

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Progress in Quantum Electronics

journal homepage: www.elsevier.com/locate/pqe

Spectral coherence, Part I: Passive-resonator linewidth, fundamental laser linewidth, and Schawlow-Townes approximation

Markus Pollnau^{a,*}, Marc Eichhorn^{b,c}^aAdvanced Technology Institute, Department of Electrical and Electronic Engineering, University of Surrey, Guildford, GU2 7XH, UK^bInstitute of Control Systems (IRS), Karlsruhe Institute of Technology (KIT), Fritz-Haber-Weg 1, 76131, Karlsruhe, Germany^cFraunhofer IOSB (Institut für Optronik, Systemtechnik und Bildauswertung), Gutleuthausstraße 1, 76275, Ettlingen, Germany

ARTICLE INFO

Keywords:

Spectral coherence
Optical resonance
Fabry-Pérot resonator
Resonator linewidth
Finesse
Lasers
Laser theory
Laser resonators
Laser linewidth
Schawlow-Townes equation

ABSTRACT

The degree of spectral coherence characterizes the spectral purity of light. It can be equivalently expressed in the time domain by the decay time τ or the quality factor Q of the light-emitting oscillator, the coherence time τ^{coh} or length l^{coh} of emitted light or, via Fourier transformation to the frequency domain, the linewidth $\Delta\nu$ of emitted light. We quantify these parameters for the reference situation of a passive Fabry-Pérot resonator. We investigate its spectral line shapes, mode profiles, and Airy distributions and verify that the sum of all mode profiles generates the corresponding Airy distribution. The Fabry-Pérot resonator is described, as an oscillator, by its Lorentzian linewidth and finesse and, as a scanning spectrometer, by its Airy linewidth and finesse. Furthermore, stimulated and spontaneous emission are analyzed semi-classically by employing Maxwell's equations and the law of energy conservation. Investigation of emission by atoms inside a Fabry-Pérot resonator, the Lorentz oscillator model, the Kramers-Kronig relations, the amplitude-phase diagram, and the summation of quantized electric fields consistently suggests that stimulated and spontaneous emission of light occur with a phase 90° in lead of the incident field. These findings question the quantum-optical picture, which proposed, firstly, that stimulated emission occurred in phase, whereas spontaneous emission occurred at an arbitrary phase angle with respect to the incident field and, secondly, that the laser linewidth were due to amplitude and phase fluctuations induced by spontaneous emission. We emphasize that the first derivation of the Schawlow-Townes laser linewidth was entirely semi-classical but included the four approximations that (i) it is a truly continuous-wave (cw) laser, (ii) it is an ideal four-level laser, (iii) its resonator exhibits no intrinsic losses, and (iv) one photon is coupled spontaneously into the lasing mode per photon-decay time τ_c of the resonator, independent of the pump rate. After discussing the inconsistencies of existing semi-classical and quantum-optical descriptions of the laser linewidth, we introduce the spectral-coherence factor, which quantifies spectral coherence in an active compared to its underlying passive mode, and derive semi-classically, based on the principle that the gain elongates the photon-decay time and narrows the linewidth, the fundamental linewidth of a single lasing mode. This linewidth is valid for lasers with an arbitrary energy-level system, operating below, at, or above threshold and in a cw or a transient lasing regime, with the gain being smaller, equal, or larger compared to the losses. By applying approximations (i)-(iv) we reproduce the original Schawlow-Townes equation. It provides the hitherto missing link between the description of the laser as an amplifier of spontaneous emission and the Schawlow-

* Corresponding author.

E-mail address: m.pollnau@surrey.ac.uk (M. Pollnau).

<https://doi.org/10.1016/j.pquantelec.2020.100255>

Available online 9 March 2020

0079-6727/© 2020 Published by Elsevier Ltd.

Townes equation. Spontaneous emission entails that in a cw lasing mode the gain is smaller than the losses. We verify that also in the quantum-optical approaches to the laser linewidth, based on the density-operator master equation, the gain is smaller than the losses. We conclude this work by presenting the derivation of the laser linewidth in a nut shell.

1. Introduction

A major success in the early days of quantum optics was the alleged proof that the laser linewidth or, more generally, the spectral-coherence properties of a laser are based on the quantum-optical principle that spontaneous emission induces amplitude and phase fluctuations. It was, therefore, concluded that the laser linewidth could only be derived in a full quantum-mechanical approach. In this first part of our investigations, we proof in a self-consistent approach that the fundamental laser linewidth is based on the semi-classical principle that the gain elongates the photon-decay time and narrows the linewidth of emitted light and that the original Schawlow-Townes linewidth is a four-fold approximation of this fundamental laser linewidth. Some of our results have previously been published [1–8]. These semi-classical investigations have become necessary, because there are significant misunderstandings, discrepancies, and confusion in the available literature on this subject, especially arising from approximations and misinterpretations that have appeared in semi-classical and quantum-optical approaches.

In Sections 2-4, we systematically characterize the Fabry-Pérot resonator by extending previous results [2] to the case with propagation losses. We clarify the mode structure of the Fabry-Pérot resonator and point out that the two counter-propagating waves at the same resonance frequency and polarization are independent modes [1]. We define the reference situation of a passive Fabry-Pérot resonator with outcoupling and propagation losses and a transparent active medium and derive the spectral-coherence properties of a single longitudinal mode. We derive the generic Airy distribution of a Fabry-Pérot resonator. All related Airy distributions are obtained by simple intensity scaling factors. We define the longitudinal mode profiles and verify that the sum of all longitudinal mode profiles generates the respective Airy distribution. We point out the different functions of the Fabry-Pérot resonator as a resonator with fixed length, characterized by its Lorentzian linewidth and finesse, and as a scanning spectrometer by varying its length, characterized by its Airy linewidth and finesse.

In Section 5, we investigate the phase aspect of stimulated and spontaneous emission [3,4]. By exploiting Maxwell's equations and the law of energy conservation, investigating stimulated emission in a Fabry-Pérot resonator, analyzing the Lorentz oscillator model, applying the Kramers-Kronig relations to the complex susceptibility, understanding the summation of quantized electric fields, and quantitatively interpreting emission and absorption in the amplitude-phase diagram, we derive a consistent semi-classical picture of the phase aspect in stimulated and spontaneous emission. Both processes occur with a phase 90° in lead of the incident field.

In Sections 6-8, we expand our investigation to the situation of an active or lasing Fabry-Pérot resonator. We introduce the Schawlow-Townes linewidth and emphasize that its derivation was entirely semi-classical but included four approximations. In their derivation of the maser linewidth [9], Gordon, Zeiger, and Townes assumed (i) a truly continuous-wave (cw) laser and (ii) an ideal four-level laser, (iii) with its resonator exhibiting no intrinsic losses. Schawlow and Townes [10] transferred the result to the optical regime by assuming that (iv) one photon is coupled spontaneously into the lasing mode per photon-decay time τ_c of the resonator, independent of the pump rate. We scrutinize existing semi-classical and quantum-optical descriptions of the laser linewidth and point out various inconsistencies. By describing the laser as an amplifier of spontaneous emission, whose gain elongates the photon-decay time and narrows the linewidth, and introducing the spectral-coherence factor Λ , which quantifies spectral coherence in an active compared to its underlying passive mode, we derive semi-classically the fundamental linewidth of a single lasing mode. This linewidth is valid for lasers with an arbitrary energy-level system, operating below, at, or above threshold and in a cw or a transient lasing regime, with the gain being smaller, equal, or larger compared to the losses. It should then not come as a surprise that, by applying approximations (i)-(iv) to the fundamental laser linewidth, we are able to reproduce the original Schawlow-Townes equation. It provides the hitherto missing link between the description of the laser as an amplifier of spontaneous emission and the Schawlow-Townes equation. Since spontaneous emission converts one atomic excitation into one photon, the spontaneous-emission rate into the lasing mode is positive. Consequently, in a cw lasing mode the gain is smaller than the losses. We verify that also in the quantum-optical approaches to the laser linewidth, based on the density-operator master equation, the gain is smaller than the losses, hence the same fundamental laser linewidth must also emerge from the quantum-optical descriptions. Finally, to summarize our results, we present the derivation of the laser linewidth in a nut shell.

2. Fabry-Pérot resonator: fundamental properties

The Fabry-Pérot resonator which was invented in 1899 [11] has proven very useful as a high-finesse interferometer in uncountable spectroscopic applications. Since 1960, it has also formed the fundamental basis for a large class of open resonators that have enabled laser oscillation. It has been described in many text books, e.g. by Vaughan [12,13], Siegman [14–17], Svelto [18,19], and Saleh and Teich [20–22].

In this Section, it is shown that the two counter-propagating waves with wave vectors $\pm|k_q|$ at the same resonance frequency ν_q and polarization constitute independent optical modes with mode indices $\pm|q|$. Often confusion arises in the literature from unprecise definitions, misinterpretation of the precise meaning of parameters, or inconsequent use of parameters, not least concerning resonators, modes, lasers, and spectral-coherence properties. Therefore, we define the reference situation, namely a passive Fabry-Pérot resonator

with a transparent active medium that exhibits discrete outcoupling losses and continuous propagation losses. We then derive the spectral-coherence properties of a single longitudinal mode inside this passive Fabry-Pérot resonator. In subsequent Sections, we will compare to this reference situation the situation of a single longitudinal mode inside a Fabry-Pérot-type resonator with an absorbing, amplifying, or lasing active medium.

2.1. Resonance frequencies and free spectral range

In our investigation, we assume a two-mirror Fabry-Pérot resonator of geometrical length ℓ , which is homogeneously filled with a medium of refractive index n_r . Light travels inside the resonator with speed

$$c = \frac{c_0}{n_r}, \quad (1)$$

where c_0 is the speed of light in vacuum. The round-trip time t_{RT} of light travelling in the resonator is given by

$$t_{RT} = \frac{2\ell}{c}. \quad (2)$$

Both ℓ and n_r are assumed to vary insignificantly over the frequency range of interest. Consequently, also the speed of light c and the round-trip time t_{RT} are assumed to be independent of frequency.

With $\phi(\nu)$ quantifying the single-pass phase shift that light exhibits when propagating from one mirror to the other, the round-trip phase shift at frequency ν accumulates to

$$2\phi(\nu) = 2\pi\nu t_{RT} = 2\pi\nu \frac{2\ell}{c}. \quad (3)$$

The difference in phase shift per round trip with frequency is [13]

$$\frac{d}{d\nu}(2\phi) = \frac{d}{d\nu}(2\pi\nu t_{RT}) = 2\pi t_{RT}. \quad (4)$$

Resonances occur at frequencies at which light exhibits constructive interference after one round trip, i.e., light at these frequencies has accumulated a phase shift that is a multiple integer of 2π . Consequently, the difference in phase shift accumulated per round trip between adjacent resonance frequencies amounts to 2π , from which the free spectral range $\Delta\nu_{FSR}$ between adjacent resonance frequencies derives as

$$\frac{2\pi}{\Delta\nu_{FSR}} = 2\pi t_{RT} \Rightarrow \Delta\nu_{FSR} = \frac{1}{t_{RT}}. \quad (5)$$

For a given polarization of light, each spectral (or longitudinal) resonator mode with its mode index q , where q is an integer number in the interval

$$q = -\infty, \dots, -1, 0, 1, \dots, \infty \quad (6)$$

is associated with a resonance frequency ν_q and wavenumber k_q ,

$$\nu_q = q \frac{1}{t_{RT}} = q \Delta\nu_{FSR} = q \frac{c}{2\ell} \Rightarrow k_q = \nu_q \frac{2\pi}{c} = q \frac{1}{t_{RT}} \frac{2\pi}{c} = q \Delta\nu_{FSR} \frac{2\pi}{c} = q \frac{\pi}{\ell}. \quad (7)$$

Two modes with opposite values $\pm q$ and $\pm k$ of modal index and wavenumber, respectively, physically representing opposite propagation directions, occur at the same absolute value $|\nu_q|$ of resonance frequency. This fact will be verified in the following Sub-section.

2.2. Mode density and counter-propagating modes

There exists some confusion within the optics community as to whether the two counter-propagating electromagnetic waves at the same frequency and polarization in a one-dimensional standing-wave Fabry-Pérot-type resonator, i.e., a resonator of finite length with significant reflection at its ends, belong to the same mode or constitute independent modes. These situations would result in either one or two longitudinal modes per resonance frequency and polarization, which may have significant consequences, e.g. for the distribution of vacuum energy among the modes, the excess noise factor [23,24] and laser linewidth [10,25,26], mode competition in lasers, and so on. The relevant issues will be discussed in part II of this paper.

The two counter-propagating waves at the same absolute resonance frequency $|\nu_q|$ have positive and negative mode indices $\pm|q|$ and wave vectors $\pm|k_q|$. It seems that a majority of scientists supports the interpretation that these counter-propagating waves belong to the same mode in a standing-wave resonator; hence, only one longitudinal mode would exist per resonance frequency and polarization [6,27–29]. For example, Saleh and Teich state explicitly [27]: “Negative values of q do not constitute independent modes

since $\sin k_{-qz} = -\sin k_{qz}$. Here, we provide evidence that the other interpretation is correct that two longitudinal modes exist per resonance frequency and polarization [1].

We commence our investigation in three-dimensional space. It is widely accepted that the three-dimensional mode density per unit frequency interval (in spherical coordinates with radial distance, $k_r \geq 0$, polar angle, $0 \leq \theta \leq \pi$, and azimuthal angle, $0 \leq \varphi \leq 2\pi$) can be calculated from an integration in k -space,

$$M^{3D}(\nu) = \frac{2}{(2\pi)^3} \int_{k_r=0}^k \int_{\theta=0}^{\pi} \int_{\varphi=0}^{2\pi} k_r^2 \sin(\theta) d\varphi d\theta dk_r = \frac{8\pi\nu^3}{3c^3}, \quad (8)$$

where the factor of two accounts for the two orthogonal polarizations, which constitute independent modes. The results presented here are independent of whether both polarizations are considered or one of them is suppressed by an intra-cavity polarizing element. The spectral mode density then derives as [30–33]

$$\tilde{M}^{3D}(\nu) = \frac{d}{d\nu} M^{3D}(\nu) = \frac{8\pi\nu^2}{c^3}. \quad (9)$$

Importantly, the result of Eq. (9) is valid not only in free space, as derived above, but also for a three-dimensional standing-wave resonator, as one can easily see from the fact that Planck's law of blackbody radiation [34] derives from the same spectral mode density. Its spectral energy density is given by

$$\tilde{u}(\nu) = \tilde{M}^{3D}(\nu) h\nu \varphi_{mode} = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{\exp(h\nu/k_B T) - 1}. \quad (10)$$

Here, h is Planck's constant, φ_{mode} is the number of photons per optical mode according to Bose-Einstein statistics [35], k_B is the Boltzmann constant, and T is the temperature.

Let us define one independent direction in k -space to include both counter-propagating directions, i.e., $\pm|k|$. These counter-propagating directions are given in spherical coordinates by the coordinate points (k_r, θ, φ) and $(k_r, \pi - \theta, \varphi + \pi)$, or in Cartesian coordinates by the coordinate points (k_x, k_y, k_z) and $(-k_x, -k_y, -k_z)$. Consequently, the number of independent directions in three-dimensional k -space increases with the value of $|k|$ or ν as the half spherical surface integral, equal to $2\pi\nu^2/c^2$. Then the number of modes per independent direction in k -space amounts to

$$\frac{\tilde{M}^{3D}(\nu)}{2\pi\nu^2/c^2} = \frac{4}{c} = \frac{4}{2\ell\Delta\nu_{FSR}}. \quad (11)$$

Per free spectral range $\Delta\nu_{FSR}$ and round-trip length 2ℓ , after which constructive interference occurs, we find that four modes exist. These four modes represent two counter-propagating modes per resonance frequency and polarization, for two orthogonal polarizations. This result is independent of whether free space or a standing-wave resonator is considered. In free space, this result coincides with the findings presented in Ref. [36], whereas in a standing-wave resonator, it agrees with Planck's law of blackbody radiation. Of course, this result comes as no surprise, because its foundation was laid in the integration ranges of the spherical integral in Eq. (9), by which we integrated over the whole sphere, thereby including the counter-propagating modes and counting them separately.

Now we turn our attention to the one-dimensional situation. Equivalent to Eq. (9), the one-dimensional spectral mode density per unit frequency interval of a Fabry-Pérot resonator (in Cartesian coordinate k_z) is given by

$$\tilde{M}^{1D}(\nu) = \frac{d}{d\nu} \left[\frac{2}{2\pi} \int_{k_z=-k}^{+k} dk_z \right] = \frac{d}{d\nu} \left[\frac{4\nu}{c} \right] = \frac{4}{c}. \quad (12)$$

Here one should avoid the mistake, which sometimes occurs in the literature, of integrating only from 0 to k . Since the coordinate points (k_r, θ, φ) and $(k_r, \pi - \theta, \varphi + \pi)$ in spherical coordinates equal the coordinate points (k_x, k_y, k_z) and $(-k_x, -k_y, -k_z)$ in Cartesian coordinates, i.e., the direction k_z includes positive and negative values, one has to integrate k_z from $-k$ to k . Equivalent to the three-dimensional situation of Eq. (9), also this spectral mode density must be valid for one-dimensional free space and for a one-dimensional standing-wave resonator.

Exactly one independent direction exists in one-dimensional k -space, independent of the value of $|k|$ or ν , as becomes immediately clear from the result of Eq. (12). The number of modes per independent direction in one-dimensional k -space amounts to the same value of

$$\tilde{M}^{1D}(\nu) = \frac{4}{c} = \frac{4}{2\ell\Delta\nu_{FSR}} \quad (13)$$

as in three-dimensional k -space [see Eq. (11)].

At this point, we have established that the two counter-propagating electromagnetic waves constitute independent modes in free space and in a standing-wave resonator and for one and three dimensions. Therefore, Eq. (13) can be interpreted as a formal proof that in a one-dimensional Fabry-Pérot resonator, the two counter-propagating electromagnetic waves per resonance frequency and polarization

constitute independent modes. The corresponding mode spectrum of a Fabry-Pérot resonator is sketched in Fig. 1 for the lowest-order modes with mode indices q between -4 and 4 .

Furthermore, we provide two rather heuristic arguments that support this result. Firstly, in one-dimensional free space, e.g. in an infinitely long fiber or waveguide, the forward- and backward-propagating electromagnetic waves at the same frequency and polarization constitute independent optical modes [36], despite the fact that their transverse electric-field components are identical. Siegman pointed out [37] the equivalence between a linear standing-wave resonator, having two concave mirrors and a resonator length within the resonator stability range, and an infinite periodic lensguide with the same focusing properties (Fig. 2). The similarity between such an infinite periodic lensguide and the free-space one-dimensional case of an infinitely long fiber or waveguide is immediately clear. In both cases, the counter-propagating electromagnetic waves constitute independent modes [36]. If Siegman was right in that the standing-wave resonator is equivalent to an infinite periodic lensguide, then also its two counter-propagating electromagnetic waves must constitute independent modes, in agreement with our findings above.

The second comparison also suggests that the two counter-propagating electromagnetic waves constitute independent modes. We emphasize the equivalence between a ring resonator and a linear resonator (Fig. 3). For the ring resonator [Fig. 3(a)], there is also a clear similarity with the free-space one-dimensional case of an infinitely long fiber or waveguide. Therefore, the counter-propagating electromagnetic waves in a ring resonator constitute independent modes [36]. These two modes can be described by the electric-field distributions

$$E_{\pm} e^{i(2\pi\nu_q t \pm k_q z)}, \quad (14)$$

where z denotes the forward direction that bends along the ring. Whereas a ring resonator has no mirrors and, thus, per se no out-coupling, in a quasi-ring resonator [Fig. 3(b)], one can experimentally demonstrate the existence of two counter-propagating modes, because these two modes couple out through the same semi-transparent mirror into spatially different directions.

The spectral mode density is the same for the ring and the linear resonator [Fig. 3(c)] and is given by Eq. (12). This can be easily seen by inserting a double-sided mirror into the ring resonator, such that resonant electromagnetic waves that have performed an integer number of sine cycles within one round trip now reverse their direction instead of continuing their path. No mode will vanish because of insertion of the mirror, because both counter-propagating modes reverse their direction along the ring.

We can also understand the equivalence between a linear Fabry-Pérot resonator, on the one hand, and a ring resonator or an infinite periodic lensguide, on the other hand, in the following situation. When we use an amplitude-phase diagram to display the amplitude and phase of an electromagnetic wave, we assume implicitly that the coordinate system rotates with the angular frequency of the electromagnetic wave, such that the arrow representing the electromagnetic wave does not accumulate an additional phase shift as time elapses but stands still in the diagram. Similarly, for a short pulse propagating in a linear Fabry-Pérot resonator, where the short pulse obviously occupies several longitudinal modes with the same propagation direction because of its spectral width, we can flip the coordinate system by reversing the z - and $-z$ -directions in each moment when the short pulse reaches one of the two mirrors, such that the sign of the k -vectors of its electromagnetic wave remains the same over time, thereby transforming the Fabry-Pérot resonator into a ring resonator or an infinite periodic lensguide. Moreover, if we launch simultaneously two short pulses from opposite sides into the linear Fabry-Pérot resonator, these two pulses will counter-propagate and, thus, have opposite signs of their k -vectors forever. If, again, we flip the coordinate system in each moment when the short pulses reach the opposite mirrors, the picture of the Fabry-Pérot resonator turns into a picture in which these two pulses counter-propagate along a ring resonator or an infinite periodic lensguide, thereby occupying the two counter-propagating modes [36].

The presented examples provide formal as well as heuristic evidence that the two counter-propagating electromagnetic waves in a linear Fabry-Pérot resonator represent two independent modes, equivalently to the independent counter-propagating modes in a ring resonator or an infinite periodic lensguide or an infinitely long fiber or waveguide.

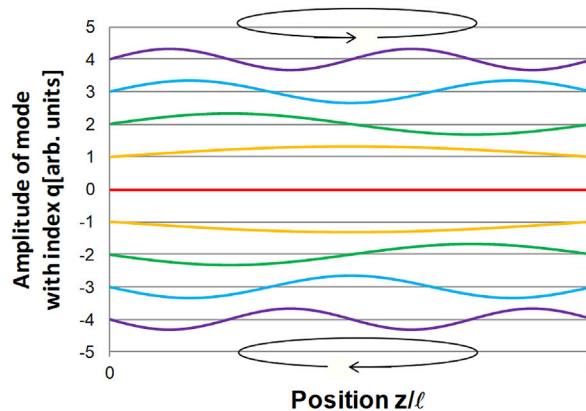


Fig. 1. Normalized electric-field distributions of the lowest-order longitudinal modes of a Fabry-Pérot resonator. The two opposite propagation directions of modes with positive and negative mode indices q are indicated. (Figure taken from Ref. [1]).

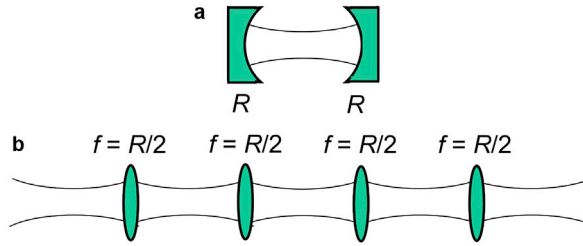


Fig. 2. (a) Two-mirror resonator with mirror radii R and (b) equivalent infinite periodic lensguide with focal lengths $f = R/2$. (Figure taken from Ref. [1]).

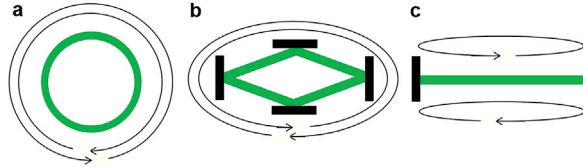


Fig. 3. Comparison of counter-propagating modes (green) in (a) ring resonator, (b) 4-mirror quasi-ring resonator, and (c) linear resonator. Propagation directions of the counter-propagating modes are indicated by arrows. (Figure taken from Ref. [1]).

2.3. Resonator losses and passive-mode properties

A schematic of the Fabry- Pérot resonator and the relevant electric fields is shown in Fig. 4. Generally, a resonator suffers from different optical losses along the light-propagation path. These losses can be of a discrete or continuous nature and usually vary with frequency ν . Discrete outcoupling losses occur due to non-perfect reflectivity of the resonator mirrors,

$$|r_i|^2 = R_i = 1 - |t_{out,i}|^2 = 1 - T_{out,i} = e^{-t_{RT}/\tau_{out,i}}. \tag{15}$$

Here, r_i and R_i are the electric-field and intensity reflectivities, respectively, and $t_{out,i}$ and $T_{out,i}$ are the electric-field and intensity transmissions, respectively, of mirror i . In the following, we assume that the mirror reflectivities and transmissions of the Fabry-Pérot resonator under investigation vary insignificantly over the frequency range of interest. The exponential decay time $\tau_{out,i}$ resulting from the outcoupling loss at mirror i , averaged over one round trip, is given by

$$\frac{1}{\tau_{out,i}} = \frac{-\ln(R_i)}{t_{RT}} = \frac{-\ln(1 - T_{out,i})}{t_{RT}}. \tag{16}$$

Siegman [14] and others named the logarithmic reflectivities, given by

$$\delta_{out,i} = -\ln(R_i) = \frac{t_{RT}}{\tau_{out,i}}, \tag{17}$$

the “mirror coupling coefficients”, referred to as the “delta notation”. In a Fabry-Pérot resonator, transmission of light through its two

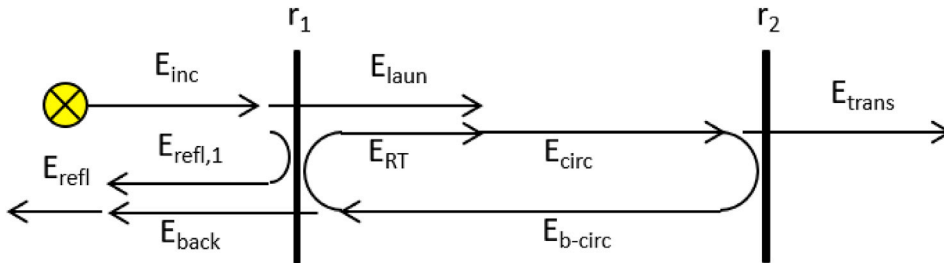


Fig. 4. Fabry-Pérot resonator with electric-field mirror reflectivities r_1 and r_2 . Indicated are the characteristic electric fields produced by an electric field E_{inc} incident upon mirror 1: $E_{refl,1}$ initially reflected at mirror 1, E_{laun} launched through mirror 1, E_{circ} and E_{b-circ} circulating inside the resonator in forward and backward propagation direction, respectively, E_{RT} propagating inside the resonator after one round trip, E_{trans} transmitted through mirror 2, E_{back} transmitted through mirror 1, and the total field E_{refl} propagating backward. Interference occurs at the left- and right-hand sides of mirror 1 between $E_{refl,1}$ and E_{back} , resulting in E_{refl} , and between E_{laun} and E_{RT} , resulting in E_{circ} , respectively. (Figure taken from Ref. [2]).

mirrors with reflectivities R_1 and R_2 yields the total outcoupling losses. These losses are quantified by the outcoupling decay-rate constant $1/\tau_{out}$:

$$R_1 R_2 = e^{-t_{RT}/\tau_{out}} \Rightarrow \frac{1}{\tau_{out}} = \frac{1}{\tau_{out,1}} + \frac{1}{\tau_{out,2}} = \frac{-\ln(R_1 R_2)}{t_{RT}}. \quad (18)$$

In terms of mirror coupling coefficients, Eq. (18) reads

$$\delta_{out} = \delta_{out,1} + \delta_{out,2} = -\ln(R_1 R_2) = \frac{t_{RT}}{\tau_{out}}. \quad (19)$$

Other individual discrete losses may arise from, e.g., diffraction losses at the mirrors of finite lateral dimensions, absorption and scattering losses in the mirror coatings, diffraction losses at the edges of an active laser medium that is shorter than the resonator length, or losses due to other optical elements inside the resonator. These losses shall not be included in our investigation.

Furthermore, continuous losses, quantified by the intensity propagation-loss coefficient α_{prop} per unit length, may be induced by the medium, e.g. scattering losses due to material imperfections, waveguide propagation losses due to interface roughness, or permanent absorption losses in the medium. In the description of lasers, the intrinsic resonator losses (which usually also include the other discrete losses, which we neglect) are commonly accounted for by the intrinsic round-trip loss L_{RT} ,

$$1 - L_{RT} = e^{-\alpha_{prop} 2\ell} = e^{-t_{RT}/\tau_{loss}}. \quad (20)$$

These losses are also assumed to vary insignificantly over the frequency range of interest. Averaged over one round trip, these losses are quantified by the intrinsic decay-rate constant $1/\tau_{loss}$:

$$\frac{1}{\tau_{loss}} = \frac{-\ln(1 - L_{RT})}{t_{RT}} = c\alpha_{prop}. \quad (21)$$

Equivalent to the mirror coupling coefficients of Eq. (17), one can introduce an intrinsic loss coefficient, equalling the logarithmic round-trip loss:

$$\delta_{loss} = -\ln(1 - L_{RT}) = \frac{t_{RT}}{\tau_{loss}}. \quad (22)$$

The decay-rate constant $1/\tau_c$ of the passive ‘‘cavity’’ or resonator, averaged over one round trip, is the sum of the above-mentioned decay-rate constants:

$$\frac{1}{\tau_c} = \frac{1}{\tau_{out}} + \frac{1}{\tau_{loss}} = \frac{-\ln[R_1 R_2 (1 - L_{RT})]}{t_{RT}} = \frac{-\ln[R_1 R_2]}{t_{RT}} + c\alpha_{prop}. \quad (23)$$

τ_c is the photon-decay time of the resonator. In terms of the coupling coefficients of Eqs. (19) and (22), Eq. (23) reads

$$\delta_c = \delta_{out} + \delta_{loss} = -\ln[R_1 R_2 (1 - L_{RT})] = \frac{t_{RT}}{\tau_c}. \quad (24)$$

According to our definition of a passive mode, we do not include the absorption loss or gain introduced by the unpumped or pumped active medium, respectively.

2.4. Photon decay

The intensity I of an electromagnetic wave is related to its electric field by

$$I = c\epsilon_0 \frac{|\vec{E}|^2}{2}, \quad (25)$$

where ϵ_0 is the vacuum permittivity. When either assuming a collimated, quasi-plane wave propagating along the resonator axis, or propagation in a simple monolithic fiber or waveguide resonator, the beam-spot size w_q of light propagating inside a mode with mode index q , centered at the resonance frequency ν_q , is constant over the resonator length ℓ . The mode volume is [11]

$$V_q = 2\pi w_q^2 \ell. \quad (26)$$

The photon number φ_q inside this mode volume is then related to the electric energy W_{stored} stored inside the mode volume, the intensity I_q , and the electric field E_q by

$$\varphi_q = \frac{2}{h\nu_q} W_{stored} = \frac{2}{ch\nu_q} \int_{V_q} I_q dV = \frac{\epsilon_0}{h\nu_q} \int_{V_q} |\vec{E}|^2 dV. \quad (27)$$

The factor of two stems from the fact that the other half of the optical energy is stored in the magnetic field.

Throughout this paper, by use of the term “photon” we will solely refer to the fundamental energy unit $h\nu$ that corresponds to the quantized energy of one photon, but we will not associate any particle or quantum-statistical properties with it. The parameter φ represents the number of these energy units called photons that is present in a classical electromagnetic field (or the expectation value of a quantum-optical coherent state, if you will). In principle, φ may assume any non-integer value. However, it is probably easy to agree upon the fact that the conservation of energy requires that an emission or absorption process by a two-level atom changes the value of φ by ± 1 energy unit $h\nu$.

The resonator can be investigated by inserting light as a short pulse or continuously. In the transient regime, the number $\varphi(t)$ of photons at frequency ν , present inside the resonator at time t , is described via the photon-decay time τ_c by the differential rate equation

$$\frac{d}{dt}\varphi(t) = -R_{decay}(t) = -\frac{1}{\tau_c}\varphi(t). \tag{28}$$

$R_{decay}(t)$ is the photon-decay rate per unit time. With a number φ_s of photons present at the starting time $t = 0$, integration of Eq. (28) delivers

$$\varphi(t) = \begin{cases} \varphi_s e^{-t/\tau_c} & t \geq 0 \\ 0 & t < 0 \end{cases}. \tag{29}$$

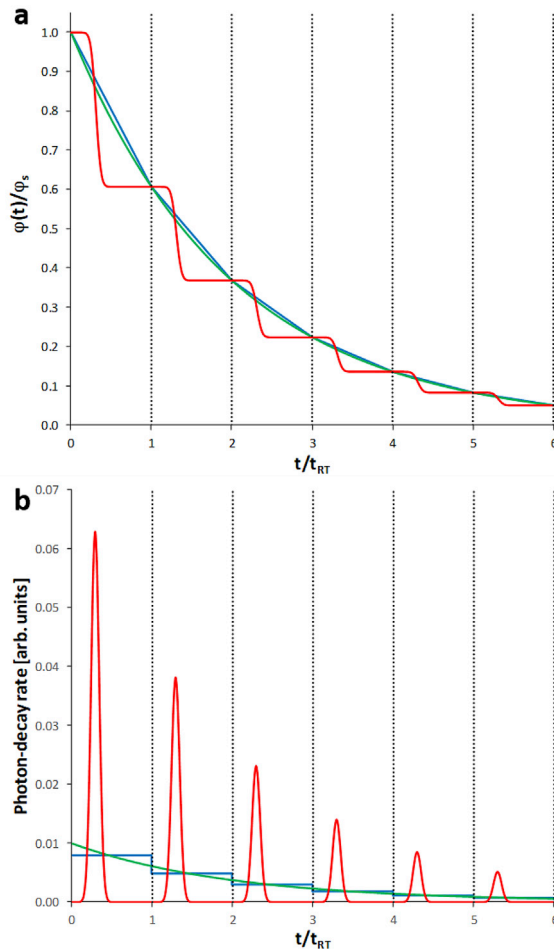


Fig. 5. (a) Relative number of photons inside the resonator as a function of time for a photon decay time of $\tau_c = 2 t_{RT}$. In case the photon decay is entirely due to intrinsic losses, the round-trip loss is $L_{RT} = 0.393$. If these losses are equally distributed within each round trip (green curve). In case the photon decay is entirely due to outcoupling losses through one mirror, the reflectivities are $R_1 = 1$ and $R_2 = 0.606$. Then the decay within one round trip depends on the distribution of light inside the resonator: decay of light homogeneously distributed over the resonator length (blue curve) and of a short Gaussian-shaped pulse with a temporal width of $t_{RT}/10$, whose peak is positioned 30% of a round trip away from the outcoupling mirror at $t = 0$ (red curve). Dotted black lines: times at which full round trips are completed. (b) Photon-decay rate as a function of time.

The change of $\varphi(t)$ per round trip is described by the integral round-trip equation

$$\varphi(t + t_{RT}) = \varphi(t)e^{-t_{RT}/\tau_c} = \varphi(t)R_1R_2(1 - L_{RT}). \quad (30)$$

The differential rate equation (28) and the integral round-trip equation (30) both average over one round-trip time t_{RT} , because the outcoupling time τ_{out} and the intrinsic decay time τ_{loss} that constitute the photon-decay time τ_c are defined in Eqs. (18) and (21), respectively, by the decay that occurs per round-trip time t_{RT} . Depending on how the light is distributed in the resonator at $t = 0$, a fraction of it is coupled out in a different temporal fashion during one round trip. However, independent of the light distribution inside the resonator, after each round trip the same fraction is coupled out and Eq. (30) is fulfilled, see Fig. 5.

In contrast, a steady-state or continuous-wave (cw) regime where the decaying photon number is continuously replenished by the incoupling of photons from an external source, such that $\varphi(t) = \varphi_s$ at all times, is described by the equation

$$R_{decay} \stackrel{cw}{=} \frac{1}{\tau_c}\varphi_s. \quad (31)$$

Consequently, the number φ_{decay} of photons lost during a time interval Δt in the transient regime,

$$\varphi_{decay}(\Delta t) \stackrel{transient}{=} \int_{t=0}^{t=\Delta t} R_{decay}(t)dt = \int_{t=0}^{t=\Delta t} \frac{1}{\tau_c}\varphi_s e^{-t/\tau_c} dt = \varphi_s(1 - e^{-\Delta t/\tau_c}) \stackrel{\Delta t=\tau_c}{=} \varphi_s(1 - e^{-1}), \quad (32)$$

and in the cw regime,

$$\varphi_{decay}(\Delta t) \stackrel{cw}{=} \int_{t=0}^{t=\Delta t} R_{decay}dt = \frac{\Delta t}{\tau_c}\varphi_s \stackrel{\Delta t=\tau_c}{=} \varphi_s, \quad (33)$$

are different. Whereas in the transient regime the photon number in the resonator decays to the fraction e^{-1} of its original value during the photon-decay time τ_c , in the cw regime in average all photons are lost during τ_c , because additionally the photons that are replenished during τ_c contribute to the decay during τ_c , see Fig. 6.

2.5. Thermal emission

Since in a steady state all photons in a resonator mode are, in average, coupled out during the photon-decay time τ_c , see Eq. (33) and Fig. 6, these photons must be replaced by coupling the same number of photons into this mode during the same time interval τ_c . In a situation, in which coupling of photons from an external light source or by spontaneous or stimulated emission from an active medium into this mode is absent, the thermal occupation of this mode by photons, therefore, equals the number of photons coupled into this mode by the thermal-emission rate R_{therm} during τ_c . As photons are bosons, the thermal occupation is given by the Bose-Einstein distribution [35]. These considerations yield

$$R_{therm}\tau_c = \frac{1}{e^{h\nu/(k_B T)} - 1}. \quad (34)$$

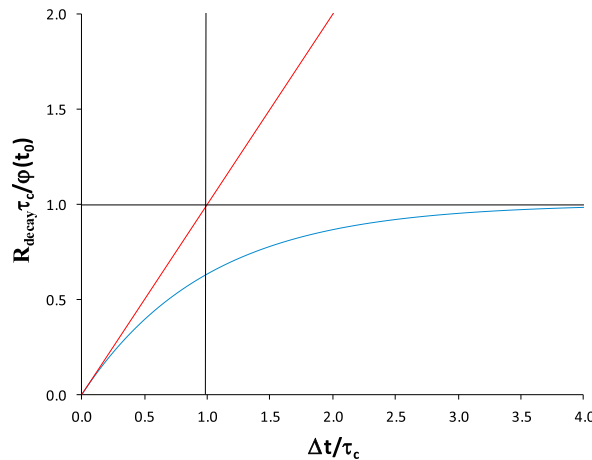


Fig. 6. Number $R_{decay}\tau_c$ of photons decaying out of the resonator, calibrated to the number $\varphi(t_0)$ of photons inside the resonator at time t_0 , as a function of time interval Δt , calibrated to the photon-decay time τ_c , in the transient regime according to Eq. (32) (blue curve) and in the cw regime according to Eq. (33) (red curve).

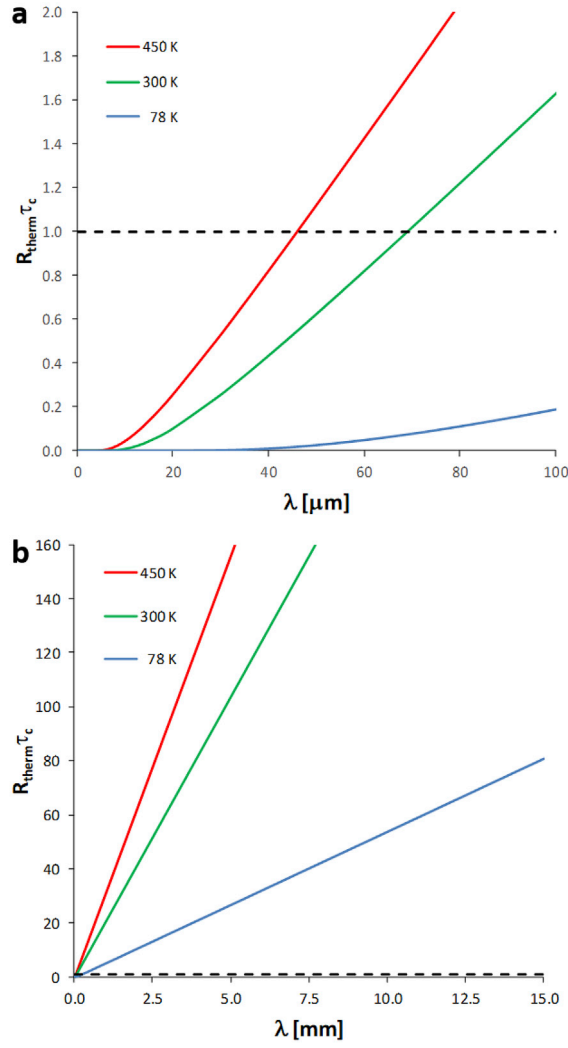


Fig. 7. Thermal-emission rate R_{therm} into the mode during τ_c as a function of wavelength at different temperatures, for wavelengths up to (a) 100 μm and (b) 15 mm. Upper limit of $R_{sp}\tau_c = 1$ of Eq. (171) in an ideal four-level laser at the fictive point where the gain would equal the losses (black dashed line).

The dependence of thermal emission on wavelength and temperature is displayed in Fig. 7. At room temperature, the thermal emission at visible and near-infrared wavelengths is extremely small compared to unity. At longer wavelengths, thermal emission into the mode increases and equals one photon per photon-decay time τ_c at a wavelength of

$$\lambda^{R_{therm}=1/\tau_c} = \frac{hc}{k_B T \ln 2} \stackrel{T=300 \text{ K}}{=} 69.2 \mu\text{m}. \tag{35}$$

2.6. Spectral line shapes

According to Eq. (29), light at frequency ν oscillating inside the resonator decays out of the resonator with a time constant of τ_c . If the resonator losses are independent of frequency, as we have assumed above, the photon-decay time τ_c is the same at all frequencies. The decaying electric field is represented by a damped harmonic oscillation with an initial amplitude of $E_{q,s}$ and a decay-time constant of $2\tau_c$. In phasor notation, it can be expressed as

$$E_q(t) = \begin{cases} E_{q,s} e^{i2\pi\nu_q t} e^{-t/(2\tau_c)} & t \geq 0 \\ 0 & t < 0 \end{cases}. \tag{36}$$

Fourier transformation of the electric field in time provides the electric field per unit frequency interval,

$$\tilde{E}_q(\nu) = \int_{-\infty}^{+\infty} E_q(t) e^{-i2\pi\nu t} dt = E_{q,s} \int_0^{+\infty} e^{-[1/(2\tau_c) + i2\pi(\nu - \nu_q)]t} dt = E_{q,s} \frac{1}{(2\tau_c)^{-1} + i2\pi(\nu - \nu_q)}. \quad (37)$$

Each mode has a normalized spectral line shape per unit frequency interval given by [6]

$$\tilde{\gamma}_q(\nu) = \frac{1}{\tau_c} \left| \frac{\tilde{E}_q(\nu)}{E_{q,s}} \right|^2 = \frac{1}{\tau_c} \left| \frac{1}{(2\tau_c)^{-1} + i2\pi(\nu - \nu_q)} \right|^2 = \frac{1}{\pi} \frac{1/(4\pi\tau_c)}{1/(4\pi\tau_c)^2 + (\nu - \nu_q)^2} \quad \text{with} \quad \int \tilde{\gamma}_q(\nu) d\nu = 1, \quad (38)$$

in units of (1/Hz). The line shape has a Lorentzian form and is centered at the resonance frequency ν_q . The normalization factor $1/\tau_c$ in Eq. (38) compensates for the broadening of the spectral line shape as the photon-decay time becomes shorter, such that the integral of the spectral line shape is normalized to unity. The spectral line shape decays to $1/2$ of its peak value at the frequencies $\nu = \nu_q \pm 1/(4\pi\tau_c)$, i.e., it has a full-width-at-half-maximum (FWHM) linewidth of

$$\Delta\nu_c = \frac{1}{2\pi\tau_c}. \quad (39)$$

Hence, Eq. (38) can be expressed in terms of either the FWHM or the half-width-at-half-maximum (HWHM) linewidth as

$$\tilde{\gamma}_q(\nu) = \frac{2}{\pi} \frac{\Delta\nu_c}{\Delta\nu_c^2 + 4(\nu - \nu_q)^2} = \frac{1}{\pi} \frac{\Delta\nu_c/2}{(\Delta\nu_c/2)^2 + (\nu - \nu_q)^2} \quad \text{with} \quad \int \tilde{\gamma}_q(\nu) d\nu = 1. \quad (40)$$

Compared to a resonator which comprises only outcoupling losses, resulting in a linewidth of $\Delta\nu_c = \Delta\nu_{out} = 1/(2\pi\tau_{out})$, for $\alpha_{prop} > 0 \text{ cm}^{-1}$ the linewidth increases by $\Delta\nu_{loss} = 1/(2\pi\tau_{loss})$ to

$$\Delta\nu_c = \Delta\nu_{out} + \Delta\nu_{loss}. \quad (41)$$

Simultaneously, its peak value at the resonance frequency ν_q decreases, such that the integral of the spectral line shape remains calibrated to unity. Calibrated to a peak height of unity, we obtain the Lorentzian lines, in units of (1):

$$\gamma_{q,L}(\nu) = \frac{\pi}{2} \Delta\nu_c \tilde{\gamma}_q(\nu) = \frac{(\Delta\nu_c)^2}{(\Delta\nu_c)^2 + 4(\nu - \nu_q)^2} = \frac{(\Delta\nu_c/2)^2}{(\Delta\nu_c/2)^2 + (\nu - \nu_q)^2} \quad \text{with} \quad \gamma_{q,L}(\nu_q) = 1. \quad (42)$$

When repeating the above Fourier transformation for all the modes with mode index q in the resonator, one obtains the full mode spectrum of the resonator. White light launched into or generated inside the resonator occupies all resonator modes. It is stored inside and emitted from the mode at ν_q with this spectral line shape $\tilde{\gamma}_c$ and linewidth $\Delta\nu_c$.

