

Amplitude response of a Fabry-Perot interferometer

Juan J Monzón and Luis L Sánchez-Soto

Departamento de Óptica, Facultad de Física, Universidad Complutense, 28040 Madrid, Spain

Abstract. The complex reflected and transmitted amplitudes from a Fabry-Perot interferometer are analyzed using a phase-space approach, in which the real and imaginary parts of those amplitudes are taken as basic variables. As functions of the optical path length of the cavity, the reflected amplitude describes a circle, whereas the transmitted is given by a hippopede. The system performance can be directly related to the geometrical parameters of these curves.

1. Introduction

The Fabry-Perot (FP) interferometer provides a superb illustration of the mysterious ways in which interference works. Despite its apparent simplicity, it plays a central role as high-resolution spectrometer, laser resonator, or spectral filter, to cite only but a few of its many relevant uses. A complete account of the subject can be found in the two comprehensive monographs by Hernandez [1] and Vaughan [2].

This variety of fields of application spawned many descriptions of the FP operation, each one capitalizing on specific aspects. The geometric treatment, in which one adds the multiple beams reflected at each of the different interfaces, is probably more instructive and, accordingly, is reproduced in almost every textbook [3]. The question can also be tackled by imposing the appropriate boundary conditions, which gives the resonant frequencies and allowed fields in the FP [4]. As the boundary conditions appear as a linear system, they can pop up under multiple guises: for example, the FP may be viewed as an optical transmission system with feedback [5], or as a direct application of the transfer matrix [6, 7], a method especially germane to deal with layered structures.

Irrespective of the approach, the focus is always on the intensity distributions (both in reflection and transmission); namely, the well-known Airy formulas. Even if the corresponding amplitudes are somehow required to determine these distributions, they are ultimately discarded on the grounds that real experiments measure intensity.

This viewpoint can be challenged, however, by a discussion of the conventional harmonic oscillator, wherein the complex amplitude is decomposed into two orthogonal in-phase and out-of-phase quadratures, also known as the dispersive and absorptive components [8]. These quadratures are of paramount importance as they convey more useful information than just the intensity. At the quantum level, for example, they play the role of the effective position and momentum of the oscillator [9]. Actually, a good deal of the latest advances in quantum information processing stem from a proper engineering of these quadratures, with homodyne detection constituting an ideal tool for their measurement, whereas squeezing them provides an efficient route to producing entanglement [10].

Inspired by this, we intend to shed light on the amplitude response of the FP. Indeed, one can define an equivalent version of the quadratures. When the parameters of the FP vary, the amplitudes trace out elementary curves: the reflected amplitude is a circle, and the

transmitted one is a hippopede, a curve with remarkable properties [11, 12]. Furthermore, the FP performance can be naturally assigned to the geometrical features of these curves.

Apart from its elegance, this approach makes a close contact with phase-space methods that pervade physics today. The derivation is straight and easy; suitable for undergraduates. Surprisingly enough, to the best of our knowledge, such simple ideas have been not hitherto explored; even if they can be important in some instances in which the phase introduced by the FP matters, as it happens in optical metrology, where the stabilization is crucial.

2. The Fabry-Perot: Basic background

The ideal FP interferometer consists of two parallel mirrors (that, for simplicity, we assume to be identical) separated at a distance d . Figure 1 is a block diagram of the system. This can be addressed by considering 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.

The complex reflection and transmission coefficients (i.e., the ratios of the reflected and the transmitted amplitudes to the incident one, respectively) are given by [6]

$$R(\Phi) = \frac{r[1 - \exp(-i2\Phi)]}{1 - r^2 \exp(-i2\Phi)}, \quad T(\Phi) = \frac{(1 - r^2) \exp(-i\Phi)}{1 - r^2 \exp(-i2\Phi)}. \quad (2.1)$$

Here, r is the Fresnel reflection coefficient for a wave travelling from the surrounding medium into the FP and

$$\Phi = \frac{2\pi}{\lambda} nd \cos \theta \quad (2.2)$$

is the plate phase thickness, with λ the wavelength in vacuum and θ the angle of refraction in the medium n , which is related to the angle of incidence according to Snell's law.

