# The many facets of the Fabry-Perot

**Luis L Sánchez-Soto[1,2], Juan J Monzón[1] and Gerd Leuchs[2,3]**

[1] Departamento de Óptica, Facultad de Física, Universidad Complutense, E-28040 Madrid, Spain
[2] Max-Planck-Institut für die Physik des Lichts, Günther-Scharowsky-Straße 1, Bau 24, D-91058 Erlangen, Germany
[3] Institut für Optik, Information und Photonik, Universität Erlangen-Nürnberg, Staudtstraße 7/B2, 91058 Erlangen, Germany

**Abstract.** We address the response, both in amplitude and intensity, of a Fabry-Perot from a variety of viewpoints. These complementary pictures conspire to achieve a comprehensive and consistent theory of the operation of this system.

*Key words:* Fabry-Perot, Interference.

## 1. Introduction

The names of Charles Fabry and Alfred Perot‡ are inextricably linked to one of the most deceptively simple setups ever devised in optics: just two parallel highly reflecting mirrors.

The instrument was the fortunate convergence of two independent developments: a long history of producing mirrors and a deeper understanding of multiple beam interference [7]. Fabry and Perot took full advantage of the potential of this setup and accomplished major scientific discoveries: some of their seminal work includes the determination of the temperature of the Orion nebula and the measurement of the gravitational redshift of light. Other discoveries are less known: the determination of the altitude and thicknesses of the atmospheric ozone layers; the calibration of the flux of different stars, the improvement of stellar spectrophotometry; the development of electrometers to measure weak potentials; the elaboration of an atlas of emission lines; the laboratory verification of the Doppler-Fizeau principle and much more [10]. A detailed biographic sketch of these two influential scientists can be found in the excellent article by Mulligan [11].

Apart from its luminosity, the distinctive feature of the Fabry-Perot is its narrow resonances. This is the basis for its extensive use in high-resolution spectroscopy, interferometry, and laser resonators. An exhaustive covering can be found in the two thorough monographs by Hernandez [12] and Vaughan [13]. This variety of contexts accounts for the diversity of terms used to name it: interferometer, spectrometer, filter, étalon and cavity, to cite only the most popular.

‡ Is it Perot or Pérot? The former accent-free spelling is indeed the official one, as confirmed by Perot's birth certificate and other official documents [1, 2]. This is also the form used by the most authoritative French authors, such as Kastler [3], Françon [4] Jacquinot [5], Chabbal [6] or Connes [7], among others. We therefore adopt here this usage. However, the accented version persists to this day in the scientific literature. Surprisingly enough, Perot referred to himself as Pérot in a few of his original works. In an awesome paper [8], Steel hypothesises that this misspelling had its origins in Parisian journals, such as Comptes Rendues, the savant editors of which might have over-zealously purported to know better than authors from the south of France. However, as Orr [9] aptly points out, it seems that the error became entrenched on the other side of the Atlantic, where Perot's name was persistently misspelt in American digests of his papers during the period 1900–1905.

Each of these tags capitalizes on specific ideas. The geometric treatment, in which one adds the multiple beams reflected at each of the different interfaces, is probably the more instructive and, accordingly, is reproduced in almost every textbook [14]. But the question can also be tackled from different standpoints. In the course of many years of teaching and research, we have brought together a number of nonconventional approaches to the Fabry-Perot. The delightful notes compiled by Jacobs [15] persuaded us that it is worthwhile to understand this remarkable device in as many different ways as possible. It is not that the classical textbooks are incomplete; rather, we seek to highlight some points that often go unnoticed and have been elevated to a position of relevance by current progress. Besides its inherent interest, we believe that this can help overcome the little cross talk between different specialists.

A final word of caution: the arrangement of a realistic Fabry-Perot imposes stringent practical requirements that the interested reader can find in the two aforementioned treatises. Our discussion, however, is limited to its basic aspects. None of the methods presented here is original by itself and we make no claim of completeness of the collection. Although, at first sight it seems difficult to find anything truly new and useful in this topic, we hope that the final picture emerging from all these complementary viewpoints gives fresh perspectives of the subtleties behind this amazing system.

## 2. Amplitude response of a Fabry-Perot

### *2.1. Multiple beam interference*

The ideal Fabry-Perot (FP) interferometer consists of two parallel mirrors (which, for simplicity, we assume to be identical) separated at a distance $d$. This is customarily modeled by a plane parallel plate of thickness $d$ and refractive index $n$ immersed in a medium of index $n'$. The plate is illuminated near normal incidence with a linearly polarized quasi-monochromatic plane wave, with the electric field lying either parallel or perpendicular to the plane of incidence. Any diffraction effect or polarization dependence are thus neglected henceforth.

We denote by $A_{\text{in}}$ the amplitude of the incident wave. At the first surface, this wave is divided into two plane waves, one reflected and the other transmitted into the plate. The latter wave is incident on the second surface and is divided into two plane waves, and the process of division of the wave remaining inside the plate continues as sketched in figure 1.

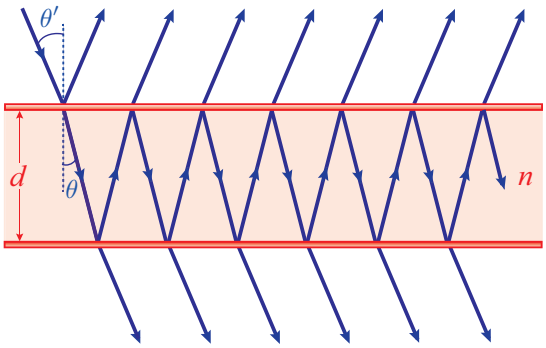


**Figure 1.** Schematic of the reflected and transmitted amplitudes in an FP, modeled as a plane parallel plate.

Let $r'$ and $t'$ be the Fresnel reflection and transmission coefficients, respectively, for a wave travelling from the surrounding medium into the plate and let $r$ and $t$ the corresponding coefficients for a wave travelling from the plate to the surrounding medium. The total reflected and transmitted waves can be jotted down as

