Oscillations, Waves and Optics

Part IB Physics A 2021/2022

Lecturer: Tijmen Euser te287@cam.ac.uk...

October 2021

Part IB Physics A: Oscillations, Waves and Optics

Welcome!
This course will build on the foundation of the 1A course, but take the material considerably further and allow us to understand many more physical phenomena.
We will cover

- Oscillations: driven and free
- Waves: the new key ideas are polarisation, group velocity, dispersion, reflection and transmission
- Applications of Fourier ideas to waves and oscillations
- Optics: mathematical theory of diffraction — in the near and far fields
- Interference of light

Damped Harmonic Oscillator

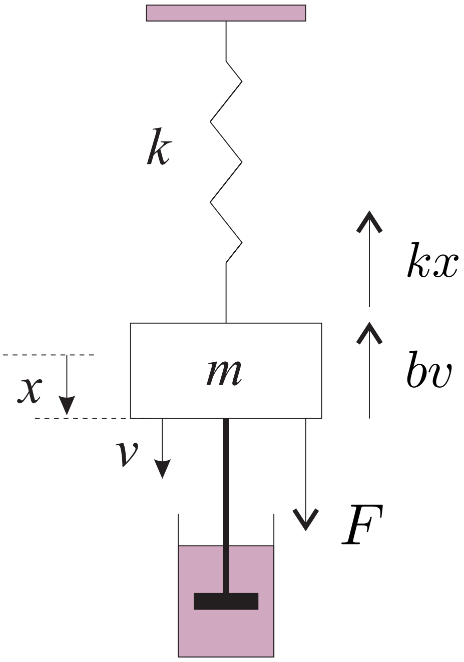


Figure 1: Simple oscillator

Consider a simple system with a restoring force proportional to displacement ($F \propto -x$). This harmonic oscillator has a quadratic potential ($V \propto x^2$).
Assume that there is a resistive drag force with a magnitude proportional to the speed of the oscillation.
Assume we apply some driving force $F(t)$. The equation of motion for a driven, damped oscillator is then

$$F(t) = m\ddot{x} + b\dot{x} + kx \tag{1}$$

b is a constant giving the strength of the damping. k specifies the strength of the restoring force (or the curvature of the quadratic potential function near equilibrium: $V = \frac{1}{2}kx^2$).

When does this equation apply?

- Simple harmonic motion with negligible drag force is common near equilibrium because the potential will most likely be quadratic locally.
- A mass on a spring undergoing small oscillations with a viscous damping force may approximate this ideal situation well.
- But a pendulum is not a simple harmonic oscillator in general:
  - the restoring force is proportional to $(-\sin\theta)$ for displacement θ which is linear only for small displacements when $\sin\theta \approx \theta$.
  - damping may not be proportional to speed (e.g. wind resistance).
- In general, numerical integration will be needed to solve such problems. (See the MATLAB example on the example sheet).

Canonical Form of Damped SHM

We define some quantities to put the equation of motion into a standard ("canonical") form:

$$\omega_0 = \sqrt{\frac{k}{m}} \tag{2}$$

This is the frequency of oscillation in the absence of damping and forcing.

$$\gamma = \frac{b}{m} \tag{3}$$

this expresses the damping; these yield the canonical form:

$$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = \frac{F}{m} \tag{4}$$

N.B. This is consistent with Main's book, but not Pain's or perhaps the IA notes (where $2\gamma\dot{x}$ appears in the equation).
Note also that γ has dimensions of inverse time. It is also useful to define a new dimensionless parameter

$$Q = \frac{\omega_0}{\gamma} \tag{5}$$

This is the Quality Factor.

Complex Notation: Phasor Diagrams

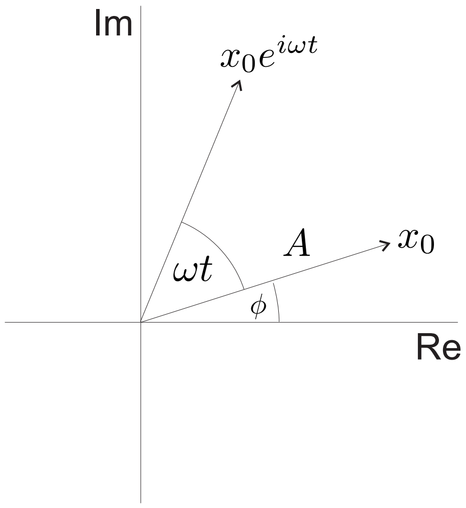


Figure 2: phasor diagram

It is often convenient (but not necessary) to represent an oscillating quantity as a vector rotating anti-clockwise in the complex plane. The observable quantity (in this case the displacement x) is by convention equal to the real part of the complex representation.

$$x = \Re\left(x_0\, e^{i\omega t}\right)$$

where the complex amplitude is $x_0 = A\,e^{i\phi}$:

$$\begin{aligned} x &= \Re\left[A\,e^{i\phi}e^{i\omega t}\right] \\ &= \Re\left[A\,e^{i(\omega t+\phi)}\right] \\ &= A\,\cos(\omega t + \phi) \end{aligned}$$

$A = |x_0|$ is the amplitude of the oscillation (a real value), and $\arg(x_0)$ is its phase.
If instead we use $\exp(-i\omega t)$, vector rotates clockwise (as in IB Electromagnetism course).

Speed and Acceleration

- Using complex representation, the speed of the particle can be derived using

$$v = \frac{\mathsf{d}}{\mathsf{d}t}\Re\left(x_0 \mathrm{e}^{\mathrm{i}\omega t}\right) = \Re\left(\mathrm{i}\omega x_0 \mathrm{e}^{\mathrm{i}\omega t}\right) = \Re\left(v_0 \mathrm{e}^{\mathrm{i}\omega t}\right)$$

  where we introduce the complex velocity $v_0 = \mathrm{i}\omega x_0$.
- Note we have interchanged the order of differentiation and taking the real part: these operations commute.
- Similarly, the complex acceleration is

$$a_0 = \mathrm{i}\omega v_0 = -\omega^2 x_0$$

Speed and Acceleration: Phasor Diagram

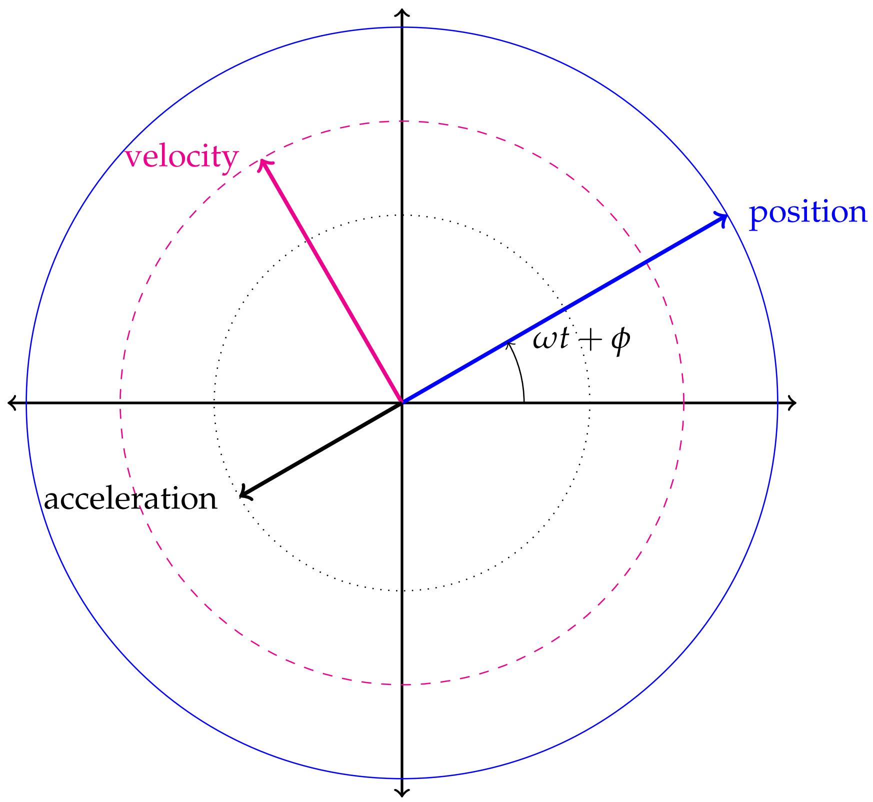


Figure 3: Phasor diagram for velocity and acceleration

Free Oscillations: Review

With no driving force, the oscillator obeys the homogeneous equation

$$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = 0 \tag{6}$$

We look for solutions of the form $x \propto \mathrm{e}^{pt}$, finding

$$p^2 + \gamma p + \omega_0^2 = 0 \tag{7}$$

$$\therefore p = \frac{-\gamma \pm \sqrt{\gamma^2 - 4\omega_0^2}}{2} = -\frac{\gamma}{2}\left(1 \pm \sqrt{1 - 4Q^2}\right)$$

where we used $Q = \omega_0/\gamma$. There are three regimes to consider, depending on the sign of the square root term:

- Light Damping: $\gamma < 2\omega_0$ $(Q > 0.5)$
- Critical Damping: has $\gamma = 2\omega_0$ $(Q = 0.5)$
- Heavy Damping: has $\gamma > 2\omega_0$ $(Q < 0.5)$

$Q = 0.5$ separates the regimes. Q is the quality factor of the oscillator. High-Q means low damping, and vice versa.

Free Oscillations: Light Damping $Q > 0.5$

- We find $p = -\frac{\gamma}{2} \pm \mathrm{i}\omega_f$, where the free oscillation frequency ω_f is

$$\omega_f = \sqrt{\omega_0^2 - \frac{\gamma^2}{4}} = \omega_0\sqrt{1 - \frac{1}{4Q^2}} \tag{8}$$

Note that ω_f is very close to ω_0 for Q values greater than a few. (For $Q = 5$, the difference is 0.5%). The solution involves 2 arbitrary constants, set by the boundary conditions. We can write equivalently:

$$x(t) = \mathrm{e}^{-\gamma t/2}\left(A\mathrm{e}^{\mathrm{i}\omega_f t} + A^*\mathrm{e}^{-\mathrm{i}\omega_f t}\right)$$
$$x(t) = \mathrm{e}^{-\gamma t/2} B\cos(\omega_f t + \phi)$$
$$x(t) = \mathrm{e}^{-\gamma t/2}\left(C_1 \cos\omega_f t + C_2 \sin\omega_f t\right)$$
$$x(t) = \Re\left(D\mathrm{e}^{-\gamma t/2}\mathrm{e}^{\mathrm{i}\omega_f t}\right)$$

where A and D are complex; and B, C_1 and C_2 are real. Simple algebra allows us to relate the A, B, $C_{1,2}$, and D.