The usual analysis proceeds by calculating the reflectivity and transmissivity (i.e., the ratios of the reflected and the transmitted intensities to the incident one). The expressions are

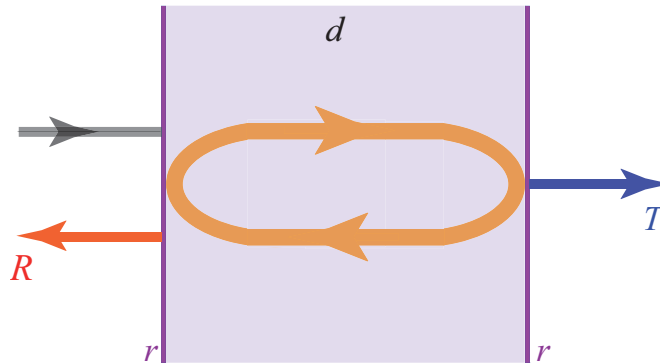


Figure 1. Schematic block diagram of an FP. The intracavity fields can be analyzed from a variety of perspectives. The interferometer plate surfaces have reflection coefficient r and they are separated at a distance d . The reflected and transmitted amplitudes are labeled by R and T , respectively, since we assume unit incident amplitude.

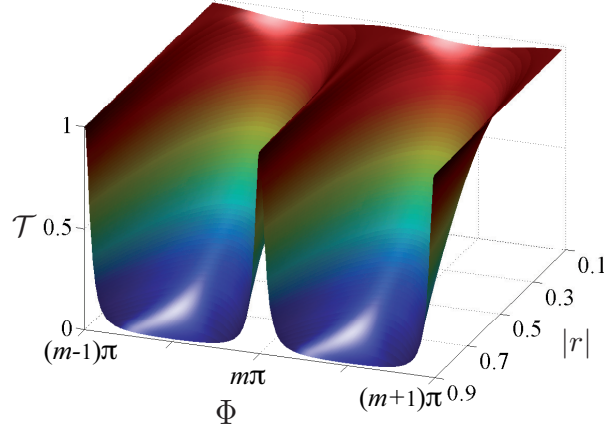


Figure 2. Transmissivity \mathcal{T} of the FP as a function of the phase shift Φ and the parameter $|r|$, whose square is the reflectivity of the plate surfaces.

obtained directly from equation (2.1) and read

$$\mathcal{R} = |R|^2 = \frac{F \sin^2 \Phi}{1 + F \sin^2 \Phi}, \quad \mathcal{T} = |T|^2 = \frac{1}{1 + F \sin^2 \Phi}, \quad (2.3)$$

where the parameter F is

$$F = \frac{4|r|^2}{(1 - |r|^2)^2}. \quad (2.4)$$

Although r is up to now a real number, we formally treat it as a complex for reasons that will become apparent soon and so $|r|^2$ is the reflectivity of the plate surfaces. Equations (2.3) constitute the time honored Airy formulas. Evidently, since there are no losses, the two patterns are complementary, in the sense that

$$\mathcal{R} + \mathcal{T} = 1. \quad (2.5)$$

In figure 2 we plot the transmissivity \mathcal{T} as a function of the phase thickness Φ and $|r|$. As $|r|$ increases, the minima of \mathcal{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.

The sharpness of the fringes is conveniently measured by their full width at half maximum (FWHM), which is the width between the points on either side of a maximum where the intensity has fallen to half its maximum value. The ratio of the separation of adjacent fringes (also called free spectral range) and the FWHM is called the finesse \mathcal{F} of the fringes. A direct calculation shows that

$$\mathcal{F} = \frac{\pi\sqrt{F}}{2}. \quad (2.6)$$

This quantity is a measure of the apparatus ability to resolve closely spaced spectral features. High values of \mathcal{F} require 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 formulas still hold provided we interpret r as the reflection coefficient of the mirror (which becomes now a complex number). This adds to the plate phase thickness Φ 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.

