \ln r_i / \ln(r_i/r_o) = T_i + (T_i - T_o) \ln(r/r_i) / \ln(r_i/r_o)$$

Does this work regardless of which surface is hotter?

Lecture 9: The two-dimensional heat equation

(Lienhard and Lienhard pp. 146-150)

We've already derived the one-dimensional heat equation, and the finite-difference approximation of the two-dimensional heat equation. If we wanted, we could consider a tiny volume and derive the two-dimensional heat equation:

$$\alpha \frac{\partial^2 T}{\partial x^2} + \alpha \frac{\partial^2 T}{\partial y^2} = \frac{\partial T}{\partial t}$$

Let's look at the steady-state case:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0$$

Example: Another lucky boundary condition

A long slab has surfaces at $x = 0$, $x = L$, $y = 0$, and $y = (\text{something very large } \approx \infty)$. The surfaces at $x = 0$ and $x = L$ are held at 0°C . The temperature distribution along the $y = 0$ surface is $T(x,0) = \beta \sin(\pi x/L)$. Determine $T(x,y)$ throughout the slab.

Solution

When T depended on x and t , we wrote $T(x,t) = X(x)\tau(t)$. Similarly, when T depends on x and y , we write $T(x,y) = X(x)Y(y)$. We plug this into the heat equation:

$$Y \frac{d^2 X}{dx^2} + X \frac{d^2 Y}{dy^2} = 0.$$

We divide both sides by XY :

$$\frac{1}{X} \frac{d^2 X}{dx^2} + \frac{1}{Y} \frac{d^2 Y}{dy^2} = 0,$$

so

$$\frac{1}{X} \frac{d^2 X}{dx^2} = -\frac{1}{Y} \frac{d^2 Y}{dy^2}.$$

The left side is a function of x . The right side is a function of y . The only way they can equal each other is if they're both equal to the same constant, which I'll again call $-\lambda^2$. This gives us two differential equations, one in x , and one in y :

$$\frac{d^2 X}{dx^2} = -\lambda^2 X$$

and

$$\frac{d^2 Y}{dy^2} = \lambda^2 Y.$$

We've seen the x equation before. The solution is

$$X = A \sin(\lambda x) + B \cos(\lambda x).$$

The solution to the y equation is

$$Y = e^{-\lambda y},$$

where I'm neglecting $e^{+\lambda y}$ because I don't want infinite temperatures as I get far from $y = 0$. I'm also neglecting the multiplicative constant because I'm going to multiply X and Y together, and I've included the constants in X .

So

$$T(x,y) = [A \sin(\lambda x) + B \cos(\lambda x)] e^{-\lambda y}.$$

Now we apply boundary conditions to determine A and B. Since $T(0,y) = 0$, $B = 0$. Since $T(L,y) = 0$,

$$\sin(\lambda L) = 0,$$

so $\lambda = n\pi/L$. Our solution is now

$$T(x,y) = A\sin(n\pi x/L)e^{-n\pi y/L},$$

and the only unknowns are n and L. Since $T(x,0) = \beta\sin(\pi x/L)$, we see that $A = \beta$ and $n = 1$, so

$$T(x,y) = \beta\sin(\pi x/L)e^{-\pi y/L}.$$

We're done, but **we were lucky**. What if the boundary condition $T(x,0)$ is anything but a multiple of $\sin(n\pi x/L)$?

Example: A triangular thermal distribution at one wall

A long slab has surfaces at $x = 0$, $x = L$, $y = 0$, and $y =$ (something very large $\approx \infty$). The surfaces at $x = 0$ and $x = L$ are held at 0°C . The temperature distribution along the $y = 0$ surface is $T(x,0) = \beta x$ for $0 < x < L/2$, and $T(x,0) = \beta(L-x)$ for $L/2 < x < L$. Determine $T(x,y)$ throughout the slab.

Solution

Just as before, we separate variables, so $T(x,y) = X(x)Y(y)$, and

$$Y \frac{d^2 X}{dx^2} + X \frac{d^2 Y}{dy^2} = 0,$$

which leads to

$$\frac{d^2 X}{dx^2} = -\lambda^2 X$$

and

$$\frac{d^2 Y}{dy^2} = \lambda^2 Y.$$

Again,

$$X = A\sin(\lambda x) + B\cos(\lambda x)$$

and

$$Y = e^{-\lambda y},$$

so just as before, a solution is

$$T(x,y) = [A\sin(\lambda x) + B\cos(\lambda x)]e^{-\lambda y}.$$

Even the boundary conditions at $x = 0$ and $x = L$ are the same, so $B = 0$ and $\lambda = n\pi/L$. Therefore a solution is

$$T(x,y) = A_n \sin(n\pi x/L) e^{-n\pi y/L},$$

but we'll need a sum of these solutions to satisfy the final boundary condition at $y = 0$. The complete solution is

$$T(x, y) = \sum_{n=1,2,3} A_n \sin\left(\frac{n\pi x}{L}\right) e^{-n\pi y/L}.$$

Now all we have to do is solve for an infinite number of coefficients A_n . We do this by using the boundary condition at $y = 0$:

$$\sum_{n=1,2,3,\dots} A_n \sin\left(\frac{n\pi x}{L}\right) = \begin{cases} \beta x & x < L/2 \\ \beta(L-x) & x > L/2 \end{cases}$$

Luckily, Fourier's ingenious idea works just as well here as it did before. We multiply both sides by $\sin(m\pi x/L)$ and integrate from 0 to L . All the terms in the sum integrate to 0 except $m = n$, so we obtain

$$A_n \int_0^L \sin^2\left(\frac{n\pi x}{L}\right) dx = \int_0^{L/2} \beta x \sin\left(\frac{n\pi x}{L}\right) dx + \int_{L/2}^L \beta(L-x) \sin\left(\frac{n\pi x}{L}\right) dx.$$

These are just integrals, and you can find the solutions using any method you like, including online integral calculators. The result is

$$A_n = 4\beta L (-1)^{(n-1)/2} / \pi^2 n^2 \quad \text{for odd } n; \quad A_n = 0 \quad \text{for even } n.$$

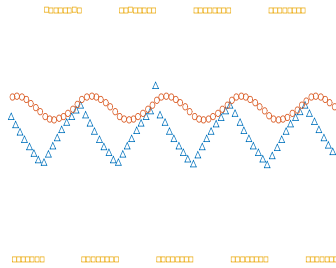
Our final solution is therefore

$$T(x, y) = \frac{4\beta L}{\pi^2} \sum_{n=1,3,5,\dots} \frac{(-1)^{(n-1)/2}}{n^2} \sin\left(\frac{n\pi x}{L}\right) e^{-n\pi y/L}$$

When I separated variables, how did I know to put the minus on the y side: $\frac{1}{X} \frac{d^2 X}{dx^2} = -\frac{1}{Y} \frac{d^2 Y}{dy^2}$?

Equivalently, how did I know to set both sides of this equation to $-\lambda^2$ instead of $+\lambda^2$?

UNIT 2: DESIGN OF AMPLIFIERS: OP AMPS, CIRCUIT ANALYSIS, AND LAPLACE TRANSFORMS



Circuit analysis doesn't have much in common with heat conduction. However, as promised, we'll continue to practice the broadly applicable skills we've learned: while analyzing circuits, we'll use Fourier series (Problem Set 15), finite difference approximations (Problem Set 16), and the Arduino (Labs 3 and 4). Also, the new skill we'll learn, Laplace transforms, will be applied to heat conduction (Problem Set 16).

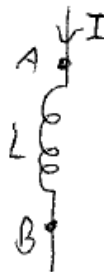
Lecture 10: Op amps

(Santiago, Ch. 10)

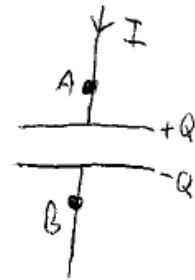
First let's review the relationship between voltage and current in resistors, inductors, and capacitors:



$$V_A - V_B = IR$$



$$V_A - V_B = L \frac{dI}{dt}$$

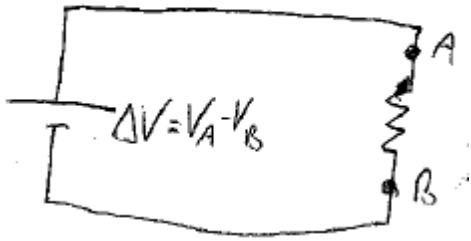


$$V_A - V_B = \frac{Q}{C}$$

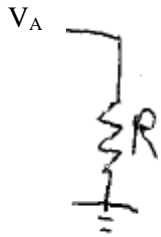
$$\frac{dV_A}{dt} - \frac{dV_B}{dt} = \frac{I}{C}$$

These equations are true if we draw the current I flowing **from A to B**. Notice that the voltage across a resistor is proportional to current, the voltage across an inductor is proportional to the derivative of current, and the voltage across a capacitor is proportional to the integral of current (which is charge).

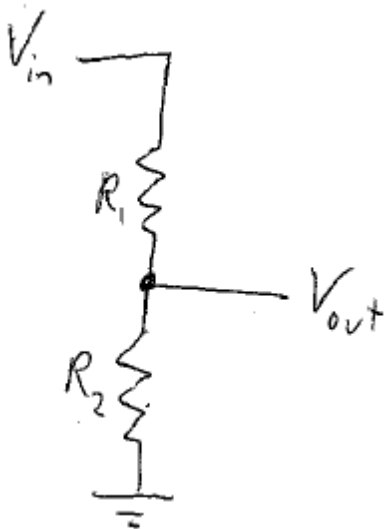
Notice, too, that these equations are true even if we don't show the rest of the circuit. We know that current only flows in a loop, but we don't have to show the whole loop. In fact, I'll almost never show the whole loop like this:



I'll imply **exactly the same thing** with this abbreviated diagram (where $V_B = 0$, ground):



I use the current-voltage relationships shown above for resistors, inductors, and capacitors; and I use Kirchhoff's current rule. **I almost never explicitly use Kirchhoff's voltage rule!** Let's see how I analyze a useful circuit:



This circuit is called a **voltage divider**, and we'll see many variations of it. Here's the key idea: **current flows from (a battery or other power supply whose voltage is) V_{in} , through R_1 and R_2 , and into ground**; we assume (for now) that **negligible current flows through the wire labeled V_{out}** . (This assumption is 100% true if nothing is connected to this wire: if it's just sticking out in space. The V_{in} wire, on the other hand, must be connected to a voltage source. If you wish, you may explicitly draw a battery whose negative terminal is connected to ground and whose positive terminal is connected to V_{in} . Even if we don't draw it, this battery—or equivalent voltage source—is implied!)

As long as we're assuming that negligible current flows along the V_{out} wire (which is therefore just a label or tag), the same current must flow through R_1 and R_2 .

The current through R_1 is

$$I = (V_{in} - V_{out})/R_1,$$

and the same current through R_2 is

$$I = (V_{out} - 0)/R_2.$$

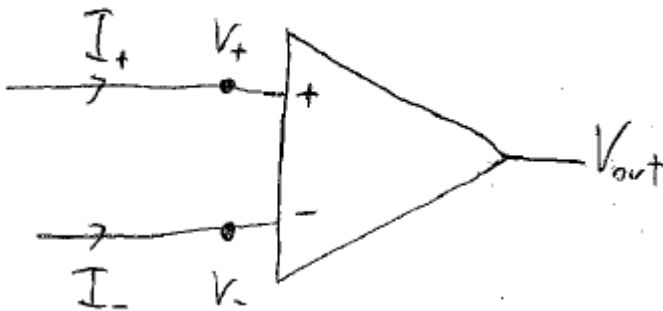
We can combine these equations to eliminate I and find V_{out} in terms of V_{in} , but the algebra is simpler if we write the current through the series combination:

$$I = (V_{in} - 0)/(R_1 + R_2).$$

Make sure you understand where these equations come from! All three of them are applications of $V_A - V_B = IR$, with current flowing from A to B. If you combine any two of the three equations to eliminate I , you find

$$V_{out} = [R_2/(R_1 + R_2)]V_{in}$$

Now we're ready to introduce the op amp, which will allow you to easily design many useful circuits. The op amp is made of many transistors and is usually packaged as an "integrated circuit" (a tiny black box with metal pins sticking out) with 8 pins. However, we usually represent only three of those pins in a circuit diagram:



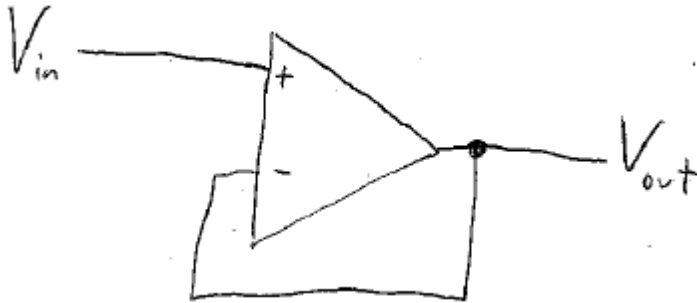
These three op amp terminals (which I label by their voltages) are

- V_+ , the noninverting input
- V_- , the inverting input
- V_{out} , the output

There are **two Golden Rules** of op amp operation:

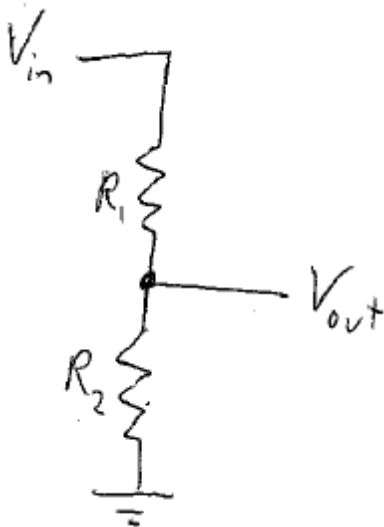
- Golden Rule for current: $I_+ = I_- = 0$. This is always (approximately) true. No current flows in or out of the op amp **inputs**. However, current may flow in or out of the **output**.
- Golden Rule for voltage: $V_+ = V_-$. This is (approximately) true for most op amp circuits, but it depends on the circuit design. More on this later.

Now let's apply these rules to some simple but practical op amp circuits.

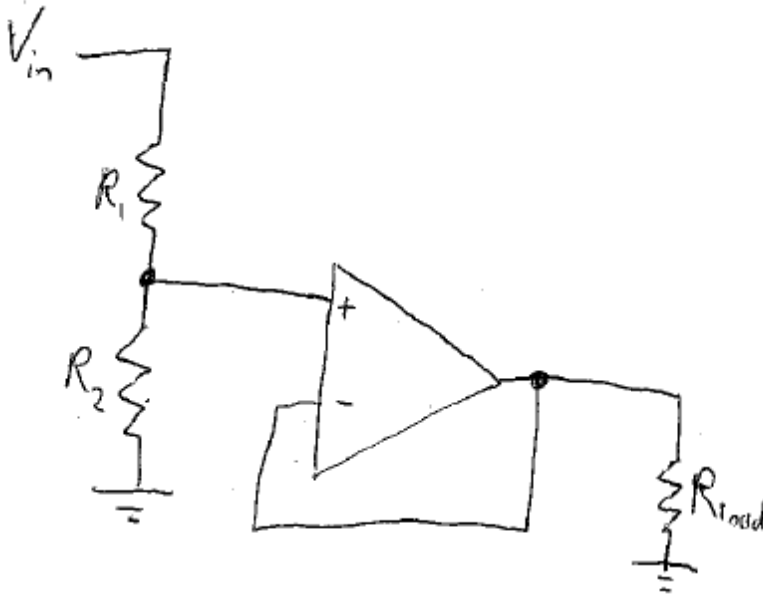
Follower

In this circuit, $V_+ = V_{in}$ and $V_- = V_{out}$. The voltage Golden Rule gives us $V_{out} = V_{in}$. The output follows the input. Although this is the simplest op amp circuit, its purpose may be the hardest to understand: What's the point of this circuit?

Think back to the voltage divider.



Remember our dubious assumption that no current flows into V_{out} ? If we connect the voltage divider's output to the follower's input, then the assumption is guaranteed to be true, by the current Golden Rule.



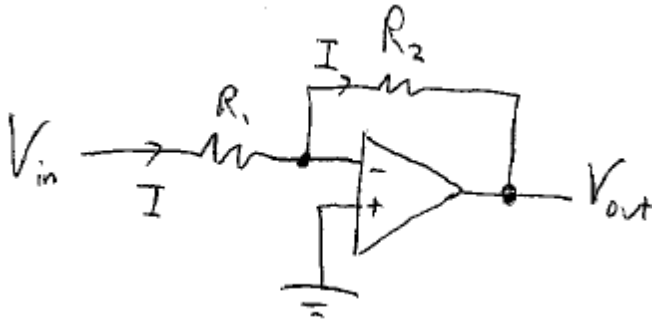
Now $V_+ = [R_2/(R_1 + R_2)]V_{in}$, the output of the voltage divider, which is also the output of the op amp, because this is a follower circuit. So we can apply our desired voltage, $[R_2/(R_1 + R_2)]V_{in}$, across any resistor (known as a “load” in electronics jargon); **the resistance of R_{load} will not affect the voltage across R_{load} .** The current through the load comes from the op amp output, not the voltage divider. **The follower circuit acts as a buffer that transmits a voltage without draining any current from the source (which would lower the output voltage).**

If we connected the load directly to the output of the voltage divider, current **would** flow through the output wire of the voltage divider (our earlier assumption would be false), and the voltage across the load would be (confirm this)

$$\frac{\frac{R_2 R_{load}}{R_2 + R_{load}}}{R_1 + \frac{R_2 R_{load}}{R_2 + R_{load}}} V_{in}.$$

Without the follower circuit, the output of the voltage divider depends on the load resistance, which is undesirable; we’d like to say, “The output of our voltage divider is $[R_2/(R_1 + R_2)]V_{in}$, no matter what we apply this voltage across.”

Inverting amplifier



The voltage Golden Rule gives $V_- = 0$ since $V_+ = 0$ (ground). The current Golden Rule requires I through R_1 to equal I through R_2 ; no current flows into the op amp input. Applying Ohm's law to each resistor:

$$I = (V_{in} - 0)/R_1$$

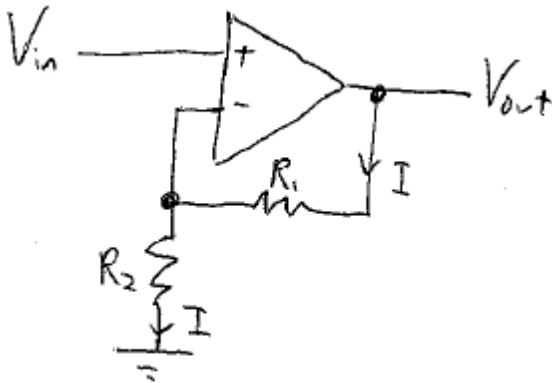
and

$$I = (0 - V_{out})/R_2,$$

so

$$V_{out} = (-R_2/R_1)V_{in}.$$

Noninverting amplifier



The voltage Golden Rule gives $V_- = V_{in}$. The current Golden Rule requires I through R_1 to equal I through R_2 . Applying Ohm's law to each resistor:

$$I = (V_{out} - V_{in})/R_1$$

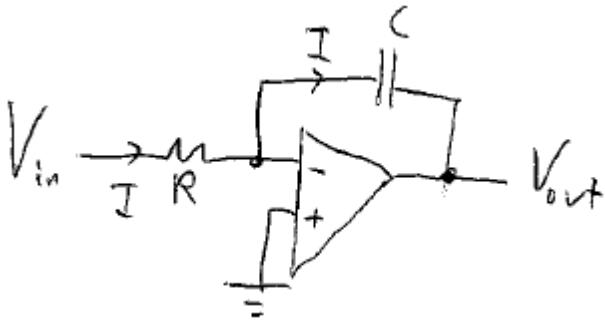
and

$$I = (V_{in} - 0)/R_2,$$

so, after some algebra,

$$V_{out} = (1 + R_1/R_2)V_{in}.$$

Integrator



Here's a circuit that does calculus! The voltage Golden Rule sets $V_- = 0$. The current Golden Rule requires I through R to equal I through C . Ohm's law gives

$$I = (V_{in} - 0)/R_1$$

and the equivalent rule for capacitors, $Q = C(0 - V_{out})$, is differentiated to give

$$I = -C dV_{out}/dt.$$

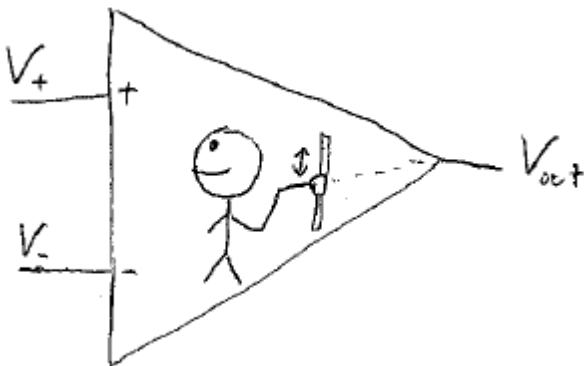
Setting them equal gives

$$dV_{out}/dt = -V_{in}/RC,$$

so $V_{out} = -(1/RC)\int V_{in}dt$.