2.7. Equivalent spectral-coherence parameters

Besides the exponential photon-decay time τ_c and the FWHM linewidth $\Delta\nu_c$ of the Lorentzian spectral line shape, there are three additional parameters that equally describe the spectral coherence of a passive resonator mode. The quality of an optical resonator is expressed by its intrinsic Q -factor [38], originally introduced as the ‘‘coil dissipation constant’’ of a resonant electric circuit [39] and later generalized [40] as the energy stored in the resonator, W_{stored} , divided by the energy lost per oscillation cycle, W_{lost} . More precisely, it should be defined as the quality factor Q_c of a single passive mode resonant at ν_q :

$$Q_c := 2\pi \frac{W_{stored}(t)}{W_{lost}(t)} = 2\pi \frac{\varphi(t)}{\frac{1}{\nu_q} \frac{d}{dt} \varphi(t)} = 2\pi\nu_q\tau_c = \frac{\nu_q}{\Delta\nu_c}. \quad (43)$$

Here, the energy of the electromagnetic field is related to the photon number as $W = h\nu_q\varphi$. Because of the Fourier transformation, resulting in the FWHM linewidth $\Delta\nu_c$ of Eq. (39), the Q -factor is equivalently defined via the FWHM linewidth $\Delta\nu_c$.

Furthermore, the complex degree of temporal coherence of the electric field emitted from the resonator at ν_q can be derived from the normalized first-order correlation function [41–45]

$$g^{(1)}(\vec{r}_1, t_1, \vec{r}_2, t_2) = \frac{\langle E(\vec{r}_1, t_1) E^*(\vec{r}_2, t_2) \rangle}{[\langle |E(\vec{r}_1, t_1)|^2 \rangle \langle |E(\vec{r}_2, t_2)|^2 \rangle]^{1/2}} = \frac{\langle E(\vec{r}_1, t_1) E^*(\vec{r}_2, t_1 + \Delta t) \rangle}{[\langle |E(\vec{r}_1, t_1)|^2 \rangle \langle |E(\vec{r}_2, t_1 + \Delta t)|^2 \rangle]^{1/2}}. \quad (44)$$

Since there exist misinterpretations as to what the parameters in Eq. (44) mean and how this general equation is related to similar equations that appear in the literature, we precise here their meanings and relations. In the general equation (44), t_1 and t_2 are the retarded times at which the two parts of an electric field start at the positions r_1 and r_2 , respectively, in order to arrive via different beam trajectories at the observation point r_0 simultaneously at the time t_0 . Then $\tau_1 = t_0 - t_1$ and $\tau_2 = t_0 - t_2$ are the time intervals required to reach the observation point, and $\Delta t = t_2 - t_1$ is the difference in retarded starting times. The calibrations $\langle |E(\vec{r}_1, t_1)|^2 \rangle$ and $\langle |E(\vec{r}_2, t_2)|^2 \rangle$ of the two parts of the electric field are executed at the starting positions r_1 and r_2 at the retarded starting times t_1 and t_2 , respectively. In

a specific situation under investigation, the choice of observation point r_0 and time t_0 determines t_1, t_2, τ_1, τ_2 , and Δt . Examples for the application of Eq. (44) are Young's early interference experiments [46] or the double-slit experiment that has been ascribed to him. The electric field may well consist of a frequency spectrum. A reduced form of Eq. (44) [47],

$$g^{(1)}\left(\vec{r}_1, t_1, \vec{r}_2, t_2\right) = \frac{\langle E(t_1)E^*(t_2) \rangle}{[\langle |E(t_1)|^2 \rangle \langle |E(t_2)|^2 \rangle]^{1/2}} = \frac{\langle E(t_1)E^*(t_1 + \Delta t) \rangle}{[\langle |E(t_1)|^2 \rangle \langle |E(t_1 + \Delta t)|^2 \rangle]^{1/2}}, \tag{45}$$

describes situations where the two parts of the electric field start from the same position, $r = r_1 = r_2$, but follow different beam trajectories and arrive at the observation point r_0 simultaneously at the time t_0 . Examples are Michelson and Mach-Zhender interferometers. Also in this case, $\tau_1 = t_0 - t_1$ and $\tau_2 = t_0 - t_2$ are the time intervals required to reach the observation point, $\Delta t = t_2 - t_1$ is the difference in retarded starting times, and the calibrations are executed at the starting position $r = r_1 = r_2$ at the retarded starting times t_1 and $t_2 = t_1 + \Delta t$.

When analyzing the complex degree of temporal coherence of a single electric field with a single beam trajectory, starting at the retarded time $t = t_1 = t_2$ at the position $r = r_1 = r_2$ and travelling time interval $\tau = \tau_1 = t_0 - t_1 = \tau_2 = t_0 - t_2$ to reach the observation point r_0 at the observation time t_0 , the difference in retarded starting times is $\Delta t = t_2 - t_1 = 0$, and the two calibrations are executed at the same position $r = r_1 = r_2$ at the same retarded time $t = t_1 = t_2$. Both Eqs. (44) and (45) then reduce to a form which is often displayed in the literature as

$$g^{(1)}(\tau) = \frac{\langle E(t)E^*(t + \tau) \rangle}{\langle |E(t)|^2 \rangle} = \frac{\int_{-\infty}^{+\infty} E(t)E^*(t + \tau) dt}{\int_{-\infty}^{+\infty} E(t)E^*(t) dt}. \tag{46}$$

According to Eqs. (36) and (29), respectively, the electric-field amplitude and intensity of a wave packet emitted from the resonator decay exponentially with time, resulting in the Lorentzian spectral line shape of Eq. (38). While the emitted electric field propagates away from the resonator, its different frequencies run out of phase. Consequently, their correlation $g^{(1)}$ is a decreasing function of the travel time τ and, therefore, the temporal coherence of emitted light decreases with elapsed time τ and distance from the resonator. Inserting Eq. (36) into Eq. (46), with the integrals starting at $t = 0$, because $E(t) = 0$ for $t < 0$, yields

$$\langle E(t)E^*(t + \tau) \rangle = \int_0^{+\infty} E(t)E^*(t + \tau) dt = \tau_c E_{q,s}^2 e^{-i2\pi\nu_q\tau} e^{-\tau/(2\tau_c)} \tag{47}$$

for the cross-spectral density given by the integral in the numerator and

$$\langle |E(t)|^2 \rangle = \int_0^{+\infty} E(t)E^*(t) dt = \tau_c E_{q,s}^2 \tag{48}$$

for the power-spectral density given by the integral in the denominator, resulting in

$$g^{(1)}(\tau) = e^{-i2\pi\nu_q\tau} e^{-\tau/(2\tau_c)}. \tag{49}$$

Since generally $0 \leq |g^{(1)}|^2 \leq 1$, with the two extremes representing complete non-correlation and complete correlation, respectively, the exponentially decaying field of Eq. (36) is partially correlated. Its coherence time is given by

$$\tau_c^{coh} = \int_{-\infty}^{+\infty} |g^{(1)}(\tau)|^2 d\tau = \int_0^{+\infty} |e^{-i2\pi\nu_q\tau} e^{-\tau/(2\tau_c)}|^2 d\tau = \int_0^{+\infty} e^{-\tau/(2\tau_c)} d\tau = 2\tau_c, \tag{50}$$

and the coherence length becomes [48]

$$\ell_c^{coh} = c\tau_c^{coh} = 2c\tau_c. \tag{51}$$

The relationship between photon-decay time τ_c and coherence time τ_c^{coh} is illustrated in Fig. 8.

Generally, the light stored inside and emitted from the resonator depends spectrally and temporally not only on the resonator properties, but also on the exciting light source. For example, if light from an ultranarrow-linewidth laser is launched into the resonator, the light transmitted has practically the same center frequency and linewidth as the laser. If the launched light is somewhat more broadband, the transmitted light is a convolution of the spectral dependence of the launched light and that of the resonator. If the light source is white, the transmitted light has the same spectral dependence as the resonator.

The five reference parameters $\tau_c, \Delta\nu_c, Q_c, \tau_c^{coh}$, and ℓ_c^{coh} defined in Eqs. (23), (39), (43), (50), and (51) equivalently quantify the first-order spectral-coherence properties of the passive mode at ν_q . Changing one of these parameters simultaneously changes all of them. Important – but unfortunately sometimes misunderstood – is the fact that these parameters characterize the passive mode independently of how it is excited by light, from outside or inside and cw or transiently, because these five parameters depend only on the resonator losses. Particularly relevant for our investigation is the situation in which a mode is excited by a broadband source with a cw rate R_{launch} per unit time and the light stored inside the resonator reaches a steady state. In this situation, light still decays out of the

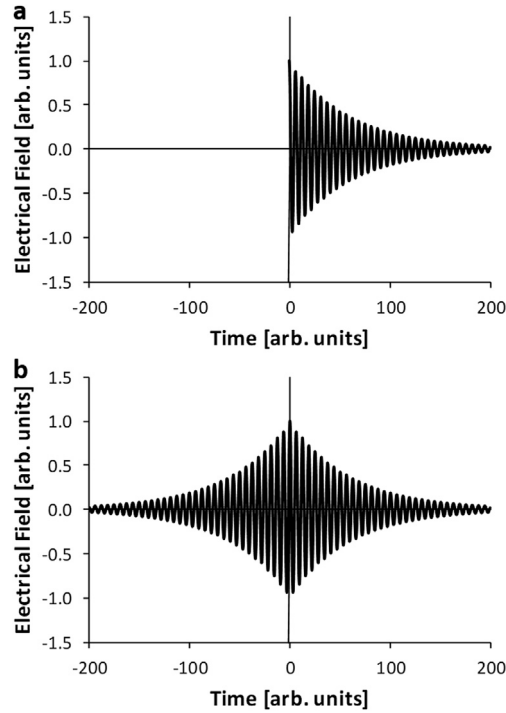


Fig. 8. Comparison of the mathematical definitions of a) photon decay time τ_c and b) coherence time τ_c^{coh} . (Figure taken from Ref. [6]).

resonator with the same photon-decay time τ_c , while being continuously replenished:

$$\frac{d}{dt}\varphi = R_{\text{launch}} - \frac{1}{\tau_c}\varphi = 0. \quad (52)$$

The definitions of τ_c , $\Delta\nu_c$, Q_c , τ_c^{coh} , and ℓ_c^{coh} remain perfectly valid and their values remain exactly the same. No Q -factor becomes infinite and no linewidth vanishes, if the number φ of photons inside the passive mode at ν_q is constant in time in a steady-state situation.

3. Airy distributions of the Fabry-Pérot resonator

In this Section, we demystify the Airy distribution. We derive the generic Airy distribution of a Fabry-Pérot resonator, which equals the internal resonance enhancement factor. All related Airy distributions are obtained by simple intensity scaling factors. The same results for a resonator without intrinsic propagation losses have previously been published [2]. We define the longitudinal mode profiles, which are Lorentzian-shaped in case of frequency-independent losses, and verify that the sum of the mode profiles of all longitudinal modes generates the respective Airy distribution.

3.1. Generic Airy distribution: the internal resonance enhancement factor

The response of the Fabry-Pérot resonator is most easily derived by use of the circulating-field approach [15], as displayed in Fig. 4. This approach assumes a steady state and derives the Airy distributions via the electric field E_{circ} circulating inside the resonator. In fact, E_{circ} is the field propagating in the forward direction from mirror 1 to mirror 2 after interference between the field E_{RT} that is circulating after one round trip, i.e., after having suffered intrinsic round-trip losses and outcoupling losses at both mirrors, and the field E_{laun} launched through the first mirror.

With the phase shift 2ϕ of Eq. (3) accumulated in one round trip, the outcoupling losses as defined in Eq. (15), and the propagation losses as defined in Eq. (20), where the amplitude propagation-loss coefficient is $\alpha_{\text{prop}}/2$, the field E_{circ} can be related to the field E_{laun} that is launched into the resonator in the situation of Fig. 4 by

$$E_{\text{circ}} = E_{\text{laun}} + E_{RT} = E_{\text{laun}} + r_1 r_2 e^{-(\alpha_{\text{prop}}/2)2\ell} e^{-i2\phi} E_{\text{circ}} \Rightarrow \frac{E_{\text{circ}}}{E_{\text{laun}}} = \frac{1}{1 - r_1 r_2 e^{-\alpha_{\text{prop}}\ell} e^{-i2\phi}}. \quad (53)$$

Exploiting the identities

$$|e^{-i\phi}|^2 = |\cos(\phi) - i \sin(\phi)|^2 = \cos^2(\phi) + \sin^2(\phi) = 1, \quad \cos(2\phi) = 1 - 2\sin^2(\phi) \quad (54)$$

and Eq. (15) yields

$$\begin{aligned} |1 - r_1 r_2 e^{-\alpha_{prop} \ell} e^{-i2\phi}|^2 &= |1 - r_1 r_2 e^{-\alpha_{prop} \ell} \cos(2\phi) + i r_1 r_2 e^{-\alpha_{prop} \ell} \sin(2\phi)|^2 \\ &= [1 - r_1 r_2 e^{-\alpha_{prop} \ell} \cos(2\phi)]^2 + r_1^2 r_2^2 e^{-2\alpha_{prop} \ell} \sin^2(2\phi) \\ &= 1 + R_1 R_2 e^{-\alpha_{prop} 2\ell} - 2\sqrt{R_1 R_2} e^{-\alpha_{prop} \ell} \cos(2\phi) \\ &= (1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell})^2 + 4\sqrt{R_1 R_2} e^{-\alpha_{prop} \ell} \sin^2(\phi). \end{aligned} \quad (55)$$

The generic Airy distribution, which considers solely the physical processes exhibited by light inside the resonator, then derives as the intensity circulating in the resonator relative to the intensity launched,

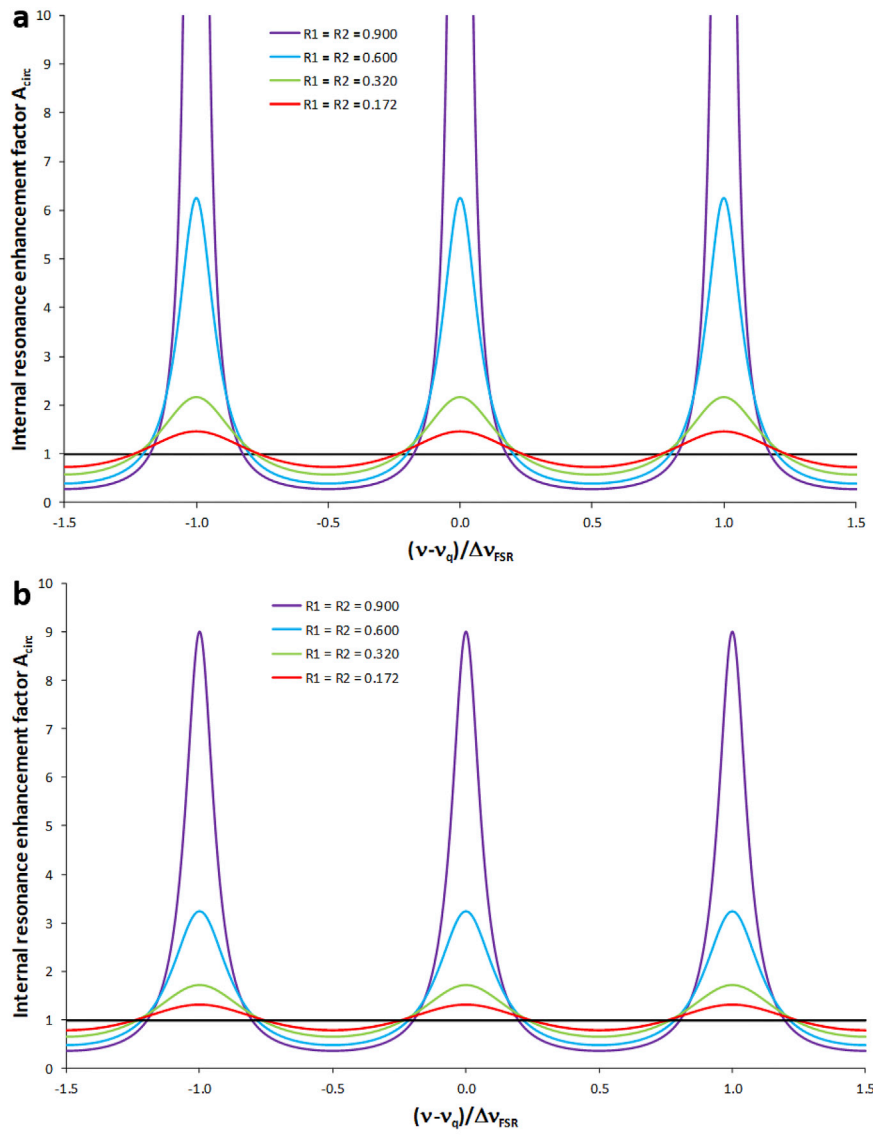


Fig. 9. Generic Airy distribution A_{circ} of Eq. (56), equaling the spectrally dependent internal resonance enhancement which the resonator provides to light that is launched into it, for different mirror reflectivities of $R_1 = R_2$ and an intensity propagation loss of (a) $\alpha_{prop} \ell = 0$ and (b) $\alpha_{prop} \ell = 0.3$. For the curve with $R_1 = R_2 = 0.9$ and $\alpha_{prop} \ell = 0$, the peak value is at $A_{circ}(\nu_q) = 100$, outside the scale of the ordinate. (Figure (a) taken from Ref. [2]).

$$A_{circ} = \frac{I_{circ}}{I_{laun}} = \frac{|E_{circ}|^2}{|E_{laun}|^2} = \frac{1}{|1 - r_1 r_2 e^{-\alpha_{prop} \ell} e^{-i2\phi}|^2} = \frac{1}{(1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell})^2 + 4\sqrt{R_1 R_2} e^{-\alpha_{prop} \ell} \sin^2(\phi)}. \tag{56}$$

Physically, A_{circ} represents the spectrally dependent internal resonance enhancement which the resonator provides to the light launched into it [49]. It is displayed for different mirror reflectivities and propagation-loss coefficients in Fig. 9. At the resonance frequencies ν_q , where $\sin(\phi)$ equals zero, the internal resonance enhancement factor is

$$A_{circ}(\nu_q) = \frac{1}{(1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell})^2}. \tag{57}$$

The integral of the internal resonant enhancement factor A_{circ} of Eq. (56) over one free spectral range, i.e., 2ϕ in the interval $[-\pi, \dots, \pi]$ or ϕ in the interval $[-\pi/2, \dots, \pi/2]$, can be written as

$$A_{circ,FSR} = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} A_{circ} d\phi = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{(1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell})^2 + 4\sqrt{R_1 R_2} e^{-\alpha_{prop} \ell} \sin^2(\phi)} d\phi = a \int_{-\pi/2}^{\pi/2} \frac{1}{b + \sin^2(\phi)} d\phi, \tag{58}$$

where a and b are given by

$$a = \frac{1}{4\pi\sqrt{R_1 R_2} e^{-\alpha_{prop} \ell}}, \quad b = \frac{(1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell})^2}{4\sqrt{R_1 R_2} e^{-\alpha_{prop} \ell}}, \tag{59}$$

respectively. This type of integral is solved by making the substitution $\phi = \arctan(x) \Rightarrow x = \tan(\phi)$ and expressing $\sin(\phi)$ and $d\phi$ in terms of x as follows:

$$\begin{aligned} \sin(\phi) &= \frac{\tan(\phi)}{\sqrt{1 + \tan^2(\phi)}} = \frac{x}{\sqrt{1 + x^2}}, \quad \cos(\phi) = \frac{1}{\sqrt{1 + \tan^2(\phi)}} = \frac{1}{\sqrt{1 + x^2}} \\ \Rightarrow \frac{dx}{d\phi} &= \frac{d \tan(\phi)}{d\phi} = \frac{1}{\cos^2(\phi)} = 1 + x^2 \quad \Rightarrow \quad d\phi = \frac{dx}{1 + x^2}. \end{aligned} \tag{60}$$

The integral can be written in a simpler form as

$$\begin{aligned} A_{circ,FSR} &= a \int_{-\pi/2}^{\pi/2} \frac{1}{b + \sin^2(\phi)} d\phi = a \int_{-\tan(\pi/2)}^{\tan(\pi/2)} \frac{1}{b + x^2/(1 + x^2)} \frac{1}{1 + x^2} dx \\ &= \frac{a}{(b + 1)} \int_{-\tan(\pi/2)}^{\tan(\pi/2)} \frac{1}{b/(b + 1) + x^2} dx = \frac{a/(b + 1)}{\sqrt{b/(b + 1)}} \left[\arctan\left(\frac{x}{\sqrt{b/(b + 1)}}\right) \right]_{-\tan(\pi/2)}^{\tan(\pi/2)} \\ &= \frac{a}{(b + 1)} \frac{\pi}{\sqrt{b/(b + 1)}}. \end{aligned} \tag{61}$$

By inserting Eq. (59), the internal resonance enhancement factor A_{circ} integrated over one free spectral range is

$$A_{circ,FSR} = \frac{1}{\pi} \int_{-\pi/2}^{+\pi/2} A_{circ} d\phi = \frac{1}{1 - R_1 R_2 e^{-\alpha_{prop} 2\ell}} \stackrel{\alpha_{prop}=0}{=} \frac{1}{1 - R_1 R_2}. \tag{62}$$

The integral resonance enhancement $A_{circ,FSR}$ is displayed as a function of $R_1 R_2$ in Fig. 10(a). For $\alpha_{prop} = 0 \text{ cm}^{-1}$, it increases from unity at zero reflectivity, because all light is simply transmitted, to infinite enhancement when the reflectivity equals unity. For increasing propagation losses, the integral resonance enhancement increases less with increasing mirror reflectivities.

As can be seen in Fig. 9, the fraction of the free spectral range over which enhancement of light occurs, narrows with increasing reflectivities. From Eq. (56), this fraction derives as

$$A_{circ} \geq 1 \Rightarrow \sin(\phi) \leq \frac{1}{2} \sqrt{2 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell}} \Rightarrow \frac{\Delta\nu_{A_{circ} \geq 1}}{\Delta\nu_{FSR}} = \frac{2}{\pi} \arcsin\left(\frac{1}{2} \sqrt{2 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell}}\right). \tag{63}$$

The fraction is displayed in Fig. 10(b). For $\alpha_{prop} = 0 \text{ cm}^{-1}$, the spectral range of enhancement decreases from $\Delta\nu_{FSR}/2$ for $R_1 R_2 = 0$ to $\Delta\nu_{FSR}/3$ for $R_1 R_2 \rightarrow 1$. With increasing propagation losses, the spectral range of enhancement decreases less.

3.2. Other airy distributions

Once the internal resonance enhancement, the generic Airy distribution of Eq. (56), is established, all other Airy distributions, i.e.,

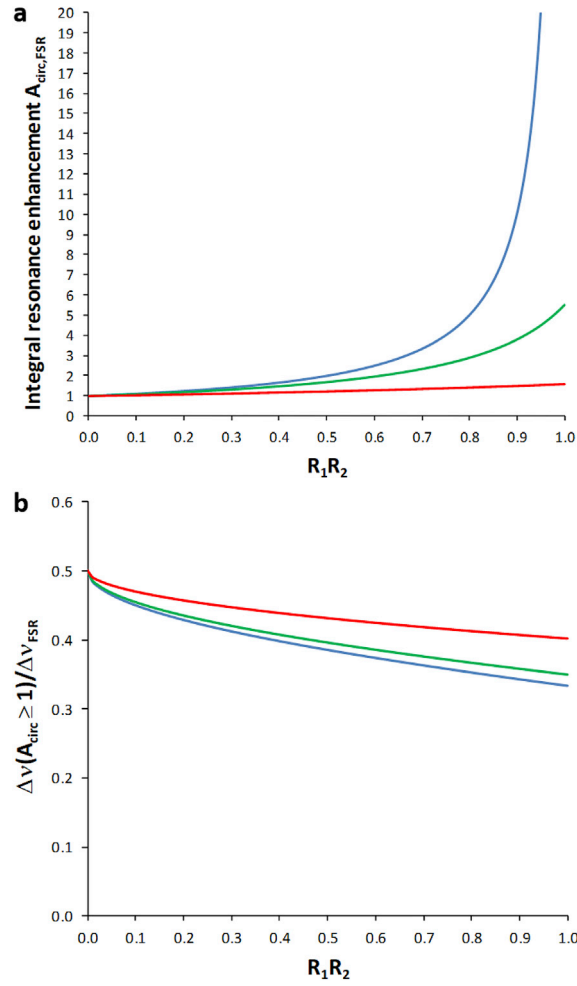


Fig. 10. (a) Integral resonance enhancement $A_{\text{circ,FSR}}$ (integrated over one free spectral range) and (b) fraction of the free spectral range over which enhancement occurs, i.e., $A_{\text{circ}} \geq 1$, as a function of the product of mirror reflectivities, $R_1 R_2$, for propagation losses of $\alpha_{\text{prop}} \ell = 0$ (blue lines), 0.1 (green lines), and 0.5 (red lines) (Figure modified from Ref. [2]).

the observed light intensities relative to the launched or initial intensity from the light source (Fig. 4), can then be straight-forwardly deduced by simple scaling factors, because no additional interference occurs (the only exception being the total intensity I_{refl} propagating backward, because an additional interference occurs between the fields $E_{\text{refl},1}$ and E_{back}). Since the intensity I_{laun} launched into the resonator equals the transmitted fraction of the intensity I_{inc} incident upon mirror 1,

$$I_{\text{laun}} = (1 - R_1) I_{\text{inc}}, \quad (64)$$

and the intensities transmitted through mirror 2, reflected at mirror 2, transmitted through mirror 1, and reflected at mirror 1 are the transmitted and reflected fractions of the intensity circulating or back-circulating inside the resonator and propagated over one resonator length,

$$\begin{aligned} I_{\text{trans}} &= (1 - R_2) e^{-\alpha_{\text{prop}} \ell} I_{\text{circ}} \\ I_{\text{b-circ}} &= R_2 e^{-\alpha_{\text{prop}} \ell} I_{\text{circ}} \\ I_{\text{back}} &= (1 - R_1) e^{-\alpha_{\text{prop}} \ell} I_{\text{b-circ}} \\ I_{\text{RT}} &= R_1 e^{-\alpha_{\text{prop}} \ell} I_{\text{b-circ}}, \end{aligned} \quad (65)$$

respectively, one obtains the following Airy distributions and intensity-scaling factors f :

$$A_{\text{b-circ}} = \frac{I_{\text{b-circ}}}{I_{\text{laun}}} = f_{\text{b-circ}} A_{\text{circ}} = R_2 e^{-\alpha_{\text{prop}} \ell} A_{\text{circ}}, \quad (66)$$

$$A_{RT} = \frac{I_{RT}}{I_{laur}} = f_{RT} A_{circ} = R_1 R_2 e^{-\alpha_{prop} 2\ell} A_{circ}, \tag{67}$$

$$A_{trans} = \frac{I_{trans}}{I_{laur}} = f_{trans} A_{circ} = (1 - R_2) e^{-\alpha_{prop} \ell} A_{circ}, \tag{68}$$

$$A_{back} = \frac{I_{back}}{I_{laur}} = f_{back} A_{circ} = (1 - R_1) R_2 e^{-\alpha_{prop} 2\ell} A_{circ}, \tag{69}$$

$$A_{emit} = A_{trans} + A_{back} = \frac{I_{trans} + I_{back}}{I_{laur}} = f_{emit} A_{circ} = [(1 - R_2) + (1 - R_1) R_2 e^{-\alpha_{prop} \ell}] e^{-\alpha_{prop} \ell} A_{circ}, \tag{70}$$

$$A'_{circ} = \frac{I_{circ}}{I_{inc}} = f_{inc} A_{circ} = (1 - R_1) A_{circ}, \tag{71}$$

$$A'_{b-circ} = \frac{I_{b-circ}}{I_{inc}} = f_{inc} f_{b-circ} A_{circ} = (1 - R_1) R_2 e^{-\alpha_{prop} \ell} A_{circ}, \tag{72}$$

$$A'_{RT} = \frac{I_{RT}}{I_{inc}} = f_{inc} f_{RT} A_{circ} = (1 - R_1) R_1 R_2 e^{-\alpha_{prop} 2\ell} A_{circ}, \tag{73}$$

$$A'_{trans} = \frac{I_{trans}}{I_{inc}} = f_{inc} f_{trans} A_{circ} = (1 - R_1) (1 - R_2) e^{-\alpha_{prop} \ell} A_{circ}, \tag{74}$$

$$A'_{back} = \frac{I_{back}}{I_{inc}} = f_{inc} f_{back} A_{circ} = (1 - R_1)^2 R_2 e^{-\alpha_{prop} 2\ell} A_{circ}, \tag{75}$$

$$A'_{emit} = A'_{trans} + A'_{back} = \frac{I_{trans} + I_{back}}{I_{inc}} = f_{inc} f_{emit} A_{circ} = (1 - R_1) [(1 - R_2) + (1 - R_1) R_2 e^{-\alpha_{prop} \ell}] e^{-\alpha_{prop} \ell} A_{circ}. \tag{76}$$

The index “emit” denotes Airy distributions that consider the sum of intensities emitted on both sides of the resonator,

$$A_{emit} = A_{trans} + A_{back}, \quad A'_{emit} = A'_{trans} + A'_{back}. \tag{77}$$

The prime denotes Airy distributions with respect to the incident intensity I_{inc} . The connection of the Airy distributions via the intensity-scaling factors can be summarized as follows:

$$\begin{aligned} A_{circ} &= \frac{1}{(1 - R_2) e^{-\alpha_{prop} \ell}} A_{trans} = \frac{1}{R_2 e^{-\alpha_{prop} \ell}} A_{b-circ} = \frac{1}{(1 - R_1) R_2 e^{-\alpha_{prop} 2\ell}} A_{back} \\ &= \frac{1}{R_1 R_2 e^{-\alpha_{prop} 2\ell}} A_{RT} = \frac{1}{[(1 - R_2) + (1 - R_1) R_2 e^{-\alpha_{prop} \ell}] e^{-\alpha_{prop} \ell}} A_{emit} = \frac{1}{1 - R_1} A'_{circ} \\ A'_{circ} &= \frac{1}{(1 - R_2) e^{-\alpha_{prop} \ell}} A'_{trans} = \frac{1}{R_2 e^{-\alpha_{prop} \ell}} A'_{b-circ} = \frac{1}{(1 - R_1) R_2 e^{-\alpha_{prop} 2\ell}} A'_{back} \\ &= \frac{1}{R_1 R_2 e^{-\alpha_{prop} 2\ell}} A'_{RT} = \frac{1}{[(1 - R_2) + (1 - R_1) R_2 e^{-\alpha_{prop} \ell}] e^{-\alpha_{prop} \ell}} A'_{emit}. \end{aligned} \tag{78}$$

As long as ℓ , n_r , R_1 , R_2 , and α_{prop} vary insignificantly over the frequency range of interest, the intensity-scaling factors that connect the individual Airy distributions with each other are frequency independent. Then the free-spectral-range integrals of all these Airy distributions are connected by the same intensity-scaling factors. For example,

$$\frac{A_{emit.FSR}}{A_{circ.FSR}} = \frac{A_{emit}}{A_{circ}} = f_{emit} = [(1 - R_2) + (1 - R_1) R_2 e^{-\alpha_{prop} \ell}] e^{-\alpha_{prop} \ell} \stackrel{\alpha_{prop}=0}{=} 1 - R_1 R_2. \tag{79}$$

With Eq. (62), this results in

$$A_{emit.FSR} = f_{emit} A_{circ.FSR} = \frac{[(1 - R_2) + (1 - R_1) R_2 e^{-\alpha_{prop} \ell}] e^{-\alpha_{prop} \ell}}{1 - R_1 R_2 e^{-\alpha_{prop} 2\ell}} \stackrel{\alpha_{prop}=0}{=} 1. \tag{80}$$

A_{emit} is the Airy distribution whose free-spectral-range integral equals unity in the absence of propagation loss. The reason is that the cosine function in the interference Eq. (123) averages to zero over all phase angles, hence in average over one free spectral range the interferences at mirror 1 disappear and the light intensity launched into the resonator also exits the resonator. However, caution should

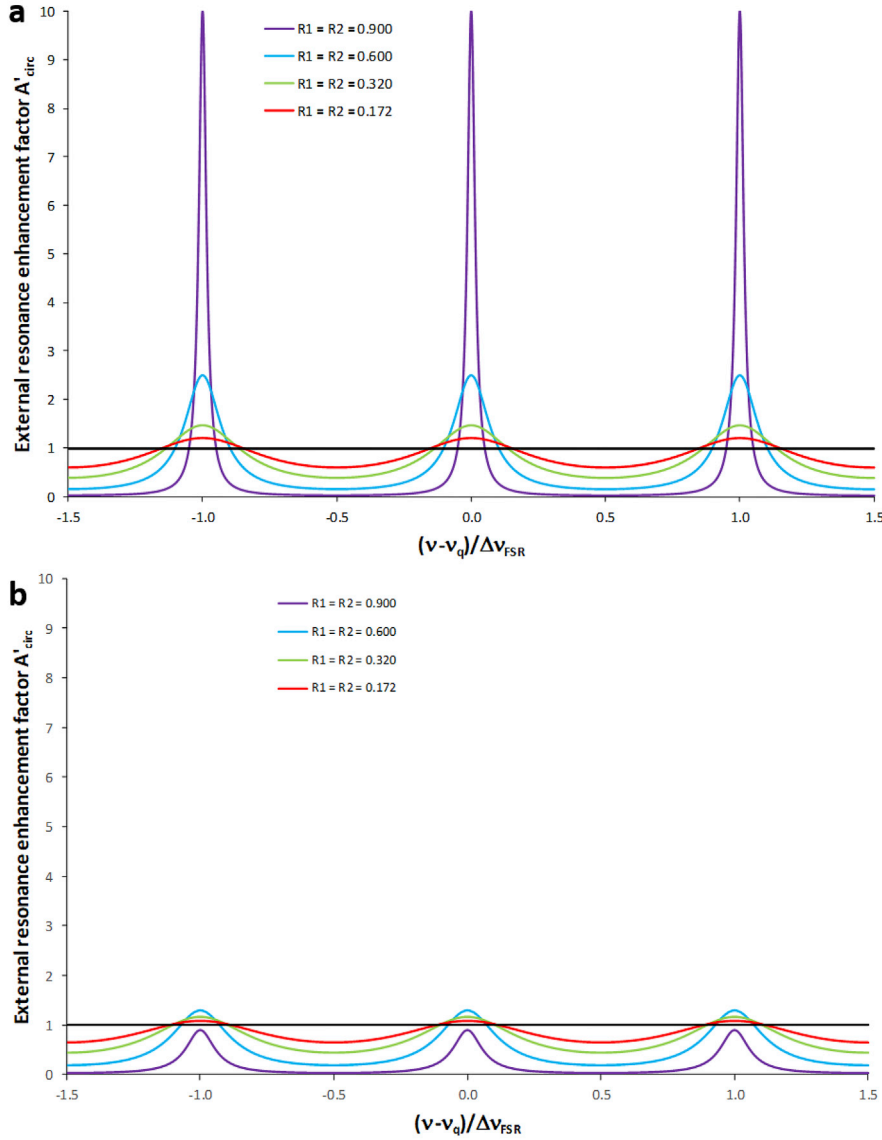


Fig. 11. Airy distribution A'_{circ} of Eq. (71), equaling the spectrally dependent external resonance enhancement which the resonator provides to light that is incident to it, for different mirror reflectivities of $R_1 = R_2$ and an intensity propagation loss of (a) $\alpha_{prop\ell} = 0$ and (b) $\alpha_{prop\ell} = 0.1$.

be exerted not to misinterpret this result as an expression of energy conservation. Otherwise, one could proof energy conservation wrong by launching a single wavelength. The integrals of all Airy distributions decrease with increasing propagation loss.

Like A_{circ} represents the internal resonance enhancement factor with respect to I_{laun} , A'_{circ} of Eq. (71) represents the external resonance enhancement factor with respect to I_{inc} (Fig. 11). The external resonance enhancement factor at the resonance frequencies is

$$A'_{circ}(\nu_q) = f_{inc} A_{circ}(\nu_q) = (1 - R_1) A_{circ}(\nu_q) = \frac{1 - R_1}{(1 - \sqrt{R_1 R_2} e^{-\alpha_{prop\ell}})^2}. \quad (81)$$

Integrated over one free spectral range, it amounts to

$$A'_{circ,FSR} = f_{inc} A_{circ,FSR} = (1 - R_1) A_{circ,FSR} = \frac{1 - R_1}{1 - R_1 R_2 e^{-\alpha_{prop} 2\ell}}. \quad (82)$$

I.e., the distributions A'_{circ} , $A'_{circ}(\nu_q)$, and $A'_{circ,FSR}$ related to the incident light are smaller by the same factor $1 - R_1$ than the same distributions related to the launched light, because only this fraction $1 - R_1$ of the light incident on mirror 1 is transmitted through mirror 1

and launched into the resonator, see Eq. (64).

3.3. Airy distribution of transmitted light

In experimental situations, often light is transmitted through a Fabry-Pérot resonator in order to characterize the resonator or to use it as a scanning interferometer (see later in Section 4). Therefore, an often applied Airy distribution is A'_{trans} of Eq. (74):

$$A'_{trans} = \frac{I_{trans}}{I_{inc}} = f_{inc} f_{trans} A_{circ} = (1 - R_1)(1 - R_2)e^{-\alpha_{prop}\ell} A_{circ} = \frac{(1 - R_1)(1 - R_2)e^{-\alpha_{prop}\ell}}{(1 - \sqrt{R_1 R_2} e^{-\alpha_{prop}\ell})^2 + 4\sqrt{R_1 R_2} e^{-\alpha_{prop}\ell} \sin^2(\phi)}. \quad (83)$$

It describes the fraction I_{trans} of the intensity I_{inc} of a light source incident upon mirror 1 that is transmitted through mirror 2, see Fig. 4. A'_{trans} is displayed in Fig. 12 (solid lines) for different values of the reflectivities $R_1 = R_2$. Its peak value at the resonance frequencies ν_q is

$$A'_{trans}(\nu_q) = f_{inc} f_{trans} A_{circ}(\nu_q) = \frac{(1 - R_1)(1 - R_2)e^{-\alpha_{prop}\ell}}{(1 - \sqrt{R_1 R_2} e^{-\alpha_{prop}\ell})^2} \stackrel{!}{=} 1. \quad (84)$$

For equal mirror reflectivities, $R_1 = R_2$, and vanishing propagation losses, $\alpha_{prop} = 0 \text{ cm}^{-1}$, all incident resonant light is transmitted.