$$\begin{aligned} A_{\text{ref}}(\varphi) &= A_{\text{in}}[r' + tt'r\mathrm{e}^{-\mathrm{i}2\varphi}(1 + r^2\mathrm{e}^{-\mathrm{i}2\varphi} + r^4\mathrm{e}^{-\mathrm{i}4\varphi} + \ldots)]\,, \\ A_{\text{trans}}(\varphi) &= A_{\text{in}}tt'\,\mathrm{e}^{-\mathrm{i}\varphi}(1 + r^2\mathrm{e}^{-\mathrm{i}2\varphi} + r^4\mathrm{e}^{-\mathrm{i}4\varphi} + \ldots)\,, \end{aligned} \tag{2.1}$$

where $\varphi = \frac{2\pi}{\lambda} nd\cos\theta$ is the plate phase thickness. Here, $\lambda$ is the wavelength in vacuum and $\theta$ the angle of refraction in the medium $n$, which is related to the angle of incidence according to Snell's law.

The complex reflection and transmission coefficients (i.e., the ratios $A_{\text{ref}}/A_{\text{in}}$ and $A_{\text{trans}}/A_{\text{in}}$, respectively) are obtained by adding all the waves in (2.1); the result is

$$R(\varphi) = r' + \frac{tt'r\exp(-\mathrm{i}2\varphi)}{1 - r^2\exp(-\mathrm{i}2\varphi)}\,, \qquad T(\varphi) = \frac{tt'\exp(-\mathrm{i}\varphi)}{1 - r^2\exp(-2\mathrm{i}\varphi)}\,. \tag{2.2}$$

The quantities $r, r', t, t'$ are given by Fresnel formulas. For the typical operation parameters of a a realistic FP, the dependence of these coefficients with the angle of incidence or the wavelength can be neglected and take them as constants. We shall make use of the Stokes relations [14]

$$r' = -r\,, \qquad tt' + r^2 = 1\,, \tag{2.3}$$

which constraints the reflectivity and transmissivity of both surfaces. Equations (2.2) can then be recast as

$$R(\varphi) = \frac{r[\exp(-\mathrm{i}2\varphi) - 1]}{1 - r^2\exp(-\mathrm{i}2\varphi)}\,, \qquad T(\varphi) = \frac{(1 - r^2)\exp(-\mathrm{i}\varphi)}{1 - r^2\exp(-2\mathrm{i}\varphi)}\,. \tag{2.4}$$

These responses can be seen as parametric curves in the complex plane. First of all, we observe that $R(\varphi)$ is a $\pi$-periodic function, while $T(\varphi)$ is $2\pi$-periodic. This translates into the fact that when $T(\varphi)$ completes a revolution, $R(\varphi)$ makes two turns.

Probably, the easiest way to draw them is by introducing polar coordinates

$$R = |R|\exp(\mathrm{i}\rho)\,, \qquad T = |T|\exp(\mathrm{i}\tau)\,. \tag{2.5}$$

For reasons that will become apparent soon, we call $\mathscr{R} = |R|^2$ and $\mathscr{T} = |T|^2$. With this parametrization, equations (2.4) read as

$$\mathscr{R} = 4a^2\cos^2\rho\,, \qquad \mathscr{T} = 1 - 4a^2\sin^2\tau\,, \tag{2.6}$$

where $a = r/(1 + r^2)$ is a real parameter verifying $0 \le a^2 \le 1/4$.

Both amplitudes lie inside the unit disk because one can immediately check that

$$\mathscr{R} + \mathscr{T} = 1\,, \tag{2.7}$$

which is just energy conservation. The reflected amplitude $R(\varphi)$ is readily identified as a circle of radius $a$ centered in the point $(a, 0)$ of the real axis. Actually, it can also be expressed as $R = a[1 + \exp(i2\rho)]$, which confirms that. The transmitted amplitude $T(\varphi)$ describes a hippopede, a curve full of stunning properties [16, 17]. For $0 < a^2 < 1/8$ it is an oval, and for $1/8 < a^2 < 1/4$ it is an indented oval, which tends to be an eight in the limit $a^2 = 1/4$. In figure 2 we represent $R(\varphi)$ and $T(\varphi)$ for several values of $r$, supporting these facts.

The reflected amplitude passes through the origin for any value of $r$: $R(\varphi)$ is zero for $\varphi = 0$ and $\pi$ and traces the circle clockwise, getting its maximum at $\varphi = \pm\pi/2$, where $\rho = 0$ and then, according equation (2.6), $\mathscr{R}_{\max} = 4a^2$.

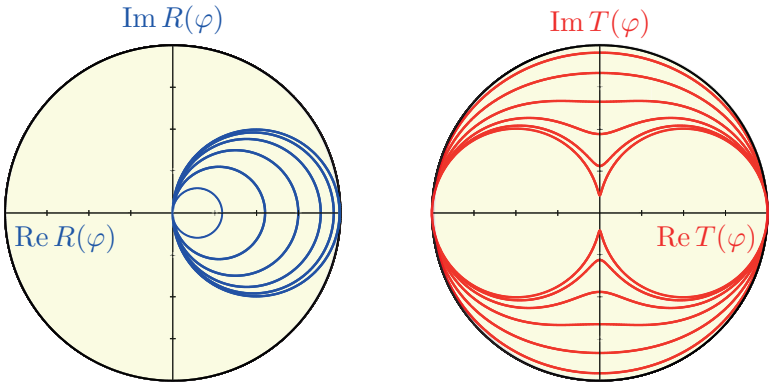


**Figure 2.** Reflected (left) and transmitted (right) amplitudes by a FP. The different curves correspond reflection coefficients ranging from $r = 0.15$ to $r = 0.90$ in steps of 0.15. For $R(\varphi)$, the curves increase in size with $r$, while the converse happens for $T(\varphi)$.