What's the difference between V_+ and V_- ?

That's enough circuits for today, but to design op amp circuits, we have to know the difference between V_+ and V_- . If we reverse them, the circuit won't work! Let's glance inside the op amp and pretend that there's a smiley stick figure who monitors the inputs and controls the output.



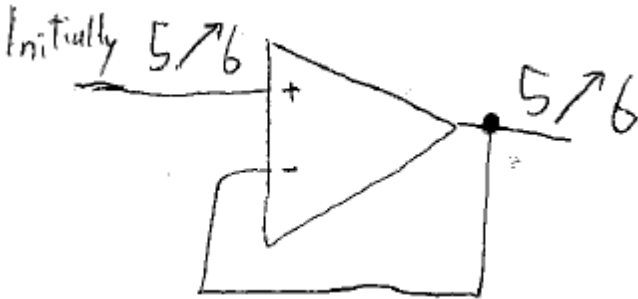
How does Stick Figure decide how to set V_{out} ? Stick Figure follows these rules:

Condition 1. If $V_+ > V_-$, V_{out} will increase.

Condition 2. If $V_- > V_+$, V_{out} will decrease.

Condition 3. If $V_+ \approx V_-$, V_{out} will stay how it is (the voltage Golden Rule is achieved!).

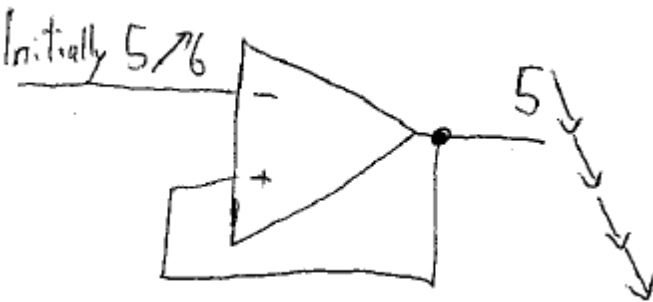
Let's see how these rules apply to a correctly designed follower.



Suppose the input and output voltages are initially 5 V. Stick Figure sees that the circuit is in condition 3: the inputs are equal, so the output voltage shouldn't change.

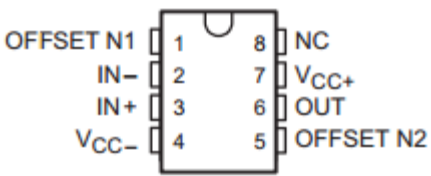
Now, suppose the input increases to 6 V. The output voltage does not change instantaneously and is still 5 V. Stick Figure sees that the circuit is in condition 1: $V_+ = 6$ V, which is greater than $V_- = 5$ V. Therefore, Stick Figure causes the output to increase. It increases to 5.1 V, but the op amp is still in condition 1. The output increases to 5.5 V, but the op amp is still in condition 1. The op amp output increases to 6 V, when finally the op amp is in condition 3, and the output doesn't need to change any more.

Next, let's analyze an incorrectly designed follower, whose + and - are reversed:



Once again, suppose that the input and output voltages are initially 5 V, so the circuit is in condition 3. Next, the input voltage increases to 6 V. Stick Figure now sees that the op amp is in condition 2: $V_- = 6$ V, which is greater than $V_+ = 5$ V. Now, poor Stick Figure (who really doesn't know any better) has to decrease the op amp output, first to 4.9 V, then 4 V, then 0 V, then even lower. But the op amp remains in condition 2 (and gets further and further from condition 3).

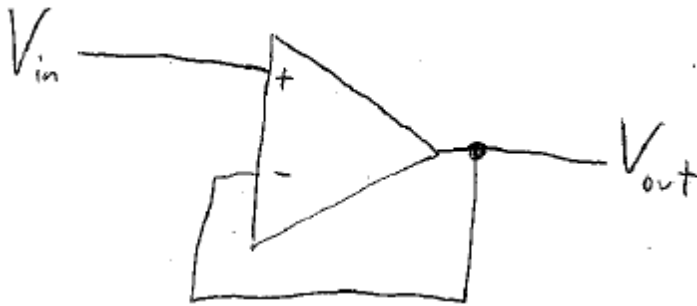
The op amp's output can't decrease forever. To predict the voltage it reaches, we must discuss two additional pins in the op amp: the **supply voltages**. Here's a pin diagram for a typical op amp:



(figure from <http://www.ti.com/lit/ds/symlink/ua741.pdf>)

In this diagram, $IN_- = V_-$, $IN_+ = V_+$, and $OUT = V_{out}$. **We must always apply voltages to V_{CC+} and V_{CC-} to power the op amp.** Typically V_{CC+} and V_{CC-} are +15 V and -15 V respectively, or ± 12 V, or, if you're using two 9 V batteries, ± 9 V. **V_{out} is constrained by the supply voltages:** $V_{CC-} < V_{out} < V_{CC+}$.

Just in case you're curious, here's one final detail: Why do I keep saying $V_+ \approx V_-$ instead of $V_+ = V_-$? If we want to be extremely precise, we can replace the voltage Golden Rule with the exact equation $V_{out} = A(V_+ - V_-)$, where A is a huge number (\sim ten million) called open-circuit gain. Apply this to the follower:



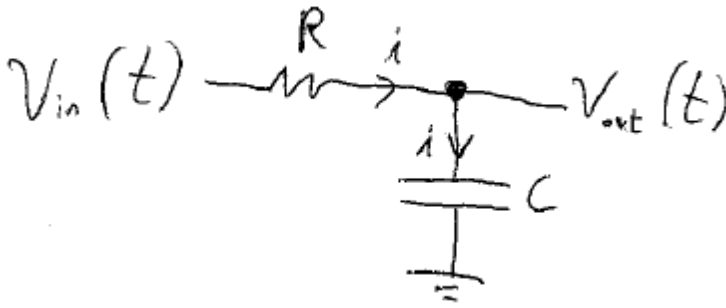
$V_{out} = A(V_+ - V_-) = A(V_{in} - V_{out})$. Solving for V_{out} : $V_{out} = V_{in}A/(A + 1) \approx V_{in}$, which is what the voltage Golden Rule gives us (much more quickly).

Look back to the R, L, and C diagrams at the beginning of the lecture. What if I is negative? Is this allowed? Are the equations still true? What if we don't know the direction of I ?

Lecture 11: An RC circuit and 3 ways to analyze it

(Santiago Ch. 13 and 15)

Consider this simple circuit:



I'm interested in a **sinusoidal input** $v_{in}(t)$, but I'm noncommittal about phase; for v_{in} , I'm equally happy with $V_{in}\sin(\omega t)$, $V_{in}\cos(\omega t)$, or $V_{in}\sin(\omega t + \delta)$, where δ is an arbitrary constant. (I'm using **lower-case letters for instantaneous voltages and currents** and **upper-case letters for constant amplitudes**.)

We can reasonably expect v_{out} to also be sinusoidal, but the output amplitude (V_{out}) will be smaller than the input amplitude (V_{in}). Also, there will be a phase shift between the input and output: the peaks of the input will not occur at the same time as the peaks of the output.

Our goal is to determine the ratio of output amplitude to input amplitude, V_{out}/V_{in} , and as a secondary interest, to determine the phase shift between input and output. We will solve this problem four ways! (Three ways today, and a fourth way in Lecture 13.)

Method 1. Write and solve the differential equation

Evaluation of this method: It's often (but not always) the most laborious method. The integrator circuit is an example of a circuit whose differential equation is so simple that this method may be the best.

Let $v_{in}(t) = V_{in}\sin(\omega t)$.

From the circuit diagram, we see that the same current i flows through both R and C . (We're assuming no "load" is connected to the output, except possibly a follower circuit, which draws no current.)

The voltage across the capacitor is

$$v_{out} = q/C.$$

The voltage across the resistor is iR . Therefore the voltage across the series combination is $q/C + iR$, and this is v_{in} :

$$V_{in}\sin(\omega t) = q/C + iR,$$

where $i = dq/dt$, so

$$V_{in}\sin(\omega t) = q/C + R dq/dt.$$

Here's our differential equation, in q . Let's guess a solution. Since the input voltage is sinusoidal, it's reasonable to expect that i and q are also sinusoidal, though we don't know the amplitude and phase. So let's guess

$$q = Q\sin(\omega t + \phi),$$

where Q and ϕ are unknown and need to be determined, by plugging this expression into the differential equation:

$$V_{in}\sin(\omega t) = Q\sin(\omega t + \phi)/C + RQ\omega\cos(\omega t + \phi)$$

We will simplify this using the trig identity $\alpha\sin\theta + \beta\cos\theta = (\alpha^2 + \beta^2)^{1/2}\sin(\theta + \tan^{-1}(\beta/\alpha))$:

$$V_{in}\sin(\omega t) = Q(1/C^2 + \omega^2 R^2)^{1/2}\sin(\omega t + \phi + \tan^{-1}(\omega RC)).$$

We see that

$$Q = V_{in}/(1/C^2 + \omega^2 R^2)^{1/2}$$

and

$$\phi = -\tan^{-1}(\omega RC),$$

so

$$q = V_{in}/(1/C^2 + \omega^2 R^2)^{1/2} \sin(\omega t - \tan^{-1}(\omega RC))$$

and

$$v_{out} = q/C = V_{in}/(1 + \omega^2 R^2 C^2)^{1/2} \sin(\omega t - \tan^{-1}(\omega RC)).$$

Therefore the output amplitude

$$V_{out} = V_{in}/(1 + \omega^2 R^2 C^2)^{1/2},$$

and the output amplitude is smaller than the input amplitude by a factor of

$$V_{out}/V_{in} = 1/(1 + \omega^2 R^2 C^2)^{1/2}.$$

The phase shift is $-\tan^{-1}(\omega RC)$.

Method 2. Phasor diagram

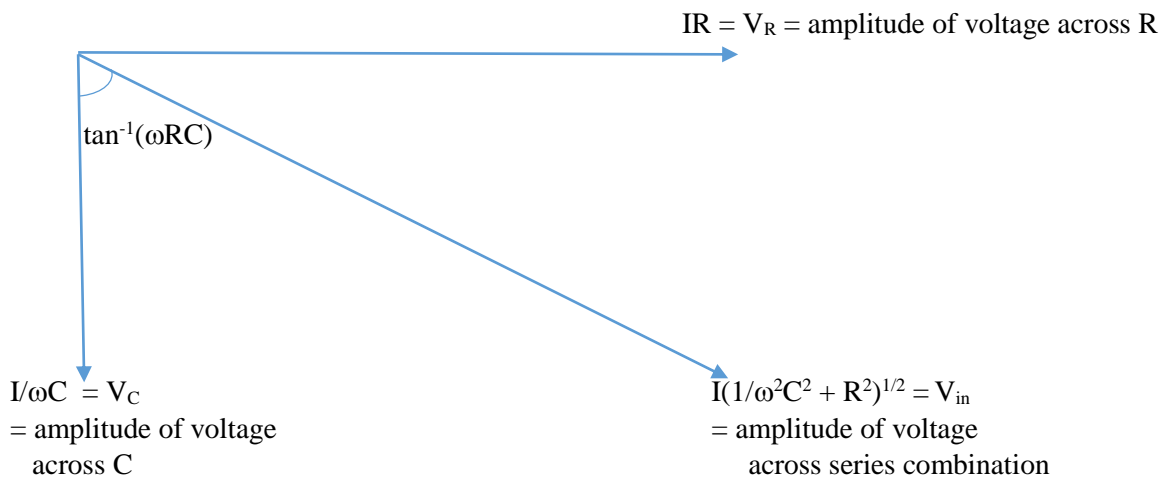
Evaluation of this method: This immediately gives V_{out}/V_{in} and the phase shift, but it works only for simple series circuits (such as the one we're analyzing right now).

Draw an **arrow to the right and call its length IR** . Draw an **arrow downward and call its length $I/\omega C$** . We immediately see that the “vector sum” of these arrows is $I(1/\omega^2 C^2 + R^2)^{1/2}$. This vector sum the amplitude of the voltage across the series combination, which is V_{in} . The downward arrow is the amplitude of the voltage across the capacitor, which is V_{out} . So we immediately obtain

$$V_{out}/V_{in} = (I/\omega C)/(I(1/\omega^2 C^2 + R^2)^{1/2}) = 1/(1 + \omega^2 R^2 C^2)^{1/2}.$$

Since we often just want this ratio, the current amplitude I is often omitted from the three arrows in the phasor diagram.

We immediately read the (magnitude of the) phase shift from the phasor diagram as well.



You might think of the phasor diagram as a visual mnemonic for the results we obtained from Method 1. If you want to take it a step further, imagine the whole diagram rotating counterclockwise with frequency ω . The projections of the arrows onto either a horizontal or vertical axis (your choice) give instantaneous voltages. See <http://www.animations.physics.unsw.edu.au/jw/AC.html>.

We can use the phasor diagram without really understanding why it works. But let's try to understand it.

We know that IR is the amplitude of the voltage across a resistor. But why is $I/\omega C$ the amplitude of the voltage across the capacitor? If

$$q = Q\sin(\omega t + \phi),$$

then

$$i = dq/dt = \omega Q\cos(\omega t + \phi),$$

thus current amplitude

$$I = \omega Q.$$

So

$$V_C = Q/C \text{ (the basic capacitor equation)}$$

$$V_C = I/\omega C.$$

Why is $I(1/\omega^2 C^2 + R^2)^{1/2}$ equal to the input amplitude? In Method 1, we found

$$Q = V_{in}/(1/C^2 + \omega^2 R^2)^{1/2},$$

so

$$V_{in} = Q(1/C^2 + \omega^2 R^2)^{1/2} = I(1/\omega^2 C^2 + R^2)^{1/2},$$

simply by substituting $Q = I/\omega$.

Method 3. Complex impedances (Ohm's law generalized)

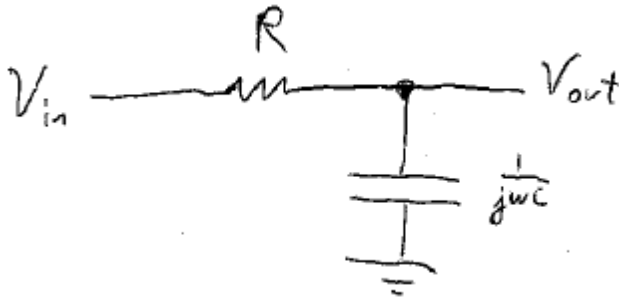
Evaluation of this method: This is often the best method for circuits of any complexity, but only for sinusoidal (or at least periodic) input.

- Write the complex impedance for each R, C, or L in the circuit:
 - The impedance of a resistor is \mathbf{R} .
 - The impedance of a capacitor is $\mathbf{1/j\omega C}$. ($j \equiv \sqrt{-1}$)
 - The impedance of an inductor is $\mathbf{j\omega L}$.
- Ohm's law generalized: If a current i flows through,
 - the voltage across a resistor is \mathbf{iR} .
 - the voltage across a capacitor is $\mathbf{i/j\omega C}$.
 - the voltage across an inductor is $\mathbf{ij\omega L}$.
- All impedances combine like resistors: a series combination is $Z_1 + Z_2$, a parallel combination is $1/(1/Z_1 + 1/Z_2)$.
- Use Ohm's law (and maybe Kirchhoff's current law) to find v_{out}/v_{in} . This will be a complex number (unless you have no C or L in your circuit).

- Find the magnitude of this complex number: $\sqrt{\left(\frac{v_{out}}{v_{in}}\right)\left(\frac{v_{out}}{v_{in}}\right)^*}$

- This magnitude is the amplitude ratio we want: $\frac{V_{out}}{V_{in}} = \sqrt{\left(\frac{v_{out}}{v_{in}}\right)\left(\frac{v_{out}}{v_{in}}\right)^*}$

Let's apply this procedure to our simple RC circuit.



This is really just a voltage divider since we get to treat $1/j\omega C$ as a resistor. We could skip to the result, but let's derive it.

The same current i flows through R and $1/j\omega C$. Applying Ohm's law to $1/j\omega C$, we see

$$v_{out} = i/j\omega C.$$

Applying Ohm's law to the series combination, we see

$$v_{in} = i(R + 1/j\omega C)$$

Combining the equations, we obtain

$$v_{out}/v_{in} = 1/(j\omega RC + 1).$$

This is a complex number, and we want to know its magnitude, so we multiply by its complex conjugate and take the square root:

$$\{[1/(j\omega RC + 1)][1/(-j\omega RC + 1)]\}^{1/2} = 1/(\omega^2 R^2 C^2 + 1)^{1/2}.$$

This is V_{out}/V_{in} , just as we found in the two previous methods.

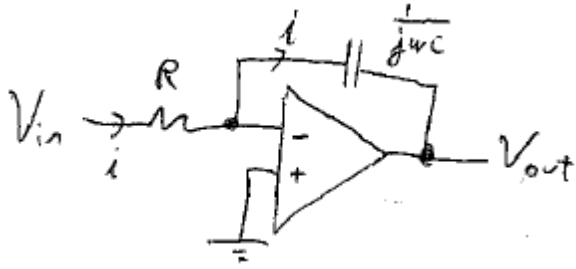
We will also use a fourth method (Laplace transforms) to obtain this result. But first, we'll practice using Method 3 with other circuits.

The rules for series and parallel **resistor** combinations are the opposite of the rules for series and parallel combinations of **capacitors**. So why do all complex impedance combine like resistors?

Lecture 12: Complex impedances and frequency response

Let's use the complex-impedance method to determine V_{out}/V_{in} for other circuits. This ratio is so useful that we have two names for it. We call it G , for "gain." Since G depends on frequency, we also call it "frequency response."

Let's determine the gain, $G = V_{out}/V_{in}$, for an integrator circuit.



Applying Ohm's law to the resistor:

$$i = (v_{in} - 0)/R.$$

Applying Ohm's law to the capacitor:

$$i = (0 - v_{out})/(1/j\omega C) = -j\omega C v_{out}.$$

Combining these equations:

$$v_{out}/v_{in} = -1/j\omega RC.$$

(We could have arrived here immediately by recognizing the circuit as an inverting amplifier with $R_1 = R$ and $R_2 = 1/j\omega C$.)