A'_{trans} has been derived in the circulating-field approach [15] by considering an additional phase shift of $e^{i\pi/2} = i$ during each transmission through a mirror (alternatively, but equivalently, often a phase shift of $e^{i\pi} = -1$ during reflection from the low-refractive-index side of a mirror is assumed) [17]:

$$E_{circ} = it_1 E_{inc} + r_1 r_2 e^{-i2\phi} E_{circ} \Rightarrow \frac{E_{circ}}{E_{inc}} = \frac{it_1}{1 - r_1 r_2 e^{-i2\phi}}, \quad (85)$$

$$E_{trans} = it_2 E_{circ} e^{-i\phi} \Rightarrow \frac{E_{trans}}{E_{inc}} = \frac{-t_1 t_2 e^{-i\phi}}{1 - r_1 r_2 e^{-i2\phi}}. \quad (86)$$

By use of Eqs. (15), (54), and (55) one obtains

$$A'_{trans} = \frac{I_{trans}}{I_{inc}} = \frac{|E_{trans}|^2}{|E_{inc}|^2} = \frac{|-t_1 t_2 e^{-i\phi}|^2}{|1 - r_1 r_2 e^{-i2\phi}|^2} = \frac{(1 - R_1)(1 - R_2)}{(1 - \sqrt{R_1 R_2})^2 + 4\sqrt{R_1 R_2} \sin^2 \phi}, \quad (87)$$

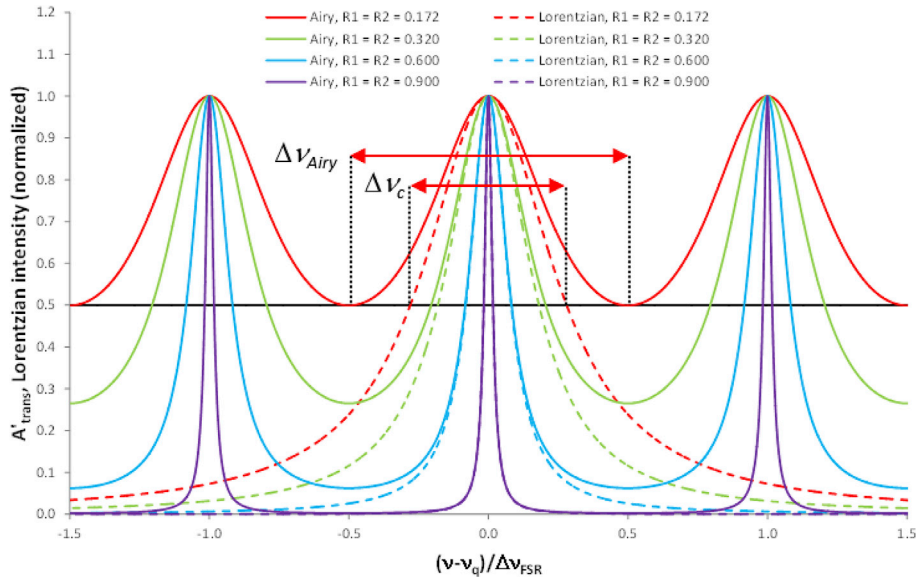


Fig. 12. Airy distribution A'_{trans} (solid lines), corresponding to light transmitted through a Fabry-Pérot resonator, calculated from Eq. (74) or (83) for different values of the reflectivities, $R_1 = R_2$, and vanishing propagation losses, $\alpha_{prop} = 0 \text{ cm}^{-1}$. Comparison with a single Lorentzian line (dashed lines) calculated from Eq. (42) for the same $R_1 = R_2$. At half maximum (black line), with decreasing reflectivities the FWHM linewidth $\Delta\nu_{Airy}$ of the Airy distribution broadens compared to the FWHM linewidth $\Delta\nu_c$ of its corresponding Lorentzian line: $R_1 = R_2 = 0.9, 0.6, 0.32, 0.172$ results in $\Delta\nu_{Airy}/\Delta\nu_c = 1.001, 1.022, 1.132, 1.717$, respectively, as calculated from Eq. (117). (Figure taken from Ref. [2]).

which equals Eqs. (74) and (83) for vanishing propagation losses, $\alpha_{prop} = 0 \text{ cm}^{-1}$. Alternatively, it can be obtained via the round-trip-decay approach [18] by tracing the infinite number of round trips that the incident electric field E_{inc} exhibits after entering the resonator and accumulating the electric field E_{trans} transmitted in all round trips. The field transmitted after the first propagation and the smaller and smaller fields transmitted after each consecutive propagation through the resonator are

$$E_{trans,1} = E_{inc}it_1t_2e^{-i\phi} = -E_{inc}t_1t_2e^{-i\phi} \quad \text{and} \quad E_{trans,m+1} = E_{trans,m}r_1r_2e^{-i2\phi}, \tag{88}$$

respectively. Exploiting

$$\sum_{m=0}^{\infty} x^m = \frac{1}{1-x} \quad \Rightarrow \quad E_{trans} = \sum_{m=1}^{\infty} E_{trans,m} = -E_{inc}t_1t_2e^{-i\phi} \sum_{m=1}^{\infty} (r_1r_2)^m e^{-im2\phi} = E_{inc} \frac{-t_1t_2e^{-i\phi}}{1-r_1r_2e^{-i2\phi}}. \tag{89}$$

This result is identical to Eq. (86), therefore one obtains the same Airy distribution of Eq. (87). Svelto [18] uses the alternative convention for the phase shift occurring at the mirrors, hence obtains no minus sign in the numerator of Eq. (89), which becomes irrelevant when squaring the electric field to calculate the intensity. A'_{trans} has also been derived in the impulse-response approach [50]. This approach is identical to the round-trip-decay approach [18], except that it assumes a specific light source that has a delta function in time, which Fourier transforms to an infinite frequency spectrum, such that A'_{trans} is probed over the entire spectral range.

Although the round-trip-decay approach is mathematically and physically correct, it suffers from a physical misinterpretation. Its calculation assumes that individual light beams interfere only after exiting the resonator. If this interpretation were true, spectral modification of the incident intensity by constructive and destructive interference would occur only after passing through the resonator. There would be no resonant enhancement or suppression of intensity inside the resonator, and the spectral shape of intensity inside the resonator would equal that of incident light, except for the simple scaling factor $(1-R_1)$. On the other hand, since the spectral shape of intensity inside the resonator would be spectrally modified when exiting the resonator, one would not be allowed to derive the intensity inside the resonator from the calculated transmitted intensity by the simple scaling factor $1/(1-R_2)$, as is usually done in the literature. As indicated in Fig. 4, interference between electric fields takes places on both sides of mirror 1.

3.4. Airy distribution of reflected light

The back-transmitted intensity I_{back} cannot be measured, because also the initially back-reflected light adds to the backward-propagating signal. In the situation of Fig. 4, the measurable case of the intensity resulting from the interference of both backward-propagating electric fields is derived as follows. The back-transmitted electric field is

$$E_{back} = it_1r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}E_{circ} = \frac{it_1r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}}{1-r_1r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}}E_{loun} = \frac{it_1r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}}{1-r_1r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}}it_1E_{inc} \quad \Rightarrow \quad \frac{E_{back}}{E_{inc}} = \frac{-t_1^2r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}}{1-r_1r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}}. \tag{90}$$

Including the initially back-reflected electric field, $E_{refl,1} = r_1E_{inc}$, the total electric field propagating backward from mirror 1 is

$$\frac{E_{refl}}{E_{inc}} = \frac{E_{refl,1} + E_{back}}{E_{inc}} = r_1 + \frac{-t_1^2r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}}{1-r_1r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}} = \frac{r_1-r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}}{1-r_1r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}}. \tag{91}$$

Exploiting Eqs. (15) and (54) yields

$$\begin{aligned} |r_1-r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}|^2 &= |r_1-r_2e^{-\alpha_{prop}\ell}\cos(2\phi) + ir_2e^{-\alpha_{prop}\ell}\sin(2\phi)|^2 \\ &= [r_1-r_2e^{-\alpha_{prop}\ell}\cos(2\phi)]^2 + [r_2e^{-\alpha_{prop}\ell}\sin(2\phi)]^2 \\ &= R_1+R_2e^{-\alpha_{prop}2\ell}-2\sqrt{R_1R_2}e^{-\alpha_{prop}\ell}\cos(2\phi) \\ &= R_1+R_2e^{-\alpha_{prop}2\ell}-2\sqrt{R_1R_2}e^{-\alpha_{prop}\ell}[1-2\sin^2(\phi)] \\ &= (\sqrt{R_1}-\sqrt{R_2}e^{-\alpha_{prop}\ell})^2 + 4\sqrt{R_1R_2}e^{-\alpha_{prop}\ell}\sin^2(\phi). \end{aligned} \tag{92}$$

With Eqs. (92) and (55), the total relative intensity propagating backward from mirror 1 amounts to

$$A'_{refl} = \frac{I_{refl}}{I_{inc}} = \frac{|E_{refl}|^2}{|E_{inc}|^2} = \frac{|r_1-r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}|^2}{|1-r_1r_2e^{-\alpha_{prop}\ell}e^{-i2\phi}|^2} = \frac{(\sqrt{R_1}-\sqrt{R_2}e^{-\alpha_{prop}\ell})^2 + 4\sqrt{R_1R_2}e^{-\alpha_{prop}\ell}\sin^2\phi}{(1-\sqrt{R_1R_2}e^{-\alpha_{prop}\ell})^2 + 4\sqrt{R_1R_2}e^{-\alpha_{prop}\ell}\sin^2\phi}. \tag{93}$$

At the resonance frequencies and in the absence of propagation losses, $\alpha_{prop} = 0 \text{ cm}^{-1}$, the back-emitted electric field E_{back} destructively interferes with the electric field $E_{refl,1}$ initially back-reflected,

$$A'_{refl}(v_q) = \frac{(\sqrt{R_1}-\sqrt{R_2}e^{-\alpha_{prop}\ell})^2}{(1-\sqrt{R_1R_2}e^{-\alpha_{prop}\ell})^2} \stackrel{R_1=R_2}{\underset{\alpha_{prop}=0}}{=} 0. \tag{94}$$

In the absence of propagation losses, $\alpha_{prop} = 0 \text{ cm}^{-1}$, when summing up the transmitted and reflected Airy distributions of Eqs. (83) and (93), respectively, one finds that the energy is conserved at all frequencies,

$$A'_{emit} = A'_{trans} + A'_{refl} = \frac{I_{trans} + I_{refl}}{I_{inc}} = 1, \quad (95)$$

as expected for a resonator that exhibits only outcoupling losses, see Fig. 13(a). It proofs that constructive and destructive interference at both sides of mirror 1 exactly compensate each other at all frequencies. In the presence of propagation losses, the Airy distribution of intensity lost during light propagation versus incident intensity,

$$A'_{prop} = \frac{I_{prop}}{I_{inc}}, \quad (96)$$

and the transmitted and reflected light intensities equals unity,

$$A'_{trans} + A'_{refl} + A'_{prop} = \frac{I_{trans} + I_{refl} + I_{prop}}{I_{inc}} = 1, \quad (97)$$

see Fig. 13(b).

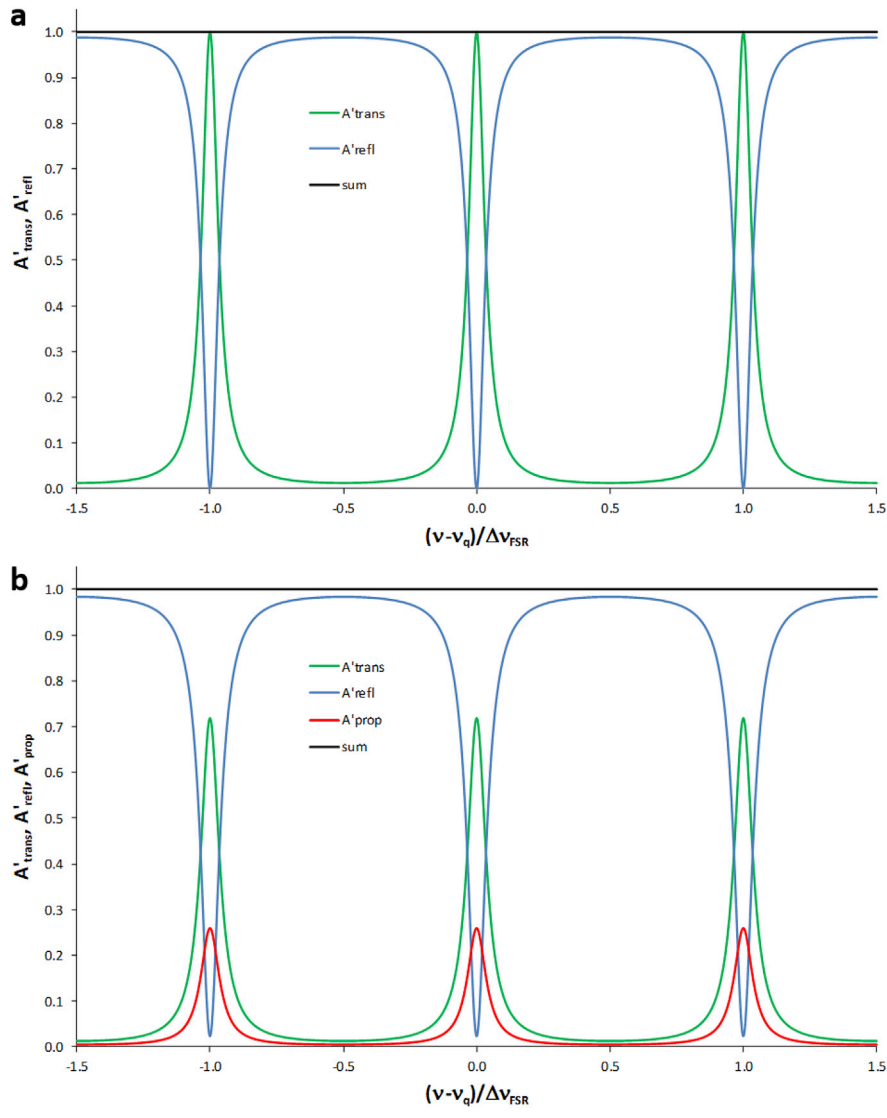


Fig. 13. Airy distributions A'_{trans} (green line), A'_{refl} (blue line), and A'_{prop} (red line), corresponding to light transmitted through, reflected by, and lost during propagation in a Fabry-Pérot resonator for $R_1 = R_2 = 0.8$ and propagation losses of (a) $\alpha_{prop}^L = 0$ and (b) $\alpha_{prop}^L = 0.04$.

3.5. Airy distribution as a sum of mode profiles

In Fig. 12 one observes that, at high reflectivity, there is almost perfect agreement between the spectral shape of the Airy distribution (solid purple line) and its underlying Lorentzian lines (dashed purple line), i.e., the former is rather well represented by the latter. This fact has prompted Saleh and Teich [20] to propose that in this case the Airy linewidth $\Delta\nu_{\text{Airy}}$ of a Fabry-Pérot resonator is similar to the linewidth $\Delta\nu_c = 1/(2\pi\tau_c)$ of its underlying Lorentzian lines, both defined as FWHM (black line). However, as is generally well known, with decreasing reflectivity the linewidth of the Airy distribution (solid lines) broadens faster than that of the underlying Lorentzian lines (dashed lines).

Svelto [19] attributes this discrepancy to Eq. (29) being only an approximation, thereby implicating that also Eq. (38) is only an approximation, such that the Airy linewidth $\Delta\nu_{\text{Airy}}$ of a Fabry-Pérot resonator can only at high reflectivity be approximated by the linewidth $\Delta\nu_c$ of its underlying Lorentzian lines. In fact, the discrepancy has a different reason [2].

According to Koppelman [51], Bayer-Helms [52] “showed that the Airy distribution can be represented exactly” by the sum of Lorentzian spectral line shapes times a calibration factor. Firstly, while being literally correct, this statement is physically misleading, secondly, the calibration factor used in Ref. [51] remains unexplained, thirdly, the equivalence is shown only for equal reflectivities, $R_1 = R_2$, and finally, the equivalence is not investigated for non-Lorentzian spectral line shapes. Kumer and Uplinger [53] compared the sum of Lorentzian spectral line shapes to an Airy distribution that is different from Eq. (56) and Eq. (74) and, not surprisingly, these authors found a difference.

We have verified [2] that, for a Fabry-Pérot resonator that exhibits only outcoupling losses, i.e., $\tau_c = \tau_{\text{out}}$, the Airy distribution is nothing else but the sum of the mode profiles of the longitudinal resonator modes, thereby revealing the physical origin of the Airy distribution. Our approach starts from the electric field E_{circ} circulating inside the resonator, considers the exponential decay in time of this field through both mirrors of the resonator, see Fig. 4, Fourier transforms it to frequency space according to Eq. (37) to obtain the normalized spectral line shapes $\tilde{\gamma}_q(\nu)$ of Eq. (38), divides it by the round-trip time t_{RT} to account for how the total circulating electric-field intensity is longitudinally distributed in the resonator and coupled out per unit time, resulting in the emitted mode profiles,

$$\gamma_{q,\text{emit}}(\nu) = \frac{1}{t_{RT}} \tilde{\gamma}_q(\nu) = \frac{1}{t_{RT}} \frac{1}{\pi} \frac{1/(4\pi\tau_c)}{1/(4\pi\tau_c)^2 + (\nu - \nu_q)^2} \quad \text{with} \quad \int \gamma_{q,\text{emit}}(\nu) d\nu = \frac{1}{t_{RT}}, \quad (98)$$

in units of (1), and then sums over the emitted mode profiles of all longitudinal modes at positive, zero, and negative frequencies, thereby explicitly considering the counter-propagating modes at each resonance frequency established in Section 2.2 and Ref. [1]. Consequently, the sum of emitted mode profiles describes an experiment that must result in the Airy distribution A_{emit} of Eq. (70).

When assuming $\tau_c = \tau_{\text{out}}$ and inserting a mirror coupling coefficient $\delta = \delta_{\text{out}}/2$ of Eq. (19), which can be a general function of frequency, the resonance frequency ν_q of Eq. (7), the round-trip phase shift 2ϕ of Eq. (3), and the free spectral range $\Delta\nu_{\text{FSR}}$ of Eq. (5), we can write the emitted mode profiles of Eq. (98) as

$$\begin{aligned} \gamma_{q,\text{emit}}(\nu) &= \frac{1}{t_{RT}} \frac{1}{\pi} \frac{\delta/(2\pi t_{RT})}{[\delta/(2\pi t_{RT})]^2 + (\nu - q\Delta\nu_{\text{FSR}})^2} \\ &= \frac{2\delta}{\delta^2 + (2\pi\nu t_{RT} - 2\pi t_{RT} q \Delta\nu_{\text{FSR}})^2} \\ &= \frac{2\delta}{\delta^2 + (2\phi - 2\pi q)^2} \\ &= \frac{\delta + i(2\phi - 2\pi q)}{\delta^2 + (2\phi - 2\pi q)^2} + \frac{\delta - i(2\phi - 2\pi q)}{\delta^2 + (2\phi - 2\pi q)^2} \\ &= \frac{1}{\delta - i(2\phi - 2\pi q)} + \frac{1}{\delta + i(2\phi - 2\pi q)} \\ &= \frac{i}{2\pi} \left[\frac{1}{(\phi + i\delta/2)/\pi - q} - \frac{1}{(\phi - i\delta/2)/\pi - q} \right]. \end{aligned} \quad (99)$$

With the mathematical series [54]

$$\sum_{q=-\infty}^{+\infty} \frac{1}{x - q} = \pi \cot(\pi x), \quad (100)$$

the sum of all mode profiles obtains as

$$\sum_{q=-\infty}^{\infty} \gamma_{q,\text{emit}}(\nu) = \frac{i}{2\pi} \left[\sum_{q=-\infty}^{+\infty} \frac{1}{(\phi + i\delta/2)/\pi - q} - \sum_{q=-\infty}^{+\infty} \frac{1}{(\phi - i\delta/2)/\pi - q} \right] = \frac{i}{2} \left[\cot\left(\phi + \frac{i\delta}{2}\right) - \cot\left(\phi - \frac{i\delta}{2}\right) \right]. \quad (101)$$

Expressing the cot functions as

$$\cot(\alpha) - \cot(\beta) = \frac{\sin(\beta - \alpha)}{\sin(\alpha)\sin(\beta)} \quad (102)$$

leads to

$$\sum_{q=-\infty}^{\infty} \gamma_{q,emit}(\nu) = \frac{i}{2} \frac{\sin(-i\delta)}{\sin[\phi + i\delta/2]\sin[\phi - i\delta/2]}. \quad (103)$$

Exploiting the trigonometric relation

$$\sin(\alpha + \beta)\sin(\alpha - \beta) = \sin^2(\alpha) - \sin^2(\beta) \quad (104)$$

results in

$$\sum_{q=-\infty}^{\infty} \gamma_{q,emit}(\nu) = \frac{i}{2} \frac{\sin(-i\delta)}{\sin^2(\phi) - \sin^2(i\delta/2)}. \quad (105)$$

Making use of the Euler identity

$$\sin(x) = \frac{e^{ix} - e^{-ix}}{2i} \quad (106)$$

yields

$$\sum_{q=-\infty}^{\infty} \gamma_{q,emit}(\nu) = \frac{i}{2} \frac{(e^\delta - e^{-\delta})/(2i)}{\sin^2(\phi) - [(e^{-\delta/2} - e^{\delta/2})/(2i)]^2} = \frac{e^\delta - e^{-\delta}}{(e^{-\delta/2} - e^{\delta/2})^2 + 4\sin^2(\phi)} = \frac{1 - e^{-2\delta}}{(1 - e^{-\delta})^2 + 4e^{-\delta}\sin^2(\phi)}. \quad (107)$$

Finally, by re-substituting $\delta_{out}/2 = \delta$ and Eq. (19) into Eq. (107), the sum of emitted mode profiles of Eq. (98) becomes

$$\sum_{q=-\infty}^{\infty} \gamma_{q,emit}(\nu) = \frac{1 - R_1R_2}{(1 - \sqrt{R_1R_2})^2 + 4\sqrt{R_1R_2}\sin^2(\phi)} = A_{emit}. \quad (108)$$

Indeed, this result is equal to Eq. (70), with Eq. (56) inserted. This equivalence is numerically illustrated in Fig. 14(a).

Each spectral line shape $\tilde{\gamma}_q(\nu)$ of Eq. (38) and mode profile $\gamma_{q,emit}(\nu)$ of Eq. (98) is extended over the infinite frequency range, consequently light at a specific frequency ν inside the resonator excites all longitudinal modes of the resonator. However, in the absence of propagation losses, $\alpha_{prop} = 0 \text{ cm}^{-1}$, the contributions from different longitudinal modes to the light at frequency ν do not interfere with each other, because all optical modes are orthogonal with each other. For this reason, the sum in Eq. (108) is over the intensity mode profiles $\gamma_{q,emit}(\nu)$ rather than over the electric fields. Because of resonant enhancement of the launched light the peak value of A_{emit} at the resonance frequencies ν_q ,

$$A_{emit}(\nu_q) = \frac{1 - R_1R_2}{(1 - \sqrt{R_1R_2})^2} = \frac{1 + \sqrt{R_1R_2}}{1 - \sqrt{R_1R_2}} > 1, \quad (109)$$

becomes larger than unity. Nevertheless, because of Eq. (95) the energy of the system is conserved at all frequencies.

The observation that with decreasing R_1 and R_2 the linewidth of the resulting Airy distribution in Fig. 12 is increasingly broader than the linewidth of the underlying Lorentzian lines simply arises from the fact that one sums up mode profiles (with the same linewidth as the Lorentzian lines) that resonate at different frequencies. It does not constitute a discrepancy, as has often been proposed.

The derivation shown here demonstrates that—from a physical point of view—the spectral line shapes and mode profiles are the fundamental spectral functions that characterize the Fabry-Pérot resonator and their sum, the Airy distribution, quantifies its spectral response. The circulating-field and round-trip-decay approaches only implicitly exploit this fact, thereby obscuring the physical nature of the Airy distribution. As we will see in Section 4, this fundamental understanding has direct consequences for the definitions of linewidth and finesse.

The same intensity-scaling factors of Eqs. (64) and (65) that provide the relations between the individual Airy distributions, see Eqs. (56) and (66)–(76), also provide the relations among $\gamma_{q,emit}(\nu)$ and the other mode profiles, albeit only in the absence of propagation losses, $\alpha_{prop} = 0 \text{ cm}^{-1}$:

$$\begin{aligned} \gamma_{q,circ} &= \frac{1}{R_2} \gamma_{q,b-circ} = \frac{1}{R_1R_2} \gamma_{q,RT} = \frac{1}{1 - R_2} \gamma_{q,trans} = \frac{1}{(1 - R_1)R_2} \gamma_{q,back} = \frac{1}{1 - R_1R_2} \gamma_{q,emit} \\ \gamma'_{q,circ} &= \frac{1}{R_2} \gamma'_{q,b-circ} = \frac{1}{R_1R_2} \gamma'_{q,RT} = \frac{1}{1 - R_2} \gamma'_{q,trans} = \frac{1}{(1 - R_1)R_2} \gamma'_{q,back} = \frac{1}{1 - R_1R_2} \gamma'_{q,emit} = (1 - R_1) \gamma_{q,circ}. \end{aligned} \quad (110)$$

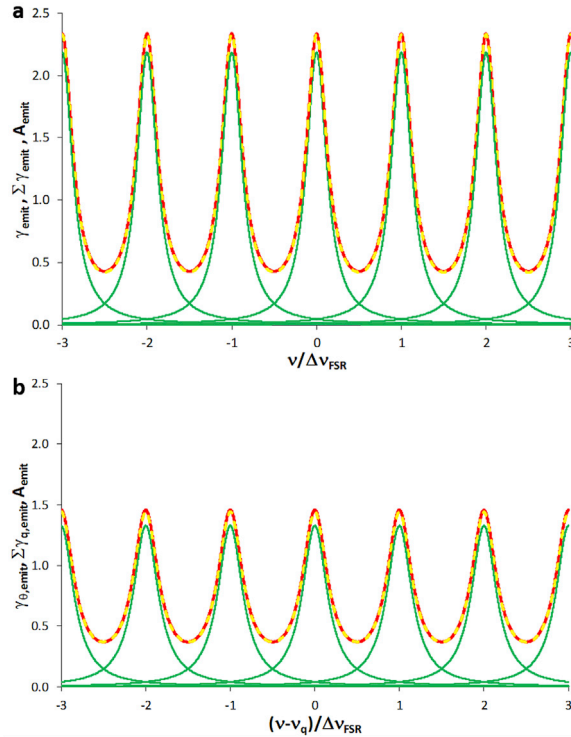


Fig. 14. Mode profiles $\gamma_{q,emit}$ (solid green lines) with mode indices $q = -3, -2, -1, 0, 1, 2, 3$ of Eq. (98), here approximated by the numerical sum (solid red line) of 31 mode profiles with mode indices q between -15 and 15 , and Airy distribution A_{emit} (dashed yellow line) of Eqs. (70) and (108). The displayed examples are for $\ell = 1$ cm, $n_r = 1$, equal mirror reflectivities of $R_1 = R_2 = 0.4$, and (a) vanishing propagation losses, $\alpha_{prop} = 0$ cm $^{-1}$, resulting in $\tau_c = \tau_{out} = 3.6 \times 10^{-11}$ s, or (b) propagation losses of $\alpha_{prop} = 0.1$ cm $^{-1}$, resulting in $\tau_c = 3.0 \times 10^{-11}$ s (Figure (a) taken from Ref. [1]).

The various mode profiles $\gamma_q(\nu)$ and $\gamma'_q(\nu)$ are calibrated with respect to the launched and the incident intensity, respectively, and the sum over one of these mode profiles at all resonance frequencies generates the corresponding Airy distribution.

In the presence of propagation loss, the physical origin of the Airy distribution is still the sum of mode profiles of the longitudinal resonator modes. The same intensity-scaling factors of Eq. (66)–(76) that provide the relations between the individual Airy distributions also provide the relations between the individual mode profiles. For example,

$$\gamma_{q,emit}(\nu) = A_{emit,FSR} \frac{1}{t_{RT}} \tilde{\gamma}_q(\nu) = f_{emit,circ,FSR} \frac{1}{t_{RT}} \tilde{\gamma}_q(\nu) = \frac{[(1 - R_2) + (1 - R_1)R_2 e^{-\alpha_{prop}\ell}] e^{-\alpha_{prop}\ell}}{1 - R_1 R_2 e^{-\alpha_{prop}2\ell}} \frac{1}{t_{RT}} \tilde{\gamma}_q(\nu), \quad (111)$$

which is displayed in Fig. 14(b). In Eq. (111), for $\alpha_{prop} = 0$ cm $^{-1}$ one obtains Eq. (98). Another example is

$$\gamma'_{q,trans}(\nu) = A'_{trans,FSR} \frac{1}{t_{RT}} \tilde{\gamma}_q(\nu) = f_{inc,trans,circ,FSR} \frac{1}{t_{RT}} \tilde{\gamma}_q(\nu) = \frac{(1 - R_1)(1 - R_2) e^{-\alpha_{prop}\ell}}{1 - R_1 R_2 e^{-\alpha_{prop}2\ell}} \frac{1}{t_{RT}} \tilde{\gamma}_q(\nu). \quad (112)$$

Situations, in which the mirror reflectivities R_1 and R_2 are frequency dependent, can be described by the same equations as above [2]. Since the photon-decay time also becomes frequency dependent, the FWHM linewidth loses its meaning. In such situations, each Airy distribution is still the infinite sum of its underlying mode profiles. However, the mode profiles are not Lorentzian-shaped anymore and, consequently, the Airy distributions become distorted [2].

3.6. 0th-order mode: no cut-off wavelength!

The resonance frequencies of the Fabry-Pérot resonator are given by Eq. (7). Particularly, $q = 0$, resulting in $\nu_q = 0$, is a physical solution of Eq. (7), and the mode profiles summed in Eq. (108) include this 0th-order mode [55]. Consequently, in wavelength space there exists no cut-off wavelength at the resonance line with $q = 1$, but an additional resonance occurs at infinite wavelength. As long as the resonator losses are low (purple line in Fig. 15), in most cases this resonance will go unnoticed. However, when the losses become significant and the mode profiles broaden, this resonance manifests itself in a steady increase of the Airy distribution A'_{trans} in the long-wavelength range (other lines in Fig. 15). While this effect looks quite dramatic in Fig. 15, one should keep in mind that the natural energy or frequency units are, when converted to wavelength units, expanded by λ^2 at the long-wavelength side.

4. Lorentzian linewidth and finesse versus Airy linewidth and finesse

The understanding that the Airy distribution describing the spectral transmission of a Fabry-Pérot resonator physically originates in the sum of mode profiles of the longitudinal resonator modes has fundamental consequences. In this Section, we point out the different functions of the Fabry-Pérot resonator as a resonator with fixed length, characterized by its Lorentzian linewidth and finesse, and as a scanning spectrometer by varying its length, characterized by its Airy linewidth and finesse.

4.1. Spectral resolution: Rayleigh and Taylor criteria

The most commonly accepted definitions of spectral resolution are the Rayleigh criterion and the Taylor criterion. The Rayleigh criterion [56], see Fig. 16(a), applies to diffraction spots and proposes that two intensity maxima can be resolved if the maximum of the first spot is spectrally superimposed on the first minimum of the second spot. For spectral lines that do not exhibit minima, such as the Lorentzian lines underlying the Airy distribution of a Fabry-Pérot resonator, as well as the Airy distribution itself within one free spectral range from $\nu_q - \Delta\nu_{FSR}/2$ to $\nu_q + \Delta\nu_{FSR}/2$, the Taylor criterion [57], see Fig. 16(b), is more appropriate. It proposes that two spectral lines can be resolved if the individual lines cross each other at half intensity. In the case of two identical, symmetric spectral lines their peaks would then be separated by their FWHM. The Taylor criterion will be utilized in the following.

4.2. Characterizing the Fabry-Pérot resonator: Lorentzian linewidth and finesse

When launching light into the Fabry-Pérot resonator in a non-scanning experiment, i.e., at fixed resonator length (and fixed angle of incidence), the Fabry-Pérot resonator becomes the object of investigation. The spectral line shapes, Lorentzian lines, and mode profiles are the fundamental functions. By measuring the sum of mode profiles, the Airy distribution, one can derive the total loss of the Fabry-Pérot resonator via recalculating the Lorentzian linewidth $\Delta\nu_c$ of Eq. (39), displayed (blue line) relative to the free spectral range in Fig. 17(a) and (c).

The underlying Lorentzian lines can be resolved as long as the Taylor criterion is obeyed (Fig. 18). Consequently, one can define a parameter which we call the Lorentzian finesse of a Fabry-Pérot resonator:

$$F_c := \frac{\Delta\nu_{FSR}}{\Delta\nu_c} = \frac{2\pi}{-\ln(R_1 R_2 e^{-\alpha_{prop} 2\ell})}. \quad (113)$$

It is displayed as blue lines in Fig. 17(b) and (d). This parameter has received little attention in the scientific literature, other than that at high reflectivity, where $\Delta\nu_{Airy} \approx \Delta\nu_c$, see the purple line in Fig. 12, it can serve as a reasonable approximation of the Airy finesse F_{Airy} (see Section 4.2). Yet, the Lorentzian finesse F_c has a fundamental physical meaning: it describes how well the Lorentzian lines underlying the Airy distribution can be resolved when measuring the Airy distribution. A Fabry-Pérot resonator generating single-longitudinal-mode laser light is characterized by its Lorentzian linewidth and finesse.

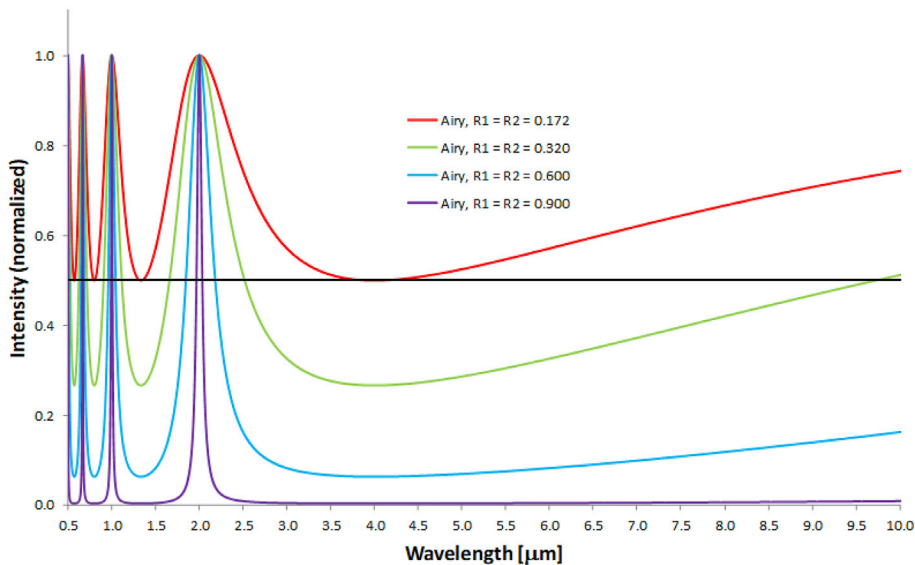


Fig. 15. Airy distribution A'_{trans} of a Fabry-Pérot micro-resonator with a length of $\ell = 1 \mu\text{m}$ versus wavelength, calculated from Eq. (74) for different values of the reflectivities $R_1 = R_2$, and vanishing propagation losses, $\alpha_{prop} = 0 \text{ cm}^{-1}$. The resonance peak with $q = 1$, in our example occurring at a wavelength of $2 \mu\text{m}$, which is often considered as a “cut-off wavelength”. At wavelengths beyond $4 \mu\text{m}$, however, the Airy distribution increases towards the resonance peak with $q = 0$, located at infinite wavelength. (Figure taken from Ref. [2]).

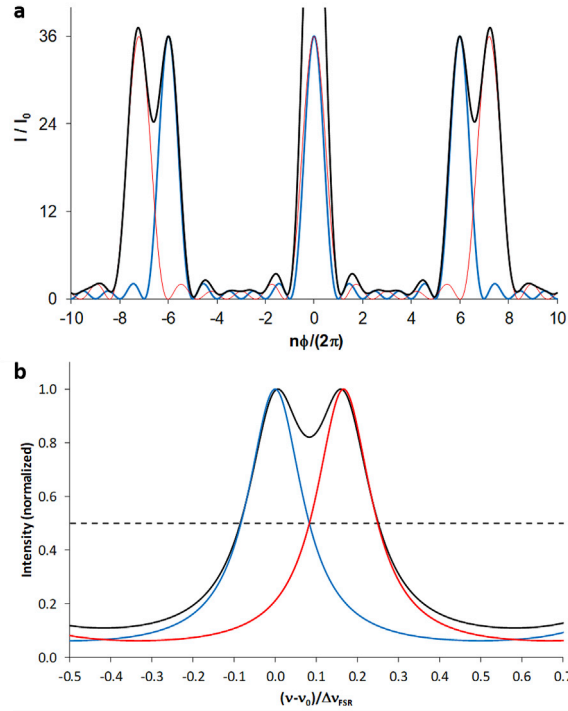


Fig. 16. Criteria for the limit of spectral resolution: (a) Rayleigh criterion and (b) Taylor criterion.

Since $\Delta\nu_c$ exists for any mirror reflectivity and propagation losses, the definition of the Lorentzian finesse does not break down at a critical value. However, at the point where

$$\Delta\nu_c = \Delta\nu_{FSR} \Rightarrow R_1 R_2 e^{-\alpha_{prop} 2\ell} = e^{-2\pi} \approx 0.001867, \quad (114)$$

equivalent to $F_c = 1$, the Taylor criterion for the spectral resolution of a single Airy distribution is reached. For equal mirror reflectivities and vanishing propagation losses, this point occurs when $R_1 = R_2 \approx 4.32\%$. Therefore, the linewidth of the Lorentzian lines underlying the Airy distribution of a Fabry-Pérot resonator can be resolved by measuring the Airy distribution, hence its resonator losses can be spectroscopically determined, until this point. Obviously, the Lorentzian finesse according to the definition of Eq. (113) plays an essential role in the characterization of low-reflectivity or otherwise high-loss Fabry-Pérot resonators.

4.3. Scanning the Fabry-Pérot resonator: Airy linewidth and finesse

A different situation occurs when the Fabry-Pérot resonator is used as a scanning interferometer, i.e., at varying resonator length (or angle of incidence), to spectroscopically distinguish spectral lines at different frequencies within one free spectral range. In this case several Airy distributions $A'_{trans}(\nu)$, each one created by an individual frequency, must be resolved. Therefore, now the Airy distribution $A'_{trans}(\nu)$ becomes the underlying fundamental function and the measurement delivers a sum of Airy distributions. The parameters that properly quantify this situation are the Airy linewidth $\Delta\nu_{Airy}$ and the Airy finesse F_{Airy} .

On either side of the peak of Eq. (84) located at ν_q , the transmitted intensity decreases to $A'_{trans}(\nu_q)/2$ when the phase shift ϕ changes by the amount $\Delta\phi$ and, accordingly, $\sin^2(\phi)$ changes from 0, such that in the denominator of A_{circ} of Eq. (56)

$$4\sqrt{R_1 R_2} e^{-\alpha_{prop} \ell} \sin^2(\Delta\phi) = (1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell})^2 \Rightarrow \Delta\phi = \arcsin\left(\frac{1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell}}{2\sqrt[4]{R_1 R_2} e^{-\alpha_{prop} \ell/2}}\right). \quad (115)$$

Exploiting Eqs. (3) and (5) to calculate $\phi = \pi\nu/\Delta\nu_{FSR}$, resulting in $\Delta\phi = \pi(\Delta\nu_{Airy}/2)/\Delta\nu_{FSR}$, then provides the FWHM linewidth $\Delta\nu_{Airy}$ of the Airy distribution $A'_{trans}(\nu)$ [15],

$$\Delta\nu_{Airy} = \Delta\nu_{FSR} \frac{2}{\pi} \arcsin\left(\frac{1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell}}{2\sqrt[4]{R_1 R_2} e^{-\alpha_{prop} \ell/2}}\right). \quad (116)$$

The Airy linewidth $\Delta\nu_{Airy}$ is displayed as green curves in Fig. 17(a) and (c) in direct comparison with the Lorentzian linewidth $\Delta\nu_c$. The

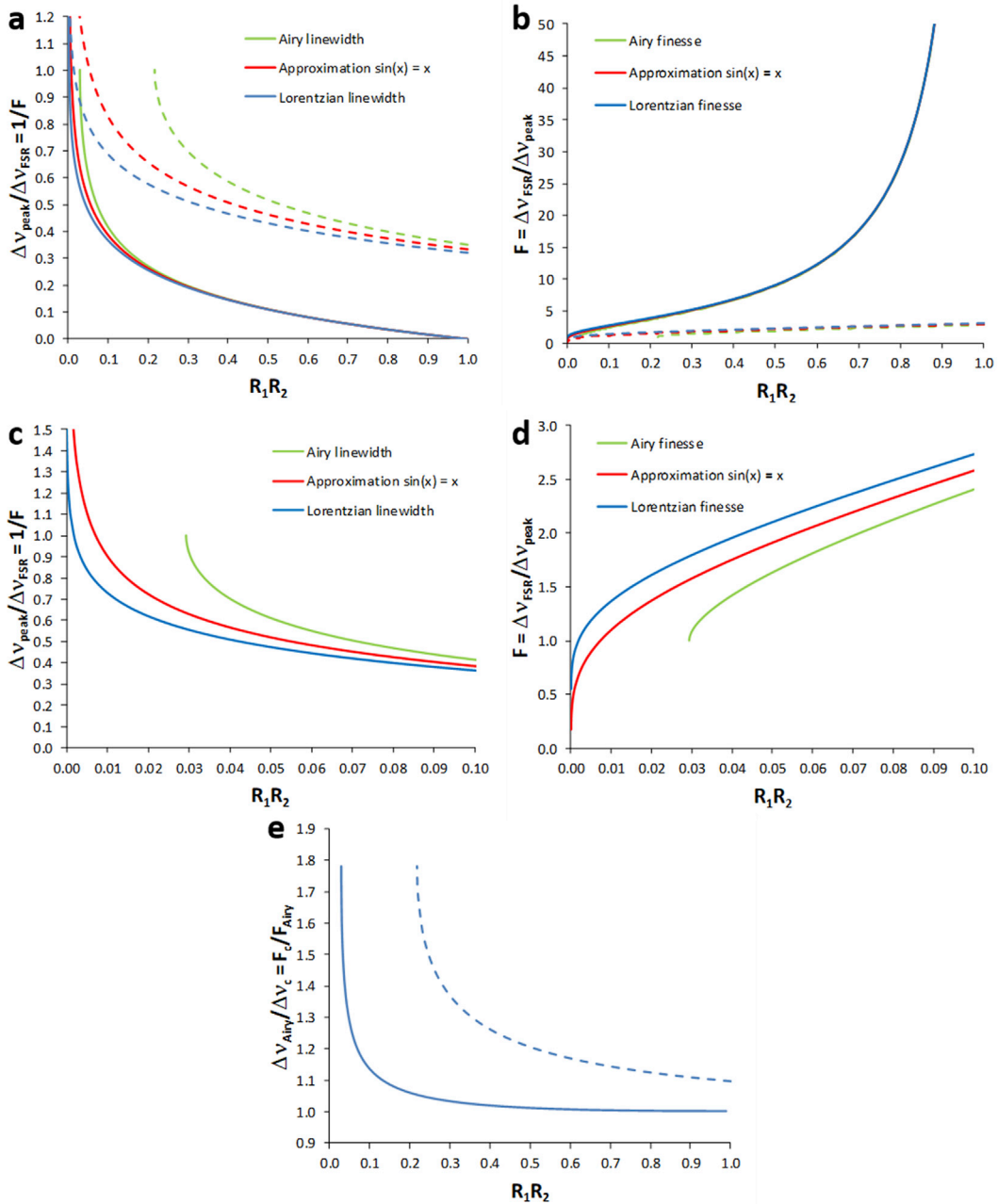


Fig. 17. (a) Relative Lorentzian linewidth $\Delta\nu_c/\Delta\nu_{\text{FSR}}$, with $\Delta\nu_c$ from Eq. (39) (blue curve), relative Airy linewidth $\Delta\nu_{\text{Airy}}/\Delta\nu_{\text{FSR}}$, with $\Delta\nu_{\text{Airy}}$ from Eq. (116) (green curve), and its approximation of Eq. (120) (red curve), and (b) Lorentzian finesse F_c of Eq. (113) (blue curve), Airy finesse F_{Airy} of Eq. (119) (green curve), and its approximation of Eq. (121) (red curve) as a function of reflectivity product $R_1 R_2$, for propagation losses of $\alpha_{\text{prop}}^{\ell} = 0$ (solid lines) and $\alpha_{\text{prop}}^{\ell} = 1$ (dotted lines). (c), (d) Zoom into the low-reflectivity region for $\alpha_{\text{prop}}^{\ell} = 0$. The exact solutions of the Airy linewidth and finesse (green curves) correctly break down at $\Delta\nu_{\text{Airy}} = \Delta\nu_{\text{FSR}}$, equivalent to $F_{\text{Airy}} = 1$, whereas their approximations (red curves) incorrectly do not break down. (e) Ratio of Eq. (117) between the Airy linewidth $\Delta\nu_{\text{Airy}}$ of Eq. (116) and the Lorentzian linewidth $\Delta\nu_c$ of Eq. (39), equaling the ratio between the Lorentzian finesse F_c of Eq. (113) and the Airy finesse F_{Airy} of Eq. (119), as a function of reflectivity product $R_1 R_2$. (Figure modified from Ref. [2]).