On the other hand, the transmitted amplitude also describes the hippopede clockwise. At $\varphi = 0$ and $\pi$, $\mathscr{T}(\varphi)$ reaches its maxima, which are in the real axis at $T = 1$ and $-1$, respectively. The minimum occurs at $\varphi = \pi/2$ and $3\pi/2$, where $\tau = -\pi/2$ and $-3\pi/2$, respectively. Therefore $\mathscr{T}_{\min} = 1 - 4a^2$, which corresponds to half the waist of the hipoppede at its indentation [18]. Note that $a^2 = 1/8$ happens for $r = \sqrt{3-\sqrt{8}} \simeq 0.4142$ and then $\mathscr{R}_{\max} = \mathscr{T}_{\min} = 1/2$.

We also observe that the quotient

$$\frac{R(\varphi)}{T(\varphi)} = \mathrm{i}\,\frac{2r}{1-r^2}\sin\varphi \tag{2.8}$$

is an imaginary number. Therefore for a transparent symmetric system, as the one we are dealing with, we have

$$\rho(\varphi) - \tau(\varphi) = \pm\frac{\pi}{2}\,, \tag{2.9}$$

so the reflected and transmitted amplitudes are always at quadrature.

To round up this discussion, we examine the local slopes $\dot{\rho}(\varphi)$ and $\dot{\tau}(\varphi)$, the dot denoting derivative respect to the parameter. They are just the rates at which the curves are traced out. Indeed, they entail a valuable physical interpretation. If we focus for simplicity at $\tau$, we can write

$$\frac{\dot{\tau}}{\mathscr{T}} = \frac{\tan\tau}{\tan\varphi} = -\frac{1+r^2}{1-r^2}\,, \tag{2.10}$$

the negative sign reflecting that the curve is oriented clockwise. This quotient is thus independent of $\varphi$: when the transmitted amplitude is large, so is the velocity and the opposite.

Moreover, $\dot{\tau}(\varphi)$ admits a crystal clear interpretation. Actually, a straight application of the chain rule gives

$$\mathrm{t}_g = \dot{\tau}(\varphi)\,\mathrm{t}_c\,, \tag{2.11}$$

where the group- and phase-delay times are, respectively, [19]

$$\mathrm{t}_g = \frac{\mathrm{d}\tau}{\mathrm{d}\omega}\,, \qquad \mathrm{t}_c = \frac{\mathrm{d}\varphi}{\mathrm{d}\omega}\,, \tag{2.12}$$

with $\omega$ the angular frequency. The time $\mathrm{t}_c$ is the single-pass time inside the plate (for a non-dispersive material, this is $nd\cos\theta/c$, with $c$ the velocity of light in vacuum), and $\mathrm{t}_g$ is a

measure of the time needed for a signal to propagate through the device. In a crude model, $t_g$ is the time a photon spends before it is ultimately transmitted (which is also called the dwelling time [20]). Hence, $\dot{\tau}(\varphi)$ may be viewed as an enhancement factor of the dwell due to the FP itself.

### *2.2. Variations on the same subject*

*2.2.1. Phasors.—* To facilitate the discussion, in what follows we shall restrict our attention to the transmitted field. One can alternatively interpret $A_{\text{trans}}$ in equation (2.1) as a sum of phasors (or rotating vectors), each one representing the field transmitted into the plate in a round trip. The length of every individual phasor is just the modulus of the corresponding term in the sum, whereas its orientation is the phase of such a term.

Figure 3 shows the phasors and the resulting transmitted amplitude for near-resonant, off-resonant, and intermediate conditions. Phasors in phase with the incident field are plotted along the horizontal axis, while those shifted by $\pi/2$ are plotted along the vertical axis.

Near resonance, all phasors acquire a phase shift close to an integer multiple of $2\pi$ after each round trip: they are nearly parallel and constructively interfere to yield a large transmitted amplitude. In off resonance, the components of the field acquire a wide range of phase shifts after circulating within the plate and tend to destructively interfere and produce a negligible overall field.

In the same vein, higher values of $r$ result in phasors that remain in the FP for a longer time. In this case, phasors can still be of a significant amplitude when, after many round trips, they have acquired enough phase to destructively interfere with the other phasors in the FP. Input fields whose frequencies are several linewidths from the resonance condition can be considered to be directly reflected, and the resulting field is small.

This approach highlights the need for exceptionally small tolerances for the mirrors flatness. The high transmission and narrow linewidth occur because all the phasors become accurately aligned when $\varphi = 0$ or $\varphi = \pi$. But if there are small errors in the phase because the reflections came from different parts of the plate, the alignment is less accurate and the linewidth increases.

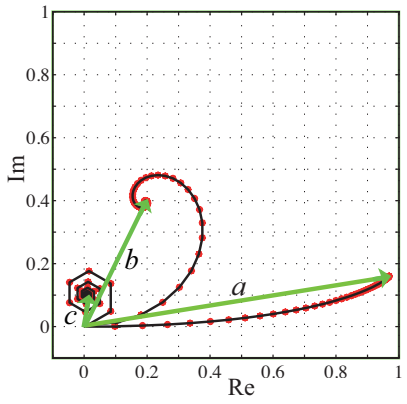


**Figure 3.** Representation of the first fifty terms in equation (2.1) as phasors, as well as the resulting transmitted amplitude for an FP interferometer. The arrows are indicated by red points and the total field is plotted in green. The horizontal axis represents signals in phase with the input field, and the vertical axis represents signals $\pi/2$ out of phase. The three examples correspond to $r = 0.95$ and phase lags $\varphi$ of (a) $\pi/360$, (b) $\pi/30$ and (c) $\pi/6$.