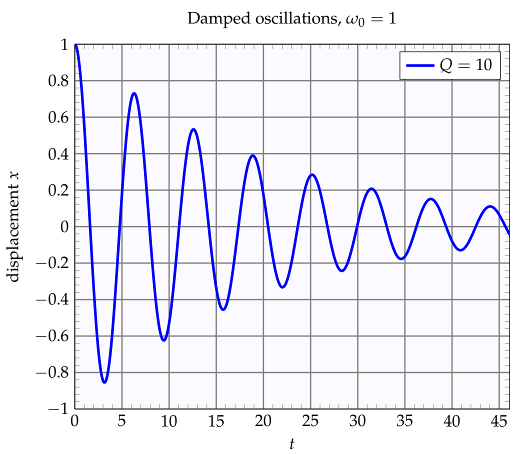


Figure 4: Lightly damped free oscillations, starting from $\dot{x} = 0, x = 1$ at $t = 0$

Light Damping: estimating Q

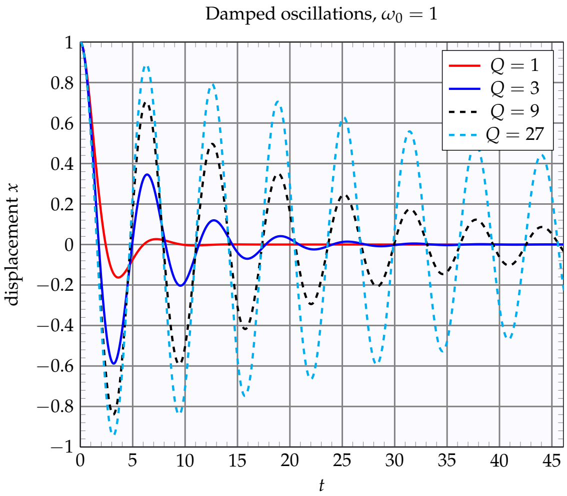


Figure 5: Lightly damped free oscillations

- Amplitude $\propto e^{-\gamma t/2} \propto e^{-\omega_0 t/(2Q)}$ i.e. decays with a time constant $\tau = 2/\gamma$.
- When $\omega_0 t = Q$, amplitude is $e^{-1/2} \approx 61\%$ of initial amplitude.
- So Q is the number of radians of phase elapsed as the amplitude falls to $e^{-0.5} \approx 61\%$ of its original value
- The energy U decays twice as fast, $U \propto e^{-\omega_0 t/Q}$.
- and the energy is then $e^{-1} \approx 37\%$ of the initial energy.

e.g. The light blue curve has amplitude ≈ 0.6 after 4 complete oscillations, so $Q \approx 4 \times 2\pi \approx 25$ (close to true value of 27).

Free Oscillations: Heavy Damping $Q < 0.5$

- In this case, $\gamma > 2\omega_0$, and p has two real values.
- No oscillations, only exponential decay occurs.

$$x(t) = C_1 e^{-\mu_1 t} + C_2 e^{-\mu_2 t}$$

$$\mu_{1,2} = \frac{1}{2}\left(\gamma \pm \sqrt{\gamma^2 - 4\omega_0^2}\right)$$

with C_1 and C_2 being real.

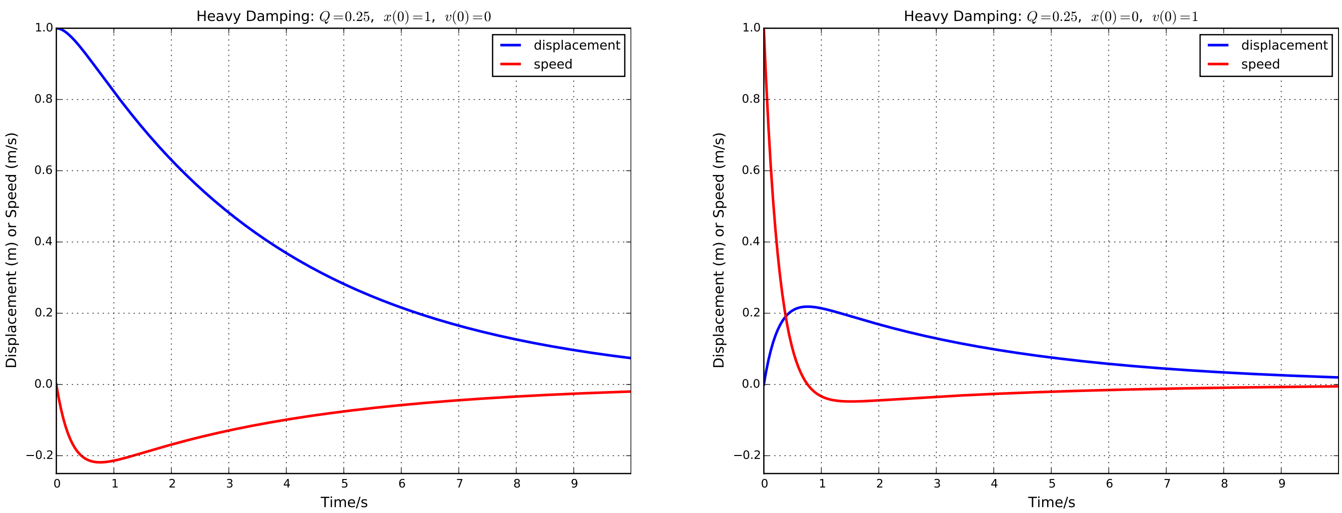


Figure 6: Transient response of a heavily damped oscillator with $m = 1\,\mathrm{kg}$, $k = 1\,\mathrm{N\,m^{-1}}$, $b = 4\,\mathrm{kg\,s^{-1}}$.

Free Oscillations: Critical Damping $Q = 0.5$

H13(270921)

- Only one solution for p, which is $p = -\gamma/2 = -\omega_0$. It is impossible to fit two boundary conditions with this single exponential solution. The general solution can be shown to be

$$\boxed{x(t) = (C_1 + C_2\omega_0 t)\,\mathrm{e}^{-\omega_0 t}}$$

- This solution yields the most rapid approach to equilibrium with no overshooting. Example application: analog speedometers.
- The system returns to equilibrium with a time constant of $1/\omega_0$ in critical damping conditions.

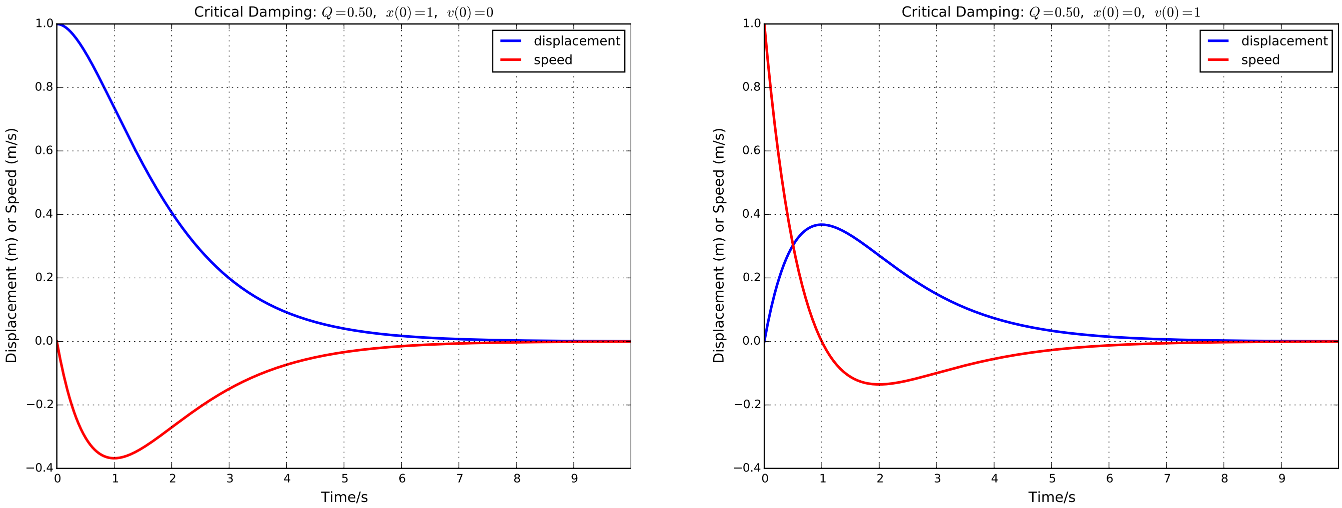


Figure 7: Transient response of a critically damped oscillator with $m = 1\,\mathrm{kg}$, $k = 1\,\mathrm{N\,m^{-1}}$, $b = 2\,\mathrm{kg\,s^{-1}}$.

Finding A Specific Solution from Initial Conditions: Example

H14(270921)

Consider a lightly-damped oscillator, with initial conditions $x(0) = 1$, $\dot{x}(0) = 0$, i.e. it is displaced by unit amplitude and released. Write general solution (with real A, ϕ):

$$x(t) = A\,\mathrm{e}^{-\gamma t/2}\cos(\omega_f t + \phi)$$

- The initial, or boundary conditions imply the following.
- Since $x(0) = 1$, we have

$$A\cos\phi = 1$$

- Since $\dot{x}(0) = 0$, we have

$$-A\omega_f\sin\phi - \frac{\gamma A}{2}\cos\phi = 0$$

- These have solution

$$\tan\phi = -\frac{\gamma}{2\omega_f}; \qquad A = \sqrt{\frac{1}{\cos^2\phi}} = \sqrt{1+\tan^2\phi} = \sqrt{1+\frac{\gamma^2}{4\omega_f^2}} = \frac{\omega_0}{\omega_f}$$

(see Eq. 8).
- Note that the initial phase is not zero, but close to zero if Q is large; and the amplitude is greater than 1.