3. Amplitude response of the Fabry-Perot

Let us look in more detail at the amplitudes (2.1). First of all, we observe that $R(\Phi)$ is a π -periodic function, while $T(\Phi)$ is 2π -periodic. Such a difference cannot be seen in the intensity response, for both \mathcal{R} and \mathcal{T} have the same period π .

To proceed further, let us rewrite equation (2.1) as

$$R(\Phi) = \frac{2|r|\sin\Phi}{N(\Phi)}[X(\Phi) + iY(\Phi)], \quad (3.1)$$

$$T(\Phi) = \frac{1 - |r|^2}{iN(\Phi)}[X(\Phi) - iY(\Phi)].$$

where we have defined

$$X(\Phi) = (1 + |r|^2)\sin\Phi, \quad Y(\Phi) = (1 - |r|^2)\cos\Phi, \quad (3.2)$$

$$N(\Phi) = (1 - |r|^2)^2 + 4|r|^2\sin^2\Phi.$$

We recall that any harmonic signal $x(t)$ can be decomposed as

$$x(t) = X \cos \omega t + Y \sin \omega t, \quad (3.3)$$

where the X and Y are the in-phase and $(\pi/2)$ out-of-phase quadratures. In this spirit, $X(\Phi)$ and $Y(\Phi)$ can be seen as sort of quadratures for $R(\Phi)$ and $T(\Phi)$. We stress, however, that at difference of the harmonic oscillator, here the pre-factors in their definition (3.1) are not constant, but depend on Φ . This arises because the complex amplitudes $R(\Phi)$ and $T(\Phi)$ do not have constant modulus, as in the oscillator.

In figure 3 we represent these complex amplitudes for several values of $|r|$. The different behavior commented above in relation with the periodicity translates into the fact that when $T(\Phi)$ completes a revolution, $R(\Phi)$ makes two turns. Both amplitudes lie inside the unit disk because of the fundamental constraint (2.5).

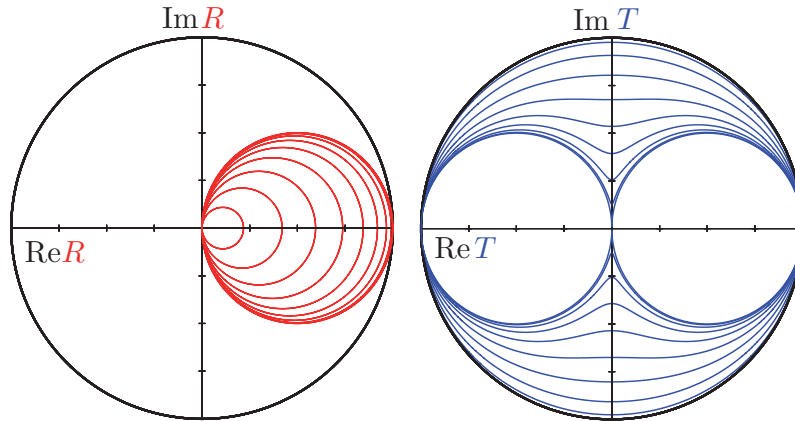


Figure 3. Phase-space trajectories for $R(\Phi)$ (left) and $T(\Phi)$ (right) as given in equation (3.1). They lie inside the unit disk. The different curves correspond to different values of the plate reflectivity ranging from $r = 0.11$ to $r = 0.99$ in steps of 0.11. For $R(\Phi)$, the curves increase in size with $|r|$, while the converse happens for $T(\Phi)$.

To get a better grasp of these amplitudes, we introduce polar coordinates as

$$R = |R| \exp(i\rho), \quad T = |T| \exp(i\tau). \quad (3.4)$$

With this parametrization, we can recast equations (3.1) as

$$\mathcal{R} = 4a^2 \cos^2 \rho, \quad \mathcal{T} = 1 - 4a^2 \sin^2 \tau, \quad (3.5)$$

where we have used the definition in (2.3) and a is a real parameter

$$a = \frac{|r|}{1 + |r|^2}, \quad (3.6)$$

so that $0 \leq a \leq 1/2$.