Phasor analysis is a useful mathematical tool for solving problems involving linear systems in which the excitation is a periodic time function [21]. Although the reader most likely has encountered this phasor notation in a variety of areas, its use in the context of FP is rare [22], even if it allows to treat situations for which the standard method, as developed in the previous subsection, fails: for example, the temporal response of a FP to changes in cavity length and frequency of the incident field [23].

*2.2.2. Equivalent diffraction grating.—* Let us think for a moment of the FP as illuminated by spherical wavefronts arising from an ideal point source $P_0$, as sketched in figure 4. The action of the FP thus appears as the interference of the wavefronts emanating from $P_0$ and the virtual sources $P_1$, $P_2$,..., separated by a distance $2d/n$ (we assume $n' = 1$), and amplitudes decreasing as in equation (2.1). This simplified picture holds only within the limits of the paraxial approximation; otherwise, aberrations appear, although this plays no major role for our discussion here.

This opens a new way of looking into multiple-beam interference: the beams being interfered can be though of as originating from equally spaced coherent point sources in a line, which bears a close resemblance to a diffraction grating. Actually, if we assume a "shaded" diffraction grating, with a pupil function of $\exp(-\eta z)$, with $z$ being the position along the axis perpendicular to the mirrors, one recovers exactly the amplitudes in (2.1) once we identify $\eta = n\ln(1/r^2)/2d$ [24], as we shall confirm soon.

It is worth observing that the separation between images, $2d/n$, is a bit of a surprise, since one expects $d$ to be multiplied and not divided by $n$. But if we use the diffraction grating equation $m\lambda = (2d/n)\cos\theta'$ (defining $\theta'$ as the angle measured from the plane of the grating), everything works out since $\cos\theta' \simeq 1 - \frac{1}{2}\theta'^2 \simeq \frac{1}{2}n^2\theta^2$.

Our physical intuition suggests that, since these secondary sources get fainter farther away from the real source, only an effective number of them contribute. This can be equivalently formulated as how many terms must be retained in the series (2.1) for it to be accurate enough. In figure 5 we have represented the amplitude resulting when one takes into account 5, 10 and 20 terms. The convergence to the ideal curve, shown in figure 2, is pretty fast, although it depends on the values of $r$. A more quantitative discussion of this question will be presented in section 3.3.2.

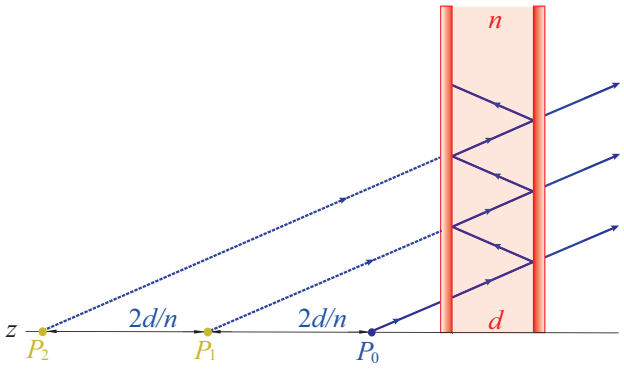


**Figure 4.** FP illuminated by an ideal point source $P_0$ and the series of virtual sources $P_1$, $P_2$,... that represent the action of the transmitted waves.

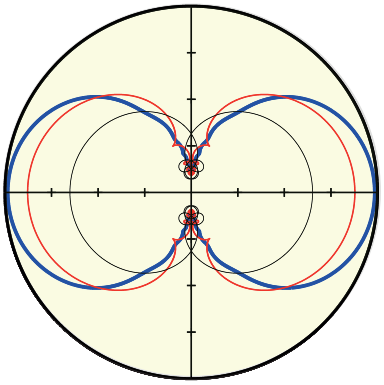

**Figure 5.** Transmitted amplitude for a transparent FP with plates coefficient $r = 0.90$. We have pondered a finite number of terms in (2.1): 5 for the black curve, 10 for the red and 20 for the blue. For this last case, the difference with the ideal hippopede is negligible.

Another practical aspect of this construction is to give intuition about what happens when the planes are not quite parallel. The points $P_0, P_1, P_2, \ldots$ are now equally spaced around a circle and the resulting diffraction grating then gives a high-order Bessel function [25], which has associated oscillations on one side of the peak, instead of the Airy function. The effective number of reflections is now replaced by the number of equivalent sources that lie in half the circle (the greatest physical extent), and this gets smaller as the angle between the mirrors increases.

From an electric-engineering perspective, these secondary sources constitute a far-field uniform linear array antenna with excitation amplitudes arranged according to a geometric sequence. Such a configuration can be addressed with the standard methods of antenna theory [26]. The high gain and high directionality of this setup appear from the amazing properties of the FP.

*2.2.3. Self-consistent field.—* Instead of looking at the field inside the plate as arising from multiple reflected waves, we can just think of it as an intracavity field which has to travel the length of the cavity twice in opposite directions to complete a round trip [27]. We take as before the $z$ axis perpendicular to the mirrors, and denote by $A(z)$ the complex amplitude of this field at an arbitrary plane $z$ inside the mirrors, as schematized in figure 6. The self-consistency condition can be concisely formulated as

$$A(0) = A(0)\, r^2 \exp(-\mathrm{i}2\varphi) + t' A_{\mathrm{in}}\,. \tag{2.13}$$

This amount to imposing that the wave circulating after the first mirror has to be reconstructed by interference of the internal wave after one full round trip and the incident wave [28]. In other words, we impose that the field distribution reproduces itself after one round trip in the cavity. From here, we have

$$A(0) = \frac{t'}{1 - r^2 \exp(-\mathrm{i}2\varphi)} A_{\mathrm{in}}\,, \tag{2.14}$$

and if we take into account that $A(d) = A(0)\exp(-\mathrm{i}\varphi)$ and the boundary condition $A_{\mathrm{trans}} = t\,A(d)$, we immediately get that the transmission coefficient $T = A_{\mathrm{trans}}/A_{\mathrm{in}}$ coincides with (2.4), obtained as a summation of infinite waves.