The Driven Harmonic Oscillator

What happens when we apply a force $F(t)$ to an oscillator?

$$F(t) = m\ddot{x} + b\dot{x} + kx \tag{9}$$

- Let us initially restrict the driving force to be sinusoidal in form.
- We will see that this is not a serious restriction once we use Fourier analysis
- Write the real driving force as $F = \Re\left[F_0 \mathrm{e}^{\mathrm{i}\omega t}\right]$
- F_0 is in general complex (driving sinusoid may not be at phase 0 at $t = 0$)
- Substitute into the equation of motion:

$$\Re\left\{[-\omega^2 + \mathrm{i}\gamma\omega + \omega_0^2]x_0 \mathrm{e}^{\mathrm{i}\omega t}\right\} = \Re\left\{\frac{F_0}{m}\mathrm{e}^{\mathrm{i}\omega t}\right\}$$

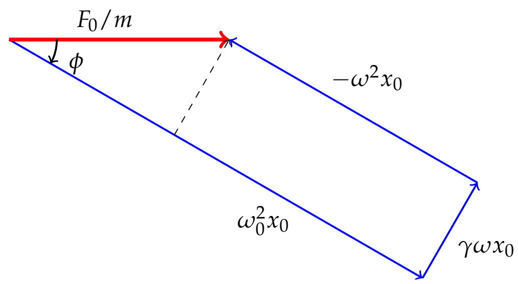


Figure 8: Phasor diagram for driven oscillator

- This implies (we can drop the $\Re$):

$$x_0 = \frac{F_0}{m[(\omega_0^2 - \omega^2) + \mathrm{i}\gamma\omega]} \tag{10}$$

Sinusoidal Driving: Response Function

- We can write the displacement as

$$x_0 = R(\omega)F_0$$

- where we define the Response Function $R(\omega)$

$$R(\omega) = \frac{x_0}{F_0} = \frac{1}{m[(\omega_0^2 - \omega^2) + \mathrm{i}\gamma\omega]} \tag{11}$$

- This can be also written as

$$R(\omega) = \frac{(\omega_0^2 - \omega^2) - \mathrm{i}\gamma\omega}{m[(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2]} \tag{12}$$

- $R(\omega)$ describes the displacement relative to the driving force:
  - $|R|$ gives the amplitude of the sinusoidal displacement per unit force
  - $\arg(R)$ gives the phase of the displacement with respect to the forcing sinusoid.

Driven Response: Amplitude

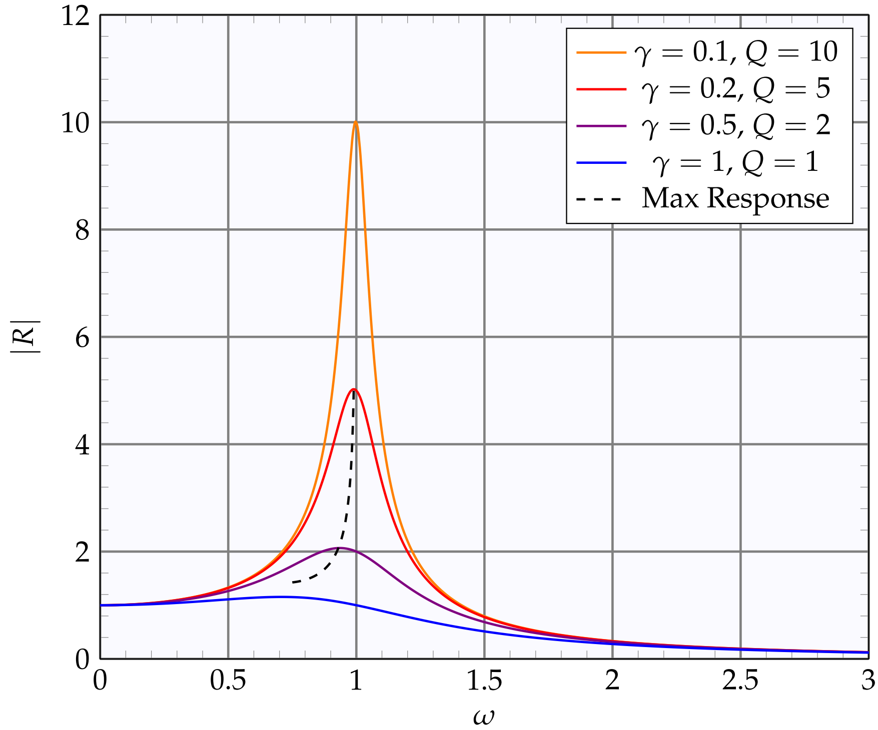


Figure 9: Driven Response –Amplitude. $m = 1\,\mathrm{kg}$, $k = 1\,\mathrm{N\,m^{-1}}$, $\omega_0 = 1\,\mathrm{rad\,s^{-1}}$

$$|R| = \frac{1}{m\sqrt{(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2}} \tag{13}$$

Driven Response: Phase

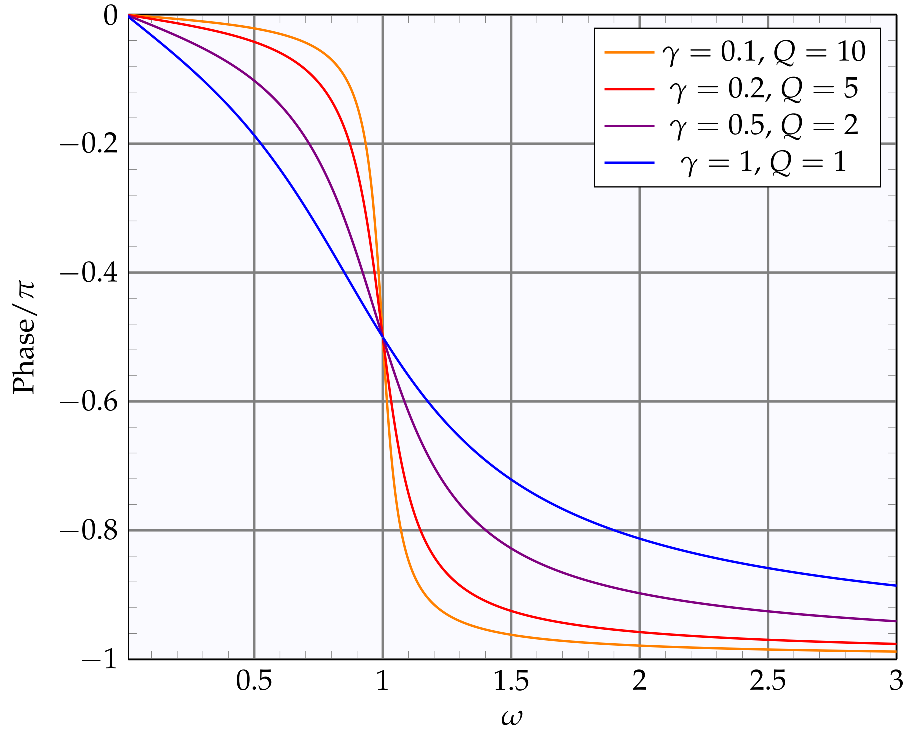


Figure 10: Driven Response – Phase

$$\arg(R) = \arctan\left[\frac{-\gamma\omega}{(\omega_0^2 - \omega^2)}\right] \tag{14}$$

Three Frequency Regimes of Driven Oscillator

- $\omega \ll \omega_0$: response in phase with the driving force. Motion controlled by spring constant/stiffness.

$$|R| \approx 1/(m\omega_0^2) = 1/k, \quad \arg R \approx 0$$

- $\omega \gg \omega_0$: response in anti-phase with the driving. Motion controlled by the inertia of the system.

$$|R| \approx 1/(m\omega^2), \quad \arg R \approx -\pi$$

- $\omega \sim \omega_0$: resonance, i.e. max value of $|R|$. Differentiate $|R|$ (Eq. 13) to show this occurs at angular frequency ω_a:

$$\omega_a = \omega_0\sqrt{1 - \frac{\gamma^2}{2\omega_0^2}} = \omega_0\sqrt{1 - \frac{1}{2Q^2}} \tag{15}$$

i.e. ω_a is smaller than ω_0 if $\gamma \neq 0$. If damping is light,

$$|R| \approx Q/(m\omega_0^2) = Q/k, \quad \arg R \approx -\pi/2$$

i.e. the response lags $\pi/2$ behind the driving force, and the response is Q times larger than $\omega \to 0$ limit. So we can also regard Q as a measure of the amplification.

Velocity Response

We can derive the particle velocity's complex amplitude via

$$\begin{aligned} v_0 &= \mathrm{i}\omega x_0 = \mathrm{i}\omega F_0 R(\omega) \\ &= \frac{\mathrm{i}\omega F_0}{m[(\omega_0^2 - \omega^2) + \mathrm{i}\gamma\omega]} \\ &= \frac{F_0}{m[(\omega_0^2 - \omega^2)/(\mathrm{i}\omega) + \gamma]} \end{aligned} \tag{16}$$

The velocity resonance occurs where $|v_0|$ is at a maximum.

$$|v_0| = \frac{|F_0|}{m\sqrt{(\omega_0^2 - \omega^2)^2/\omega^2 + \gamma^2}}, \tag{17}$$

has maximum value at $\omega_v = \omega_0$ independent of the damping. From Eq. 16 we can see that, at this frequency, the velocity is exactly in phase with the driving force, and has magnitude $F_0/\gamma m$.

