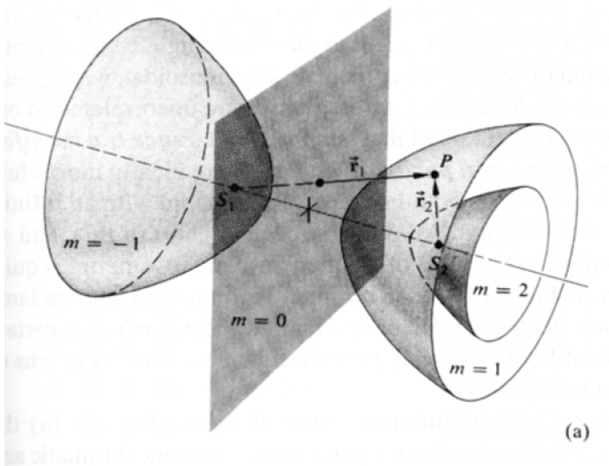
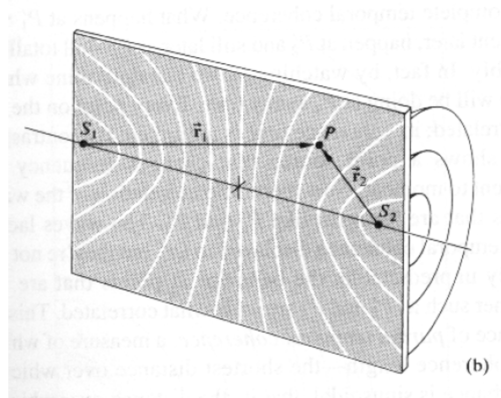
Interference and Interferometry [Pedrotti^3 Ch. 7 & Ch. 8]

Note: Read Ch. 3 & 7 E&M Waves and Superposition of Waves and Meet with TAs and/or Dr. Lai if necessary.





Condition for interference to occur

1. Same frequency.
2. Coherent. Constant phase difference.

Conventional quasimonochromatic sources produce light that is a mix of photon wavetrains. At each illuminated point in space there is a net field that oscillates nicely for less than 10 ns or so before it randomly changes phase. The interval over which the lightwave resembles a sinusoid is a measure of its **temporal coherence.** The corresponding spatial extent over which the lightwave oscillates in a regular, predictable way is the coherence length.

Amplitude-Splitting Interferometers

If a lightwave is split to two and bring back together again at a detector, interference would result, as long as the original coherence between the two had not been destroyed.

Examples: Dielectric films, Newtons' Rings

Wavefront-Splitting Interferometers

The main problem introducing interference is the sources: they must be coherent.
And yet separate, independent, adequately coherent sources, other that the modern laser, don't exist.
Thomas Young in his double-beam experiment took a single wavefront, split off from it two coherent portions, and had them interfere.

Examples: Young's double slit interferometer, Fresnel's double mirror/prism

Mirrored Interferometers

Examples: Michelson interferometer, Mach-Zehnder Interferometer, Sagnac Interferometer

Multiple Beam Interference

Example: Fabry-Perot Interferometer

General Consideration

Suppose there are two plane wave $\vec{E}_1$ and $\vec{E}_2$ described by

$$\vec{E}_1=(\vec{r},t)=\vec{E}_{01}\cos(\vec{k}_1\cdot\vec{r}-\omega t+\varepsilon_1)$$
$$\vec{E}_2=(\vec{r},t)=\vec{E}_{02}\cos(\vec{k}_2\cdot\vec{r}-\omega t+\varepsilon_2)$$

The irradiance $I=\epsilon v<\vec{E}^2>_T$.

Since $\vec{E}=\vec{E}_1+\vec{E}_2$,

$$\vec{E}\cdot\vec{E}=\vec{E}_1\cdot\vec{E}_1+\vec{E}_2\cdot\vec{E}_2+2\vec{E}_1\cdot\vec{E}_2$$

Let

$$I=I_1+I_2+I_{12}$$

where

$$I_1=\epsilon v<\vec{E}_1^2>_T$$
$$I_2=\epsilon v<\vec{E}_2^2>_T$$
$$I_{12}=2\epsilon v<\vec{E}_1\cdot\vec{E}_2>_T$$

I_{12} is known as the interference term.

$$\begin{aligned}\vec{E}_1\cdot\vec{E}_2 &= \vec{E}_{01}\cdot\vec{E}_{02}\cos(\vec{k}_1\cdot\vec{r}-\omega t+\varepsilon_1)\times\cos(\vec{k}_2\cdot\vec{r}-\omega t+\varepsilon_2)\\ &= \vec{E}_{01}\cdot\vec{E}_{02}\left[\cos(\vec{k}_1\cdot\vec{r}+\varepsilon_1)\cos\omega t+\sin(\vec{k}_1\cdot\vec{r}+\varepsilon_1)\sin\omega t\right]\times\left[\cos(\vec{k}_2\cdot\vec{r}+\varepsilon_2)\cos\omega t+\sin(\vec{k}_2\cdot\vec{r}+\varepsilon_2)\sin\omega t\right]\end{aligned}$$

Thus

$$\left\langle\vec{E}_1\cdot\vec{E}_2\right\rangle_T=\frac{1}{2}\vec{E}_{01}\cdot\vec{E}_{02}\cos(\vec{k}_1\cdot\vec{r}+\varepsilon_1-\vec{k}_2\cdot\vec{r}-\varepsilon_2)$$

and

$$I_{12}=\epsilon v\vec{E}_{01}\cdot\vec{E}_{02}\cos(\vec{k}_1\cdot\vec{r}+\varepsilon_1-\vec{k}_2\cdot\vec{r}-\varepsilon_2)=\epsilon v\vec{E}_{01}\cdot\vec{E}_{02}\cos(\delta)$$

where δ is the phase difference arising from a combined path length and initial phase angle difference. If $\vec{E}_{01}$ and $\vec{E}_{02}$ are perpendicular, $I_{12}=0$. When $\vec{E}_{01}$ and $\vec{E}_{02}$ are parallel

$$I_{12}=\epsilon vE_{01}E_{02}\cos\delta=2\sqrt{I_1I_2}\cos\delta$$

Therefore,

$$I=I_1+I_2+2\sqrt{I_1I_2}\cos\delta$$

Maximum occur when $\delta=0,\pm2\pi,\pm4\pi,...$

$$I_{max}=I_1+I_2+2\sqrt{I_1I_2}$$

This is called **total constructive interference**. The phase difference between the two waves is an integer multiple of 2π. This is called in-phase.

Minimum occur when $\delta=\pm\pi,\pm3\pi,\pm5\pi,...$

$$I_{min}=I_1+I_2-2\sqrt{I_1I_2}$$

This is called **total destructive interference**.

When $0<\cos\delta<1$, the waves are out-of-phase, $I_1+I_2<I<I_{max}$, the result is **constructive interference**.

When $\delta=90^\circ$, $\cos\delta=0$, the result is 90° out-of-phase, $I=I_1+I_2$.

When $-1<\cos\delta<0$, the waves are out-of-phase, $I_1+I_2>I>I_{min}$, the result is **destructive interference**.

If $E_{01}=E_{02}$, $I=4I_0\cos^2\frac{\delta}{2}$, $I_{min}=0$, $I_{max}=4I_0$.

For spherical wave,

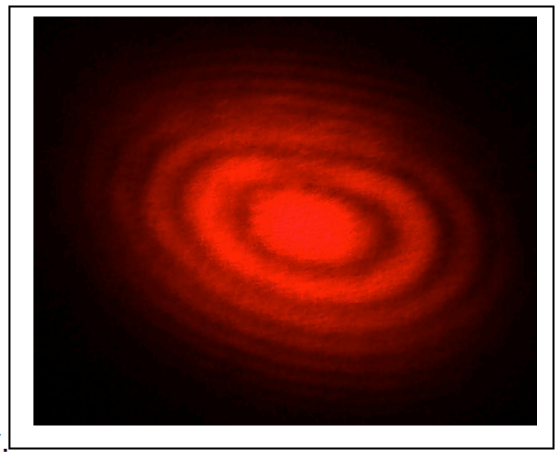
$$\vec{E}_1=(r_1,t)=\vec{E}_{01}(r_1)\cos(kr_1-\omega t+\varepsilon_1)$$
$$\vec{E}_2=(r_2,t)=\vec{E}_{02}(r_2)\cos(kr_2-\omega t+\varepsilon_2)$$

Similar to previous section, we have

$$\delta=k(r_1-r_2)+(\varepsilon_1-\varepsilon_2)$$

If the separation of the two sources is small in comparison to r_1 and r_2, and the the interference region is also small in the same sense, $\vec{E}_{01}$ and $\vec{E}_{02}$ may be considered independent of position. If the sources are of equal strength, we have

$$I=4I_0\cos^2\frac{1}{2}\left[k(r_1-r_2)+(\varepsilon_1-\varepsilon_2)\right]$$



Maximum occurs when

$$(r_1-r_2)=\frac{2\pi m+\varepsilon_2-\varepsilon_1}{k}$$

Minimum occurs when

$$(r_1-r_2)=\frac{(2m+1)\pi+\varepsilon_2-\varepsilon_1}{k}$$

where m is an integer.

