Simulations of Optical Thin Films

Thin film interference, despite being an interesting physical phenomena in its own right, has numerous applications to various forms of industry. A few examples are optical communication, precision scientific equipment and even in photovoltaic cells (solar panels). In fact, it is widely believed that the implementation of thin film technologies to solar cells could greatly increase their effectiveness and hence reduce costs¹. For this reason, and of course just for the sack of doing some awesome physics, it is important to develop a solid understanding of the phenomena of thin film interference. To this end, we propose to construct a mathematical model of light traveling, from air, into a series of thin film layers of high and low index of refraction and try to understand the observed behaviour. In order to test our model, we propose to actually construct a device called a Fabry-Perot interference filter (which is a direct application of thin film interference, these devices are actually an important part of producing spectroscopic images, more about applications will be discussed later), and make a few measurements which will compare to the predictions of our model/simulations. Such a filter is composed constructed by depositing varying layers of high and low "index of refraction" onto a glass substrate. The exact details of how all this will be accomplished will be discussed later, but for now we must first understand exactly what all this interference business is about.

First, we would like to consider our model of light (since ultimately we are trying to understand how light is transmitted through a Fabry-Perot filter, we must have a way to understand light!). The model we use is a macroscopic one, treating light as an electromagnetic phenomenon, ignoring any Quantum effects (this is allowable because we are interested in the macroscopic behaviour or light, we don't really care about how each photon is transmitting through the device since in all practical applications, we are dealing with huge ensembles of photons). First, we must consider Maxwell's equations inside of a dielectric medium (a dielectric medium is basically an insulator, which is the type of material our interference filter will be constructed out of).

¹ See for example, Y.Hamakawa, Thin-Film Solar Cells, Next Generation Photovoltaics and its Applications. Springer 2004.

$$\nabla \cdot \vec{D} = \rho_f$$
$$\nabla \cdot \vec{B} = 0$$
$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$
$$\nabla \times \vec{H} = \vec{J}_f + \frac{\partial \vec{D}}{\partial t}$$

Where:

 \vec{D} , is the so called "electric displacement field", \vec{B} , is the magnetic field, \vec{E} is the electric field, \vec{H} , is the so called "Auxiliary field", \vec{J}_f is the free current density (we say "free", because these currents are not produced by any magnetization of the medium, i.e. they are put in the medium by some external force that we can control) and ρ_f is the free charge (free for the same reason as the free current density). For our purposes (the construction of an optical device), it suffices to supposed that there are no free currents or free charges in our materials. This is reasonable simply because we are dealing with a free standing device sitting in air, there really isn't any plausible source for any free currents or charges! Hence, our equations reduce too:

$$\nabla \cdot \vec{D} = 0$$
$$\nabla \cdot \vec{B} = 0$$
$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$
$$\nabla \times \vec{H} = \frac{\partial \vec{D}}{\partial t}$$

We also note that the displacement and auxiliary fields are related to the electric and magnetic fields respectively by:

$$\vec{D} = \varepsilon \vec{E}$$
$$\vec{H} = \frac{\vec{B}}{\mu}$$

Where ε is the "electric permittivity" and μ is the "magnetic permeability". ε is essentially a measure of how well a material transmits (or "permits") and electric field. μ is a little more complicated, it is defined as the degree of magnetization of a material the responds linearly to an applied magnetic field. When a magnetic field is applied to a medium, so called "bound currents" can form inside the material (essentially, they are result of electrons being pushed around in medium by the applied field). The beautiful thing about the way we wrote down Maxwell's equations is that we actually don't have to consider these currents (notice that the equations only care about free currents)! That being said, we can ignore their effect.

The next critical assumption that we make is that our materials are linear (actually, we already made this assumption when we wrote down the relationships between D and E and between H and B above). What exactly does this mean? It means that the values of ε and μ do not vary inside the medium, they are the same from point to point, in other words they are constants. If we didn't make this assumption, then the values of ε and μ would vary from place to place (in fact, they would have to be represented by tensors, making the math all that much more complicated). Finally, with all this information, we can finally write:

$$\nabla \cdot \vec{E} = 0$$
$$\nabla \cdot \vec{B} = 0$$
$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$
$$\nabla \times \vec{B} = \mu \varepsilon \frac{\partial \vec{E}}{\partial t}$$

If we take the curl of the last two of these equations (and using the first two when needed), we find the following:

$$\nabla^{2}\vec{E} = \mu\varepsilon \frac{\partial^{2}\vec{E}}{\partial t^{2}}$$
$$\nabla^{2}\vec{B} = \mu\varepsilon \frac{\partial^{2}\vec{B}}{\partial t^{2}}$$

Which we recognize as wave equations having wave velocity (in fact it is this very wave nature of light that allows for interference, just like standing waves on a string can interfere with each other):

$$v = \frac{1}{\sqrt{\varepsilon\mu}} = \frac{c}{n}$$

Where c is the speed of light in vacuum and n is a constant called the index of refraction (which depends on the properties of the material we are interested in), and is defined as:

$$n = \sqrt{\frac{\varepsilon\mu}{\varepsilon_o\mu_o}}$$

Where ε_o and μ_o are the electric permittivity and magnetic permeability of vacuum, having numerical values of $8.85 \cdot 10^{-12} s^4 A^2 / m^3 kg$ and $1.26 \cdot 10^{-6} m kg / s^2 A^2$ respectively (where m is meters, A is Ampere's, s is seconds and kg is kilograms).