Velocity Response: Amplitude

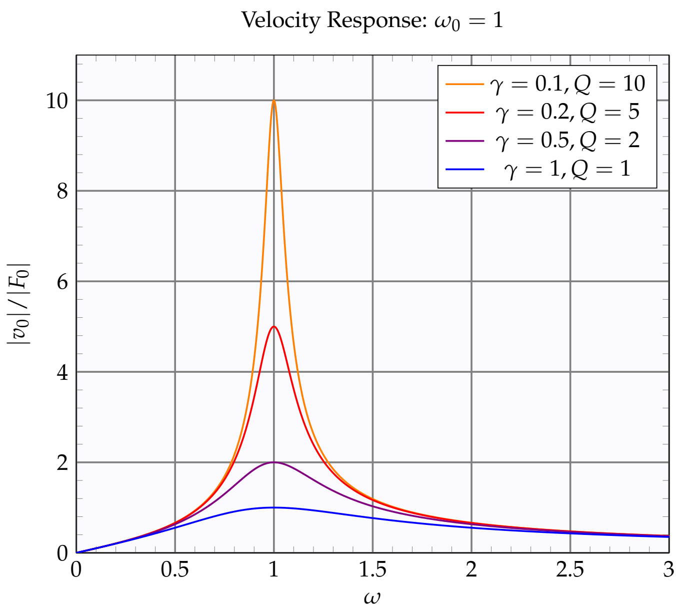


Figure 11: Velocity Response - Amplitude

$$\boxed{\left|\frac{v_0}{F_0}\right| = \frac{1}{m\sqrt{\frac{(\omega_0^2-\omega^2)^2}{\omega^2} + \gamma^2}}} \tag{18}$$

This has a peak value at $\omega = \omega_0$

Velocity Response: Phase

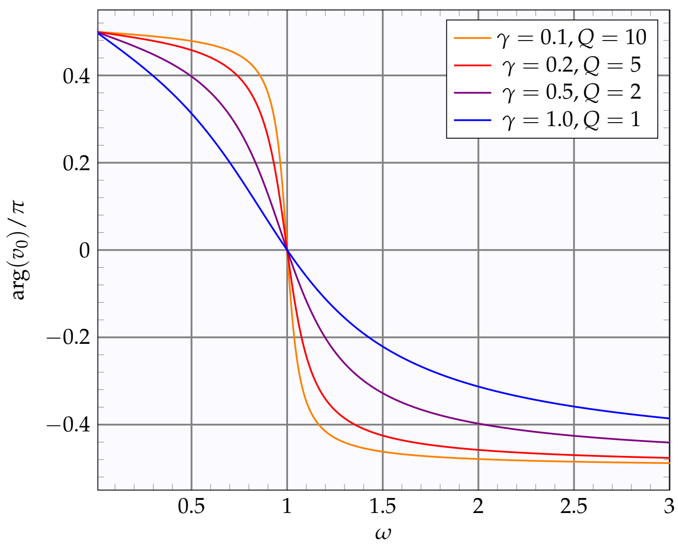


Figure 12: Velocity Response - Phase

$$\boxed{\arg\left(\frac{v_0}{F_0}\right) = \arctan\left[\frac{(\omega_0^2 - \omega^2)}{\gamma\omega}\right]} \tag{19}$$

Maximum response at $\omega = \omega_0$; velocity in-phase with the driving force at this velocity resonance.

Acceleration Response

We can derive the particle acceleration via

$$\begin{aligned} a_0 &= -\omega^2 x_0 = -\omega^2 F_0 R(\omega) \\ &= \frac{-\omega^2 F_0}{m[(\omega_0^2-\omega^2)+\mathrm{i}\gamma\omega]} \\ &= \frac{-F_0}{m[(\omega_0^2-\omega^2)/(\omega^2)+\frac{\mathrm{i}\gamma}{\omega}]} \end{aligned} \tag{20}$$

The acceleration resonance occurs where $|a_0|$ is at a maximum. Differentiate to find that the maximum acceleration occurs at frequency:

$$\omega_{\rm acc} = \omega_0 \left(1-\frac{1}{2Q^2}\right)^{-1/2}$$

Note this is above the frequency of undamped, unforced oscillations ω_0 and that $\omega_a\omega_{\rm acc}=\omega_0^2$, using $\omega_a=\omega_0\sqrt{1-\frac{1}{2Q^2}}$ (see Eq. 15).

Power: Multiplying Quantities using Complex Notation

The rate at which the driving force does work is "force $\times$ velocity": $P=Fv$. We have $F(t)=\Re\left[F_0\mathrm{e}^{\mathrm{i}\omega t}\right]$ and $v(t)=\Re\left[v_0\mathrm{e}^{\mathrm{i}\omega t}\right]$
How should we evaluate the product? "Take the real part" does not commute with "multiply". In general, to multiply two quantities represented in complex form, we must. . .

Take the real parts first!

$$\begin{aligned} \Re\{\mathbf{A}\}\Re\{\mathbf{B}\} &= \tfrac{1}{2}(\mathbf{A}+\mathbf{A}^*)\tfrac{1}{2}(\mathbf{B}+\mathbf{B}^*) \\ &= \tfrac{1}{4}(\mathbf{AB}+\mathbf{A}^*\mathbf{B}^*+\mathbf{AB}^*+\mathbf{A}^*\mathbf{B}) \\ &= \tfrac{1}{2}\Re\{\mathbf{AB}+\mathbf{AB}^*\} \end{aligned}$$

Now, if the quantities both vary as $\exp(\mathrm{i}\omega t)$, then

$$P=Fv=\Re\left[F_0\mathrm{e}^{\mathrm{i}\omega t}\right]\Re\left[v_0\mathrm{e}^{\mathrm{i}\omega t}\right]=\tfrac{1}{2}\Re\left[F_0v_0\mathrm{e}^{2\mathrm{i}\omega t}+F_0v_0^*\right].$$

Time average:

$$\langle P\rangle = \tfrac{1}{2}\Re\left[F_0v_0^*\right] \tag{21}$$

where$\langle\rangle$ denotes taking the mean value.

Phase differences and power factor

- Mean power input depends on phase difference between the force and the velocity: if $F_0 = |F_0|\mathrm{e}^{\mathrm{i}\phi_F}$ and $v_0 = |v_0|\mathrm{e}^{\mathrm{i}\phi_v}$ then

$$\langle P \rangle = \frac{1}{2}\Re[F_0 v_0^*] = \frac{1}{2}\Re\left[|F_0||v_0|\mathrm{e}^{\mathrm{i}(\phi_F-\phi_v)}\right] = \frac{1}{2}|F_0||v_0|\cos(\phi_F-\phi_v)$$

- $\cos(\phi_F - \phi_v)$ is the power factor in electrical circuits (after replacing force and velocity by voltage and current).
- If the force and velocity are are in phase ($\phi_F = \phi_v$), $\langle P \rangle = \frac{1}{2}|F_0||v_0|$
- if they are $\pi/2$ out of phase (in quadrature), then $\langle P \rangle = 0$.

Power in oscillators

What is the mean power required to drive a damped oscillator? Using equation 16:

$$\langle P \rangle = \frac{1}{2}\Re\left\{F_0 v_0^*\right\} = \frac{1}{2}\Re\left\{v_0 v_0^* m[(\omega_0^2-\omega^2)/(\mathrm{i}\omega)+\gamma]\right\} = \frac{1}{2}m\gamma|v_0|^2 = \frac{1}{2}b|v_0|^2 \quad (22)$$

Where does this energy go? In steady state, the average rate of working must equal the average power dissipated. The damping force bv is always in phase with the velocity v, so that

$$\langle P_{\mathrm{dissipated}} \rangle = \frac{1}{2}|F_r||v_0| = \frac{1}{2}b|v_0|^2 \quad (23)$$

as required.

Power Resonance and Bandwidth

The mean power input is $\frac{1}{2}b|v_0|^2$, so power resonance occurs at a frequency ω_P which is equal to the velocity resonance frequency, i.e. $\omega_P = \omega_v = \omega_0$.

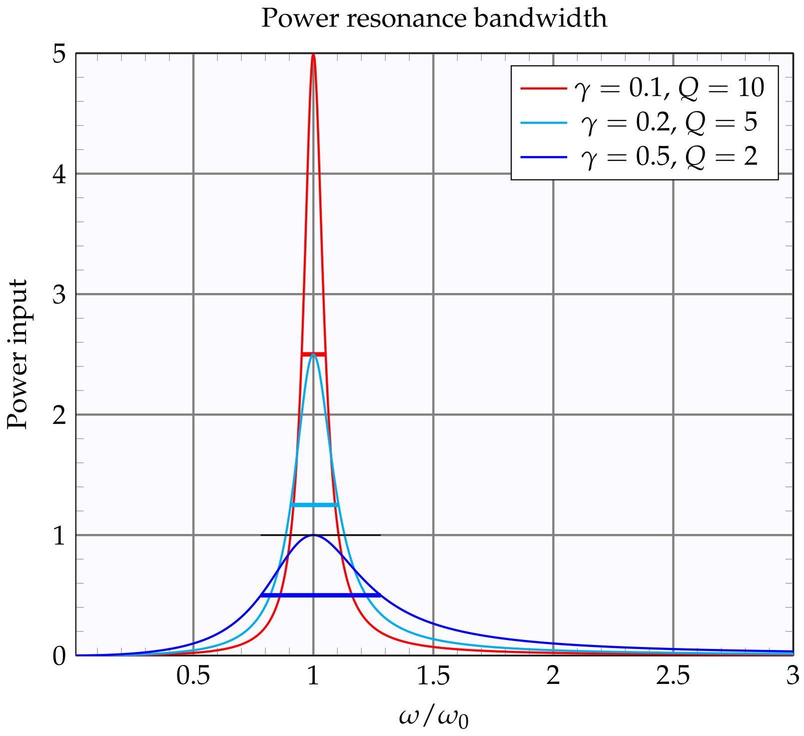


Figure 13: Bandwidth of a driven damped oscillator

Power Resonance and Bandwidth

- We can characterise the width of the resonance via its *half power points*, where power absorbed is half the value at power/velocity resonance.
- This corresponds to the two frequencies ω_+ and ω_- at which $|v_0| = |v_{\max}|/\sqrt{2}$. Using

$$|v_0|^2 = \frac{|F_0|^2}{m^2[(\omega_0^2 - \omega^2)^2/\omega^2 + \gamma^2]}.$$

- we see that

$$\gamma\omega_\pm = \mp(\omega_0^2 - \omega_\pm^2) \tag{24}$$

- Simple algebra shows the half-power bandwidth of the resonance is

$$\Delta\omega = \omega_+ - \omega_- = \gamma \tag{25}$$

- As expected, the peak gets wider as the damping goes up.
- The dimensionless ratio Q gives the ratio of the resonant frequency to the bandwidth

$$Q = \frac{\omega_0}{\gamma} = \frac{\omega_0}{\Delta\omega}.$$

- This provides another way of measuring Q.