If $\varepsilon_1=\varepsilon_2$, the above can be simplified as below.

Maximum occurs when

$$(r_1-r_2)=m\lambda$$

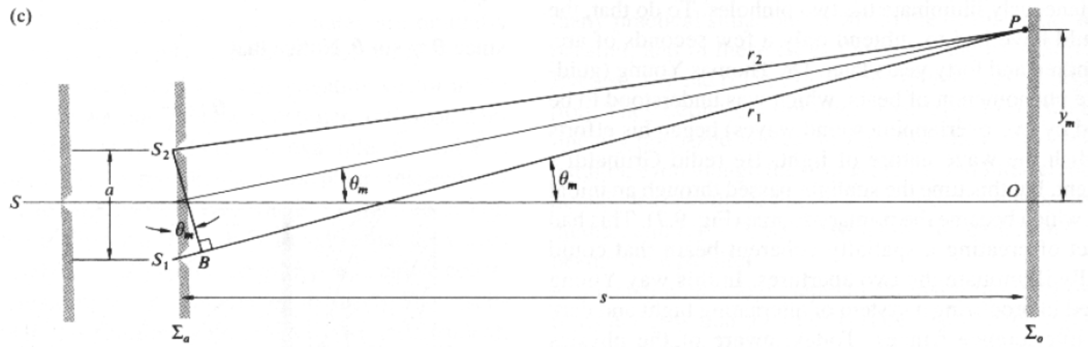
Minimum occurs when

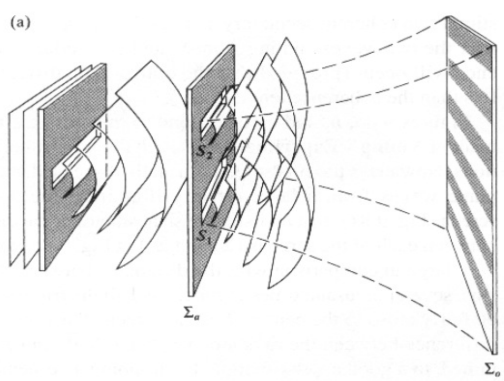
$$(r_1-r_2)=\frac{2m+1}{2}\lambda$$

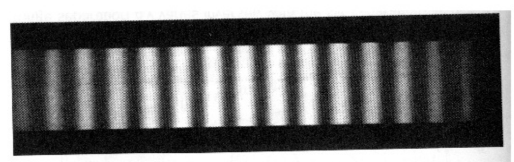
The dark and light zones that would be seen on a screen placed in the region of interference are known as **interference fringes**

Wavefront-Splitting Interferometers

The main problem introducing interference is the sources: they must be coherent.
And yet separate, independent, adequately coherent sources, other that the modern laser, don't exist. Thomas Young in his double-beam experiment took a single wavefront, split off from it two coherent portions, and had them interfere.







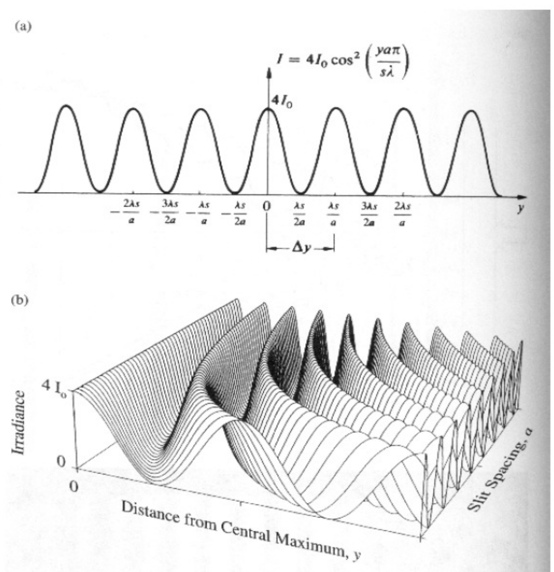


Figure 9.9 (*a*) Idealized irradiance versus distance curve. (*b*) The fringe separation Δy varies inversely with the slit separation, as one might expect from Fourier considerations; remember the inverse nature of spatial intervals and spatial frequency intervals. (Reprinted from "Graphical Representations of Fraunhofer Interference and Diffraction," *Am. J. Phys* **62**, 6 (1994), with permission of A.B. Bartlett, University of Colorado and B. Mechtly, Northeast Missouri State University and the American Association of Physics Teachers.)

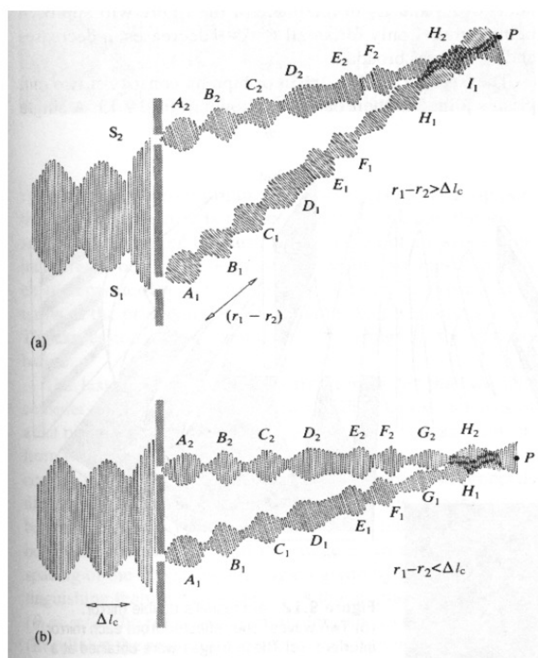


Figure 9.10 A schematic representation of how light, composed of a progression of wavegroups with a coherence length Δl_c, produces interference when (*a*) the path length difference exceeds Δl_c and (*b*) the path length difference is less than Δl_c.

The path difference

$$\overline{S_1B}=\overline{S_1P}-\overline{S_2P}=r_1-r_2\approx a\sin\theta\approx a\theta\approx a\frac{y}{s}$$

For constructive interference:

$$r_1-r_2=m\lambda \Rightarrow y_m\approx\frac{s}{a}m\lambda \text{ or } \theta_m=\frac{m\lambda}{a}.$$

The difference between consecutive maxima is

$$\Delta y=y_{m+1}-y_m\approx\frac{s}{a}\lambda$$

The intensity

$$I=4I_0\cos^2\frac{k(r_1-r_2)}{2}\approx4I_0\cos^2\frac{ya\pi}{s\lambda}$$

Note that the above discussion requires that path difference be smaller than the coherence length.

