The Fabry-Perot Interferometer

The FabryPerot interferometer, simply referred to as the Fabry-Perot, is an important application of multiple wave interference in optics. It consists of two partially reflecting surfaces aligned with each other in such a way that many waves of light derived from the same incident wave can interfere. The resulting interference patterns may be used to analyze the spectral character of the incident beam.

4.1 The Fabry-Perot Equation

Consider a Fabry-Perot consisting of two parallel reflecting surfaces, separated by a distance d as shown in Figure (4.1)[1]. Let n be the index of refraction of the medium between the mirrors.

A plane monochromatic wave is incident on the FabryPerot plate at an angle ϕ. Let $\overline{E_0A_1}$ be the ray representing the direction of propagation of the incident wave. At the first surface, this wave is divided into two plane waves, one reflected in the direction of $\overline{A_1E_1'}$, and the other transmitted into the plate in the direction $\overline{A_1B_1}$. This latter wave is incident on the second surface at angle θ and is there divided into two plane waves, one transmitted in the direction $\overline{B_1E_1}$, the other reflected back at the Fabry-Perot in the direction $\overline{B_1A_2}$. This process of division of the wave remaining inside the plate continues as shown in Figure (4.1). The total transmitted field can be calculated by adding the contributions from each of the transmitted waves. To carry out this sum we need to know their relative phases and amplitudes. These can be calculated as follows.

We note that the phase of each transmitted wave differs from that of the preceding

[1] In its simplest form a Fabry-Perot could be a glass plate of fixed thickness with parallel polished surfaces. This form of the Fabry-Perot is referred to as an etalon.

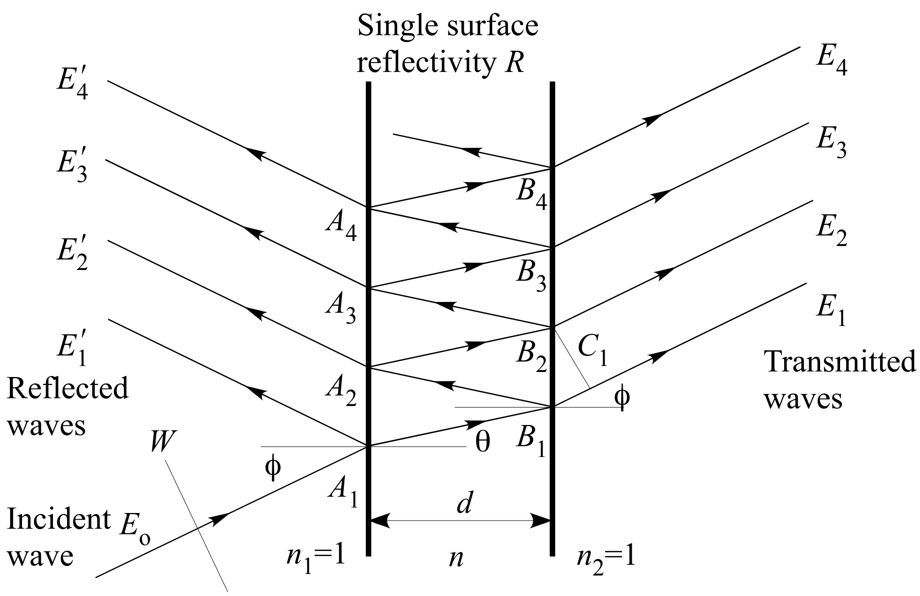


Figure 4.1: A Fabry-Perot gives rise to multiple beam interference both in transmission and reflection.

wave by an amount corresponding to the difference in their paths. For example, the optical paths of E_2 and E_1 differ by

$$n\,\overline{B_1A_2B_2} - \overline{B_1C_1} = \frac{2nd}{\cos\phi} - \frac{2d}{\cos\phi}\sin\phi\sin\theta = \frac{2nd}{\cos\phi} - \frac{2d}{\cos\phi}\sin\phi(n\sin\phi)$$
$$= \frac{2nd}{\cos\phi}\left(1-\sin^2\phi\right) = 2nd\cos\phi\,. \tag{4.1}$$

This path difference corresponds to a phase difference

$$2\delta = 2\pi \times \frac{2nd\cos\theta}{\lambda} = \frac{4\pi nd\cos\theta}{\lambda} = \frac{\omega 2nd\cos\theta}{c}\,. \tag{4.2}$$

We can now write down the relative amplitudes of transmitted waves. We shall assume that both reflecting surfaces are identical and each surface, when it alone is present, reflects a fraction $\mathcal{R}$ of the intensity of the light

$$\frac{I_{R}}{I_{inc}} = \mathcal{R}\,. \tag{4.3}$$

Then the ratios of reflected-to-incident and transmitted-to-incident electric field amplitudes at each interface are [2]

$$\frac{E_{ref}}{E_{inc}} = -\sqrt{\mathcal{R}}\,, \tag{4.4}$$

$$\frac{E_{trans}}{E_{inc}} = \sqrt{1-\mathcal{R}}\,. \tag{4.5}$$

We are now ready to calculate the transmission of the Fabry-Perot. We will write down all the fields at the same instant of time t.

The field of a plane wave is of the form

$$E(t) = E_0 e^{-i\omega t + ikz} \tag{4.6}$$

where $k = 2\pi n/\lambda$ and z is the distance of propagation. Let us set $z = 0$ at point A_1, at the input face so that the incident wave can be written as

$$E_{inc} = E_0 e^{-i\omega t}\,. \tag{4.7}$$

The transmitted field amplitude across the first surface, according to Eq.(4.5), is $\sqrt{1-\mathcal{R}}\ E_0$ and after crossing the second surface, it is $\sqrt{1-\mathcal{R}}\ \cdot \sqrt{1-\mathcal{R}}E_0 = (1-\mathcal{R})\ E_0$. The transmitted field E_1 at B_1 is then

$$E_1 = (1-\mathcal{R})E_0 e^{-i(\omega t - k\,\overline{A_1B_1})}\,. \tag{4.8}$$

where the distance $\overline{A_1B_1} = d/\cos\phi$. Writing the constant phase $k\,\overline{A_1B_1} = kd/\cos\phi = \delta_0$ we can write E_1 as

$$E_1 = (1-\mathcal{R})E_0 e^{-i(\omega t - \delta_0)}\,. \tag{4.9}$$

To write down E_2 we note that the wave transmitted across the first surface is reflected twice inside the plate before being transmitted. Its amplitude will thus gather amplitude factors $\sqrt{1-\mathcal{R}}\,(\sqrt{\mathcal{R}})\,(\sqrt{\mathcal{R}})\,\sqrt{1-\mathcal{R}} = (1-\mathcal{R})\mathcal{R}$ and an additional phase factor $e^{i2\delta}$ relative to E_1,

$$E_2 = (1-\mathcal{R})\mathcal{R}E_0 e^{-i(\omega t - \delta_0)} e^{i2\delta} \tag{4.10}$$

[2] The minus sign takes into account the phase change at reflection from a denser medium.