Now we've got two partial differential equations that need to be solved, this of course means that we need some boundary conditions, which get by considering how the electric and magnetic fields are related on opposite sides of two media (say 1 and 2). These are easily supplied by:

$$\varepsilon_{1}E_{1}^{\perp} = \varepsilon_{2}E_{2}^{\perp}$$
$$B_{1}^{\perp} = B_{2}^{\perp}$$
$$\vec{E}_{1}^{\perp} = \vec{E}_{2}^{\perp}$$
$$\vec{B}_{1}^{\parallel} = \vec{B}_{1}^{\parallel}$$
$$\mu_{1}^{\parallel} = \frac{\vec{B}_{1}^{\parallel}}{\mu_{2}}$$

Here, the symbol \perp means the perpendicular component of the fields (i.e. perpendicular to the interface between the two media) and \parallel means the parallel components. For the sake of simplicity (and brevity!) we omit the derivation of these boundary conditions, they can be found in pretty much any textbook on classical electrodynamics (see for example Introduction to

Electrodynamics by David J. Griffiths pages 331 to 333 for all the details). In order to get to the point of our model, we only need a few more pieces of background information, so alas, patience!

Next, we propose another assumption about our system to help simply (in fact, greatly) the math. We assume that source of light (in our experimental measurements, this will be a laser) is far enough away from the film so that the incident light is approximately a plane wave. This assumption makes sense, if we have a light source, which is emitting in light in some shape (a good example to keep in mind is that of a light bulb, or the sun, light is emitted in a spherical fashion), once we get far enough away from the source, the light waves will have spread out to the point that any small area on the surface of the wavefront is approximately flat (think of the surface of the Earth, from high up it looks pretty curved, but as you get closer it looks suspiciously flat). This assumption allows us to propose solutions of the wave equations that take the following particularly simple form:

$$\vec{\widetilde{E}}(\vec{r},t) = \vec{\widetilde{E}}_{O}e^{i(\vec{k}\cdot\vec{r}-\omega t)}$$
$$\vec{\widetilde{B}}(\vec{r},t) = \vec{\widetilde{B}}_{O}e^{i(\vec{k}\cdot\vec{r}-\omega t)}$$

Here, we use the tilde to indicate that the quantities (actually they are wave amplitudes) may be complex. It's easy to see that these equations verify the wave equations for E and B. We have also used the notation \vec{k} to represent the "wave vector", which is a vector of magnitude $2\pi/\lambda$, where λ is the wavelength of the wave. This vector points in the direction of propagation of the wave. The quantity ω is the so called "angular frequency" of the wave (measuring how much the wave oscillates, since the complex exponentials are of course sines and cosines). Seeing how these fields are complex, the actual physical fields are simply the real parts.

Finally, the last result we need to consider is the so called "Poynting Vector", \vec{S} . The relation is:

$$\vec{S} = \varepsilon_o c^2 \left(\vec{E} \times \vec{B} \right)$$

This quantity gives the direction of energy propagation of an electromagnetic wave (again, the proof is omitted for brevity; the reference to David J. Griffths' book has all the relevant details pages 345 to 347). Since this vector tells us the direction in which energy is traveling, it tells us in fact the direction of propagation of the wave! Notice that this direction is

perpendicular to both the electric and magnetic fields. In fact, all we have to do is specify one of the field directions and, given the direction of propagation, we can determine the orientation of the remaining field. Traditionally, it is the direction of the electric field that is specified (called the polarization of the wave), and we will follow this tradition in what follows. Using the Poynting vector, we can express the magnetic field in terms of the electric field. It will be convenient to determine the magnitude of the field in what follows. It is readily shown that:

$$E = vB$$

Now we have all the background information that we need. All we need to now (easier said than done) is to apply the boundary conditions that apply to our specific case of a Fabry-Perot interference filter. The general method for attacking problems like this is to setup the magnitudes of the electric and magnetic fields on all sides of the system and taking into account the reflected and transmitted parts. Consider the following:



We assume that our thin film is homogeneous and isotropic, and that the film thickness is of the order of the wavelength of light. In our experiment, the thin films are of ZnS (Zinc Sulphide) and LiF (Lithium Fluoride), both of which are homogeneous and isotropic, and the thicknesses (as will be later determined), are measured in angstroms, which is approximately of the order of the laser light used in this experiment, 632.8 nm. This is important so that the pathlength difference between multiply reflected and transmitted beams remains small compared with the coherence length of the monochromatic light.

Also, the tangential components of incident light on an interface, from Maxwell's equations, must be continuous across the interface. Thus, the boundary conditions for the electric field at the two interfaces (a) and (b) becomes

$$E_a = E_0 + E_{r1} = E_{t1} + E_{i1} \tag{1}$$

$$E_b = E_{i2} + E_{r2} = E_{i2} \tag{2}$$

And the corresponding equations for the magnetic field are:

$$B_a = B_0 \cos \theta_0 - B_{r1} \cos \theta_0 = B_{t1} \cos \theta_{t1} - B_{i1} \cos \theta_{t1}$$
(3)

$$B_b = B_{i2}\cos\theta_{i1} - B_{r2}\cos\theta_{i1} = B_{i2}\cos\theta_{i2}$$

$$\tag{4}$$

We, however, know that the electric field and the magnetic field are related in the following way:

$$B = \frac{E}{v} = \left(\frac{n}{c}\right)E = n\sqrt{\varepsilon_0\mu_0}E$$
, where *n* is the index of refraction.

Thus equations (3) and (4) can be expressed in terms of electric field

$$B_{a} = n_{0}\sqrt{\varepsilon_{0}\mu_{0}}E_{0}\cos\theta_{0} - n_{0}\sqrt{\varepsilon_{0}\mu_{0}}E_{r1}\cos\theta_{0} = n_{1}\sqrt{\varepsilon_{0}\mu_{0}}E_{t1}\cos\theta_{t1} - n_{1}\sqrt{\varepsilon_{0}\mu_{0}}E_{i1}\cos\theta_{t1}$$
$$B_{b} = n_{1}\sqrt{\varepsilon_{0}\mu_{0}}E_{i2}\cos\theta_{t1} - n_{1}\sqrt{\varepsilon_{0}\mu_{0}}E_{r2}\cos\theta_{t1} = n_{s}\sqrt{\varepsilon_{0}\mu_{0}}E_{t2}\cos\theta_{t2}$$