In this way, the reflected amplitude $R(\Phi)$ can be immediately identified as a circle of radius a centered in the point $(a, 0)$ of the real axis.

The transmitted amplitude $T(\Phi)$ describes a hippopede (which literally means “horse fetter”). It was first investigated by Proclus [13] and later on by Booth [14], hence their names are sometimes attached to this stunning curve. For $0 < a < 1/\sqrt{8}$ it is an oval, and for $1/\sqrt{8} < a < 1/2$ it is an indented oval, which tends to be an eight in the limit $a = 1/2$ (i.e., $|r| \rightarrow 1$). Yet not so well known in Physics, it has a truly amazing set of properties that the reader can look up in the abundant literature on the subject [15–17]. We merely quote that the hippopede can be defined as the curve formed by the intersection of a torus and a plane parallel to the axis of the torus and tangent to it on the interior circle. It is thus a spiric section [18].

For every value of $|r|$, the reflected amplitude passes through the origin: $R(\Phi)$ is zero for $\Phi = 0$ and π and traces the circle clockwise, getting its maximum at $\Phi = \pm\pi/2$, where $\rho = 0$ and then, according equation (3.5), $\mathcal{R}_{\max} = 4a^2$.

On the other hand, the transmitted amplitude also describes the hippopede clockwise. At $\Phi = 0$ and π , $T(\Phi)$ reaches its maxima, which are in the real axis at $T = 1$ and -1 , respectively. The minimum occurs at $\Phi = \pi/2$ and $3\pi/2$, where $\tau = -\pi/2$ and $-3\pi/2$, respectively. Therefore $\mathcal{T}_{\min} = 1 - 4a^2$, which corresponds to half the waist of the hippopede at its indentation.

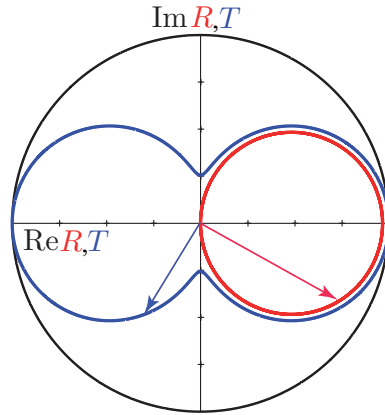


Figure 4. Reflected (red) and transmitted (blue) amplitudes $R(\Phi)$ and $T(\Phi)$ for a transparent FP with $|r| = 0.77$. For every fixed value of the parameter Φ the corresponding position vectors are orthogonal.

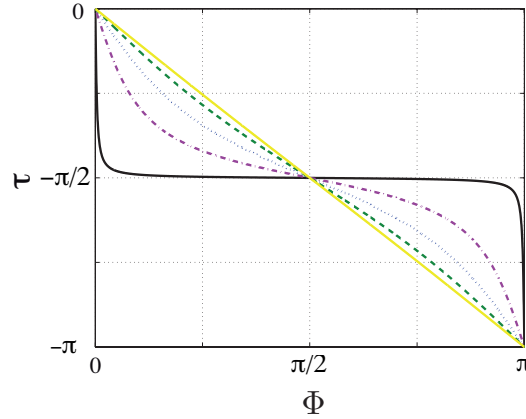


Figure 5. Transmitted phase lag τ by an FP as a function of Φ for different values of $|r|$, from 0.11 (yellow full line) to 0.99 (black full line) in steps of 0.22. The curves bend more as $|r|$ increases.