Other Interferometers

Fresnel's double mirror

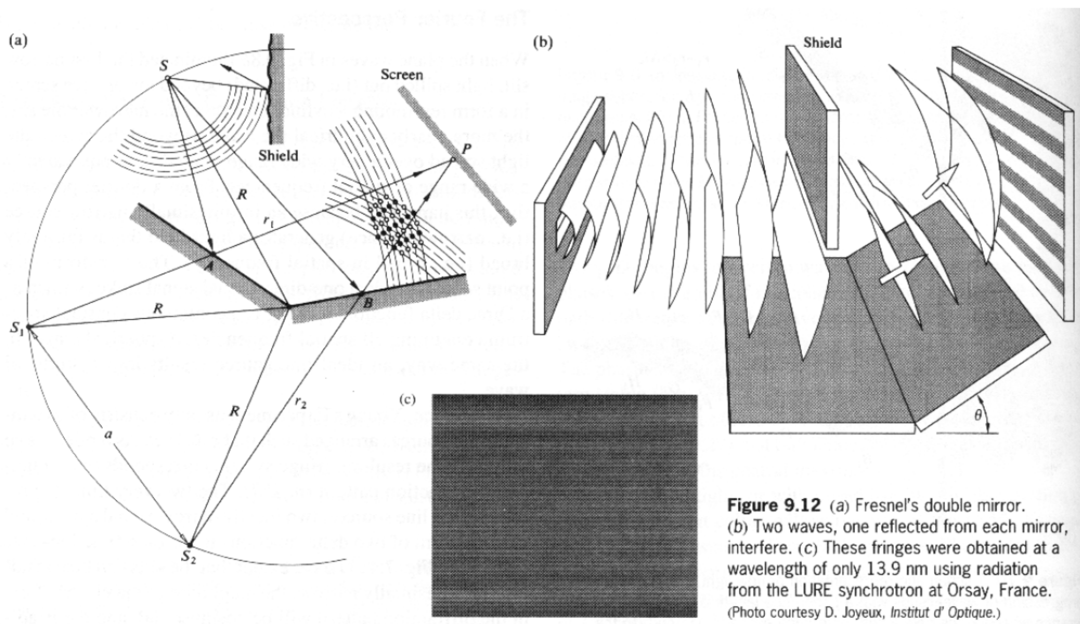
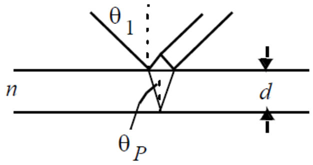


Figure 9.12 (*a*) Fresnel's double mirror. (*b*) Two waves, one reflected from each mirror, interfere. (*c*) These fringes were obtained at a wavelength of only 13.9 nm using radiation from the LURE synchrotron at Orsay, France. (Photo courtesy D. Joyeux, *Institut d' Optique*.)

Thomas Young performed his famous double slit experiment which seemed to prove that light was a wave. This experiment had profound implications, determining most of nineteenth century physics and resulting in several attempts to discover the ether, or the medium of light propagation. Though the experiment is most notable with light, the fact is that this sort of experiment can be performed with any type of wave, such as water.

In the early 1800's (1801 to 1805, depending on the source), Thomas Young conducted his experiment. There was not laser or lamps, how did Young achieve that?

Amplitude-Splitting Interferometer & Anti-Reflection (AR) Coating



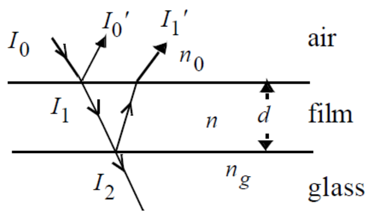
thin film, thickness d, index n

The phase difference between the reflected rays can be shown to be

$$\boxed{\delta = \pi + \frac{4\pi n}{\lambda} d \cos\theta_P}$$

For $\boxed{\delta = 2m\pi}$, we get a bright fringe; for $\boxed{\delta = (2m+1)\pi}$, we get a dark fringe.

Variations in d, λ, n, or θ give rise to fringes.



The Fresnel reflection coefficient at the top surface is

$$R_o = \left(\frac{n - n_0}{n + n_0}\right)^2 \qquad I_o' = R_o I_o$$

where the typical value for R_o is ~ 4%.

At the bottom surface:

$$R_g = \left(\frac{n_g - n}{n_g + n}\right)^2 \qquad I_o' \cong I_o R_g$$

I_1' and I_o' interfere *destructively* if

$$\boxed{\delta = \frac{4\pi n d}{\lambda} = (2m+1)\pi \qquad m = 0, 1, 2, \ldots}$$

or $nd = (2m+1)\frac{\lambda}{4} = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, \ldots$

"quarter wave"

The *net* reflected intensity is zero if I_1' and I_o' are equal, but out of phase.

So, $$\frac{n - n_o}{n + n_o} = \frac{n_g - n}{n_g + n}$$

$$(n - n_o)(n_g + n) = (n_g - n)(n + n_o)$$

$$n^2 - n_o n_g - n_o n + n n_g = n_g n - n^2 - n n_o + n_o n_g$$

$$2n^2 = 2n_o n_g$$

$$\boxed{n = \sqrt{n_o n_g}}$$

Dielectric Films--Double-Beam interference

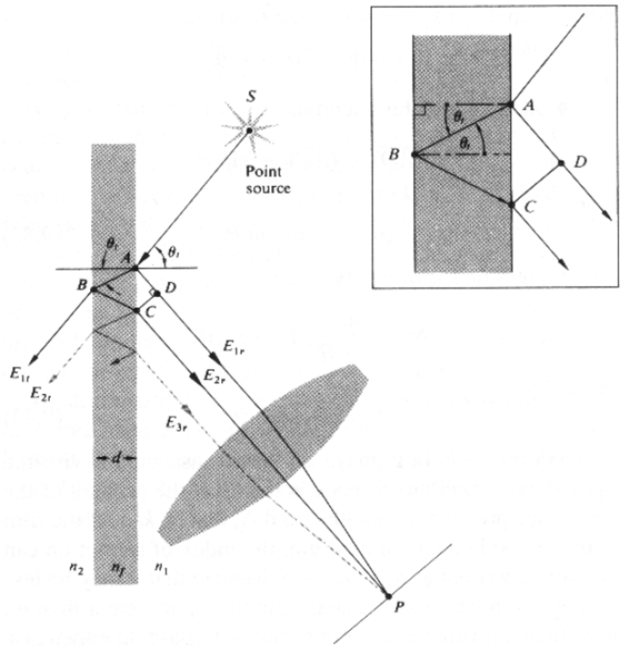


Figure 9.17 Fringes of equal inclination.

Consider only the first and the second reflected beams.
The optical path length difference is

$$\Lambda = n_f[\overline{AB}+\overline{BC}] - n_1\overline{AD}$$

Since $\overline{AB}=\overline{BC}=\dfrac{d}{\cos\theta_t}$,

$$\Lambda = \frac{2n_f d}{\cos\theta_t} - n_1\overline{AD}.$$

Also

$$\overline{AD}=\overline{AC}\sin\theta_i.$$

By Snell's Law,

$$\overline{AD}=\overline{AC}\frac{n_f}{n_1}\sin\theta_t = 2d\tan\theta_t\frac{n_f}{n_1}\sin\theta_t.$$

Therefore,

$$\Lambda = \frac{2n_f d}{\cos\theta_t}(1-\sin^2\theta_t) = 2n_f d\cos\theta_t$$

If the dielectric is immersed in the same media, the two reflection coefficients will have a phase difference of π. Considering this and the optical path length difference, the total phase difference will be

$$\delta = k_0\Lambda \pm \pi = \frac{4\pi n_f}{\lambda_0} d\cos\theta_t \pm \pi = \frac{4\pi n_f}{\lambda_0} d\sqrt{n_f^2 - n^2\sin^2\theta_i} \pm \pi$$

Maximum occurs at

$$d\cos\theta_t = (2m+1)\frac{\lambda_f}{4}$$

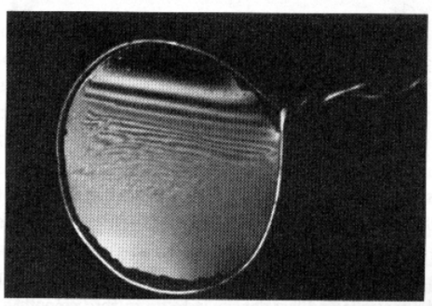
Minimum occurs at

$$d\cos\theta_t = 2m\frac{\lambda_f}{4}$$

Haidinger Fringes.

Fringes of Equal Thickness

A whole class of interference fringes exists for which the optical thickness, $n_f d$, is the dominant parameter rather than θ_i. There are referred to as fringes of equal thickness. Examples, soap bubbles, oil slicks, and oxidized metal surfaces. Each fringe is the locus of all points n the film for which the optical thickness is a constant.



A wedge-shaped film made of liquid dishwashing soap. (Photo by E. H.)