Or, defining $\gamma_0 = n_0 \sqrt{\varepsilon_0 \mu_0} \cos \theta_0$, $\gamma_1 = n_1 \sqrt{\varepsilon_0 \mu_0} \cos \theta_{t1}$ and $\gamma_s = n_s \sqrt{\varepsilon_0 \mu_0} \cos \theta_{t2}$,

$$B_a = \gamma_0 (E_0 - E_{r1}) = \gamma_1 (E_{t1} - E_{i1})$$
(5)

$$B_{b} = \gamma_{1}(E_{i2} - E_{r2}) = \gamma_{s}E_{t2}$$
(6)

Notice now that E_{i2} and E_{i1} differ due to a phase difference φ that develops due to one transversal of the film. We know that the optical path length difference Δ_1 due to one transversal of a film is $\Delta_1 = n_1 l \cos \theta_{i1}$, while the phase difference φ which occurs in this case is

 $\varphi = k_0 \Delta_1 = \left(\frac{2\pi}{\lambda_0}\right) n_1 l \cos \theta_{t_1}$, where k_0 is the wave number associated with the corresponding

wavelength λ_0 .

Thus we have $E_{i2} = E_{i1}e^{-i\varphi}$ (7) and similarly $E_{i1} = E_{r2}e^{-i\varphi}$ (8).

By substituting equations (7) and (8) into equations (2) and (6), we are able to eliminate the fields E_{i2} and E_{r2} in the boundary conditions at the (b) interface:

$$E_{b} = E_{t1}e^{-i\varphi} + E_{r2}e^{-i\varphi} = E_{t2}$$
(9)

$$B_{b} = \gamma_{1} (E_{t1} e^{-i\varphi} - \frac{E_{i1}}{e^{-i\varphi}}) = \gamma_{1} (E_{t1} e^{-i\varphi} - E_{i1} e^{i\varphi}) = \gamma_{s} E_{t2}$$
(10)

Equations (9) and (10) above may be solved to get E_{t1} and E_{t1} in terms of E_b and B_b :

$$E_{t1} = \left(\frac{\gamma_1 E_b + B_b}{2\gamma_1}\right) e^{i\varphi}$$
(11)
$$E_{i1} = \left(\frac{\gamma_1 E_b - B_b}{2\gamma_1}\right) e^{-i\varphi}$$
(12)

Finally, substituting equations (11) and (12) into the expressions (1) and (5) for the boundary conditions of (a), we get

$$E_{a} = \left(\frac{\gamma_{1}E_{b} + B_{b}}{2\gamma_{1}}\right)e^{i\varphi} + \left(\frac{\gamma_{1}E_{b} - B_{b}}{2\gamma_{1}}\right)e^{-i\varphi}$$
$$B_{a} = \gamma_{1}\left[\left(\frac{\gamma_{1}E_{b} + B_{b}}{2\gamma_{1}}\right)e^{i\varphi} - \left(\frac{\gamma_{1}E_{b} - B_{b}}{2\gamma_{1}}\right)e^{-i\varphi}\right]$$

But keeping in mind that Euler's Identities are:

$$2\cos\varphi = e^{i\varphi} + e^{-i\varphi}$$

$$2i\sin\varphi = e^{i\varphi} - e^{-i\varphi}$$

We get:

$$E_a = E_b \cos \varphi + B_b \left(\frac{i \sin \varphi}{\gamma_1}\right) \tag{a}$$

$$B_a = E_b(i\gamma_1 \sin\varphi) + B_b \cos\varphi \tag{b}$$

Equations (a) and (b) may be rewritten in matrix form as

$$\begin{bmatrix} E_a \\ B_a \end{bmatrix} = \begin{bmatrix} \cos\varphi & \frac{i\sin\varphi}{\gamma_1} \\ i\gamma_1\sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} E_b \\ B_b \end{bmatrix}$$
(c)

Or, for normal incidence, $\theta_0 = \theta_{t1} = \theta_{t2} = 0$ and thus $\gamma_0 = n_0 \sqrt{\varepsilon_0 \mu_0}$, $\gamma_1 = n_1 \sqrt{\varepsilon_0 \mu_0}$ and $\gamma_s = n_s \sqrt{\varepsilon_0 \mu_0}$, then also $\varphi = k \Delta_1 = \left(\frac{2\pi}{\lambda}\right) n_1 l$, where φ is the phase difference between each succeeding reflection at each interface.

Thus the matrix equation (c), at normal incidence, becomes

$$\begin{bmatrix} E_a \\ B_a \end{bmatrix} = \begin{bmatrix} \cos\left[\left(\frac{2\pi}{\lambda}\right)n_1l\right] & \frac{i\sin\left[\left(\frac{2\pi}{\lambda}\right)n_1l\right]}{n_1\sqrt{\varepsilon_0\mu_0}} \\ in_1\sqrt{\varepsilon_0\mu_0}\sin\left[\left(\frac{2\pi}{\lambda}\right)n_1l\right] & \cos\left[\left(\frac{2\pi}{\lambda}\right)n_1l\right] \end{bmatrix} \begin{bmatrix} E_b \\ B_b \end{bmatrix}$$
(d)

However, λ is the wavelength of light *inside* the medium. Using Snell's Law, we have λ in terms of λ_0 :

$$\lambda = \frac{\lambda_0}{n_1}$$

and similarly, for a second layer adjacent to the first, with wavelength λ' , we have:

$$n'\lambda' = n_1\lambda \qquad \Rightarrow \qquad \lambda' = \frac{n_1\lambda}{n'} = \frac{n_1\lambda_0}{n'n_1} = \frac{\lambda_0}{n'} \qquad \Rightarrow \qquad \lambda' = \frac{\lambda_0}{n'}$$

Also, since in our experiment we are taking the thickness of the film, l, where λ the wavelength of the light inside the medium, $k\Delta_1$ becomes

$$k\Delta_1 = \left(\frac{2\pi}{\lambda}\right) n_1 l = \frac{2\pi n_1}{\lambda} l$$
 (x)

Thus the matrix equation (d) may be simplified even more to be

$$\begin{bmatrix} E_a \\ B_a \end{bmatrix} = \begin{bmatrix} \cos\left[\frac{2\pi n}{\lambda}l\right] & \frac{i\sin\left[\frac{2\pi n}{\lambda}l\right]}{n_1\sqrt{\varepsilon_0\mu_0}}\\ in_1\sqrt{\varepsilon_0\mu_0}\sin\left[\frac{2\pi n}{\lambda}l\right] & \cos\left[\frac{2\pi n}{\lambda}l\right] \end{bmatrix} \begin{bmatrix} E_b \\ B_b \end{bmatrix}$$

The 2x2 matrix is called the Transfer matrix.