One can check that

$$\frac{R(\Phi)}{T(\Phi)} = i\sqrt{F} \sin \Phi, \quad (3.7)$$

which, in turn, implies that for a transparent symmetric system, as the one we are dealing with, we have

$$\rho(\Phi) - \tau(\Phi) = \pm \frac{\pi}{2}. \quad (3.8)$$

Consequently, for every value of Φ the reflected and transmitted amplitudes are at quadrature. This is illustrated in figure 4, where we see that the position vectors of $R(\Phi)$ and $T(\Phi)$ are orthogonal, as it is somehow implicit in equation (3.1). Apropos of this, it is noteworthy to mention that quite similar relations may be derived under the general assumptions of symmetric and lossless systems [19–21].

We next examine the local slopes $\dot{\rho}(\Phi)$ and $\dot{\tau}(\Phi)$, the dot denoting derivative respect to the parameter. They are the “rates” at which the curves are traced out. Indeed, they entail a valuable physical interpretation. If we focus for simplicity at τ , we can write

$$\frac{d\tau}{d\omega} = \dot{\tau}(\Phi) \frac{d\Phi}{d\omega}, \quad (3.9)$$

with ω being the angular frequency. Now, $d\Phi/d\omega$ is the single-pass time inside the cavity medium (for a non-dispersive material, this is $nd \cos \theta/c$, where c is the velocity of light in vacuum), and $d\tau/d\omega$ is the time flight through the FP, which incorporates the feedback. Hence, $d\tau/d\omega$ may be viewed as an enhancement factor of the time flight due to the FP [22, 23].

Because of equation (3.8), both are equal for a lossless medium: $\dot{\rho}(\Phi) = \dot{\tau}(\Phi)$. In addition, we have

$$\frac{\dot{\tau}}{\mathcal{F}} = -\frac{1 + |r|^2}{1 - |r|^2}, \quad (3.10)$$

where the negative sign indicates that the curve is oriented clockwise. This quotient is thus independent of Φ : where the transmitted amplitude is large, so is the velocity and the opposite.

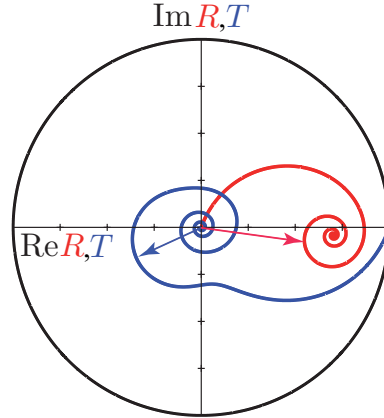


Figure 6. Same as in figure 4, but for an absorbing FP made of germanium with complex refractive index $N = 5.588 - i 0.933$ at a wavelength of $0.6199 \mu\text{m}$. We take normal incidence and the film thickness d varying between 0 and $0.35 \mu\text{m}$. The marked position vectors correspond to $d = 0.052 \mu\text{m}$.

To gain further insight into this issue, in figure 5 we plot τ as a function of Φ for several values of $|r|$. Note that the range of variation of τ is from -2π to 0 when Φ goes from 0 to 2π , but ρ ranges from $-\pi/2$ to $\pi/2$, as one can directly infer at a glance from figure 4. When $|r|$ is small, τ is almost a straight line, with a slope pretty constant: this is the case when the hippede is almost an oval, without indentation. However, as $|r|$ increases, τ starts bending near $\Phi = -\pi/2$, which is precisely at the waist. In the limit $|r| \rightarrow 1$, τ becomes almost horizontal, with slope zero almost everywhere, except a narrow interval around the maxima of \mathcal{S} , where it quickly gets large.

The previous considerations can be extended to a lossy (or gain) cavity medium, a case of particular interest in laser physics. The medium is now specified by a complex refractive index (and so a complex Φ), whose imaginary part accounts for the losses. The resulting trajectories, for the simple case of a plate of germanium, are shown in figure 6 and they turn out to be loxodromics [24], a universal feature of absorption. They start at $R = 0$ and $T = 1$ (when there is no film) and tend to $R \rightarrow r$ and $T \rightarrow 0$ (when the film becomes opaque). We can appreciate that the orthogonality between position vectors does not hold true, neither the complementary relation equation (2.5).