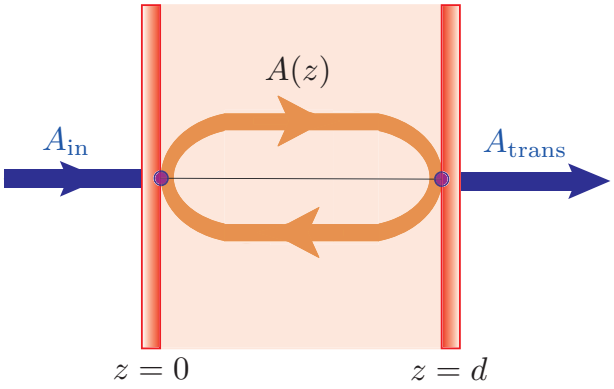


**Figure 6.** Schematic block diagram of an FP. We indicate the intracavity field, as well as the input and transmitted ones (the reflected is omitted for clarity). The axis $z$ is perpendicular to the mirrors.

We stress the leading role played by the boundary conditions in this idea. The method is at the realm of the transfer matrix approach [29], which is of great value in dealing with multilayered structures.

These results can also be reinterpreted using elementary notions of the theory of linear systems [30]. Indeed, the FP can be envisioned as a closed-loop system [31], as outlined in figure 7. Apart from the mirrors, we have a main box, which is just a delay line representing the propagation of the intracavity field. We have as well a feedback path with a frequency response $r^2\exp(-\mathrm{i}\varphi)$. A direct application of the standard rules leads to the same transmission coefficient $T(\varphi)$. Note that $T(\varphi)$ appears here as the transfer function and accordingly its poles and zeros provide full information of the response of the system. From this perspective, the FP is a linear, time-invariant and stable system.

The self-consistency condition can be worked out in the time domain, instead of the frequency domain. The intracavity field at an arbitrary time $\mathrm{t}$ is determined by the condition [32]

$$A(\mathrm{t}) = A(\mathrm{t}-2\mathrm{t}_c)\, r^2 \exp(-\mathrm{i}2\varphi) + t' A_{\rm in}\,, \tag{2.15}$$

where the single-pass time $\mathrm{t}_c$ is given in (2.12). Taking the Fourier transform on both sides of this equation yields the basic cavity response function as before. This formulation is especially suitable for cavities with moving mirrors, which is the basic scheme for the recent detection of gravitational waves [33].

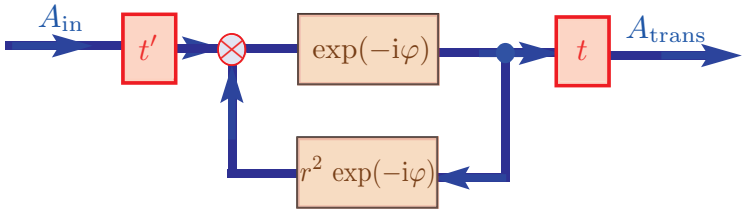


**Figure 7.** Structure diagram of the FP interferometer in figure 1.

*2.2.4. Impulse response.—* The previous example suggests to examine the general FP response in the time domain [34]. In this way, we can deal with dynamical situations, such as, e.g., pulse propagation in a FP. To this end, we apply a short pulse whose amplitude can be approximated as $A_{\rm in}\delta(\mathrm{t})$, where $\delta$ stands for the Dirac delta function. Neglecting any material dispersion, the transmitted amplitude, for $\mathrm{t}>0$, is

$$A_{\rm trans}(\mathrm{t}) = A_{\rm in}\, tt' \sum_{m=0}^{\infty} \delta(\mathrm{t}-\mathrm{t}_c-2m\mathrm{t}_c)\exp[-\gamma_c(\mathrm{t}-\mathrm{t}_c)]\,, \tag{2.16}$$

and we have defined $\exp(-\gamma_c \mathrm{t}_c) = r^2$. This $A_{\rm trans}(\mathrm{t})$ consists of a train of short pulses separated in time by $2\mathrm{t}_c$, but in each round trip the pulse amplitude is decreased by $r^2$. Rather than Fourier transforming (2.16) and summing up the resulting series, we follow an alternative route and rewrite (2.16) as

$$A_{\rm trans}(\mathrm{t}) = A_{\rm in}\,\frac{tt'}{r^2}\,\mathrm{comb}_{2\mathrm{t}_c}(\mathrm{t})\,\exp(-\gamma_c \mathrm{t})\,H(\mathrm{t})\,. \tag{2.17}$$

Here the Dirac comb is $\mathrm{comb}_{\mathrm{t}_c}(\mathrm{t}) = \sum_{m=-\infty}^{\infty}\delta(\mathrm{t}-m\mathrm{t}_c)$ and $H(\mathrm{t})$ is the Heaviside step function that ensures that the exponential is only for positive times. Since the Fourier transform of a Dirac comb is a Dirac comb, the convolution theorem, used backwards, suggests that the frequency response should be a convolution of an infinite comb of frequencies and a building block that is the transform of the one-sided exponential decay: this is just $1/(\gamma_c+\mathrm{i}\omega)$. Moreover, since the convolution with a delta function $\delta(\mathrm{t}-m\mathrm{t}_c)$ is equivalent to shifting the function by $m\mathrm{t}_c$, convolution with the Dirac comb corresponds to replication or periodic summation. In consequence, we get

$$T(\varphi) = \sqrt{\frac{1-r^2}{1+r^2}}\sum_{m=-\infty}^{\infty}\frac{1}{(\frac{1}{2}\Gamma)+\mathrm{i}(\varphi-m\pi)}\,, \tag{2.18}$$

where we have used the adimensional variable $\varphi$ and $\Gamma = \gamma_c \mathrm{t}_c = \ln(1/r^2)$ is the decay $\gamma_c$ in adimensional units. $T(\varphi)$ thus appears as a sum of infinity curves displaced by integer values of $\pi$.