Now, for two different layers of material of a high index of refraction material (H) and a low index of refraction material (L) of thickness $l_H = \frac{\lambda_H}{4} = \frac{\lambda_0}{4n_H}$ and $l_L = \frac{\lambda_L}{4} = \frac{\lambda_0}{4n_L}$ respectively,

where $\lambda_0 = 632.8nm$, and where $n_H(\lambda)$ and $n_L(\lambda)$ are the indices of refraction respectively (as functions of wavelength), we have the following transfer matrices:

$$H = \begin{bmatrix} \cos\left[\frac{2\pi n_{H}(\lambda)}{\lambda}l_{H}\right] & \frac{i\sin\frac{2\pi n_{H}(\lambda)}{\lambda}l_{H}}{n_{H}\sqrt{\varepsilon_{0}\mu_{0}}}\\ in_{H}\sqrt{\varepsilon_{0}\mu_{0}}\sin\left[\frac{2\pi n_{H}(\lambda)}{\lambda}l_{H}\right] & \cos\left[\frac{2\pi n_{H}(\lambda)}{\lambda}l_{H}\right] \end{bmatrix}$$
(h)

$$L = \begin{bmatrix} \cos\left[\frac{2\pi m_{L}(\lambda)}{\lambda}l_{L}\right] & \frac{i\sin\left[\frac{2\pi m_{L}(\lambda)}{\lambda}l_{L}\right]}{n_{L}\sqrt{\varepsilon_{0}\mu_{0}}} \\ in_{L}\sqrt{\varepsilon_{0}\mu_{0}}\sin\left[\frac{2\pi m_{L}(\lambda)}{\lambda}l_{L}\right] & \cos\left[\frac{2\pi m_{L}(\lambda)}{\lambda}l_{L}\right] \end{bmatrix}$$
(1)

A Fabry-Perot interference filter consists of the following arrangement of thin layers,

$(HL)^{m}(HH)^{l}(LH)^{m}$

where H and L are the transfer matrices corresponding to each layer, and m and l are natural numbers. Thus, it is possible to get a resultant matrix, M, after multiplying the corresponding transfer matrices in the order above, with certain entries X, Y, Z and W.

$$M = (HL)^{m} (HH)^{l} (LH)^{m} = \begin{bmatrix} X & Y \\ Z & W \end{bmatrix}$$

Thus the collection of the many layers of the Fabry-Perot Interference filter, in matrix notation, can be treated as one entity, expressed by the resultant matrix M, and namely, for the Fabry-Perot filter, we have from the matrix expression (d)

$$\begin{bmatrix} E_a \\ B_a \end{bmatrix} = M \begin{bmatrix} E_b \\ B_b \end{bmatrix} = \begin{bmatrix} X & Y \\ Z & W \end{bmatrix} \begin{bmatrix} E_b \\ B_b \end{bmatrix}$$

However, utilizing expressions (1), (2), (5) and (6), we get

$$\begin{bmatrix} E_0 + E_{r_1} \\ n_0 \sqrt{\varepsilon_0 \mu_0} (E_0 - E_{r_1}) \end{bmatrix} = \begin{bmatrix} X & Y \\ Z & W \end{bmatrix} \begin{bmatrix} E_{t_2} \\ n_s \sqrt{\varepsilon_0 \mu_0} E_{t_2} \end{bmatrix}$$

This gives us the equivalent system of equations:

$$E_{0} + E_{r1} = XE_{t2} + Yn_{s}\sqrt{\varepsilon_{0}\mu_{0}}E_{t2}$$
(13)

$$n_0 \sqrt{\varepsilon_0 \mu_0} (E_0 + E_{r1}) = Z E_{t2} + W \gamma_s E_{t2}$$
(14)

Next we divide both equations (13) and (14) by E_0 , and noting that the reflection coefficient, r, and transmission coefficient, t, are defined as

$$r = \frac{E_{r1}}{E_0} \quad \text{and} \quad t = \frac{E_{t2}}{E_0}$$

Thus we obtain

$$1 + r = Xt + Yn_s \sqrt{\varepsilon_0 \mu_0} t \tag{15}$$

$$n_0 \sqrt{\varepsilon_0 \mu_0} (1 - r) = Zt + W n_s \sqrt{\varepsilon_0 \mu_0} t$$
⁽¹⁶⁾

Solving the above system of equations (15) and (16) for r and t, we get the following expressions

$$t = \frac{2n_0\sqrt{\varepsilon_0\mu_0}}{n_0\sqrt{\varepsilon_0\mu_0}X + n_0\sqrt{\varepsilon_0\mu_0}n_s\sqrt{\varepsilon_0\mu_0}Y + Z + n_s\sqrt{\varepsilon_0\mu_0}W}$$
$$r = \frac{n_0\sqrt{\varepsilon_0\mu_0}X + n_0\sqrt{\varepsilon_0\mu_0}n_s\sqrt{\varepsilon_0\mu_0}Y - Z - n_s\sqrt{\varepsilon_0\mu_0}W}{n_0\sqrt{\varepsilon_0\mu_0}X + n_0\sqrt{\varepsilon_0\mu_0}n_s\sqrt{\varepsilon_0\mu_0}Y + Z + n_s\sqrt{\varepsilon_0\mu_0}W}$$

The reflectance, R, and transmittance, T, are defined as

$$R = |r|^2 = r \cdot r^*$$
 $T = |t|^2 = t \cdot t^*$

The transmittance is a measure of how much light is transmitted through the medium; it can take values between 0 and 1. Similarly, R is a measure of how much light is reflected from the medium, it can also only take values between 0 and 1. In fact, R + T = 1. Transmittance is generally defined as the ratio of the intensity of light that made it through the medium to the intensity of light incident on the medium (where "intensity" is a measure of the time average "energy flux". In our present case, it is the time average of the energy being transported by the electromagnetic fields as they propagate through space).