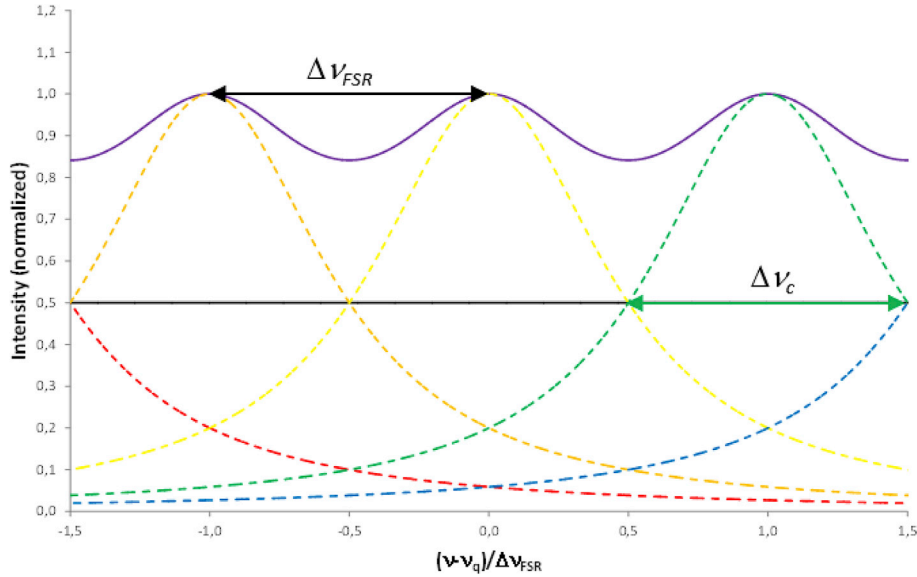


Fig. 18. Illustration of the physical meaning of the Lorentzian finesse F_c of a Fabry-Pérot resonator. Displayed is the situation for vanishing propagation losses, $\alpha_{prop} = 0 \text{ cm}^{-1}$, and $R_1 = R_2 \approx 4.32\%$, at which $\Delta\nu_c = \Delta\nu_{FSR}$ and $F_c = 1$, i.e., two adjacent Lorentzian lines (dashed colored lines, only 5 lines are shown for clarity) cross at half maximum (solid black line) and the Taylor criterion for spectrally resolving two peaks in the resulting Airy distribution (solid purple line) is reached. (Figure taken from Ref. [2]).

ratio between $\Delta\nu_{Airy}$ of Eq. (116) and $\Delta\nu_c$ of Eq. (39), which equals the ratio between the Lorentzian finesse F_c of Eq. (113) and the Airy finesse F_{Airy} of Eq. (119),

$$\frac{\Delta\nu_{Airy}}{\Delta\nu_c} = \frac{F_c}{F_{Airy}} = \frac{4 \arcsin\left(\frac{1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell}}{2 \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell / 2}}\right)}{-\ln(R_1 R_2) + I_{RT} C \alpha_{prop}}, \quad (117)$$

is displayed in Fig. 17(e).

The concept of defining the linewidth of the Airy peaks as FWHM breaks down at $\Delta\nu_{Airy} = \Delta\nu_{FSR}$ (solid red line in Fig. 12 and green line in Fig. 19), because at this point the Airy linewidth instantaneously jumps to an infinite value (Fig. 19). For lower reflectivity values $R_1 R_2$, the FWHM linewidth of the Airy peaks is undefined and other definitions or concepts would have to be utilized to describe the situation. The limiting case occurs at

$$\Delta\nu_{Airy} = \Delta\nu_{FSR} \Rightarrow \frac{1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell}}{2 \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell / 2}} = 1 \stackrel{\alpha_{prop}=0}{\Rightarrow} R_1 R_2 = 17 - 12\sqrt{2} \approx 0.02944. \quad (118)$$

For vanishing propagation losses, $\alpha_{prop} = 0 \text{ cm}^{-1}$, and equal mirror reflectivities, this point is reached when $R_1 = R_2 \approx 17.2\%$ (solid red line in Fig. 12 and green line in Fig. 19). A number of fiber and channel waveguide lasers that produce very high gain due to their strong pump confinement are operated in resonators that rely solely on the Fresnel reflection on both ends of the resonator, which is typically below this value. I.e., we are operating numerous laser resonators whose passive Airy linewidth $\Delta\nu_{Airy}$ is undefined. Nevertheless, no problem arises from it, because the appropriate parameters in the situation of the lasing Fabry-Pérot resonator are its Lorentzian linewidth $\Delta\nu_c$ and finesse F_c .

Definitions can never be right or wrong. They can only be more or less meaningful and, in the worst case, contradict with another more or less meaningful definition. The finesse of the Airy distribution of a Fabry-Pérot resonator, which is displayed as green curves in Fig. 17(b) and (d) in direct comparison with the Lorentzian finesse F_c , is properly defined as

$$F_{Airy} \stackrel{!}{=} \frac{\Delta\nu_{FSR}}{\Delta\nu_{Airy}} = \frac{\pi}{2} \left[\arcsin\left(\frac{1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell}}{2 \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell / 2}}\right) \right]^{-1}. \quad (119)$$

When scanning the length of the Fabry-Pérot resonator (or alternatively the angle of incident light), the Airy finesse lucidly quantifies the maximum number of Airy distributions created by light at individual frequencies ν_m within the free spectral range of the Fabry-Pérot resonator, whose adjacent peaks can be unambiguously distinguished spectroscopically, i.e., they do not overlap at their FWHM (Fig. 20). Consequently, this definition of the Airy finesse is consistent with the Taylor criterion of the resolution of a spectrometer, see Fig. 16(b), and is, therefore, denoted by an exclamation mark in Eq. (119). Since the concept of the FWHM linewidth breaks down at $\Delta\nu_{Airy} = \Delta\nu_{FSR}$, consequently the Airy finesse is defined only until $F_{Airy} = 1$, see Fig. 17(d), because the arcsin function in Eq. (119) cannot

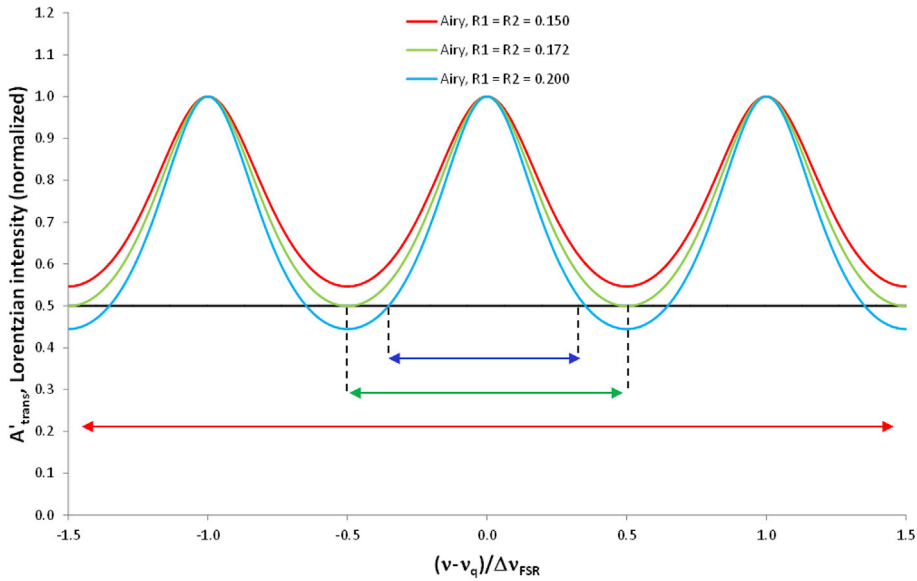


Fig. 19. Breakdown of Airy linewidth. When the minimum of the Airy distribution crosses half its maximum (critical case: green line), which is the case when, e.g., $\alpha_{prop} = 0 \text{ cm}^{-1}$ and $R_1 = R_2 \approx 17.2\%$, the linewidth instantaneously jumps to infinity.

produce values above $\pi/2$.

In 2014 the Lorentzian finesse was proposed [58] as a better definition of the finesse of a Fabry-Pérot resonator than the Airy finesse, because it does not break down at low reflectivity. However, as we have seen above, firstly, these two finessses describe two fundamentally different physical situations, i.e., they do not compete with each other and, therefore, one cannot be better than the other, and, secondly, the Airy finesse must break down because of the commonly accepted definition of a linewidth, and any attempt to circumvent this problem from the side of the Airy finesse literally puts the cart before the horse.

Generally, one can argue that the Taylor criterion leaves some ambiguity to the definition of the limit of spectral resolution, because it does not state whether it requires the absence or allows for the presence of additional resolvable spectral lines. In the example of Fig. 20, the sum of two adjacent “resolvable” Airy distributions (dashed gray line) is better resolvable than the sum of all “resolvable” Airy distributions (dashed black line), because in the latter case additional Airy distributions contribute, see the dashed lines within the red circle in Fig. 20. If either of the two options defines the limit of spectral resolution, the other cannot.

4.4. Questionable approximations and definitions

Often the unnecessary approximation $\sin(\phi) \approx \phi$ is made when deriving from A'_{trans} of Eq. (74) or Eq. (87) the Airy linewidth [12,18]. In contrast to the exact Eqs. (115) and (116), it leads to

$$4\sqrt{R_1 R_2} e^{-\alpha_{prop} \ell} (\Delta\phi)^2 \approx (1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell})^2 \Rightarrow \Delta\nu_{Airy} \approx \Delta\nu_{FSR} \frac{2}{\pi} \Delta\phi = \Delta\nu_{FSR} \frac{1}{\pi} \frac{1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell}}{\sqrt{R_1 R_2} e^{-\alpha_{prop} \ell / 2}}. \quad (120)$$

This approximation of the Airy linewidth, displayed as red curves in Fig. 17(a) and (c), deviates from the correct curve at low reflectivities and incorrectly does not break down when $\Delta\nu_{Airy} > \Delta\nu_{FSR}$. This approximation is then typically also inserted into Eq. (119) to calculate the Airy finesse [18], resulting in

$$F_{Airy} := \frac{\Delta\nu_{FSR}}{\Delta\nu_{Airy}} \approx \pi \frac{\sqrt{R_1 R_2} e^{-\alpha_{prop} \ell / 2}}{1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell}}. \quad (121)$$

Vaughan [12] and Siegman [16] even defined the Airy finesse by its approximation of Eq. (121),

$$\widehat{F}_{Airy} := \pi \frac{\sqrt{R_1 R_2} e^{-\alpha_{prop} \ell / 2}}{1 - \sqrt{R_1 R_2} e^{-\alpha_{prop} \ell}} \approx \frac{\Delta\nu_{FSR}}{\Delta\nu_{Airy}}, \quad (122)$$

thereby depriving this parameter of its lucid meaning. Since the definition of Eq. (122) does not comply with the Taylor criterion, it is denoted by a question mark. Saleh and Teich [21,22] also proposed Eq. (122) for the Airy finesse, but from their derivation it remains unclear whether they consider it as a definition or as an approximation of Eq. (119).

Whereas the meaningful definitions of Airy linewidth and finesse correctly break down at $\Delta\nu_{Airy} = \Delta\nu_{FSR}$ and $F_{Airy} = 1$, the Lorentzian

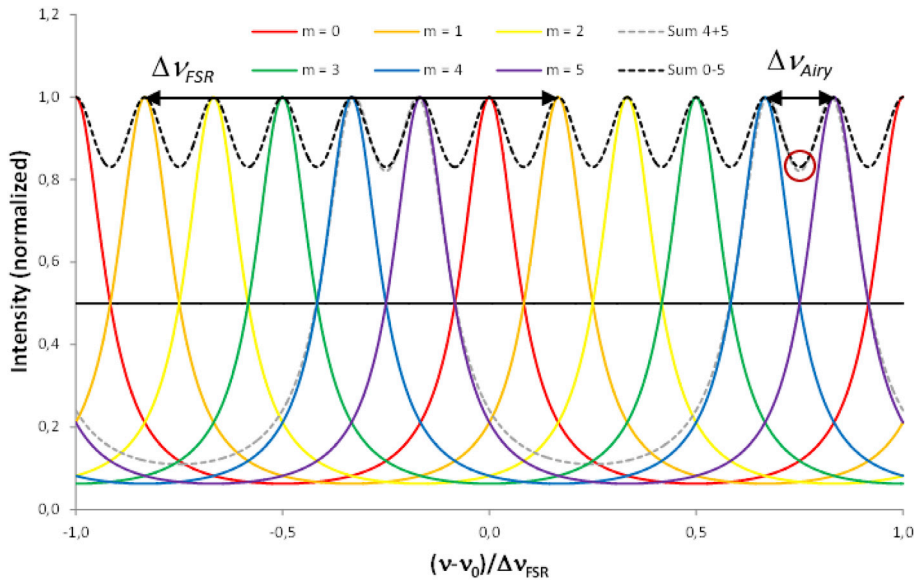


Fig. 20. Illustration of the physical meaning of the Airy finesse F_{Airy} of a Fabry-Pérot resonator. When scanning the Fabry-Pérot length (or alternatively the angle of incident light), Airy distributions (solid lines) are created by signals at individual frequencies. If the signals occur at frequencies $\nu_m = \nu_q + m\Delta\nu_{\text{Airy}}$, where m is an integer starting at q , the Airy distributions at adjacent frequencies are separated from each other by the linewidth $\Delta\nu_{\text{Airy}}$, thereby fulfilling the Taylor criterion for the spectroscopic resolution of two adjacent peaks. The maximum number of signals that can be resolved is F_{Airy} . Since in the specific example displayed in this figure, $\alpha_{\text{prop}} = 0 \text{ cm}^{-1}$ and the reflectivities $R_1 = R_2 = 0.59928$ have been chosen such that $F_{\text{Airy}} = 6$ is an integer, the signal for $m = F_{\text{Airy}}$ at the frequency $\nu_q + F_{\text{Airy}}\Delta\nu_{\text{Airy}} = \nu_q + \Delta\nu_{\text{FSR}}$ coincides with the signal for $m = q$ at ν_q . In this example, a maximum of $F_{\text{Airy}} = 6$ peaks can be resolved when applying the Taylor criterion. However, the sum of two adjacent “resolvable” peaks (dashed gray line) exhibits a deeper dip between the adjacent peaks to be resolved, i.e., a better resolution, than the sum of all “resolvable” peaks (dashed black line), see the difference highlighted in the red circle. (Figure taken from Ref. [2]).

linewidth $\Delta\nu_c$ and finesse F_c do not break down at a critical value. Moreover, the latter reach their Taylor limit at higher losses than the former. Therefore, useful and well-defined information can be extracted from non-scanning Fabry-Pérot measurements beyond the limit of the Airy linewidth and finesse (which are irrelevant quantities for non-scanning measurements).

5. Phase aspect in photon emission and absorption

In his centennial paper [59], Einstein exploited a semi-classical rate-equation approach including the rates of stimulated and spontaneous emission and absorption of electromagnetic radiation to provide the physical foundation of Planck’s law of blackbody radiation [34], thereby predicting the existence of stimulated emission, which was confirmed experimentally in 1928 [60]. By assuming conservation of energy and momentum, Einstein showed that an incident electromagnetic field at frequency ν triggers a two-level atom in its excited state to emit an additional electromagnetic field with an energy $h\nu$ that equals the energy gap between the two levels, such that the energy of the incident field increases by this energy during the interaction. Only a very small energy mismatch occurs due to recoil of the atom. The emitted field has the same frequency, the same direction, and the same polarization as the incident field, hence it is emitted into the same optical mode. In addition, spontaneous emission into the same optical mode occurs. In his original work, Einstein did not specify the phase difference between the incident and emitted electromagnetic field.

In 1995 Lamb reprimanded [61] Einstein for a fundamental conceptual mistake in his seminal paper from 1917 [59], when introducing this “new” (set within quotation marks by Lamb [61]) process of stimulated emission, quantified by the Einstein B coefficient. Lamb criticized Einstein for not having realized that, whereas the Einstein A coefficient of spontaneous emission cannot be a result of classical electromagnetic theory, “the classical Maxwell electrodynamics already made provision for both of the Einstein B coefficients of absorption and stimulated emission” [61]. Lamb then becomes explicit: “It would have made for much better physics if Einstein had recognized this fact, and had used his theory to calculate the value of the A coefficient for spontaneous emission in 1917, instead of leaving it to Dirac in 1927 to get the A coefficient from the quantum theory of radiation” [61]. Here Lamb referred to Ref. [62], albeit without citing it.

In a 1st-year university course on classical electrodynamics we are typically instructed in the following way (Fig. 21).

- (A) Spontaneous emission is a process, in which an atom in its excited state emits an electromagnetic field, comprising the energy of one photon, with an arbitrary phase with respect to a potentially existing incident field (but the students would learn later that spontaneous emission is actually driven by so-called vacuum fluctuations [63]).
- (B) Stimulated emission (absorption) is a process, in which an incident electromagnetic field, comprising the energy of one photon, forces an atom in its excited (ground) state to oscillate and emit a second electromagnetic field, comprising the energy of one

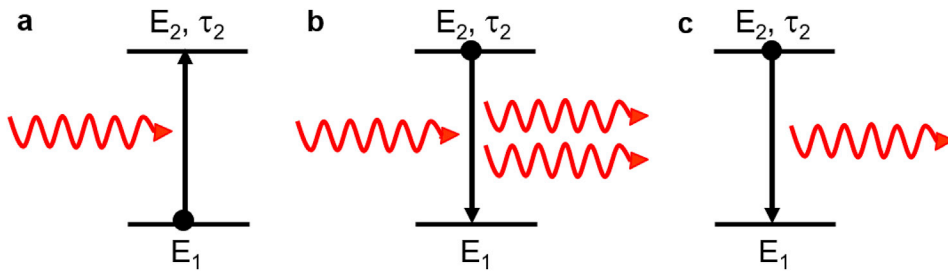


Fig. 21. The processes of (a) absorption, (b) stimulated emission, and (c) spontaneous emission by a two-level atom with level energies E_1 and E_2 and an upper-state luminescence lifetime of τ_2 , as typically depicted in an optics text book. (Figure taken from Ref. [4]).

photon, that is in phase (anti-phase) with the incident field, such that these two fields interfere constructively (destructively) with each other and the energy is emitted (absorbed) by the atom.

In the literature, four different accounts of the phase difference between the incident and the emitted field can be found. (i) In line with the above-mentioned instructions, the amplitude-phase diagram of Fig. 22(a) proposes that stimulated emission is in phase with the incident field (red solid arrows), whereas according to Lax [25], Haken [26], Henry [64], and others spontaneous emission occurs at an arbitrary phase angle θ with the incident field (red dashed-dotted arrow). (ii) The amplitude-phase diagram of Fig. 22(b) proposes that spontaneous emission occurs within a limited phase angle but with a magnitude of the emitted electric field that is adjusted, such that, compared to the incident electric field, the resulting electric field comprises an additional energy that exactly matches the emitted photon. (iii) Quantum-optically stimulated and spontaneous emission are both described by the same creation operator [68], hence either (iii,a) both processes must occur with the same phase difference or (iii,b) the phase difference is not explicitly considered when applying the creation operator. (iv) The semi-classical Lorentz oscillator model [69] predicts that stimulated emission is in quadrature, i.e., 90° out of phase with the incident field [70–73].

As is often the case in science, we must carefully distinguish between different, yet compatible explanations and explanations that exclude each other. Fundamentally, it is clear that these four models cannot simultaneously be correct. The only exceptions are that model (iii) in its interpretation (a) would not exclude model (iv) and in its interpretation (b) would not exclude models (ii) and (iv).

Despite their obvious incompatibility concerning the phase aspect, all four models have been applied to understand important optical phenomena. Model (ii) can be considered a kind of standard model in contemporary quantum optics. Model (iv) has been exploited to derive the Kramers-Kronig relations [74,75] between susceptibility and absorption. In a simple rate-equation approach equivalent to Einstein's [59] without phase considerations, i.e., in line with model (iii,b), the power behavior of semiconductor lasers around the laser threshold has been explained [76–79]. A combination of models (iii) and (i) has been applied to calculate quantum-mechanically the fundamental laser linewidth and its reduction compared to the Schawlow-Townes linewidth [10] by an additional factor of two around laser threshold [25,26,65–67]. Model (i) has served to justify this reduction of laser linewidth around the laser threshold [25,26,64]. According to Fig. 22(a), spontaneous emission is supposed to induce amplitude and phase fluctuations (projection onto the axes with $\theta_a = 0^\circ$ – 180° and $\theta_p = \pm 90^\circ$, respectively), of which the latter constitute the quantum noise that determines the fundamental laser linewidth, whereas the former are damped out in a laser, thereby reducing the laser linewidth by a factor of 2. It has also served to derive Henry's α -factor, which quantifies broadening of the laser linewidth due to amplitude-phase coupling via the refractive index [64].

Since these four incompatible versions seem to describe various optical phenomena, does then the phase aspect matter at all? The arguments presented in this Section suggest that the phase is of fundamental importance and, consequently, there must not occur an ambiguity concerning the phase difference. By exploiting Maxwell's equations and the law of energy conservation, investigating stimulated emission in a Fabry-Pérot resonator, analyzing the Lorentz oscillator model, applying the Kramers-Kronig relations to the complex susceptibility, understanding the summation of quantized electric fields, and quantitatively interpreting emission and absorption in the amplitude-phase diagram, we derive a consistent semi-classical picture of the phase aspect in stimulated and spontaneous emission, as well as absorption.

5.1. Semi-classical versus quantum-optical description

How far a quantum-optical treatment can confirm or conflicts with this semi-classical picture is not a subject of this paper. Nevertheless, a few fundamental remarks on this matter deem appropriate. Emission and absorption of a photon are quantized processes. When judging the present work from a quantum-optical point of view, one should keep in mind the following points. (a) Only on a sufficiently short time scale can a process violate the law of energy conservation according to the uncertainty principle [80,81]. At longer time scales, the law of energy conservation applies in quantum mechanics. (b) The classical Maxwell equations [82] maintain their full validity in quantum optics, but with the additional requirement of a quantization of optical energy, as was demonstrated by Dirac [62]. (c) The amplitude-phase diagram is not a sloppy way of sketching some processes but can—and should—be understood as a quantitative vectorial description. (d) Einstein's semi-classical rate-equation approach [59] not only confirmed Planck's law [34], but also delivered the Einstein A and B coefficients of spontaneous and stimulated emission, which were found entirely consistent with a full

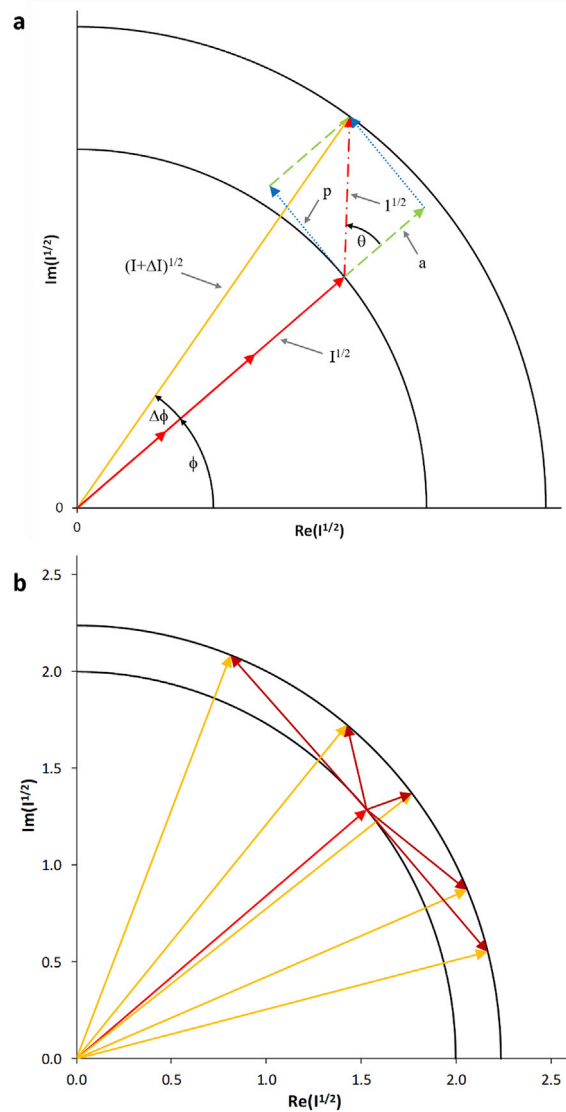


Fig. 22. (a) Amplitude-phase diagram visualizing the interpretation of quantum noise and laser linewidth by Lax [25], Haken [26], and Henry [64]. Quantum noise in a laser is said to be induced by adding with an arbitrary phase difference θ a spontaneously emitted photon (red dashed-dotted arrow) of intensity 1 to the intra-cavity laser field (red solid arrows) of intensity I and phase ϕ , resulting in an intra-cavity laser field of intensity $I + \Delta I$ (orange solid arrow) and inducing a phase shift $\Delta\phi$. Below laser threshold all phase differences θ are proposed to generate noise, whereas, above laser threshold, amplitude fluctuations a ($\theta = 0$ or π , i.e., the projection of noise onto the direction of the green dashed arrow) are rapidly damped out by relaxation oscillations, and only phase fluctuations p ($\theta = \pm\pi/2$, i.e., the projection of noise onto the direction of the blue dotted arrow) contribute to noise, thereby reducing the laser linewidth compared to the Schawlow-Townes linewidth [10] by a factor of 2 [25,26,64–67]. (b) Amplitude-phase diagram visualizing the interpretation of quantum noise that allows a limited range of phase angles and an accordingly adjusted magnitude of the emitted electric field, such that the energy is conserved. (Figure (a) taken from Ref. [3]).

quantum-mechanical treatment; see, e.g., Schiff [83]. (e) The Kramers-Kronig relations [74,75] are bidirectional mathematical relations between the real and the imaginary part of any complex function that is analytic in the upper half of the complex plane. The analyticity condition is a consequence of causality in physical systems. Consequently, the Kramers-Kronig relations apply to the complex susceptibility, semi-classically as well as quantum-optimally. (f) The Lorentz oscillator model [69] is an approximation to the quantum theory that is equivalent to the standard perturbation-theoretical approach to absorption and stimulated emission, see, e.g., Schiff [84], in which a weak oscillator strength is assumed, such that the transition rate is essentially constant and the atom is still in its initial state (upper or lower state of the transition) after some time interval that is long compared to the oscillation period (Fermi's golden rule [85], originally formulated by Dirac [62]). (g) Despite its obvious crudeness, the Lorentz oscillator model has enormous prediction power in optics, as was pointed out by Feynman et al. [86] and Weisskopf [87]. (h) Based upon the Lorentz oscillator model, stimulated emission and absorption may be understood in the semi-classical approach of treating the atom, which in the simplest case is considered a

two-level system, quantum mechanically, and the electromagnetic field, which is in near resonance with the transition between the two atomic levels, classically, as was emphasized by Milonni et al. [71,72]. Therefore, it is not *a priori* clear that semi-classical models insufficiently describe quantized emission and absorption processes.

5.2. Maxwell's equations, interference, and conservation of energy

Based upon Maxwell's equations [82], superposition of two co-propagating electromagnetic waves at the same frequency ν , with electric-field amplitudes E_1 and E_2 and a phase difference θ , yields

$$|\vec{E}_1 + \vec{E}_2|^2 = |E_1|^2 + 2|E_1 E_2| \cos(\theta) + |E_2|^2 \stackrel{\cos(\theta)=0}{=} |E_1|^2 + |E_2|^2. \tag{123}$$

Since the intensity is related to the electric field by Eq. (25), we can rewrite Eq. (123) as

$$I_{1+2} = I_1 + 2\sqrt{I_1 I_2} \cos(\theta) + I_2 \stackrel{\cos(\theta)=0}{=} I_1 + I_2, \tag{124}$$

and since the photon number is related to the intensity by Eq. (27), we can rewrite Eq. (124) as

$$\varphi_{1+2} = \varphi_1 + 2\sqrt{\varphi_1 \varphi_2} \cos(\theta) + \varphi_2 \stackrel{\cos(\theta)=0}{=} \varphi_1 + \varphi_2. \tag{125}$$

For $\cos(\theta) \neq 0$, the interference term does not vanish and the law of energy conservation is obviously violated in all three equations above. Two fields with $E_1 = E_2$ constructively (destructively) interfere to produce four (zero) times the intensity of each field alone. When adding a field representing one photon ($E_2 \propto 1^{1/2} = 1$) to a field representing 100 photons ($E_1 \propto 100^{1/2} = 10$), then averaged over all phase angles, $\overline{\cos(\theta)} = 0$, the expected energy of 101 photons emerges, whereas constructive (destructive) interference yields the energy of 121 (81) photons; see Fig. 23. Nevertheless, energy can be conserved, if there is a concrete source or drain of energy in the system, for example, enhanced scattering of a perpendicular external light source into a resonant cavity [88]. In the absence of such a source or drain, energy can only be conserved if the opposite interference occurs in another location of the optical system, i.e., constructive or destructive interference never occurs alone!

Let us analyze the processes in Fig. 22(a) with respect to energy conservation. If stimulated emission occurred in phase with the incident field, $\cos(\theta) = 1$, then according to Eq. (123) each stimulated-emission event would generate an excess of photons, thereby violating the law of energy conservation. If spontaneous emission occurred at an arbitrary phase angle, then, according to Eq. (123), each spontaneous-emission event would either generate or annihilate extra photons (Fig. 23), and in the amplitude-phase diagram (Fig. 22(a)) the added intensity ΔI would not correspond to the intensity generated by one photon. As pointed out by Henry [64], only when averaging over many spontaneous-emission events, $\overline{\cos(\theta)} = 0$, energy would be conserved (green dashed line in Fig. 23). In a lasing resonator, such a violation of energy conservation by stimulated and spontaneous emission could be experimentally manifested by cavity dumping of the stored optical energy within a single resonator round trip. Since energy must be conserved, this interpretation of stimulated and spontaneous emission is obviously questionable. The assumption that the energy difference magically comes out of or dissipates into the universe can safely be discarded.

Presumably to circumvent the obvious violation of energy conservation, an alternative physical picture, displayed in Fig. 22(b), has been proposed in quantum optics. An arbitrary phase angle occurs between the incident and resulting field. The difference in energy between the incident and resulting field amounts exactly to one photon. However, the assumption that the emitted electric field

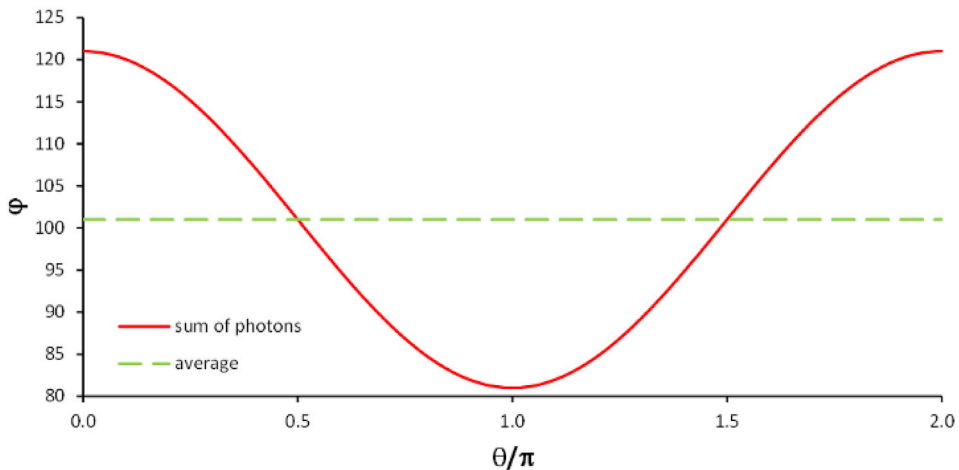


Fig. 23. Number φ of photons resulting from the interference according to Eq. (123) between one photon and 100 photons (red solid curve) versus phase difference θ and medium of 101 photons averaged over all θ (green dashed line). (Figure taken from Ref. [3]).

magically adjusts its magnitude such that the energy is conserved violates Maxwell's equations. Besides, this picture cannot explain the Schawlow-Townes linewidth. In summary, the physical picture of Fig. 22(a) obeys Maxwell's equations but violates the law of energy conservation, whereas the physical picture of Fig. 22(b) obeys the law of energy conservation but violates Maxwell's equations. Neither is acceptable.

5.3. Stimulated emission in a Fabry-Pérot resonator

The problem of energy conservation in stimulated emission manifests itself in a fundamentally important optical system, the Fabry-Pérot resonator, which we have discussed in detail in Sections 2-4. In Fig. 24(a), a monochromatic external light source continuously launches light into a resonator whose only losses are the outcoupling losses through its two mirrors. A steady state of light launched into, circulating inside, and emitted from the resonator is established, described by the Airy distributions

$$\begin{aligned}
 A_{circ} &= I_{circ}/I_{laun} = \frac{1}{(1 - \sqrt{R_1 R_2})^2 + 4\sqrt{R_1 R_2} \sin^2(\Delta\phi_{RT}/2)} \\
 A'_{trans} &= I_{trans}/I_{inc} = (1 - R_1)(1 - R_2)A_{circ} \\
 A'_{back} &= I_{back}/I_{inc} = (1 - R_1)^2 R_2 A_{circ} \\
 A'_{refl} &= I_{refl}/I_{inc} = [(\sqrt{R_1} - \sqrt{R_2})^2 + 4\sqrt{R_1 R_2} \sin^2(\Delta\phi_{RT}/2)] A_{circ} \\
 A'_{trans} + A'_{refl} &= (I_{trans} + I_{refl})/I_{inc} = 1.
 \end{aligned}
 \tag{126}$$

$\Delta\phi_{RT}$ is the phase shift accumulated over one round trip. In the situation of Fig. 24(a), it is identical to 2ϕ from Eq. (3). The different electric fields are displayed in Fig. 24(a). For arbitrary $\Delta\phi_{RT}$ the energy is conserved, $A'_{trans} + A'_{refl} = 1$, see Eq. (95), because the interference between E_{RT} and E_{laun} is compensated by the opposite interference between $E_{refl,1}$ and E_{back} .

Now we move the light source into the resonator [Fig. 24(b)]. A light source inside the mode must be transparent in order not to block the propagating light; consequently, the propagating light will penetrate and interact with the light source. Let us assume that the light source is a pumped inverted medium (an atomically thin gain sheet placed close to mirror 1, oriented perpendicular to the resonator axis) that continuously generates E_{gen} via stimulated emission triggered by E_{RT} , resulting in the combined field E_{circ} . For simplicity, we neglect spontaneous emission and assume that several stimulated-emission processes can occur simultaneously in different lateral regions of the gain sheet without influencing each other. Furthermore, for comparison with Fig. 24(a), we assume a race-track resonator with unidirectional light propagation, such that the backward-circulating field E_{b-circ} does not penetrate the active medium and the stimulated-emission process is uni-directional. If stimulated emission occurred "in phase" and we pumped the medium to a desired inversion, such that in Fig. 24(a) and (b) the same field $E_{gen} = E_{laun}$ interfered constructively with E_{RT} at the resonance frequency ν_q , then the same steady state of light generated, circulating inside, and emitted from the resonator would be established and

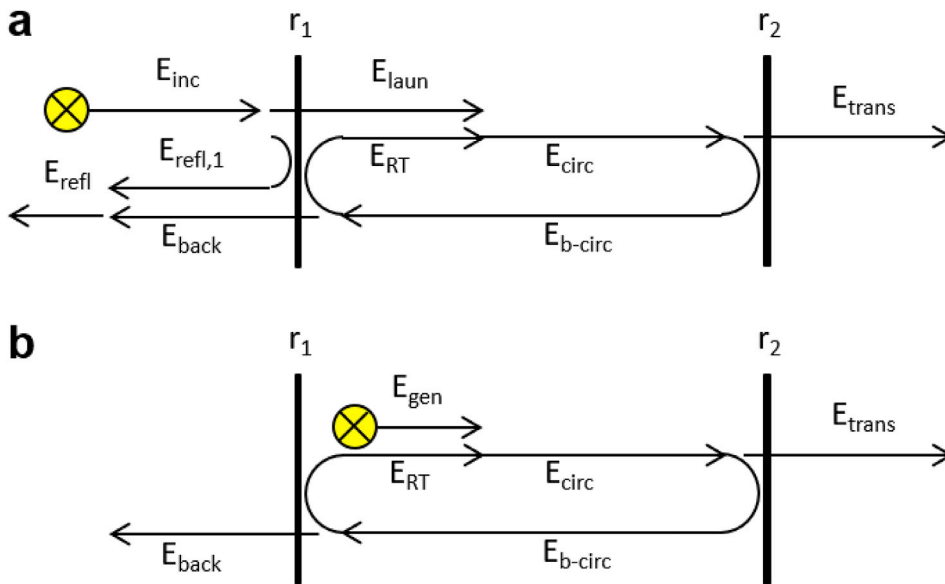


Fig. 24. Schematic of a Fabry-Pérot resonator and the relevant electric fields E for (a) light launched from outside [2] and (b) light generated inside the resonator. (Figure taken from Ref. [3]).

the same fields E_{trans} and E_{back} would be emitted in both situations. However, since in Fig. 24(b) E_{back} cannot destructively interfere, because there is no $E_{refl,1}$, the constructive interference is not compensated for. Therefore, if the phase shift $\Delta\phi_{em}$ potentially induced by stimulated emission between E_{RT} and E_{circ} equals zero, energy is not conserved:

$$\begin{aligned}
 A_{circ} &= I_{circ}/I_{gen} = \frac{1}{(1 - \sqrt{R_1 R_2})^2 + 4\sqrt{R_1 R_2} \sin^2[(\Delta\phi_{RT} + \Delta\phi_{em})/2]} \\
 A_{trans} &= I_{trans}/I_{gen} = (1 - R_2)A_{circ} \\
 A_{back} &= I_{back}/I_{gen} = (1 - R_1)R_2 A_{circ} \\
 A_{emit} &= A_{trans} + A_{back} = (I_{trans} + I_{back})/I_{gen} = (1 - R_1 R_2)A_{circ} \\
 A_{emit} &> 1 \quad \text{for} \quad \Delta\phi_{RT} = \Delta\phi_{em} = 0.
 \end{aligned}
 \tag{127}$$

A_{trans} and A_{back} are the Airy distributions of I_{trans} and I_{back} , respectively, here with respect to I_{gen} instead of I_{laun} . Their sum A_{emit} is displayed in Fig. 25(a) as a function of $(R_1 R_2)^{1/2}$. Since light builds up inside the resonator and stimulates emission around the resonance frequency ν_q (either in a broadband gain medium or by tuning the resonator length, such that the resonance frequency ν_q coincides with the emission frequency), $\Delta\phi_{RT}$ becomes a multiple of 2π . Consequently, the phase shift indicated in the legend of Fig. 25(a) is solely due to $\Delta\phi_{em}$. Energy conservation requires $A_{emit} = 1$ (black line).

If stimulated emission occurred in phase, $\Delta\phi_{em} = 0$ [solid gray curve in Fig. 25(a)], the law of energy conservation would be violated;

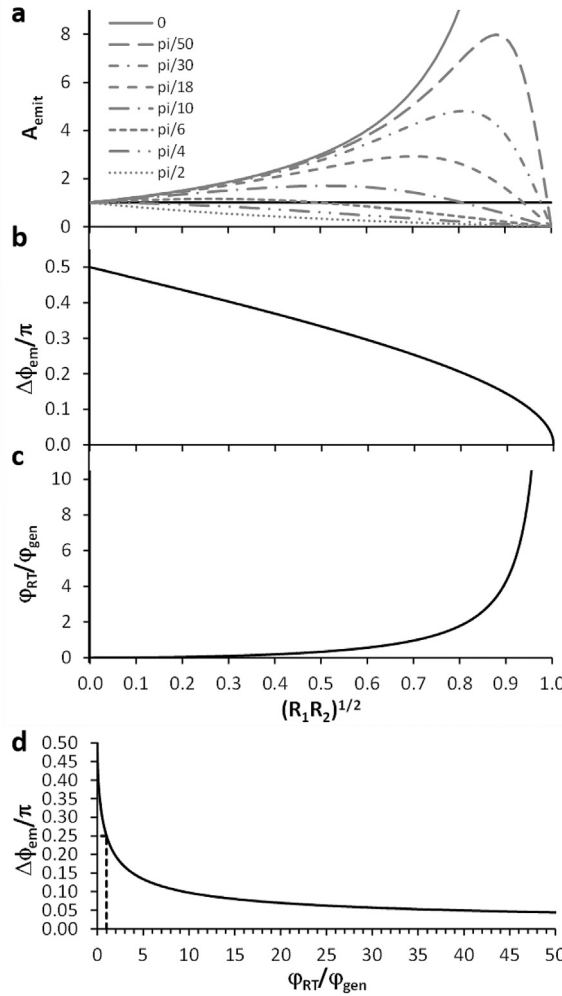


Fig. 25. (a) Requirement of energy conservation, $A_{emit} = 1$ (black line), and violation of energy conservation for different potential phase shifts $\Delta\phi_{em} = 0, \pi/50, \pi/30, \pi/18, \pi/10, \pi/6, \pi/4,$ and $\pi/2$ (see legend) from the incident field E_{RT} to the resulting field E_{circ} induced by interference of E_{RT} with the field E_{gen} generated by stimulated emission, (b) the phase shift $\Delta\phi_{em}$ that is required to obtain energy conservation, and (c) the ratio $\varphi_{RT}/\varphi_{gen}$ of incident photon number φ_{RT} over generated photon number φ_{gen} as a function of $(R_1 R_2)^{1/2}$. (d) Phase shift $\Delta\phi_{em}$ induced by stimulated emission as a function of the ratio $\varphi_{RT}/\varphi_{gen}$. For $\varphi_{RT}/\varphi_{gen} = 1$, one obtains $\Delta\phi_{em} = \pi/4$ (dashed lines). (Figure taken from Ref. [3]).

e.g., for $R_1 = R_2 = 0.7$, $A_{emit} = 5.7$ times the light generated by stimulated emission would be emitted through both mirrors. For all phase shifts $\Delta\phi_{em} \neq 0$ energy is also not conserved, except for one specific value of $\Delta\phi_{em}$ for each value of $(R_1R_2)^{1/2}$; see Fig. 25(a).

Assuming resonance, i.e., $\Delta\phi_{RT}$ is a multiple of 2π , to ensure energy conservation, i.e., $A_{emit} = 1$, the condition

$$\begin{aligned}
 A_{emit} &= (1 - R_1R_2)A_{circ} = \frac{1 - R_1R_2}{(1 - \sqrt{R_1R_2})^2 + 4\sqrt{R_1R_2}\sin^2[(\Delta\phi_{RT} + \Delta\phi_{em})/2]} = 1 \\
 \Rightarrow 2\sin^2(\Delta\phi_{em}/2) &= 1 - \sqrt{R_1R_2} \Rightarrow \cos(\Delta\phi_{em}) = \sqrt{R_1R_2} \\
 \Rightarrow \tan(\Delta\phi_{em}) &= \frac{\sin(\Delta\phi_{em})}{\cos(\Delta\phi_{em})} = \sqrt{\frac{1 - \cos^2(\Delta\phi_{em})}{\cos^2(\Delta\phi_{em})}} = \sqrt{\frac{1 - R_1R_2}{R_1R_2}}
 \end{aligned} \tag{128}$$

must be fulfilled; see Fig. 25(b). From this condition, we derive that, in order to obtain energy conservation, a phase shift $\Delta\phi_{em}$ induced by stimulated emission is required that differs from zero and depends on R_1 and R_2 . We convert this reflectivity dependence to a photon dependence. Since according to Eq. (27) each intensity is proportional to the corresponding photon number, the ratio between the numbers φ_{RT} of photons triggering stimulated emission and φ_{gen} of photons generated by stimulated emission equals the Airy distribution A_{RT} , from which we derive the phase shift as a function of photon numbers,

$$\begin{aligned}
 \varphi_{RT}/\varphi_{gen} &= I_{RT}/I_{gen} = A_{RT} = R_1R_2A_{circ} = \frac{R_1R_2}{(1 - \sqrt{R_1R_2})^2 + 4\sqrt{R_1R_2}\sin^2[(\Delta\phi_{RT} + \Delta\phi_{em})/2]} = \frac{R_1R_2}{(1 - \sqrt{R_1R_2})^2 + 2\sqrt{R_1R_2}(1 - \sqrt{R_1R_2})} = \frac{R_1R_2}{1 - R_1R_2} \\
 \Rightarrow \tan(\Delta\phi_{em}) &= \sqrt{\frac{1 - R_1R_2}{R_1R_2}} = \sqrt{\varphi_{gen}/\varphi_{RT}},
 \end{aligned} \tag{129}$$

where we used $\Delta\phi_{RT} = 0$ and Eq. (128). The ratio $\varphi_{RT}/\varphi_{gen}$ as a function of $(R_1R_2)^{1/2}$ is displayed in Fig. 25(c) and the final result of Eq. (129) is shown in Fig. 25(d). Only for an infinite ratio $\varphi_{RT}/\varphi_{gen}$ is the phase shift $\Delta\phi_{em}$ induced by stimulated emission equal to zero. The smaller the number of photons triggering stimulated emission and the larger the number of photons generated, the further the induced phase shift $\Delta\phi_{em}$ increases to $\pi/2$. If one photon triggers stimulated emission of one photon, then $\Delta\phi_{em} = \pi/4$ [dashed lines in Fig. 25(d)]. Since the second angle in the vector diagram comprising the fields E_{RT} , E_{gen} , and E_{circ} is also known, the phase difference between incident and emitted field equals

$$\theta = \pi - \arctan\sqrt{\varphi_{gen}/\varphi_{RT}} - \arctan\sqrt{\varphi_{RT}/\varphi_{gen}} = \pi/2. \tag{130}$$

In resonance, stimulated emission occurs with a phase difference of $\theta = 90^\circ$ between driving and generated field, its direct consequence being the phase shift $\Delta\phi_{em}$ of Eq. (129) between driving and transmitted field.