## 3. Intensity response of a Fabry-Perot

### *3.1. The Airy distribution*

The intensity response of the FP (i.e., the ratio of the transmitted intensity to the incident one) is just $\mathscr{T} = |T|^2$, a notation that was already anticipated in section 2.1. Consequently, it can be readily obtained from our previous discussion on the amplitudes. The final result is

$$\mathscr{T}(\varphi) = \frac{1}{1+\mathfrak{f}\sin^2\varphi}\,, \tag{3.1}$$

where the parameter $\mathfrak{f}$ is

$$\mathfrak{f} = \frac{4r^2}{(1-r^2)^2}\,. \tag{3.2}$$

Equation (3.1) is the time honored Airy distribution. In figure 8 we plot $\mathscr{T}$ as a function of the phase thickness $\varphi$ and $r$. As $r$ increases, the minima of $\mathscr{T}$ fall and the maxima become sharper. In the limit of high $r$, the pattern consists on narrow bright fringes on an almost completely dark background. Moreover, since we are not considering absorption in the plate, the peak transmittance is unity (for any value of $r$).

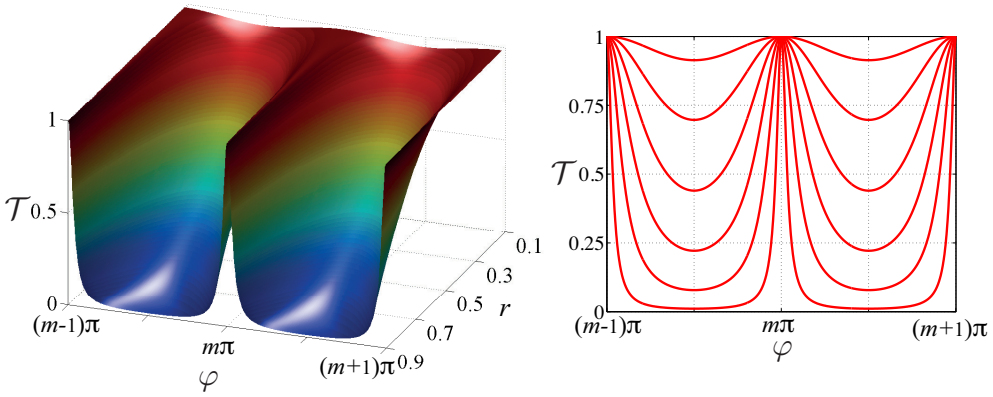


**Figure 8.** (Left) Transmissivity $\mathcal{T}$ of the FP as a function of the phase shift $\varphi$ and the parameter $r$. (Right) Cuts of the previous plot for reflection coefficients ranging from $r = 0.15$ to $r = 0.90$ in steps of 0.15.

Consequently, one requires an increased reflectivity $r^2$ and this is accomplished by coating the plates surfaces with a mirror. In what follows, we assume that such a mirror is lossless. In that case, the Airy formula still holds provided we interpret $r$ as the reflection coefficient of the mirror (which becomes, in general, a complex number). This adds to the plate phase thickness $\varphi$ a phase change on the reflection at the mirrors. In general, both modulus and phase of the complex $r$ depend on the angle of incidence and the dispersion properties of the material, albeit such a variation can be disregarded for most practical purposes.

The function $\mathcal{T}(\varphi)$ is $\pi$-periodic. The frequency separation of adjacent fringes is called the free spectral range (FSR)

$$\Delta\omega_{\mathrm{FSR}} = \frac{\pi c}{d}\,. \tag{3.3}$$

The sharpness of the fringes is conveniently measured by their full width at half maximum (FWHM), which is the frequency width between the two points on either side of a maximum where the intensity falls to half its maximum value. When $\mathfrak{f}$ is sufficiently large, this width is [14]

$$\Delta\omega_{\mathrm{FWHM}} = \frac{c}{d}\frac{1-r^2}{r} = \frac{2c}{d}\frac{1}{\sqrt{\mathfrak{f}}}\,. \tag{3.4}$$

The quotient between fringe separation and fringe width,

$$\mathcal{N} = \frac{\Delta\omega_{\mathrm{FSR}}}{\Delta\omega_{\mathrm{FWHM}}} = \frac{\pi\sqrt{\mathfrak{f}}}{2}\,, \tag{3.5}$$

is known as the finesse.

Since $\Gamma = \ln(1/r^2)$, $\mathcal{T}(\varphi)$ in equation (3.1) can be recast as

$$\mathcal{T}(\varphi) = \frac{1-r^2}{1+r^2}\left(\frac{\sinh\Gamma}{\cosh\Gamma-\cos 2\varphi}\right)\,. \tag{3.6}$$

The term in parentheses is just a wrapped Lorentz distribution [35], so that

$$\mathcal{T}(\varphi) = \frac{1-r^2}{1+r^2}\sum_{m=-\infty}^{\infty}\frac{\frac{1}{2}\Gamma}{(\varphi-m\pi)^2+(\frac{1}{2}\Gamma)^2}\,. \tag{3.7}$$

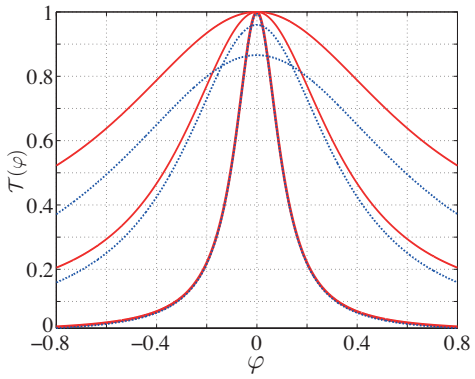


**Figure 9.** Comparison between the Airy distribution (in continuous red line) and the Lorentzian approximation (in broken blue lines) for the values $r = 0.5, 0.7$ and $0.9$ (from top to the bottom).