The magnitude of v_{out}/v_{in} is

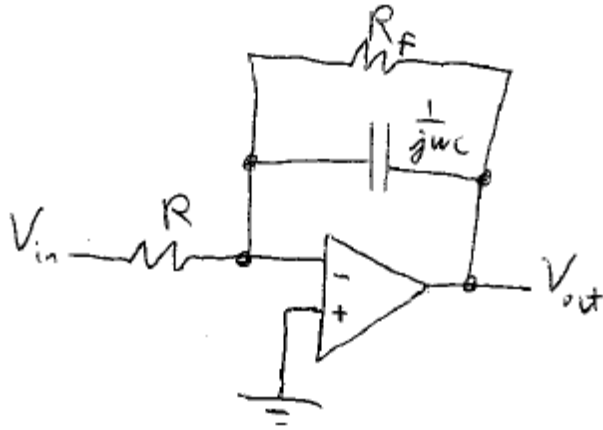
$$G = V_{out}/V_{in} = 1/\omega RC.$$

So the gain is inversely proportional to frequency. We can understand this conceptually. The greater the frequency, the less time the capacitor has to charge; the input voltage reverses before the capacitor accumulates much charge. Since the output voltage is the capacitor's voltage, the output voltage is small when the input frequency is high.

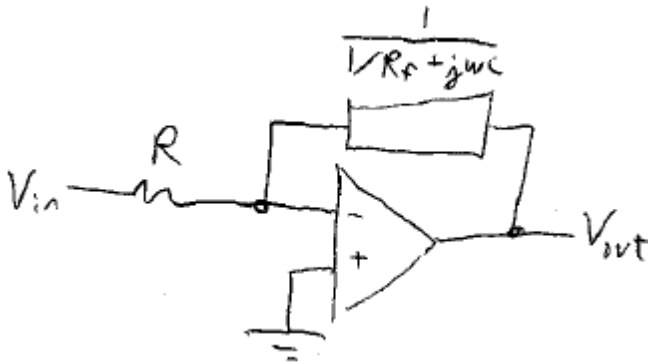
Conversely, when the input frequency is low, the capacitor has a lot of time to charge. And we see that when frequency is 0 (constant, or DC, input voltage), the gain is infinite: the capacitor keeps charging forever (well, only until the op amp output reaches a supply voltage).

Now, this is a problem for any input voltage: any input voltage is bound to have a DC "component." Even if the input is sinusoidal, and nearly centered on 0 V, in reality, it's probably not perfectly centered on 0 V, so the actual input is a sinusoid plus a small constant. This small constant causes the capacitor to charge more than discharge over each cycle, until finally the op amp output reaches a supply voltage, which we want to avoid.

How can we prevent this problem? In other words, how can we eliminate the infinite gain at $\omega = 0$? By putting a resistor in parallel with the capacitor:



The new resistor, R_f , is in parallel with the capacitor. We can form the parallel combination:



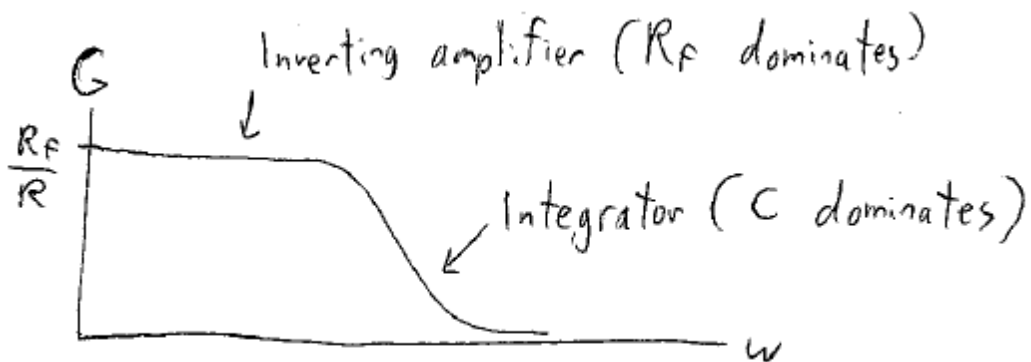
The circuit is now an inverting amplifier with $R_1 = R$ and $R_2 = 1/(1/R_f + j\omega C)$, so

$$V_{out}/V_{in} = -1/(R/R_f + j\omega RC)$$

and the magnitude of this is

$$G = V_{out}/V_{in} = 1/(R^2/R_f^2 + \omega^2 R^2 C^2)^{1/2}$$

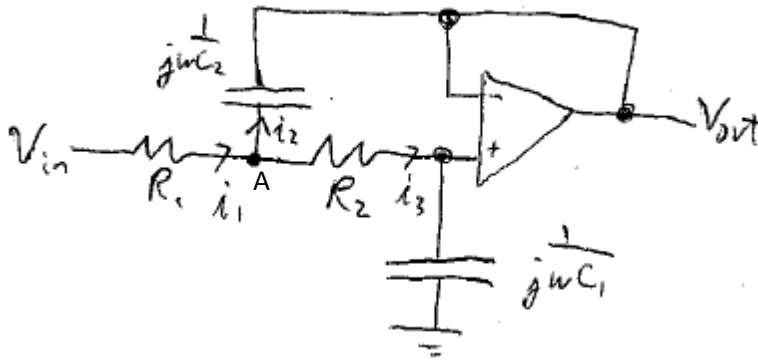
Let's plot the gain as a function of frequency (the frequency response):



We see that for low ω , $G \approx R_f/R$. How could we have predicted this without writing anything down? If we look at the circuit, we see that the capacitor's impedance is huge at low ω . R_f dominates the parallel combination; the capacitor is negligible, so the circuit looks like an inverting amplifier at low ω .

As high ω , $G \approx 1/\omega RC$, which is the gain for an integrator. We could have predicted this by recognizing that the impedance of the capacitor is very small at high ω ; the capacitor dominates, R_f has negligible effect, and the circuit looks like an integrator.

Now let's try a fancier circuit, called a Sallen-Key low-pass filter.



We're going to have a lot of unknowns, but we'll have enough equations to eliminate all the unknowns except v_{out} and v_{in} .

First, let's use Ohm's law (generalized) to find each of the three currents:

$$i_1 = (v_{in} - v_A)/R_1$$

$$i_2 = (v_A - v_{out})j\omega C_2$$

$$i_3 = v_{out}j\omega C_1$$

Kirchhoff's current law at point A lets us write

$$i_1 = i_2 + i_3$$

$$(v_{in} - v_A)/R_1 = (v_A - v_{out})j\omega C_2 + v_{out}j\omega C_1$$

How do we eliminate v_A ? We haven't yet used the fact that i_3 flows through R_2 as well as C_1 :

$$(v_A - v_{out})/R_2 = v_{out}j\omega C_1,$$

so

$$v_A = v_{out}(j\omega R_2 C_1 + 1).$$

Plug this into the equation that came from Kirchhoff's current law:

$$[v_{in} - v_{out}(j\omega R_2 C_1 + 1)]/R_1 = (j\omega R_2 C_1 v_{out})j\omega C_2 + v_{out}j\omega C_1$$

Nothing's left but algebra.

$$v_{in}/R_1 - v_{out}(j\omega R_2 C_1 + 1)/R_1 = -\omega^2 R_2 C_1 C_2 v_{out} + v_{out} j\omega C_1$$

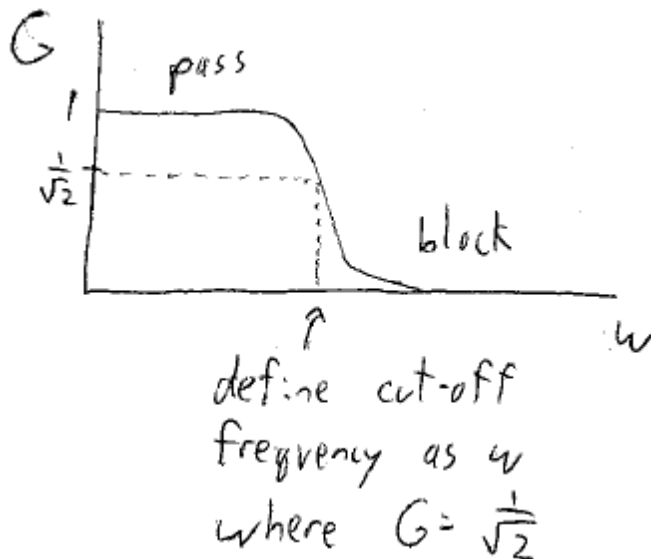
So

$$\begin{aligned} v_{out}/v_{in} &= 1/(j\omega R_2 C_1 + 1 + j\omega C_1 R_1 - \omega^2 R_1 R_2 C_1 C_2) \\ &= 1/[1 - \omega^2 R_1 R_2 C_1 C_2 + j\omega C_1 (R_1 + R_2)]. \end{aligned}$$

Multiply this by its complex conjugate and take the square root to find the magnitude:

$$G = 1/[(1 - \omega^2 R_1 R_2 C_1 C_2)^2 + \omega^2 C_1^2 (R_1 + R_2)^2]^{1/2}.$$

If we sketch this function, we obtain



We see that $G = 1$ at low frequency and $G = 0$ at high frequency. This is called a low-pass filter: low input frequencies pass through, but high frequencies are blocked. At what frequency does the transition occur, from “pass” to “block”? We define this transitional frequency as the frequency where $G = 1/2^{1/2}$. This frequency is called ω_{3dB} because a gain of $1/2^{1/2}$ is -3 decibels: $dB = 20\log G$.

Let's think about how to choose resistances and capacitances for a desired ω_{3dB} . For simplicity, let's choose $R_1 = R_2 = R$ and $C_1 = C_2 = C$. Now

$$\begin{aligned} G &= 1/[(1 - \omega^2 R^2 C^2)^2 + 4\omega^2 C^2 R^2]^{1/2} \\ &= 1/(1 + \omega^2 R^2 C^2) \end{aligned}$$

So

$$\omega = (G - 1)^{1/2}/RC$$

and

$$\omega_{3dB} = (2^{-1/2} - 1)^{1/2}/RC = 0.644/RC$$

Although angular frequency ω [rad/s] is mathematically convenient, in the lab we set and measure frequency f [Hz], so $f = \omega/2\pi$, and in this case,

$$f_{3dB} = 0.1/RC$$

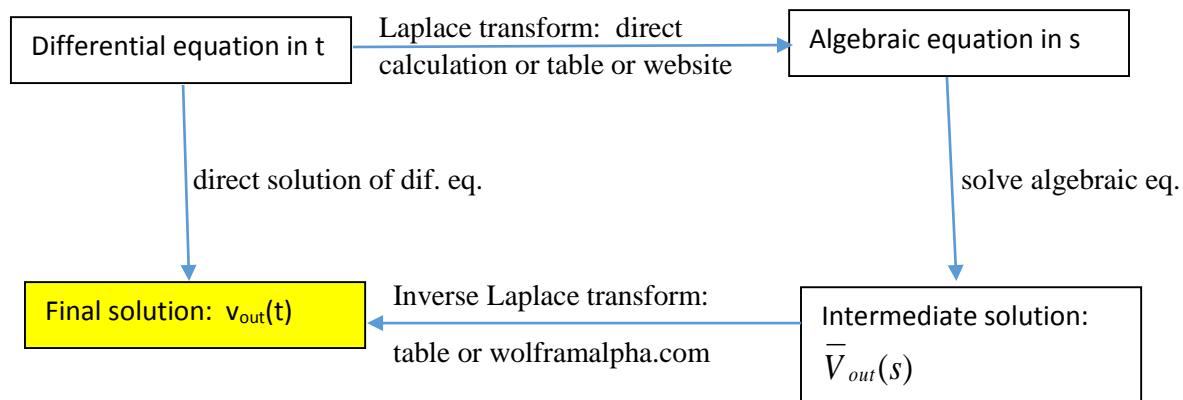
To set our cut-off frequency to 1 kHz, we could choose $R = 10 \text{ k}\Omega$ and $C = 0.01 \text{ }\mu\text{F}$.

The integrator is also a low-pass filter. Why is the Sallen-Key low-pass filter better?

Lecture 13. Laplace transforms

(Santiago Ch. 16)

The Laplace transform is a method for solving differential equations. It transforms a differential equation into an algebraic equation. This diagram compares the Laplace transform with the direct method of solving a differential equation:



The definition of a Laplace transform is

$$L[f(t)] \equiv \bar{F}(s) \equiv \int_0^{\infty} f(t)e^{-st} dt$$

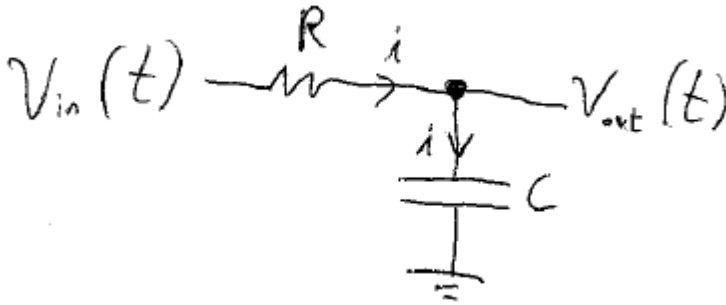
There's no new math in this definition; the Laplace transform is just an integral. The original variable t is integrated away. A new variable s appears. **s exists only in intermediate calculations**, kind of like the u in u substitution. Our final answer won't have s in it; the inverse Laplace transform gets rid of s and restores t . In this course, our method for finding inverse Laplace transforms will be to use a table or an online tool (wolframalpha.com). Analogously, we use calculators to determine $\cos 39^\circ$; we don't do it by hand. (There are lots of techniques for manipulating Laplace transforms until they're in a form found in a table. But I think it's more likely that you'll have internet access than a hard copy of a table of Laplace transforms. And since wolframalpha.com accepts Laplace transforms in any form, we won't spend time learning how to massage Laplace transforms into the form found in a table.)

Let's use Laplace transforms to analyze the RC circuit to which we already applied three methods. So this is

Method 4: Laplace transforms

Evaluation of this method: This method is not ideal for sinusoidal inputs. However, this method is useful for most non-sinusoidal inputs, especially voltage "spikes" (infinite voltages of infinitesimal duration: impulse functions).

We're applying Laplace transforms to the same RC circuit to make sure we get the same result we got three other ways:



Since we want to solve the same problem we solved before, the input voltage has to be sinusoidal. But remember, I was noncommittal about the phase of the input voltage, so this time, let $v_{in}(t) = V_{in}\cos(\omega t - \delta)$, where $\delta = \tan^{-1}(1/\omega RC)$. Let's further assume that the **capacitor is initially uncharged**, so $v_{out}(0) = 0$.

The output voltage plus the voltage across the resistor is the input voltage:

$$V_{in}\cos(\omega t - \delta) = iR + v_{out}$$

I want to write i in terms of v_{out} . v_{out} is the voltage across the capacitor, q/C , so

$$dv_{out}/dt = i/C. \text{ So}$$

$$V_{in}\cos(\omega t - \delta) = RC(dv_{out}/dt) + v_{out}$$

Now we have a differential equation that we want to solve, so we'll take the Laplace transform of both sides of this equation. The fast way to do this is to use a table of Laplace transforms. Here's a short table of the transforms that we need:

$f(t)$	$\bar{F}(s)$
$\cos(\omega t - \delta)$	$(\omega \sin \delta + s \cos \delta)/(s^2 + \omega^2)$
df/dt	$s\bar{F}(s) - f(0)$

Since the Laplace transform is an integral, the usual linearity properties apply:

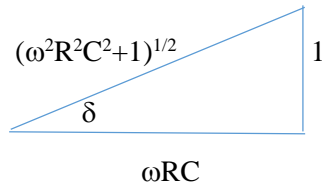
$$L[\alpha f(t) + \beta g(t)] = \alpha \bar{F}(s) + \beta \bar{G}(s).$$

Now we're ready to take the Laplace transform of

$$V_{in}\cos(\omega t - \delta) = RC(dv_{out}/dt) + v_{out}:$$

$$V_{in}(\omega \sin \delta + s \cos \delta)/(s^2 + \omega^2) = RCs\bar{V}_{out}(s) + \bar{V}_{out}(s),$$

where I used $v_{out}(0) = 0$. I can simplify this by evaluating $\sin \delta$ and $\cos \delta$. Since $\delta = \tan^{-1}(1/\omega RC)$ from the problem statement, I can draw the right triangle



So I see that $\sin\delta = 1/(\omega^2R^2C^2+1)^{1/2}$ and $\cos\delta = \omega RC/(\omega^2R^2C^2+1)^{1/2}$, and

$$[V_{in}/(\omega^2R^2C^2+1)^{1/2}]\omega(1 + sRC)/(s^2 + \omega^2) = RCs\bar{V}_{out}(s) + \bar{V}_{out}(s),$$

where the term in brackets is just a constant (no s).

The next step is to solve for $\bar{V}_{out}(s)$:

$$\bar{V}_{out}(s) = [V_{in}/(\omega^2R^2C^2+1)^{1/2}]\omega/(s^2 + \omega^2).$$

The final step is to take the inverse Laplace transform of this. I see in a table that

<u>f(t)</u>	<u>$\bar{F}(s)$</u>
$\sin(\omega t)$	$\omega/(s^2 + \omega^2)$

So the final result is

$$v_{out}(t) = [V_{in}/(\omega^2R^2C^2+1)^{1/2}]\sin(\omega t).$$

As before, the output amplitude is found to be a factor of $1/(\omega^2R^2C^2+1)^{1/2}$ times the input amplitude.

Now, let's practice using the definition of the Laplace transform just to demystify it; it really is just an integral. Let's return to our differential equation

$$V_{in}\cos(\omega t - \delta) = RC(dv_{out}/dt) + v_{out}$$

and take the Laplace transform the long way, by using $L[f(t)] \equiv \bar{F}(s) \equiv \int_0^{\infty} f(t)e^{-st} dt$. Multiplying by e^{-st}

and integrating both sides:

$$\int_0^{\infty} V_{in} \cos(\omega t - \delta) e^{-st} dt = \int_0^{\infty} \left[RC \frac{dv_{out}}{dt} + v_{out} \right] e^{-st} dt$$

$$V_{in} \int_0^{\infty} \cos(\omega t - \delta) e^{-st} dt = RC \int_0^{\infty} \frac{dv_{out}}{dt} e^{-st} dt + \int_0^{\infty} v_{out} e^{-st} dt$$

The first integral can be evaluated with the help of an online integral calculator. The final integral is the definition of $\bar{V}_{out}(s)$.

$$V_{in} \frac{(\omega \sin \delta + s \cos \delta)}{s^2 + \omega^2} = RC \int_0^{\infty} \frac{dv_{out}}{dt} e^{-st} dt + \bar{V}_{out}(s).$$

The remaining integral requires some work. We need to use integration by parts:

$$\int_0^{\infty} u dv = [uv]_0^{\infty} - \int_0^{\infty} v du$$

where

$$u = e^{-st}, \text{ so } du = -se^{-st} dt$$

and

$$dv = dv_{out}, \text{ so } v = v_{out}:$$

$$\begin{aligned} \int_0^{\infty} \frac{dv_{out}}{dt} e^{-st} dt &= [e^{-st} v_{out}]_0^{\infty} + \int_0^{\infty} v_{out} s e^{-st} dt \\ &= -v_{out}(0) + s \int_0^{\infty} v_{out} e^{-st} dt \\ &= s \bar{V}_{out}(s) \end{aligned}$$

So we've proven that the Laplace transform of our differential equation is

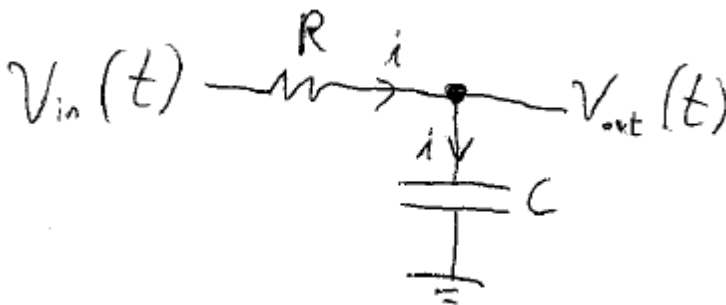
$$V_{in}(\omega \sin \delta + s \cos \delta)/(s^2 + \omega^2) = RCs \bar{V}_{out}(s) + \bar{V}_{out}(s).$$

How would the Laplace transform differ if the capacitor had an initial charge?