For this derivation we have only assumed the validity of Maxwell's equations and the law of energy conservation. Although the curve in Fig. 25(d) is continuous, one can easily impose a quantization of energy by allowing only integer values of φ_{RT} and φ_{gen} in the ratio $\varphi_{RT}/\varphi_{gen}$.

5.4. Lorentz oscillator model and Kramers-Kronig relations

The Lorentz oscillator model describes the motion of electrons with electric charge e and mass m_e , bound as a cloud with electron density N_e within an atom, as a damped harmonic oscillation with an angular resonance frequency ω_0 and gain/damping rate constant γ_e (positive for stimulated emission, negative for absorption) displaced by a distance x from its rest position by an external driving electric field E_{ext} oscillating with angular frequency ω_{ext} . The situation is illustrated in Fig. 26. The magnetic force is neglected. Its mathematical treatment is equivalent to that of a mechanical spring oscillator. The linear second-order ordinary differential equation of motion is solved, yielding the atomic polarization P_e and phase difference θ between the driving electric field and the polarization:

$$\begin{aligned}
 m_e\ddot{x}(t) + 2\gamma_e m_e\dot{x}(t) + \omega_0^2 m_e x(t) &= -eE_{ext} \exp[-i\omega_{ext}t] \\
 P_e = -N_e e x(t) &= \frac{N_e e^2/m_e}{\omega_0^2 - \omega_{ext}^2 - i2\gamma_e\omega_{ext}} E_{ext} \exp[-i(\omega_{ext}t - \theta)] \\
 \tan(\theta) &= \frac{2\gamma_e\omega_{ext}}{\omega_0^2 - \omega_{ext}^2}.
 \end{aligned} \tag{131}$$

The same phase difference θ as from the Lorentz oscillator model in Eq. (131) obtains from the Kramers-Kronig relations [74,75] between the real part (susceptibility) χ'_e and the imaginary part (gain or absorption) χ''_e of the complex susceptibility χ_e [Fig. 27(a)–(d)]:

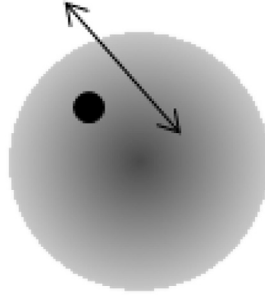


Fig. 26. Lorentz oscillator model: positive core (black spot) and negative oscillating electron cloud (gray sphere).

$$\begin{aligned}
 P_e &= \epsilon_0 \chi_e E_{ext} \exp[-i(\omega_{ext}t - \theta)] \\
 \chi_e &= \chi_0 \frac{\omega_0^2}{\omega_0^2 - \omega_{ext}^2 - i2\gamma_e \omega_{ext}}, \quad \chi_0 = \frac{N_e e^2}{\epsilon_0 m_e \omega_0^2} \\
 \chi_e' &= \chi_0 \frac{-(\omega_0^2 - \omega_{ext}^2)\omega_0^2}{(\omega_0^2 - \omega_{ext}^2)^2 + (2\gamma_e \omega_{ext})^2} \\
 \chi_e'' &= \chi_0 \frac{2\gamma_e \omega_{ext} \omega_0^2}{(\omega_0^2 - \omega_{ext}^2)^2 + (2\gamma_e \omega_{ext})^2} \\
 \tan(\theta) &= \frac{\chi_e'}{\chi_e''} = \frac{2\gamma_e \omega_{ext}}{\omega_0^2 - \omega_{ext}^2}.
 \end{aligned} \tag{132}$$

χ_0 is the frequency-independent part of the complex susceptibility.

As is well known from mechanical oscillators, when the driving frequency ω_{ext} is significantly lower (higher) than the resonance frequency ω_0 , the oscillation is in (out of) phase with the driving field. In resonance, the phase difference θ crosses the value of $\pi/2$; see Fig. 27(c). Very recently, Sommer et al. [89] found experimentally in the interaction of an attosecond pulse with an atom the following signatures. When the relative phase of an atomic polarization is in lead of the electromagnetic field, energy is transferred from the atom to the field (stimulated emission), whereas when the phase of the atomic polarization lags behind the electromagnetic field, energy is transferred from the field to the atom (absorption). This experimental finding confirms the Lorentz oscillator model and the Kramers-Kronig relations. This is the first part of a stimulated-emission or -absorption process.

The second part is the emission of an electromagnetic wave by this oscillating electron cloud. It is sometimes argued that the processes of (a) an oscillating electromagnetic field driving the oscillation of an atom and (b) an oscillating atom driving the oscillation of an electromagnetic field are equivalent and can, therefore, be described by the same Eq. (131) or (132). This argument is, in principle, correct. However, the apparent conclusion that, if the oscillating atom experiences a $\pi/2$ phase shift with respect to the driving oscillating field, also the emitted oscillating field must experience a $\pi/2$ phase shift with respect to the driving oscillating atom, thereby adding up to a total phase shift of $\theta = 0$ in stimulated emission and of $\theta = \pi$ in absorption, is incorrect. The simple reason for the non-equivalence of these two processes is that in the former case the electron with mass m_e is the driven oscillator, whereas in the latter case the photon is the driven oscillator. The photon has a rest mass of zero and a relativistic mass of $m_{ph} = h\nu/c^2$. For visible and near-infrared frequencies, m_{ph} is orders of magnitude smaller than m_e , hence its resonance frequency is orders of magnitude larger than ω_0 . If one indeed applies Eqs. (131) and (132) to the case of the oscillating electron driving the photon, the electron mass m_e oscillates with the resonance frequency ω_0 of the atomic system. The lateral oscillation of the photon is then driven at that same frequency ω_0 , which is much smaller than the resonance frequency of the photon, hence Eqs. (131) and (132) predict that the emitted oscillating field must be practically in phase with the driving atom. The phase of radiation emitted by an oscillating electric dipole has been derived in classical electrodynamics. Whereas the near-field radiation of an oscillating electric dipole is rather complex [90,91], with the phases of longitudinal and transverse components of the electric field changing with increasing distance from the dipole, the far-field radiation emitted by an atomic dipole is dominated by the transverse component and is in phase with the atomic dipole oscillation [90–92], as predicted by Eqs. (131) and (132).

By combining the above-described parts one and two of the process of stimulated emission, this derivation confirms that in stimulated emission the emissive part of the generated dipole field is in quadrature with the driving electromagnetic field [70–73], i.e., the total phase difference amounts to $\theta = 90^\circ$. Bruhat and Kastler [93,94], as well as Bertolotti [95], also emphasized the necessity of a 90° phase difference in stimulated emission in order for the energy to be conserved. Likewise, in resonant absorption the phase difference between the atomic oscillation and the transmitted field is $\theta = -\pi/2$, i.e., the atomic oscillation lags 90° behind.

Consequently, in a resonant stimulated-emission process the interference term in Eqs. (123)–(125) vanishes, the two individual intensities add up, and the energy is conserved. For exactly this reason, (i) Einstein was allowed to neglect interference in his semi-classical rate-equation derivation [59] of Planck's law [34] and (ii) laser performance can be described—and important laser parameters, such as threshold and slope efficiency, can be obtained—with a distributed-intensity-gain coefficient g by a differential rate equation

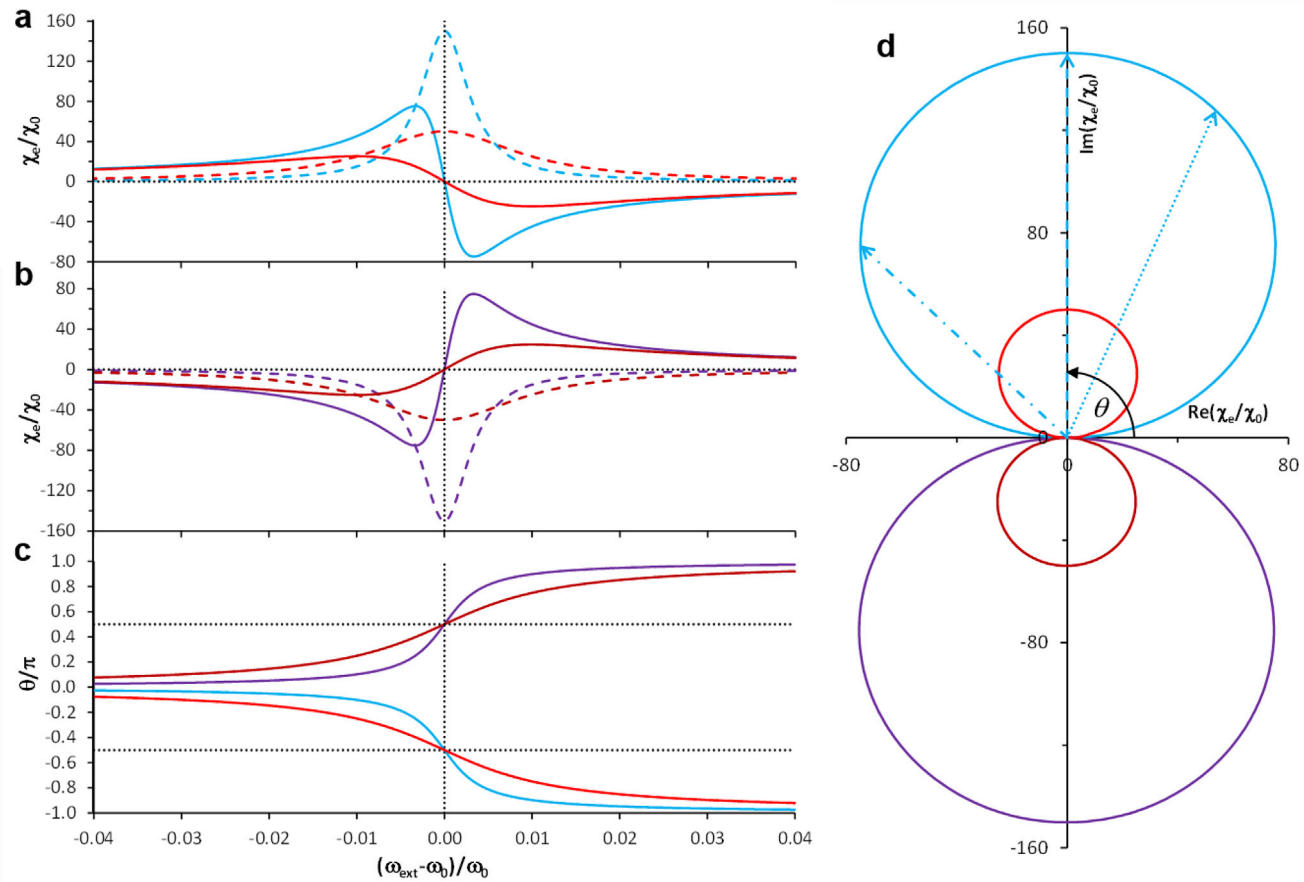


Fig. 27. Real part χ_e' (solid lines) and imaginary part χ_e'' (dashed lines) of the susceptibility, calibrated to χ_0 , for $\gamma_e = \pm 0.00333 \omega_0$ (blue curves) and $\gamma_e = \pm 0.01 \omega_0$ (red curves) in (a) stimulated emission and (b) absorption. (c) Phase difference θ between the amplitudes of a driven atomic oscillator and its driving electric field as a function of driving frequency. (d) Complex susceptibility χ_e , calibrated to χ_0 , as a function of ω_{ext} for the four examples displayed in (a) and (b). For the examples of $\gamma_e = \pm 0.00333 \omega_0$, the arrows indicate the situations of $(\omega_{\text{ext}} - \omega_0)/\omega_0 = -1.38 \times 10^{-3}$, 0 (resonance), and 3.34×10^{-3} , resulting in $\theta = 3/8\pi$ (dotted arrow), $\pi/2$ (resonance, dashed arrow), and $3/4\pi$ (dashed-dotted arrow), respectively. The phase difference in resonance of $\theta = \pi/2$ is indicated by the black curved arrow, which points in the direction of increasing ω_{ext} . (Figure taken from Ref. [3]).

that calculates the photon number φ but neglects interference, see later in Eq. (155). The quantum-optical laser master equation also neglects interference [96].

5.5. Amplitude-phase diagram and quantized electric fields

Fig. 28 and several subsequent figures display amplitude-phase diagrams. The coordinate axes are in units of the electric-field amplitude, calibrated to the square root of the number φ of photons at the frequency ν that is resonant with the atomic transition. Consequently, the quarter circles denote the amplitudes of electric fields comprising 1 (dark-red), 2 (red), 3 (orange), 4 (yellow), etc. photon energies $h\nu$. The x - and y -axes display the real and imaginary part of the amplitude, respectively. The coordinate system rotates with $e^{i2\pi\nu t}$. For φ_{ext} incident photons triggering the emission of $\varphi_{gen} = 1$ photon, the situation of $\theta = \pi/2$ is illustrated in the amplitude-phase diagram of Fig. 28(a). With $\varphi_{ext} = \varphi_{RT}$ in Eq. (129), we find that the resulting phase shift $\Delta\phi_{em}$ between the incident and transmitted electric field is

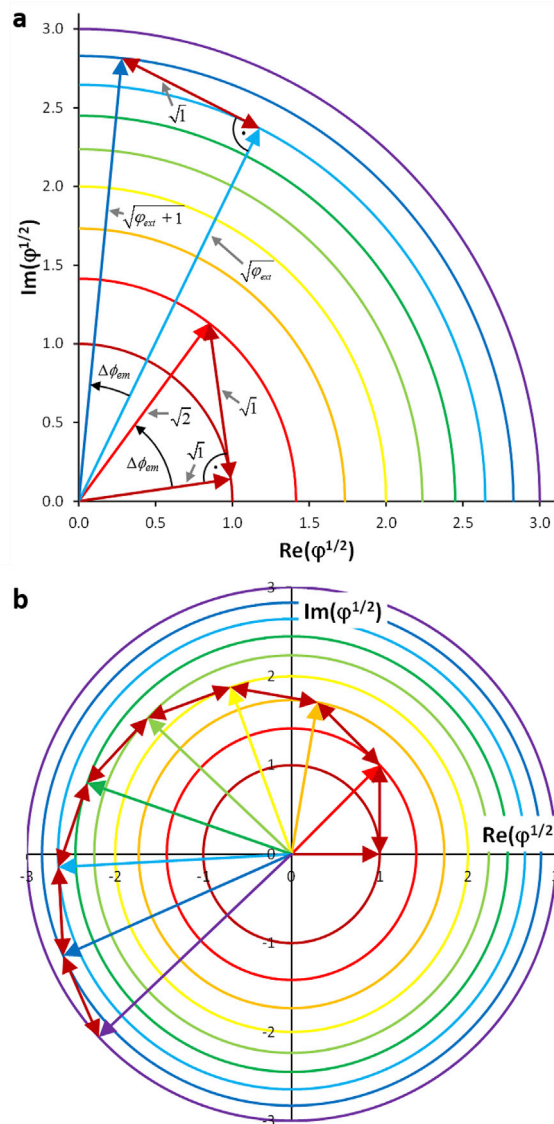


Fig. 28. (a) Quadrant of the amplitude-phase diagram illustrating the process of stimulated emission (with the dark-red arrows of the emitted field pointing towards the upper left): an incident field of φ_{ext} photons triggers an atom in its excited state to emit $\varphi_{em} = 1$ photon, in the two situations of (i) $\varphi_{ext} = 1$ and (ii) $\varphi_{ext} = 7$. In both situations, the indicated right angle is $90^\circ = 180^\circ - \theta$, hence $\theta = 90^\circ$. The color code denotes the amplitude in units of $\varphi^{1/2}$, from $\varphi = 1$ photon (dark red) to $\varphi = 9$ photons (violet). The same diagram holds true for absorption (with the dark-red arrows of the absorbed field pointing toward the lower right). (b) Build-up of a light beam by the consecutive addition of single photons in the amplitude-phase diagram. (Figure taken from Ref. [3]).

$$\tan(\Delta\phi_{em}) = \sqrt{1/\varphi_{ext}}. \tag{133}$$

Assuming an incident electromagnetic field containing the energy of $\varphi_{ext} = 1$ photon, the build-up of a larger electromagnetic field by consecutive stimulated emission of electromagnetic fields, each containing the energy of $\varphi_{gen} = 1$ photon, is displayed in Fig. 28(b). The total phase shift accumulated by the consecutive stimulated emission of $n - 1$ photons by one initial photon with an arbitrary phase, resulting in an electromagnetic wave containing n photons, amounts to

$$\Delta\phi_n = \sum_{i=1}^{n-1} \Delta\phi_{em,i} = \sum_{i=1}^{n-1} \arctan\left(1/\sqrt{i}\right) \text{ for } n \geq 1. \tag{134}$$

The total phase shift of Eq. (134) establishes a relation among all these states.

Simultaneous independent stimulated emission of several photons, $\varphi_{gen} > 1$, induces a phase shift $\Delta\phi_{em}$ in Eq. (129) that is smaller than the phase shift $\Delta\phi_n$ of Eq. (134) accumulated by consecutive stimulated emission of single photons (Fig. 29):

$$\Delta\phi_{em}(\varphi_{em} > 1) < \Delta\phi_n(n = \varphi_{em} > 1). \tag{135}$$

For investigating the Fabry-Pérot resonator in Sub-section 5.3, we chose an atomically thin gain sheet and assumed a simultaneous independent stimulated emission of several photons into the same mode. It is an interesting question whether this assumption is physically justified. If true, the total phase shift induced when building up a light beam depends on the way the photons are generated, consecutively or simultaneously. If not true, i.e., the simultaneous emission of several photons within the same mode is correlated, such that these photons must obey the law of energy conservation also with respect to each other, then the total phase shift of Eq. (134) establishes a unique relation between all photon numbers φ . Such a correlation could lead to the phenomenon of superradiance [97–99].

One should not fall into the trap of believing that this relative phase shift might be arbitrary. Of course, there is an arbitrary starting

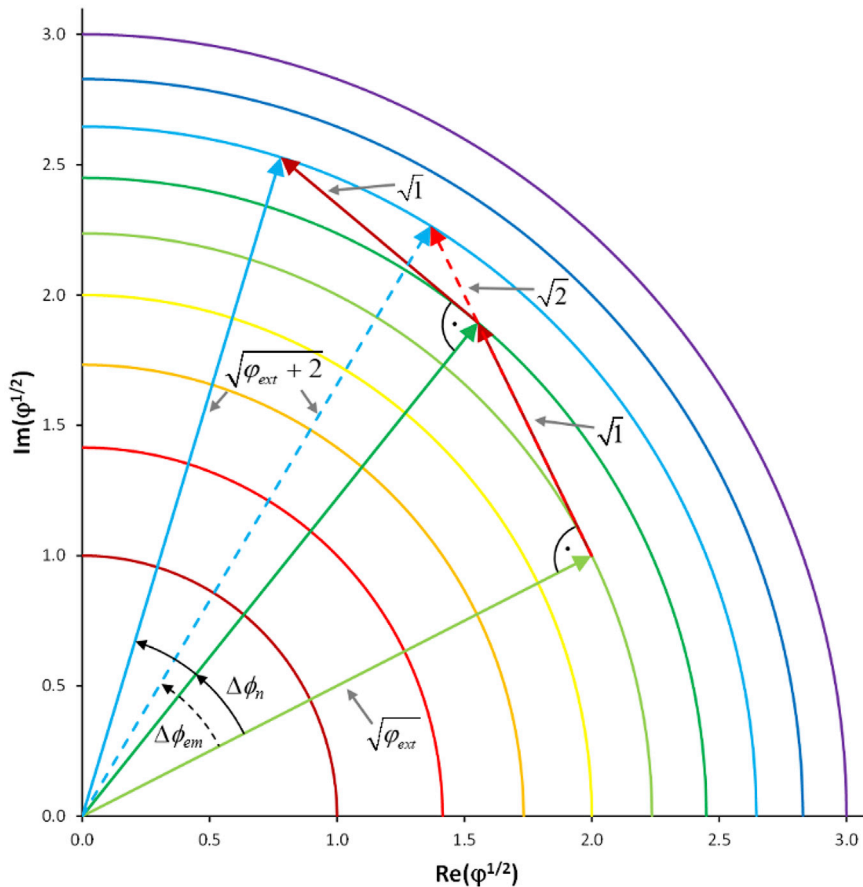


Fig. 29. Quadrant of the amplitude-phase diagram comparing the simultaneous, independent addition of several photons (dashed arrows; here, an incident field of $\varphi_{ext} = 5$ photons triggers atoms in their excited state to simultaneously emit $\varphi_{em} = 2$ photons), resulting in a phase shift $\Delta\phi_{em}$, with the consecutive addition of 2 single photons (solid arrows), resulting in an accumulated phase shift $\Delta\phi_n > \Delta\phi_{em}$. (Figure taken from Ref. [3]).

phase. By rotating the coordinate system by an appropriate fixed amount, the starting phase is set to zero in Fig. 28(b), whereas, in the examples of Fig. 28(a), it is assumed different from zero. In addition, the coordinate system can be rotated time-dependently. In fact, in Fig. 28 the real and imaginary axes are rotated by the oscillatory term $e^{i2\pi\nu t}$, such that the displayed arrows that rotate with this angular speed stand still in the graph. Beyond that, no further freedom exists in this semi-classical treatment.

When an incident electromagnetic field that was generated by individual atomic emission processes triggers stimulated emission of another electromagnetic field by an atom, it requires only two ingredients for a correct quantization, namely, (i) each individual amplitude must be proportional to the square root $\sqrt{\varphi}$ of an integer photon number φ , and (ii) the phase difference must be $\theta = 90^\circ$, the automatic consequence being the phase shift $\Delta\phi_{em}$ of Eq. (129) between driving and transmitted field. An example of $\varphi_{ext} = 16$ photons triggering the emission of $\varphi_{gen} = 1$ photon is illustrated in Fig. 30(a), its result being quantitatively equivalent to the same process when displayed in the amplitude-phase diagram, as shown for other photon numbers Fig. 28(a). The build-up of an electromagnetic field containing φ photons by consecutive stimulated emission of electromagnetic fields containing the energy of single photons is displayed in Fig. 30(b), its result being quantitatively equivalent to Fig. 28(b).

Recent progress in another field dealing with resonant photonic systems, namely plasmonic nanostructures, also clearly indicates that in resonance a 90° phase difference occurs, and that considering this phase difference is essential for the interpretation and understanding of experimental results [100]. Besides, a similar 90° phase difference is found in optical parametric amplification and oscillation [101,102], when energy is to be conserved. When the idler is generated by the signal (equivalent to stimulated emission of idler light or absorption of signal light), a 90° phase difference occurs between signal and idler. When the energy returns from the idler to the signal (equivalent to absorption of idler light or stimulated emission of signal light), the opposite 90° phase difference occurs, leaving the resulting idler and signal phases unchanged, nevertheless at a 90° phase difference with respect to each other. It is similar to the situation in Fig. 28(a), where the dark-red double arrows resembling stimulated emission and absorption occur at opposite 90° phase difference, leaving the phase of the electric field resulting from the two subsequent processes unchanged, nevertheless at a 90° phase difference with respect to the atomic polarization. The same phase shift of 90° is found also in coupling between fibers and spherical or

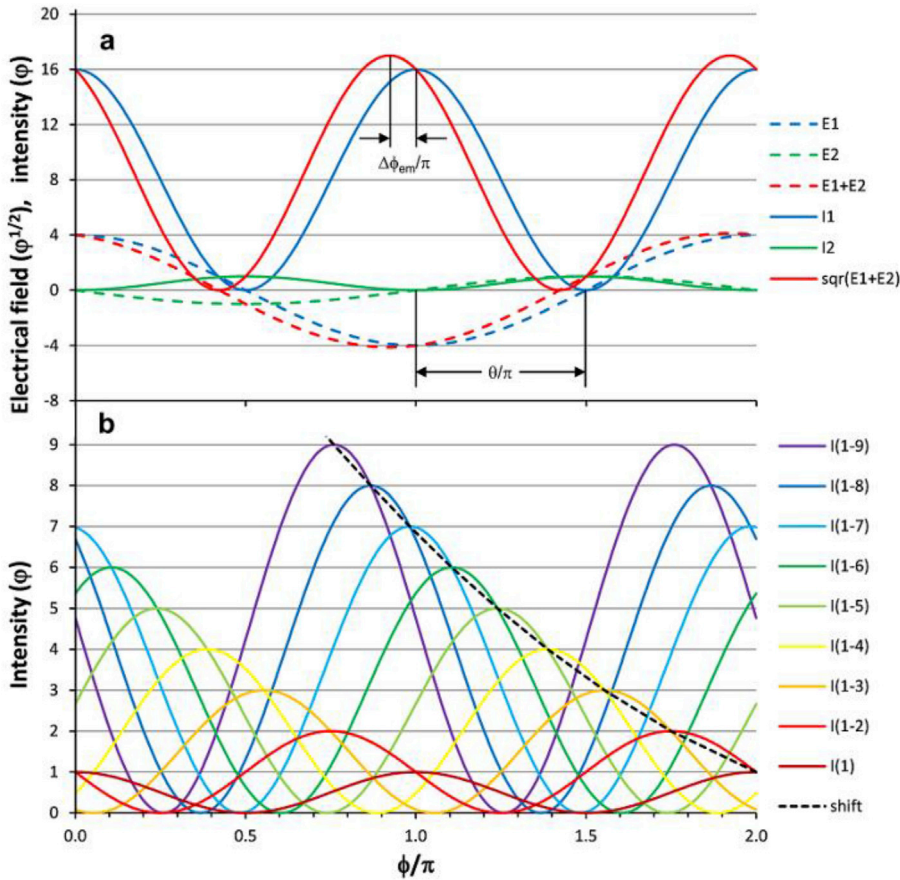


Fig. 30. (a) Example of the summation according to Eq. (123) of two electric fields of amplitudes equivalent to $\varphi^{1/2} = 4$ and 1, with a phase difference of $\theta = \pi/2$. The phase shift of $\Delta\phi_{em} = 0.1476 \pi$ between incident and resulting electric field is equal to Eq. (133), and the energy is conserved, as $16 + 1 = 17$ photons emerge. (b) Consecutive addition of single photons to an existing electromagnetic field. Intensity of the light beam in units of φ . The black dashed line calculated from Eq. (134) indicates the phase shift $\Delta\phi_n = \Sigma (\Delta\phi_{em})$ accumulated with increasing number φ of photons. (Figure taken from Ref. [3]).

ring micro-resonators [103–105].

A mode is defined by its resonance frequency and spectral mode shape (which can be Lorentzian, but also highly distorted [2]), its transverse spatial mode shape (in an open two-mirror resonator these are the Hermite-Gaussian TEM_{xy} modes), and its polarization (in the simplest case, two linear polarizations). If all modes are orthogonal with each other in these three properties (which is not necessarily the case [24,106–108]), photons in the same mode share these three properties and trigger stimulated emission into this mode. In contrast, a mode is not distinguished from other modes by a unique phase, and light propagating in this mode can assume any phase value. Consequently, there is no need for an electromagnetic field triggered by stimulated emission to have the same phase as the incident electric field. As we have reconfirmed above in multiple ways, semi-classically it must be in quadrature with the incident field.

5.6. Vacuum fluctuations and spontaneous emission in the amplitude-phase diagram

A vacuum fluctuation that occurs during a very short time scale (being ~ 1 fs in the visible spectral range) can violate the law of energy conservation according to the uncertainty principle [80,81]. If vacuum fluctuations appear in an empty mode, they generate a time-averaged zero-point energy corresponding to half a photon. If an electromagnetic field exists in the mode, the vacuum field adds onto that field with an arbitrary phase difference θ .

Spontaneous emission is the consequence of vacuum fluctuations in the presence of optically active, excited species [63]. In contrast to vacuum fluctuations, each individual spontaneous-emission event must increase the field amplitude and obey the law of energy conservation in the same manner as stimulated emission, because one atomic excitation is converted to an electromagnetic energy equivalent to one photon that survives at a macroscopic time scale. This statement agrees with the quantum-optical description of spontaneous emission, in which the creation operator \hat{a}^\dagger increases the photon number by one. If in a spontaneous-emission event the photon is emitted into an unoccupied mode, it enters this mode under a phase that is determined solely by the triggering vacuum fluctuations. However, if the mode is occupied by an electromagnetic field, emission under an arbitrary phase difference θ with respect to the existing field would violate the law of energy conservation; see Fig. 23. Consequently, spontaneous emission must occur with the same phase difference of $\theta = \pi/2$ relative to the total field as stimulated emission.

Only under this condition was it possible for Einstein, by use of a semi-classical rate-equation approach [59] that neglects interference, to confirm Planck's law [34] and derive the Einstein A and B coefficients of spontaneous and stimulated emission. If spontaneous emission occurred with an arbitrary phase difference with respect to an existing electromagnetic field, Einstein's derivation would have had to take interference into account.

Here we present a physical picture of spontaneous and stimulated emission (Fig. 31) that is consistent with all aspects discussed above. Emission of a photon into a resonator mode takes longer than the resonator round-trip time [6]. At this time scale, the energy and phase fluctuations of the many extremely fast vacuum fluctuations average out to half a vacuum photon added to the existing field of φ_{ext} photons. This averaged total field of $\varphi_{\text{ext}} + 1/2$ photons triggers atoms in their excited state to emit a photon, leading, according to Eqs. (129) and (133), to a phase shift between incident and resulting field of

$$\tan(\Delta\phi_{\text{em}}) = \sqrt{1/(\varphi_{\text{ext}} + 1/2)}. \quad (136)$$

Spontaneous and stimulated emission are undistinguishable, because the total field interacts with each atom. Nevertheless, the increase in emission rate due to the additional half vacuum photon is quantified and measurable.

5.7. Amplitude versus phase fluctuations, laser linewidth, and Henry's α -factor

The assumptions that stimulated emission (absorption) occurs in (out of) phase with its driving field, whereas spontaneous emission occurs with an arbitrary phase difference, and their corresponding interpretation in the amplitude phase diagram of Fig. 22(a), have served as the foundation for the quantum-optical derivation of the laser linewidth in the 1960s. From the semi-classical point of view established above, one comes to the following judgement. The interpretation proposed by Lax [25], Haken [26], Henry [64], and others that spontaneous emission occurs with an arbitrary phase difference θ relative to an existing electromagnetic field, thereby introducing amplitude and phase fluctuations [projection onto the axes with $\theta = 0^\circ$ – 180° and $\theta = \pm 90^\circ$, respectively, in Fig. 22(a)], is not supported semi-classically, because it violates the law of energy conservation. Ironically, the $\theta = 90^\circ$ phase difference of the spontaneously emitted photon that these authors considered to be a “pure phase fluctuation” manifests semi-classically exactly the amplitude addition by one photon [Fig. 28(a) or Fig. 31(a)] required to conserve the energy. From a semi-classical point of view, these authors have confused vacuum fluctuations with spontaneous emission.

If—in contrast to vacuum fluctuations—spontaneous emission induces neither phase nor amplitude fluctuations, these fluctuations cannot explain any of the following phenomena. (i) The laser linewidth, i.e., the Schawlow-Townes linewidth [10] or any of its extended versions, is not a result of phase fluctuations. Without the existence of amplitude fluctuations, these cannot be damped out. Consequently, such a mechanism cannot explain (ii) the predicted [25,26,65–67] and, in a few cases, experimentally observed [109,110] reduction of laser linewidth by a factor of 2 around threshold, nor will they (iii) induce the refractive-index changes and resulting amplitude-phase coupling and linewidth broadening originally proposed by Lax [25] and later quantified by Henry via his α -factor [64]. This finding does not exclude, however, that technical or other amplitude “fluctuations” occur, with all the described consequences.

Whatever aspect we have discussed semi-classically, (i) Maxwell's equations and the resulting interference term in the superposition of electromagnetic waves, (ii) energy conservation in a Fabry-Pérot resonator, (iii) the Lorentz oscillator model, (iv) the Kramers-Kronig relations applied to the complex susceptibility, (v) the amplitude-phase diagram, or (vi) simply adding up sine waves in a quantized

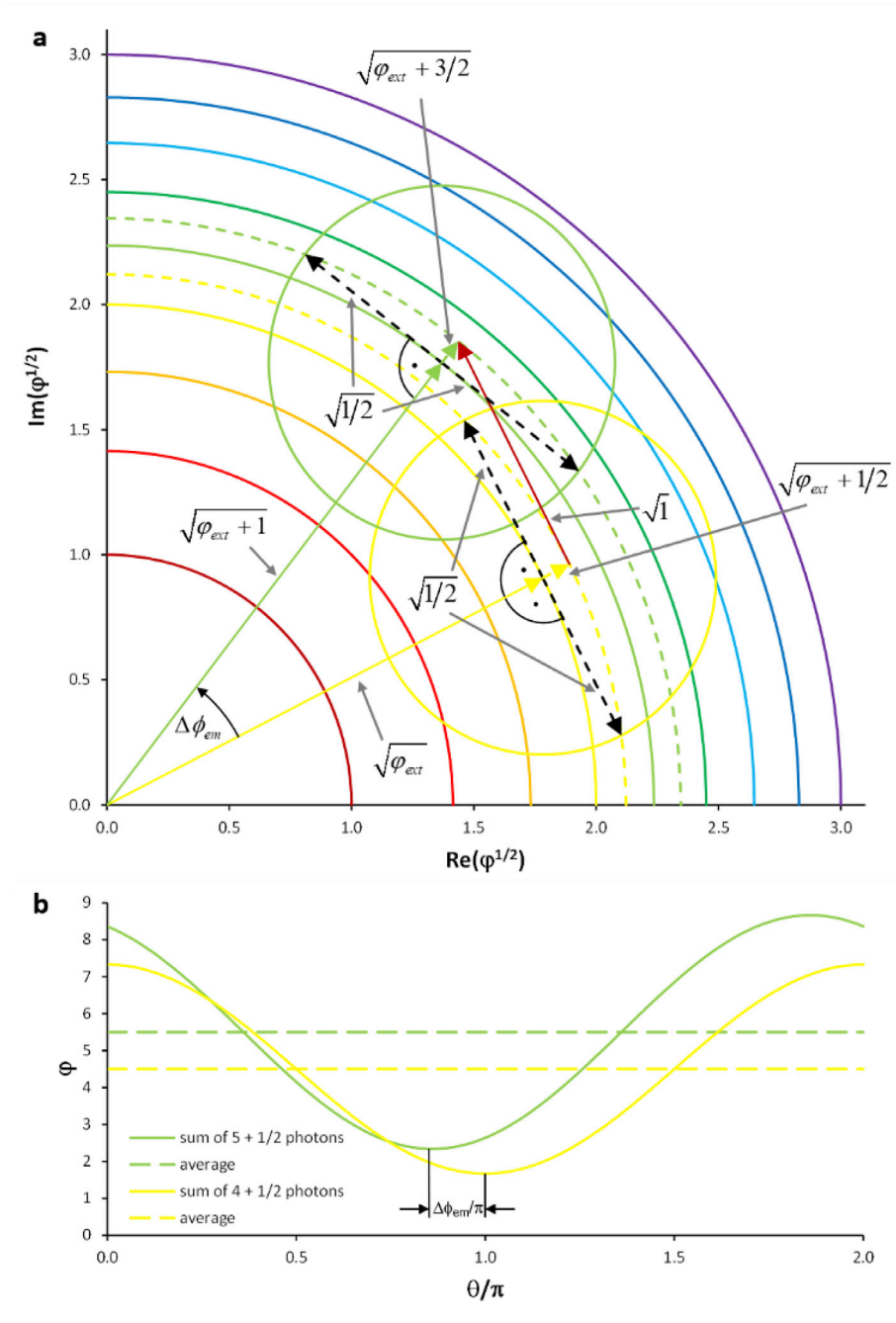


Fig. 31. Relation between vacuum fluctuations, violating the conservation of energy, and spontaneous emission, obeying the conservation of energy. (a) Quadrant of the amplitude-phase diagram illustrating the process of (stimulated and spontaneous) emission in the presence of both an external field and vacuum fluctuations. In the example, an existing field representing 4 photons (solid yellow line and arrow) plus the vacuum fluctuation, i.e., in average 4.5 photons (dashed yellow line and arrow), is increased by the emission of a photon to a field representing 5 photons (solid light-green line and arrow) plus the vacuum fluctuation, i.e., in average 5.5 photons (dashed light-green line and arrow). The dashed black arrows indicate the addition of half a vacuum photon to the field of real photons, which can occur under any phase difference θ with the existing field, resulting in a state on the (yellow or green) circle, but averaging out over many such extremely short-lived events to a state on the (yellow or green) dashed line. Although occurring under all phase differences θ , the dashed black arrows are shown only for the two specific cases of $\pm 90^\circ$ phase difference, where energy conservation happens not to be violated, equivalently to the intensity resulting from averaging over all phase angles. The dark-red arrow represents the field of the one emitted photon. The phase shift $\Delta\phi_{em}$ is calculated from Eq. (136). (b) Number φ of photons resulting from the interference according to Eq. (123) between $\frac{1}{2}$ vacuum photon and 4 photons (yellow solid curve) or 5 photons (green solid curve) versus phase difference θ . Medium of 4.5 photons (yellow dashed line) or 5.5 photons (green dashed line) averaged over all θ . The deviations in amplitude from the average are consistent with the deviations predicted by the yellow and green rings in part (a). (Figure taken from Ref. [3]).

electric-wave picture, we have arrived at the conclusion that stimulated and spontaneous emission both occur under a 90° phase difference with the incident field. This semi-classical result also holds true as a standard perturbation-theoretical approach to absorption and stimulated emission in quantum optics [70–73,84].

Naturally, the following questions arise: Can we find this 90° phase difference when applying the creation operator to a coherent state in quantum optics? If not, can we introduce it to the quantum-optical description, do we even have to rethink the quantum-optical description? Or is there a possibility that the semi-classical limit of quantum mechanics produces a 90° phase difference, whereas the more we enter the quantized world, the angle changes to 0° ? It is hard to imagine a physical mechanism that allows us to derive a positive answer to the latter question.

The current explanation of the laser linewidth, as well as its narrowing and broadening, is in fundamental contradiction with the obtained and experimentally supported semi-classical picture. Recently, we have made significant progress toward establishing an explanation of the laser linewidth [7,8] that coincides with the semi-classical picture of stimulated and spontaneous emission established in this Section. We will present these results in Section 7.

5.8. Are absorption and spontaneous or stimulated emission inverse processes?

We return now to Lamb’s argument [61] about Einstein’s approach [59] to spontaneous and stimulated emission mentioned in the beginning of this Section. To Lamb it was crystal clear that absorption and stimulated emission are the two truly inverse processes. This seemed obvious, as (i) both processes are driven by a real incident electromagnetic field, (ii) their strength is quantified by the same Einstein B coefficient, and (iii) these two processes occur with a defined, opposite phase. In this Sub-section, we will defend the position that the situation is, indeed, a bit more subtle – and that Einstein intuitively had a point.

We emphasized in this Section and Ref. [3] that, in stark contrast to the instructions (A) and (B) mentioned at the beginning of this Section, all quantitative semi-classical models suggest that, generally, emission of a photon at the resonance frequency of an atomic transition must occur with a phase that is 90° in lead of an incident field, whereas absorption must occur with a phase that lags 90° behind the resulting field. Only in this way energy is conserved.

In the amplitude-phase diagram of Fig. 32, all red vectors denote electric fields comprising one photon, whereas the orange vector denotes an electric field comprising two photons. The angle between the two sets of red vectors is 90° . Four processes can be identified, including a maximum of either one (1) or two (2) photons and describing an absorption (abs) or emission (em) process. The four processes are (1-abs) absorption of the electric field of a single incident photon, (1-em) spontaneous emission of the electric field of a photon into an empty mode, (2-abs) absorption of the electric field of one out of two incident photons, and (2-em) stimulated emission of the electric field of a photon driven by a single incident photon. In each case, shown are the initial electric field, the absorbed or emitted electric field (labelled), and the resulting electric field. All four processes obey the law of energy conservation. The figure suggests that the processes (1-abs) and (1-em) are inverse and the processes (2-abs) and (2-em) are inverse.