This is the intensity counterpart of equation (2.18) and could be obtained therefrom with a little of algebraic effort [36]. The Airy distribution appears in this way as a sum of Lorentzian profiles of width $\Gamma$ displaced by $\pi$ (which is just the FSR, in adimensional units). If the width $\Gamma$ is very small compared to the FSR, the contribution of every Lorentzian far from its peak can be neglected. Consequently, around the $m$th order, we can use

$$\mathscr{T}_{\rm np}(\varphi) \simeq \frac{1-r^2}{1+r^2}\,\frac{\frac{1}{2}\Gamma}{\varphi^2+(\frac{1}{2}\Gamma)^2}\,, \tag{3.8}$$

so that each fringe has a Lorentzian shape. This is called the narrow-peak approximation [5], and now the finesse can be estimated as

$$\mathscr{N}_{\rm np} = \frac{\pi}{\Gamma} = \frac{\pi}{\ln(1/r^2)} \simeq \frac{\pi}{1-r^2}\,, \tag{3.9}$$

which is equivalent to the expression (3.5) in the limit $r \to 1$, which is precisely when this approximation holds. This gives a better understanding of the phenomenon of peak broadening when $r$ decreases. In figure 9 we compare a single peak in the Airy function with the corresponding narrow-peak approximation, corroborating that it works well only for $r$ high enough.

### *3.2. Resolving power*

The resolving power gives the ability to discriminate between different wavelengths, to accomplish which it is first necessary to define mathematically a separation criterion between two very close maxima. Rayleigh criterion [37] is widely employed for that purpose: it 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 [14]. The resulting resolution limit seems to be rather arbitrary and is based on resolving capabilities of the human eye. Even more important, it cannot be directly applied to the Airy intensity profile because of the slow decrease of these values, so the minimum is located far from the maximum [38].

In our context, the most appropriate way of proceeding is perhaps to adopt the Taylor criterion [39, 40] (also called the FWHM criterion), which asserts that the separation of the

maxima is equal to the half-maximum width. For the FP, it gives a spectral resolving power (SRP)

$$\mathrm{SRP} \equiv \left| \frac{\lambda}{\Delta\lambda_{\mathrm{FWHM}}} \right| = \left| \frac{\omega}{\Delta\omega_{\mathrm{FWHM}}} \right| = \frac{2d}{\lambda}\mathscr{N} = m\mathscr{N}\,, \tag{3.10}$$

where $m = 2d/\lambda$ is the interference order at normal incidence. The SRP gives higher values for fringes near the centre of the interferometric pattern and for values of $r \to 1$, which means that closer wavelengths can be discriminated.

### *3.3. Resolution revisited*

*3.3.1. Q factor.—* When we consider the FP as an energy-storing system, a pivotal parameter is the quality factor $Q$, defined for any damped oscillator as

$$Q = \omega \times \frac{\text{energy stored}}{\text{power dissipated}}\,. \tag{3.11}$$

It is instructive to derive this expression for a FP and compare with the SRP we have found before.

If the energy in the field is $W$, the factor $Q$ can be formulated as

$$Q = \omega \frac{W}{W\dfrac{c}{d}\delta_{\mathrm{loss}}} = \omega \frac{d}{c}\frac{1}{\delta_{\mathrm{loss}}}\,, \tag{3.12}$$

where $\delta_{\mathrm{loss}}$ is the fractional loss per transit. Since we are neglecting any diffractional effect, losses are chiefly due to reflections in the mirrors and $\delta_{\mathrm{loss}} \simeq (1 - r^2)$ [22]. If we recall the definition of $\Delta\omega_{\mathrm{FWHM}}$, we arrive at

$$Q = \frac{2\pi d}{\lambda}\frac{1}{1-r^2}\,. \tag{3.13}$$

This coincides with equation (3.10) within a factor $r$, which is accounted for by the small-loss approximation assumed before. In consequence, as expected, the spectral resolving power is synonymous with the quality factor.

This comparison further establishes that if the energy losses are due entirely to reflection losses at the mirrors, the resolving power may be rigourously derived according to the multiple-beam treatment. When losses due to diffraction become appreciable, the effective resolving power must decrease, as $Q$ itself decreases.

This can be complemented with an easy uncertainty-principle argument: a wave in the FP will have a decay time $\Delta\mathrm{t}$ determined by the fractional loss per transit $\delta_{\mathrm{loss}}$ and can be roughly estimated as

$$\Delta\mathrm{t} = \frac{d}{c}\frac{1}{\delta_{\mathrm{loss}}}\,. \tag{3.14}$$

For such a decaying wave, the Fourier-transform determined bandwidth is

$$\Delta\omega \simeq \frac{1}{\Delta\mathrm{t}}\,, \tag{3.15}$$

which gives again the same value of $Q$.

*3.3.2. Equivalent sources.—* We recall that in 2.2.2 we have pinned down the FP action as being equivalent to an infinite linear array of sources. However, we anticipated that only an effective number of them contribute. Indeed, if we bear in mind that for a diffraction grating made of $N$ identical radiators one has [14]

$$\mathrm{SRP}_{\mathrm{grating}} = mN\,, \tag{3.16}$$

where $m$ is the order, this suggests, after a glance at (3.10), that the effective number of sources is precisely the finesse $\mathscr{N}$.

This can be further confirmed from the time-domain approach in 2.2.4. For an incident short pulse, the FP transmitted train of pulses fall in amplitude by a factor $1/\mathrm{e}$ when $r^{2N} = 1/\mathrm{e}$; that is, after

$$N = \frac{1}{\ln(1/r^2)} \simeq \frac{1}{1-r^2} = \frac{\mathscr{N}_{\mathrm{np}}}{\pi} \tag{3.17}$$

trips in the cavity. Therefore, the number of effective round trips is precisely the finesse (except for the factor $\pi$), and gives the number of equivalent sources considered before. The output pulse train has a length (from the starting to the 1/e point) of $n2d \times N$, and we can loosely say that the coherence length after traversing the FP has been increased by $N$ [24].