Lecture 14. Transfer functions

(Santiago Ch. 11)

Let's again consider the same circuit, though with a non-sinusoidal input:



Even without specifying $v_{in}(t)$, we see that

$$v_{in}(t) = iR + v_{out}$$

where the capacitor's charge $q = Cv_{out}$, so

$$i = C(dv_{out}/dt),$$

so

$$v_{in}(t) = RC(dv_{out}/dt) + v_{out}.$$

We can take the Laplace transform of this even though we haven't specified $v_{in}(t)$. Assuming the capacitor is initially uncharged, $v_{out}(0) = 0$, and

$$\bar{V}_{in}(s) = RCs\bar{V}_{out}(s) + \bar{V}_{out}(s):$$

Now define the transfer function

$$\bar{T}(s) = \frac{\bar{V}_{out}(s)}{\bar{V}_{in}(s)}.$$

The transfer function is defined for initial conditions of zero: capacitors are initially uncharged (and there's no initial current through any inductors).

The transfer function for our RC circuit is

$$\bar{T}(s) = \frac{1}{1 + sRC}$$

For any chosen $v_{in}(t)$, we can

- find the Laplace transform $\bar{V}_{in}(s)$ of $v_{in}(t)$
- find $\bar{V}_{out}(s) = \bar{T}(s)\bar{V}_{in}(s)$
- find $v_{out}(t)$, the inverse Laplace transform of $\bar{V}_{out}(s)$

Example: Step function

In the RC circuit we've been studying, suppose $v_{in}(t) = V_A u(t)$, where $u(t) = 0$ for $t < 0$, and $u(t) = 1$ for $t > 0$. $u(t)$ is called a step function. Determine $v_{out}(t)$.

Solution

First we use the definition of the Laplace transform to find $\bar{V}_{in}(s)$:

$$\bar{V}_{in}(s) \equiv \int_0^{\infty} v_{in}(t)e^{-st} dt = \int_0^{\infty} V_A e^{-st} dt = \frac{V_A}{s}$$

Next we multiply by the transfer function (which we already found) to obtain

$$\bar{V}_{out}(s) = \bar{T}(s)\bar{V}_{in}(s):$$

$$\bar{V}_{out}(s) = \frac{V_A}{s(1 + sRC)}.$$

Finally, take the inverse Laplace transform of this, using a table or wolframalpha.com:

$$v_{out}(t) = V_A(1 - e^{-t/RC})$$

This is the familiar RC charge curve: the voltage across the capacitor asymptotically approaches V_A as the capacitor charges up.

Example: Impulse function (voltage spike)

Considering the same circuit, suppose $v_{in}(t) = A\delta(t)$, where $\delta(t)$ is the impulse function (Dirac delta function). Roughly speaking, $\delta(0) = \infty$, and $\delta(t) = 0$ for all other t . The useful property of the impulse function is the sifting property: $\int_0^{\infty} f(t)\delta(t)dt = f(0)$. Determine $v_{out}(t)$.

Solution

First we use the definition of the Laplace transform to find $\bar{V}_{in}(s)$:

$$\bar{V}_{in}(s) \equiv \int_0^{\infty} v_{in}(t)e^{-st} dt = \int_0^{\infty} A\delta(t)e^{-st} dt = A$$

Next we multiply by the transfer function (which we already found) to obtain

$$\bar{V}_{out}(s) = \bar{T}(s)\bar{V}_{in}(s):$$

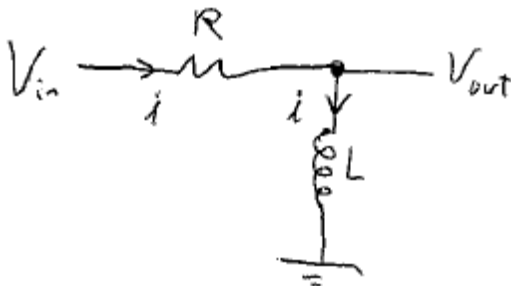
$$\bar{V}_{out}(s) = \frac{A}{(1 + sRC)}.$$

Finally, take the inverse Laplace transform of this, using a table or wolframalpha.com:

$$v_{out}(t) = Ae^{-t/RC}/RC$$

Example: RL circuit

Assuming initial current $i(0) = 0$, determine $v_{out}(t)$ if $v_{in}(t) = Au(t)$ and if $v_{in}(t) = A\delta(t)$.



Solution

First let's find the transfer function, by finding the equation relating input and output. The input voltage is the output voltage plus the voltage across the resistor:

$$v_{in}(t) = iR + v_{out}(t),$$

where

$$v_{out}(t) = L(di/dt),$$

so

$$i(t) = \frac{1}{L} \int_0^t v_{out}(t') dt', \text{ assuming } i(0) = 0.$$

The equation relating input and output voltages becomes

$$v_{in}(t) = \frac{R}{L} \int_0^t v_{out}(t') dt' + v_{out}(t).$$

Now we can take the Laplace transform, using the fact that $L[\int_0^t f(t') dt'] = \bar{F}(s)/s$:

$$\bar{V}_{in}(s) = \frac{R}{sL} \bar{V}_{out}(s) + \bar{V}_{out}(s).$$

The transfer function is

$$\bar{T}(s) = \frac{\bar{V}_{out}}{\bar{V}_{in}} = \frac{1}{1 + R/Ls}.$$

But it's even more convenient to solve for \bar{V}_{out} :

$$\bar{V}_{out}(s) = \frac{1}{1 + R/Ls} \bar{V}_{in}(s) = \frac{s}{s + R/L} \bar{V}_{in}(s)$$

If $v_{in}(t) = Au(t)$, we already know that its Laplace transform is A/s . Plug this into the equation above:

$$\bar{V}_{out}(s) = \frac{A}{s + R/L}.$$

The inverse Laplace transform (from a table or website) is

$$v_{out}(t) = Ae^{-Rt/L}.$$

Now we consider the other input, $v_{in}(t) = Atu(t)$. The Laplace transform of this (from a table, website, or direct calculation) is A/s^2 . In this case,

$$\bar{V}_{out}(s) = \frac{A}{s(s + R/L)}.$$

The inverse Laplace transform of this is

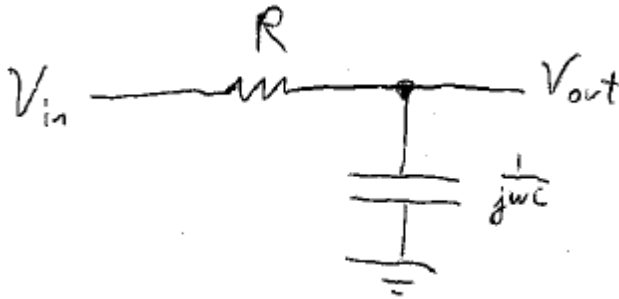
$$v_{out}(t) = \frac{AL}{R} (1 - e^{-Rt/L}).$$

How do things change if the initial current is not 0?

Lecture 15. A shortcut to the transfer function

(Santiago Ch. 18)

We've seen that we can find the transfer function for any circuit if we can write the differential or integral equation relating input and output voltages. There's a shortcut to the transfer function that's **valid whenever initial conditions are zero** (no initial charge in capacitors, no initial current in inductors). Consider once again the RC circuit that we've analyzed four ways:



We've seen that the method of complex impedances allows us to obtain

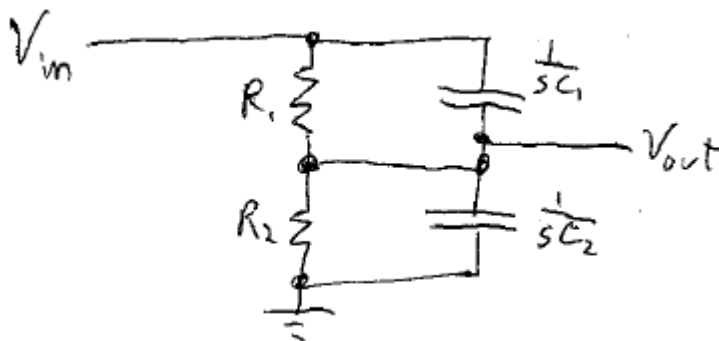
$$v_{out}/v_{in} = 1/(j\omega RC + 1).$$

We've also seen that the transfer function is

$$\bar{T}(s) = \frac{1}{1 + sRC}$$

We see that they're the same, if we make the substitution $s = j\omega$. In fact, this always works: To obtain the transfer function, obtain v_{out}/v_{in} from Ohm's law generalized, replacing $j\omega$ with s . In other words, **use $1/sC$ as the impedance of a capacitor, and use sL as the impedance of an inductor.**

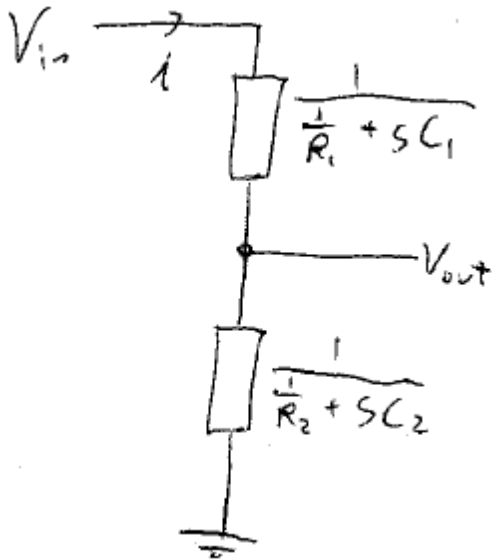
Example



Initial charges are 0. Determine $v_{out}(t)$ if $v_{in}(t) = Au(t)$.

Solution

First, let's form the parallel combinations:



Now we see that this is a voltage divider, so we can immediately (or gradually) obtain

$$\begin{aligned} v_{out}/v_{in} &= [1/(1/R_2 + sC_2)]/[1/(1/R_1 + sC_1) + 1/(1/R_2 + sC_2)] \\ &= 1/[(1/R_2 + sC_2)/(1/R_1 + sC_1) + 1] \\ &= (1/R_1 + sC_1)/(1/R_1 + sC_1 + 1/R_2 + sC_2) \end{aligned}$$

So this is the transfer function:

$$\bar{T}(s) = (1/R_1 + sC_1)/(1/R_1 + sC_1 + 1/R_2 + sC_2).$$

We multiply by $\bar{V}_{in}(s) = A/s$ to obtain

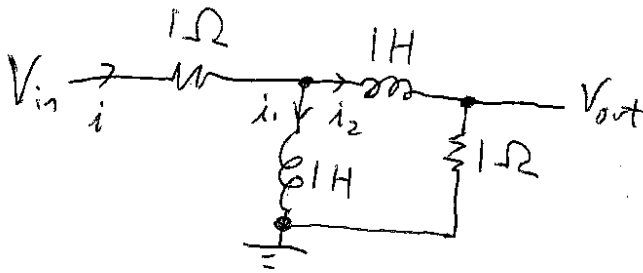
$$\bar{V}_{out}(s) = A(1/R_1 + sC_1)/[s(1/R_1 + sC_1 + 1/R_2 + sC_2)].$$

The inverse Laplace transform of this is messy but can be obtained from wolfram alpha:

$$v_{out}(t) = \frac{AC_1}{C_1 + C_2} e^{-t/(R_1 + 1/R_2)/(C_1 + C_2)} + \frac{AR_2}{(R_1 + R_2)} \left(1 - e^{-t/(R_1 + 1/R_2)/(C_1 + C_2)} \right)$$

Example

In the following figure, initial current i_1 (down) is 1 A and i_2 (to the right) is 1 A. Determine $v_{out}(t)$ if $v_{in}(t) = (2 \text{ V})u(t)$.



Solution

We can't use the shortcut to the transfer function because the initial currents are nonzero. So we have to find the differential equation the long way.

Kirchhoff's current rule:

$$i = i_1 + i_2$$

Kirchhoff's voltage rule on the right:

$$di_1/dt - di_2/dt - i_2 = 0,$$

using the fact that i_2 flows through the resistor.

Kirchhoff's voltage rule on the implied loop on the left:

$$v_{in} - i - di_1/dt = 0$$

Take the Laplace transform of all three of these equations:

$$\bar{I} = \bar{I}_1 + \bar{I}_2$$

$$s\bar{I}_1 - i_1(0) - (s\bar{I}_2 - i_2(0)) - \bar{I}_2 = 0$$

$$\bar{V}_{in} - \bar{I} - (s\bar{I}_1 - i_1(0)) = 0$$

The initial currents are both 1 A, so the three equations become

$$\bar{I} = \bar{I}_1 + \bar{I}_2$$

$$s\bar{I}_1 - s\bar{I}_2 - \bar{I}_2 = 0$$

$$\bar{V}_{in} - \bar{I} - (s\bar{I}_1 - 1) = 0$$

We have three equations in three unknowns. Ultimately we want i_2 because $v_{out} = i_2 \times (1 \Omega)$. Plug the first equation into the third:

$$\bar{V}_{in} - \bar{I}_1 - \bar{I}_2 - (s\bar{I}_1 - 1) = 0$$

Solve for \bar{I}_1 from the middle equation:

$$\bar{I}_1 = \bar{I}_2(s+1)/s$$

Plug this into $\bar{V}_{in} - \bar{I}_1 - \bar{I}_2 - (s\bar{I}_1 - 1) = 0$:

$$\bar{V}_{in} - \bar{I}_2(s+1)/s - \bar{I}_2 - (\bar{I}_2(s+1) - 1) = 0$$

Solve for \bar{I}_2 :

$$\bar{I}_2 = \frac{1 + \bar{V}_{in}}{1 + 1/s + 1 + s + 1} = \frac{1 + \bar{V}_{in}}{1/s + s + 3}$$

Since $v_{in}(t) = (2 \text{ V})u(t)$, $\bar{V}_{in} = 2/s$, and

$$\bar{I}_2 = \frac{1 + 2/s}{1/s + s + 3} = \frac{s + 2}{1 + s^2 + 3s}$$

Using wolframalpha.com to do the inverse Laplace transform,

$$i_2(t) = \frac{-e^{\left(-\frac{3}{2} - \frac{\sqrt{5}}{2}\right)t} + \sqrt{5} e^{\left(-\frac{3}{2} - \frac{\sqrt{5}}{2}\right)t} + e^{\left(\frac{\sqrt{5}}{2} - \frac{3}{2}\right)t} + \sqrt{5} e^{\left(\frac{\sqrt{5}}{2} - \frac{3}{2}\right)t}}{2\sqrt{5}}$$

which is numerically the same as $v_{out}(t) = i_2(t) \times (1 \Omega)$. Doing the arithmetic,

$$v_{out}(t) = (0.774 \text{ V})e^{-2.618t} + (0.276 \text{ V})e^{-0.382t}$$

where t is in seconds.

Why was the Laplace transfer of Kirchhoff's current rule so simple? Why was there no explicit s in the equation?

Lecture 16. Chaos in electronics

The third problem in Problem Set 16 is so complicated that I want to spend this whole lecture on it. The problem is to simulate V_1 vs. V_2 in this chaotic circuit described in "Simple Autonomous Chaotic Circuits" by Piper and Sprott:

$$V_2 - V_1 = L \, dI_4/dt, \text{ so}$$

$$dI_4/dt = (V_2 - V_1)/L. \quad (4)$$

$$I_5 = -V_2/R. \quad (5)$$

The final goal, algebraically, is to find **an equation containing a single voltage and its time derivatives**; this is the differential equation we have to solve numerically (there are no exact analytical solutions). Which voltage should we choose? We should choose V_2 because V_3 can be written only in terms of V_2 . Specifically, the upper op amp doesn't obey the voltage Golden Rule, so when $V_+ > V_-$, its output increases as much as it can (approximately to the positive supply voltage, which is 9 V if we use 9 V batteries), and when $V_- > V_+$, its output decreases as much as it can (approximately to -9 V). And since $V_+ = 0$ and $V_- = V_2$, $V_3 = 9$ V when $V_2 < 0$, and $V_3 = -9$ V when $V_2 > 0$. More concisely,

$$V_3 = -k \operatorname{sgn}(V_2),$$

where $k \approx 9$ V and $\operatorname{sgn}(V_2) = 1$ when $V_2 > 0$, and $\operatorname{sgn}(V_2) = -1$ when $V_2 < 0$.

Let's plug Eqs. (1)-(5) into Eqs. (A) and (B) to eliminate currents. Actually, we need to differentiate Eq. (B) to use Eq. (4).

Plug (1), (3), and (5) into (A):

$$V_3/R_C = -C_1 \, dV_1/dt - V_2/R. \quad (6)$$

Plug (2), (4), and (5) into d(B)/dt:

$$-(dV_2/dt)/R = C_2 \, d^2V_2/dt^2 + (V_2 - V_1)/L \quad (7)$$

Next I want to combine these equations to eliminate V_1 . From Eq. (6),

$$dV_1/dt = -V_2/RC_1 - V_3/R_C C_1.$$

Plug this into d(7)/dt:

$$-(d^2V_2/dt^2)/R = C_2 \, d^3V_2/dt^3 + (dV_2/dt)/L + V_2/LRC_1 - V_3/LR_C C_1$$

Finally, I can use $V_3 = -k \operatorname{sgn}(V_2)$ to obtain the final result:

$$d^3V_2/dt^3 = -(d^2V_2/dt^2)/RC_2 - (dV_2/dt)/LC_2 + V_2/LRC_1 C_2 - k \operatorname{sgn}V_2/LR_C C_1 C_2$$

This equation has no exact (analytical) solution, so Laplace transforms aren't helpful. Instead, we have to use an approximation (numerical) technique to replace the derivatives with finite differences. The simplest technique is the Euler technique:

$$V_2(t+\Delta t) \approx V_2(t) + (dV_2/dt)\Delta t$$

Similarly,

$$(dV_2/dt)|_{t+\Delta t} \approx (dV_2/dt)|_t + (d^2V_2/dt^2)\Delta t$$

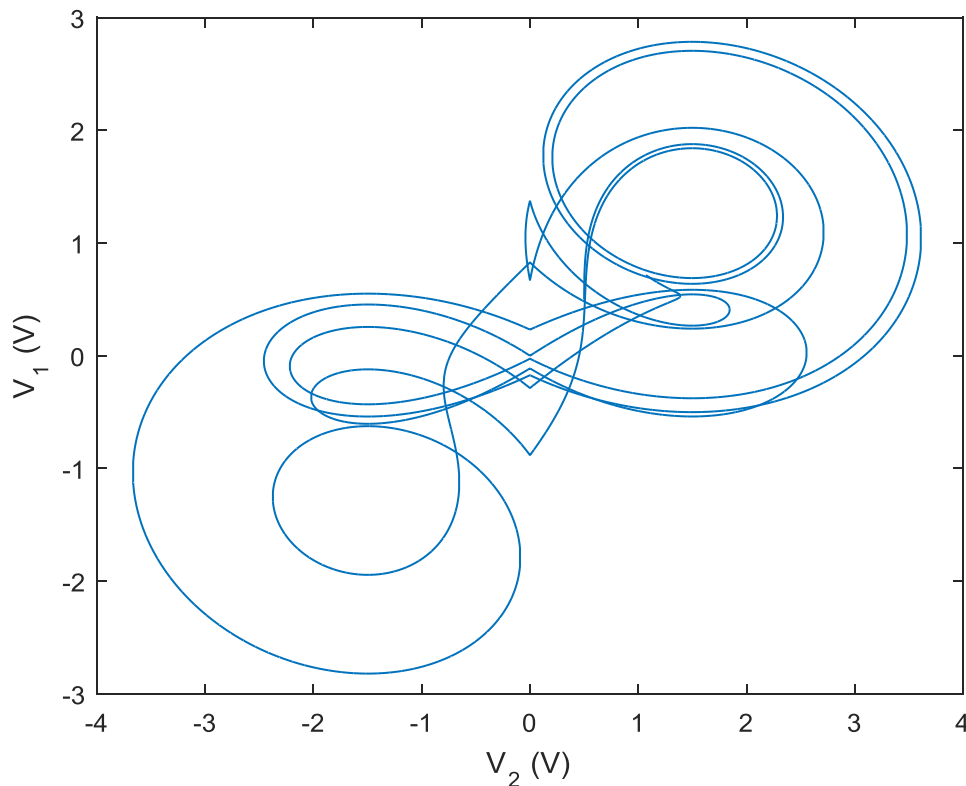
and

$$(dV_2^2/dt^2)|_{t+\Delta t} \approx (dV_2^2/dt^2)|_t + (d^3V_2/dt^3)\Delta t$$

d^3V_2/dt^3 is what we just derived! So if we know initial values of V_2 , dV_2/dt , and d^2V_2/dt^2 , Euler's method allows us to step forward in time:

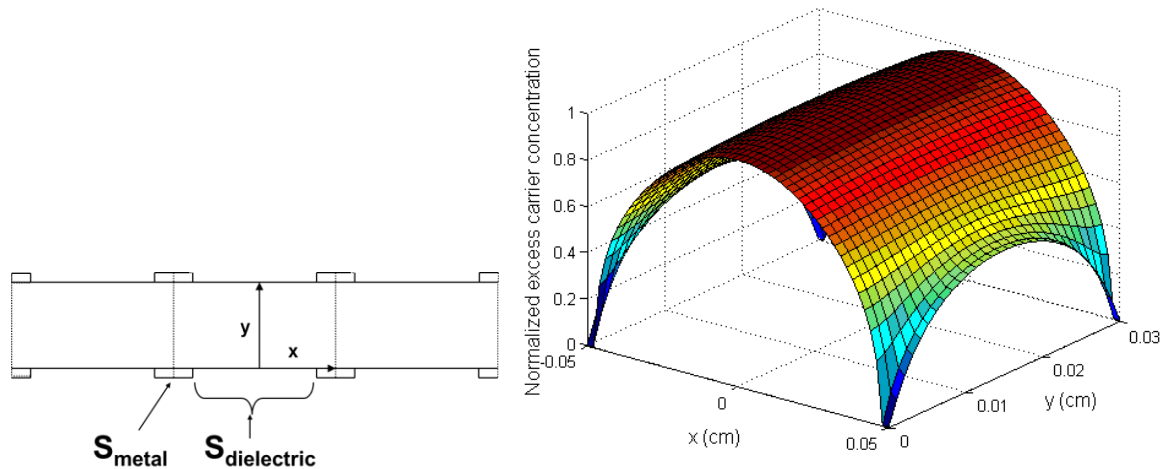
- Calculate d^3V_2/dt^3 from the equation we derived, using current values of V_2 , dV_2/dt , and d^2V_2/dt^2 .
- Determine the next value of d^2V_2/dt^2 , using $(dV_2^2/dt^2)|_{t+\Delta t} \approx (dV_2^2/dt^2)|_t + (d^3V_2/dt^3)\Delta t$.
- Determine the next value of dV_2/dt , using $(dV_2/dt)|_{t+\Delta t} \approx (dV_2/dt)|_t + (d^2V_2/dt^2)\Delta t$.
- Determine the next value of V_2 , using $V_2(t+\Delta t) \approx V_2(t) + (dV_2/dt)\Delta t$.
- These "next values" are now our current values. Return to the first step and iterate as long as desired.

How do you know how small Δt has to be? I explored this empirically by reducing Δt until the result basically stabilized.



Looking back at the circuit diagram, can you explain where I_1 and I_4 go? Why didn't I apply Kirchhoff's current rule to the junction at V_1 ?

UNIT 3: DESIGN OF SOLAR CELLS: SEMICONDUCTOR PHYSICS AND FURTHER APPLICATIONS OF FOURIER SERIES, LAPLACE TRANSFORMS, FINITE DIFFERENCE APPROXIMATIONS, AND ARDUINO!



We'll find that the skills we've learned have prepared us well for mathematical, computational, and experimental investigation of simple semiconductor structures.

Lecture 17: Semiconductors

(Honsberg and Bowden, Ch. 3.1)

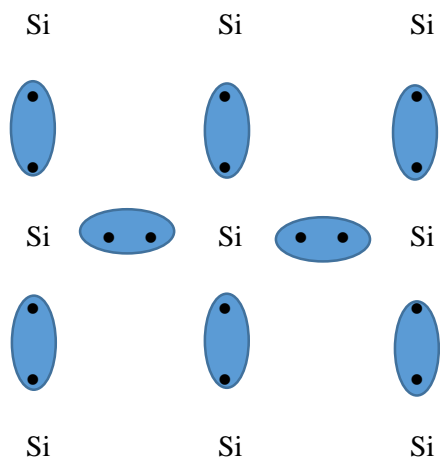
Silicon is a semiconductor: it's an insulator at low temperatures and a conductor at high temperatures.

It has four valence electrons in its outer shell:

-
- Si •
-

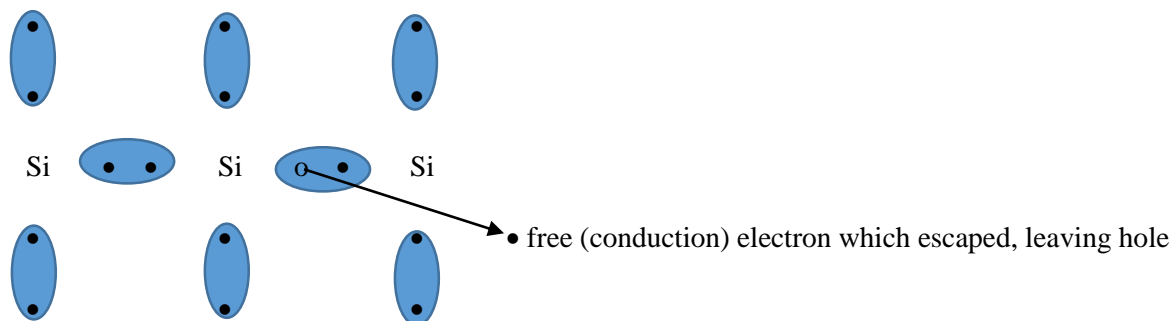
Actually, the electrons are arranged on the corners of a tetrahedron; they're not all in the same plane, but we draw it that way.

So, each silicon atom likes to bond with four others:

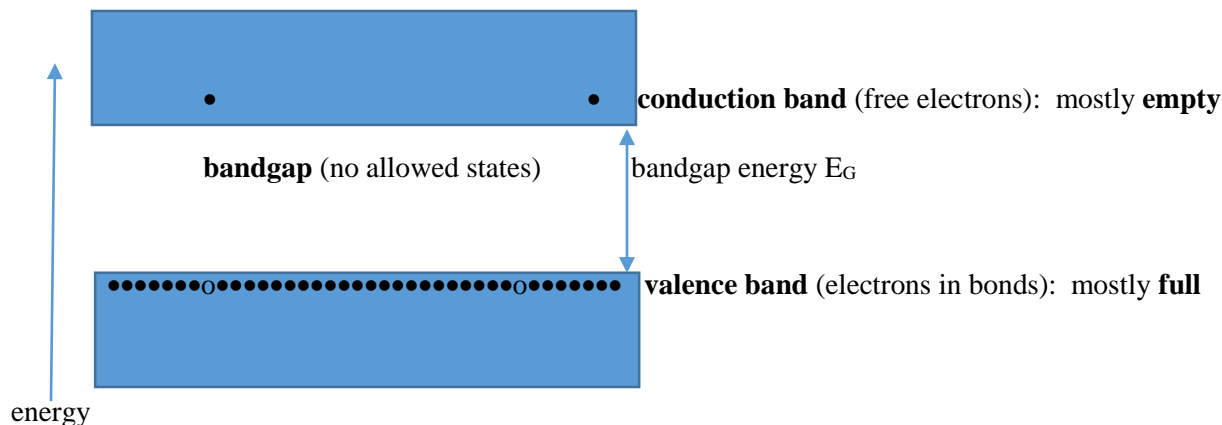


At a temperature of 0 K, all the valence electrons are stuck in the valence bands; no current can flow.

As temperature increases, some electrons acquire enough energy to escape the bonds. These free electrons leave behind “holes,” which behave as positive charges that can move around:



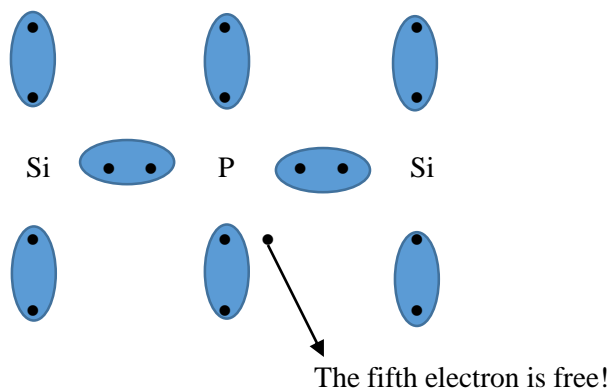
Free electrons can conduct electricity. These electrons are said to be in the “**conduction band**.” The electrons that are stuck in bonds are in the **valence band**. Free electrons have more energy than the electrons in bonds, so the conduction band is at higher energies than the valence band. This is represented in a band diagram:



How many free electrons are there in pure silicon? This is the **intrinsic carrier concentration**, n_i . n_i increases with temperature. At room temperature, $n_i = 8.3 \times 10^9 \text{ cm}^{-3}$. (This seems like a lot, but the number of silicon atoms per cubic centimeter is larger by a factor of almost 10^{13} . So an extremely small fraction of atoms contributes free electrons to the conduction band.)

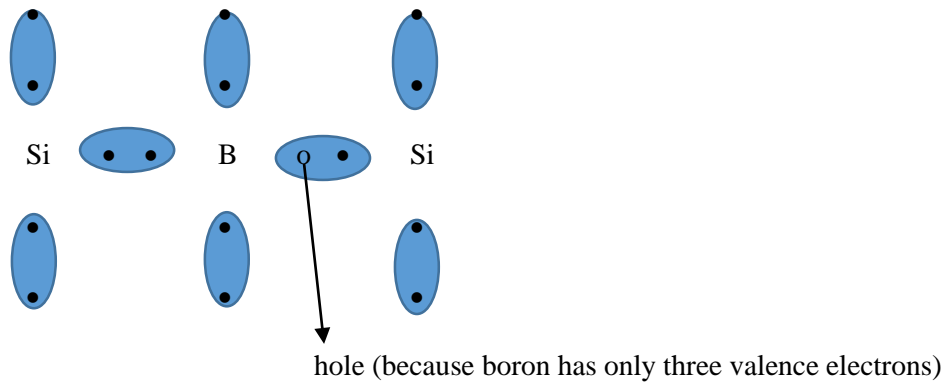
Semiconductors are useful because we can precisely engineer the concentration of free electrons and holes by **adding small amounts of another element**. This process is called **doping**.

In **n-type doping**, we add an element like phosphorous that has more than four valence electrons. Phosphorous has five valence electrons:



Phosphorous in this case is called a **donor** because it donates a free electron. The concentration of donors is called N_D . Except at extremely low temperatures, each donor atom donates one free electron. Assuming $N_D \gg n_i$, the free electron concentration n_0 equals the donor concentration: $n_0 = N_D$. In n-type silicon, electrons are called **majority** carriers, and holes are called **minority** carriers.

In **p-type doping**, we add an element like boron that has fewer than four valence electrons. Boron has three valence electrons:



Boron in this case is called an acceptor because it can accept an electron from another atom, effectively moving a hole to the other atom. The concentration of acceptors is called N_A . Except at extremely low temperatures, each acceptor atom produces one hole. Assuming $N_A \gg n_i$, the hole concentration p_0 equals the acceptor concentration: $p_0 = N_A$. In p-type silicon, holes are called **majority** carriers, and electrons are called **minority** carriers.

If we know the doping level (N_D or N_A), we know the concentration of majority carriers. But how do we determine the concentration of minority carriers? It turns out, in equilibrium it's always true that

$$n_0 p_0 = n_i^2$$

This is exactly analogous to a fact you may have learned in chemistry: $[H_3O^+][OH^-] = (10^{-7} \text{ M})^2$.

So for an **n-type** semiconductor:

$$n_0 = N_D \text{ and } p_0 = n_i^2 / N_D$$

And for a **p-type** semiconductor:

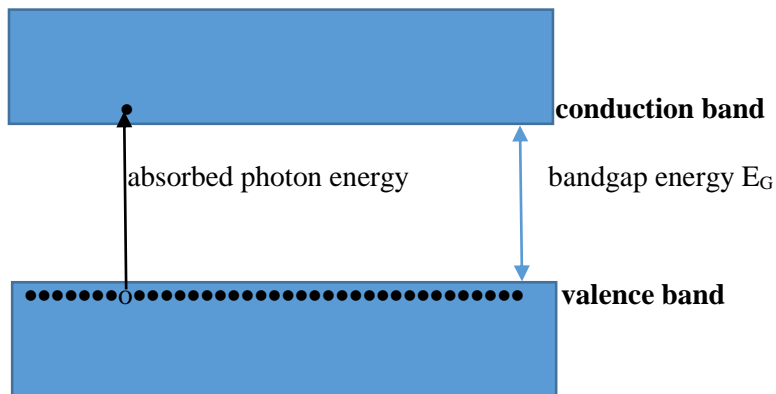
$$p_0 = N_A \text{ and } n_0 = n_i^2 / N_A$$

What do you suppose happens if a material is doped with both donors and acceptors, but there are far more of one than the other?

Lecture 18: Generation, recombination, and diffusion

(Honsberg and Bowden, Ch. 3.2 and 3.3)

If a semiconductor absorbs a photon with larger energy than the bandgap energy, then the photon energy can be transferred to an electron in the valence band, boosting it to the conduction band:



A hole is left behind in the valence band, so **the absorption of a photon creates an electron-hole pair**. The concentrations of electrons and holes increase above equilibrium level:

$$\text{electron concentration } \mathbf{n = n_0 + \Delta n}$$

$$\text{hole concentration } \mathbf{p = p_0 + \Delta p}$$

where (for our purposes) $\Delta n = \Delta p$, and this is called the **excess carrier concentration**.

At low levels of illumination, such that the excess carrier concentration is less than the dopant concentration, only the minority carrier concentration increases significantly:

$$\text{n-type: } n = N_D + \Delta n \approx N_D, p = p_0 + \Delta p$$

$$\text{p-type: } p = N_A + \Delta p \approx N_A, n = n_0 + \Delta n$$

Consider the **photon flux** N_0 [$\text{cm}^{-2}\text{s}^{-1}$] of light shining on the surface of silicon. As we look deeper into the silicon, the photon flux decreases because some of the photons have already been absorbed. The photon flux decreases with depth according to

$$N(x) = N_0 e^{-\alpha x},$$

where α is called the **absorption coefficient**. α depends strongly on wavelength:

short wavelengths (blue light) have **large α** and are **absorbed near the surface**

long wavelengths (red light) have **small α** and are **partially absorbed throughout**

These facts are very significant in the design of solar cells, which are intended to absorb all wavelengths as efficiently as possible.

Now we're ready to define the **generation rate** G [$\text{cm}^{-3}\text{s}^{-1}$]: the rate at which electron-hole pairs are generated per unit volume. For monochromatic illumination (a single wavelength and therefore a single α), the generation rate depends on depth according to

$$G(x) = \alpha N_0 e^{-\alpha x}.$$

Electron-hole pairs are created by the absorption of photon energy. The electrons in the conduction band spontaneously drop back into the holes in the valence band. This process is called recombination. The **recombination rate** U [$\text{cm}^{-3}\text{s}^{-1}$] is proportional to excess carrier concentration Δn :

$$U = \Delta n / \tau = (\text{excess carrier concentration}) / (\text{minority carrier lifetime})$$

The **minority carrier lifetime** is the average time between the generation of an electron-hole pair and its spontaneous recombination. τ can be different for electrons and holes and so is often specified as τ_n or τ_p . There are different recombination mechanisms that reduce lifetime. In crystalline silicon solar cells, much recombination occurs through defects: energy levels in the bandgap due to imperfections in the crystal structure. A big challenge in solar cell engineering is to reduce defects as cheaply as possible, or to minimize their impact.

Free electrons and holes can move around. In the absence of electric fields, they move around randomly. **Random motion** causes a **net movement** of charge carriers **from high concentration to low concentration**. This is called **diffusion**. The **diffusion length** is the average distance an electron or hole travels between generation and recombination:

$$L = \sqrt{D\tau},$$

where D is the **diffusivity** [cm^2/s]. Since D and τ can be different for electrons and holes, L is also different for electrons and holes. We specify the carrier type with subscripts:

$$L_n = \sqrt{D_n \tau_n} \quad \text{and} \quad L_p = \sqrt{D_p \tau_p}$$

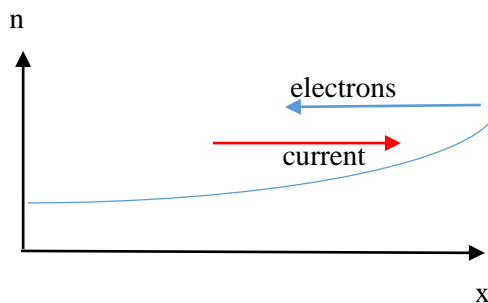
The movement of charge is current. The **diffusion current density** [$\text{C cm}^{-2} \text{s}^{-1}$] is given as

$$J_n = qD_n(dn/dx) \quad \text{for electrons}$$

and

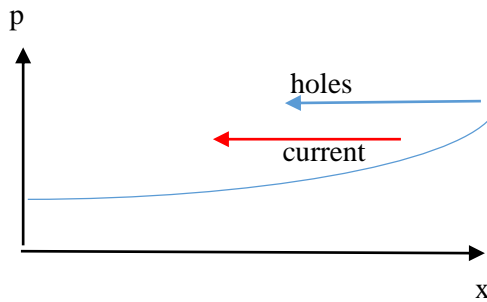
$$J_p = -qD_p(dp/dx) \quad \text{for holes.}$$

Why is there a minus sign in one case but not the other? Let's first consider an electron concentration that increases in the positive x direction:



Electrons spontaneously diffuse from high concentration (high x) to low concentration (low x). Since electrons have a negative charge, conventional current is in the opposite direction as the electrons' motion. Therefore, current is in the positive direction and should have a positive value. Since dn/dx is positive, $qD_n(dn/dx)$ is positive as needed.

Now consider a hole concentration that increases in the positive x direction:



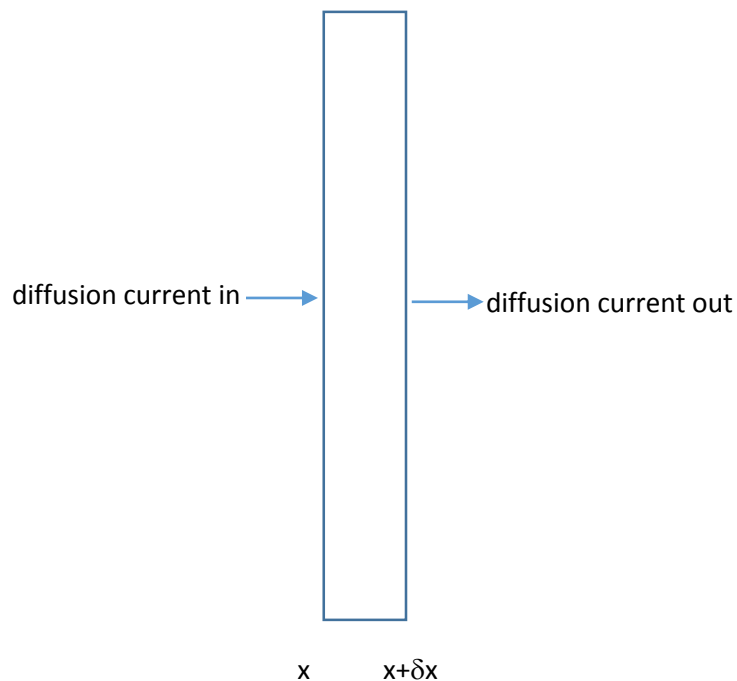
Holes spontaneously diffuse from high concentration (high x) to low concentration (low x). Since holes have a positive charge, conventional current is in the same direction as the holes' motion. Therefore, current is in the negative direction and should have a negative value. Since dp/dx is positive, $qD_p(dp/dx)$ is positive, which is why we need to add a minus sign for hole diffusion current.