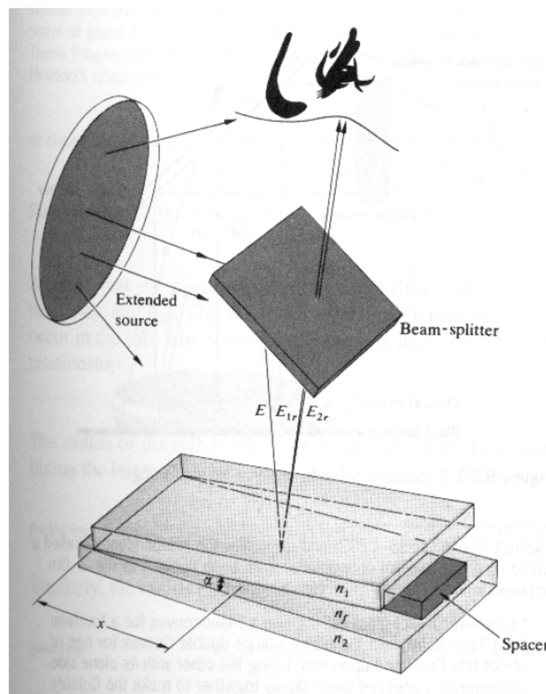
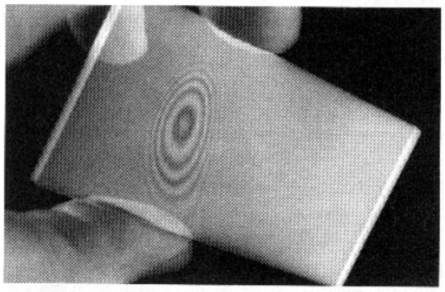


Figure 9.22 Fringes from a wedge-shaped film.



Newton's rings with two microscope slides. The thin film of air between the slides creates the interference pattern. (Photo by E. H.)

Newton's Rings

From the figure, if $R \gg d$, then

$$x^2 R^2-(R-d)^2 \Rightarrow x^2 \approx 2Rd$$

The interference maximum will occur if

$$2n_f d_m=(m+\frac{1}{2})\lambda_0$$

Thus, the radius of the bring rings are

$$x_m=\sqrt{(m+\frac{1}{2})\lambda_f R}$$

Similarly, the radius of dark rings are

$$x_m=\sqrt{m\lambda_f R}$$

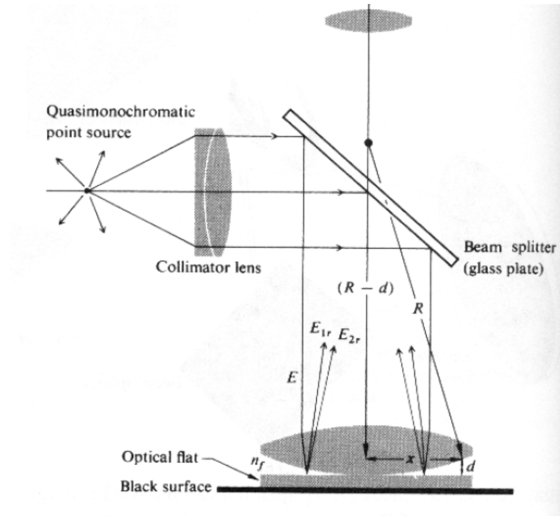
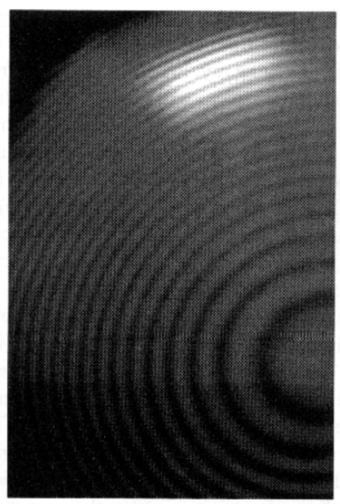


Figure 9.23 A standard setup to observe Newton's rings



Interference from the thin air film between a convex lens and the flat sheet of glass it rests on. The illumination was quasimonochromatic. These fringes were first studied in depth by Newton and are known as Newton's rings. (Photo by E.H.)

Applications of Single and Multilayer Films

Mathematical Treatment

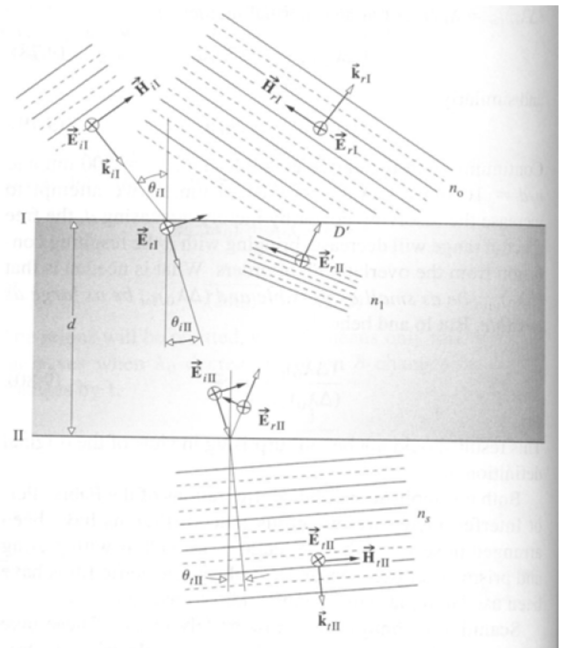


Figure 9.49 Fields at the boundaries.

At boundary I,

$$E_I = E_{iI} + E_{rI} = E_{tI} + E'_{rII}$$

$$H_I = \sqrt{\frac{\epsilon_0}{\mu_0}}(E_{iI} - E_{rI})n_0\cos\theta_{iI} = \sqrt{\frac{\epsilon_0}{\mu_0}}(E_{tI} - E'_{rII})n_1\cos\theta_{iII}$$

At boundary II,

$$E_{II} = E_{iII} + E_{rII} = E_{tII}$$

$$H_{II} = \sqrt{\frac{\epsilon_0}{\mu_0}}(E_{iII} - E_{rII})n_1\cos\theta_{iII} = \sqrt{\frac{\epsilon_0}{\mu_0}}E_{tII}n_1\cos\theta_{tII}$$

Note that

$$E_{iII} = E_{tI}e^{-k_0 h}$$

$$E_{rII} = E'_{rII}e^{+k_0 h}$$

where $h = n_1 d\cos\theta_{iII}$

Thus,

$$E_{II} = E_{tI}e^{jk_0 h} + E'_{rII}e^{+k_0 h}$$

and

$$H_{II} = (E_{tI}e^{-jk_0 h} - E'_{rII}e^{jk_0 h})\sqrt{\frac{\epsilon_o}{\mu_0}}n_1\cos\theta_{iII}$$

Solving for the relationship between (E_I, H_I) and (E_{II}, H_{II}), we have

$$\begin{bmatrix} E_I \\ H_I \end{bmatrix} = \begin{bmatrix} \cos k_0 h & (j\sin k_0 h)/\Upsilon_1 \\ \Upsilon_1 j\sin k_0 h & \cos k_0 h \end{bmatrix}\begin{bmatrix} E_{II} \\ H_{II} \end{bmatrix} = M_I\begin{bmatrix} E_{II} \\ H_{II} \end{bmatrix}$$

where

$$\Upsilon_1 = \sqrt{\frac{\epsilon_0}{\mu_0}}n_1/\cos\theta_{iII}$$

M_I is called the characteristic matrix which relates the fields at two adjacent boundaries.

Antireflection Coatings

Consider normal incidence, that is,

$$\theta_{iI}=\theta_{iII}=\theta_{tII}=0$$

then,

$$r_1=\frac{n_1(n_0-n_s)\cos k_0h+j(n_0n_s-n_1^2)\sin k_0h}{n_1(n_0+n_s)\cos k_0h+j(n_0n_s+n_1^2)\sin k_0h}$$

The reflectance

$$R_1=|r_1|^2=\frac{n_1^2(n_0-n_s)^2\cos^2k_0h+(n_0n_s-n_1^2)^2\sin^2k_0h}{n_1^2(n_0+n_s)^2\cos^2k_0h+(n_0n_s+n_1^2)^2\sin^2k_0h}$$

If $k_0h=\frac{2n+1}{2}\pi$ or equivalently $d=\frac{2n+1}{4}\lambda_f$, then

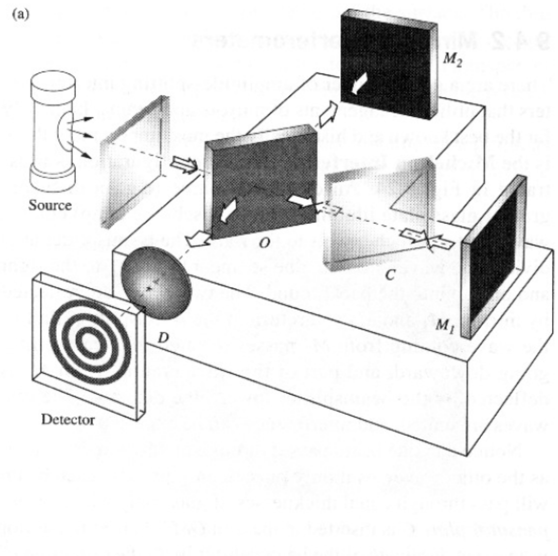
$$R_1=\frac{(n_0n_s-n_1^2)^2}{(n_0n_s+n_1^2)^2}$$

Therefore,

$$n_1^2=n_0n_s \Rightarrow R_1=0$$

Usually d is so chosen to be $\frac{\lambda_f}{4}$ in the yellow-green portion of the visible spectrum, where the eye is most sensitive. $MgF_2(n=1.38)$ is frequently used due to its durability. On a glass substrate $(n\approx 1.5)$, even the refraction index of MgF_2 doesn't satisfy the zero reflectance equation, however, it reduce the reflectance of glass from 4% to about 1% over the visible spectrum.