Similar considerations for other waves lead to the following expressions for the transmitted waves

$$\begin{aligned}
E_1 &= E_0(1-\mathcal{R})e^{-i(\omega t-\delta_0)} \\
E_2 &= E_0(1-\mathcal{R})\mathcal{R}e^{-i(\omega t-\delta_0)}e^{i2\delta} \\
E_3 &= E_0(1-\mathcal{R})\mathcal{R}^2e^{-i(\omega t-\delta_0)}e^{i4\delta} \\
E_4 &= E_0(1-\mathcal{R})\mathcal{R}^3e^{-i(\omega t-\delta_0)}e^{i6\delta} \\
&\cdots \\
E_N &= E_0(1-\mathcal{R})\mathcal{R}^{N-1}e^{-i(\omega t-\delta_0)}e^{i(N-1)2\delta}
\end{aligned} \tag{4.11}$$

If the incident wave and the plates are wide enough and reflectivity is high there will be a large number of contributions. For all practical purposes we can take the number of transmitted waves to be infinitely large. The total transmitted field is then obtained by summing the infinite geometric series [3]

$$\begin{aligned}
E_T &= E_1 + E_2 + E_3 + \cdots \\
&= E_0(1-\mathcal{R})e^{-i(\omega t-\delta_0)}\left(1+\mathcal{R}e^{i2\delta}+\mathcal{R}^2e^{i4\delta}+\cdots\right) \\
&= \frac{E_0(1-\mathcal{R})e^{-i(\omega t-\delta_0)}}{1-\mathcal{R}e^{i2\delta}}\,.
\end{aligned} \tag{4.12}$$

Since the (time averaged) intensity is proportional to the modulus squared of the field amplitude, the transmitted intensity I_T is given in terms of the incident wave intensity I_0 as

$$I_T = \frac{I_0(1-\mathcal{R})^2}{1+\mathcal{R}^2-2\mathcal{R}\cos 2\delta} = \frac{I_0(1-\mathcal{R})^2}{1+\mathcal{R}^2-2\mathcal{R}+2\mathcal{R}(1-\cos 2\delta)} \tag{4.13}$$

$$= \frac{I_0(1-\mathcal{R})^2}{(1-\mathcal{R})^2+4\mathcal{R}\sin^2\delta} \tag{4.14}$$

This can be written in the form

$$I_T = \frac{I_0}{1+F\sin^2\delta}\,, \tag{4.15}$$

where F is the coefficient of finesse

$$F = \frac{4\mathcal{R}}{(1-\mathcal{R})^2}\,, \tag{4.16}$$

[3] An infinite geometric series $1+x+x^2+x^3+\cdots$ with $x<1$ has the sum $1/(1-x)$.

and the phase δ from Eq. (4.2) is given by

$$\delta = \frac{2\pi n d \cos\theta}{\lambda} . \tag{4.17}$$

From Eq. (4.15) we see that the transmitted intensity is a periodic function of δ that varies between a maximum and a minimum as δ changes

$$[I_T]_{max} = I_0 , \qquad \delta = p\pi , \quad p \text{ an integer} \tag{4.18}$$

$$[I_T]_{min} = \frac{I_0}{1+F} , \qquad \delta = \left(p + \frac{1}{2}\right)\pi \tag{4.19}$$

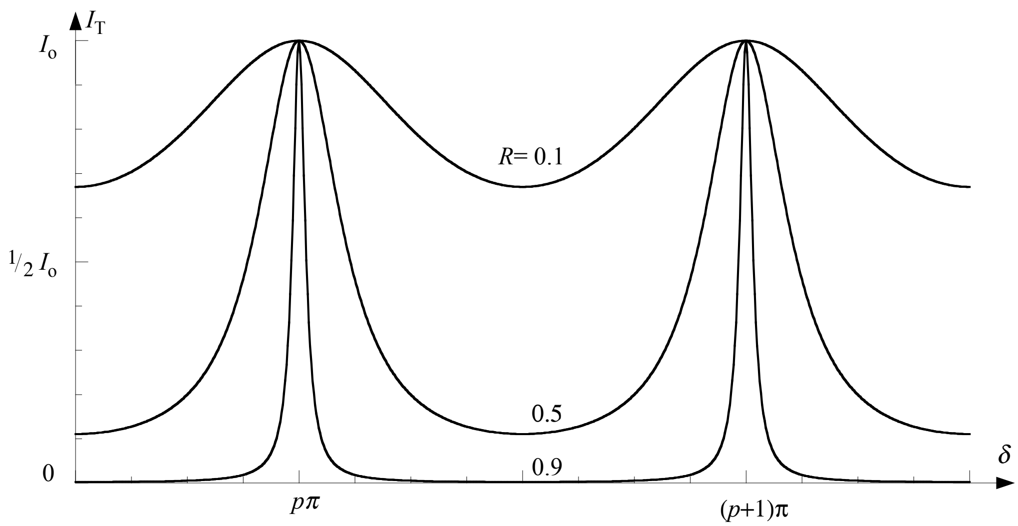


Figure 4.2: Fabry-Perot transmission as a function of δ.

Figure (4.2) shows the transmitted intensity I_T as a function of δ. Note that the peaks get narrower as the mirror reflectivity (and therefore the coefficient of finesse F) increases. When peaks are very narrow, light can be transmitted only if the plate separation d, refractive index n, and the wavelength λ satisfy the precise relation

$$\delta = \frac{2\pi n d \cos\theta}{\lambda} = \text{integer} \times \pi \equiv p\pi , \tag{4.20}$$

otherwise no light is transmitted. It is this property that permits the Fabry-Perot to act as very **narrow band-pass filter** for fixed d.

If the incident light contains many wavelengths of varying intensities, we can analyze its spectrum (wavelength/frequency and intensity) by scanning the length d of the Fabry-Perot because for a given separation d, the Fabry-Perot transmits only the wavelength that satisfies Eq. (4.20). In this mode the Fabry-Perot is referred to as a spectrum analyzer. Also note that as we scan the Fabry-Perot the transmission pattern will repeat when δ increases by π.

Integer p in Eq. (4.20) is referred to as the order of the transmission peak (or the fringe). Note that p has its maximum value for $\theta = 0$. If the width of the pump beam was very large, and all the rays of light were incident at the same angle, we would not receive any light at angles other than those satisfying Eq. (4.20). In practice, the incident light often has some divergence so that the relation (4.20) is satisfied for other angles as well [see Figure (4.3)]. Consequently, in transmitted light we will see concentric rings corresponding to rays entering the FabryPerot at angles θ_1 and θ_2, if Eq. (4.20) is satisfied.

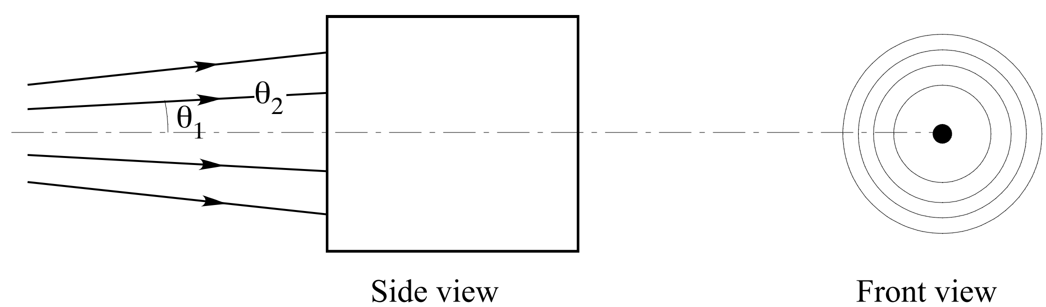


Figure 4.3: The output of a Fabry-Perot illuminated by a diverging set of rays consists of concentric rings.