General Response: driven and transient behaviour

- Driven SHO equation has two kinds of solutions:

$$F(t) = m\ddot{x} + b\dot{x} + kx \tag{26}$$

  - free oscillations, $F = 0$: solution characterised by 2 free parameters
  - steady-state response for $F \neq 0$ — from Eq.12: no free parameters
- Because the equation is linear in x we can add any free solution to the driven solution and still satisfy the equation.
- The general solution is thus the sum of the complementary function, x_c which is a solution to the homogeneous equation

$$m\ddot{x}_c + b\dot{x}_c + kx_c = 0 \tag{27}$$

  and a function that solves equ. (26), which is termed the particular integral
- This must be the general solution as it will include two free parameters.
- These are fixed by the initial conditions.

Driven

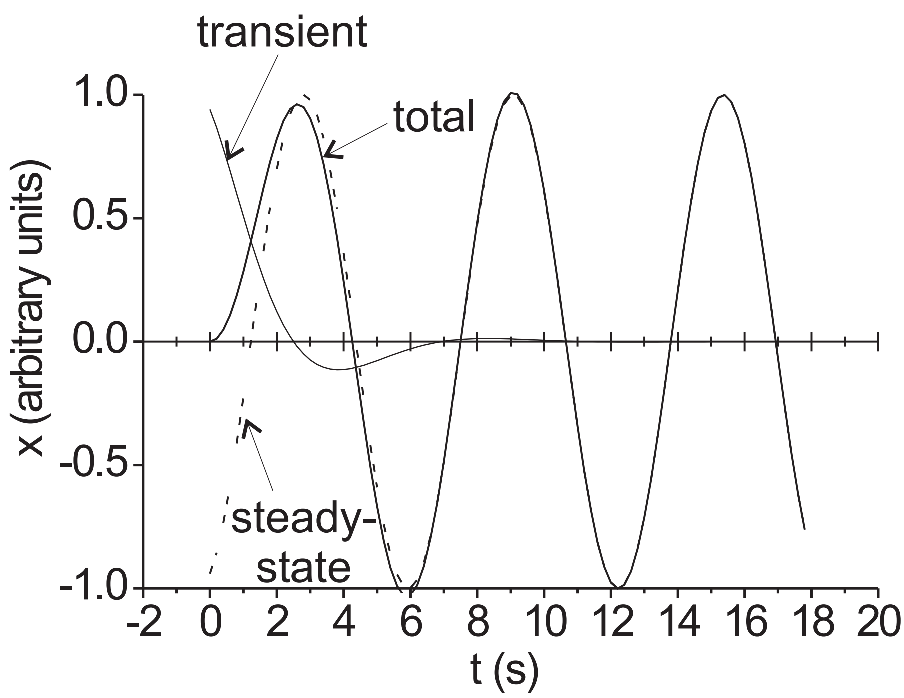


Figure 14: Oscillator driven from rest at $x = 0$ with $Q = 1$

Driven near the resonance

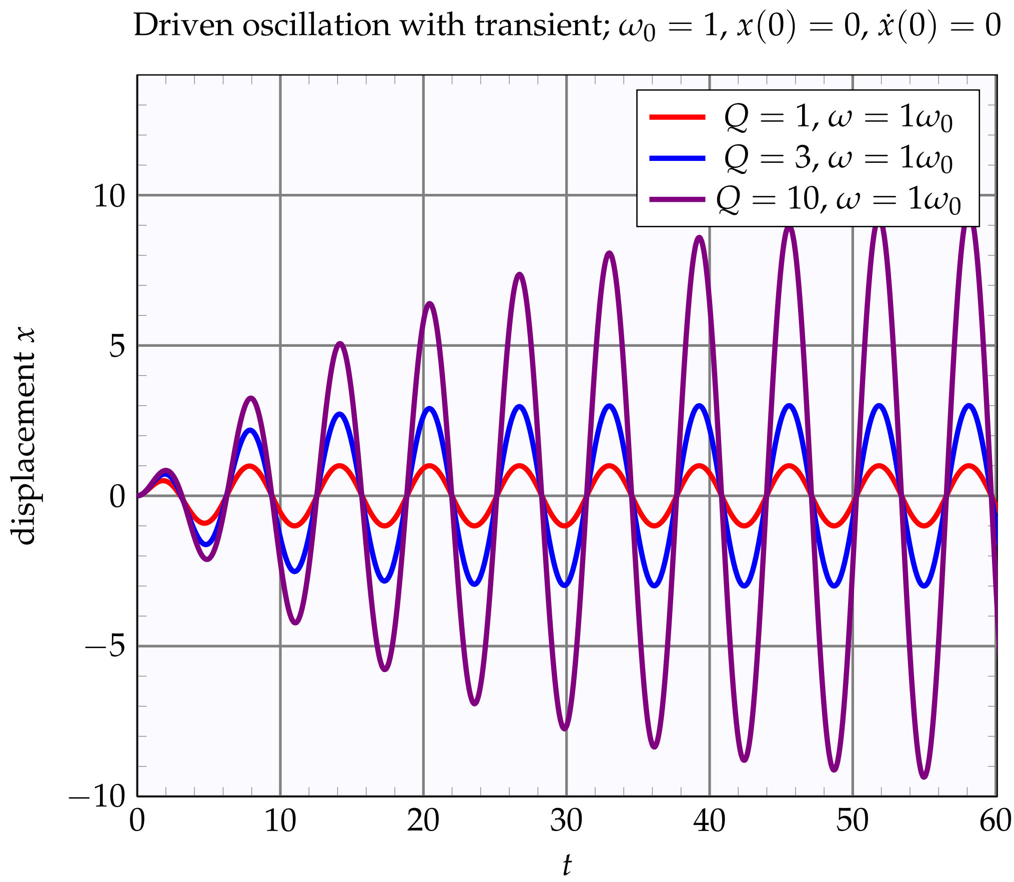


Figure 15: Oscillator driven from rest at $x = 0$

Driven Oscillation: Light damping

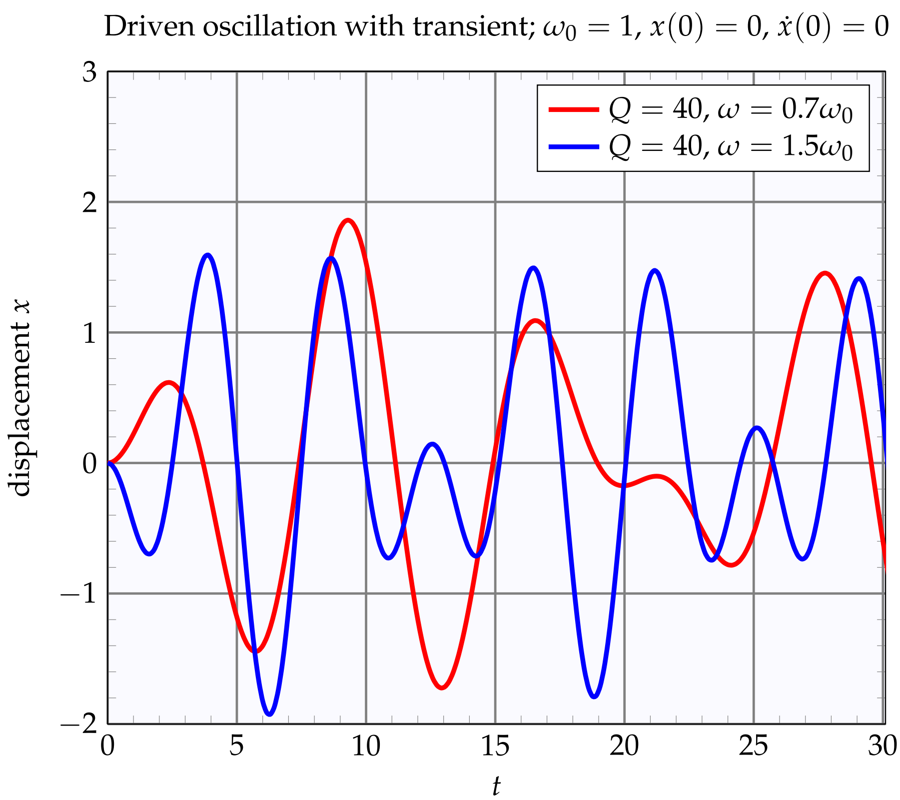


Figure 16: Oscillator driven from rest

Driven Oscillation: Light damping

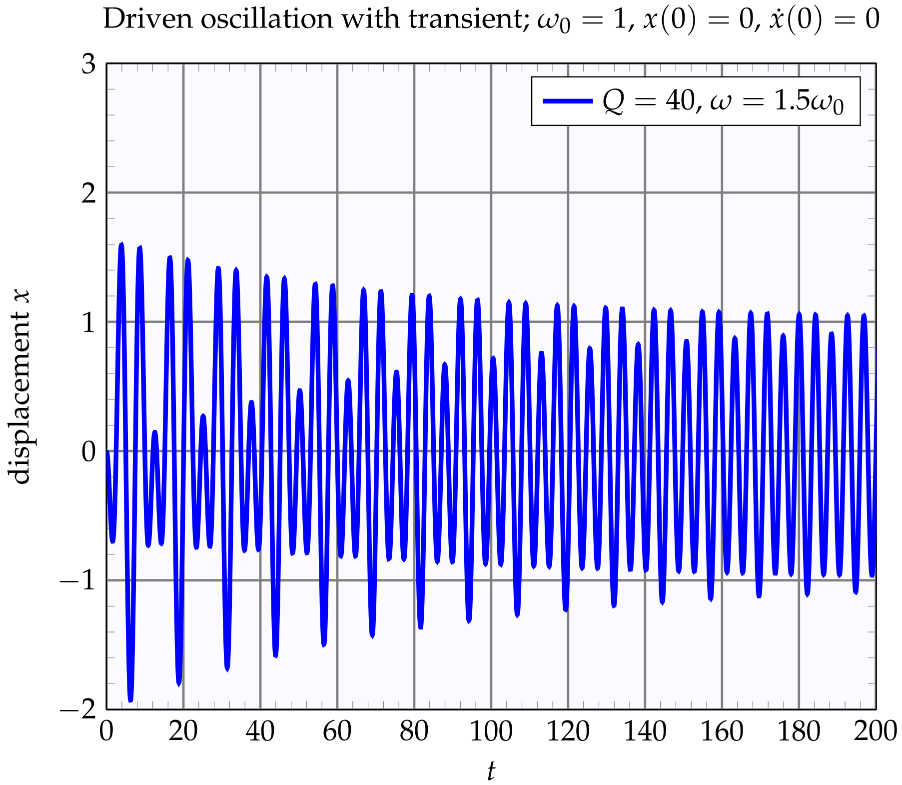


Figure 17: Oscillator driven from rest

Driven Electrical Circuits

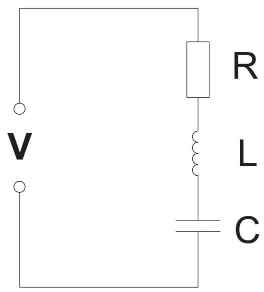


Figure 18: L-C-R circuit

$$V_L + V_R + V_C = V(t)$$

$$L\ddot{q} + R\dot{q} + \frac{1}{C}q = V(t)$$

Our analysis of a driven mechanical oscillator can be used in other areas of physics – electrical circuits, in particular.
We simply need to make some translations:

displacement, x	charge, q
velocity, v	current, I
force, F	voltage, V
mass, m	inductance, L
damping, b	resistance, R
spring constant, k	$(\text{capacitance})^{-1}$, $1/C$

Canonical Form of LCR circuit equation

We obtain the canonical form equ. (4)

$$\begin{aligned}\omega_0 &= \frac{1}{\sqrt{LC}} \\ \gamma &= R/L \\ Q &= \frac{\omega_0}{\gamma} = \frac{1}{R}\sqrt{\frac{L}{C}}\end{aligned}$$

so that we have

$$\boxed{\ddot{q} + \gamma\dot{q} + \omega_0^2 q = V/L}$$

and the current response (using equ. (18)) becomes:

$$\frac{|I|}{|V|} = \frac{1}{\sqrt{(\omega L - \frac{1}{C\omega})^2 + R^2}} \tag{28}$$

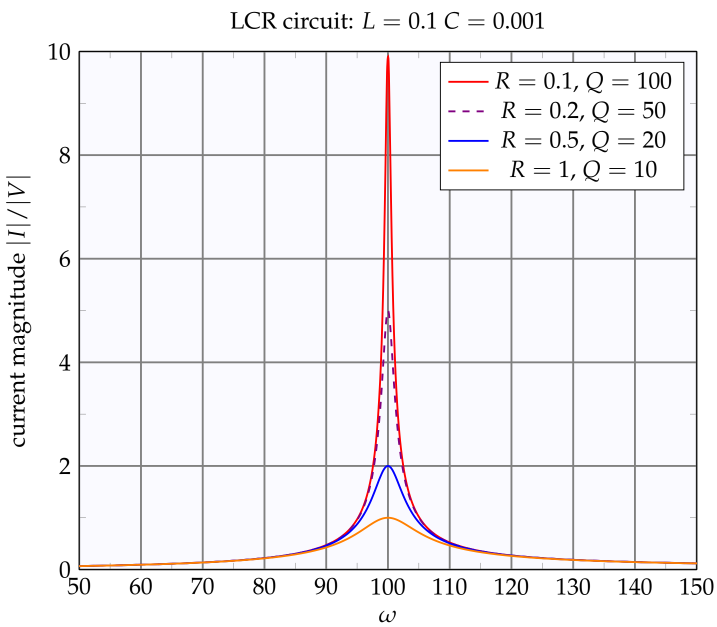


Figure 19: LCR resonance

- Current peaks at the resonant frequency $\omega_0 = (LC)^{-0.5}$, as does V_R ('velocity')
- V_C: 'displacement' (peaks just below ω_0)
- V_L: 'acceleration' (peaks above ω_0)
- Power Bandwidth $\Delta\omega$ given by $Q = \omega_0/\Delta\omega$ (a route to measuring Q)

Impedance

Impedance relates current to voltage: if $V = \Re\{V_0 e^{i\omega t}\}$ gives rise to a current of $I = \Re\{I_0 e^{i\omega t}\}$, where V_0 and I_0 are complex, then the impedance Z is given by

$$Z = V_0/I_0$$

The electrical impedance of the $L - C - R$ series circuit

$$Z = i\omega L + \frac{1}{i\omega C} + R = Z_L + Z_C + Z_R. \tag{29}$$

The power dissipated is given by

$$\begin{aligned} \langle P \rangle &= \frac{1}{2}\Re\{V_0 I_0^*\} \\ &= \frac{1}{2}|V_0|^2\Re\{1/Z\} = \frac{1}{2}|I_0|^2\Re\{Z\} \\ &= \frac{1}{2}|I_0|^2 R. \end{aligned} \tag{30}$$

Mechanical Impedance

Impedance is a useful concept in most oscillating systems. For the mechanical oscillator, we define Z to be the ratio of the complex force to the complex velocity:

$$Z = \frac{F_0}{v_0}$$

which from Eq. 16 is given by

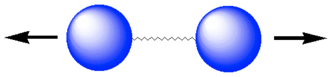
$$Z = m\left[\frac{\omega_0^2 - \omega^2}{i\omega} + \gamma\right]$$

We can write down expressions for the mean power dissipated in terms of the impedance as

$$\begin{aligned} \langle P \rangle &= \frac{1}{2}\Re\{F_0 v_0^*\} \\ &= \frac{1}{2}|F_0|^2\Re\{1/Z\} = \frac{1}{2}|v_0|^2\Re\{Z\}, \end{aligned} \tag{31}$$

which can be compared with the electrical equivalents.

Atoms as Oscillators



- How does light interact with atomic matter?
- Consider a semi-classical model of a hydrogen atom: a uniform spherical electron cloud of radius a. It is driven by an electric field $E_0 \mathrm{e}^{\mathrm{i}\omega t}$ (e.g. from a light wave). How does it respond?
- If the proton moves a distance x from the centre of the electron cloud, it feels a restoring force
$$F = \left(\frac{x^3}{a^3} e\right) \times e \times \frac{1}{4\pi\epsilon_0 x^2} = kx$$
where the effective spring constant is thus $k = e^2/4\pi\epsilon_0 a^3$
- We can thus find the natural undamped oscillation frequency $\omega_0^2 = k/m$:
$$\omega_0^2 = \frac{k}{m} = \frac{e^2}{4\pi\epsilon_0 m_e a^3}$$
- For $a = 50\,\mathrm{pm}$ (Bohr radius) we get $f_0 \approx 1 \times 10^{16}$ Hz.
- Optical light has $f_{\mathrm{blue}} \approx 8 \times 10^{14}$ Hz; suggests atoms should behave as oscillators driven below resonance when illuminated by light — stiffness controlled; amplitude independent of ω. (This is a simple model for scattering).

Damping in Atomic Oscillators

- What is the source of the damping? If a charge moves as $x_0 \mathrm{e}^{\mathrm{i}\omega t}$, it will radiate energy at an average rate
$$P_{\mathrm{rad}} = \frac{e^2 a_0^2}{12\pi\epsilon_0 c^3} = \frac{e^2 x_0^2 \omega^4}{12\pi\epsilon_0 c^3}$$
[Larmor's Radiation formula — you don't need to know this (Part II).]
- The mean energy of the electron oscillating is $W = m_e \omega^2 x_0^2/2$ (time average of KE + PE). §
- Hence
$$\frac{\mathrm{d}W}{\mathrm{d}t} = -P_{\mathrm{rad}} = -\frac{e^2\omega^2}{6\pi\epsilon_0 c^3 m_e} W$$
Thus the energy decays exponentially.
- But a lightly damped oscillator's energy decays with time constant $1/\gamma$, hence
$$\gamma = \frac{e^2\omega^2}{6\pi\epsilon_0 c^3 m_e} \approx 1.3 \times 10^{10}\,\mathrm{s}^{-1}$$
- Hence we estimate $Q = \omega_0/\gamma$ for the atom as $\sim 10^6$.

Anharmonic Oscillations*(non-examinable)

- In real world, perfect harmonic oscillations are rare: need exactly quadratic potential.
- The equation becomes non-linear: cannot superpose solutions
- For example, a pendulum has a restoring force $\propto \sin\theta \approx \theta - \theta^3/3! + \ldots$. If we assume the force is proportional to θ, the fractional error in the force $(\theta^2/6)$ is of order 0.5% for a swing of 10°, and $\sim 5\%$ for a swing of 30°
- Consider a symmetric, anharmonic potential

$$V(x) = \frac{kx^2}{2}\left(1 + \frac{\alpha}{2}x^2\right)$$

So the restoring force is $kx(1+\alpha x^2)$ and the equation of motion is then

$$\ddot{x} + (1+\alpha x^2)\omega_0^2 x = 0 \qquad (32)$$

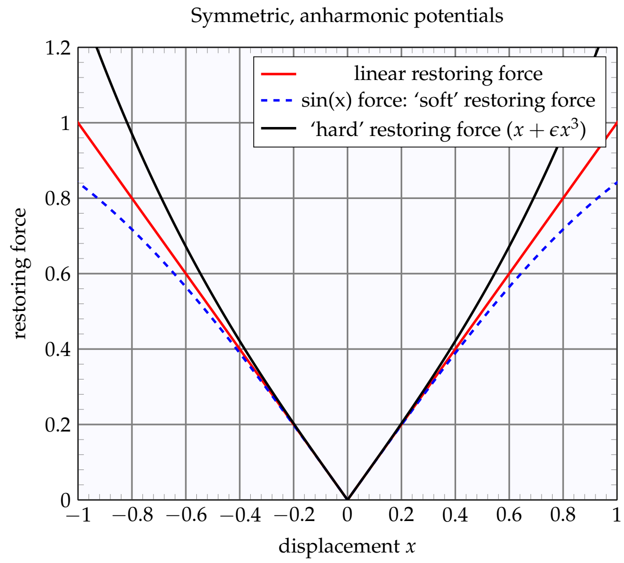


Figure 20: Soft and Hard Restoring Forces

- If $\alpha > 0$, the potential is hard — the spring stiffens with extension. $\alpha < 0$ implies a soft potential.A pendulum is soft, for example.