We have written a MATLAB code that will allow use to make a few simulations of what we expect to see for various values of m and l (remember than the number m and l control how many layers). The following graphs demonstrate what the theory predicts for a few values of m and l (note that we also need to know the indices of refraction of the layers used, see page 20 for all the details):













The last graph shown above is very encouraging. It shows that our simulations predict a constant transmittance through optical system if we don't have any layers of high and low index of refraction. This makes perfect sense physically, since if there are no layers of material, then all we have is the glass substrate upon which the system is sitting, in other words, we just have one medium for light to travel through, hence we shouldn't expect anything other than a constant transmittance! We should also note that the order of the layers is important, and it must follow the convention set out on page 10. If we don't use such a convention, we no longer have an interference filter (in fact, if one were to follow a layering pattern like $g(HL)^nHa$, you would end up with a device called a dielectric mirror. Unlike our interference filter, that tries to block out a range of wavelengths, while letting one pass through, the dielectric mirror attempts to reflected all the light in a certain range)!

If m > l, the interference filter blocks the transmission of light in a certain range, allowing for the transmission of light in a narrow region around the peak (this is called the band pass). The peak transmission occurs for one wavelength in this range, in our case this corresponds to 632.8 nanometers, since this is the wavelength of light we are going to use in our experiment, it makes sense that we should construct our simulations around this value! The phase shift at $\lambda_0 = 632.8nm$ is an integer multiple of the wavelength, which results in a maximum transmission. Phase shifts at all other wavelengths are fractional values, which do not provide maximum transmission, as can be seen from the results of the simulation.

Also, as can be seen from the graphs where m>l, the spectral transmissions of each of those Fabry-Perot interference filter is such that wavelengths in the range of ~530nm to ~750nm are blocked, with a transmission peak centered at ~630nm. Thus, in this experiment, the Fabry-Perot filter (designed for a wavelength of $\lambda_0 = 632.8nm$) blocks wavelengths from most of the visible spectrum and a little of the infrared spectrum (green light to infrared). Hence depending on the value for which the Fabry-Perot filter is designed, not only is one able to choose the wavelength that will be transmitted (transmission peak), but also the range of wavelengths to be blocked.

Notice that as m increase (still in the cases where m>l), the transmission peak around our special wavelength of 632.8 nm becomes sharper. This means that, as we increase the value of m, the filter blocks out more and more of the wavelengths in the surrounding spectral region and we see that the wavelength that we have constructed the system around really becomes the only significant source of transmission, just as we want it to! In other words, increasing m decreases the band pass. As m continues to increase (such as in the cases for which m = 13 and m = 100), the transmission peak vanishes nothing at all is transmitted in the range of about 550 nm to 750 nm (i.e. the optical range), in fact for the case l = 2, we find by experimenting with various values of m in our simulation code, that complete blockage starts for m = 8.

Lastly, if l>m, the number of transmission peaks increases. In fact, this number of peaks seems to correspond approximately to the value chosen for l, which are all evenly spaced. In case that m = 4, l = 21, however, the number of peaks is smaller, the band pass for each is narrower, the minimum transmission for the wavelengths blocked is also lower, and these regions of blocked wavelengths are also more sharply defined (this can be attributed to an increase in m, as noted above). Thus, the number, band pass, and sharpness of transmission peaks in a region of blocked wavelengths may be characterized by the ratio of m to l.

Armed with our simulations, we would now like to see how these compare to a real situation. We will construct two Fabry-Perot interferences filter with m = 3 and l = 1 and measure their transmittance as a function of wavelength. In order to achieve this we make use of a high vacuum deposition technique, in which material (ZnS and LiF) are evaporated and deposited onto two glass slides. First, we need to determine how thick we need our layers to be if we are to use light of 632.8 nm.

Using $n_0 = 1$ (that of air), the material of high index of refraction is ZnS and the material of low index of refraction is LiF, where n_{ZnS} and n_{LiF} vary as a function of wavelength in the following way²

$$n_{ZnS} = \sqrt{5.131 + \frac{1.275 \cdot 10^7}{\lambda^2 - 0.732 \cdot 10^7}}$$

where λ is measured in angstroms (10⁻¹⁰m), and

$$n_{LiF} = \sqrt{1 + \frac{0.92549 \cdot \lambda^2}{\lambda^2 - (0.07376)^2} + \frac{6.96747 \cdot \lambda^2}{\lambda^2 - (32.79)^2}}$$

where λ is measured in micrometers (10⁻⁶m).

For this experiment, $\lambda_0 = 632.8nm$, therefore the thicknesses of the ZnS and the LiF layers are

$$l_{ZnS} = \frac{\lambda_{ZnS}}{4} = \frac{\lambda_0}{4n_{ZnS}} = \frac{\lambda_0}{4 \cdot \sqrt{5.131 + \frac{1.275 \cdot 10^7}{\lambda_0^2 - 0.732 \cdot 10^7}}}$$
(!)
$$= \frac{6328 ang.}{4 \cdot \sqrt{5.131 + \frac{1.275 \cdot 10^7}{(6328 ang.)^2 - 0.732 \cdot 10^7}}} = 673 ang. = 0.673 kiloang$$
$$l_{LiF} = \frac{\lambda_{LiF}}{4} = \frac{\lambda_0}{4n_{LiF}} = \frac{\lambda_0}{4 \cdot \sqrt{1 + \frac{0.92549 \cdot \lambda^2}{\lambda^2 - (0.07376)^2} + \frac{6.96747 \cdot \lambda^2}{\lambda^2 - (32.79)^2}}} = \frac{6.328 \cdot 10^{-5} \mu m}{4 \cdot \sqrt{1 + \frac{0.92549 \cdot (0.6328 \mu m)^2}{(0.6328 m)^2 - (0.07376)^2} + \frac{6.96747 \cdot (0.6328 \mu m)^2}{(0.6328 \mu m)^2 - (32.79)^2}}}$$

 $= 1.138 \cdot 10^{-1} \mu m = 1.138 kiloang$

² See Palik and Hunter and E. Palik and A Addamiano or E. Palik. "Handbook of Optical Constants of Solids. Vol. I and II. Academic Press. 1985/1989.