4.2 FabryPerot Finesse

The peaks in Figure (4.2) are not infinitely sharp because the surfaces cannot be made perfectly reflecting. This limits the instrument's (spectral) resolution [Its ability to tell two closely spaced wavelengths or frequencies apart]. A number $\mathcal{F}$ called "finesse" describes the resolution of the instrument. Its significance is shown in Figure (4.4). Note that successive transmission maxima are separated by $\Delta\delta = \pi$ and the peak width δ_c is defined be the full width at half maximum (FWHM). The finesse for the instrument is defined as the maximum number of resolvable peaks that can be inserted in the interval π,

$$\mathcal{F} = \frac{\pi}{\delta_c} \tag{4.21}$$

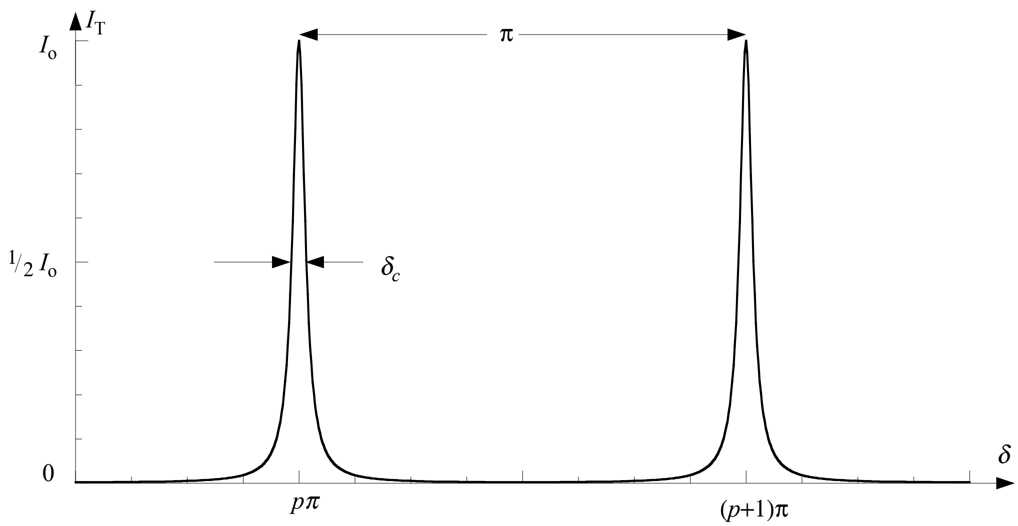


Figure 4.4: Finesse is a measure of the sharpness of transmission peaks.

To determine δ_c, we look for the values of δ, for which the transmitted intensity is reduced to half of its peak value [See Figure (4.5)]. Near a transmission peak of integral order p, the points where the intensity falls to half of its maximum value are at $\delta = p\pi \pm \frac{1}{2}\delta_c$. At these points we obtain from Eq. (4.15)

$$\frac{1}{2}I_0 = \frac{I_0}{1 + F\sin^2(\frac{1}{2}\delta_c)}\,. \tag{4.22}$$

This equation easily gives

$$\sin\left(\frac{1}{2}\delta_c\right) = \frac{1}{\sqrt{F}} = \frac{1-\mathcal{R}}{2\sqrt{\mathcal{R}}} \tag{4.23}$$

In most cases of practical interest, the width of the peak is small compared to the free spectral range, $\delta_c \ll \pi$. This is the case when F is large. We can then use the approximation $\sin(\frac{1}{2}\delta_c) \approx \frac{1}{2}\delta_c$ to obtain

$$\delta_c = \frac{2}{\sqrt{F}} = \frac{(1-\mathcal{R})}{\sqrt{\mathcal{R}}}\,. \tag{4.24}$$

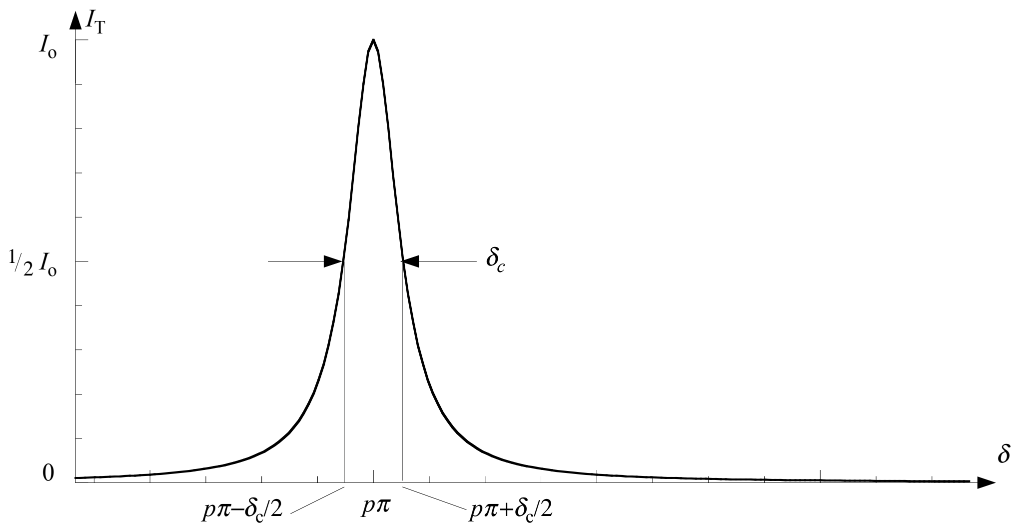


Figure 4.5: Peak width δ_c is defined to be the full width at half maximum.

From our definition of finesse (4.20) then we find

$$\mathcal{F} = \frac{\pi}{\delta_c} = \frac{\pi\sqrt{\mathcal{R}}}{1-\mathcal{R}} \equiv= \frac{\pi\sqrt{F}}{2}\,. \tag{4.25}$$

As an example, consider a Fabry-Perot consisting of two mirrors with $\mathcal{R} = 0.99$ separated by an air gap ($n = 1$) of 2.500000 cm. The source of light is a hypothetical laser emitting light of wavelength 500.0000 nm, incident normally ($\theta = 0$) on the mirrors. Then from the expression for δ (Eq.(4.17)) we find

$$\delta = 10,000\,\pi \tag{4.26}$$

The transmission should be maximum under this condition.

What happens if we gradually increase the air gap? To answer this, suppose we begin at a transmission peak at $\theta = 0$. Then the wavelength λ of maximum transmission and Fabry-Perot separation d satisfy the condition

$$\delta \equiv \frac{2\pi d}{\lambda} = p\pi \quad \Rightarrow \quad \lambda = \frac{2d}{p} \tag{4.27}$$

where p is an integer. As we vary the Fabry-Perot separation, the wavelength of maximum transmission will change. Thus if the Fabry-Perot separation changes by an amount Δd, the wavelength transmitted will be (fixed p)

$$\lambda + \delta\lambda = \frac{2(d + \delta d)}{p} \,. \tag{4.28}$$

If we continue to increase the separation, then at some point, when the separation has increased by an amount $\mathit{\Delta} d$, the original wavelength λ will be transmitted again but this time satisfying the relation

$$\lambda = \frac{2(d + \Delta d)}{p + 1} \,. \tag{4.29}$$

Note that this will be in addition to another wavelength $\lambda + \Delta\lambda$ satisfying

$$\lambda' \equiv \lambda + \Delta\lambda = \frac{2(d + \Delta d)}{p} \,. \tag{4.30}$$

Equations (4.29) and (4.30) imply that when the separation can simultaneously fit $p + 1$ half wavelengths of λ and p half wavelengths of $\lambda + \Delta\lambda$ both wavelengths will be transmitted. In this case we would not be able to tell if the Fabry-Perot is transmitting a new wavelength or the original one. So how much can we scan before this confusion or overlap of orders arises? The answer from Eqs. (4.30) is

$$\Delta d = \frac{\lambda}{2} \,. \tag{4.31}$$

Using this result in Eq. (4.30) and substituting value of p from Eq. (4.25) [or (4.29)][4] we obtain the range of wavelengths that can be freely scanned without overlapping of orders is

$$\Delta\lambda = \frac{\lambda^2}{2d} \tag{4.32}$$

Scanning beyond this point will simply repeat the pattern already encountered and is of little interest to us. Wavelength scanning range $\Delta\lambda$ free of any repetition is called the free spectral range (FSR) of the Fabry-Perot. This FSR in terms of scanned wavelength corresponds to the free spectral range in terms of phase, $\delta_{FSR} = \pi$.
Exercise: Compute FSR in terms of frequency $\nu = c/\lambda$.

In our numerical example, the free spectral range turns out to be 0.005 nm, which tells us that we are capable of looking very closely into the spectrum of the laser in

[4]It does not matter which equation is used as long as p is large.