Anharmonic Oscillations*

- Look for solutions when $|\alpha| \ll 1$; and assume that the mass is at rest at $t = 0$ for simplicity. We need a solution which has the correct periodicity and symmetry: a suitable form is

$$x(t) = A(\cos\omega_f t + \epsilon\cos 3\omega_f t + \ldots)$$

- Two unknowns: for small $|\alpha|$, expect $\omega_f \approx \omega_0$, and ϵ to be small compared to unity.

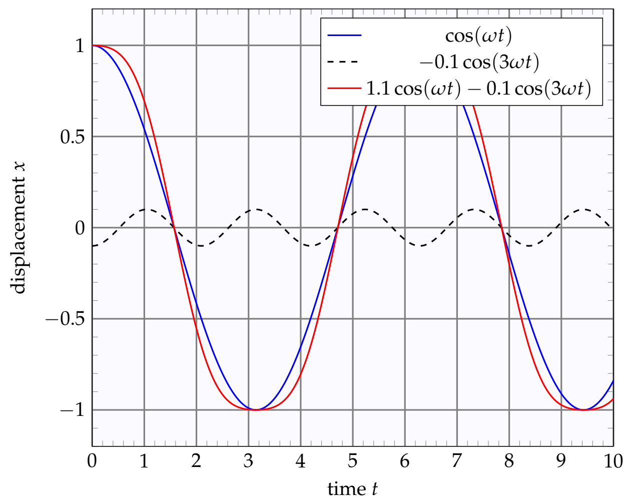


Figure 21: Approximate solution for symmetric anharmonic potential

Anharmonic Oscillations*

- The amplitude is of order A, so the anharmonic correction is small if $|\alpha A^2| \ll 1$
- Now substitute this expression into the equation of motion, working to first order in αA^2 and ϵ. We find that this obeys the equation if

$$\omega_f \approx \omega_0 (1 + \frac{3}{8}\alpha A^2)$$

$$\epsilon \approx \frac{1}{32}\alpha A^2$$

(See Main's book (CH 7, pp 93-99) for all the algebra)

- So that the period is slightly longer than in harmonic potential if $\alpha < 0$, as we expect.
- Moreover, the period is now amplitude-dependent.

Application to a Pendulum*

- The equation of motion is, denoting the angle of swing by x:

$$0 = \ddot{x} + \frac{g}{l}\sin x \approx \ddot{x} + \frac{g}{l}x\left(1 - \frac{x^2}{6}\right)$$

- So we have, from equ. (32), $\alpha = -1/6$ i.e. a soft potential. This yields

$$\omega_f \approx \omega_0 \left(1 - \frac{A^2}{16}\right), \quad \epsilon \approx -\frac{A^2}{192}$$

- So if we put $A = 10^\circ \approx 0.175\,\mathrm{rad}$, we get a fractional change in period of 0.19% compared to the harmonic solution. (This is about 164 s per day).

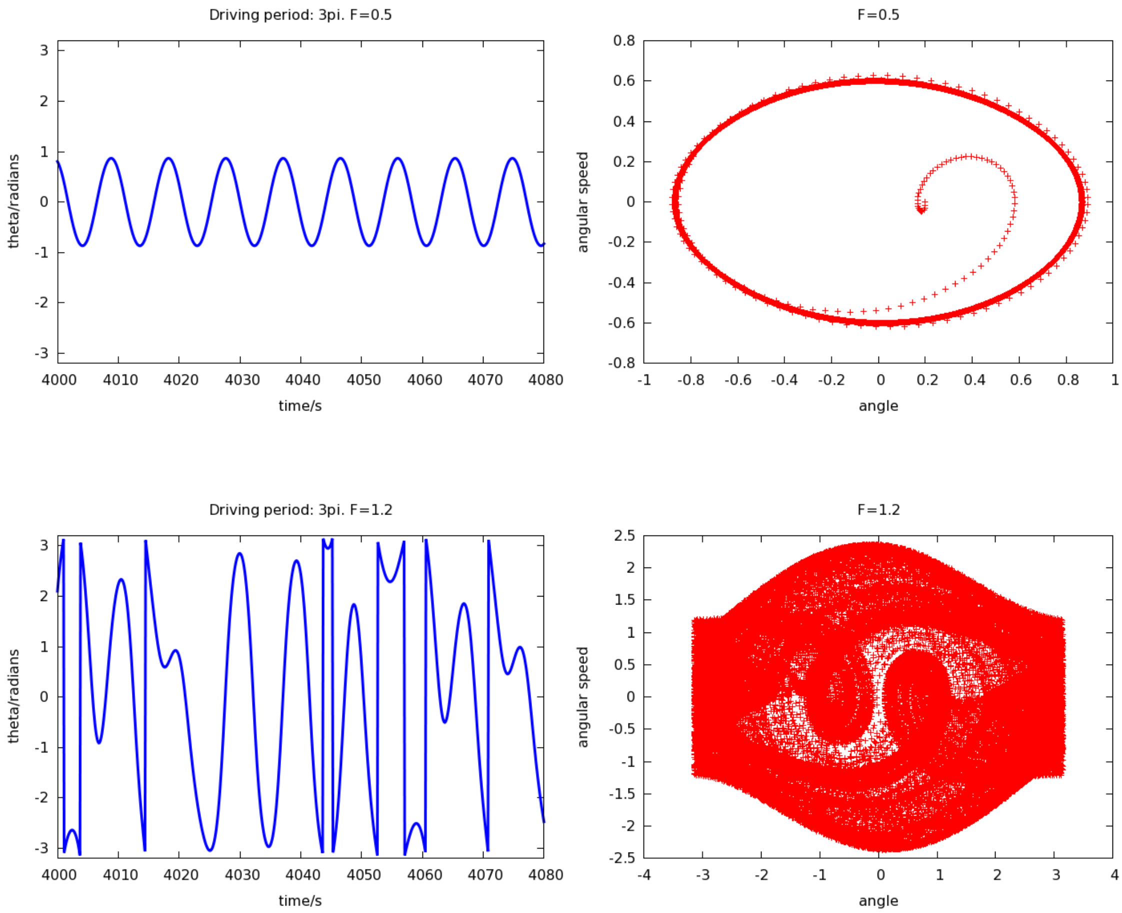
We could also examine asymmetric potentials, in which case our expansion would include even harmonics. In the end, we will need to use numerical integration of the differential equation for large amplitude anharmonic oscillations.

Driven Anharmonic Oscillators*

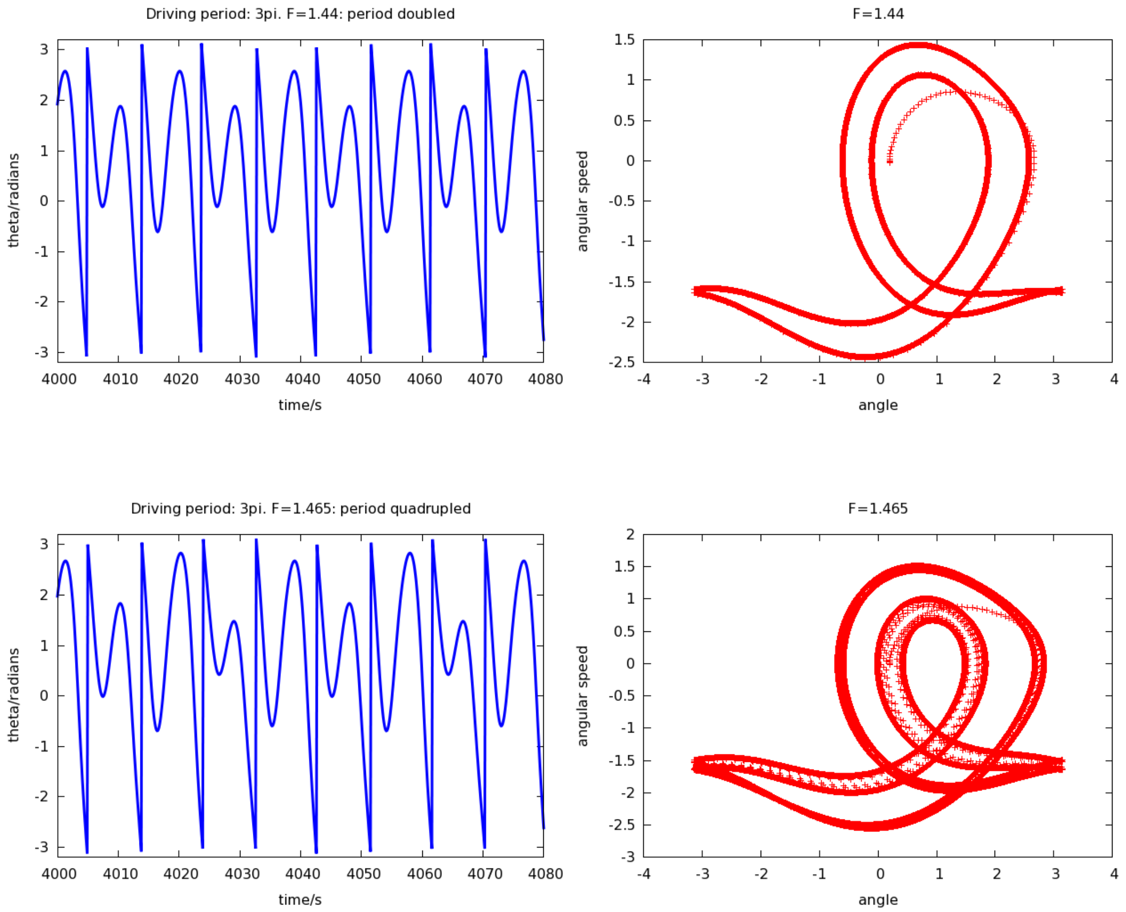
What happens if we drive a non-linear oscillator with a sinusoidal forcing term? A whole rich set of dynamical phenomena emerge. Here's an on-line demo.
http://www.myphysicslab.com/pendulum2.html
New behaviours:

- 'Deterministic Chaos': driven response can look very complex
- Furthermore, even if initial conditions extremely close, solutions will diverge — The Butterfly Effect
- Period doubling, quadrupling etc: the final driven solution can contain *sub-harmonics* of the driving period.