And the indices of refraction at this specific wavelength are

$$n_{ZnS}(\lambda_0) = \sqrt{5.131 + \frac{1.275 \cdot 10^7}{\lambda_0^2 - 0.732 \cdot 10^7}} = \sqrt{5.131 + \frac{1.275 \cdot 10^7}{(6328ang.)^2 - 0.732 \cdot 10^7}} = 2.350$$

And

$$n_{LiF}(\lambda_0) = \sqrt{1 + \frac{0.92549 \cdot \lambda_0^2}{\lambda_0^2 - (0.07376)^2} + \frac{6.96747 \cdot \lambda_0^2}{\lambda_0^2 - (32.79)^2}}$$
$$= \sqrt{1 + \frac{0.92549 \cdot (0.6328 \mu m)^2}{(0.6328 \mu m)^2 - (0.07376)^2} + \frac{6.96747 \cdot (0.6328 \mu m)^2}{(0.6328 \mu m)^2 - (32.79)^2}} = 1.391$$

In order to actually evaporate the material, we place it inside of a crucible within the vacuum chamber. We wrap wires around the crucible and (once the vacuum has been established) pass a current through the wires. This has the effect of heating up the material to the point (eventually) of evaporation. The evaporated material then rises from its crucible and encounters the glass slides that are placed above it. In order to determine the thickness of the layers (which is supposed to be precise), we use a thickness monitor that is placed near the glass slides (the monitor is also able to tell us the rate at which the material is being deposited). Once the desired thickness is achieved, a shutter inside the evaporator is closed; this shutter is place between the glass slides and the material, so that it can block the further deposition of material onto the slides. We can then do the exact same thing for the second material, and proceeding in this way, depositing layer after layer, we can construct our interference filter. Of course, like any experiment, this procedure is riddled with potential problems, which will be discussed later.

Once we have our filters, we can quite try to measure their transmission spectrum to see if they match what our simulations predict. In order to do this we shine light from a Mercury lamp (which has a structured spectrum, this way we can test our filter at various different wavelengths at the same time) on our filters. The light is then collected and fed into a spectrometer. The spectrometer can scan over a range of wavelengths (in our experiment this range corresponds to 380 nm to 800 nm, which is from the violet end of the spectrum into the ultra violet), and pass whatever light it gets through at each wavelength to a photovoltaic cell. This cell will then produce a current that is proportional to the intensity of light that has been received. The cell functions based on the photoelectric effect which demonstrates that light on a metallic surface will produce a current, and that this current is proportional to the intensity of light incident on the surface. Now that we have a current, we can measure a voltage (since there is a resistance as well), and this value is recorded by a computer.

The spectrometer scans wavelengths at an interval of 100nm/min, taking a measurement about every 0.1 seconds. With this scan rate (and knowing that we have scanned over an interval of 380 nm to 800 nm); we can determine exactly which wavelengths of light were tested as follows:

$$\lambda(t) = 380nm + (100nm / min) \cdot t = 380nm + (\frac{5}{3}nm / sec) \cdot t$$

With this, we can associate that each scanned wavelength with a voltage measured by the computer. We know that the voltages are proportional to the transmittance, but we still need to determine exactly how. In order to do this, we do the same experiment expect with a clear glass slide. We then take the voltages measured for the interference filters at each wavelength and divide each one by the corresponding voltage measured at the same wavelength for the clear slide. This ratio tells us difference in the proportion of light that was transmitted through the filter compared to when the filter was not present for each wavelength measured, in other words this ratio is exactly the transmittance value we are looking for!

When building our filters we named them "window" and "door". The names come from the proximity of the filters to either a window in the lab or the door to lab as they sat inside the evaporator. The reason for naming them is to be able to keep track of the results for each individual filter and to see if their location inside the evaporator had any effect on their overall behaviour (which it most certainly did). Below, we show graphs of the results obtained for each filter plotted against the predicted graph of our simulation for m = 3, l = 1, as well as our simulated graph:





("Door")



("Window")

We summarize some of the relevant features in the following tables:

Theoretical Fabry-Perot		Constructed Fabry-Perot		Constructed Fabry-Perot	
		(Door)		(Window)	
Blocked	With the	Blocked	With the	Blocked	With the
Spectrum	exception of a	Spectrum	exception of a	Spectrum (nm)	exception of a
(nm)	Peak centered	(nm)	Peak centered		Peak centered
	at about (nm)		at about (nm)		at about (nm)
500-800	632.8	490-780	571.7	430-780	659.9

|--|

Table 2: Varied Parameters for the Theoretical Simulation of the Fabry-Perot Interference Filter

Simulation for F	ilter near DOOR	Simulation for Filter near WINDOW		
Inputted thickness of	Inputted thickness of	Inputted thickness of	Inputted thickness	
ZnS layer	LiF layer	ZnS layer	of LiF layer	
(kiloangstroms)	(kiloangstroms)	(kiloangstroms)	(kiloangstroms)	
0.479	0.786	0.553	0.911	

Theoretical Fabry-Perot		Constructed Fabry-Perot		Constructed Fabry-Perot	
		(Door)		(Window)	
Maximum	Minimum	Maximum	Minimum	Maximum	Minimum
Transmittance	Transmittance	Transmittance	Transmittance	Transmittance	Transmittance
0.630	0.02	0.240	0	0.531	0
Percent Difference of Maximum Transmittance from Theoretical Value (%)		61.9		15.7	

Table 3: Maximum Transmittance Data of the Fabry-Perot Interference Filter

The values in these tables deserve some explanation, since they tell us a lot about what is happening physically.