4. Concluding remarks

In summary, we have thoroughly explored the amplitude response of an FP interferometer. Despite its basic nature and its simplicity, this topic has been snubbed in the literature, which has strengthened instead the role of the associated intensity.

Given the relevant role played by the Airy formulas, one might have expected that their “square root” counterparts should have remarkable features. This is indeed the case, as our results indicate: the reflected amplitude traces a circle, while the transmitted one is a hippede, an intriguing curve full of nice mathematical properties.

We finally stress that the phase-space methods employed here are quite appealing for they have branched into offshoots of importance for modern physical theories.

Acknowledgments

Many of the ideas in this paper originated from a long cooperation with the late Alberto G. Barriuso, who unexpectedly passed away before being able to guide this work to completion. This paper is dedicated to his memory. Over the years, these ideas have been further developed and expanded with questions, suggestions, criticism, and advice from many colleagues. Particular thanks for help in various ways goes to G. Björk, J. F. Cariñena, H. de Guise, P. de la Hoz, A. B. Klimov, G. Leuchs, and J. M. Montesinos-Amilibia. This work is partially supported by the Spanish MINECO (Grant FIS2011-26786).

- [1] Hernandez G 1986 *Fabry-Perot Interferometers* (Cambridge: Cambridge Univ. Press)
- [2] Vaughan J M 1989 *The Fabry-Perot Interferometer: History, Theory, Practice, and Applications* (Bristol: Adam Hilger)
- [3] Born M and Wolf E 1999 *Principles of Optics* (Cambridge: Cambridge Univ. Press)
- [4] Saleh B A E and Teich M C 2007 *Fundamentals of Photonics* (New York: Wiley)
- [5] Chen C T 2012 *Linear System Theory and Design* (Oxford: Oxford Univ. Press)
- [6] Yeh P 2005 *Optical Waves in Layered Media* (New York: Wiley)
- [7] Sánchez-Soto L L., Monzón J J, Barriuso A G, and Cariñena J 2012 *Phys. Rep.* **513** 191–227
- [8] Crawford Jr F S 1968 *Waves (Berkeley Physics Course Vol. 3)* (New York: MacGraw-Hill)
- [9] Schleich W P 2001 *Quantum Optics in Phase Space* (Berlin: Wiley)
- [10] Braunstein S L and van Loock P 2005 *Rev. Mod. Phys.* **77** 513–577
- [11] Lawrence J D 1972 *Catalog of Special Plane Curves* (New York: Dover)
- [12] Shikin E V 1995 *Handbook and Atlas of Curves* (London: CRC)
- [13] Proclus, 1992 *A Commentary on the First Book of Euclid's Elements*. (Princeton: Princeton Univ. Press)
- [14] Booth J 1877 *A Treatise on Some New Geometrical Methods*, vol. II (London: Longmans, Green, Reader and Dyer)
- [15] Ferréol R *Encyclopédie des formes mathématiques remarquables* <http://www.mathcurve.com>.
- [16] Coffman A *National Curve Bank* <http://users.ipfw.edu/CoffmanA/pov/spiric.html>.
- [17] Wassenaar J *Mathematical Curves*, <http://www.2dcurves.com>.
- [18] Brieskorn E and Knörrer H 1986 *Plane Algebraic Curves* (Basel: Birkhäuser)
- [19] Degiorgio V 1980 *Am. J. Phys.* **48**, 81–82
- [20] Zeilinger A 1981 *Am. J. Phys.* **49** 882–883
- [21] Ou Z Y and Mandel L 1989 *Am. J. Phys.* **57** 66–67
- [22] Yariv A and Yeh P 2006 *Photonics: in Modern Communications* 6th edition (Oxford: Oxford University Press)
- [23] Schwelb O 2004 *J. Lightwave Technol.* **22** 1380–1394
- [24] Monzón J J, Barriuso A G, Sánchez-Soto L L and Montesinos-Amilibia J M 2011 *Phys. Rev. A* **84** 023830