These four processes are displayed individually on the left-hand side of Fig. 33. In (1-abs), the rightward-pointing red vector is the incident field, the leftward-pointing red vector is the emitted field, and the resulting field is a vector of length zero at the origin. In (1-em), the incident field is a vector of length zero at the origin, whereas the red vector simultaneously indicates the emitted field and the

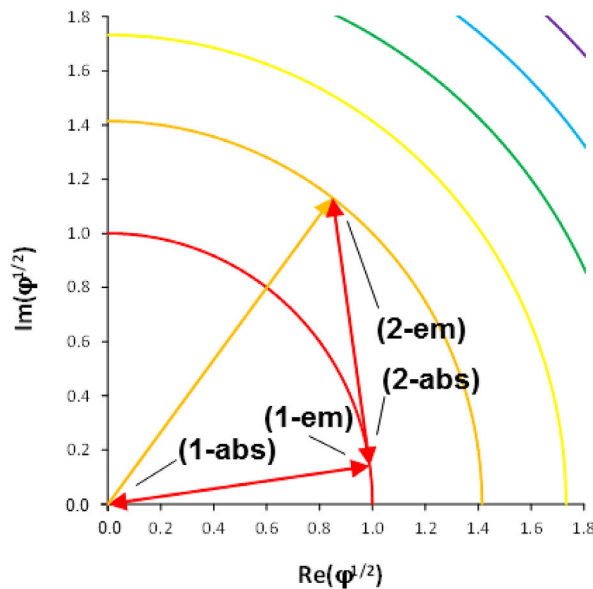
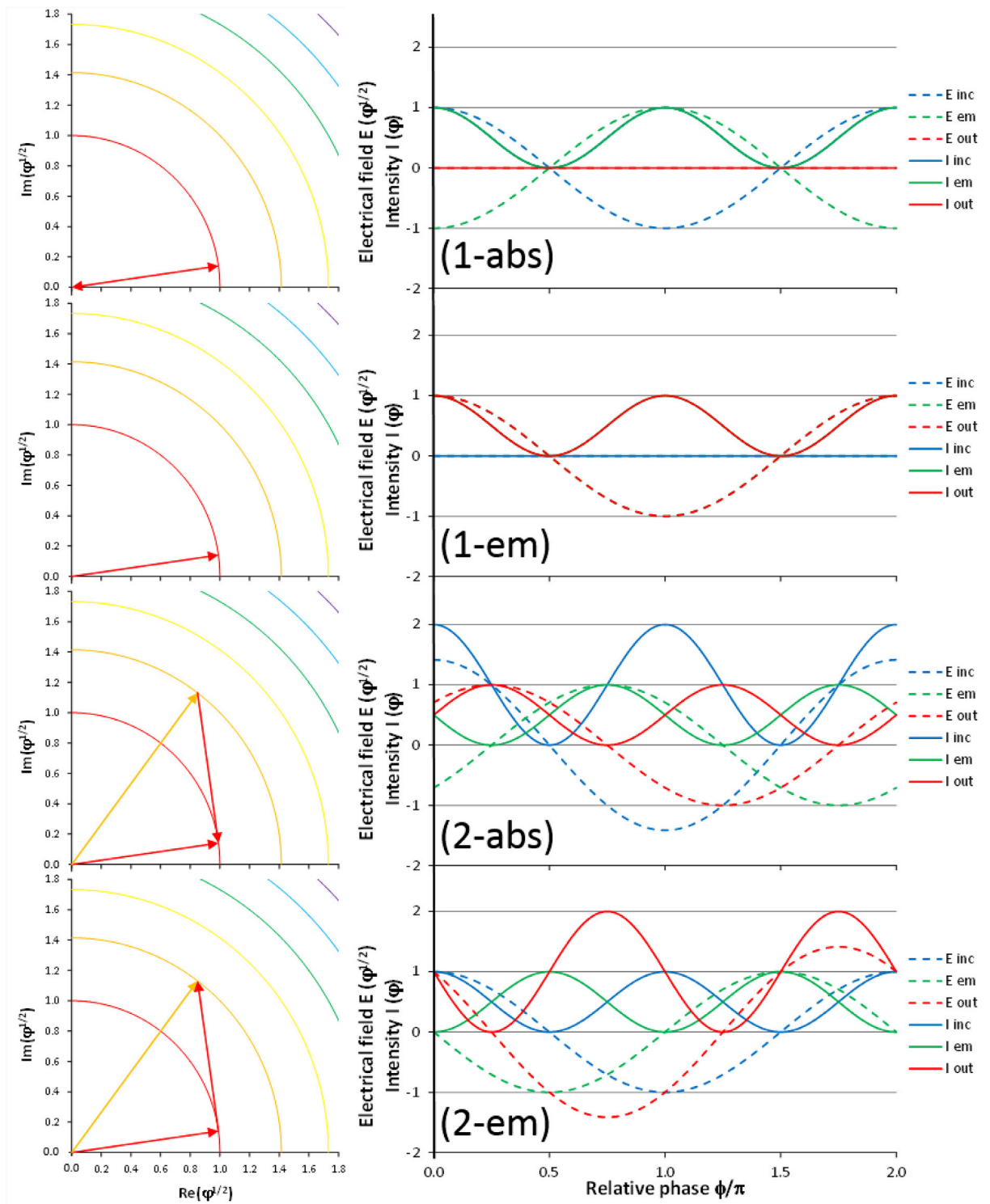


Fig. 32. Quadrant of the amplitude-phase diagram. Incident and resulting fields, as well as absorption and emission processes are represented by electric-field vectors. Labels: A maximum of (1) or (2) photons is involved in an absorption (abs) or emission (em) process. (Figure taken from Ref. [4]).



(caption on next page)

Fig. 33. From top to bottom: The four processes of (1-abs) absorption of the electric field of a single incident photon, (1-em) spontaneous emission of the electric field of a photon into an empty mode, (2-abs) absorption of the electric field of one out of two incident photons, and (2-em) stimulated emission of the electric field of a photon driven by a single incident photon. Left column: Individual processes in the amplitude-phase diagram (equivalent to Fig. 32). Right column: Electric fields E (dashed lines, with the ordinate calibrated as square root of number of photons) and intensities I (solid lines, with the ordinate calibrated as number of photons) versus relative phase ϕ in units of π . Incident (*inc*, blue) wave, displayed with absolute phase angle zero, as well as emitted (*em*, green) and resulting (*out*, red) wave. For (1-abs), blue and green solid lines are identical and red dashed and solid lines are identical. For (1-em), blue dashed and solid lines, green and red dashed lines, and green and red solid lines, respectively, are identical. (Figure taken from Ref. [4]).

resulting field. In (2-abs), the orange vector is the incident field, the downward-pointing red vector is the emitted field, and the rightward pointing red vector is the resulting field. In (2-em), the rightward-pointing red vector is the incident field, the upward-pointing red vector is the emitted field, and the orange vector is the resulting field. The four processes are quantified on the right-hand side of Fig. 33 as a sum of sine waves by taking into account the different phase angles emerging from the left-hand side; see also above and Ref. [3]. An additional phase angle of 8° has been introduced in the amplitude-phase diagrams in order to move the arrows away from the x -axis for improved visibility. This angle is not considered in the sine-wave diagrams on the right-hand side of Fig. 33. The incident field (blue dashed line) induces the emitted field (green dashed line), thereby generating the resulting field (red dashed line). The corresponding intensities, calibrated as number of photons, are shown as solid lines. In all 4 examples the energy is conserved, because the expected number of photons results.

The answer to the question whether absorption and spontaneous or stimulated emission are inverse processes is subtle. The truly inverse processes are those, in which (abs) an absorbing atom removes one photon from an incident field containing $\varphi + 1$ photons, such that the resulting field contains φ photons, and (em) an emitting atom adds one photon to an incident field containing a number φ of photons, such that the resulting field contains $\varphi + 1$ photons. Both processes comprise the same number of photons and are described by the same vector triangle, with the same phase angles, in the amplitude-phase diagram (Fig. 32). The reader can easily extend the diagram in Fig. 32 to larger numbers of involved photons; see also the previous Sub-sections and Ref. [3]. This identification of inverse processes is resembled by the results of the Jaynes-Cummings model [111].

In the four cases displayed in Fig. 33, the relative phase angle of emitted field with respect to incident (or resulting) field is, from top to bottom, $-\pi$ (arbitrary), arbitrary (0), -0.75π (-0.5π), 0.5π (0.25π), respectively. The common pattern of the two corresponding processes with the same number of photons involved is the following. The relative phase of emitted with respect to resulting field in the absorption process lags by a phase difference of π behind the relative phase of emitted with respect to incident field in the emission process or, equivalently (but not adding any additional information), the relative phase of emitted with respect to resulting field in the absorption process lags by a phase difference of π behind the relative phase of emitted with respect to incident field in the emission process. For the two sets of corresponding processes displayed in Fig. 33, namely (1-abs) versus (1-em) or (2-abs) versus (2-em), we find the following:

(1-abs, w.r.t. incident)	$-\pi$	vs.	(1-em, w.r.t. resulting)	0	$\Delta\phi = -\pi$
(1-abs, w.r.t. resulting)	arbitrary	vs.	(1-em, w.r.t. incident)	arbitrary	$\Delta\phi = -\pi$
(2-abs, w.r.t. incident)	-0.75π	vs.	(2-em, w.r.t. resulting)	0.25π	$\Delta\phi = -\pi$
(2-abs, w.r.t. resulting)	-0.5π	vs.	(2-em, w.r.t. incident)	0.5π	$\Delta\phi = -\pi$

In the second line of this comparison both angles are arbitrary, because after absorption of a single photon no field exists anymore and before emission of a single photon into empty space no field existed. In this situation, the relative phase difference of $-\pi$ obtains only when we assume that the incident photon in (1-abs) and the resulting photon in (1-em) have the same absolute phase. As we have already seen in Figs. 28 and 30, the same relative phase difference of $-\pi$ occurs in the emission and absorption of a single photon in fields containing more than one or two photons [3].

This leads us to the simple conclusion that the inverse process of (1-abs) absorption of a single incident photon is (1-em) spontaneous emission into an empty mode, whereas the inverse process of (2-abs) absorption of one photon out of an incident field containing two photons is (2-em) stimulated emission of a second photon induced by a single incident photon.

It then also becomes clear that the two processes of (1-abs) absorption and (2-em) stimulated emission induced by one incident photon are not inverse processes. They are neither in phase and anti-phase with the driving field, respectively, nor are they even in opposite phase with each other. In (1-abs) absorption of a single incident photon, the induced atomic oscillation (and the quasi-emitted photon) is in anti-phase with the incident field, thereby extinguishing the incident field. In (2-em) stimulated emission induced by a single photon, the phase of atomic oscillation and emitted photon is 90° in lead of the incident field (Fig. 32).

The two-level atoms Einstein considered [59] for the description of blackbody radiation [34] can each only absorb one photon, which they would most likely re-emit by spontaneous emission. Therefore, it is natural to assume one incident photon when describing the process of absorption and no incident photon when describing the process of spontaneous emission, in their simplest form. It is then only consequent to also consider stimulated emission by an excited atom for the simplest case of one incident photon. When comparing these three simplest possible processes, Einstein rightfully identified absorption and spontaneous emission as inverse processes because of their symmetry: one photon either disappears or appears. In contrast, the process of stimulated emission involves two photons. Einstein interpreted stimulated emission as a “new” phenomenon, because it was hitherto unknown and because it does not have a symmetry with the two other processes. Einstein certainly had a point! Unfortunately, although Maxwell’s electromagnetic theory was

known, Einstein did not investigate the phase aspect of these processes, which would have revealed the full picture.

On the one hand, the two processes of absorption and stimulated emission of the electromagnetic field of one photon by an atom driven by the electromagnetic field of a single incident photon are connected by the facts that they are both (i) driven by an external electromagnetic field and (ii) quantified by the same Einstein B coefficient. However, these two processes are unrelated in terms of the number of photons involved and do not exhibit opposite phase difference between driving and driven field. On the other hand, the two processes of absorption of a single incident photon by an atom and spontaneous emission of one photon from an atom into an empty mode are connected by the facts that (iii) they are inverse processes in the amplitude-phase diagram, with the same phase angles, and (iv) the same number of photons is involved. The truly inverse process of stimulated emission of one photon by an excited atom driven by a single incident photon is the absorption of one out of two incident photons by an atom in its ground state, as both processes (i) are driven by an external electromagnetic field, (ii) are quantified by the same Einstein B coefficient, (iii) are opposite processes in the amplitude-phase diagram, with the same phase angles, and (iv) involve the same number of photons. Hence, neither Einstein nor Lamb was right or wrong.

6. Theories of the laser linewidth and their contradicting statements

It is a point of fundamental physical logic that an experimental phenomenon that is based on a quantum-mechanical effect can by no means be explained semi-classically; conversely, if an experimental phenomenon can be explained semi-classically, it cannot be the result of a quantum-mechanical effect but must have a semi-classical reason. It has often been claimed that the laser linewidth can only be derived in a full quantum-electrodynamics approach by considering amplitude and phase fluctuations induced by spontaneous emission. Historically, however, the original Schawlow-Townes equation was first derived semi-classically [9,10], without considering quantum fluctuations. The direct, undeniable consequence is that the Schawlow-Townes linewidth cannot be the result of quantum fluctuations but must have a different, semi-classical reason. Therefore, it should have been immediately clear in the 1960s when the quantum-mechanical explanations of the Schawlow-Townes equation surfaced that such an approach cannot lead to an improved or deeper understanding of the semi-classical reason for the fundamental laser linewidth.

In the following Sections, we excavate the semi-classical physical principle underlying the original Schawlow-Townes equation: The gain elongates the photon-decay time, thereby reducing the linewidth of the passive resonator introduced in Section 2 to the fundamental laser linewidth. This principle reveals that the fundamental laser linewidth does not directly depend on spontaneous emission. It describes four- and three-level, transient and continuous-wave (cw) lasers above, at, and below the laser threshold. We introduce the spectral-coherence factor Λ , which relates the photon-decay time, linewidth, Q -factor, and coherence time and length of a mode with gain or a lasing mode to the respective values of the underlying passive mode. In a cw laser, the inherent consequence of spontaneous emission into the lasing mode is that the gain is smaller than the losses, thereby causing photons inside this mode to decay out of the resonator, hence explaining the finite linewidth value in a cw laser. The original Schawlow-Townes equation then emerges from the fundamental laser linewidth by applying the same four approximations exploited in its original derivation.

6.1. Original Schawlow-Townes equation

One of the key features of a laser that is equally of theoretical and practical interest is its spectral coherence, manifested in the laser linewidth. Understanding the linewidth of a single longitudinal resonator mode under lasing conditions is, therefore, of utmost importance. In 1951 Purcell and Pound [112] showed that in a nuclear spin system one can achieve a negative temperature – an early interpretation of population inversion by use of Boltzmann statistics. Inspired by this result, in 1954 Gordon, Zeiger, and Townes demonstrated experimentally the first man-made coherent light source, the ammonia maser [113], which operated on the ammonia inversion line at a wavelength of 12.5 mm [114]. A year later they derived theoretically the linewidth $\Delta\nu_M$ of their ammonia maser [9],

$$\Delta\nu_M^* = \frac{4\pi k_B T (\Delta\nu_c^*)^2}{P_{out}} \Leftrightarrow \Delta\nu_M = \frac{2\pi k_B T (\Delta\nu_c)^2}{P_{out}}, \quad (137)$$

originally defined as the half width at half maximum (HWHM), denoted here by an asterisk and transformed to full width at half maximum (FWHM). k_B is Boltzmann's constant, T is the temperature, $\Delta\nu_c$ is the linewidth of the passive-resonator mode, and P_{out} is the maser output power. Equation (137) was derived [9] under the reasonable approximations that it is (i) a truly continuous-wave (cw) maser, (ii) an ideal four-level maser, and that its microwave resonator suffers from outcoupling losses, but (iii) exhibits no intrinsic losses. In 1958, two years before the first laser was experimentally demonstrated [115], Schawlow and Townes transferred this approximation of the maser linewidth to the optical regime [10] by replacing the thermal energy $k_B T$ by the photon energy $h\nu_L$ at the laser frequency ν_L . As we will show later, this replacement introduces the additional approximation that (iv) one photon is coupled spontaneously into the lasing mode per photon-decay time τ_c of the resonator, independent of the available pump power. Their approximation of the laser linewidth has become famous as the ‘‘Schawlow-Townes linewidth’’,

$$\Delta\nu_{L,ST}^* = \frac{4\pi h\nu_L (\Delta\nu_c^*)^2}{P_{out}} \Leftrightarrow \Delta\nu_{L,ST} = \frac{2\pi h\nu_L (\Delta\nu_c)^2}{P_{out}}. \quad (138)$$

The procedure by which Eq. (137) was derived [9] and subsequently Eq. (138) was deduced from it [10] was semi-classical, in the sense that it did not consider quantum fluctuations in the derivation. The Schawlow-Townes equation existed even before quantum scientists

became interested in the matter.

After the demonstration of the maser and the first derivation of its linewidth, several theories of the laser linewidth were proposed, namely (a) a semi-classical description of the laser as an amplifier of noise, (b) various extensions to the original Schawlow-Townes equation, and (c) quantum theories based on different methods. These three types of approaches contradict each other in several important points. Their fundamental discrepancies have never been clarified and often go unnoticed by scientists working in the field. In order to appreciate the existing problems in understanding the laser linewidth, it is helpful to compare the relevant theories developed since the first demonstration of the maser. In the following Sub-sections, we discuss briefly the existing theories and point out the discrepancies between them.

6.2. Description of the laser as an amplifier of noise

Obviously, the original Schawlow-Townes equation relies on a semi-classical physical principle; otherwise it would have been all but impossible to derive it semi-classically [9,10]. Unfortunately, the first account by Gordon, Zeiger, and Townes [9] is lengthy and complex, and the underlying physical principle remains somewhat obscure. Moreover, the transfer from the microwave regime of Eq. (137) to the optical regime of Eq. (138) has implications that are not straight-forward to understand. Ultimately, Schawlow and Townes may have added confusion to the interpretation of the laser linewidth by claiming in the very paragraph [10], in which they introduce Eq. (138), that “spontaneous emission into this mode adds waves of random phase to the electromagnetic oscillation, and hence produces a finite frequency width which may be obtained by analogy with expression (137)”. However, in stark contrast to this claim, nowhere in Ref. [9] or Ref. [10] occurs a single equation that takes account of spontaneous emission as a wave, or phases, or interference, i.e., Gordon, Zeiger, and Townes [9] and subsequently Schawlow and Townes [10] have manifestly not considered in their derivation of the laser linewidth what the latter claim needs to be considered. Instead, Gordon [116] pointed out that a fundamental physical concept, namely that of the cw maser as a saturated amplifier of noise, whose gain is smaller than its losses, was already inherently present in the first derivation of the maser linewidth [9]. Therefore, it must also underly the original Schawlow-Townes equation (138).

Inspired by the first derivation of the linewidth of a maser [9] and a laser [10], between 1957 and 1964 semi-classical approaches were formulated [116–121] that described the cw maser or laser as a saturated amplifier of noise, whose gain is smaller than its losses; hence the acronym MASER or LASER, in which the letter A emphasizes the principle of “amplification”.

In 1957, before the concept of the maser linewidth was transferred to the optical regime and the laser was experimentally demonstrated, Pound [117] based his approach to noise in a maser on an analogy with electrical circuit theory, which was well established at that time. He modeled the noise equivalently to spontaneous emission from excited atoms, in analogy to blackbody radiation, however at a negative temperature that directly reflects population inversion. Although Pound associated a broader sense than just population inversion with this interpretation, we would rather argue that the pump rate drives the system out of its thermal equilibrium at the given temperature, hence Boltzmann statistics does not apply. Pound’s derivation shows that the system operates at a point where the gain is smaller than the total circuit losses and that the Q -factor of the oscillating system remains finite. However, Pound did not make an attempt to verify the maser linewidth of Eq. (137), and it would potentially have constituted an impossible task because of the approximations he made in his derivations.

In 1961, Wagner and Birnbaum [118] exploited the established knowledge on atomic spectroscopy to consider the spectral output characteristics of the laser for the case of many longitudinal modes with different losses. In their derivation, the Q -factor of each – lasing or non-lasing – mode also remains finite. For a single lasing mode, they derived a linewidth that is more general than the Schawlow-Townes linewidth, in the sense that it avoids approximation (ii), i.e., it is valid also for lasers exhibiting reabsorption from the lower laser level and considers the overlap integral between the lasing mode and a narrow atomic emission line. When approximating a four-level laser and, for the emission line, a delta function centered at the modal resonance frequency, one obtains the Schawlow-Townes linewidth.

In 1962, by comparison with the theory of fission in a nuclear reactor in a subcritical, critical, or supercritical state, which was also well established at that time, Blaquièrre [120] derived semi-classically half the maser linewidth of Eq. (137), which consequently translates into half the original Schawlow-Townes linewidth. This result would be remarkable, because the quantum theories also predict, compared to the Schawlow-Townes linewidth, a factor-of-two decrease above the laser threshold (see Section 6.3). However, it is likely that Blaquièrre made a simple mistake by relating the passive-resonator linewidth $\Delta\nu_c^*$, which he does not specify as HWHM or FWHM, and the photon-decay time τ_c as $2\Delta\nu_c^* = 1/(\pi\tau_c)$ in his third hypothesis and his Eq. (5). Dividing both sides by 2 would, indeed, result in the correct FWHM linewidth of Eq. (39). However, in the early 1960s it was still rather common to use the HWHM linewidth, Blaquièrre also cites the maser and laser linewidths in their HWHM form, and the factor of 2 on the left-hand side of the equation (instead of in the denominator of the right-hand side) also suggests that $2\Delta\nu_c^*$ means the FWHM linewidth. The correct FWHM linewidth of Eq. (39) would translate into a HWHM linewidth of $2\Delta\nu_c^* = 1/(2\pi\tau_c)$, i.e., Blaquièrre possibly erred by a factor of two.

In 1963, Fleck performed a purely optical investigation. He distinguished between a laser operating based on a few round trips, which he likened to multi-eigenmode oscillation in a Fabry-Pérot resonator, or in a single mode established by many round trips, according to the Fox and Li method [122] for resonators that suffer diffraction losses owing to mirrors with finite lateral dimensions. For both cases, Fleck derived a laser linewidth in which one factor is similar to the Schawlow-Townes equation. However, in the former case the passive-resonator linewidth is replaced by the atomic linewidth of a gas laser. The other factor includes a correction due to diffraction at the end mirrors. This factor is different for the two cases. Whereas the former case does not describe the situation of a laser operating in a steady state, Fleck concluded that even in the latter case and in the absence of other noise sources the linewidth must be

larger than the Schawlow-Townes linewidth.

In 1964, Gordon [116] exploited the transmission-line matrix formalism in microwave networks, which was also well established at that time, including a thermal noise source at a temperature T that, in the optical regime, represents spontaneous emission. Gordon explicitly pointed out that in a steady-state oscillator, including a cw laser, the gain is always smaller than the losses. He also insisted that a true oscillator differs from an ordinary saturated amplifier of noise in that the gain exhibits a faster relaxation time than the time constant of amplitude fluctuations of the initial noise signal, which is approximately the inverse of its bandwidth, such that, when the unsaturated gain exceeds the total resonator losses, $\Delta\nu_L \ll \Delta\nu_c$. Gordon derived an approximation of the laser linewidth that differs from the original Schawlow-Townes equation (138), as it includes reabsorption from the lower laser level by the same factor as already found earlier by Shimoda [119], thereby avoiding approximation (ii). However, Gordon's result also contains additional, rather complicated factors related to the mirror reflectivities. When further approximating that the gain equals the losses, the linewidth seemingly turns into half the Schawlow-Townes linewidth, which is easily explained by the fact that Gordon used the FWHM definition; see Eq. (138).

Each of these investigations has its specific merits, and the comparison with non-optical systems reveals additional insight into physics by emphasizing that a laser is very similar in its behavior to other true oscillators. However, it is particularly difficult to understand what an omission or approximation, made in the respective theory to facilitate a solution, exactly corresponds to on the optical side. Therefore, these comparisons do not provide a simple understanding of the relevant optical processes that determine the laser linewidth. More importantly, none of these early investigations obtained the fundamental maser or laser linewidth by avoiding all the approximations (i)–(iv), consequently also none of them managed to establish the relationship between the fundamental linewidth and the Schawlow-Townes approximation. This lack of generality and completeness of the semi-classical derivations of the laser linewidth left room for the quantum-optical interpretations of the laser linewidth that emerged in the mid 1960s (Sub-section 6.3).

Later Siegman [123] also proposed that the original derivation describes a laser that operates below threshold inversion. Coldren et al. [124] suggested the same explanation as – in their own wording – an “intuitive” understanding of the laser linewidth, albeit immediately turning their back on it and referring to quantum fluctuations as the correct reason instead. Ghatak and Thyagarajan [125] called essentially the same semi-classical approach “heuristic”. Later, Stéphan [126], Verdeyen [127], Khurgin and Sun [128], we [5–8], and probably others have pointed in the direction of a semi-classical explanation of the laser linewidth.

6.3. Semi-classical extensions to the original Schawlow-Townes equation

Following its publication in 1958 [10], the original Schawlow-Townes equation (138) was extended in various ways. These extended equations often trade under the same name, the “Schawlow-Townes linewidth”, thereby creating a veritable confusion in the available literature on the laser linewidth, as it is often unclear which particular extension of the original Schawlow-Townes equation the respective authors refer to.

Firstly, the original Schawlow-Townes equation was extended to eliminate some – but not all – of the approximations (i)–(iv) made during its first derivation. Most extended versions eliminate approximation (iii). Most of them also address approximation (ii) by considering reabsorption, albeit not in a general manner, but in a simplified three-level system. Approximations (i) and (iv) are typically still contained in the extended versions. In these extensions, no quantum fluctuations were considered, i.e., these extended versions remained semi-classical.

Secondly, in 1979 Petermann introduced [23] the K -factor to account for the increased linewidth of a gain-guided compared to an index-guided semiconductor laser, which was later shown to be a result of transverse-mode non-orthogonality and generalized by Siegman [106,107]. Similarly, non-orthogonality between longitudinal [24] and polarization [108] modes was addressed by Woerdman et al. All these effects were derived semi-classically. As a consequence, a K -factor is sometimes added to the linewidth equation *a posteriori*.

6.4. Quantum-optical extensions to the original Schawlow-Townes equation

Mostly in the 1960s, different quantum theories of the laser were developed by Haken [26], Lax [25,67], Sargent, Scully, and Lamb [129], Henry [64], and others. These theories proposed that in a cw laser the gain equals the losses, in fundamental contrast to the early semi-classical description of the laser as an amplifier of noise [9,10,116–121]. The quantum theories typically resulted in the original Schawlow-Townes equation (138), hence they include the same four approximations (i)–(iv) and do not consider its semi-classical extensions, which were sometimes added *a posteriori*. On the other hand, these quantum theories suggested two additional effects, which have been derived only quantum-optically.

Firstly, a reduction of linewidth by an additional factor of two around the laser threshold was predicted [26,67]. It was attributed to spontaneous emission coupling into the lasing mode with an arbitrary phase angle, thereby inducing amplitude and phase fluctuations of the oscillating electromagnetic field [Fig. 22(a)]. Above laser threshold, amplitude fluctuations are damped by relaxation oscillations and do not contribute to the linewidth anymore, whereas phase fluctuations exhibit no restoring force, thereby evoking the reduced laser linewidth. Consequently, nowadays half the original Schawlow-Townes equation is usually found in the literature. However, among thousands of investigations experimental evidence of an additional reduction in laser linewidth directly above laser threshold has been found only in a handful of cases, e.g. in the investigations by Gerhardt et al. [109,110].

Secondly, amplitude fluctuations, even if damped above laser threshold, nevertheless induce refractive-index fluctuations, which in turn induce phase fluctuations, thereby increasing the laser linewidth, as quantified by Henry's α -factor in 1982 [64]. As a consequence, this α -factor is sometimes added to the linewidth equation *a posteriori*.

However, Henry pointed out [64] that, if spontaneous emission coupled into the lasing mode with an arbitrary phase angle, thereby

inducing amplitude and phase fluctuations of the oscillating electromagnetic field, each single spontaneous-emission event would violate the law of energy conservation, thereby implicitly questioning the quantum-mechanical explanation of the laser linewidth and his own α -factor. To conserve the energy, i.e., one atomic excitation energy is converted into one photon in the lasing mode, spontaneous and stimulated emission must occur with a phase that is 90° in lead of the driving field; see Section 5 and Ref. [3]. The unavoidable consequence is that each spontaneous-emission event does not induce arbitrary amplitude and phase fluctuations but adds to the resonator exactly the one photon that is emitted by the decaying atom. Spontaneous emission of real photons must not be confused with vacuum fluctuations in the lasing mode. The latter can, indeed, violate the conservation of energy within a time duration allowed by the uncertainty principle, and their addition to the real photon field leads to temporal deviations from the expectation value of the photon number, resulting in the Poissonian photon statistics of a coherent state [68].

Thanks to the quantum-optical investigations and explanations [25,26,64,67,129], nowadays many scientists claim that the laser linewidth can only be derived in a full quantum-electrodynamics approach, see e.g. Ref. [130]. Whereas this claim may hold true for the additional physical effects mentioned in this Sub-section, it is demonstrably incorrect for the fundamental laser linewidth and its four-fold approximation, the original Schawlow-Townes equation. However, the relationship between (a) the semi-classical descriptions of the laser as an amplifier of noise, (b) the original Schawlow-Townes equation and its extended versions, and (c) the quantum theories has never been clarified. Besides, a description of the time-dependent linewidth of a lasing mode for the important class of transient lasers, thus avoiding approximation (i), is still lacking. These three approaches contradict each other in several important points, namely whether the gain equals or is smaller than the losses, which physical phenomenon is responsible for the finite laser linewidth, and the quantitative result for the laser linewidth. In the following Section, we will clarify these points.

7. Fundamental laser linewidth and Schawlow-Townes approximation

In this Section we excavate the physical principle underlying the fundamental laser linewidth: the gain elongates the photon-decay time. Particularly, in a cw-lasing mode the gain is smaller than the losses, resulting in an exponential decay of photons out of the resonator. We derive straight-forwardly the fundamental linewidth equation of a single longitudinal lasing mode, which avoids all the approximations (i)-(iv) discussed in Section 6. It accounts for four- and three-level, transient and cw lasers above, at, and below the laser threshold. We then apply the approximations (i)-(iv), thereby deriving the original Schawlow-Townes equation. This result underlines the correctness of the derived fundamental laser linewidth and clarifies the relationship between the description of the laser as an amplifier of noise, the original Schawlow-Townes equation, and several of its extended versions. Finally, we point out that the existing quantum theories would have provided the same result, had its investigators not missed the important result that also the quantum theories predict that the gain is smaller than the losses.

The results we present in this Section do by no means deny that quantum fluctuations exist, nor do they question whether such fluctuations have an additional influence on the laser linewidth. Nevertheless, since the physical principle behind the fundamental laser linewidth and the Schawlow-Townes approximation is semi-classical, our investigations concentrate on the semi-classical point of view. A detailed investigation of the existing quantum theories is beyond the scope of this paper.

We investigate the spectral properties of a single mode with longitudinal-mode index q_L , centered at the resonance frequency

$$\nu_L = q_L \frac{1}{l_{RT}} = q_L \Delta\nu_{FSR} = q_L \frac{c}{2\ell}, \quad (139)$$

at which eventually laser oscillation will occur when the gain produced by pumping the active medium is high enough. We call this mode a passive mode when the active medium is transparent at ν_L . The losses of a passive mode comprise its outcoupling losses and its intrinsic losses. According to our definition of a passive mode, we do not include the absorption loss or gain introduced by an unpumped or pumped active medium, respectively. Consequently, the active medium is transparent at ν_L . This is the reference situation, which we defined in Section 2.

7.1. Active-mode properties

In this Sub-section, we assume a situation, in which the active medium absorbs or amplifies light at frequency ν_L , depending on excitation of the active species of the active medium. Just like for Section 2, these results are, in principle, well known. Again, confusion has arisen in the literature from unprecise definitions, misinterpretation of the precise meaning of parameters, or inconsequent use of parameters. Perhaps surprising to many scientists, it is exactly the equations we introduce in this Sub-section about amplifier loss and gain that also explain the spectral-coherence properties of the lasing mode. This intriguing fact makes precise definition of the relevant parameters important. Furthermore, we introduce the spectral-coherence factor Λ , which quantifies the relationship of spectral coherence in an active with respect to its underlying passive mode and, in the subsequent Sub-section, will be shown to be the fundamental parameter that defines the spectral-coherence properties of a lasing mode.

In an active medium, the strengths of emission from the upper and absorption from the lower level of a transition resonant at ν_L are quantified by the effective emission and absorption cross sections $\sigma_e(\nu)$ and $\sigma_a(\nu)$, respectively [6]. We assume that these cross sections vary insignificantly over the Lorentzian spectral line shape of the mode at ν_L , i.e.,

$$\begin{aligned} \sigma_e(\nu) &\approx \sigma_e(\nu_L) = \sigma_e \\ \sigma_a(\nu) &\approx \sigma_a(\nu_L) = \sigma_a, \end{aligned} \quad (140)$$

as is typically the case for bulk visible and near-infrared solid-state, semiconductor, and dye lasers. The small-signal gain per unit length on the laser transition between upper and lower laser level or, in the presence of crystal-field splitting, upper and lower manifold with population densities N_2 and N_1 , respectively, can then be approximated by [6]

$$g = \sigma_e N_2 - \sigma_a N_1 = -\alpha_{abs}. \tag{141}$$

Equivalently, we can express the situation by the small-signal absorption coefficient α_{abs} on the laser transition. In an ideal four-level system, $N_1 = 0$. In the passive mode described in Section 2, g and α_{abs} equal zero, and the medium is transparent at ν_L . When the medium amplifies light, $g > 0$ and $\alpha_{abs} < 0$; when it absorbs light, $g < 0$ and $\alpha_{abs} > 0$. We assume that the gain or absorption introduced by the active medium is distributed homogeneously over the resonator length, i.e., its value is independent of position z along the resonator axis.

In the presence of gain or absorption due to the active medium, the total decay-rate constant $1/\tau_L$ of photons inside the resonator becomes

$$\frac{1}{\tau_L} = \frac{1}{\tau_c} - cg = \frac{1}{\tau_c} + c\alpha_{abs}. \tag{142}$$

Equivalently to Eq. (36)–(39), Fourier transformation results in a Lorentzian spectral line shape with a FWHM linewidth given by

$$\Delta\nu_L = \left| \frac{1}{2\pi\tau_L} \right| = \left| \Delta\nu_c - \frac{cg}{2\pi} \right| = \left| \Delta\nu_c + \frac{c\alpha_{abs}}{2\pi} \right|. \tag{143}$$

Although not necessary at this point, but in anticipation of the following Sub-section, the absolute values are taken in Eq. (143). For $g > 0$, this linewidth decrease is known as gain narrowing. For $g < 0$, the linewidth increase is known as absorption broadening. Furthermore, equivalently to Eqs. (43), (50), and (51), respectively,

$$Q_L = 2\pi\nu_L\tau_L, \tag{144}$$

$$\tau_L^{coh} = \frac{\ell_L^{coh}}{c} = 2\tau_L. \tag{145}$$

Because of gain or absorption in the active medium, all five reference parameters τ_c , $\Delta\nu_c$, Q_c , τ_c^{coh} , and ℓ_c^{coh} change by the same factor

$$\Lambda := \frac{1}{1 - cg\tau_c}, \tag{146}$$

which we introduce here as the spectral-coherence factor Λ , resulting in the five parameters τ_L , $\Delta\nu_L$, Q_L , τ_L^{coh} , and ℓ_L^{coh} , which now equivalently quantify the spectral coherence of the active mode at ν_L :

$$\tau_L = \Lambda\tau_c, (\Delta\nu_L)^{-1} = |\Lambda|(\Delta\nu_c)^{-1}, Q_L = \Lambda Q_c, \tau_L^{coh} = \Lambda\tau_c^{coh}, \ell_L^{coh} = \Lambda\ell_c^{coh}. \tag{147}$$

The situation is summarized in Table 1. When the value of Λ is in the range $0 < \Lambda < 1$, the active medium absorbs at ν_L ; for $\Lambda = 1$, the active medium is transparent at ν_L , delivering the coherence parameters of the passive mode, whereas for $1 < \Lambda < \infty$ the active medium amplifies light at ν_L . Vice versa, we can express the ratio between gain-rate constant, cg , and decay-rate constant, $1/\tau_c$, in terms of Λ :

$$cg\tau_c = 1 - \frac{1}{\Lambda}. \tag{148}$$

Important – but unfortunately sometimes misunderstood – is the fact that also these new parameters characterize the active mode independently of how it is excited by small-signal light, from outside or inside and cw or transiently, because these five parameters

Table 1
Relationship between the parameters of passive and active mode (Table taken from Ref. [7]).

Mode	Passive		Active	Relation
Relation		$\times \Lambda$		
Q-factor	Q_c	\rightarrow	Q_L	
Inverse linewidth	$(\Delta\nu_c)^{-1}$	\rightarrow	$(\Delta\nu_L)^{-1}$	$\times \nu_L$
Photon-decay time	τ_c	\rightarrow	τ_L	$\times 2\pi$
Coherence time	τ_c^{coh}	\rightarrow	τ_L^{coh}	$\times 0.5$
Coherence length	ℓ_c^{coh}	\rightarrow	ℓ_L^{coh}	$\times 1/c$

depend only on the resonator losses and the gain. The gain remains practically unchanged in case of a small-signal source. Particularly, if small-signal cw white light is launched into the resonator at a rate R_{launch} per unit time and the light stored inside the resonator reaches a steady state, light still decays out of the resonator with the same photon-decay time τ_L , while being continuously replenished:

$$\frac{d}{dt}\varphi = R_{launch} - \frac{1}{\tau_L}\varphi = 0. \quad (149)$$

The definitions of τ_L , $\Delta\nu_L$, Q_L , τ_L^{coh} , and ℓ_L^{coh} remain perfectly valid and their values remain exactly the same. No Q -factor becomes infinite and no linewidth vanishes, if the number φ of photons inside the active mode at ν_L is constant in time in a steady-state situation. Equivalently, no Airy distribution – the sum of all mode profiles – collapses to a series of delta functions, if we shine small-signal cw white light through an active Fabry-Pérot resonator [126].

An important result has potentially gone unnoticed by one or the other reader. We have already derived the fundamental laser linewidth, namely in Eq. (143). In the following, let us elaborate a bit more on this result.

7.2. Fundamental laser linewidth

Thanks to the proper introduction of the relevant parameters of a passive and an active longitudinal mode and their relationship, it is now straight-forward to understand the spectral-coherence properties of a lasing mode. With the stimulated-emission rate R_{st} and photon-decay rate R_{decay} into and out of the mode at ν_L , respectively,

$$R_{st} = c g \varphi, \quad (150)$$

$$R_{decay} = \frac{1}{\tau_c} \varphi, \quad (151)$$

we can rewrite the spectral-coherence factor of Eq. (146) for this mode as

$$\Lambda = \frac{R_{decay}}{R_{decay} - R_{st}}, \quad (152)$$

its photon-decay time of Eq. (142) as

$$\tau_L = \Lambda \tau_c = \frac{R_{decay}}{R_{decay} - R_{st}} \tau_c, \quad (153)$$

and its linewidth of Eq. (143) as

$$\Delta\nu_L = \frac{1}{|\Lambda|} \Delta\nu_c = \left| \frac{R_{decay} - R_{st}}{R_{decay}} \right| \Delta\nu_c. \quad (154)$$

Equations (146) and (152) equivalently describe spectral coherence in a single longitudinal lasing mode. Equations (143) and (154) equivalently represent the fundamental linewidth of a single longitudinal lasing mode. The laser can have any energy-level system ranging from the extreme situation of an ideal 4-level laser to that of an ideal 3-level laser and covering any intermediate situation, e.g., of a quasi-4-level or quasi-3-level laser. The lasing mode can operate below, at, or above threshold and in a cw or a transient lasing regime, with its gain being smaller, equal, or larger compared to its losses. The situation is displayed in Fig. 34.

Equation (154) highlights the important fact that, in a transient lasing regime, the laser linewidth does not generally depend on the spontaneous-emission rate or on the photon number φ inside the lasing mode (or, equivalently, the output power). However, for a cw lasing mode we will find in the second-next Sub-section that it does so, because the spontaneous-emission rate determines the difference $R_{decay} - R_{st}$ in Eq. (154).

It is sometimes argued that in a number of situations the linewidth of a single longitudinal mode would not matter anymore. For example, it has been proposed that (a) below laser threshold the emission from the resonator consists mostly of noise resulting from spontaneous emission and, consequently, the linewidth approaches the luminescence bandwidth, (b) at and slightly above laser threshold, the laser may lase on multiple longitudinal modes, resulting in a significantly broader linewidth than that of a single mode, (c) in a Q -switched or mode-locked laser, the linewidth is approximately inverse to the pulse duration, or (d) in super-continuum generation, the linewidth is given by the generated super-continuum itself. All these statements seem valid when observing emission from the resonator with a low-resolution spectrometer, e.g., a grating monochromator whose spectral resolution is typically a few orders of magnitude lower than the spectral features of the Fabry-Pérot resonator, from which the light is emitted. When employing a resolving method that exceeds the spectral resolution of the resonator, however, it immediately becomes evident that in all the above-mentioned situations the broadband spectrum consists of the individual longitudinal modes of the Fabry-Pérot resonator and that the linewidth of each mode depends directly on the linewidth of its underlying passive mode and its gain or absorption according to Eq. (154). While the Schawlow-Townes equation has always been understood as describing a single longitudinal mode operating far above laser threshold, Eq. (154) holds true even far below threshold, where the output is dominated by spontaneous emission and for the important transition

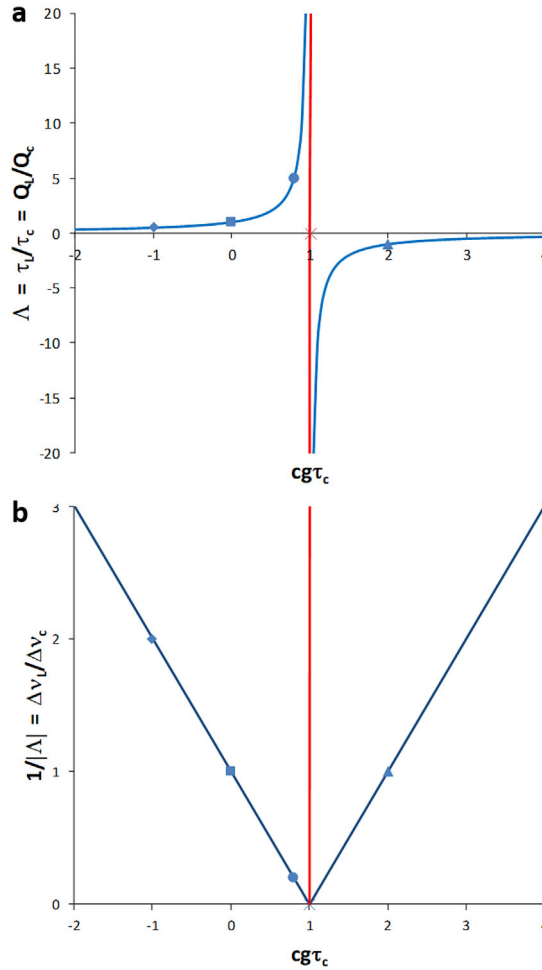


Fig. 34. (a) Spectral-coherence factor Δ , quantifying the spectral coherence of a lasing mode, i.e., enhancement of the photon-decay time, Q-factor, coherence time and length of the lasing mode compared to its underlying passive mode, and (b) its inverse, $1/|\Delta|$, quantifying narrowing of the linewidth of the lasing mode compared to its underlying passive mode, as a function of gain-rate constant cg calibrated to decay-rate constant $1/\tau_c$ (red line). We can distinguish five different situations. (i) Absorbing medium: $\sigma_e N_2 < \sigma_a N_1$ and $g < 0$, resulting in $0 < \Delta < 1$, $\tau_L < \tau_c$, $Q_L < Q_c$, and $\Delta\nu_L > \Delta\nu_c$ (e.g., diamond: medium with absorption losses equaling resonator losses, $cg\tau_c = -1$, resulting in $\Delta = 0.5$). (ii) Passive mode, i.e., resonator with transparent gain medium: $\sigma_e N_2 = \sigma_a N_1$ and $g = 0$, resulting in $\Delta = 1$, $\tau_L = \tau_c$, $Q_L = Q_c$, and $\Delta\nu_L = \Delta\nu_c$ (square). (iii) Amplifying medium with gain smaller than losses: $\sigma_e N_2 > \sigma_a N_1$ and $0 < cg\tau_c < 1$, resulting in $\Delta > 1$, $\tau_L > \tau_c$, $Q_L > Q_c$, and $\Delta\nu_L < \Delta\nu_c$, and the photons inside the mode decay with time (e.g., circle: medium with $cg\tau_c = 0.8$, resulting in $\Delta = 5$). (iv) Amplifying medium with gain equal to losses: $\sigma_e N_2 > \sigma_a N_1$ and $cg\tau_c = 1$, resulting in Δ , τ_L , $Q_L \rightarrow \pm\infty$ and $\Delta\nu_L = 0$, and photons inside the mode exactly reproduce themselves while decaying out of the resonator (cross). (v) Amplifying medium with gain larger than losses (e.g., in the initial phase of a Q-switch pulse): $\sigma_e N_2 > \sigma_a N_1$ and $cg\tau_c > 1$, resulting in $\Delta < 0$, $\tau_L < 0$, $Q_L < 0$, but $\Delta\nu_L > 0$, and photons inside the lasing mode more than reproduce themselves while decaying out of the resonator (e.g., triangle: medium with gain twice as high as resonator losses, $cg\tau_c = 2$, resulting in $\Delta = -1$).

between both extremes. Depending on the application, the linewidth of a single longitudinal mode may matter a lot, even if the overall emission stems from several or many modes.