Michelson Interferometer



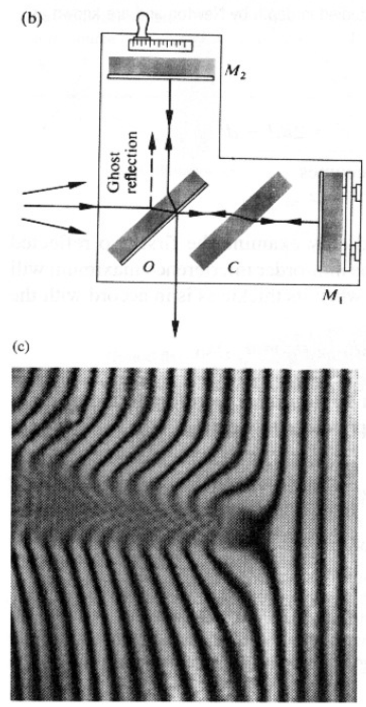


Figure 9.24 The Michelson Interferometer. (*a*) Circular fringes are centered on the lens. (*b*) Top view of the interferometer showing the path of the light. (*c*) A wedge fringe pattern was distorted when the tip of a hot soldering iron was placed in one arm. Observe the interesting perceptual phenomenon whereby the region corresponding to the iron's tip appears faintly yellow. (Photo by E. H.)

From the figure below, the optical patch difference is $2d\cos\theta$. Due the difference in reflections, the two waves have an extra phase difference π. Thus destructive interference will exist when

$$2d\cos\theta_m = m\lambda_0$$

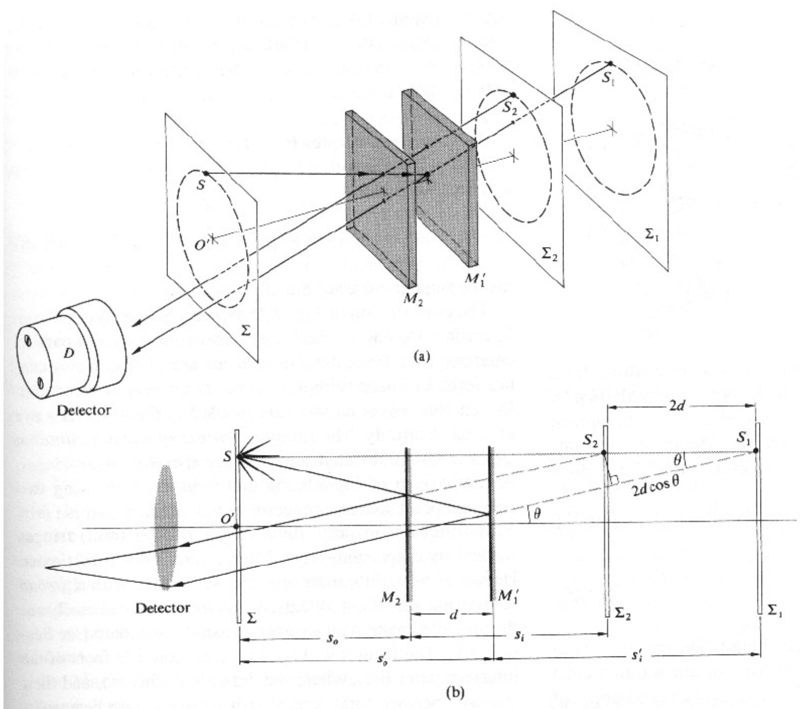


Figure 9.25 A conceptual rearrangement of the Michelson Interferometer.

As M_2 is moved toward $M_1^{\prime}, d$ decreases, $\cos\theta_m$ must increases to satisfy the equation. Thus, θ_m decreases. The rings shrink toward the center, with the highest-order one disappearing whenever d decreases by $\lambda_0/2$ Each remaining ring broadens as more and more fringes vanish at the center. By the time $d=0$, the cntral fringe will have spread out, filling the entire field of view. Moving M_2 further causes the fringes to reappear.

Notice that the central dark fringe for which $\theta_m=0$ can be represented by

$$2d=m_0\lambda_0.$$

Let the p-th ring be

$$2d\cos\theta_p=(m_0-p)\lambda_0,$$

then,

$$2d(1-\cos\theta_p)=p\lambda_0$$

If θ_p is small, $\cos\theta_p=\sqrt{1-\sin\theta_p}\approx 1-\frac{\theta_p^2}{2}$.

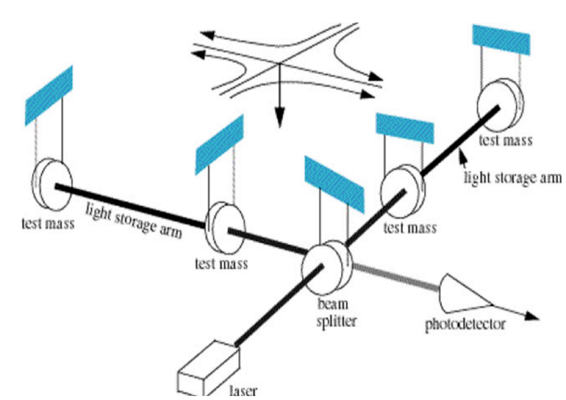
Therefore, the angular radius of the p-th fringe is

$$\theta_p\approx\sqrt{\frac{p\lambda_0}{d}}$$

The Michelson Interferometer can be used to make extremely accurate length measurements. . As the moveable mirror is displaced by $\frac{\lambda_0}{2}$, each fringe will move to the position previously occupied by an adjacent fringe. One only need to count the number of fringes N pass a reference point to determine the distance traveled by the mirror Δd, that is,

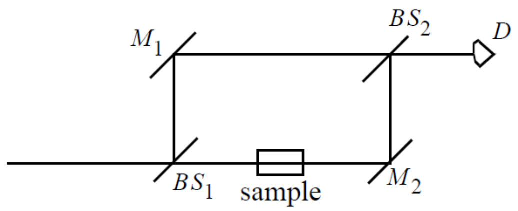
$$\Delta d=N\frac{\lambda_0}{2}$$

LIGO: http://www.ligo.caltech.edu/



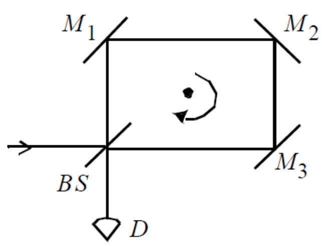
Other Mirrored Interferometers

Mach-Zender



This interferometer can be used for measuring material properties. If the index of refraction of the sample varies, then the phase difference varies and the intensity at D varies. As an example, one can determine the temperature dependence of the index of refraction n for air or other gases.

Sagnac interferometer (modified Mach-Zender)

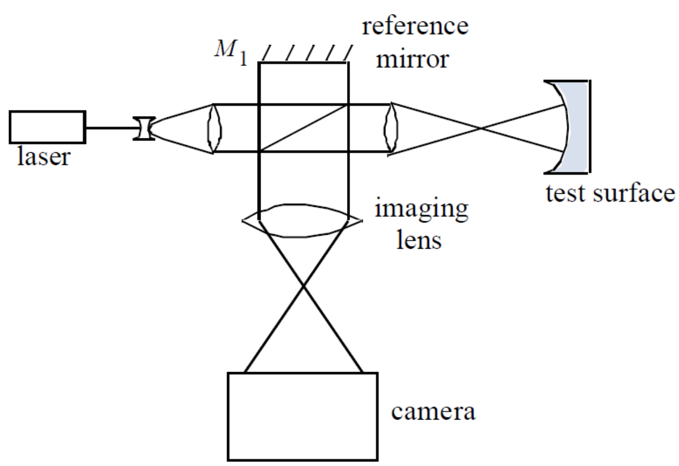


If the interferometer is rotating clockwise, the clockwise light has a longer time-of-flight than the opposite direction.