First of all, if we take a look at the graphs comparing the experimental data with the predictions of our simulation, we can see that they don't exactly match. Indeed the data for the "door" filter is significantly different from the simulated curve; at least this is how it seems at first glance. However, it one is to look closely at the curves, we can see that, although the peaks are not at the same height nor are they centered at exactly the right wavelength, the overall behaviour of the systems are exactly the same. There are many reason why the curves don't line up exactly though, and we will discuss some of them in detail. First of all, it should be noted that, in the end of the day, the reason why are simulations don't match with the experimental values exactly is because are simulations are of an idealized world, in which all things are equal, unfortunately for us the real world is hardly ever so neatly packaged.

As is summarized in table 3, the minimum transmittances for the simulated data and the experimental data fit pretty closely (0.02 compared with 0), but the major issue of concern is the maximum transmittances. We can see that the maximum for the "door" filter is 0.240, whereas for the "window" filter it is 0.531. When compared to the simulated value of 0.630, these don't seem so great; in fact they lead to percent errors of:

$$\% difference = \frac{| theo.value - exp.value |}{theo.value} \cdot 100\% = \frac{| 0.630 - 0.240 |}{0.630} \cdot 100\% = 61.9\%$$
And
$$\% difference = \frac{| 0.630 - 0.531 |}{0.630} \cdot 100\% = 15.7\%$$

While the first error (for the "door" filter) is not so great, the last one for the "window" filter actually isn't that bad considering all the possible source of error, of which there are many.

One of the primary reasons is the differences in thickness that seem to have arisen between the values that we wanted to build for each layer, compared to the projected thickness' that were actually built. These results are summarized in table 2, we know that (as calculated above) that the desired thickness values were 0.673 kilo-angstroms for ZnS and 1.138 kiloangstroms for LiF, whereas it seems like the actual thickness' were closer to 0.479 and 0.786 kilo-angstroms for the "door" filter and 0.533 and 0.911 kilo-angstroms for the "window" filter. These values were extrapolated from our simulations by taking the experimental results for the peak transmission wavelengths (summarized in table 1) and outputting calculated thickness' (these outputs were suppressed in the original program because we already knew what the thickness' should be). These values probably aren't exactly the real thicknesses values, but they at least give us an approximation to tell us roughly how much we were off in our construction of the filters. In order to determine the actual thickness, we would have to go back to the lab and measure them using a laser (actually, measuring the thicknesses is not far off from what we have done in our experiment, the idea is to use light interference again in a cleaver way, this topic is covered in pretty much any standard textbook on optics). The new plots, of the simulations fitted to the experimental data, are shown below:



("door")



The differences in the thicknesses can help to account for the shift in peak transmission wavelength, but how exactly did these difference occur? This problem is readily explained simply by the way in which the filters were constructed. When we built these things, we had to be able to close the "shutter" quickly once the thickness monitor had indicated that the desired thickness had been achieved, the shutter was closed by hand. This leads to a lot of uncertainty because we couldn't actually see the shutter closing, we had to estimate its position, and the only indication that deposition had been successfully blocked was when the thickness monitor stopped recording any changes. This is another major problem, because the detector inside the apparatus was placed off to the side of the glass plates. This means that, while the shutter may have been completely blocking the detector, it may not actually have been blocking both slides. Conversely, when the detector was uncovered and reading changes in thickness, there is no guarantee that the slides were completely uncovered, so it is extremely difficult to say whether or not the layer thickness measured by the monitor is exactly what got deposited on the slides.

There are some other problems that we have to consider as well. For one, the simulations were designed assuming that there was nothing in the surrounding environment other than air,

this is of course a pretty grand approximation. In fact, any little bit of dirt of dust on the filters could greatly affect the measurement of transmission. A little bit of dust wouldn't make much of a difference from an optical point of view, but from the stance of a precision measurement, a layer of dust can certainly block a noticeable amount of light from getting through the filter into the detector. Unfortunately, we can't really account for this since we can't measure how much dust is on the surface.

This dust/dirt also possess another problem, if dust is to collect on the surface of the filter and one tries to wipe it off, there is actual physical damage being done to the filters. It may not seem like it from the point of view of someone who cleaning the filter, but remember that the layers of deposited material have thicknesses on the order of the wavelength of light! Any small disturbance will actually physically damage the surface by dislodging some of the material from the top layer, this has the effect of introducing in-homogeneities into the material which, if you recall, was one of our primary assumptions in the construction of our model!

That being said, we can't definitely say that our layers were even deposited completely evenly either! Since we have no control over the rate at which material is evaporated onto our slides, it is entirely possible that some regions of the filters received more material than others, which again will cause in-homogeneities in our filters, ultimately affected our final results.

In order to try and take this into account, we have written a program that attempts to correct for some variations of the thickness of layers. The program allows the user to into certain tolerances (or errors) on the thicknesses, as well as a theoretical peak wavelength. The program then generates some thicknesses as follows. Suppose the error given by the user is Δl (which is input as a fraction), then the program computes (letting l be the thickness):

$$l - l \cdot \Delta l$$
$$2l \cdot \Delta l$$

The last line being the length of the error "interval". Added together, we get:

$$l - l \cdot \Delta l + 2l \cdot \Delta l$$

We can then generate thicknesses inside this interval by using the random number generator in MATLAB, this has the effected of simulating a pseudo-random thickness for the high and low layers. While this obviously doesn't deal with random (and in-homogenous) alternating layers of material, it at least gives us some idea of what happens when the layer thicknesses are shifted

around a bit from their theoretical ideal which, let's face it, is extremely difficult to achieve in practice without very expensive equipment. The final expression we use for the simulation is hence:

$$l - l \cdot \Delta l + 2l \cdot \Delta l \cdot rand(1,1)$$

Where we apply this to both the high and low layer (based on the user input for the Δl value). For example, our program allows for user input:



We can choose the theoretical wavelength for which we would like to construct our filter, this produces:



We can then enter the tolerance range that we are interested in (i.e. the fraction error that we would like to us). Once we have entered these values, we push the plot button and we get back a plot with the error taken into account.