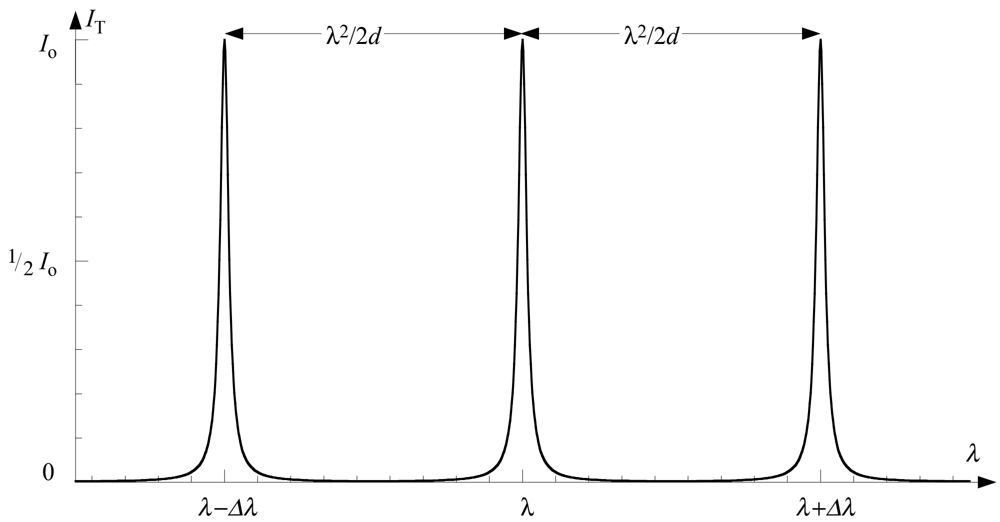


Figure 4.6: Transmitted intensity as a function of wavelength λ.

the range 500.0000 ± 0.0050 nm. This is particularly helpful in the event that the output of the laser actually consists of several closely spaced spectral lines.

According to Eq. (4.25), with 97% reflectivity mirrors, a finesse of 100 can obtained. This means that we can ultimately resolve two spectral lines separated by one-hundredth of the Free Spectral Range, or by 0.00005 nm, indeed a small separation. By contrast, the resolution of the best spectrometers available today can only reach .01 nm.

4.3 Applications

Fabry-Perots are particularly useful in the field of spectroscopy whenever high resolution is needed. Other applications include astronomy and intracavity laser-line narrowing. A traditional grating spectrometer is useful when the needed resolution is of the order of 0.01 nm while the Fabry-Perot finds application in the 0.1 nm to 10^{-6} nm range. To go below 10^{-6} nm, a third technique called "heterodyne detection" is used. This involves beating an unknown frequency with a reference light wave and measuring the beat frequency.

Part 1

The outcome of this experiment will be a transmission curve $I_T(\delta)$ displayed on an oscilloscope. To understand the experimental procedure described in Part 2, let us first perform some calculations.

> The reflectivity of the mirrors is given by the manufacturer to be 97.5%. Assuming that your set up is perfect, plot the curve I_T/I_0 [See Eq. (4.15)] for this particular reflectivity and calculate the finesse of the Fabry-Perot.

You will notice that the plotted peaks are quite sharp, but in the experiment they will be somewhat broader. This is due to imperfections in the experimental set up. We shall consider a few of these.

In the discussion of the theory, we assumed that the detection system was capable of looking at a single point at the output of the Fabry-Perot. In practice, the detector looks at a small area of the output plane and this can profoundly affect our results. To see this, let us assume that our experimental set up consists of a laser, a Fabry-Perot, a pinhole, and a detector as shown in Figure (4.7)

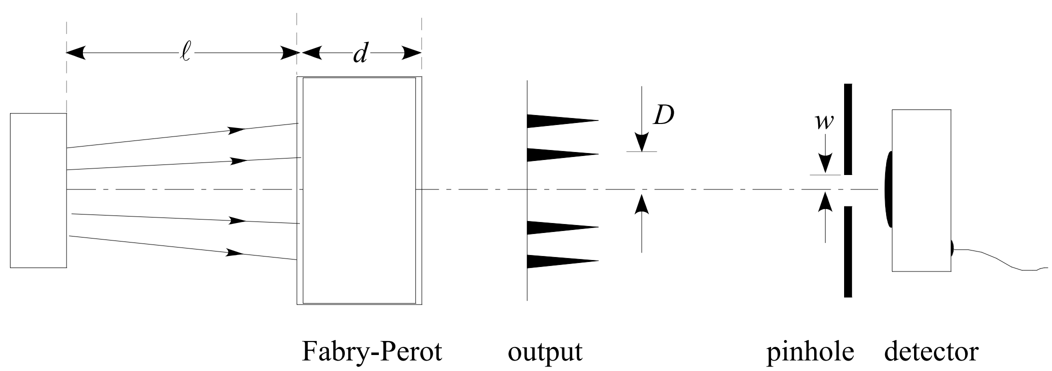


Figure 4.7: Detector aperture must not look at more than one fringe (transmission maximum) at a time. This can be accomplished by placing a pin hole of suitable size in front of the detector.

We shall assume that the laser behaves as a point source. The output of the Fabry-Perot then consists of concentric rings corresponding to the transmission peaks [Fig. (4.3)]. Figure (4.7) shows a side view of the output. The distance D is the separation of fringes[5]. The pinhole in front of the detector is used to limit the effective size of the detector so that the detector monitors only one fringe at a time. How small should we make the pinhole? Well, it depends on the resolution we want to achieve. It is important to understand this that even if the Fabry-Perot is able to produce distinct fringes in its output for two closely spaced wavelengths, we will not be able to resolve them if the detector cannot separately view these fringes one at a time. To specify the detector performance we introduce the concept of pinhole finesse$\mathcal{F}_p$, defined as the number of times the pinhole width w can fit into D,

$$\mathcal{F}_p = \frac{D}{w}\,. \tag{4.33}$$

The central fringe ($\theta = 0$) obeys the relation

$$2d = p\lambda \tag{4.34}$$

for some p. Similarly, the first fringe away from the optical axis satisfies[6]

$$2d\cos\theta_1 = (p-1)\lambda\,. \tag{4.35}$$

From Eqs. (4.34) and (4.35) we find that the first fringe outward from the center is located at an angle θ_1 given by

$$\cos\theta_1 = 1 - \frac{\lambda}{2d}\,. \tag{4.36}$$

Since the angle θ is usually small we can use $\cos\theta \approx 1 - \theta^2/2$ and find the angular position of the first fringe from the center

$$\theta_1 = \sqrt{\frac{\lambda}{d}}\,. \tag{4.37}$$

Now assuming that $d \ll \ell$, the fringe sepration D is given by

$$D \approx \ell \sin\theta_1 \approx \ell\theta_1 = \ell\sqrt{\frac{\lambda}{d}}\,. \tag{4.38}$$

Finally, from our definition (4.33), the pinhole finesse is expressed as

$$\mathcal{F}_p = \frac{\ell}{w}\sqrt{\frac{\lambda}{d}}\,. \tag{4.39}$$

[5] The spacing between the fringes in a Fabry-Perot decreases as we move away from the optical axis.

[6] The order p decreases as we move away from the optical axis.

- As a general rule, the pinhole finesse should be at least three times greater than the Fabry-Perot finesse. Using this rule, and assuming that $\ell = 1$ m, $d = 6$ cm and $\lambda = 632.8$ nm, calculate the largest diameter pinhole allowed for our system.

We should also point out in this discussion that the lack of flatness of the mirrors also affects the overall finesse of the system. (For example, if the plate flatness is $\lambda/200$, then the Finesse is limited to 100.) This is simply because the distance between the mirrors cannot be well defined when surface "errors" are present.