Would you expect a very thin solar cell to have a better efficiency under red or blue illumination?

Lecture 19: The minority-carrier diffusion equation

(Honsberg and Bowden, Ch. 3.4)

Now we're ready to derive the minority-carrier diffusion equation. The derivation is much like the derivation of the heat equation! As before, we consider a very tiny volume of thickness δx and cross-section area A (perpendicular to the diagram):



Let's take the case of n-type silicon, so we'll derive the minority-carrier diffusion equation for holes. We simply have to keep track of the change in the number of holes in our volume:

- Some holes diffuse through the left side at x . They diffuse in if $J_p(x) > 0$.
- Some holes diffuse through the right side at $x+\delta x$. They diffuse out if $J_p(x+\delta x) > 0$.
- Some holes may be generated within the volume ($A\delta x$)
- Some holes may recombine within the volume ($A\delta x$).

We need an algebraic expression for each term.

- $J_p(x)$ is charge per unit area per unit time. We want the number of holes per unit time flowing through the left edge, which is $AJ_p(x)/q$.
- Similarly the number of holes flowing through the right edge is $AJ_p(x+\delta x)/q \approx (A/q)[J_p(x) + (dJ_p/dx)\delta x]$.
- The number of holes generated per unit time in the tiny volume is G times the volume: $GA\delta x$.
- Similarly, the numbers of holes recombining per unit time is $UA\delta x = (\Delta p/\tau_p)A\delta x$.

So

rate of change in number of holes = rate of holes flowing in – rate of holes flowing out

+ rate of generation of holes – rate of recombination of holes

where the number of holes is the hole concentration times the volume:

$$\partial(pA\delta x)/\partial t = AJ_p(x)/q - (A/q)[J_p(x) + (\partial J_p/\partial x)\delta x] + GA\delta x - (\Delta p/\tau_p)A\delta x$$

Divide by $A\delta x$:

$$\partial p/\partial t = -(\partial J_p/\partial x)/q + G - \Delta p/\tau_p$$

Now use the expression for diffusion current, $J_p = -qD_p(\partial p/\partial x)$:

$$\partial p/\partial t = D_p(\partial^2 p/\partial x^2) + G - \Delta p/\tau_p$$

This is the **minority carrier diffusion equation** for holes. Since $p = p_0 + \Delta p$, and p_0 is a constant, we can write the diffusion equation in terms of Δp and its derivatives:

$$\partial \Delta p/\partial t = D_p(\partial^2 \Delta p/\partial x^2) + G - \Delta p/\tau_p$$

Similarly, for electrons in p-type material,

$$\partial \Delta n/\partial t = D_n(\partial^2 \Delta n/\partial x^2) + G - \Delta n/\tau_n$$

Notice how similar this is to the heat equation! G is analogous to internal energy generation. The recombination term, however, has no counterpart in the heat equation.

Example

Consider p-type silicon with an acceptor concentration of 10^{16} cm^{-3} . The sample is illuminated with infrared light such that there is a uniform generation of $10^{10} \text{ cm}^{-3}\text{s}^{-1}$. There is a uniform electron concentration of 10^6 cm^{-3} . What is the minority carrier lifetime? At $t = 0$, the light is blocked, so the generation rate becomes 0. Assume electron concentration remains uniform. Find $n(t)$.

Solution

We start with the diffusion equation, where the derivatives are 0 because the electron concentration is uniform and also constant (while the light shines on the silicon):

$$0 = G - \Delta n / \tau_n$$

In steady-state, generation equals recombination and $\tau_n = \Delta n / G = (n - n_0) / G \approx n / G = 10^{-4} \text{ s}$. (I knew $n \gg n_0$ because $n_0 = n_i^2 / N_A$, and I approximated $n_i \approx 10^{10} \text{ cm}^{-3}$ to find $n_0 \approx 10^4 \text{ cm}^{-3}$.)

At $t = 0$, the light is blocked, so $G = 0$ and Δn must change, so the time derivative is no longer 0:

$$d\Delta n / dt = -\Delta n / \tau_n$$

The solution to this differential equation is $\Delta n = A \exp(-t / \tau_n)$, where A is a constant determined by the initial condition: $\Delta n(0) = 10^6 \text{ cm}^{-3} = A$.

Example

In thick p-type silicon in the dark, electrons arrive from the left at $x = 0$ such that $n(0) = 10^{12} \text{ cm}^{-3}$ in steady-state. Find $n(x)$ and $J_n(x)$ for positive x .

Solution

We start with the minority carrier diffusion equation, with the time derivative zero because we're in steady state. Generation rate is also 0.

$$0 = D_n (d^2 \Delta n / dx^2) - \Delta n / \tau_n$$

So Δn is proportional to its second derivative. The solution is

$$\Delta n = A e^{kx} + B e^{-kx},$$

where A , B , and k are constants that need to be determined. We can immediately set $A = 0$ because we don't want infinite Δn at infinite x . (We're allowed to do this **only** because we're told that the silicon is **thick**. If the silicon were thin, we'd need a boundary condition for the surface.) Plugging $\Delta n = B e^{-kx}$ into the differential equation yields

$$k^2 = 1 / (D_n \tau_n), \text{ so } k = 1 / L_n \text{ and } \Delta n = B \exp(-x / L_n).$$

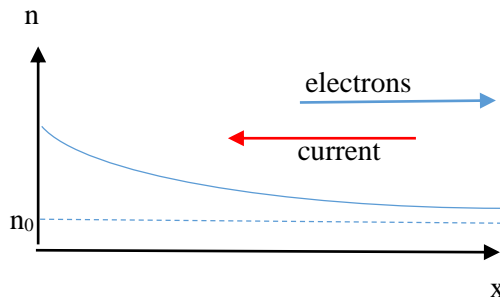
Now we just need to find B from the boundary condition, $n(0) = n_0 + \Delta n(0) \approx \Delta n(0) = 10^{12} \text{ cm}^{-3} = B$.

$$\text{So } \Delta n = (10^{12} \text{ cm}^{-3})\exp(-x/L_n)$$

and

$$n = n_0 + \Delta n = n_0 + (10^{12} \text{ cm}^{-3})\exp(-x/L_n)$$

n_0 and L_n are unknown, so we'll just leave them as algebraic symbols.



Far from $x = 0$ where excess electrons are arriving, the electron concentration is n_0 . Electrons recombine as they diffuse away from $x = 0$.

To find diffusion current, just plug n into the equation for J_n :

$$J_n = qD_n(dn/dx) = -(qD_n/L_n)(10^{12} \text{ cm}^{-3})\exp(-x/L_n)$$

Why does the magnitude of current decrease as x increases? Where does the lost current “go”? Why must the current through a metal wire remain constant, and what makes a semiconductor different?

Lecture 20: Nonhomogeneous differential equations

The minority carrier diffusion equation has generation rate G in it. G isn't proportional to excess carrier concentration or its derivatives; mathematicians say that G is a **nonhomogeneous** term. So far, we haven't solved any nonhomogeneous differential equations. But we're about to.

Example

Beginning at time $t = 0$, p-type silicon is illuminated with gradually dimming light such that there's a uniform generation rate $G(t) = G_0e^{-bt}$, where G_0 and b are constants. Find $\Delta n(t)$, assuming equilibrium at $t < 0$, and neglecting surface recombination. (Uniform generation combined with negligible surface recombination means that n is uniform.)

Solution: Laplace transform!

Write the minority carrier diffusion equation, where the spatial derivative is zero because n is uniform:

$$d\Delta n/dt = G_0e^{-bt} - \Delta n/\tau_n$$

Take the Laplace transform:

$$s\overline{\Delta N} = \frac{G_0}{s+b} - \frac{1}{\tau_n} \overline{\Delta N}$$

Solve for $\overline{\Delta N}$:

$$\overline{\Delta N} = \frac{G_0}{(s+b)(s+\tau_n)}$$

Take the inverse Laplace transform (using a table or Wolfram Alpha):

$$\Delta n = G_0 \frac{e^{-bt} - e^{-t/\tau_n}}{(1/\tau_n) - b}$$

If you plot this, you find that it starts at 0 (because the silicon is in equilibrium before the light comes on). Then it rises (because generation exceeds recombination). Then it decays to 0 (because generation decays to 0).

The Laplace transform made this problem easy, but some problems can be solved more easily by another method. So let's learn an alternative method.

Alternative solution: particular and complementary solutions

We can build the complete solution Δn as a sum of two functions: the **particular solution** Δn_p and the **complementary solution** Δn_c . How do we find these two solutions? Read on!

First we attempt to find a **particular solution** to the differential equation: we stare at the nonhomogeneous term and try to guess a solution:

$$d\Delta n/dt = G_0 e^{-bt} - \Delta n/\tau_n$$

Since the nonhomogeneous term is proportional to e^{-bt} , it seems likely that we can solve this equation by guessing that Δn is also proportional to e^{-bt} :

$$\Delta n_p = A e^{-bt}$$

where A is a constant that we'll need to determine, and I've used the subscript p to emphasize that this is called the particular solution.

Plug Δn_p into the differential equation:

$$-bAe^{-bt} = G_0 e^{-bt} - Ae^{-bt}/\tau_n$$

The exponential terms conveniently divide out, and we can solve for A :

$$A = \frac{G_0}{(1/\tau_n) - b}$$

Now we need the **complementary solution** Δn_c : this is the solution to the equation obtained by dropping the nonhomogeneous term:

$$d\Delta n_c/dt = -\Delta n_c/\tau_n$$

Δn_c is evidently $B \exp(-t/\tau_n)$, where B is a constant that we need to determine from the initial condition.

The complete solution $\Delta n = \Delta n_p + \Delta n_c$ is

$$\Delta n = \frac{G_0}{(1/\tau_n) - b} e^{-bt} + B e^{-t/\tau_n}$$

To find B, use the initial condition $\Delta n(0) = 0$:

$$0 = \frac{G_0}{(1/\tau_n) - b} + B,$$

so

$$B = -\frac{G_0}{(1/\tau_n) - b}$$

and we obtain the same solution that Laplace transforms gave us previously.

Example

A constant monochromatic photon flux of N_0 illuminates the surface ($x = 0$) of thick p-type Si. Find $\Delta n(x)$, neglecting surface recombination. (This means $d\Delta n/dx = 0$ at the surface, as we'll see in the next lecture.)

Solution

Recall that the generation rate $G = \alpha N_0 e^{-\alpha x}$.

Write the minority carrier diffusion equation, where the time derivative is 0 because nothing is changing.

$$0 = D_n(d^2\Delta n/dx^2) + \alpha N_0 e^{-\alpha x} - \Delta n/\tau_n$$

This is a nonhomogeneous equation because of the generation term. Since the nonhomogeneous term is proportional to $e^{-\alpha x}$, let's guess that the particular solution

$$\Delta n_p = A e^{-\alpha x}$$

Plug this into the differential equation:

$$0 = D_n \alpha^2 A e^{-\alpha x} + \alpha N_0 e^{-\alpha x} - A e^{-\alpha x} / \tau_n$$

Solving for A:

$$A = \frac{\alpha N_0}{(1/\tau_n) - \alpha^2 D_n}$$

Now we look for the complementary solution Δn_c by dropping the nonhomogeneous term:

$$0 = D_n (d^2 \Delta n_c / dx^2) - \Delta n_c / \tau_n$$

The solution is

$$\Delta n_c = B \exp(-x/L_n),$$

where I've dropped the $\exp(+x/L_n)$ solution because the silicon is thick and Δn can't approach infinity as x gets very large.

So the complete solution is

$$\Delta n = \frac{\alpha N_0}{(1/\tau_n) - \alpha^2 D_n} e^{-\alpha x} + B e^{-x/L_n},$$

and we need to use the boundary condition to solve for B. At $x = 0$, $d\Delta n/dx = 0$:

$$0 = -\frac{\alpha^2 N_0}{(1/\tau_n) - \alpha^2 D_n} - \frac{B}{L_n},$$

so

$$B = -\frac{\alpha^2 N_0 L_n}{(1/\tau_n) - \alpha^2 D_n}$$

Finally

$$\Delta n = \frac{\alpha N_0}{(1/\tau_n) - \alpha^2 D_n} \left(e^{-\alpha x} - \alpha L_n e^{-x/L_n} \right)$$

The algebra's getting messy. In the next lecture, it gets worse. Much worse.

Look at the solution above. Mathematically and conceptually, is it possible for excess carrier concentration to be negative at any location?

Lecture 21: Surface recombination

Every silicon atom wants to bond with four atoms surrounding it. At the surface of silicon, there are no more atoms in one direction, so the surface atoms can't achieve a fourth bond: there are **dangling bonds** at the surface. The dangling bonds create energy levels within the bandgap, and this causes a lot of recombination at the surfaces.

The **surface recombination rate** [$\text{cm}^{-2}\text{s}^{-1}$] is the recombination rate at the surface per unit area. All of the diffusion current into the surface results in recombination because the current can't keep going. So

$$\text{surface recombination rate} = |\text{minority carrier current density}|/q$$

We divide by q to obtain the number of minority carriers flowing into the surface (per unit area per unit time), rather than the total charge of these carriers flowing into the surface (per unit area per unit time).

Also, the surface recombination rate is proportional to the excess carrier concentration at the surface:

$$\text{surface recombination rate} = S\Delta n(\text{surface}), \text{ or } S\Delta p(\text{surface})$$

where S is called the **surface recombination velocity** (if only because it has units of velocity).

Combining these expressions for surface recombination yields

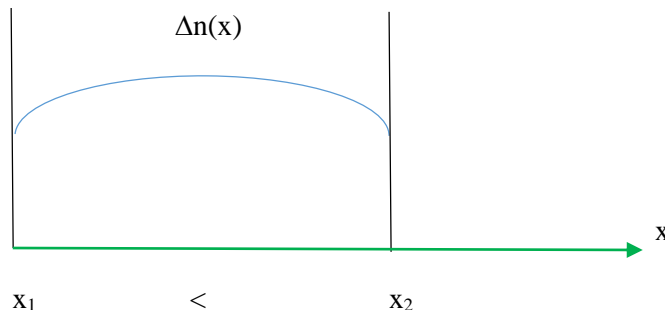
$$S\Delta n(\text{surface}) = D_n \left. \frac{d\Delta n}{dx} \right|_{\text{surface}}$$

for electrons in p-type semiconductors.

(Notice that this is exactly analogous to a convective boundary condition with $T_\infty = 0$, from the very first

lecture: $\bar{h}(T_{\text{surface}} - T_\infty) = k \left. \frac{dT}{dx} \right|_{\text{surface}}$)

Let's see how we can get rid of the absolute value sign. Carriers must flow **into** surfaces where recombination occurs because the surface recombination is gobbling up carriers, not sending them out. If we plot Δn vs. x , Δn must slope down towards surfaces: if x increases to the right, the slope must be positive at the left surface (x_1) and negative at the right surface (x_2):



We see that $d\Delta n/x$ is positive at x_1 and negative at x_2 . Therefore

$$S\Delta n(x_1) = D_n(d\Delta n/dx)|_{x_1} \text{ and } S\Delta n(x_2) = -D_n(d\Delta n/dx)|_{x_2}$$

Similarly, for holes in n-type silicon,

$$S\Delta p(x_1) = D_p(d\Delta p/dx)|_{x_1} \text{ and } S\Delta p(x_2) = -D_p(d\Delta p/dx)|_{x_2}$$

For the rest of this lecture, let's drop the p and n subscripts from D, τ , and L.

Example

Find $\Delta n(x)$ in p-type silicon with constant uniform generation G and identical surface recombination S at both surfaces ($x = \pm W/2$).

Solution

We start with the minority carrier diffusion equation, where the time derivative is 0.

$$0 = D(d^2\Delta n/dx^2) + G - \Delta n/\tau$$

First we look for a particular solution. Since the nonhomogeneous term is a constant, it's a good guess that Δn_p is a constant (and the spatial derivative is 0):

$$0 = G - \Delta n_p/\tau$$

So $\Delta n_p = G\tau$.

Next, we look for the complementary solution, dropping the nonhomogeneous term:

$$0 = D(d^2\Delta n_c/dx^2) - \Delta n_c/\tau$$

We've seen this before. But now, let's write the solution in terms of cosh and sinh:

$$\Delta n_c = A\cosh(x/L) + B\sinh(x/L)$$

Generally, when a material is (infinitely) thick, it's most convenient to use exponential functions. But when a material is thin, it's most convenient to use cosh and sinh.

Because the silicon and the illumination are symmetrical about $x = 0$, we know that Δn must be an even function, so B must be 0. The complete solution is then

$$\Delta n = G\tau + A\cosh(x/L),$$

where we must determine A from the boundary conditions.

At $x = W/2$,

$$S\Delta n(W/2) = -D(d\Delta n/dx)|_{W/2}$$

Plug Δn into the boundary condition:

$$S\tau G + SA \cosh(W/2L) = -(DA/L) \sinh(W/2L)$$

So

$$A = -\frac{S\tau G}{\frac{D}{L} \sinh\left(\frac{W}{2L}\right) + S \cosh\left(\frac{W}{2L}\right)}$$

Example

In p-type silicon with the same surface recombination velocity on both surfaces ($x = \pm W/2$), the excess carrier concentration, due to illumination, is initially uniform: $\Delta n(x,0) = C$. At $t = 0$, the illumination ceases. Determine $\Delta n(x,t)$.

Solution: Fourier series!

Write the minority carrier diffusion equation, where $G = 0$ for $t > 0$:

$$\partial \Delta n / \partial t = D(\partial^2 \Delta n / \partial x^2) - \Delta n / \tau$$

Assume $\Delta n(x,t)$ is a product of a function of x and a function of t :

$$\Delta n(x,t) = X(x)T(t)$$

Plug this into the differential equation:

$$X(dT/dt) = DT(d^2X/dx^2) - XT/\tau$$

Divide by XT :

$$(1/T)(dT/dt) = (D/X)(d^2X/dx^2) - 1/\tau$$

For convenience, I'll rearrange a little:

$$[(1/T)(dT/dt) + 1/\tau]/D = (1/X)(d^2X/dx^2)$$

How can a function of T equal a function of X ? If both equal the same constant. Let's call it $-\lambda^2$. So we now have two equations:

$$[(1/T)(dT/dt) + 1/\tau]/D = -\lambda^2$$

and

$$(1/X)(d^2X/dx^2) = -\lambda^2.$$

The t equation can be written

$$dT/dt = -(1/\tau + \lambda^2 D)T$$

The solution is

$$T = \exp[-t(1/\tau + \lambda^2 D)]$$

(I'll put a multiplicative constant in X, so I don't need it in T.)

The solution to the x equation is

$$X = A \cos(\lambda x),$$

where I dropped the sine because of the symmetry of the problem.