## 4. Conclusions

In summary, we have thoroughly explored several complementary viewpoints on the FP response, both in amplitude and intensity. Traditional discussions generally pinpoint the intensity, for this is the variable measured in experiments. However, amplitude is the natural arena to deal with the FP, as it is the variable for which the superposition principle holds. Even though translating this response into intensity is direct, the physical picture of the approximations involved can be better appreciated using amplitude, as we have illustrated in this work.

Yet very basic in nature, we hope that these results may be helpful in updating the modern views on the operation of such a relevant setup. We finally emphasize that the methods employed here are quite appealing for they have branched into offshoots of importance for many other modern physical theories.

## Acknowledgments

The original ideas in this paper have been developed and completed with questions, suggestions, criticism, and advice from many students and colleagues. Particular gratitude for help in various ways go to G. Björk, J. F. Cariñena, H. de Guise, and A. B. Klimov. We also thank two referees for their insightful comments that motivated us to get deeper into the most delicate aspects of the problem treated here. Financial support from the Spanish Ministerio de Economía y Competitividad (MINECO Grant FIS2015-67963-P) is gratefully acknowledged.

[1] Georgelin Y P and Amram P 1995 A review of Fabry and Perot discoveries *3D Optical Spectroscopic Methods in Astronomy* (*ASP Conference Series* vol 71) ed Marcelin G C M (SanFrancisco: Astronomical Society of the Pacific) pp 382–394
[2] Métivier F 2006 *Photoniques* **25** 2
[3] Kastler A 1974 *Nouv. Rev. Opt.* **5** 133
[4] Françon M 1966 *Optical Interferometry* (New York: Academic Press)
[5] Jacquinot P 1960 *Rep. Prog. Phys.* **23** 267–312
[6] Jacquinot P and Chabbal R 1956 *J. Opt. Soc. Am.* **46** 556–557
[7] Connes P 1986 *J. Opt.* **17** 5
[8] Steel W H 1999 *AOS News* **13** 5
[9] Orr B J 2006 *AOS News* **20** 16–23
[10] Amram P and Georgelin Y 2000 *Physica Scripta* **T86** 76–82
[11] Mulligan J F 1998 *Am. J. Phys.* **66** 797–802
[12] Hernandez G 1986 *Fabry-Perot Interferometers* (Cambridge: Cambridge University Press)
[13] Vaughan J M 1989 *The Fabry-Perot Interferometer: History, Theory, Practice, and Applications* (Bristol: Adam Hilger)
[14] Born M and Wolf E 1999 *Principles of Optics* (Cambridge: Cambridge University Press)
[15] Jacobs S F http://fp.optics.arizona.edu/stephen.f.jacobs/pdf/fabryperot.pdf
[16] Lawrence J D 1972 *Catalog of Special Plane Curves* (New York: Dover)
[17] Shikin E V 1995 *Handbook and Atlas of Curves* (London: CRC Press)
[18] Monzón J J and Sánchez-Soto L L 2015 *Eur. J. Phys.* **36** 045021
[19] Yu J, Yuan S, Gao J Y and Sun L 2001 *J. Opt. Soc. Am. A* **18** 2153–2160
[20] Hauge E H and Støvneng J A 1989 *Rev. Mod. Phys.* **61** 917–936
[21] Feynman R P, Leighton R B and Sands M L 2011 *The Feynman Lectures on Physics* New Millenium ed (Reading: Addison Wesley)
[22] Siegman A E 1986 *Lasers* (Sausalito: University Science Books)
[23] Lawrence M J, Willke B, Husman M E, Gustafson E K and Byer R L 1999 *J. Opt. Soc. Am. B* **16** 523–532
[24] Brooker G 2003 *Modern Classical Optics* (Oxford: Oxford University Press)
[25] Mcgloin D and Dholakia K 2005, *Contemp. Phys.* **46** 15 – 28
[26] Balanis C A 2016 *Antenna Theory: Analysis and Design* 4th ed (Wiley)
[27] Macleod H A 2010 *Thin-Film Optical Filters* 4th ed (Boca Raton: CRC Press)
[28] Meschede D 2004 *Optics, Light and Lasers* (Weinheim: Wiley-VCH)
[29] Sánchez-Soto L L, Monzón J J, Barriuso A G and Cariñena J F 2012 *Phys. Rep.* **513** 191–227
[30] Chen C T 2012 *Linear System Theory and Design* (Oxford: Oxford University Press)
[31] Donges A 1997 *Eur. J. Phys.* **18** 338–342
[32] Rakhmanov M, Jr R L S, Reitze D and Tanner D 2002 *Phys. Lett. A* **305** 239–244
[33] Abbott B P et al (LIGO Scientific and Virgo Collaborations) 2016 *Phys. Rev. Lett.* **116** 061102
[34] Kafri O, Kimel S and Lipson S G 1973 *Appl. Opt.* **12** 854–857
[35] Mardia K V and Jupp P E 2000 *Directional Statistics* (Chichester: Wiley)
[36] Koppelmann G 1969 Multiple-beam interference and natural modes in open resonators *Progress in Optics* vol VII ed Wolf E (Amsterdam: North-Holland) pp 3–66
[37] Rayleigh L 1879 *Phil. Mag.* **8** 261–274, 403–411, 477–486
[38] Juvells I, Carnicer A, Ferré-Borrull J, Martín-Badosa E and Montes-Usategui M 2006 *Eur. J. Phys.* **27** 1111–1119
[39] Klein M V and Furtak T E 1986 *Optics* (New York: Wiley)
[40] Fowles G R 1989 *Introduction to Modern Optics* (NewYork: Dover)