Finally, we mention that the HeliumNeon laser that we will use for our experiment emits light of wavelength 632.8 nm. A closer look with a FabryPerot will reveal that the spectrum actually consists of three equidistant lines, as shown in Figure (4.6). They are referred to as the (longitudinal) modes of the laser. Just as the transmission peaks of a FabryPerot, these modes are separated in wavelength by $\lambda^2/2L$, where L is the separation of mirrors inside the laser.

Part 2

The set up used in this experiment is similar to that described in Part 1. If the diameter of the pinhole available is significantly larger than the one you calculated in Part 1, you can compensate for this by increasing the distance between the detector and FabryPerot [See Figure (4.8)]

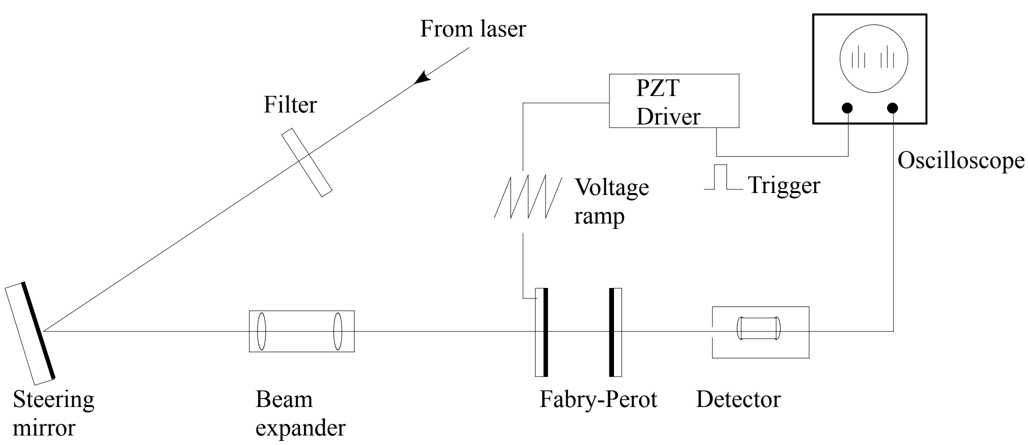


Figure 4.8: An outline of the experimental setup.

One of the FabryPerot mirrors is held in place by a piezoelectric transducer (PZT). By applying a voltage across the PZT we can change the length of the Fabry-Perot. The PZT driver supplies a time varying repetitive voltage in the form of a sawtooth or a triangular wave. The F-P length, therefore, increases linearly until it reaches the end of the ramp and then decreases linearly to its original length. This scanning sequence is repeated. A trigger pulse sent to the oscilloscope synchronizes the horizontal scan of the oscilloscope with the FabryPerot scan. A portion of the transmitted light passes through the pinhole and falls on the detector which generates a voltage proportional to the incident intensity. The detector output is displayed on the oscilloscope. The oscilloscope display is then a plot of intensity as a function of mirror separation.

You can calibrate the oscilloscope horizontal sweep in terms of frequency as follows. Increase the scanning voltage slowly until you see the transmitted mode pattern repeated in the output. The horizontal distance between two points on the oscilloscope trace that you see repeated equals the free-spectral range of the Fabry-Perot.

Before assembling the apparatus, make a preliminary alignment of the FabryPerot mirrors. Set the mirror separation to be about 3 cm. Arrange the laser and steering mirror in such a way that the beam hits the input mirror of the FabryPerot and travels back to the laser and just misses it. This procedure insures that the beam is perpendicular to the first mirror and no reflection goes back to the laser. Next, use the micrometric adjustments on the output mirror to align it until all multiple reflections at the output are superimposed.

Assemble the apparatus as shown in Figure (4.8). Ask your lab instructor for help if you do not know how to operate the equipment.

- Use the oscilloscope to measure the finesse of your Fabry-Perot. If the value obtained is low, your system is not correctly aligned. Realign and report the highest value obtained for the finesse. How does it compare to the theoretical value? How does the finesse change when the input beam is not properly aligned? Explain.

- Calculate the free spectral range of the Fabry-Perot. Use this information to measure the mode separation of the laser. From this compute the separation of the mirrors inside the laser? Is this a reasonable value for your laser? Calculate the ratio of the laser mode separation to the free spectral range of the Fabry-Perot.

- Change the FabryPerot mirror separation to 6 cm. Measure the finesse again. Do you expect it to be different from your earlier measurement? What is the expected ratio of the laser mode separation to the free spectral range of the Fabry-Perot? How does this compare with the measured value?

INSTITUTE OF PHYSICS PUBLISHING EUROPEAN JOURNAL OF PHYSICS

Eur. J. Phys. **27** (2006) 1111–1119 doi:10.1088/0143-0807/27/5/010

Understanding the concept of resolving power in the Fabry–Perot interferometer using a digital simulation

I Juvells[1], A Carnicer[1], J Ferré-Borrull[2], E Martín-Badosa[1] and M Montes-Usategui[1]

[1] Departament de Física Aplicada i Òptica, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain

[2] Departament d'Enginyeria Electrònica, Universitat Rovira i Virgili, Elèctrica i Automàtica. Av. Països Catalans 26, Campus Sescelades 43007 Tarragona, Spain

E-mail: ignasi.juvells@ub.edu

Received 15 March 2006, in final form 4 June 2006
Published 17 July 2006
Online at stacks.iop.org/EJP/27/1111

Abstract

The resolution concept in connection with the Fabry–Perot interferometer is difficult to understand for undergraduate students enrolled in physical optics courses. The resolution criterion proposed in textbooks for distinguishing equal intensity maxima and the deduction of the resolving power equation is formal and non-intuitive. In this paper, we study the practical meaning of the resolution criterion and resolution power using a computer simulation of a Fabry–Perot interferometer. The light source in the program has two monochromatic components, the wavelength difference being tunable by the user. The student can also adjust other physical parameters so as to obtain different simulation results. By analysing the images and graphics of the simulation, the resolving power concept becomes intuitive and understandable.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The multiple beam interferometer was proposed by Charles Fabry (1867–1945) and Alfred Perot (1863–1925) in 1899. The multiple beam interferences generated between two glass plates, internally covered by a highly reflective film and illuminated by an extended quasi-monochromatic source, produce an intensity distribution consisting of concentric circles. It can be demonstrated, using the mathematical law that describes the intensity profile of the interference pattern, that a given wavelength generates a specific set of maxima. Consequently, circles resulting from different wavelengths can be discriminated. This capability makes the Fabry–Perot interferometer a valuable tool in high-resolution spectroscopy.

The spectral resolving power $|\lambda/\Delta\lambda|$ quantifies the capability of an interferometer to resolve two close wavelengths. It is defined as the ratio of the wavelength of the source λ to the minimum difference of wavelengths $\Delta\lambda$ that generates two circle series that can be discriminated. To calculate the resolving power formula, it is necessary to know the law that describes the interference pattern and set a separation criterion to decide when the circles are viewed separately [1, 2].

Our experience in teaching optics to second year university physics students shows that the explanation of the resolution concept is barely understood. Very often, students merely retain the formula and are unable to grasp the underlying concepts. Because of this reason, we propose to use a Java applet to show how the Fabry–Perot interferometer works. These simulations make possible to modify the parameters involved in the physical problem, so the effect of the changes in the variables is visualized automatically.

As the mathematical description of the interferometer is relatively simple, the implementation of a computer simulation of this device is a straightforward task. Consequently, it is possible to find in the Internet multiple computer implementations of the interferometer (see, for instance, [3, 4]). Nevertheless, these Fabry–Perot applets do not allow you to analyse problems related with resolving power because they are not able to deal with more than one wavelength. For this reason, a simple application to emphasize this topic [5] has been developed.