So a possible solution is

$$\Delta n = A \cos(\lambda x) \exp[-t(1/\tau + \lambda^2 D)]$$

Next, we use the boundary conditions to determine possible values of λ . At $x = W/2$:

$$S \Delta n(W/2, t) = -D(\partial n / \partial x)|_{x=W/2}$$

Plug Δn into the boundary condition (and divide out the time function):

$$S A \cos(\lambda W/2) = D A \lambda \sin(\lambda W/2)$$

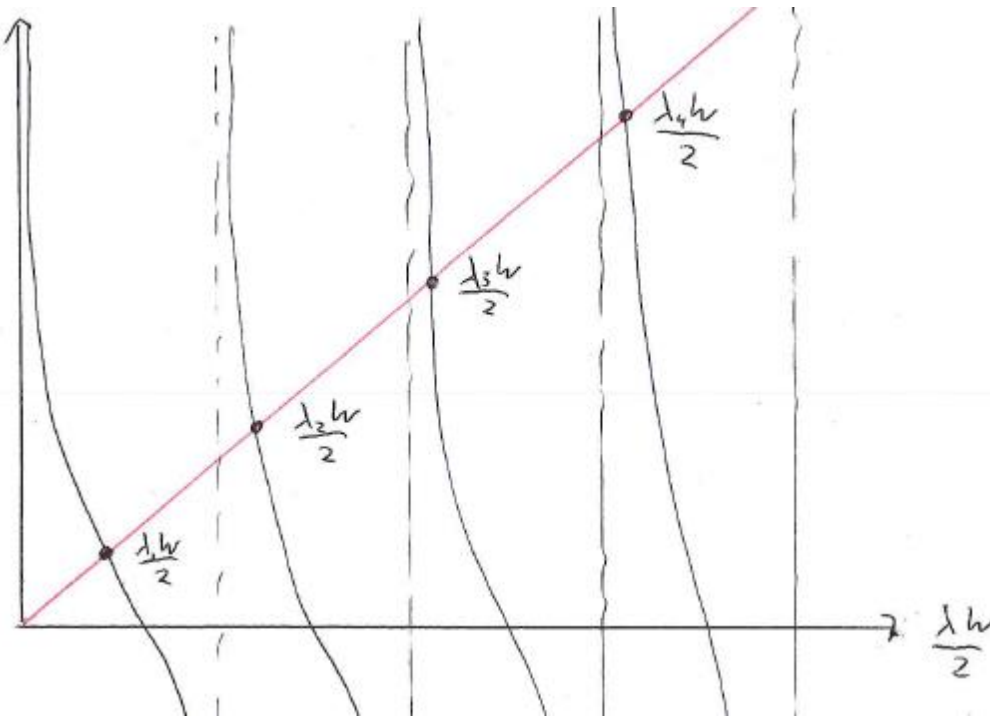
So

$$\cot(\lambda W/2) = D \lambda / S$$

but I want to rewrite the right side to see a factor of $\lambda W/2$:

$$\cot(\lambda W/2) = (2D/SW)(\lambda W/2)$$

This is a transcendental equation. It has no exact analytical solutions! We can solve it with a computer. We can also understand it better by solving it graphically: plot $\cot(\lambda W/2)$, the black curves below, and also $(2D/SW)(\lambda W/2)$, the red line, as functions of $\lambda W/2$:



Where the red line intersects a black curve, we've found a value of $\lambda W/2$ that satisfies the transcendental equation. As you can see, there are infinite possible values of λ , and we label them with subscripts.

So the complete solution we need is a sum of solutions with all possible values of λ :

$$\Delta n = \sum_i A_i \cos(\lambda_i x) \exp \left[-t \left(\frac{1}{\tau} + \lambda_i^2 D \right) \right]$$

How do we determine A_i for each of the infinite terms? We use the initial condition $\Delta n(x,0) = C$:

$$C = \sum_i A_i \cos(\lambda_i x)$$

What's next? Fourier to the rescue! We multiply both sides by $\cos(\lambda_j x)$ and integrate from $-W/2$ to $W/2$. All the terms integrate to 0 except for $i = j$. (It's not obvious that this should be true, but you can confirm it by evaluating the integrals where i isn't j .)

For $i = j$:

$$\int_{-W/2}^{W/2} C \cos(\lambda_j x) dx = \int_{-W/2}^{W/2} A_j \cos^2(\lambda_j x) dx$$

Since both integrands are even functions, let's simplify by changing the lower limits to 0 (we've divided both sides by 2):

$$\int_0^{W/2} C \cos(\lambda_i x) dx = \int_0^{W/2} A_i \cos^2(\lambda_i x) dx$$

Evaluating the integrals:

$$\frac{C}{\lambda_i} \sin(\lambda_i W / 2) = A_i \frac{\lambda_i W + \sin(\lambda_i W)}{4\lambda_i}$$

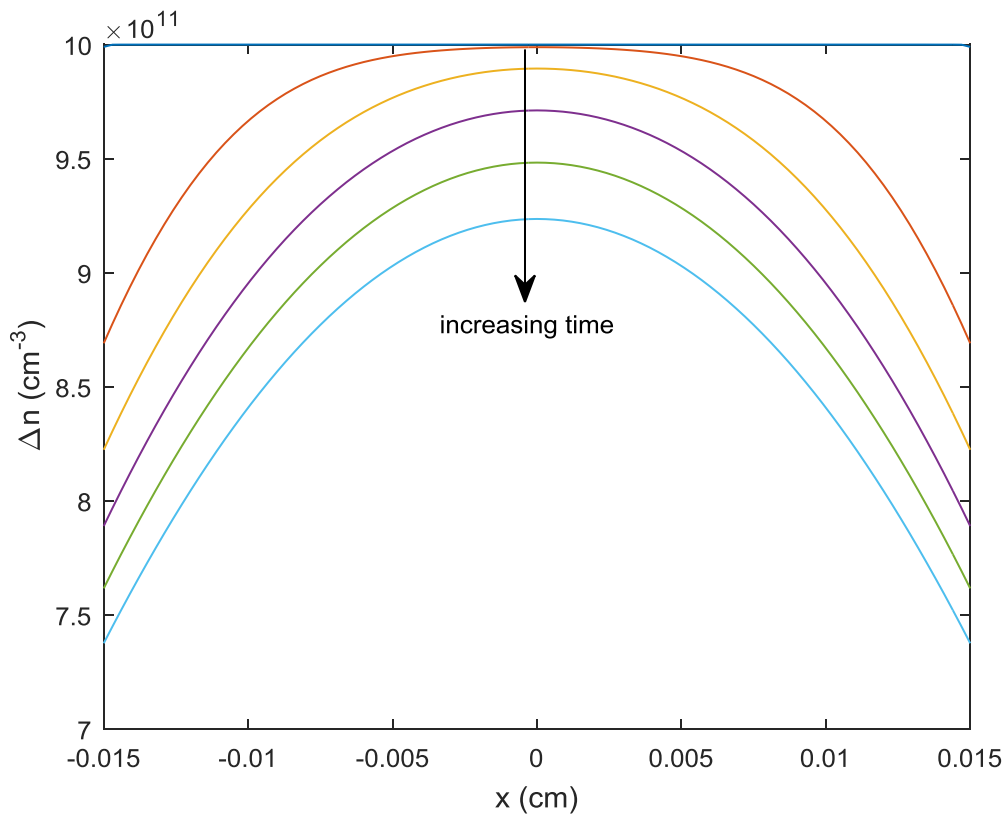
And solve for A_i :

$$A_i = \frac{4C \sin(\lambda_i W / 2)}{\lambda_i W + \sin(\lambda_i W)}$$

So the final result is

$$\Delta n(x, t) = 4C \sum_{i=1}^{\infty} \frac{\sin(\lambda_i W / 2)}{\lambda_i W + \sin(\lambda_i W)} \cos(\lambda_i x) \exp[-t(\lambda_i^2 D + 1/\tau)]$$

where λ_i are the solutions to $\cot(\lambda_i W / 2) = D\lambda_i / S$.



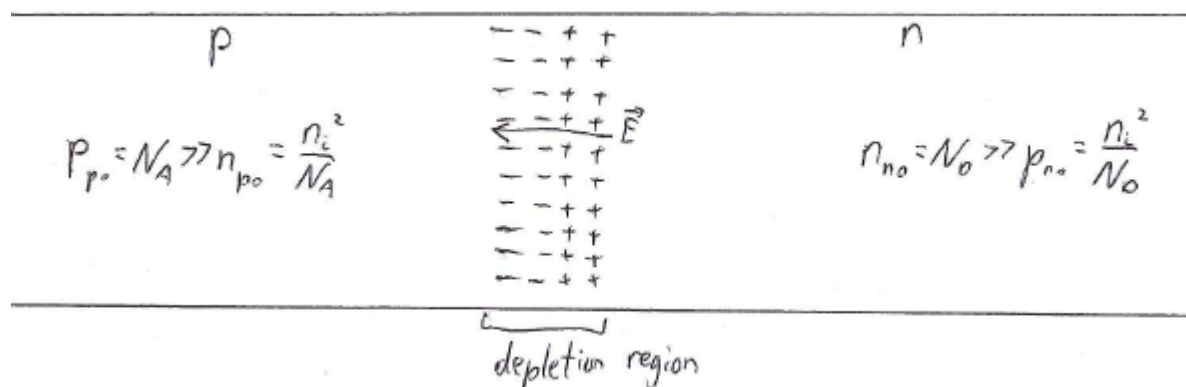
Why did I use hyperbolic trig functions in the first example, but circular (regular) trig functions in the second?

Lecture 22: p-n junction: the heart of solar cell, LED, or other diode

(Honsberg and Bowden, Ch. 3.5 and 3.6)

Suppose we have some p-type silicon, and we cause a high concentration of donor atoms to diffuse into the material on one side. The donor concentration overwhelms the acceptor concentration and locally transforms the material to n-type. But the opposite side of the silicon is still p-type. Where n-type and p-type meet, we have **p-n junction**, and the device so formed is called a **diode**. The physics of the p-n junction dominates the behavior of solar cells, LEDs, and even transistors.

So what happens at a p-n junction?

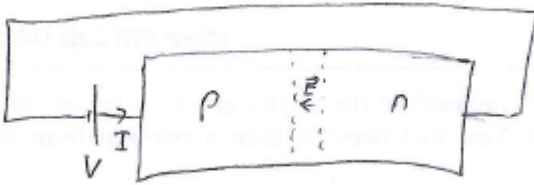


Free electrons in the n region cheerfully drop into holes in the p region. This process can't continue forever because electric charge builds up: Negative charge builds up on the p side, and positive charge is left behind on the n side. This creates an electric field (pointing, as always, from positive charge to negative charge). The electric field prevents additional electrons from moving to the p side (electrons are repelled by the electrons that have already gone over).

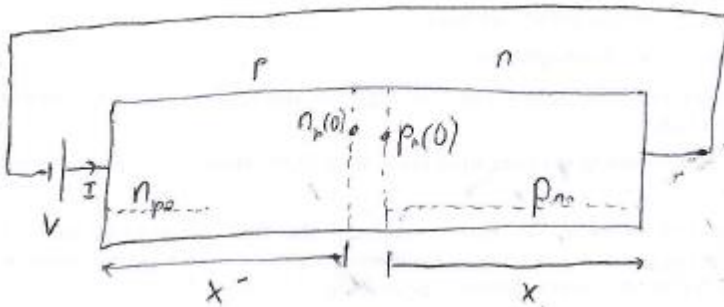
The charged region is called the **depletion region** because it's depleted of mobile charge carriers. (The free electrons have left the n-side, and the holes on the p-side recombined with those electrons.)

Recall that voltage is the (negative of the) integral over an electric field, and the electric field points from high potential to low. So there's a **built-in voltage ψ** across the depletion region, with the higher potential on the right.

What happens when an **external** voltage is applied across a p-n junction? Suppose the positive end of the battery is connected to the p-type material (this is called **forward bias**):



We want to derive an equation for I through the diode as a function of applied voltage V across it. The result is called the **ideal diode law**. Let's take a look at carrier concentrations on both sides of the junction:



The equilibrium minority carrier concentrations are $n_{p0} = n_i^2/N_A$ on the p side and $p_{n0} = n_i^2/N_D$ on the n side. We assume that these **equilibrium** concentrations exist **far from the depletion region**. However, at the **edge of the depletion region**, the minority carrier concentrations are affected by applied voltage V . We need one new equation (for each carrier type), the **law of the junction** (minority carrier concentration at junction edges increases exponentially with V):

$$n_p(0) = n_{p0}e^{qV/kT} \text{ and } p_n(0) = p_{n0}e^{qV/kT}$$

where k is the Boltzmann constant.

The law of the junction gives us boundary conditions, so now we simply solve the minority carrier diffusion equation in each region. We assume steady-state conditions and no illumination (for now). So in the n-type region the diffusion equation is

$$0 = D_p(d^2\Delta p/dx^2) - \Delta p/\tau_p,$$

and the boundary conditions are $\Delta p(0) = p_n(0) - p_{n0} = p_{n0}(e^{qV/kT} - 1)$, and $\Delta p(x)$ doesn't get infinitely large at infinitely large x . (This boundary condition only works for **thick** diodes: the **surfaces** are very **far** from the **junction**. For thin diodes, we'd need to account for surface recombination.) We've already solved a nearly identical problem! The solution is

$$\Delta p(x) = p_{n0}(e^{qV/kT} - 1)\exp(-x/L_p)$$

Plug it into the diffusion equation and the boundary conditions to confirm!

Similarly, in the p-type region,

$$\Delta n(x') = n_{p0}(e^{qV/kT} - 1)\exp(-x'/L_n)$$

For convenience, I use a different coordinate x' .

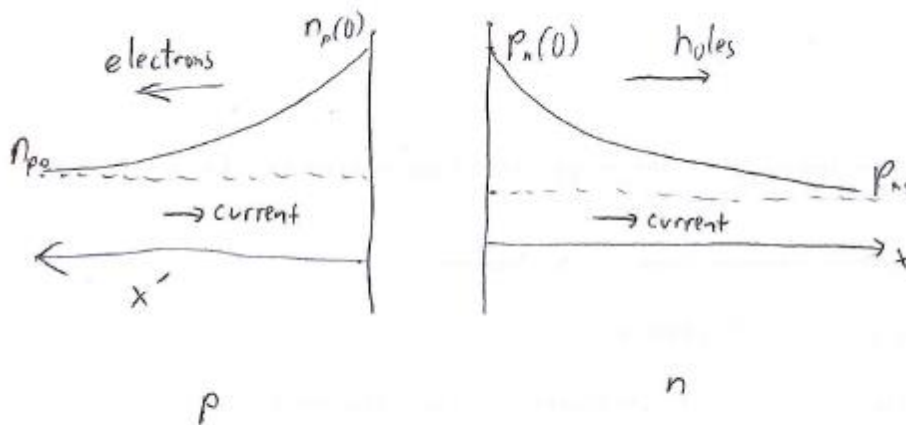
We can calculate minority carrier diffusion currents. On the n side:

$$J_p(x) = -qD_p(d\Delta p/dx) = (qD_p p_{n0}/L_p)(e^{qV/kT} - 1)\exp(-x/L_p)$$

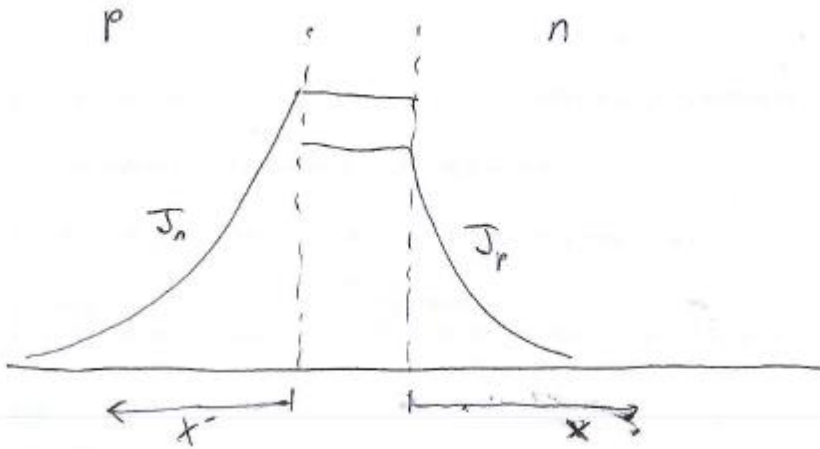
And on the p side:

$$J_n(x') = qD_n(d\Delta n/dx) = -(qD_n n_{p0}/L_n)(e^{qV/kT} - 1)\exp(-x'/L_n)$$

Both currents are in the same direction (to the right):



We want to know the total current density, $J = J_p + J_n$. The total current is uniform across the diode because otherwise charge would accumulate somewhere. So if we can determine total current anywhere, we know total current everywhere. So far we know minority carrier currents, but not majority carrier currents. We now assume that the depletion region is so narrow that the each current density (hole and electron) stays approximately constant across it:



So we know the total current density $J = |J_n(0)| + |J_p(0)|$ in the depletion region, and this is the total current density everywhere:

$$J = q(D_p p_{n0}/L_p + D_n n_{p0}/L_n)(e^{qV/kT} - 1)$$

We can multiply by cross-sectional area A to get current:

$$I = qA(D_p p_{n0}/L_p + D_n n_{p0}/L_n)(e^{qV/kT} - 1) = I_0(e^{qV/kT} - 1)$$

where we gathered all the constants into one constant called I_0 , “dark saturation current.” We have now obtained the ideal diode law!

To derive the I vs. V equation for a solar cell, the only modification we have to make is to add illumination. If we assume a uniform generation rate G , the diffusion equation becomes

$$0 = D_p(d^2\Delta p/dx^2) + G - \Delta p/\tau_p.$$

The particular solution is $G\tau_p$ and the complementary solution is $B\exp(-x/L_p)$, where we must determine B from the boundary condition (the law of the junction). Since

$$\Delta p(x) = G\tau_p + B\exp(-x/L_p)$$

and

$$\Delta p(0) = p_n(0) - p_{n0} = p_{n0}(e^{qV/kT} - 1),$$

$$p_{n0}(e^{qV/kT} - 1) = G\tau_p + B$$

and finally

$$\Delta p(x) = G\tau_p + [p_{n0}(e^{qV/kT} - 1) - G\tau_p]\exp(-x/L_p).$$

The hole current density on the n side is now

$$J_p(x) = -qD_p(d\Delta p/dx) = (qD_p p_{n0}/L_p)(e^{qV/kT} - 1)\exp(-x/L_p) - (qD_p G\tau_p/L_p)\exp(-x/L_p)$$

and

$$J_p(0) = (qD_p p_{n0}/L_p)(e^{qV/kT} - 1) - qL_p G$$

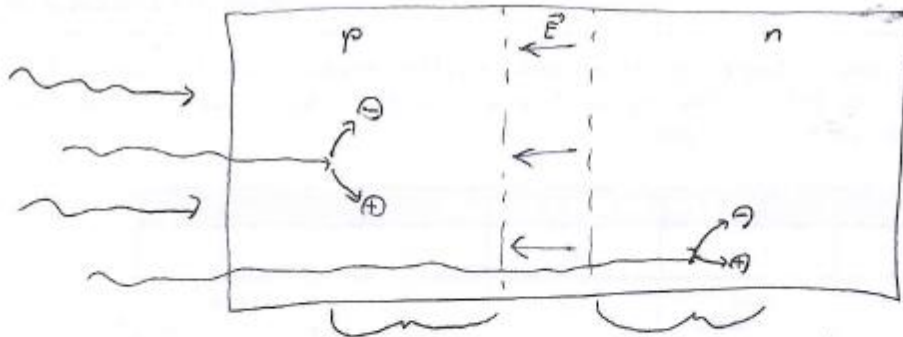
where I've used $D_p\tau_p = L_p^2$. Similarly,

$$J_n(0) = (qD_n n_{p0}/L_n)(e^{qV/kT} - 1) - qL_n G$$

and the total current is

$$I = qA(D_p p_{n0}/L_p + D_n n_{p0}/L_n)(e^{qV/kT} - 1) - qG(L_p + L_n) = I_0(e^{qV/kT} - 1) - I_{sc}$$

where $I_{sc} = qG(L_p + L_n)$ is the short circuit current (when $V = 0$), which can also be called the light-generated current. Minority carriers generated (by light absorption) within a diffusion length of the junction are swept across the junction and contribute to light-generated current.



electrons generated within L_n of junction diffuse there and are swept across

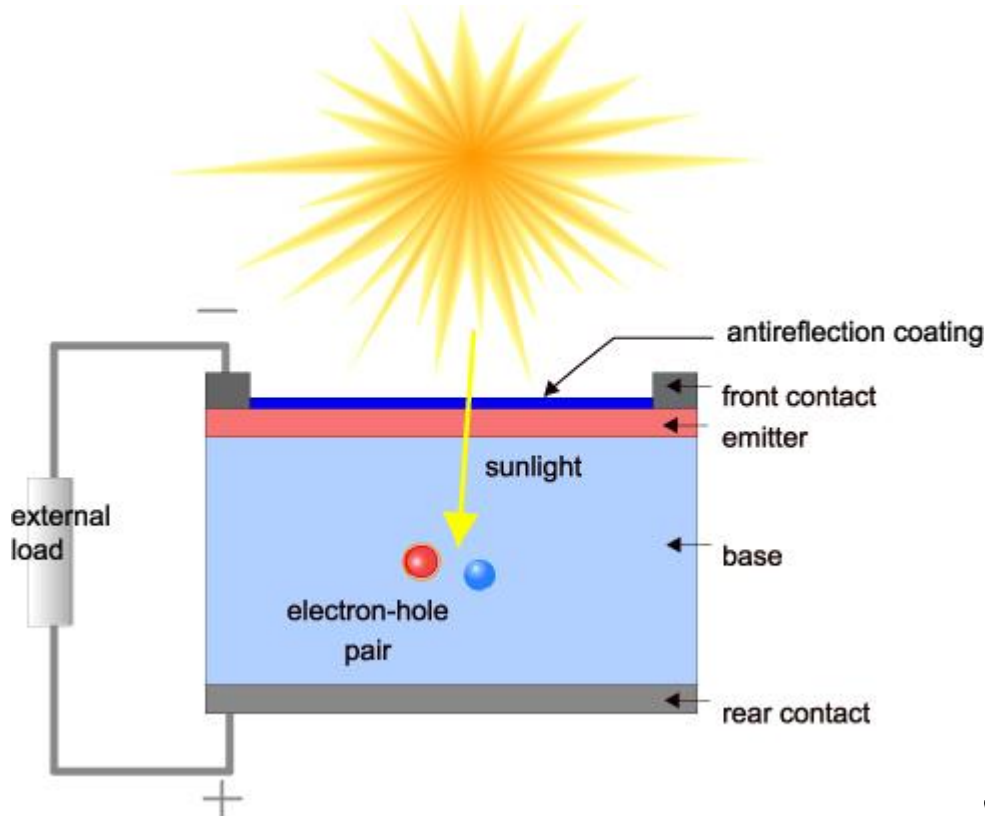
holes generated within L_p of junction diffuse there and are swept across

← light-generated current
(opposite the direction of forward-bias current!)