\# of fringes shift $N = \dfrac{4A\Omega}{c\lambda}$ A: area
Ω: rot vel

By using a spool of fiber instead of discrete mirrors, a very stable arrangement can be made and sensitivity is increased by n, the number of turns of fiber on the spool. This is called the "fiber-ring gyro," very popular in inertial navigation.

Twyman-Green interferometer



Basic Coherence Theory [Pedrotti^3 Ch. 12]

Let the disturbances at two points in space S_1 and S_2 are $E_1(t)$ and $E_2(t)$. At point P, the total electric field is

$$E_P(t)=k_1E_1(t-t_1)+K_2E_2(t-t_2)$$

where $t_1=\frac{r_1}{c}$ and $t_2=\frac{r_2}{c}$. The quantities K_1 and K_2 which are known as propagators, depend on the size of the apertures and their relative locations with respect to P.

Then, without considering the constant coefficients, the irradiance at P is

$$I=\left\langle E_P(t)E_P^*(t)\right\rangle_T=K_1K_1^*\left\langle E_1(t-t_1)E_1^*(t-t_1)\right\rangle_T+K_2K_2^*\left\langle E_2(t-t_2)E_2^*(t-t_2)\right\rangle_T$$
$$K_1K_2^*\left\langle E_1(t-t_1)E_2^*(t-t_2)\right\rangle_T+K_{21}K_1^*\left\langle E_2(t-t_2)E_1^*(t-t_1)\right\rangle_T$$

Assume stationary,

$$\left\langle E_1(t-t_1)E_1^*(t-t_1)\right\rangle_T=\left\langle E_1(t)E_1^*(t)\right\rangle_T=I_{S_1}$$
$$\left\langle E_2(t-t_2)E_2^*(t-t_2)\right\rangle_T=\left\langle E_2(t)E_2^*(t)\right\rangle_T=I_{S_2}$$
$$\left\langle E_1(t-t_1)E_2^*(t-t_2)\right\rangle_T=\left\langle E_1(t+\tau)E_2^*(t)\right\rangle_T$$
$$\left\langle E_2(t-t_2)E_1^*(t-t_1)\right\rangle_T=\left\langle E_1(t+\tau)^*E_2(t)\right\rangle_T$$

where $t_2-t_1=\tau$

The last two terms in the irradiance equation becomes

$$2\Re\left[K_1K_2^*\left\langle E_1(t+\tau)E_2^*(t)\right\rangle_T\right]$$

Let

$$\Gamma_{12}\equiv\left\langle E_1(t+\tau)E_2^*(t)\right\rangle_T \text{ (Mutual coherence function)}$$

Then

$$I=|K_1|^2I_{S_1}+|K_2|^2I_{S_2}+2\Re\left[K_1K_2^*\Gamma_{12}(\tau)\right]$$

Again, ignoring multiplicative constants,

$$I=I_1+I_2+2\sqrt{I_1I_2}\Re\left[e^{j(\angle K_1+\angle K_2)}\gamma_{12}(\tau)\right]=I_1+I_2+2\sqrt{I_1I_2}|\gamma_{12}(\tau)|\Re\left[e^{j(\angle K_1+\angle K_2+\angle\gamma_{12}(\tau))}\right]$$

where

$$\gamma_{12}(\tau)\equiv\frac{\Gamma_{12}(\tau)}{\sqrt{\Gamma_{11}(0)\Gamma_{22}(0)}}=|\gamma_{12}(\tau)|e^{j\angle\gamma_{12}(\tau)} \text{ (Complex degree of coherence)}$$
$$I_1=|K_1|^2\Gamma_{11}(0)$$
$$I_2=|K_2|^2\Gamma_{22}(0)$$

$|\gamma_{12}(\tau)|$ is the degree of coherence.

1. $|\gamma_{12}(\tau)|=1$, coherent limit,
2. $|\gamma_{12}(\tau)|=0$, incoherent limit,
3. $0<|\gamma_{12}(\tau)|<1$, partial coherent.

Define Visibility as

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$$

then

$$V = \frac{2\sqrt{I_1}\sqrt{I_2}}{I_1 + I_2} |\gamma_{12}(\tau)|$$

If $I_1 = I_2$,

$$V = |\gamma_{12}(\tau)|$$

Thus measuring visibility can determine the degree of coherence.

Coherence of Light

As shown on previous pages, most available light sources emit a range wavelengths. **For many photonic applications it is desirable to have a "monochromatic" (ideal) point light source.** An ideal light source would emit only one exact frequency f_0 and its physical size would be infinitely small (an ideal point source). Such a source cannot exist in reality. Every real source of light has some emission uncertainty which appears as linewidth Δf. Also, every real light source has some (non-zero) physical size.

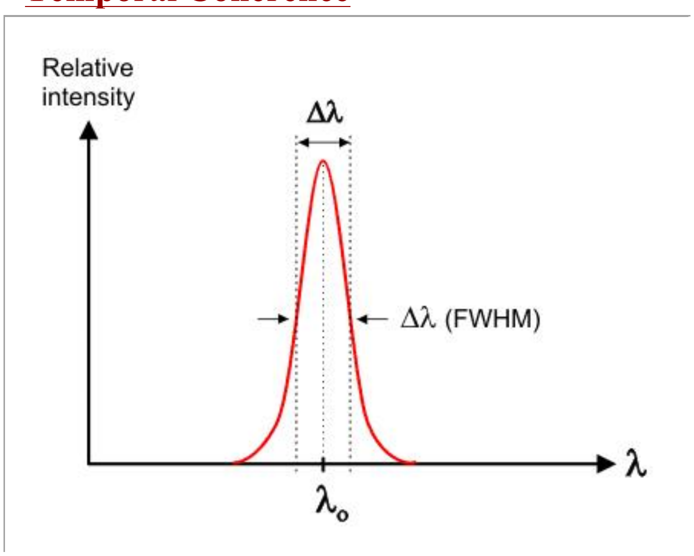
Coherence is a concept that that establishes the limits within which a real light source can be considered ideal.

There are two types of coherence:

(1) Temporal Coherence (related to the emitted linewidth)

(2) Spatial Coherence (related to the physical size of the source).

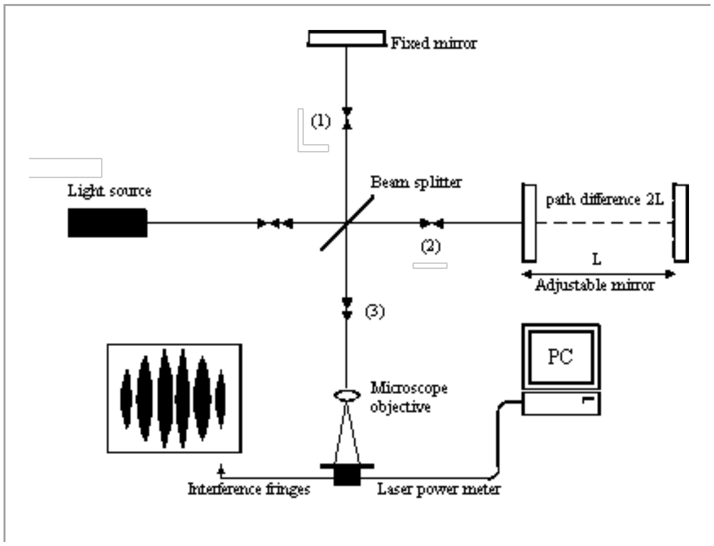
Temporal Coherence

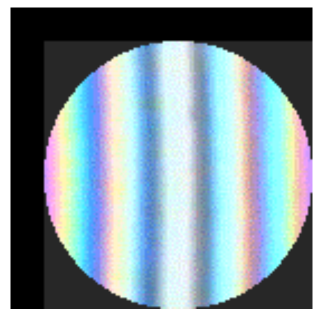


Most quasi-monochromatic sources of light have spectral intensity profile that can be approximated by a Gaussian curve. Temporal coherence is determined by the coherence length $\mathbf{L_c}$ which depends on the linewidth $\mathbf{\Delta\lambda}$ (wavelength uncertainty within FWHM of relative irradiance) of central wavelength $\mathbf{\lambda_0}$. It can also be expressed in terms of frequency bandwidth (uncertainty) $\mathbf{\Delta f}$ of the central optical frequency $\mathbf{f_0}$.
(Note that when $\Delta f \ll f_0$ and $\Delta\lambda \ll \lambda_0$, then $\Delta f/f_0 = \Delta\lambda / \lambda_0$).
In free space, coherence length $\mathbf{L_c = c/\Delta f = \lambda_0^2 / \Delta\lambda}$.
Coherence length $\mathbf{L_c}$ is the distance in the direction of wavefront propagation within which the

amplitude and phase of the wave can be considered well defined, predictable, and therefore subject to possible wave interference. An ideal light source would have an infinite coherence length. On the other hand, thermal sources of light (such as the sun or a light bulb) cover a relatively broad range of wavelengths $\Delta\lambda$ (see Planck's radiant function) and therefore have very small coherence length. For that reason, under normal lighting conditions, we can observe interference only in very thin regions, such as the thickness of a soap bubble.