7.3. Linewidth of a transient laser

In a transient lasing regime, for example in a Q-switched laser during the build-up or decay time of its giant pulse, the gain changes with time t , $dg(t)/dt \neq 0$. Then also Δ and, thus, τ_L , $\Delta\nu_L$, Q_L , τ_L^{coh} , and ℓ_L^{coh} become time-dependent. In such a situation, Eq. (141)–(148) and (150)–(154) correctly describe the instantaneous spectral-coherence properties of the transient lasing mode. Even a situation where in a transient lasing mode the gain is larger than the losses, $cg > 1/\tau_c$, e.g. during the build-up time of a Q-switch pulse, is correctly described. Whereas $\Delta\nu_L$ always has a positive value [Fig. 34(b)], Δ , τ_L , Q_L , τ_L^{coh} , and ℓ_L^{coh} then assume negative values [Fig. 34(a)], which makes physical sense, as an exponential increase of light intensity occurs. $cg > 1/\tau_c$, resulting in $\Delta < 0$, describes transient situations that

are temporally unsustainable, because over time the number of photons in the lasing mode would increase to infinity.

This raises the usual question about how one can practically measure the linewidth of a lasing mode with a detector, particularly in a time-dependent situation. If the detector were fast enough to measure the decay during one oscillation cycle, then per definition of the Q -factor, namely energy stored over energy lost per oscillation cycle, one would, in principle, be able to determine the spectral-coherence properties within one oscillation cycle. However, the definition of outcoupling, intrinsic, and photon-decay time of Eq. (18)–(23), respectively, and, therefore, all spectral-coherence properties, are only defined over one round-trip time t_{RT} (Fig. 5). Consequently, the photon-rate equation, i.e., Eqs. (52) and (149) above, or (155) and (160) below, average over one round-trip time. Moreover, the longer the coherence time and length of the emitted light, the larger is the extension of a wave packet associated with one photon and the larger is the number of wave packets generated at different times that contribute to the electromagnetic field that hits the detector in an instant. Therefore, while the above equations correctly describe the instantaneous spectral-coherence properties of a transient lasing mode averaged over one round-trip time, their measurement typically averages over a longer duration.

7.4. Linewidth of a continuous-wave laser

In this Sub-section, we will look in more detail into the situation of cw lasers, because we can compare our results with existing theories.

In a steady-state situation, Eq. (142)–(146) make physical sense until one reaches the point where $cg = 1/\tau_c$, i.e., where the stimulated-emission rate equals the photon-decay rate, or the gain equals the losses, such that photons inside the lasing mode perfectly reproduce themselves while decaying out of the resonator. At exactly this point, the spectral-coherence factor Λ becomes infinite, hence the photon-decay time, the Q -factor, and the coherence time and length diverge, whereas the linewidth vanishes.

It has been proven experimentally that a single-mode cw laser has a finite (a) linewidth, (b) coherence time, and (c) coherence length. According to the above parameter definitions, it must necessarily also have a finite (d) photon-decay time and (e) Q -factor. The former three aspects have been widely accepted by the scientific community, whereas the latter two have been disputed for long. Nevertheless, the fact that the Q -factor of a cw-lasing mode assumes a finite value was already outlined in 1955 by Gordon, Zeiger, and Townes [9], see also Yariv [131]. Hence, it remains to clarify what the finite photon-decay time of a cw-lasing mode is. As we will see in the following, we can trust our definitions of Eq. (142)–(146), because all five parameters express the same spectral-coherence properties, quantified by the spectral-coherence factor Λ . A cw laser with infinite spectral coherence does not exist, hence the value of Λ must be positive and finite for a cw-lasing mode, $0 < \Lambda < \infty$. Therefore, from Eq. (142)–(146) we can readily exclude that a cw-lasing mode operates at the point where the gain equals the losses.

In fact, it has been known for decades that a cw-lasing mode operates in a region where the gain is smaller than the losses. The emission behavior of semiconductor lasers around their laser threshold has been well explained based on this principle, see e.g. Ref. [76–79], and this explanation has also been accepted in quantum optics. It becomes clear from those investigations that the gain is not only smaller than the losses in a narrow power region close to the laser threshold, but this situation continues until the lasing mode reaches infinite output power. It is scientifically inconsistent to accept this argument when it comes to explaining the laser threshold behavior but reject it when it comes to explaining the laser linewidth and, instead, claim that the gain equals the losses. In the following derivation, we will understand how this physical principle determines the linewidth of a cw-lasing mode.

The semi-classical rate equation for the number ϕ of photons inside the active or lasing mode at ν_L reads:

$$\frac{d}{dt}\phi = R_{sp} + R_{st} - R_{decay} = R_{sp} + cg\phi - \frac{1}{\tau_c}\phi = R_{sp} - \frac{1}{\tau_L}\phi = R_{sp} - 2\pi\Delta\nu_L\phi \stackrel{cw}{=} 0. \quad (155)$$

Here, R_{sp} is the spontaneous-emission rate into the active or lasing mode (not to be confused with the total spontaneous-emission rate from the upper laser level). Since each spontaneous-emission event converts one atomic excitation energy into one photon, R_{sp} is a positive rate (see Section 5 and Ref. [3]). In the situation of a lasing mode operating far above threshold, this rate is often neglected in the photon-rate equation, because the stimulated-emission rate becomes orders of magnitude larger than this rate. However, it is exactly this small rate that causes the linewidth of a cw-lasing mode to be finite. A cw-lasing mode is defined by a time-independent value of ϕ , i.e., the photon-rate equation equals zero. We see immediately from Eq. (155) that the stimulated-emission rate does not equal, but is smaller than the photon-decay rate,

$$R_{decay} - R_{st} \stackrel{cw}{=} R_{sp} > 0 \quad \Rightarrow \quad cg\tau_c = 1 - \frac{1}{\Lambda} \stackrel{cw}{<} 1, \quad (156)$$

i.e., in a cw-lasing mode the gain due to stimulated emission is smaller than the losses, because the spontaneous- and stimulated-emission rate together compensate the losses. Exploiting Eq. (142) in Eq. (155) shows that photons decay out of the lasing mode with the photon-decay time τ_L , and inserting Eq. (143) into Eq. (155) shows that this entails a finite laser linewidth $\Delta\nu_L$. If in a cw-lasing mode R_{sp} equaled zero, the gain would equal the losses, the spectral-coherence factor, the photon-decay time, the Q -factor, and the coherence time and length would become infinite, and the laser linewidth would vanish.

To emphasize the photon decay in a cw-lasing mode further, the photon-rate equation (155) can be split into two parts,

$$\frac{d}{dt}\phi \Big|_{source} + \frac{d}{dt}\phi \Big|_{gain/loss} \stackrel{cw}{=} 0, \quad (157)$$

where

$$\left. \frac{d}{dt} \varphi \right|_{\text{source}} = R_{sp}^{cw} > 0, \quad (158)$$

$$\left. \frac{d}{dt} \varphi \right|_{\text{gain/loss}} = R_{st} - R_{decay} = cg\varphi - \frac{1}{\tau_c} \varphi = -\frac{1}{\tau_L} \varphi < 0. \quad (159)$$

Equations (155) and (157) describe the photon balance inside the mode. Equation (158) describes the photon source, whose rate does not directly depend on the number φ of photons inside the mode. Equation (159) describes the photon action inside the active or lasing mode. We see directly from Eq. (159) that the photons inside the mode decay out of the resonator faster than they reproduce themselves by stimulated emission.

In analogy to a cavity-ring-down experiment in a passive mode that would deliver τ_c as the photon-decay time, one can directly observe the decay of photons present in the lasing mode at time t_1 with photon-decay time τ_L by interfering, in a self-heterodyne experiment [132,133], the laser signal emitted at t_1 with the laser signal emitted at any later time. The longer one waits, the smaller is the fraction of each photon, and its reproduction by stimulated emission, that is still contained in the mode and interferes with the original signal, i.e., the interference between the two signals becomes weaker with increasing delay time.

We emphasize here again, now for the lasing situation: Important – but unfortunately quite often misunderstood – is the fact that the five parameters that characterize the lasing mode depend only on the resonator losses and the gain. However, now the gain reacts very sensitively on any, even small-signal, light source. Particularly, if the light source is spontaneous emission into the lasing mode at a rate R_{sp} per unit time, i.e., the external light source with its rate R_{launch} in Eq. (149) is replaced by the internal light source with its rate R_{sp} , and the light stored inside the lasing mode reaches a steady state, light still decays out of the lasing resonator with the same photon-decay time τ_L (because in a cw-lasing mode the gain is smaller than the losses), while being continuously replenished:

$$\frac{d}{dt} \varphi = R_{sp} - \frac{1}{\tau_L} \varphi \stackrel{cw}{=} 0. \quad (160)$$

The definitions of τ_L , $\Delta\nu_L$, Q_L , τ_L^{coh} , and ρ_L^{coh} remain perfectly valid and their values remain exactly the same. No Q-factor becomes infinite and no linewidth vanishes, if the number φ of photons inside the lasing mode at ν_L is constant in time in a steady-state lasing situation. Equivalently, no Airy distribution – the sum of all mode profiles – collapses to a series of delta functions, if cw white light is generated by spontaneous emission inside a lasing Fabry-Pérot resonator [126].

As mentioned in Section 2.7, the light stored inside and emitted from the resonator depends spectrally and temporally also on the exciting light source, in this case the spontaneous emission into the lasing mode. It raises the question whether the defined $\Delta\nu_L$ represents not only the linewidth of the lasing mode, but also the linewidth of the actually emitted laser light, i.e., whether the two are the same. In bulk visible and near-infrared solid-state, semiconductor, and dye lasers, spontaneous emission is spectrally rather white light compared to the linewidth $\Delta\nu_L$ of the lasing mode. The gain is spectrally similarly broadband as – and in an ideal four-level laser without reabsorption identical to – the spectral line shape of spontaneous emission. Besides, a single spontaneous-emission event by an atom occurs at a time scale that is long compared to one round-trip time t_{RT} [6]. Consequently, spontaneous emission by many atoms has the character of cw white light.

However, presence of the resonator spectrally modifies the spontaneous emission into the lasing mode [6]. If electromagnetic radiation is emitted by an atom in spectral resonance and in phase with the atomic oscillation [90–92] at a frequency that lies exactly between two resonance frequencies, then after one round trip through the resonator it meets the still emitting atom in anti-phase. Consequently, it slows down the atomic oscillation, thereby being reabsorbed by the atom [4]. In other words, the part of the photon which has conducted one round trip and the part of the photon which is currently emitted interfere destructively. Therefore, the spontaneous emission is effectively suppressed, and the energy remains in the atom. If the electromagnetic radiation is emitted at a frequency that, after one round trip, is 90° out of phase with the atomic oscillation, it will interfere destructively after two round trips, i.e., it is slightly less but still suppressed. The described phenomenon is the same as for white light from an external source launched into a Fabry-Pérot resonator [2], the only difference being that the light is generated inside the resonator. Gain then further narrows the linewidth over many round trips, leading to light with the same linewidth being spontaneously emitted as the linewidth of the lasing mode.

As we will see later, when applying the approximations (i)–(iv) mentioned in Section 6.1 to the linewidth of Eq. (143) or (154), we obtain exactly the original Schawlow-Townes equation.

The general dependence between the spectral-coherence factor Λ and the ratio of gain-rate constant, cg , and decay-rate constant, $1/\tau_c$, given by Eqs. (146) and (148), is displayed in Fig. 35. This curve universally describes a cw-lasing mode of any laser.

If we interpret spontaneous emission as stimulated emission driven by the zero-point electromagnetic energy [63], represented by the number φ_{vac} of vacuum photons per resonance frequency and polarization, the spontaneous-emission rate into the lasing mode at ν_L becomes

$$R_{sp} = c\sigma_e N_2 \varphi_{vac}. \quad (161)$$

Sometimes an “extra” photon, representing one vacuum photon, $\varphi_{vac} = 1$, is introduced in the photon-rate equation to account for

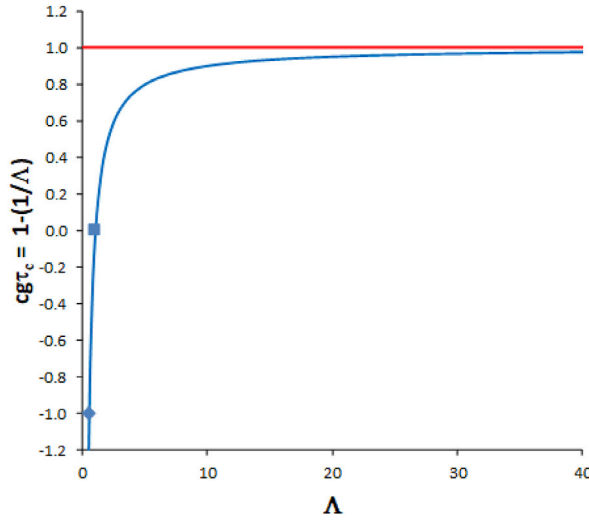


Fig. 35. Gain-rate constant cg (blue curve) calibrated to decay-rate constant $1/\tau_c$ (red curve), as a function of spectral-coherence factor Λ according to Eq. (148) in a cw-lasing mode. This figure is equivalent to the upper left quadrant of Fig. 34(a), with abscissa and ordinate exchanged. A curve commencing at the point, where $cg\tau_c = -1$ and, consequently, $\Lambda = 0.5$ (diamond), describes the example of a three-level laser, whose gain medium introduces absorption losses which, in the unpumped case, equal the resonator losses. An ideal four-level laser is described by a curve commencing at the transparency point, where $cg\tau_c = 0$ and accordingly $\Lambda = 1$ (square). (Figure taken from Ref. [7]).

spontaneous emission into the lasing mode [134–136]. Inserting Eqs. (159), (161), (150), and (141) into Eq. (152) provides a physical expression for the spectral-coherence factor Λ :

$$\Lambda \stackrel{cw}{=} \frac{R_{decay}}{R_{sp}} = \frac{R_{st}}{R_{sp}} + 1 = \left(1 - \frac{\sigma_a N_1}{\sigma_e N_2}\right) \frac{\varphi}{\varphi_{vac}} + 1 \stackrel{4-level}{=} \frac{\varphi}{\varphi_{vac}} + 1 = \frac{\varphi + \varphi_{vac}}{\varphi_{vac}}. \tag{162}$$

For an ideal 4-level laser, the spectral-coherence factor takes a particularly simple form: $\Lambda = \varphi/\varphi_{vac} + 1$. For a laser exhibiting reabsorption of laser light from the occupied lower laser level, because this level has a finite lifetime or is the ground state, the spectral-coherence factor is diminished. Because in a lasing mode φ and, thus, Λ can become very large, the gain very closely approaches the losses, see Fig. 35, and the linewidth of this lasing mode can be orders of magnitude narrower than the passive-mode linewidth. However, if the gain equaled the losses, equivalent to either an infinite number φ of photons inside the lasing mode or neglecting the spontaneous-emission rate into the lasing mode in the photon-rate equation (155), the laser linewidth would vanish. Inserting Eq. (162) into Eqs. (153) and (154) yields the photon-decay time τ_L and linewidth $\Delta\nu_L$ of a cw-lasing mode,

$$\tau_L \stackrel{cw}{=} \frac{R_{decay}}{R_{sp}} \tau_c, \tag{163}$$

$$\Delta\nu_L \stackrel{cw}{=} \frac{R_{sp}}{R_{decay}} \Delta\nu_c, \tag{164}$$

respectively. For an ideal 4-level laser, we obtain

$$\tau_L \stackrel{cw, 4-level}{=} \frac{\varphi + \varphi_{vac}}{\varphi_{vac}} \tau_c, \tag{165}$$

$$\Delta\nu_L \stackrel{cw, 4-level}{=} \frac{\varphi_{vac}}{\varphi + \varphi_{vac}} \Delta\nu_c. \tag{166}$$

The essence of Eq. (166) is the same as can be found in, e.g., Refs. [127,128].

7.5. Schawlow-Townes approximation

When applying the approximations (i)-(iv), mentioned in Section 6.1, to the fundamental laser linewidth of Eq. (154), one obtains straight-forwardly the original Schawlow-Townes equation. Approximating (i) a cw-lasing mode results in Eq. (164). Assuming (ii) an ideal 4-level laser and setting $\varphi_{vac} = 1$ yields

$$cg = c\sigma_e(N_2 - N_1) \stackrel{4-level}{=} c\sigma_e N_2 = R_{sp}. \tag{167}$$

Assuming that (iii) the resonator has no intrinsic losses, $1/\tau_{\text{loss}} = 0$, allows one to set the outcoupling losses equal to the total losses:

$$\frac{1}{\tau_c} = \frac{1}{\tau_{\text{out}}} + \frac{1}{\tau_{\text{loss}}} \stackrel{1/\tau_{\text{loss}}=0}{=} \frac{1}{\tau_{\text{out}}}. \quad (168)$$

The laser output power

$$P_{\text{out}} = h\nu_L R_{\text{out}} = h\nu_L \frac{\varphi}{\tau_{\text{out}}} \quad (169)$$

can then, by use of Eq. (168), be rewritten as

$$P_{\text{out}} \stackrel{1/\tau_{\text{loss}}=0}{=} h\nu_L \frac{\varphi}{\tau_c} = h\nu_L \varphi 2\pi \Delta\nu_c. \quad (170)$$

Via (iv) replacing the thermal energy $k_B T$ by the photon energy $h\nu_L$, Schawlow and Townes implicitly assumed that one photon is coupled spontaneously into the lasing mode per photon-decay time τ_c of the resonator:

$$R_{\text{sp}} = c\sigma_e N_2 \stackrel{4\text{-level}}{=} \frac{R_{\text{st}}}{\varphi} = c g \stackrel{R_{\text{st}}=R_{\text{decay}}}{=} \frac{1}{\tau_c} \stackrel{1/\tau_{\text{loss}}=0}{=} \frac{1}{\tau_{\text{out}}}. \quad (171)$$

Since, in the fictive case of a 4-level cw laser operating at the unreachable point where the gain would equal the losses, each photon inside the lasing mode would exactly reproduce itself once while decaying out of the resonator, also the vacuum photon would produce exactly one copy of itself during the photon-decay time τ_c , i.e., $R_{\text{sp}} = 1/\tau_c$. Applying these four approximations consecutively to the fundamental laser linewidth of Eq. (154) results in

$$\begin{aligned} \Delta\nu_L &= \frac{R_{\text{decay}} - R_{\text{st}}}{R_{\text{decay}}} \Delta\nu_c \\ &\stackrel{\text{cw}}{=} \frac{R_{\text{sp}}}{R_{\text{decay}}} \Delta\nu_c = \frac{c\sigma_e N_2}{\varphi/\tau_c} \Delta\nu_c \\ &\stackrel{4\text{-level}}{=} \frac{c g}{\varphi/\tau_c} \Delta\nu_c \\ &\stackrel{1/\tau_{\text{loss}}=0}{=} \frac{c g}{\varphi/\tau_{\text{out}}} \Delta\nu_c = \frac{c g h\nu_L}{P_{\text{out}}} \Delta\nu_c \\ &\stackrel{R_{\text{st}}=R_{\text{decay}}}{=} \frac{(1/\tau_c) h\nu_L}{P_{\text{out}}} \Delta\nu_c = \frac{2\pi h\nu_L (\Delta\nu_c)^2}{P_{\text{out}}} = \Delta\nu_{L,ST} \end{aligned} \quad (172)$$

Quod erat demonstrandum.

Our derivation of the laser linewidth has clarified the following points. (a) The original Schawlow-Townes equation is a four-fold approximation of the fundamental laser linewidth of Eq. (154) and a three-fold approximation of the cw-laser linewidth of Eq. (164). (b) All three linewidths, the fundamental linewidth of Eq. (154), the cw linewidth of Eq. (164), and the original Schawlow-Townes approximation of Eqs. (172) and (138) are derived semi-classically, the latter two by considering a semi-classical spontaneous-emission rate into the lasing mode. (c) Since such a semi-classical spontaneous-emission rate does not fluctuate temporally but is constant in time, it does not contain any “noise”. It is important at this point to remind the reader that, according to Eq. (157)–(159), the spontaneous-emission rate is the source term in the photon-rate equation and not an independent noise. (d) Quantum fluctuations that could potentially turn the spontaneous-emission rate into a temporally unstable “noise” are not part of this derivation. Neither do any quantum fluctuations appear in the first, semi-classical derivation of the original Schawlow-Townes equation [9,10] or any of its semi-classical extensions. Therefore, the original Schawlow-Townes equation and all its semi-classical extensions have absolutely nothing to do with “noise”.

As one of the misunderstandings in laser physics, it has been re-iterated thousands of times in the literature that the laser linewidth be due to “spontaneous-emission noise”, often even called “Schawlow-Townes noise”. Our investigations demonstrate that this expression is a *contradictio in adiecto*, a contradiction between the noun “noise” and its explaining adjective “Schawlow-Townes”, because the Schawlow-Townes linewidth is not a result of “noise”. “Schawlow-Townes noise” in a physical sense does not exist. The expression “Schawlow-Townes noise” should, therefore, be avoided.

Nowhere in the derivation of the fundamental laser linewidth, the cw-laser linewidth, and the Schawlow-Townes approximation have we considered amplitude and phase fluctuations, or phase angles, or interference between electric fields. As we have shown in Section 5 and Ref. [3], the reason is that stimulated and spontaneous emission by an atom occur with a 90° phase difference with respect to the incident field, hence the interference term in Eqs. (123)–(125) vanishes and intensities or photon numbers add up. To underline this fact further, let us follow the derivation that leads to Eq. (172) in the opposite direction, e.g. directly within Eq. (172) by going backward from the right-hand side to the left-hand side. By use of Eqs. (39) and (169) one can rewrite the original Schawlow-Townes equation as

$$\Delta\nu_{L,ST} = \frac{2\pi h\nu_L(\Delta\nu_c)^2}{P_{out}} = 2\pi\Delta\nu_c \frac{h\nu_L}{P_{out}} \Delta\nu_c = \frac{1}{\tau_c} \frac{1}{R_{out}} \Delta\nu_c. \quad (173)$$

The right-hand side is still the original Schawlow-Townes equation, and it still contains all four approximations. When consecutively eliminating approximations (iv) that the gain equals the losses,

$$cg\varphi \stackrel{R_{st}=R_{decay}}{=} \frac{1}{\tau_c} \varphi \Rightarrow \Delta\nu_L = cg \frac{1}{R_{out}} \Delta\nu_c, \quad (174)$$

(iii) that there occur no intrinsic resonator losses,

$$R_{decay} \stackrel{1/\tau_{loss}=0}{=} R_{out} \Rightarrow \Delta\nu_L = cg \frac{1}{R_{decay}} \Delta\nu_c, \quad (175)$$

and (ii) that it is an ideal four-level laser,

$$cg = c\sigma_e(N_2 - N_1) \stackrel{4\text{-level}}{=} c\sigma_e N_2 = R_{sp} \Rightarrow \Delta\nu_L = \frac{R_{sp}}{R_{decay}} \Delta\nu_c, \quad (176)$$

from the original Schawlow-Townes equation, the generalized Schawlow-Townes equation for a cw-lasing mode reads

$$\Delta\nu_{L,ST} \stackrel{=}{=} \frac{R_{sp}}{R_{decay}} \Delta\nu_c. \quad (177)$$

It equals the general linewidth of a cw-lasing mode, given in Eq. (164). The fundamental linewidth of a cw-lasing mode is simply the ratio between the constant spontaneous-emission rate into the mode and the constant photon-decay rate out of the mode, i.e., it depends on intensities rather than amplitudes. This point was already clear from its original derivation [9,10] but somehow got lost in the scientific community in subsequent years.

At its operation temperature of $T = 78$ K, the four-level ammonia maser [113] oscillating at $\nu = 24$ GHz, i.e., $\lambda = 12.5$ mm, exhibits a thermal emission rate into the lasing mode, given by Eq. (34), of $R_{therm} = 67/\tau_c$, see Fig. 7(b), which is large compared to the spontaneous-emission rate, $R_{sp} = 1/\tau_c$, of an ideal four-level laser at the fictive point where the gain would equal the losses (black line in Fig. 7). Consequently, only the thermal emission was considered in the approximation of its linewidth [9], whereas spontaneous emission was neglected. For far-infrared lasers, e.g. hydrogen, deuterium, and iodide cyanide lasers [137] or THz quantum-cascade lasers [138], the thermal contribution also has to be considered. For a laser at $\lambda = 10$ μm , e.g. a CO₂ laser, the thermal emission of photons is still relevant. Schawlow and Townes replaced the thermal energy $k_B T$ in Eq. (137) by the photon energy $h\nu_L$, leading to Eq. (138). For a room-temperature ($T = 300$ K) laser at $\lambda = 1$ μm , e.g. a YAG:Yb³⁺ or InGaAs laser, Eq. (34) yields $R_{therm} = 1.5 \times 10^{-21}/\tau_c$, i.e., for such a laser the thermal emission of photons into the lasing mode is negligible compared to the influence of spontaneous emission, see Fig. 7(a). This conclusion holds even when partial conversion of absorbed pump power to heat elevates the temperature of the gain medium by more than 100 K [139].

We can learn more from the derivation of Eq. (172). Firstly, whereas the derivation of Eq. (172) applies consecutively the approximations (i)-(iv), we can deliberately first apply approximation (iv) by use of Eq. (171) and then approximation (iii) by use of Eq. (168),

$$\Delta\nu_L \stackrel{cw}{=} \frac{R_{sp}}{R_{decay}} \Delta\nu_c \stackrel{4\text{-level}}{=} \frac{cg}{\varphi/\tau_c} \Delta\nu_c \stackrel{R_{st}=R_{decay}}{=} \frac{1}{\varphi} \Delta\nu_c = \frac{1}{\tau_c} \frac{h\nu_L}{P_{out}} \Delta\nu_c \stackrel{1/\tau_{loss}=0}{=} \frac{1}{\tau_c} \frac{h\nu_L}{P_{out}} \Delta\nu_c = \frac{2\pi h\nu_L(\Delta\nu_c)^2}{P_{out}} = \Delta\nu_{L,ST} \quad (178)$$

to obtain the intermediate approximation $\Delta\nu_L = \Delta\nu_c/\varphi$. Comparison with Eq. (166), which for $\varphi_{vac} = 1$ becomes $\Delta\nu_L = \Delta\nu_c/(\varphi + 1)$, shows that the term “+ 1” is missing here in the denominator. This is, because replacement of the thermal energy $k_B T$ in Eq. (137) by the photon energy $h\nu_L$ in Eq. (138) assumes that one photon is coupled spontaneously into the mode per photon-decay time τ_c of the resonator, independent of the available pump power. For $P_{out} \rightarrow 0$, this incorrectly pump-independent term leads to divergence of the original Schawlow-Townes equation, $\Delta\nu_L \rightarrow \infty$, a “zero-photon catastrophe”, so to speak. Consequently, the Schawlow-Townes approximation overestimates the laser linewidth for any photon number, particularly for small photon numbers, see Fig. 36. Basically, all extended Schawlow-Townes equations found in the literature include this approximation. Instead, Eq. (166) leads to the physically correct result of $\Delta\nu_L = \Delta\nu_c$ for $g = 0$, at the transparency point.

Secondly, by setting the outcoupling losses equal to the total losses in Eq. (168) and then inserting the value of the passive linewidth $\Delta\nu_c$ into the Schawlow-Townes equation, one counts the intrinsic losses twice. Consequently, the original Schawlow-Townes equation overestimates the linewidth with respect to intrinsic losses, see Fig. 36(a). On the other hand, it represents a lower limit with respect to linewidth broadening due to reabsorption from the lower laser level, see Fig. 36(b).

In this Section, we have provided arguments that the fundamental laser linewidth and its four-fold approximation, the original Schawlow-Townes equation, are based on the semi-classical physical principle that the gain elongates the photon-decay time. The fundamental laser linewidth describes any three-level or four-level, transient or cw laser below or above its threshold, with its gain being smaller, equal, or larger compared to its losses. In a cw laser the gain is smaller than the losses, because the spontaneous-emission rate

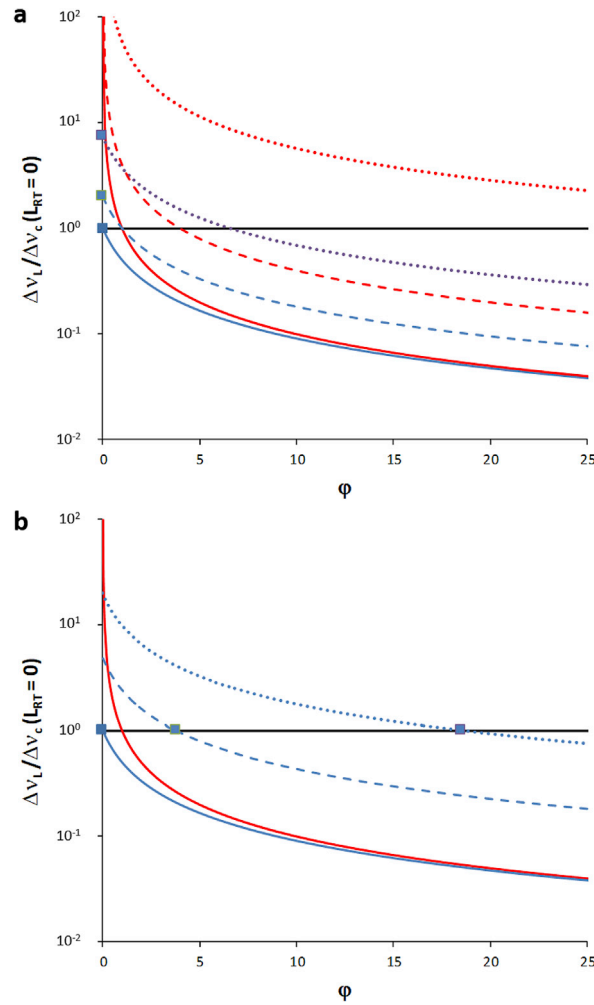


Fig. 36. Linewidth of a cw-lasing mode of Eq. (154) (blue curves) and Schawlow-Townes approximation of Eq. (138) or (172) [red curves, in (b) identical with each other], normalized to the linewidth of the underlying passive mode without intrinsic resonator losses, as a function of photon number φ . (a) Linewidth of an ideal four-level laser with increasing intrinsic resonator losses, $L_{RT} = 0$ (solid curve), $L_{RT} = 0.1$ (dashed curve), and $L_{RT} = 0.5$ (dotted curve), resulting in $\tau_c = 6.33 \times 10^{-9}$ s and $\Delta\nu_c = 2.52 \times 10^7$ s $^{-1}$ (normalization point of ordinate), $\tau_c = 3.16 \times 10^{-9}$ s and $\Delta\nu_c = 5.03 \times 10^7$ s $^{-1}$, or $\tau_c = 8.35 \times 10^{-10}$ s and $\Delta\nu_c = 1.91 \times 10^8$ s $^{-1}$, respectively, in a resonator with $\ell = 0.1$ m and $T_{out} = 0.1$. For the transparency point (squares), at which $\Lambda = 1$ and in an ideal four-level laser $\varphi = 0$ and $P_{out} = 0$, from Eq. (154) in each case the linewidth of the underlying passive mode correctly obtains. The incorrect consideration of intrinsic resonator losses by the Schawlow-Townes approximation increases the linewidth compared to the correct laser linewidth with the same intrinsic resonator losses, while its fictive assumption of $R_{sp} = 1/\tau_c$ for a four-level laser increases the linewidth for small photon numbers and forces the Schawlow-Townes approximation to diverge for $\varphi = 0$. (b) Linewidth of a three-level laser with increasing absorption cross section σ_a from the lower laser level, $\sigma_a = 0$ cm 2 (i.e., an ideal 4-level laser; solid curve), $\sigma_a = 2 \times 10^{-22}$ cm 2 (dashed curve), and $\sigma_a = 1 \times 10^{-21}$ cm 2 (dotted curve), for the resonator described above with $L_{RT} = 0$ and a gain medium with values of $N = 1 \times 10^{20}$ cm $^{-3}$, $\sigma_e = 4 \times 10^{-20}$ cm 2 . Again, for the transparency point (squares), at which $\Lambda = 1$, but in a three-level laser $\varphi > 0$ and $P_{out} > 0$, from Eq. (154) in each case the linewidth of the underlying passive mode correctly obtains. The deviation of the Schawlow-Townes approximation (solid, dashed, and dotted red curves, which are all identical with each other, because the Schawlow-Townes approximation neglects absorption from the lower laser level) are shown. Again, the Schawlow-Townes approximation diverges for $\varphi = 0$, whereas for large φ it underestimates the linewidth when reabsorption on the laser transition from the lower laser level occurs.

into the lasing mode is positive. Neither the fundamental laser linewidth, nor its four-fold approximation, the original Schawlow-Townes equation, is due to amplitude and phase fluctuations.

8. Comparison with other derivations

We have already seen in Section 6 that numerous, often contradicting accounts of the laser linewidth exist. In this Section, we compare our results of the fundamental laser linewidth with a few other derivations available in the literature.

8.1. Comparison with original linewidth derivation by Townes and co-workers

As emphasized above, the five parameters τ_L , $\Delta\nu_L$, Q_L , τ_c^{coh} , and ℓ_c^{coh} describing an absorbing mode ($g < 0$, $\Lambda < 1$), a passive mode ($g = 0$, $\Lambda = 1$, hence equaling τ_c , $\Delta\nu_c$, Q_c , τ_c^{coh} , and ℓ_c^{coh} , respectively), or an amplifying or lasing mode ($g > 0$, $\Lambda > 1$) equivalently quantify the spectral-coherence properties of that mode. If one of these parameters is known, then all of them are. The original work by Gordon, Zeiger, and Townes [9] described spectral coherence by utilization of electromagnetic powers and Q -factors and finally derived the linewidth of their ammonia maser via a Fourier transformation. From their derivation it becomes clear that the Q -factor remains finite in the cw-lasing case, see also Yariv [131]. Siegman [123] correctly concluded that Gordon, Zeiger, and Townes described a maser that oscillates below “threshold inversion”, by which he means the inversion at which the gain equals the losses. Because of the equivalence of the five parameters quantifying spectral coherence, the result by Gordon, Zeiger, and Townes [9] and subsequently Schawlow and Townes [10] directly translates into our results of Eq. (142) that the photon-decay time τ_L of a cw-lasing mode is finite and of Eq. (156) that in a cw-lasing mode the gain is smaller than the losses.

8.2. Comparison with Henry's derivation of his α -factor

In the situation of a cw-lasing mode, the result that the gain is smaller than the losses – or, equivalently, the stimulated-emission rate is smaller than the photon-decay rate – crucially depends on the fact that the spontaneous-emission rate into the lasing mode is positive, in the same way as the stimulated-emission rate is. In order to conserve energy, one atomic excitation must be converted into electromagnetic energy equaling one photon, and this energy must appear in the mode. The interference equation, which quantifies the interference between (a) the incident electromagnetic field and (b) the electromagnetic field emitted by the atom, can be expressed by either Eq. (123) in terms of the incident, emitted, and resulting electric fields \vec{E}_1 , \vec{E}_2 , and $\vec{E}_1 + \vec{E}_2$, respectively, Eq. (124) in terms of the incident, emitted, and resulting intensities I_1 , I_2 , and $I_1 + I_2$, respectively, or Eq. (125) in terms of the incident, emitted, and resulting photon numbers φ_1 , φ_2 , and $\varphi_1 + \varphi_2$, respectively. In spectral resonance, conservation of energy inevitably requires that the interference term vanishes, hence $\cos(\theta) = 0$ or $\theta = 90^\circ$, i.e., the emitted field occurs with a 90° phase difference with respect to the incident field. As has been shown in multiple ways in Section 5 and Ref. [3], this is indeed the case in stimulated and spontaneous emission. For $\cos(\theta) = 0$, each single emission process adds exactly the one additional photon, $\varphi_2 = 1$, that is emitted by the atom to the oscillating beam, such that $\varphi_1 \rightarrow \varphi_1 + 1$. Even in the picture sketched in Fig. 22(a), in which spontaneous emission would cause amplitude and phase fluctuations, thereby violating the conservation of energy in each single spontaneous-emission event, as has been pointed out by Henry [64], at least averaged over many such events the one photon expected from one spontaneous-emission event is indeed generated [3,64], because $\cos(\theta)$ in Eq. (123)–(125) would average out to zero. On the second page of his paper [64], Henry phrases it as follows: “The second term in [the equation $\Delta I_i = 1 + 2I_i^{1/2} \cos(\theta_i)$] averages to zero, so that on the average spontaneous emission causes a field intensity change equivalent to adding one photon to the mode”. Consequently, also in that picture, a cw-lasing mode would operate at a point where the gain is smaller than the losses, at least in average. Therefore, the same fundamental laser linewidth must derive.

8.3. Comparison with Weichel's derivation of the laser linewidth by use of the uncertainty principle

Weichel's concise derivation [140] of the Schawlow-Townes equation makes use of the uncertainty principle. We translate his equations into the notation used in the present work: his n equals φ , his K equals $c\sigma_e$, and his $\hbar\omega$ equals $\hbar\nu_L$.

Weichel commences with the same photon-rate equation as Eq. (155), with Eq. (161) and $\varphi_{vac} = 1$ inserted, but later sets the time derivative to zero for cw lasers, which is approximation (i) made by Gordon, Zeiger, and Townes, and derives $c\sigma_e$ for a laser operating far above threshold, where $\varphi \gg 1$, thereby neglecting spontaneous emission, $R_{sp} = 0$, which is equivalent to approximation (iv) made by Schawlow and Townes, $R_{st} = R_{decay}$:

$$\frac{d}{dt}\varphi = c\sigma_e N_2 + c\sigma_e(N_2 - N_1)\varphi - \frac{1}{\tau_c}\varphi \stackrel{R_{sp}=0}{\stackrel{cw}{\rightleftharpoons}} c\sigma_e(N_2 - N_1)\varphi = \frac{1}{\tau_c}\varphi \Rightarrow c\sigma_e = \frac{1}{\tau_c} \frac{1}{N_2 - N_1}. \tag{179}$$

He then makes use of the uncertainty principle, in the form

$$\Delta E \Delta t \geq \hbar = \frac{h}{2\pi}, \tag{180}$$

but obviously misses a factor of $\frac{1}{2}$ on the right-hand side of this inequality. For a coherent state, which has the minimum possible uncertainty, this inequality becomes an equality. Unlike Weichel, we will use the equality in the following. Inspired by the quantum-optical approaches and their interpretation of spontaneous emission by Fig. 22(a), which he understands as the ultimately correct description of the laser linewidth, Weichel derives from the optical energy E the energy uncertainty ΔE :

$$E = \varphi h\nu_L \Rightarrow \Delta E = \Delta\varphi h\nu_L + \varphi h\Delta\nu_L \approx \varphi h\Delta\nu_L. \tag{181}$$

Weichel interprets the two terms as amplitude fluctuations $\Delta\varphi$ and phase fluctuations $\Delta\nu_L$ caused by spontaneous emission. By use of the argument discussed in connection with Fig. 22(a) that above laser threshold amplitude fluctuations are suppressed by relaxation oscillations, he neglects the first term. Inserting Eq. (181) into Eq. (180) and eliminating h on both sides yields

$$2\pi\Delta\nu_L = \frac{1}{\varphi} \frac{1}{\Delta t}. \quad (182)$$

Weichel then interprets Δt as the time interval between two spontaneous-emission events. Consequently, $1/\Delta t$ should equal the spontaneous-emission rate,

$$\frac{1}{\Delta t} = c\sigma_e N_2. \quad (183)$$

Inserting Eq. (183) and the result of Eq. (179) into Eq. (182) yields

$$2\pi\Delta\nu_L \stackrel{R_{st}=R_{decay}}{=} \frac{1}{\varphi} c\sigma_e N_2 = \frac{1}{\varphi\tau_c} \frac{N_2}{N_2 - N_1}. \quad (184)$$

Weichel then inserts Eq. (170), thereby applying approximation (iii) made by Gordon, Zeiger, and Townes to the correct Eq. (169),

$$2\pi\Delta\nu_L \stackrel{1/\tau_{loss}=0}{=} \frac{N_2}{N_2 - N_1} \frac{h\nu_L}{P_{out}} \frac{1}{\tau_c^2}, \quad (185)$$

and the Fourier transformation of Eq. (39), resulting in

$$\Delta\nu_L = \frac{N_2}{N_2 - N_1} \frac{2\pi h\nu_L (\Delta\nu_c)^2}{P_{out}}. \quad (186)$$

This is the extended Schawlow-Townes equation that avoids approximation (ii) of an ideal four-level maser made by Gordon, Zeiger, and Townes. When setting $N_1 = 0$, thereby applying approximation (ii), one obtains the original Schawlow-Townes equation of Eq. (138).

What do we learn from this derivation? Firstly, it is again the same approximations (i)-(iv) that lead to the original Schawlow-Townes equation. Secondly, had Weichel not made the error in Eq. (180), he would have derived half the original Schawlow-Townes linewidth of Eq. (138) or its extended version of Eq. (186). In his interpretation of Eq. (181), Weichel claimed that he neglected amplitude fluctuations, whereas phase fluctuations caused by spontaneous emission were the physical reason for the laser linewidth. Therefore, his derivation seems to yield a simple proof of the quantum-optical interpretation of spontaneous emission by Fig. 22(a) and derivation of the laser linewidth, including the factor-of-two decrease in linewidth above threshold, because amplitude fluctuations are suppressed. However, Weichel made the same mistake that was previously made in the quantum-optical approaches. Instead of correctly interpreting the time interval Δt in the uncertainty principle as the duration of one vacuum fluctuation (being ~ 1 fs in the visible spectral range), by introducing Eq. (183) Weichel incorrectly interprets Δt as being related to spontaneous emission and, thereby, the uncertainty principle itself as a justification for the violation of energy conservation by spontaneous emission. It is, therefore, repeated here what was said in Section 5, namely that spontaneous emission converts one atomic excitation into one photon and this photon must appear in the mode. The uncertainty principle explains vacuum fluctuations, which cause the Poissonian photon statistics around the expectation value of the photon number, but it does not explain a deviation from the expectation value of the photon number by spontaneous emission. On the other hand, the fact that the spontaneous-emission rate in Eqs. (179) and (183) is a positive rate, resulting in the gain being smaller than the losses in Eq. (179) and leading to the fundamental laser linewidth, is not taken into account by Weichel, because the spontaneous-emission rate is considered negligible in Eq. (179). The correct evaluation of Eq. (179) yields

$$c\sigma_e = \frac{1}{\tau_c} \frac{1}{\varphi} \frac{1}{N_2 - N_1}. \quad (187)$$

8.4. Comparison with quantum-optical derivations of the laser linewidth

Although not subject to this paper, we briefly discuss here the relationship of our results with previous quantum-optical investigations of the laser linewidth. Obviously, the photon number φ in the mode corresponds to the mean value of the photon-statistical distribution in the quantum-optical picture:

$$\varphi \leftrightarrow \bar{n} = \sum_{n=0}^{\infty} n\rho_{nn}. \quad (188)$$

Here, ρ_{nn} is the diagonal element of the reduced field-density operator in photon-number representation.