In section 2, we explain how to calculate the spectral resolving power in a Fabry–Perot spectrometer and in section 3, we describe a visual methodology to determine the resolving power from the data generated by the applet. We compare visual results with those obtained directly from the formula presented in section 2.

2. Fabry–Perot resolution

2.1. How a Fabry–Perot interferometer works

The Fabry–Perot interferometer consists of two glass plates with parallel plane surfaces, separated at a distance d. The media between the glass plates is air ($n = 1$). If a monochromatic wave impinges upon the plate at angle ϵ, multiple reflections are generated. If the inner surfaces are covered by a highly reflective film, the reflections in the glass plates are negligible. Therefore, only the interferences produced by the multiple beams in the air plate are observable. The intensity of the interference patterns is described by the following expression:

$$I = \frac{a^2}{1 + \frac{4r^2}{(1-r^2)^2}\sin^2\left(\frac{\delta}{2}\right)}, \tag{1}$$

where a is the amplitude of the incident wave, r is the reflection coefficient of the coating film and δ is the phase difference between two consecutive waves. Let Δ be the optical length difference. In this case, the optical length is $\Delta = 2d\cos\epsilon$. As usual, the optical length and phase difference are related by $\delta = \frac{2\pi}{\lambda}\Delta$.

The intensity depends on the thickness d, the reflection coefficient r, the wavelength λ, the intensity a^2 of the incident plane wave and the incidence angle ϵ. If the light comes from an extended source from all possible directions ϵ, and taking into account that the geometry of the light distribution only depends on ϵ, the intensity pattern should exhibit rotational symmetry.

The intensity maximum ($I_{max} = a^2$) is obtained if the following condition is verified:

$$\sin^2\frac{\delta}{2} = 0 \qquad \text{thus} \qquad \delta = 2m\pi \qquad \text{and} \qquad \Delta = 2d\cos\epsilon = m\lambda, \tag{2}$$

where $m = 0, \pm 1, \pm 2, \ldots$. The intensity profile therefore exhibits a sequence of maxima and minima and, consequently, the interference pattern is characterized by a set of light circles. The index m in the previous equation labels each circle. For instance, the maximum m value is found when ϵ tends to zero. In particular, if a maximum is found when the incidence angle is zero ($\epsilon = 0$), then $m = 2d/\lambda$.

2.2. The Rayleigh criterion

As we have explained before, the intensity pattern is a function of the wavelength. Each wavelength coming from a light source generates a set of light circles. The resolving power is a measure of the ability to discriminate between sets of circles generated by different wavelengths. Moreover, it is also necessary to define mathematically a separation criterion between two very close maxima. This criterion represents one's visual ability to distinguish two concentric circles.

Different resolution criteria are described in physical optics textbooks. Several authors such as Born and Wolf [1] and Hecht [2] use Rayleigh's criterion [6]. It was first introduced by Lord Rayleigh in 1879 to determine whether two diffraction spots can be distinguished or not. The criterion states that two intensity maxima are separated if the maximum value of the first spot is superimposed on the first minimum of the second spot.

Rayleigh's resolution limit seems to be rather arbitrary and is based on resolving capabilities of the human eye. This limit was set to guarantee a clear distinction between two close spots using the eye. When visual inspection is replaced by detectors, other less restrictive criteria can be used. For instance, Sparrow's criterion states that two diffraction spots are just distinguished when the minimum between the two intensity maxima of the composite intensity is undetectable [7].

Rayleigh's criterion cannot be directly applied to the Fabry–Perot intensity profile because of the slow decrease of these values. Consequently, the minimum is located far from the maximum. To avoid this problem, an alternative definition is proposed: let $I_\lambda(x)$ and $I_{\lambda+\Delta\lambda}(x)$ equal the intensity profiles along a diameter, generated by wavelengths λ and $\lambda + \Delta\lambda$ respectively [1]. Two maxima are resolved if the minimum value of $I_\lambda(x) + I_{\lambda+\Delta\lambda}(x)$ verifies the following condition:

$$\min\{I_\lambda(x) + I_{\lambda+\Delta\lambda}(x)\} = 0.81\max\{I_\lambda(x)\}. \tag{3}$$

When the Rayleigh criterion is applied to diffraction in a slit, we obtain the condition shown in equation (3).

2.3. The Taylor criterion

Other authors of optical textbooks (Klein–Furtak [8], Pérez [9], Françón [10], Fowles [11]) use the full-width half-maximum criterion (FWHM). This states that the separation of the maxima is equal to the half-maximum width. This criterion is also known as the Taylor criterion (TC). In practice, both criteria give very similar results. For instance, Hecht [2] introduces the Rayleigh criterion but, after an approximation, uses the TC. We also use TC in this paper because it is easier to handle the mathematics involved.

Figure 1 shows an example of the use of this criterion for different values of the reflective coefficient r. In figure 1(a) ($r = 0.65$), the intensity maxima profiles are not separated

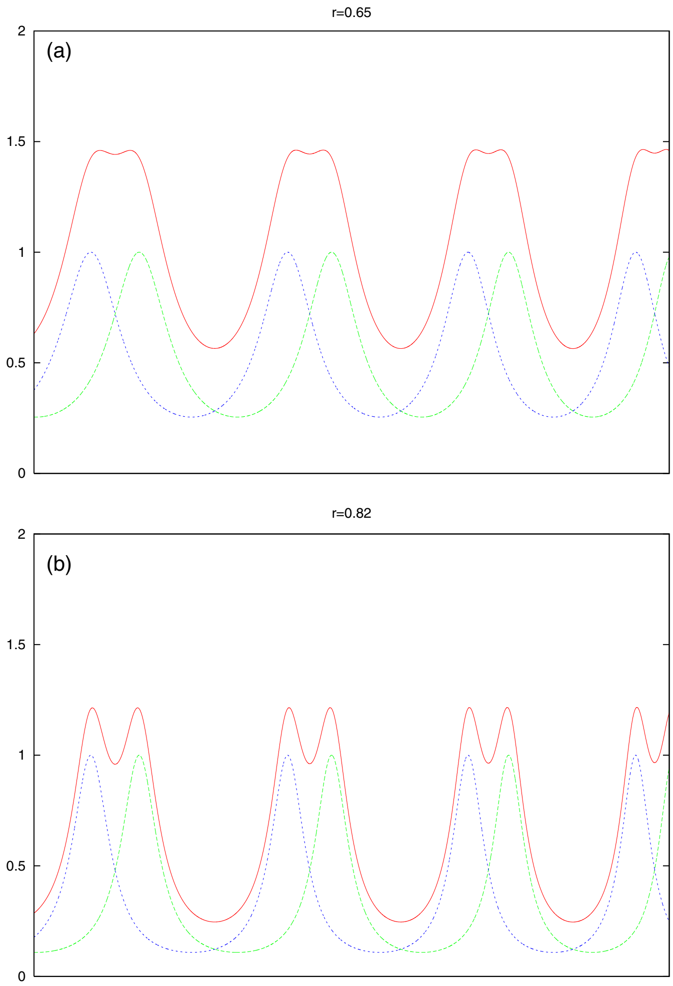


Figure 1. Taylor criterion as a function of the reflective coefficient. The figure shows the intensity profile (equation (1)) for two close wavelenghts (blue and green curves). The composite intensity is shown in red. (a) $r = 0.65$, the maxima cannot be discriminated and (b) $r = 0.82$, the maxima fulfils the criterion.

enough to be visually differentiated whereas in figure 1(b) ($r = 0.82$), the maxima fulfil the requirements of the TC; thus, the interference fringes are considered separated.

Using equations (1) and (2) (intensity and position of maxima), and applying the TC, a formula that gives the minimum wavelength difference ($\Delta\lambda$) for the fringes produced by two monochromatic sources to be separated can be deduced as

$$\mathrm{SRP} = \left|\frac{\lambda}{\Delta\lambda}\right| = \frac{\pi\,\mathrm{mr}}{1-\mathrm{r}^2}. \tag{4}$$

This equation is called the spectral resolving power (SRP) of the interferometer. The derivation of this formula is shown in the appendix.