Note that in a dielectric medium with refractive index **n,** the speed of light, wavelength, and therefore also the coherence length, is decreased by the factor **n**
thus $\mathbf{L_c = c/(n\Delta f) = \lambda^2 / (n\Delta\lambda)}$.





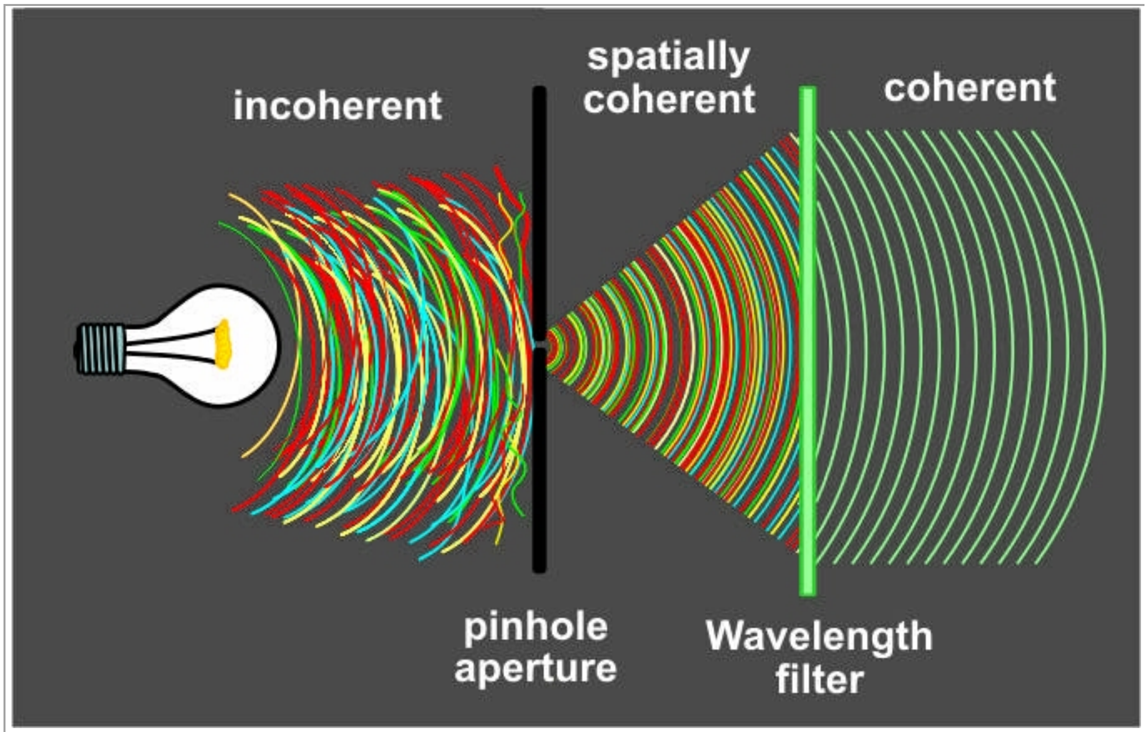
Michelson interferometer is an instrument in which a collimated beam of light is split into two beams travelling separate paths (1) and (2), and then reassembled as (3). If the path difference $\Delta L = 2L = \text{path}(2) - \text{path}(1)$ is smaller than the coherence length L_c, interference can be observed. If $\Delta L > L_c$ the interference pattern disappears (the phase relation between the reassembled beams no longer exists). We can thus determine the coherence length $L_c = \lambda_0^2/\Delta\lambda$ by increasing the distance (2) and finding the path difference $\Delta L = 2L$ at which the interference pattern disappears. From this measurement we can determine the linewidth $\Delta\lambda$ and the quality of the quasi-monochromatic light source λ_0.

Spatial Coherence

Spatial coherence is determined by the coherence width
$\mathbf{W_c = k\,(\lambda/D)\,R}$**,** where **k** is a constant dependent on the shape of the source (for a circular source k = 1.22), **D** is the approximate diameter of the source and **R** is the distance from the source. Coherence width is the distance along the wavefront (perpendicular to the direction of propagation), within which the amplitude and phase of the wave can be considered well defined and therefore predictable. An ideal light source would be a point (D = 0) generating ideal spherical wavefronts. An ideal point source would have infinite coherence width.

The degree of spatial coherence can be estimated by inspection of a shadow cast by an illuminated object. The sharper the shadow, the better spatial coherence of the source.

Note: Coherence of a source can be improved by various physical arrangements and optical components (increasing the distance from the source, focusing and passing light through a small pinhole aperture, etc.). Improved coherence, however, results in drastic reduction of light intensity. When the concept of coherence is mentioned in literature, it usually refers to temporal coherence.



Idealized improvement of coherence from an incoherent light source

Temporal coherence (From Wikipedia)

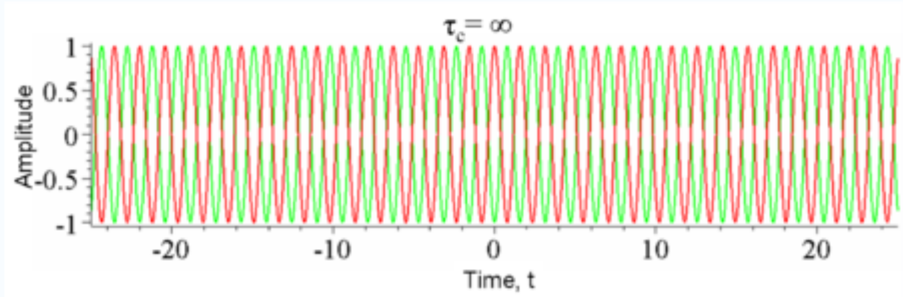


Figure 1: The amplitude of a single frequency wave as a function of time *t* (red) and a copy of the same wave delayed by τ(green). The coherence time of the wave is infinite since it is perfectly correlated with itself for all delays τ.

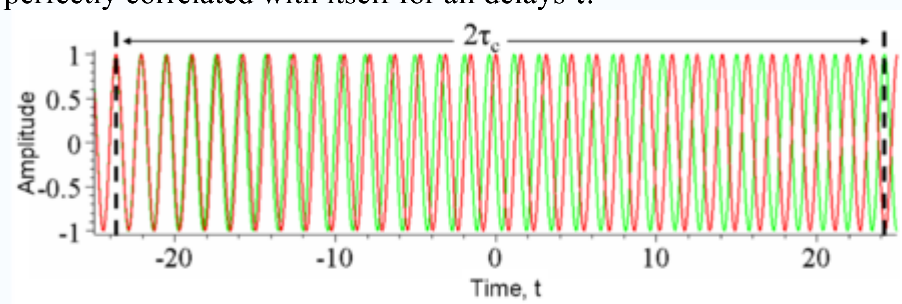


Figure 2: The amplitude of a wave whose phase drifts significantly in time τ_c as a function of time *t* (red) and a copy of the same wave delayed by $2\tau_c$(green). At any particular time t the wave can interfere perfectly with its delayed copy. But, since half the time the red and green waves are in phase and half the time out of phase, when averaged over t any interference disappears at this delay.

Temporal coherence is the measure of the average correlation between the value of a wave at any pair of times, separated by delay τ. Temporal coherence tells us how monochromatic a source is. In other words, it characterizes how well a wave can interfere with itself at a different time. The delay over which the phase or amplitude wanders by a significant amount (and hence the correlation decreases by significant amount) is defined as the coherence time τ_c. At τ=0 the degree of coherence is perfect whereas it drops significantly by delay τ_c. The coherence length L_c is defined as the distance the wave travels in time τ_c.