The clear button allows the user to clear the graph and start again fresh if they would like. The exit button closes the program.

Despite the experimental issues we have seen so far, we can generalize our simulations to take into account any theoretical wavelength as well as the values for the indices of refraction for the substrate materials (in our experiment, and earlier simulations these were air and glass) as well as the number of layers used.



The program shown above allows the user to alter the theoretical wavelength using a slider, as well as the other values shown. Every time a value is changed, the graph is updated automatically to the new one.



The graphs on the right hand side are plots of the indices of refraction of the high and low layers. These values depend on wavelength and this is taken into account by the program. Every time the wavelength is changed by the user, the program automatically moves the red cursors on these graphs to point to the current indices of refraction. Here are some examples, which were generated by varying the parameters:



(Here, the initial wavelength has been changed, notice how it shifts the transmission peak to the new value, as it should (or else it wouldn't be much of a transmittance maximum)).



(Here the parameter m has been changed; this is equivalent to varying the number of layers used).



(This time, the parameter 1 is varied, this too is equivalent to varying the layers used (i.e. the number of high refractive index layers in the middle of the system).



(This one above has a different initial refractive index than the others; here it is 1.43 as opposed to 1 in the other screen shots, notice how it changes the transmission curve).



(This screen shot has an updated value for the final refractive index, again notice how it drastically changes the transmission graph (i.e. having a higher substrate index will decrease the amount of light transmitted through the system!).

The exit button closes the program.

Finally, we should consider another one of our assumptions. We designed our interference filters to have peak transmission at one particular wavelength (632.8 nm), this is not some arbitrary number, and in fact it is, conveniently enough, the wavelength of light produced by a He-Ne laser. In theory, if we hit our filter with such a beam we would see exactly the transmission predicted (0.63) and nothing else. However, in reality this is not the case, since even a laser is not this perfect, in fact most laser beams are what is called a Gaussian beam. It can be shown that³ the intensity of a Gaussian beam is given by:

$$I(r,z) = I_0 \left(\frac{\omega_0}{\omega(z)}\right)^2 \exp\left(\frac{-2r}{\omega^2(z)}\right)$$

Where I_o is the intensity at the centre of the beam, $\omega(z)$ is the radius at which the field amplitude and intensity drop to e⁻¹ and e⁻² of their axial values, respectively, ω_0 is the so called "waist size" (this is the location along the beam axis where $\omega(z)$ is at its minimum value) and r is the distance

³ See for example, Pedrotti³, Introduction to Optics, Pearson Education inc, 2007. In particular, chapter 27 on characteristics of laser beams.

from the centre axis of the beam. The take home message is that the intensity is not uniformly distributed over the spread of the beam! Indeed, the transmittance that we have been measuring is simply the ratio of transmitted light intensity (how much got through) to incident light intensity. Thus, if the incident light intensity is not uniformly spread along the beam, the transmittance measured will not necessarily correspond to what we had predicted theoretically (since the incident intensity will be essentially "spread out"). The following program demonstrates the Gaussian beam graphically.





We can choose any values we like for the min and max values on the axes.

Moving the sliders around changes the relevant parameters and the graph automatically updates to take these changes into account. Here are a few plots for some values, notice how the curve can be really flat a spread out, or can be very sharp, peaked about a small region (this achieved by varying the w0 and lambda0 parameters).







Moving the I0 slider around changes the z axis. Also, we can change the min and max values to whatever we would like, here is an example:



Anytime one of the parameters is varied, the plot updates automatically to take these changes into account. The exit button allows the user to close the program whenever he or she so desires.

One of the major applications of interference filters (our Fabry-Perot style filter finding application here as well), is in astronomy. A branch of astronomy called photometry is concerned with measuring the intensity of astrophysical phenomena. If one is able to determine the intensity of some radiating phenomena (a classic example is variable stars), and the luminosity of this object is known then its distance may be calculated via an inverse square law. Indeed, one of the best known examples is that of a type 1a supernova, which always erupts with the same, known, luminosity (the reason for this is that a type 1a supernova occurs on white dwarf stars that reach a certain mass (by sucking material off of a companion star), once this mass has been achieved there is enough mass for thermonuclear fusion to being again (a white dwarf star is for intensive purpose a dead star in which all nuclear fusion has ceased), the result is massive nuclear reaction that always produces the same amount of energy and so the luminosity may be calculated quite accurately). If we measure the intensity of this supernova (using an optical device that makes use of our handy interference filters) we can determine the distance of the supernova. This is one of the primary methods for determining distance in the universe. This is just one of many other applications of such filters, other notable applications are in fiber optics, and as mentioned previously, in solar cells.

In conclusion, our project presents a model and simulations for the transmission of light through a Fabry-Perot interference filter. We found that simulations predicted somewhat different results than were measured in a laboratory experiment on the real thing. We noted however, that the overall behaviour of our simulations matched rather well with what was actually observed and that the primary reason for our simulations to not predict exactly what would happen in the lab is several sources of massive experimental error. Of course, our simulation and model is like any other, simply an approximation of the physical world. Physical phenomena are generally of a rather complex nature, with many variables and so many, seemingly, unpredictable possibilities, that all we could ever hope for is a halfway decent approximation. If we were to improve our experimental technique, we feel confident that our simulations would better represent the situation. For example, if were to construct and store our filters in a highly sanitized environment, free of dust and dirt (as in customary in precision optics) we could eliminate a major source of error. Better construction techniques would also be required to reduce the errors we encountered in the thickness of the layers. Nevertheless, our model serves to demonstrate and helps to understand a very interesting and applicable phenomenon.