For a specific wavelength, the SRP depends on the maximum order m and the reflective coefficient r. The SRP gives higher values for fringes near the centre of the interferometric pattern and for values of r tending to 1, which means that closer wavelengths can be discriminated.

3. Procedure

We propose to analyse the meaning of the SRP concept, using the applet. First, we study the dependence of SRP with coefficient r, for a fixed m. Initially, we set the source wavelength ($\lambda_1 = 500.1$ nm) and the thickness ($d = 7.5$ mm). From equation (2) we get $m = 29993$. We use these values just as an example.

Taking into account that the best resolving conditions can be found for central fringes (maximum m values), we adjust the screen size to display only the inner circle of the interferometric pattern (focal length 1 = 500 mm, focal length 2 = 500 mm, source size = 10 mm, screen size = 5 mm.) If, for instance, we want to discriminate a second wavelength $\lambda_2 = 500.103$ nm, then $\Delta\lambda = 0.003$ nm. These are arbitrary values and have been selected to show the ability of the Fabry–Perot interferometer to distinguish two really close wavelengths.

Now, we increase the value of r starting from $r = 0.6$. As shown in figure 2(a), the circles generated by λ_1 and λ_2 cannot be separated visually. In practice, two fringes are distinguished according to Rayleigh's criterion if they can be clearly discriminated by the observer's eye. When the values of r reach the interval $0.75 \leqslant r \leqslant 0.78$, the fringes appear clearly separated (see figure 2(b)). Figure 2(c) displays the intensity profile. The plot shows that the separation of the maxima is equal to the half-maximum width when $r = 0.76$.

Now, we can verify this result with the use of the equation predicted by theory (equation (4)). Setting the variables $\lambda = 500.1$ nm, $\Delta = 0.003$ nm and $m = 29993$, and isolating r, we obtain the following equation:

$$r^2 + \left(\pi m \left|\frac{\Delta\lambda}{\lambda}\right|\right) r - 1 = 0. \tag{5}$$

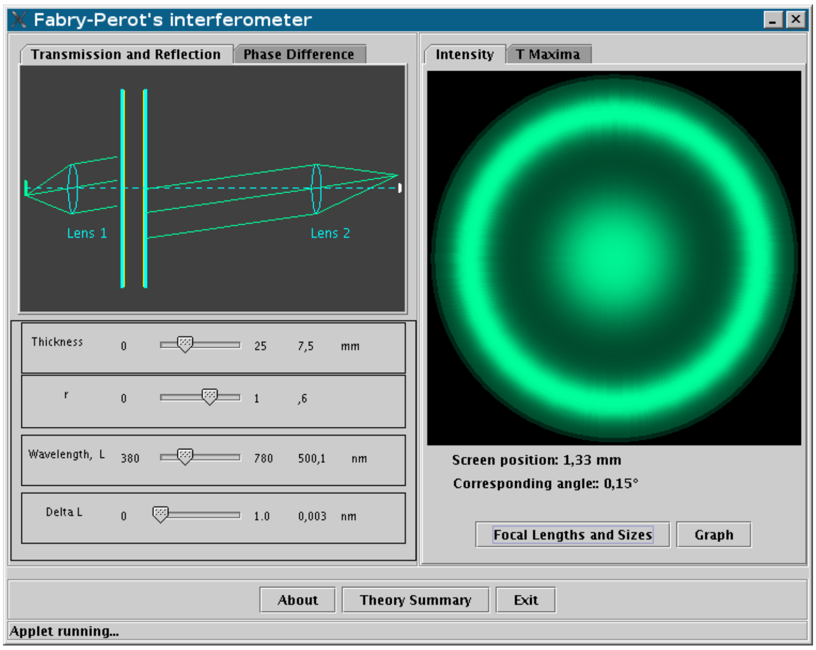
Solving equation (5), we get a value of $r = 0.757$, very similar to that obtained visually using the applet.

We suggest a second exercise to show how the SRP varies with the thickness d. Here, the variables are set to the following values: focal length 1 = 100 mm, focal length 2 = 1000 mm, source size = 5 mm, screen size = 35 mm, $r = 0.7$. The wavelengths selected are $\lambda_1 = 429.248$ nm and $\lambda_2 = 429.286$ nm, which correspond to two equal intensity lines of the emission spectrum of sodium (Na II) (see, for instance, [12]). We start from $d = 0.3$ mm, and the thickness increases until 0.7. Figure 3 shows the interferogram at $d = 0.4$, 0.45, 0.5, 0.55, 0.6 and 0.65 mm. The inner circles appear to be distinguished at $d = 0.55$ mm or $d = 0.6$ mm. Again, numerical verification agrees with the visual inspection of simulated interferograms because the limit value of d that verifies the Rayleigh criterion is $d = 0.56$ mm.

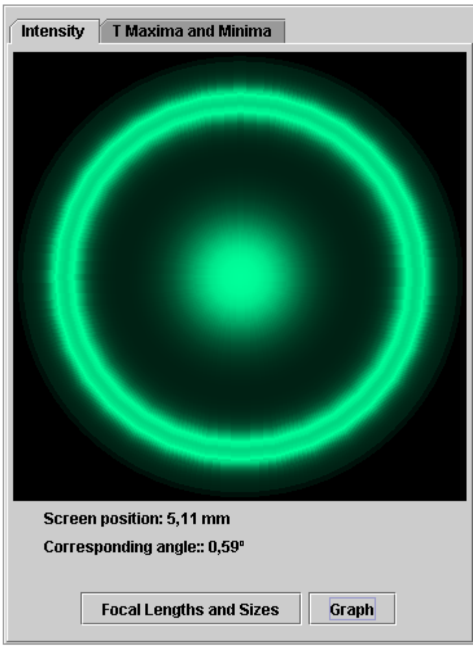
4. Concluding remarks

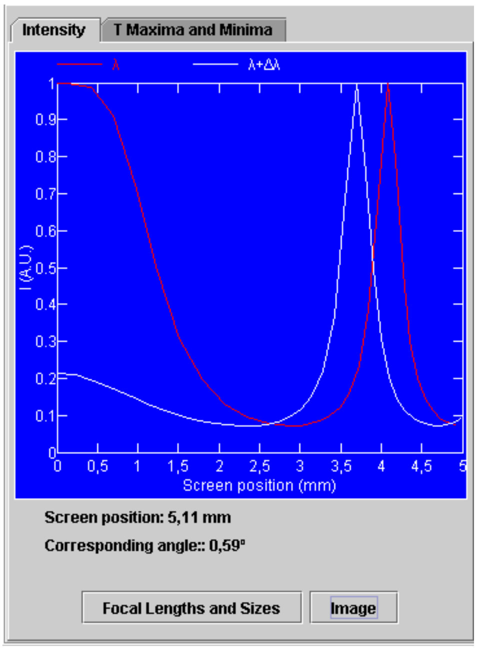
The most common physical optics textbooks analyse the spectral resolution power concept in interferometry from the point of view of Rayleigh's resolution criterion. The use of this criterion is appropriate when qualitative observation of some physical phenomena is carried out using the human visual system. Nowadays, undergraduate experiments often include data acquisition systems, so the use of Rayleigh criterion seems to be rather arbitrary and the students find it confusing to apply.

Computer simulations of interferometric devices allow the user to handle the variables involved in the physical problem. A digital implementation of the Fabry–Perot interferometer can be a helpful tool to analyse the connection between spatial resolution power, resolution criteria and their mathematical formalism. In our opinion, the use of such programs, combined with a suitable methodology, helps students to understand these concepts better.