The quantum-optical approach [129] to the laser linewidth starts from the photon-density-matrix description and accounts for the “flows of probability” between states with different photon number caused by the presence of excited atoms in a laser resonator, called the equation of motion [141] (or master equation):

$$\dot{\rho}_{nm} = - \left(\frac{N'_{nm} A}{1 + N_{nm} \frac{B}{A}} \right) \rho_{nm} + \left(\frac{\sqrt{nm} A}{1 + N_{n-1, m-1} \frac{B}{A}} \right) \rho_{n-1, m-1} - \frac{1}{2} \frac{\omega}{Q_c} (n+m) \rho_{nm} + \frac{\omega}{Q_c} \sqrt{(n+1)(m+1)} \rho_{n+1, m+1}. \quad (189)$$

Here, written in terms of variables used throughout this paper and preceding papers closely related to the subject (see, e.g., Ref. [6]),

$$A = 2R_P (\pi \nu_{Rabi} \tau_{atom}^{coh})^2 \quad (190)$$

is the unsaturated linear gain proportional to the pump rate R_P , with ν_{Rabi} being the vacuum Rabi frequency, i.e., the coupling strength between the atomic transition and the electromagnetic field [6], and τ_{atom}^{coh} being the atomic excitation coherence time, which takes into account de-coherence caused by spontaneous emission and external de-coherence caused by, e.g., phononic perturbations of the wavefunction of the excited state in a solid-state laser medium [6].

$$B = 4 (\pi \nu_{Rabi} \tau_{atom}^{coh})^2 A = 8R_P (\pi \nu_{Rabi} \tau_{atom}^{coh})^4 \quad (191)$$

is a saturation parameter, and

$$N_{nm} = \frac{1}{2} (n+m+2) + \frac{1}{16} (n-m)^2 \frac{B}{A}, \quad (192)$$

$$N'_{nm} = \frac{1}{2} (n+m+2) + \frac{1}{8} (n-m)^2 \frac{B}{A}. \quad (193)$$

The decay of photons out of the resonator is described by the terms containing $\omega/Q_c = 1/\tau_c$.

Therefrom, we obtain the master equation for the diagonal elements by setting $N_{nm} = N'_{nm} = n+1$:

$$\dot{\rho}_{nn} = - \frac{(n+1)A}{1 + (n+1) \frac{B}{A}} \rho_{nn} + \frac{nA}{1 + n \frac{B}{A}} \rho_{n-1, n-1} - \frac{\omega}{Q_c} n \rho_{nn} + \frac{\omega}{Q_c} (n+1) \rho_{n+1, n+1}. \quad (194)$$

Assuming a low gain saturation, $nB/A \ll 1$, we can simplify the denominators by use of $1/(1+x) \approx 1-x$ and obtain an approximate master equation,

$$\dot{\rho}_{nn} \approx - [A - B(n+1)] (n+1) \rho_{nn} + (A - Bn) n \rho_{n-1, n-1} - \frac{\omega}{Q_c} n \rho_{nn} + \frac{\omega}{Q_c} (n+1) \rho_{n+1, n+1}, \quad (195)$$

from which we can derive the quantum-optical rate equation for the average photon number \bar{n} :

$$\begin{aligned} \frac{d\bar{n}}{dt} &= \sum_{n=0}^{\infty} n \dot{\rho}_{nn} \approx - \sum_{n=0}^{\infty} n [A - B(n+1)] (n+1) \rho_{nn} + \sum_{n=1}^{\infty} n (A - Bn) n \rho_{n-1, n-1} - \sum_{n=0}^{\infty} n \frac{\omega}{Q_c} n \rho_{nn} + \sum_{n=0}^{\infty} n \frac{\omega}{Q_c} (n+1) \rho_{n+1, n+1} \\ &= - \sum_{n=0}^{\infty} n [A - B(n+1)] (n+1) \rho_{nn} + \sum_{k=0}^{\infty} (k+1) [A - B(k+1)] (k+1) \rho_{kk} - \frac{\omega}{Q_c} \bar{n} \\ &= \sum_{k=0}^{\infty} [A - B(k+1)] (k+1) \rho_{kk} - \frac{\omega}{Q_c} \bar{n} = \sum_{k=0}^{\infty} A (k+1) \rho_{kk} - \sum_{k=0}^{\infty} B (k+1) (k+1) \rho_{kk} - \frac{\omega}{Q_c} \bar{n} \\ &= A(\bar{n}+1) - \sum_{k=0}^{\infty} B(k^2 + 2k + 1) \rho_{kk} - \frac{\omega}{Q_c} \bar{n} = A(\bar{n}+1) - B(2\bar{n}+1) - B\bar{n}^2 - \frac{\omega}{Q_c} \bar{n}. \end{aligned} \quad (196)$$

Re-ordering the terms allows for a direct comparison between the quantum-optical and our semi-classical description in terms of rate equations using Eqs. (155) and (161):

$$\begin{aligned} \frac{d\bar{n}}{dt} &\approx A - B + (A - 2B)\bar{n} - \frac{\omega}{Q_c} \bar{n} - B\bar{n}^2 \\ \frac{d\phi}{dt} &= R_{sp} + cg\phi - \frac{1}{\tau_c} \phi. \end{aligned} \quad (197)$$

When inspecting this result, we immediately see that there is no direct general correspondence between the two descriptions, as the quantum-optical result contains a dependence on the average of the squared photon number. However, when we assume a Poissonian photon-number distribution, valid far above threshold, i.e.,

$$\rho_{nn} \approx e^{-\bar{n}} \frac{\bar{n}^n}{n!}, \quad (198)$$

we obtain $\overline{n^2} = \bar{n}(\bar{n}+1) \approx \bar{n}^2$ and can simplify the quantum-optical rate equation to

$$\frac{d\bar{n}}{dt} \approx A - B + (A - 2B)\bar{n} - \frac{\omega}{Q_c}\bar{n} - B\bar{n}^2 = A + A\bar{n} - B - 2B\bar{n} - B\bar{n}^2 - \frac{\omega}{Q_c}\bar{n} = A(\bar{n} + 1) - B(\bar{n} + 1)(\bar{n} + 1) - \frac{\omega}{Q_c}\bar{n}. \quad (199)$$

Thus, we can compare the result with our semi-classical rate equation:

$$\begin{aligned} \frac{d\bar{n}}{dt} &\approx [A - B(\bar{n} + 1)] + [A - B(\bar{n} + 1)]\bar{n} - \frac{\omega}{Q_c}\bar{n} \\ &\quad \downarrow \quad \quad \downarrow \quad \quad \downarrow \\ \frac{d\phi}{dt} &= c\sigma_e N_2 \quad + \quad cg\phi \quad - \frac{1}{\tau_c}\phi. \end{aligned} \quad (200)$$

Now a direct relation between the different terms is possible and the following conclusions can, thus, be derived. Firstly, the photon decay out of the resonator is implemented in identical manner in both equations. Secondly, by identifying the “gain” terms $A - B(\bar{n} + 1) \triangleq cg$, we obtain a correspondence between the first term in both equations only in the case when $c\sigma_e N_2 = cg$, which is the case for an ideal four-level laser. This is consistent with the fact that the master equation given here was deduced [141] under the assumption that excited atoms are “injected” into a resonator at the rate R_p , thereby implicitly assuming that there is no initially populated lower laser level that causes reabsorption. Since a direct correspondence can be drawn, we can deduce that also in the quantum-optical description of the cw laser the “gain” that is acting on the photons inside the lasing resonator mode is smaller than the “losses” of that lasing resonator mode and that the difference is exactly compensated for by the positive “spontaneous-emission” term. Based upon the same arguments as in Sections 7.2 and 7.4, the fact that quantum optically the gain is smaller than the losses must result in a finite linewidth for a cw laser, independent of the potential occurrence of amplitude and phase fluctuations, which quantum mechanically are still allowed in a cw laser, because the definition of “cw” in this case means $d\bar{n}/dt = 0$ and not $\dot{\rho}_{nn} = 0 \forall n$. Thus, the photon distribution can still fluctuate in time around the approximately Poissonian distribution, however with a temporally constant mean value \bar{n} , which is in accordance with the conservation of energy.

Based upon the master equation, two equivalent descriptions were proposed that allow one to derive a quantum-optical laser linewidth. The first description attributes the laser linewidth to the decay of the ensemble average of the electric field, which results from the decay of the off-diagonal elements of the density matrix [129]. The second description considers the Fokker-Planck equation of the coherent-state representation derived from the master equation and the assumption that spontaneous emission into the lasing mode induces amplitude and phase fluctuations. Above laser threshold, only phase fluctuations are considered to contribute to the laser linewidth in the form of phase diffusion, whereas amplitude fluctuations are damped or restricted in a steady state [129,142]. Thereby, both descriptions yield an equation for the quantum-optical laser linewidth that is exactly one-half of the original Schawlow-Townes equation [26,67,129,142].

We emphasize that these derivations also lead to a zero-photon catastrophe, which may be a result of exploiting several approximations for large photon numbers. Moreover, it is not the case that the semi-classical explanation given above is a simpler, incomplete or – as Coldren et al. put it [124] – more “intuitive” version of the more complex quantum-optical explanation. As we have demonstrated above, the physical principle underlying the original Schawlow-Townes equation is the fact that the gain is smaller than the losses because of the positive spontaneous-emission rate into the lasing mode, which turns out to be the same for the quantum-optical approach, too.

In 1964 Lamb [143] published a semi-classical theory of the laser, in which he addressed various aspects but not the laser linewidth. The need for a “better” quantum-optical treatment was subsequently justified by Scully and Lamb [144] by the observation that “the foregoing theory implies that laser radiation in an ideal steady state is absolutely monochromatic”. Since Lamb omitted the spontaneous-emission rate into the lasing mode in the semi-classical photon-rate equation of his theory [143], this observation is not surprising, as it is exactly this rate that induces the fundamental linewidth of a cw laser (Section 7.4). Obviously, this is not a failure of semi-classical physics, but a deliberate simplification by the author. Apparently, Scully and Lamb were not aware that the spontaneous-emission rate into the lasing mode induces the fundamental laser linewidth and its omission is the reason for the vanishing linewidth in Lamb’s semi-classical theory [143], because in their quantum-optical theory [129] they missed the same point (see this Sub-section), as did the other proponents [25,26,64,66,67] of the necessity to treat the fundamental laser linewidth quantum-optically. We emphasize again at the end of our investigation that this does by no means exclude that quantum fluctuations exist and have an influence on the laser linewidth, but these fluctuations do not explain the fundamental laser linewidth (Sections 7.3 and 7.4), its four-fold approximation, the original Schawlow-Townes equation (Section 7.5), or their reduction by a factor of two near the laser threshold (Section 5 and Ref. [3]).

9. Fundamental laser linewidth in a nut shell

In this final Section we summarize the derivation leading to the fundamental laser linewidth and its four-fold approximation, the original Schawlow-Townes equation. By omitting many of the details discussed above, we emphasize the physical principle underlying its derivation.

We assume a Fabry-Pérot-type resonator with mirror reflectivities of R_1 and R_2 and intrinsic round-trip losses of L_{RT} , including a

transparent active medium, i.e., without gain or absorption. The photon-decay time τ_c and, by Fourier transformation of the exponential decay in time, the FWHM linewidth $\Delta\nu_c$ of the Lorentzian-shaped spectral distribution peaked at the resonance frequency ν_L are given by Eqs. (23) and (39):

$$\frac{1}{\tau_c} = \frac{1}{\tau_{out}} + \frac{1}{\tau_{loss}} = \frac{-\ln[R_1 R_2 (1 - L_{RT})]}{t_{RT}} \Rightarrow \Delta\nu_c = \frac{1}{2\pi\tau_c}. \quad (201)$$

When the active medium inside the resonator is pumped, the gain g per unit length elongates the photon-decay time to τ_L and narrows the FWHM linewidth to $\Delta\nu_L$, represented by Eqs. (142), (143), and (147):

$$\frac{1}{\tau_L} = \frac{1}{\tau_c} - cg, \quad \tau_L = \Lambda\tau_c \Rightarrow \Delta\nu_L = \left| \frac{1}{2\pi\tau_L} \right| = \left| \Delta\nu_c - \frac{cg}{2\pi} \right| = \frac{1}{|\Lambda|} \Delta\nu_c. \quad (202)$$

The spectral-coherence factor Λ , by which the spectral coherence of the active mode increases due to the gain, is defined by Eq. (146):

$$\Lambda := \frac{1}{1 - cg\tau_c}. \quad (203)$$

In a lasing mode, by inserting the stimulated-emission rate R_{st} and the photon-decay rate R_{decay} , given by Eqs. (150) and (151),

$$R_{st} = cg\varphi, \quad R_{decay} = \frac{1}{\tau_c}\varphi, \quad (204)$$

respectively, into the spectral-coherence factor of Eq. (146), resulting in Eq. (152),

$$\Lambda = \frac{R_{decay}}{R_{decay} - R_{st}}, \quad (205)$$

the photon-decay time τ_L of Eq. (142) and the linewidth $\Delta\nu_L$ of the lasing mode of Eqs. (143) and (147) are represented by Eqs. (153) and (154):

$$\tau_L = \Lambda\tau_c = \frac{R_{decay}}{R_{decay} - R_{st}}\tau_c, \quad \Delta\nu_L = \frac{1}{|\Lambda|}\Delta\nu_c = \left| \frac{R_{decay} - R_{st}}{R_{decay}} \right| \Delta\nu_c. \quad (206)$$

This equation quantifies the fundamental linewidth of a single longitudinal lasing mode. The laser can have any energy-level system ranging from the extreme situation of an ideal 4-level laser to that of a 3-level laser and covering any intermediate situation, e.g., of a quasi-4-level or quasi-3-level laser. The lasing mode can operate below, at, or above threshold and in a cw or a transient lasing regime. As we see from Eqs. (153) and (154), spontaneous emission does not necessarily influence the linewidth, particularly not in a transient situation.

In a cw-lasing mode, the spontaneous-emission rate R_{sp} and the stimulated-emission rate R_{st} into this mode together compensate the photon-decay rate R_{decay} out of the resonator, resulting in the photon-decay time τ_L , according to the photon-rate equation (155) or (160):

$$\frac{d}{dt}\varphi = R_{sp} + R_{st} - R_{decay} = R_{sp} - \frac{1}{\tau_L}\varphi \stackrel{cw}{=} 0. \quad (207)$$

Consequently, the gain is smaller than the losses, as shown in Eq. (156),

$$R_{decay} - R_{st} \stackrel{cw}{=} R_{sp} > 0 \Rightarrow cg\tau_c = 1 - \frac{1}{\Lambda} \stackrel{cw}{<} 1, \quad (208)$$

and the photon-decay time τ_L of Eq. (142) and the linewidth $\Delta\nu_L$ of the lasing mode of Eqs. (143) and (147) become Eqs. (163) and (164), respectively:

$$\tau_L \stackrel{cw}{=} \frac{R_{decay}}{R_{sp}}\tau_c, \quad \Delta\nu_L \stackrel{cw}{=} \frac{R_{sp}}{R_{decay}}\Delta\nu_c. \quad (209)$$

The original Schawlow-Townes linewidth $\Delta\nu_{L,ST}$ is then straight-forwardly derived from the fundamental laser linewidth $\Delta\nu_L$ of Eq. (154) by applying the same four approximations applied during its first derivation, namely that the laser (i) is a truly continuous-wave (cw) laser, leading to Eq. (164), (ii) is an ideal four-level laser, (iii) exhibits no intrinsic losses, and (iv) one photon is coupled spontaneously into the lasing mode per photon-decay time τ_c of the resonator, independent of the available pump power, as shown in Eq. (172):

$$\Delta\nu_L \stackrel{cw}{=} \frac{R_{sp}}{R_{decay}} \Delta\nu_c \stackrel{4\text{-level}}{=} \frac{c\mathcal{G}}{\phi/\tau_c} \Delta\nu_c \stackrel{1/\tau_{loss}=0}{=} \frac{c\mathcal{G}h\nu_L}{P_{out}} \Delta\nu_c \stackrel{R_{st}=R_{decay}}{=} \frac{2\pi h\nu_L (\Delta\nu_c)^2}{P_{out}} = \Delta\nu_{L,ST}. \quad (210)$$

The original Schawlow-Townes approximation $\Delta\nu_{L,ST}$ and its various extended versions are merely of historical interest. The fundamental laser linewidth $\Delta\nu_L$ without these approximations is given by Eq. (154). It results from the fact that the gain elongates the losses and, in a cw lasing mode, the gain is smaller than the losses because of the spontaneous-emission rate, resulting in Eq. (164). This evidence holds true also in a quantum-optical derivation.

This is the most straight-forward and simplest derivation of the fundamental laser linewidth $\Delta\nu_L$ and its four-fold approximation, the original Schawlow-Townes equation $\Delta\nu_{L,ST}$. This derivation is entirely semi-classical. Since in the transient lasing situation of Eq. (154) the spontaneous-emission rate is not necessarily relevant and in a cw-lasing situation of Eq. (164) it is a semi-classical, temporally constant rate, obviously quantum fluctuations do not play a role. They do play a role by evoking the Poissonian photon statistics around the expectation value of the photon number in the lasing mode that has been the focus of this paper.

10. Conclusions

In this paper, we have summarized and put into context with each other several fundamental results, which enable us to draw a semi-classical picture of spectral coherence in passive, active, and lasing resonators that is consistent with Maxwell's equations and the law of energy conservation. Based upon the description of spectral coherence in a passive Fabry-Pérot resonator, a description of spontaneous and stimulated emission that correctly considers the phase aspect, and the semi-classical physical principle that in an active mode the gain elongates the photon-decay time and narrows the linewidth, as quantified by the spectral-coherence factor Λ , we have derived the fundamental laser linewidth and shown that the original Schawlow-Townes equation is a four-fold approximation of this linewidth. The fundamental laser linewidth describes lasers that have any energy-level system ranging from the extreme situation of an ideal 4-level laser to that of an ideal 3-level laser and covering any intermediate situation, e.g., of a quasi-4-level or quasi-3-level laser. The lasing mode can operate below, at, or above threshold and in a cw or a transient lasing regime, with its gain being smaller, equal, or larger compared to its losses. We have shown that, semi-classically as well as quantum-optically, in a cw-lasing mode the gain is smaller than the losses, because the spontaneous-emission rate into the lasing mode is positive. Neither the fundamental laser linewidth nor its four-fold approximation, the original Schawlow-Townes equation, is due to amplitude and phase fluctuations.

This raises two important questions that obviously need to be answered in order to establish a complete picture of spectral coherence in a lasing mode. Firstly, quantum fluctuations exist, so how precisely do they influence the fundamental laser linewidth? One might speculate that they lead to broadening of the fundamental laser linewidth. Secondly, what is the physical explanation of the quantum-mechanically derived factor-of-two reduction in laser linewidth near threshold [26,67]? We have already obtained a possible semi-classical answer to this second question and will discuss it in part II of this paper. Interestingly, it is disconnected from the first question, whose answer seems to be a bit more involved.

Acknowledgments

The authors thank Nur Ismail, Cristine Kores, and Dimitri Geskus for their contributions to the results and (in alphabetical order) Stephen Arnold, Gunnar Björk, Georges Boulon, Michael Damzen, John Dudley, Jérôme Faist, Hermann Haken, Franz Kärtner, Ursula Keller, Adam Kinos, Stefan Kröll, Olivier Martin, Jesper Mørk, Lukas Novotny, Valdas Pasiskevicius, Pepijn Pinkse, Gerhard Rempe, Rene de Ridder, Günter Steinmeyer, Martin Wegener, Kerry Vahala, Amnon Yariv and several other colleagues for helpful discussions and useful feedback. M.P. acknowledges financial support by the European Research Council via the ERC Advanced Grant "OPUS" No. 341206.

References

- [1] M. Pollnau, Counter-propagating modes in a Fabry-Pérot-type resonator, *Opt. Lett.* 43 (2018) 5033–5036.
- [2] N. Ismail, C.C. Kores, D. Geskus, M. Pollnau, Fabry-Pérot resonator: spectral line shapes, generic and related Airy distributions, linewidths, finesses, and performance at low or frequency-dependent reflectivity, *Opt. Express* 24 (2016) 16366–16389.
- [3] M. Pollnau, Phase aspect in photon emission and absorption, *Optica* 5 (2018) 465–474.
- [4] M. Pollnau, Are absorption and spontaneous or stimulated emission inverse processes? The answer is subtle!, *Appl. Phys. B* 125 (2019) 25.
- [5] M. Eichhorn, M. Pollnau, "Coherence manifestation in a continuous-wave laser," in *Laser Sources and Applications II*, in: J.I. Mackenzie, H. Jelínková, T. Taira, M.A. Ahmed (Eds.), *Proc. SPIE 9135*, SPIE, Bellingham, WA, 2014 paper 91350A.
- [6] M. Eichhorn, M. Pollnau, Spectroscopic foundations of lasers: spontaneous emission into a resonator mode, *IEEE J. Sel. Top. Quantum Electron.* 21 (2015) 9000216.
- [7] M. Pollnau, M. Eichhorn, Laser eigenvalue, coherence time, Q-factor, and linewidth, in: *Conference on Lasers and Electro-Optics*, Technical Digest, Optical Society of America, Washington, DC, 2015 paper JTh2A.93.
- [8] M. Pollnau, M. Eichhorn, The Schawlow-Townes linewidth – a threefold approximation, in: *Conference on Lasers and Electro-Optics Europe*, Conference Digest, European Physical Society, Mulhouse, 2015, p. 88, paper CA-P.39.
- [9] J.P. Gordon, H.J. Zeiger, C.H. Townes, "The maser—New type of microwave amplifier, frequency standard, and spectrometer, *Phys. Rev.* 99 (1955) 1264–1274.
- [10] A.L. Schawlow, C.H. Townes, Infrared and optical masers, *Phys. Rev.* 112 (1958) 1940–1949.
- [11] C. Fabry, A. Perot, Théorie et applications d'une nouvelle méthode de spectroscopie interférentielle, *Ann. Chim. Phys.* 16 (1899) 115–144.
- [12] J.M. Vaughan, *The Fabry-Perot Interferometer*, Adam Hilger, Bristol and Philadelphia, 1989, pp. 97–102. Ch. 3.2.3.
- [13] J.M. Vaughan, *The Fabry-Perot Interferometer*, Adam Hilger, Bristol and Philadelphia, 1989, p. 105. Ch. 3.3.2.
- [14] A.E. Siegman, *Lasers*, University Science Books, Mill Valley, CA, 1986, pp. 428–430. Ch. 11.4.
- [15] A.E. Siegman, *Lasers*, University Science Books, Mill Valley, CA, 1986, pp. 413–428. Ch. 11.3.
- [16] A.E. Siegman, *Lasers*, University Science Books, Mill Valley, CA, 1986, pp. 432–440. Ch. 11.5.

- [17] A.E. Siegman, Lasers, University Science Books, Mill Valley, CA, 1986, pp. 398–408. Ch. 11.1.
- [18] O. Svelto, Principles of Lasers, fifth ed., Springer, New York, 2010, pp. 142–146. Ch. 4.5.1.
- [19] O. Svelto, Principles of Lasers, fifth ed., Springer, New York, 2010, pp. 169–171. Ch. 5.3.
- [20] B.E.A. Saleh, M.C. Teich, Fundamentals of Photonics, second ed., Wiley-Interscience, Hoboken, NJ, 2007, pp. 571–572. Ch. 15.1.
- [21] B.E.A. Saleh, M.C. Teich, Fundamentals of Photonics, second ed., Wiley-Interscience, Hoboken, NJ, 2007, pp. 254–257. Ch. 7.1B.
- [22] B.E.A. Saleh, M.C. Teich, Fundamentals of Photonics, second ed., Wiley-Interscience, Hoboken, NJ, 2007, pp. 62–66. Ch. 2.5B.
- [23] K. Petermann, Calculated spontaneous emission factor for double-heterostructure injection lasers with gain-induced waveguiding, *IEEE J. Quantum Electron.* QE-15 (1979) 566–570.
- [24] W.A. Hamel, J.P. Woerdman, Nonorthogonality of the longitudinal eigenmodes of a laser, *Phys. Rev. A* 40 (1989) 2785–2787.
- [25] M. Lax, Classical noise. V. Noise in self-sustained oscillators, *Phys. Rev.* 160 (1967) 290–307.
- [26] H. Haken, Laser Theory, Vol. XXV/2c of Encyclopedia of Physics, Springer, 1970.
- [27] B.E.A. Saleh, M.C. Teich, Fundamentals of Photonics, second ed., Wiley-Interscience, Hoboken, NJ, 2007, p. 368. Ch. 10.1.
- [28] O. Svelto, Principles of Lasers, fifth ed., Springer, New York, 2010, p. 164. Ch. 5.1.
- [29] Wikipedia, Longitudinal mode. https://en.wikipedia.org/wiki/Longitudinal_mode (downloaded 2018-09-05).
- [30] A. Yariv, Quantum Electronics, third ed., Wiley, New York, 1989, pp. 99–100. Ch. 5.7.
- [31] R. Dändliker, The concept of modes in optics and photonics, *Proc. SPIE* 3831 (2000) 193–198.
- [32] O. Svelto, Principles of Lasers, Springer, New York, 2010, pp. 19–22. Ch. 2.2.1.
- [33] L.A. Coldren, S.W. Corzine, M.L. Masanovic, Diode Lasers and Photonic Integrated Circuits, second ed., Wiley, Hoboken, NJ, 2012, pp. 559–561. Ch. A4.1.
- [34] M. Planck, Ueber das Gesetz der Energieverteilung im Normalspectrum, *Ann. Phys.* 309 (1901) 553–563.
- [35] Bose, Plancks Gesetz und Lichtquantenhypothese, *Z. Phys.* 26 (1924) 178–181.
- [36] T.E. Murphy, Design, fabrication and measurement of integrated Bragg grating optical filters, Ph.D. Thesis, Massachusetts Institute of Technology (2001) 58–59. Ch. 2.3.
- [37] A.E. Siegman, Lasers, University Science Books, Mill Valley, CA, 1986, pp. 559–561. Ch. 14.1.
- [38] E.I. Green, The story of Q, *Am. Sci.* 43 (1955) 584–594.
- [39] K.S. Johnson, Transmission Circuits for Telephonic Communication, fifth ed., D. Van Nostrand, Inc., New York, 1931, pp. 174–179. Ch. XV.
- [40] H.H. Hansen, A type of electrical resonator, *J. Appl. Phys.* 9 (1938) 654–663.
- [41] N. Wiener, Generalized harmonic analysis, *Acta Math.* 55 (1930) 117–258.
- [42] A. Khintchine, Korrelations-theorie der stationären stochastischen Prozesse, *Math. Ann.* 109 (1934) 604–615.
- [43] R.J. Glauber, The quantum theory of optical coherence, *Phys. Rev.* 130 (1963) 2529–2539.
- [44] L. Mandel, E. Wolf, Coherence properties of optical fields, *Rev. Mod. Phys.* 37 (1965) 231–287.
- [45] L. Mandel, E. Wolf, Spectral coherence and the concept of cross-spectral purity, *J. Opt. Soc. Am. B* 66 (1976) 529–535.
- [46] T. Young, An account of some cases of the production of colours, not hitherto described, *Phil. Trans. Roy. Soc. Lond.* XCII 12 (1802) 387–397.
- [47] J.M. Dudley, G. Genty, S. Coen, Supercontinuum generation in photonic crystal fiber, *Rev. Mod. Phys.* 78 (2006) 1135–1184.
- [48] B.E.A. Saleh, M.C. Teich, Fundamentals of Photonics, second ed., Wiley-Interscience, Hoboken, NJ, 2007, pp. 408–409. Ch.11.1.
- [49] A. Kastler, “Atomes à l’intérieur d’un interféromètre Perot-Fabry, *Appl. Opt.* 1 (1962) 17–24.
- [50] J.O. Stoner, PEPSIOS purely interferometric high-resolution scanning spectrometer. III. Calculation of interferometer characteristics by a method of optical transients, *J. Opt. Soc. Am.* 56 (1966) 370–376.
- [51] G. Koppelman, Multiple-beam interference and natural modes in open resonators, in: E. Wolf (Ed.), Progress in Optics, 7, North Holland, Amsterdam and London, 1969, pp. 1–66. Ch. 1.
- [52] F. Bayer-Helms, Analyse von Linienprofilen. I Grundlagen und Messeinrichtungen, *Z. Angew. Phys.* 15 (1963) 330–338.
- [53] J.B. Kumer, W.G. Uplinger, Elsasser-related approximation to the Airy function, *Appl. Opt.* 22 (1983) 3675–3676.
- [54] R. Remmert, Convergent series of meromorphic functions, in: S. Axler, F.W. Gehring (Eds.), Theory of Complex Functions, fourth ed., Springer, New York, 1998, pp. 321–341. Ch. 11.
- [55] J.M. Llorens, J. Buencuerpo, P.A. Postigo, Absorption features of the zero frequency mode in an ultra-thin slab, *Appl. Phys. Lett.* 105 (2014) 231115.
- [56] J.W. Strutt (Lord Rayleigh), Investigations in optics, with special reference to the spectroscopy, *Phil. Mag.* 8 (1879), 261–274, 403–411, 477–486.
- [57] I. Juvells, A. Carnicer, J. Ferré-Borrull, E. Martín-Badosa, M. Montes-Usategui, Understanding the concept of resolving power in the Fabry-Perot interferometer using a digital simulation, *Eur. J. Phys.* 27 (2006) 1111–1119.
- [58] M. Suter, P. Dietiker, Calculation of the finesse of an ideal Fabry-Perot resonator, *Appl. Opt.* 53 (2014) 7004–7010.
- [59] A. Einstein, Zur Quantentheorie der Strahlung, *Phys. Z.* 18 (1917) 121–128.
- [60] H. Kopfermann, R. Ladenburg, Untersuchungen über die anomale Dispersion angeregter Gase. II. Teil. Anomale Dispersion in angeregtem Neon (Einfluß von Strom und Druck, Bildung und Vernichtung angeregter Atome), *Z. Phys.* 48 (1928) 26–50.
- [61] W.E. Lamb Jr., Anti-photon, *Appl. Phys. B* 60 (1995) 77–84.
- [62] P.A.M. Dirac, The quantum theory of the emission and absorption of radiation, *Proc. R. Soc. A* 114 (1927) 243–265.
- [63] V. Weisskopf, Probleme der neueren Quantentheorie des Elektrons, *Naturwissenschaften* 23 (1935) 631–637, 647–653, 669–674.
- [64] C.H. Henry, Theory of the linewidth of semiconductor lasers, *IEEE J. Quantum Electron.* 18 (1982) 259–264.
- [65] P. Grivet, A. Blaquièrre, Masers and classical oscillators, in: J. Fox (Ed.), Symposium on Optical Masers, Polytechnique Institute, Brooklyn, 1963, pp. 69–93.
- [66] H. Risken, Correlation function of the amplitude and of the intensity fluctuations for a laser model near threshold, *Z. Phys.* 191 (1966) 302–312.
- [67] R.D. Hempstead, M. Lax, Classical noise. VI. Noise in self-sustained oscillators near threshold, *Phys. Rev.* 161 (1967) 350–366.
- [68] R.J. Glauber, Coherent and incoherent states of the radiation field, *Phys. Rev.* 131 (1963) 2766–2788.
- [69] H.A. Lorentz, Über die Beziehung zwischen der Fortpflanzungsgeschwindigkeit des Lichtes und der Körperdichte, *Ann. Phys.* 245 (1880) 641–665.
- [70] M. Sargent III, M.O. Scully, W.E. Lamb Jr., Laser Physics, sixth ed., Westview Press, Boulder and Oxford, 1993, pp. 34–42. Ch. 3.2.
- [71] P.W. Milonni, Semiclassical and quantum-electrodynamical approaches in nonrelativistic radiation theory, *Phys. Rep.* 25 (1976) 1–81.
- [72] M. Cray, M.L. Shih, P.W. Milonni, Stimulated emission, absorption, and interference, *Am. J. Phys.* 50 (1982) 1016–1021.
- [73] P.W. Milonni, J.H. Eberly, Lasers, Wiley, New York, 1988, pp. 27–33. Ch. 2.2.
- [74] R. de L. Kronig, On the theory of the dispersion of X-rays, *J. Opt. Soc. Am.* 12 (1926) 547–557.
- [75] H.A. Kramers, La diffusion de la lumière par les atomes, in: Atti del Congresso Internazionale della Fisica (Transactions of Volta Centenary Congress) Como, vol. 2, 1927, pp. 545–557.
- [76] H.S. Sommers, Spontaneous power and the coherent state of injection lasers, *J. Appl. Phys.* 45 (1974) 1787–1793.
- [77] H.S. Sommers, Threshold and oscillation of injection lasers: a critical review of laser theory, *Solid State Electron.* 25 (1982) 25–44.
- [78] A.E. Siegman, Lasers, University Science Books, Mill Valley, CA, 1986, pp. 510–524. Ch. 13.
- [79] G. Björk, Y. Yamamoto, Analysis of semiconductor microcavity lasers using rate equations, *IEEE J. Quantum Electron.* 27 (1991) 2386–2396.
- [80] W. Heisenberg, Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik, *Z. Phys.* 43 (1927) 172–198.
- [81] H.P. Robertson, The uncertainty principle, *Phys. Rev.* 34 (1929) 163–164.
- [82] J.C. Maxwell, VIII. A dynamical theory of the electromagnetic field, *Phil. Trans. Roy. Soc. Lond.* 155 (1865) 459–512.
- [83] L.I. Schiff, Quantum Mechanics, third ed., McGraw-Hill, New York, 1968, pp. 531–533. Sec. 57.
- [84] L.I. Schiff, Quantum Mechanics, third ed., McGraw-Hill, New York, 1968, pp. 398–406. Sec. 44.
- [85] E. Fermi, Nuclear Physics, The University of Chicago Press, Chicago, 1950 formula VIII.2.
- [86] R.P. Feynman, R.B. Leighton, M. Sands, The Feynman Lectures on Physics, I, Addison-Wesley, Reading, MA, 1963, pp. 31–34. Ch. 31.
- [87] V.F. Weisskopf, How light interacts with matter, *Sci. Am.* 219 (1968) 60–71.

- [88] M. Motsch, M. Zeppenfeld, P.W.H. Pinkse, G. Rempe, Cavity-enhanced Rayleigh scattering, *New J. Phys.* 12 (2010) 063022.
- [89] A. Sommer, E.M. Bothschafter, S.A. Sato, C. Jakubeit, T. Latka, O. Razskazovskaya, H. Fattahi, M. Jobst, W. Schweinberger, V. Shirvanyan, V.S. Yakovlev, R. Kienberger, K. Yabana, N. Karpowicz, M. Schultze, F. Krausz, Attosecond nonlinear polarization and light-matter energy transfer in solids, *Nature* 534 (2016) 86–90.
- [90] L. Novotny, *Lecture Notes on Electromagnetic Fields and Waves*, ETH Zurich, Switzerland, 2013, pp. 82–83. Ch. 6.3.
- [91] L. Novotny, B. Hecht, *Principles of Nano-Optics*, Cambridge University Press, Cambridge, 2006, pp. 261–265. Ch. 8.3.1.
- [92] D.J. Griffiths, *Introduction to Electrodynamics*, third ed., Prentice Hall, Upper Saddle River, NJ, 1999, pp. 444–451. Ch. 11.1.2.
- [93] G. Bruhat, A. Kastler, *Optique*, sixth ed., Masson, Paris, 1965. Sect. 34, pp. 59–61, Sect. 55, pp. 95–96, and Sect. 103, pp. 174–176.
- [94] A. Kastler, Review: masers and lasers, *Eur. J. Phys.* 7 (1986) 69–76.
- [95] M. Bertolotti, *Masers and Lasers: an Historical Approach*, second ed., CRC Press, Boca Raton, FL, 2015, p. 6. Ch. 1.
- [96] M. Sargent III, M.O. Scully, W.E. Lamb Jr., *Laser Physics*, sixth ed., Westview Press, Boulder and Oxford, 1993, pp. 242–255. Ch. XV “Coherent states”.
- [97] R.H. Dicke, Coherence in spontaneous radiation processes, *Phys. Rev.* 93 (1954) 99–110.
- [98] M. Gross, S. Haroche, Superradiance: An essay on the theory of collective spontaneous emission, *Phys. Rep.* 93 (1982) 301–396.
- [99] M. Scheibner, T. Schmidt, L. Worschech, A. Forchel, G. Bacher, T. Passow, D. Hommel, Superradiance of quantum dots, *Nat. Phys.* 3 (2007) 106–110.
- [100] T.V. Raziman, O.J.F. Martin, Does the real part contain all the physical information? *J. Opt.* 18 (2016) 095002.
- [101] J.A. Armstrong, N. Bloembergen, J. Ducuing, P.S. Pershan, Interactions between light waves in a nonlinear dielectric, *Phys. Rev.* 127 (1962) 1918–1939.
- [102] E. Rosencher, C. Fabre, Oscillation characteristics of continuous-wave optical parametric oscillators: beyond the mean-field approximation, *J. Opt. Soc. Am. B* 19 (2002) 1107–1116.
- [103] G. Griffel, S. Arnold, D. Taskent, A. Serpengüzel, J. Connolly, N. Morris, Morphology-dependent resonances of a microsphere-optical fiber system, *Opt. Lett.* 21 (1996) 695–697.
- [104] B.E. Little, J.P. Laine, H.A. Haus, Analytic theory of coupling from tapered fibers and half-blocks into microsphere resonators, *J. Lightwave Technol.* 17 (1999) 704–715.
- [105] M. Cai, O. Painter, K.J. Vahala, Observation of critical coupling in a fiber taper to a silica-microsphere whispering-gallery mode system, *Phys. Rev. Lett.* 85 (2000) 74–77.
- [106] A.E. Siegman, Excess spontaneous emission in non-Hermitian optical systems. I. Laser amplifiers, *Phys. Rev. A* 39 (1989) 1253–1263.
- [107] A.E. Siegman, Excess spontaneous emission in non-Hermitian optical systems. II. Laser oscillators, *Phys. Rev. A* 39 (1989) 1264–1268.
- [108] A.M. van der Lee, N.J. van Druten, A.L. Mieremet, M.A. van Eijkelenborg, Å.M. Lindberg, M.P. van Exter, J.P. Woerdman, Excess quantum noise due to nonorthogonal polarization modes, *Phys. Rev. Lett.* 79 (1997) 4357–4360.
- [109] H. Gerhardt, H. Welling, A. Güttner, Measurements of the laser linewidth due to quantum phase and quantum amplitude noise above and below threshold, *J. Z. Phys.* 253 (1972) 113–126.
- [110] H. Gerhardt, H. Welling, A. Güttner, Observation of quantum-phase and quantum-amplitude noise for a laser below and above threshold, *Phys. Lett.* 40 (1972) 191–193.
- [111] E.T. Jaynes, F.W. Cummings, Comparison of quantum and semiclassical radiation theories with application to the beam maser, *Proc. IEEE* 51 (1963) 89–109.
- [112] E.M. Purcell, R.V. Pound, A nuclear spin system at negative temperature, *Phys. Rev.* 81 (1951) 279–280.
- [113] J.P. Gordon, H.J. Zeiger, C.H. Townes, Molecular microwave oscillator and new hyperfine structure in the microwave spectrum of NH_3 , *Phys. Rev.* 95 (1954) 282–284.
- [114] C.H. Townes, The ammonia spectrum and line shapes near 1.25-cm wave-length, *Phys. Rev.* 70 (1946) 665–671.
- [115] T.H. Maiman, Stimulated optical radiation in Ruby, *Nature* 187 (1960) 493–494.
- [116] E.I. Gordon, Optical maser oscillators and noise, *Bell Syst. Tech. J.* 43 (1964) 507–539.
- [117] R.V. Pound, Spontaneous emission and the noise figure of maser amplifiers, *Ann. Phys.* 1 (1957) 24–32.
- [118] W.G. Wagner, G. Birnbaum, Theory of quantum oscillators in multimode cavity, *J. Appl. Phys.* 32 (1961) 1185–1193.
- [119] K. Shimoda, Theory of masers for higher frequencies, *Sci. Pap. Inst. Phys. Chem. Res.* 55 (1961) 1–6.
- [120] A. Blaquière, Largeur de raie d’un oscillateur laser, considère comme le siège d’une réaction en chaîne, *C. R. Acad. Sci.* 255 (1962) 3141–3143.
- [121] J.A. Fleck Jr., Linewidth and conditions for steady oscillation in single and multiple element lasers, *J. Appl. Phys.* 34 (1963) 2997–3003.
- [122] A.G. Fox, T. Li, Resonant modes in a maser interferometer, *Bell Syst. Tech. J.* 40 (1961) 453–488.
- [123] A.E. Siegman, *Lasers*, University Science Books, Mill Valley, CA, 1986, pp. 447–454. Ch. 11.7.
- [124] L.A. Coldren, S.W. Corzine, M.L. Mašanović, *Diode Lasers and Photonic Integrated Circuits*, second ed., Wiley, Hoboken, NJ, 2012, pp. 292–294. Ch. 5.5.2.
- [125] A.K. Ghatak, K. Thyagarajan, *Optical Electronics*, Cambridge University Press, Cambridge, 1989, pp. 252–256.
- [126] G.M. Stéphan, Semiclassical study of the laser transition, *Phys. Rev. A* 55 (1997) 1371–1384.
- [127] J.T. Verdeyen, *Laser Electronics*, third ed., Prentice Hall, Englewood Cliffs, NJ, 1995. Ch. 8.8.
- [128] J.B. Khurgin, G. Sun, Comparative analysis of spasers, vertical-cavity surface-emitting lasers and surface-plasmon-emitting diodes, *Nat. Photonics* 8 (2014) 468–473.
- [129] M. Sargent III, M.O. Scully, W.E. Lamb Jr., *Laser Physics*, sixth ed., Westview Press, Boulder and Oxford, 1993, pp. 281–295. Ch. 17.
- [130] O. Svelto, *Principles of Lasers*, fifth ed., Springer, New York, 2010, p. 298, ch. 7.9.
- [131] A. Yariv, *Quantum Electronics*, third ed., Wiley, New York, 1989, pp. 577–582. Ch. 21.2.
- [132] T. Okoshi, K. Kikuchi, A. Nakayama, Novel method for high resolution measurement of laser output spectrum, *Electron. Lett.* 16 (1980) 630–631.
- [133] J.W. Dawson, N. Park, K.J. Vahala, An improved delayed self-heterodyne interferometer for linewidth measurements, *IEEE Photon. Technol. Lett.* 4 (1992) 1063–1066.
- [134] A.E. Siegman, *Lasers*, University Science Books, Mill Valley, CA, 1986, pp. 502–504. Ch. 13.
- [135] O. Svelto, *Principles of Lasers*, Springer, New York, 2010, pp. 256–257, ch. 7.2.1.
- [136] A. Yariv, *Quantum Electronics*, third ed., Wiley, New York, 1989, pp. 196–200. Ch. 9.3.
- [137] H. Steffen, F.K. Kneubühl, Resonator interferometry of pulsed submillimeter-wave lasers, *IEEE J. Quantum Electron.* QE-4 (1968) 992–1008.
- [138] J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A.L. Hutchinson, A.Y. Cho, Quantum cascade laser, *Science* 264 (1994) 553–556.
- [139] M. Pollnau, P.J. Hardman, M.A. Kern, W.A. Clarkson, D.C. Hanna, Upconversion-induced heat generation and thermal lensing in Nd:YLF and Nd:YAG, *Phys. Rev. B* 58 (1998) 16076–16092.
- [140] H. Weichel, The uncertainty principle and the spectral width of a laser beam, *Am. J. Phys.* 44 (1976) 839–840.
- [141] M. Sargent III, M.O. Scully, W.E. Lamb Jr., *Laser Physics*, sixth ed., Westview Press, Boulder and Oxford, 1993, p. 285. Ch. 17, Eq. (16).
- [142] R. Paschotta. http://www.rp-photonics.com/Schawlow-Townes_linewidth.pdf (version 2013-06-27).
- [143] W.E. Lamb Jr., Theory of an optical maser, *Phys. Rev.* 134 (1964) A1429–A1450.
- [144] M.O. Scully, W.E. Lamb Jr., Quantum theory of an optical maser. I. General theory, *Phys. Rev.* 159 (1967) 208–226.