One should be careful not to confuse the coherence time with the time duration of the signal, nor the coherence length with the coherence area (see below).

The relationship between coherence time and bandwidth

It can be shown that the faster a wave decorrelates (and hence the smaller τ_c is) the larger the range of frequencies Δf the wave contains. Thus there is a tradeoff:

$$\tau_c \Delta f \approx 1.$$

In terms of wavelength ($f\lambda = c$) this relationship becomes,

$$\frac{L_c}{\Delta\lambda} \approx 1$$

Formally, this follows from the convolution theorem in mathematics, which relates the Fourier transform of the power spectrum (the intensity of each frequency) to its autocorrelation.

Examples of temporal coherence

We consider four examples of temporal coherence.

- A wave containing only a single frequency (monochromatic) is perfectly correlated at all times according to the above relation. (See Figure 1)
- Conversely, a wave whose phase drifts quickly will have a short coherence time. (See Figure 2)
- Similarly, pulses (wave packets) of waves, which naturally have a broad range of frequencies, also have a short coherence time since the amplitude of the wave changes quickly. (See Figure 3)
- Finally, white light, which has a very broad range of frequencies, is a wave which varies quickly in both amplitude and phase. Since it consequently has a very short coherence time (just 10 periods or so), it is often called incoherent.

The most monochromatic sources are usually lasers; such high monochromaticity implies long coherence lengths (up to hundreds of meters). For example, a stabilized helium-neon laser can produce light with coherence lengths in excess of 5 m. Not all lasers are monochromatic, however (e.g. for a mode-locked Ti-sapphire laser, $\Delta\lambda \approx 2$ nm - 70 nm). LEDs are characterized by $\Delta\lambda \approx 50$ nm, and tungsten filament lights exhibit $\Delta\lambda \approx 600$ nm, so these sources have shorter coherence times than the most monochromatic lasers.

Holography requires light with a long coherence time. In contrast, Optical coherence tomography uses light with a short coherence time. (Why?)

Measurement of temporal coherence

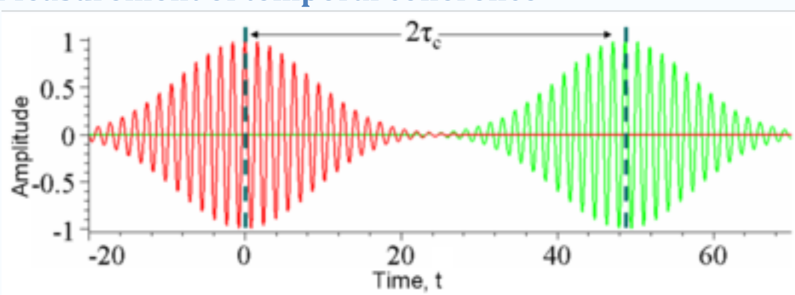


Figure 3: The amplitude of a wavepacket whose amplitude changes significantly in time τ_c (red) and a copy of the same wave delayed by $2\tau_c$(green) plotted as a function of time *t*. At any particular time the red and green waves are uncorrelated; one oscillates while the other is constant and so there will be no interference at this delay. Another way of looking at this is the wavepackets are not overlapped in time and so at any particular time there is only one nonzero field so no interference can occur.

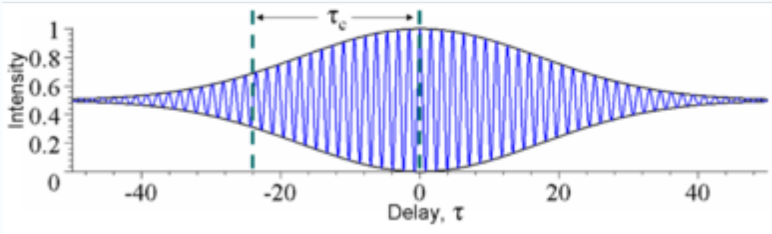


Figure 4: The time-averaged intensity (blue) detected at the output of an interferometer plotted as a function of delay τ for the example waves in Figures 2 and 3. As the delay is changed by half a period, the interference switches between constructive and destructive. The black lines indicate the interference envelope, which gives the degree of coherence. Although the waves in Figures 2 and 3 have different time durations, they have the same coherence time.

In optics, temporal coherence is measured in an interferometer such as the Michelson interferometer or Mach-Zehnder interferometer. In these devices, a wave is combined with a copy of itself that is delayed by time τ. A detector measures the time-averaged intensity of the light exiting the interferometer. The resulting interference visibility (e.g. see Figure 4) gives the temporal coherence at delay τ. Since for most natural light sources, the coherence time is much shorter than the time resolution of any detector, the detector itself does the time averaging. Consider the example shown in Figure 3. At a fixed delay, here $2\tau_c$, an infinitely fast detector would measure an intensity that fluctuates significantly over a time *t* equal to τ_c. In this case, to find the temporal coherence at $2\tau_c$, one would manually time-average the intensity.

Spatial coherence

In some systems, such as water waves or optics, wave-like states can extend over one or two dimensions. Spatial coherence describes the ability for two points in space, x_1 and x_2, in the extent of a wave to interfere, when averaged over time. More precisely, the spatial coherence is the cross-correlation between two points in a wave for all times. If a wave has only 1 value of amplitude over an infinite length, it is perfectly spatially coherent. The range of separation between the two points over which there is significant interference is called the coherence area, A_c. This is the relevant type of coherence for the Young's double-slit interferometer. It is also used in optical imaging systems and particularly in various types of astronomy telescopes. Sometimes people also use "spatial coherence" to refer to the visibility when a wave-like state is combined with a spatially shifted copy of itself.

Examples of spatial coherence

Spatial coherence

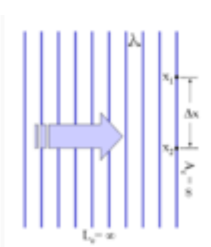


Figure 5: A plane

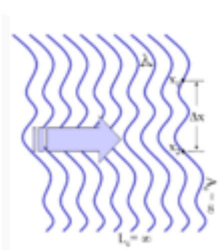


Figure 6: A wave

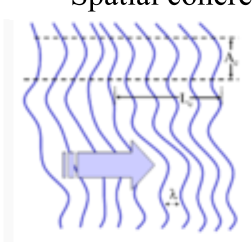


Figure 7: A wave with

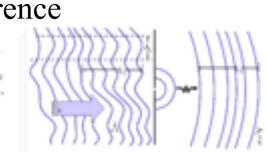


Figure 8: A wave with finite coherence area is incident on a pinhole (small aperture). The wave will diffract out of

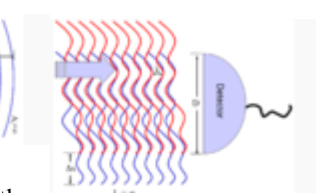


Figure 9: A wave with infinite coherence area is combined with a

wave with an infinite coherence length.	with a varying profile (wavefront) and infinite coherence length.	a varying profile (wavefront) and finite coherence length.	the pinhole. Far from the pinhole the emerging spherical wavefronts are approximately flat. The coherence area is now infinite while the coherence length is unchanged.	spatially-shifted copy of itself. Some sections in the wave interfere constructively and some will interfere destructively. Averaging over these sections, a detector with length D will measure reduced interference visibility. For example a misaligned Mach-Zehnder interferometer will do this.

Consider a tungsten light-bulb filament. Different points in the filament emit light independently and have no fixed phase-relationship. In detail, at any point in time the profile of the emitted light is going to be distorted. The profile will change randomly over the coherence time τ_c. Since for a white-light source such as a light-bulb τ_c is small, the filament is considered a spatially incoherent source. In contrast, a radio antenna array, has large spatial coherence because antennas at opposite ends of the array emit with a fixed phase-relationship. Light waves produced by a laser often have high temporal and spatial coherence (though the degree of coherence depends strongly on the exact properties of the laser). Spatial coherence of laser beams also manifests itself as speckle patterns and diffraction fringes seen at the edges of shadow.

Holography requires temporally and spatially coherent light. Its inventor, Dennis Gabor, produced successful holograms more than ten years before lasers were invented. To produce coherent light he passed the monochromatic light from an emission line of a mercury-vapor lamp through a pinhole spatial filter.

Can you explain why Gabor's method indeed generated a temporally and spatially coherent light?