(a)





(b) (c)

Figure 2. Results: (a) program window showing two unresolved fringes ($r = 0.6$) and (b) limit of visual resolution ($r = 0.76$). The two circles are clearly distinguishable. (c) Intensity profile of (b).

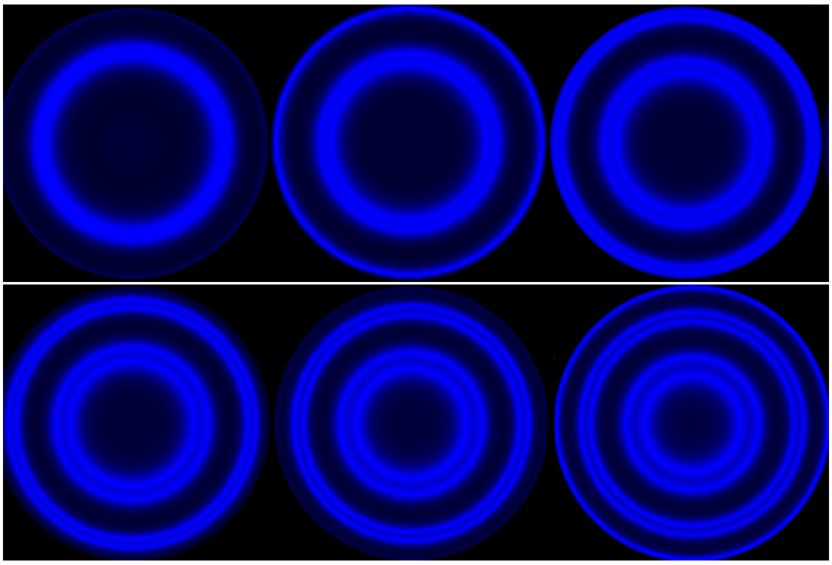


Figure 3. Interferograms at $d = 0.4, 0.45, 0.5, 0.55, 0.6$. Note that for a fixed m, the radius of the fringes decreases when the thickness increases.

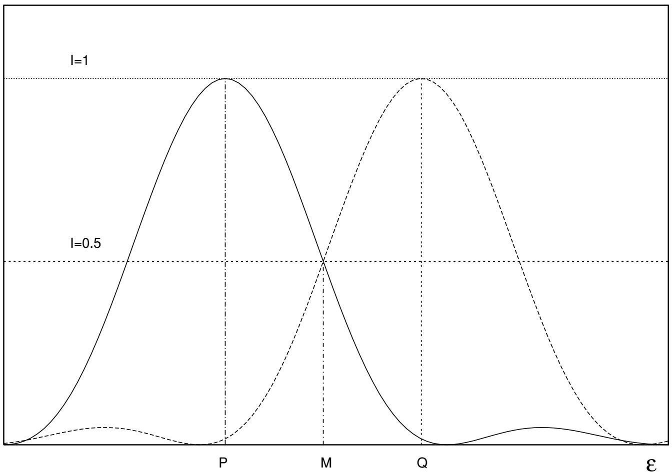


Figure 4. Full width at half-maxima criterion.

In this paper, we have suggested two exercises to study the dependence of the spatial resolution power with the reflectance r and the thickness d in a Fabry–Perot interferometer. We suggest these exercises as classroom activities or as homework.

The analysis of the results obtained shows the practical meaning of the Rayleigh criterion. Far from being an arbitrary condition, Rayleigh's formula truly captures, in a mathematical form, the resolution limit of the human visual system. The applet clearly shows that the two fringes can be distinguished by the eye when the Rayleigh criterion is fulfilled, which helps students to appreciate the resolving power of the Fabry–Perot interferometer. Moreover, the students enjoy this approach because they understand better and more easily the underlying concepts.

Acknowledgments

This paper has been partially supported by the Departament d'Universitats, Recerca i Societat de la Informacio (DURSI), Government of Catalonia (project # 2005MQD 00127) and the Institut de Ciències de l'Educació de la Universitat de Barcelona (project # REDICE 0401-14).

Appendix: Deduction of the spectral resolving power formula

The intensity of the interference patterns is described by

$$I = \frac{a^2}{1 + \frac{4r^2}{(1-r^2)^2}\sin^2\left(\frac{\delta}{2}\right)}, \tag{A.1}$$

where

$$\delta = \frac{2\pi}{\lambda}\Delta = \frac{2\pi}{\lambda}2d\cos\epsilon. \tag{A.2}$$

The maximum of intensity ($I_{\max} = a^2$) is obtained if the following condition is verified:

$$\sin^2\frac{\delta}{2} = 0, \quad \delta = 2m\pi \quad \text{then} \quad \Delta = 2d\cos\epsilon = m\lambda \quad \text{for} \quad m = 0, \pm 1, \pm 2, \ldots. \tag{A.3}$$

Now, we want to determine the minimum $\Delta\lambda$ that generates two fringes separated to the extent that they fulfil the Taylor criterion (figure 4).

Let $\Delta\delta$ be the change in a phase associated with the change in the wavelength $\Delta\lambda$: $\delta_Q = \delta_P + \Delta\delta$. As P corresponds to a maximum, $\delta_P = 2m\pi$. Taking into account that the phase variation is linear for small changes in the phase, then

$$\delta_M = \delta_P + \Delta\delta/2 = 2m\pi + \Delta\delta/2. \tag{A.4}$$

The resolving criterion states that $I_M = I_P/2 = a^2/2$, and introducing the condition in equation (A.1),

$$\sin^2\frac{\delta_M}{2} = \frac{(1-r^2)^2}{4r^2} \qquad \text{then} \qquad \sin\frac{\Delta\delta}{4} = \frac{1-r^2}{2r} \tag{A.5}$$

and approximating $\sin(\Delta\delta/4)$ with $\Delta\delta/4$, then

$$\frac{\Delta\delta}{2} = \frac{1-r^2}{r}. \tag{A.6}$$

Differentiating equation (A.2),

$$\Delta\delta = -\frac{2\pi}{\lambda^2}2d\cos\epsilon\,\Delta\lambda, \tag{A.7}$$

and using the relation $2d\cos\epsilon = m\lambda$ results in $\Delta\delta = -\frac{2\pi}{\lambda^2}m\Delta\lambda$. Replacing $\Delta\delta$ as a function of the reflective coefficient (equation (A.6)), we get the formula of the spectral resolving power:

$$\left|\frac{\lambda}{\Delta\lambda}\right| = \frac{\pi m r}{1-r^2}. \tag{A.8}$$

References

[1] Born M and Wolf E 1999 *Principles of Optics* (Cambridge: Cambridge University Press)
[2] Hecht E 1998 *Optics* (Reading, MA: Addison-Wesley)
[3] Mzoughi T, Davis-Herring S, Foley J T, Morris M J and Gilbert P J 2006 WebTOP: A 3d interactive system for teaching and learning optics *Comput. Educ.* **46** at press Online at http://dx.doi.org/10.1016/j.comedu.2005.06.008/
[4] http://www.physics.uq.edu.au/people/mcintyre/applets/fabry/fabry.html
[5] http://www.ub.edu/javaoptics/docs_applets/Doc_FabryEn.html
[6] Strutt J W (Lord Rayleigh) 1879 Investigations in optics, with special reference to the spectroscope *Phil. Mag.* **8** 261–74, 403–411, 477–486
[7] Dekker A J and van den Bos A 1997 Resolution: a survey *J. Opt. Soc. Am.* A **14** 547–57
[8] Klein M V and Furtak T E 1986 *Optics* (New York: Wiley)
[9] Pérez J P 1994 *Optique Geometrique et Ondulatoire* (Paris: Masson)
[10] Françon M 1966 *Optical Interferometry* (New York: Academic)
[11] Fowles G R 1989 *Introduction to Modern Optics* (New York: Dover)
[12] Nist atomic spectra database. http://physics.nist.gov/cgi-bin/AtData/lines_form