A photon absorbed at what location in the solar cell is most likely to contribute to photocurrent?

Lecture 23: Solar cells

(Honsberg and Bowden, Ch. 4.1, 4.2, 5.1)



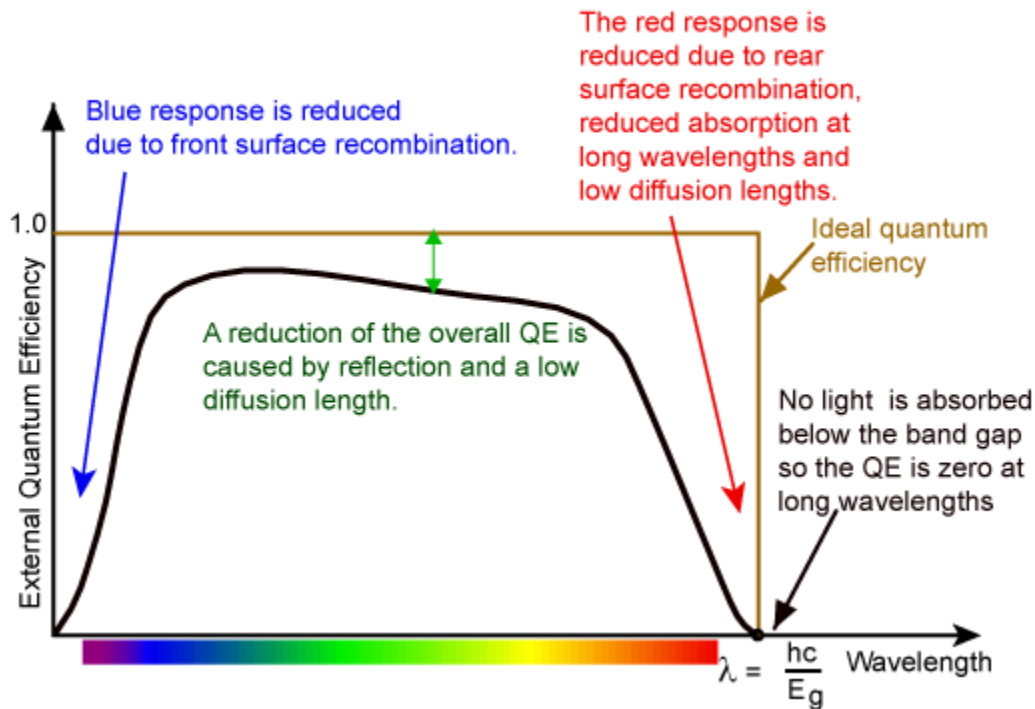
(figure from <http://pveducation.org/pvcdrom/solar-cell-operation/solar-cell-structure>)

Suppose the **emitter** is n-type silicon and the **base** is p-type silicon. The p-n junction is where the emitter and base meet. Why is the junction so close to the surface? Most sunlight is absorbed within about $10\ \mu\text{m}$ of the surface. We want the light absorbed as close as possible to the p-n junction. If light is absorbed more than a diffusion length away from the junction, the electron-hole pairs recombine before they get a chance to contribute to photocurrent.

The **antireflection coating** is very important because bare silicon reflects about a third of the photons that arrive. The antireflection coating is a thin film designed to minimize reflection. As you learned in introductory physics, reflection is reduced if **destructive interference** is achieved: light reflecting off the top of the thin film should interfere destructively with light reflecting off the bottom of the thin film.

To improve solar cell efficiency, we must reduce other loss mechanisms as well. We can analyze the **quantum efficiency** of a solar cell: the quantum efficiency is the fraction of photons (of a particular wavelength) that generate carriers that contribute to photocurrent. The quantum efficiency at different **wavelengths** informs us about loss mechanisms at different **depths** within the solar cell: **short**

wavelengths are absorbed near the **front** (the top), and **long wavelengths** are absorbed near the **back** (the bottom).



(figure from <http://pveducation.org/pvcdrom/solar-cell-operation/quantum-efficiency>)

Very short wavelengths are absorbed too close to the front surface, and surface recombination takes a heavy toll. The highest quantum efficiency occurs at wavelengths absorbed near the junction. Long wavelengths lose quantum efficiency for three reasons:

- long wavelengths have low absorption coefficients (some infrared light passes right through the silicon)
- long wavelengths are absorbed far from the front surface, more than a diffusion length below the junction
- long wavelengths are absorbed near the back surface, and carriers generated there are subject to surface recombination

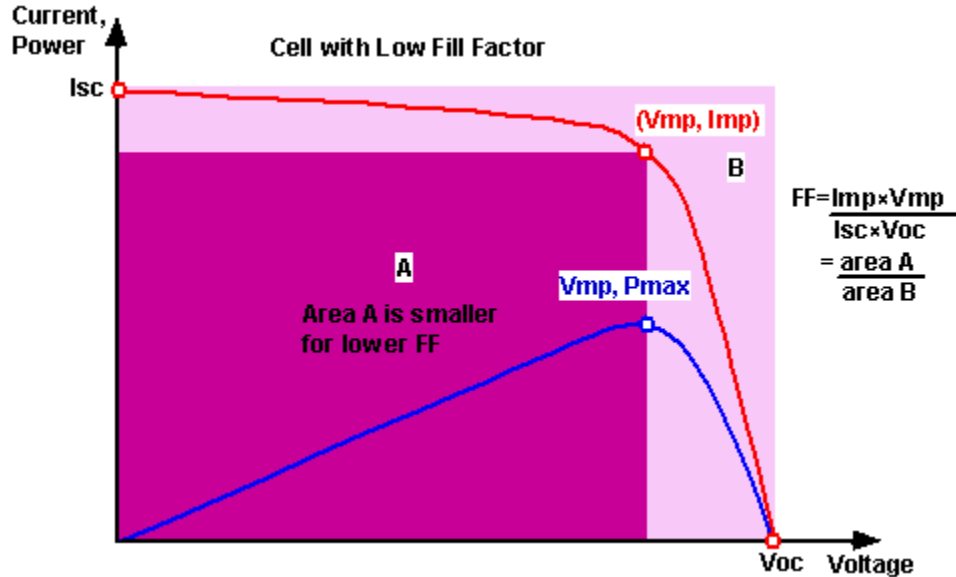
The **electric power** delivered by a solar cell is

$$P = |I|V = [I_{sc} - I_0(e^{qV/kT} - 1)]V$$

and depends on

- the intensity of light shining on a solar cell
- the efficiency of the solar cell
- the resistance of the external load (the thing that the solar cell is powering, maybe a motor)

If no load is attached, no current flows (we call this **open circuit**), and $P = 0$. On the other hand, if the load resistance is 0 (we **short circuit** the two ends of the solar cell together with a wire), voltage is 0 and therefore $P = 0$. So the solar cell delivers useful power at intermediate conditions. To understand the optimal load resistance, let's look at $|I|$ vs. V and P vs. V :



(figure from <http://pveducation.org/pvcdrom/solar-cell-operation/fill-factor>)

The red curve is current $|I|$. As we know from the equation we derived,

$$|I| = I_{sc} - I_0(e^{qV/kT} - 1).$$

We see that when $V = 0$, $|I| = I_{sc}$, the largest possible current. $|I|$ gradually decreases as V increases, until $I = 0$ at open-circuit voltage (V_{oc}). We're most concerned with power, the product of current and voltage (the blue curve). The power is maximized at the **maximum power point**, (V_{mp}, I_{mp}) . A quantity called the **fill factor** is a measure of the "squareness" of the IV curve:

$$FF = I_{mp}V_{mp}/(I_{sc}V_{oc})$$

We want the fill factor to be as large as possible (as close to 1 as possible).

The efficiency η of the solar cell is defined as the maximum electric power that can be delivered by the solar cell, as a fraction of the incoming light power:

$$\eta = I_{mp}V_{mp}/(\text{light power in}) = I_{sc}V_{oc}FF/(\text{light power in})$$

To increase solar cell efficiency, we try to increase short-circuit current, open-circuit voltage, and fill factor. Let's briefly consider design strategies for increasing each of these.

We recall that short circuit current (for a thick solar cell) is

$$I_{sc} = qG(L_p + L_n)$$

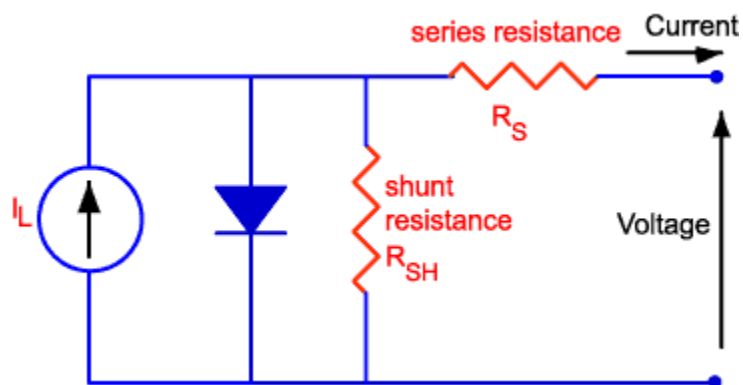
We can try to increase generation by reducing optical losses (reflection and shading due to front contacts). We can try to increase diffusion length by increasing carrier lifetime (by finding ways to eliminate “trap” energy levels in the bandgap, caused by crystalline defects).

If we look at our I vs. V equation and plug in $I = 0$ (open circuit), we find

$$V_{oc} = (kT/q)\ln(I_{sc}/I_0+1)$$

It looks as though increasing temperature would be helpful, but unfortunately I_0 (which we want to minimize) also increases with temperature. So actually V_{oc} decreases with temperature. This is why some cooling mechanism (possibly heat fins!) is required for **concentrator cells**, on which mirrors focus concentrated sunlight.

The fill factor is influenced by resistance in the solar cell. A solar cell can be modeled as a diode, a current source, and two resistors:



(figure from <http://pveducation.org/pvcdrom/solar-cell-operation/impact-of-both-resistances>)

The series resistance is due to the resistivity of the emitter and the base, as well as the contact resistance of the metal on the silicon. The series resistance is ideally 0. The shunt resistance, however, is ideally infinite.

The typical efficiency of commercial solar cells is 14-19%. These are usually multicrystalline silicon solar cells; multicrystalline silicon is cheaper (and lower quality) than the monocrystalline silicon used for computer chips.

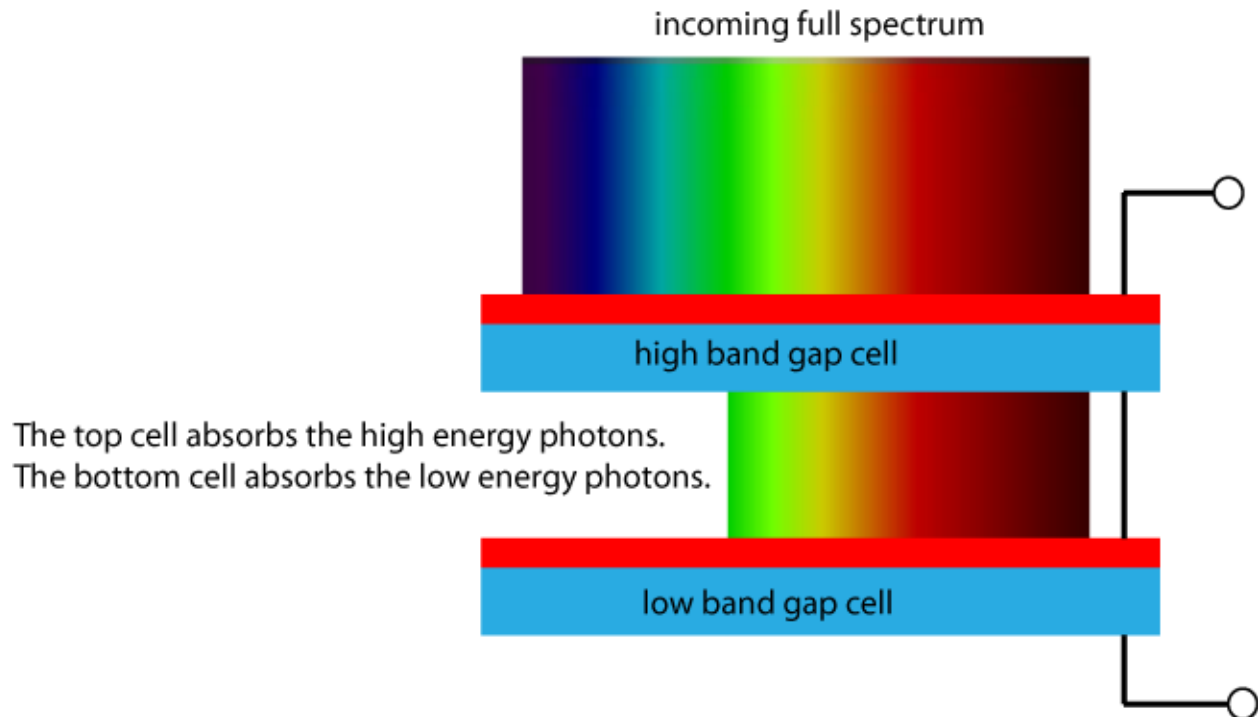
The highest efficiency of a monocrystalline silicon solar cell is 27.6%.

The best efficiency for a solar cell with a single p-n junction is 29.1%. It's made of gallium arsenide.

The world record efficiency is 46%, for a solar cell with four p-n junctions.

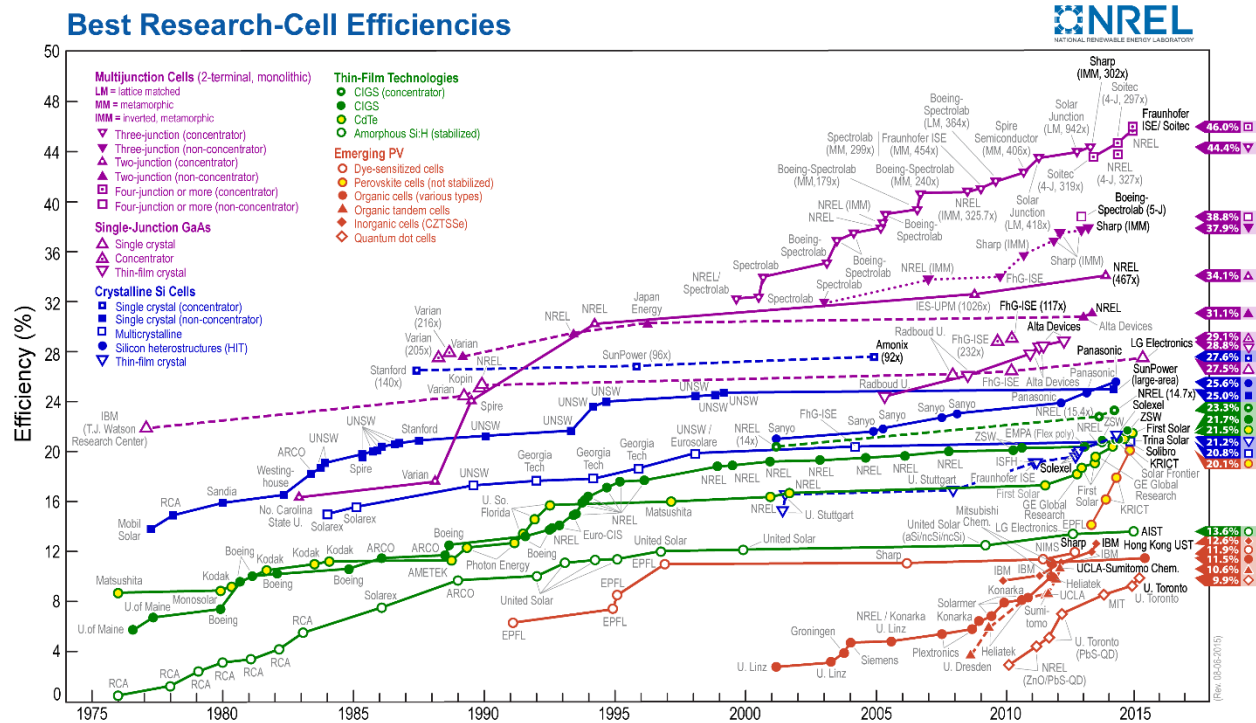
Let's understand the limitations of a solar cell with a single p-n junction. If the solar cell is made of a single material (crystalline silicon, for example), there's a single bandgap energy. Photons with less energy than the bandgap energy cannot generate electron hole pairs; the energy of these photons is lost.

Photons with more energy than the bandgap energy create electron-hole pairs, but photon energy in excess of the bandgap energy is converted to heat. So a particular bandgap energy is optimized for a particular photon energy. A solar cell with four p-n junctions is made of four different materials, each with a different bandgap energy. The material with the highest bandgap is on the top, and it absorbs the photons with the highest energies. The photons with lower energies pass right through the top material and are absorbed by a material with a lower bandgap energy. The following diagram illustrates how this works in a solar cell with two junctions:



(figure from <http://pveducation.org/pvcdrom/solar-cell-operation/tandem-cells>)

Here's a chart of record efficiencies:



(This plot is courtesy of the National Renewable Energy Laboratory, Golden, CO.)

Georgia Tech had a string of multicrystalline-silicon record efficiencies in the mid-nineties. (I can't take any credit because I joined the photovoltaics research group in 1998.)

To reduce reflection losses in a solar cell, why can't we just paint it black? (This is a question I actually asked a more experienced student when I was new at Georgia Tech.)