Increasing the driving force*



Increasing the driving force*



Dependence on initial conditions*

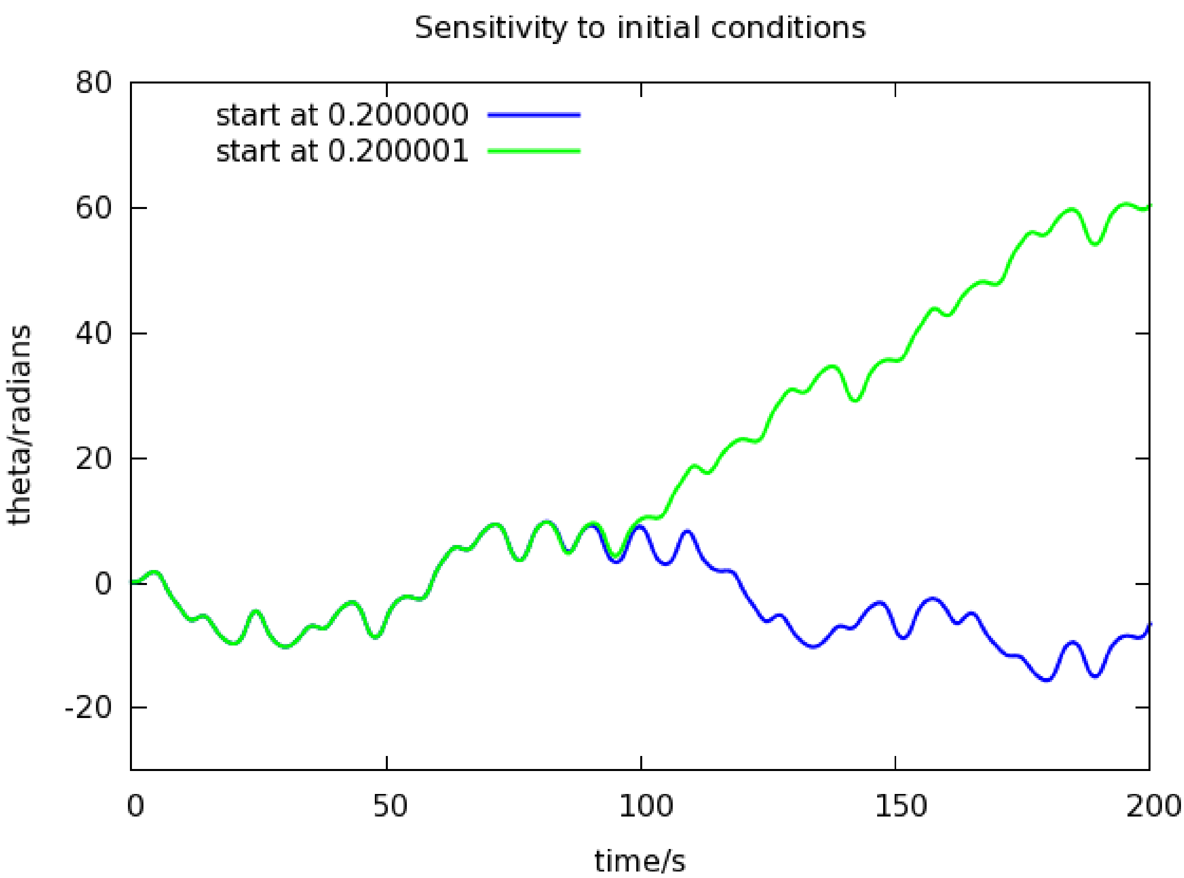


Figure 22: The Butterfly Effect: Initial amplitudes differ by only 1 part in 10^5

Linear System Driven at Multiple Frequencies

Consider driving a system with sum of two harmonic forces:

$$F(t) = F_1 \cos(\omega_1 t + \alpha_1) + F_2 \cos(\omega_2 t + \alpha_2).$$

Because the equation of motion is linear, the solution is found by adding the responses as if the individual driving forces acted separately.
Initially, suppose the driving forces have the same angular frequency $\omega_1 = \omega_2 = \omega$. Then

$$x = x_1 + x_2 = A_1 \cos(\omega t + \alpha_1 + \phi) + A_2 \cos(\omega t + \alpha_2 + \phi)$$

with $A_i = F_i|R|$ and $\phi = \arg(R)$. We can draw a phasor diagram because the frequencies are the same.

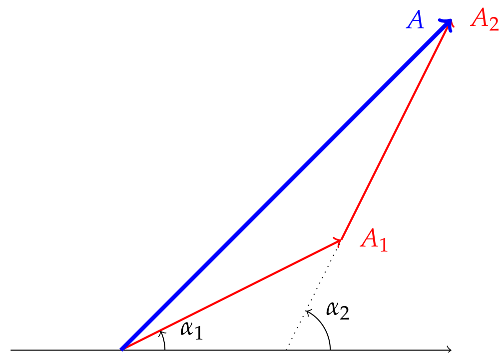


Figure 23: Summing responses from two driving forces

Coherence

The resulting amplitude depends on the relative phase $\alpha_2 - \alpha_1$.

$$A^2 = A_1^2 + A_2^2 + 2A_1A_2\cos(\alpha_2 - \alpha_1)$$

If we choose equal driving forces, $F_1 = F_2$, then

$$A^2 = 2A_1^2\left[1 + \cos(\alpha_2 - \alpha_1)\right]$$

- In-phase driving: double the amplitude response, quadruple the power
- Anti-phase driving: no response (net force is zero)

However, in practice, unless the driving terms F_1 and F_2 are highly stable, the term $\cos(\alpha_2 - \alpha_1)$ is likely to drift and average to zero over time. For example, two independent pendulums of equal length will eventually drift in phase due to tiny differences in e.g. their lengths. In this case, the driven response is the quadrature sum of the individual responses:

$$A = \sqrt{A_1^2 + A_2^2}$$

We say the forces are incoherent, and in this case we add the energies not the amplitudes.To achieve coherence, we probably need to derive the driving forces from a common source.Two oscillations may be coherent over short periods (so that $\cos(\alpha_2 - \alpha_1)$ does not average to zero), but incoherent over longer times. Next year we will quantify this.
Note also that the changing phase $\alpha(t)$ implies the existence of extra frequency components in the oscillation.

Different Driving Frequencies

The response is, as before, simply the sum of the individual responses:

$$x = x_1 + x_2 = A_1 \cos(\omega_1 t + \alpha_1 + \phi) + A_2 \cos(\omega_2 t + \alpha_2 + \phi)$$

In general, we get beating between the different frequencies, with a fast oscillation at angular frequency $\frac{1}{2}(\omega_1 + \omega_2)$ modulated by a slower envelope of angular frequency $\frac{1}{2}(\omega_1 - \omega_2)$

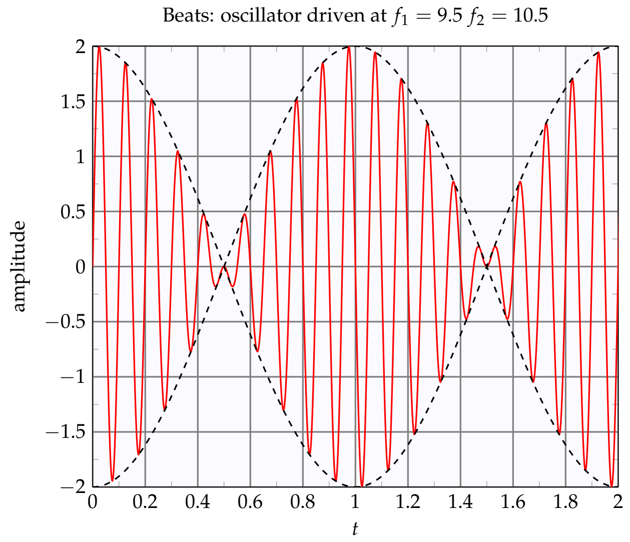
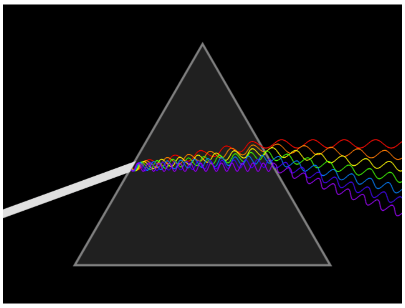
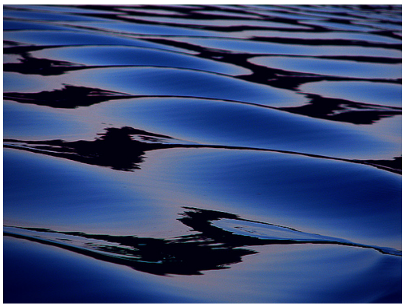


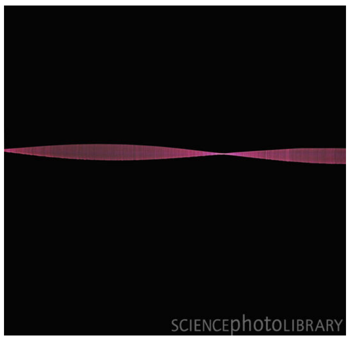
Figure 24: Beats for equal amplitude responses $A_1 = A_2$

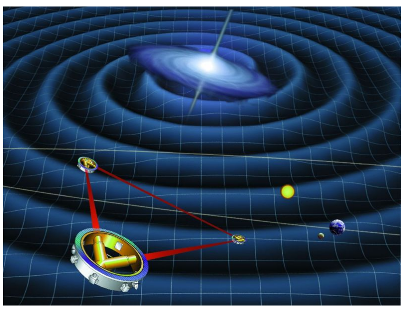
Waves

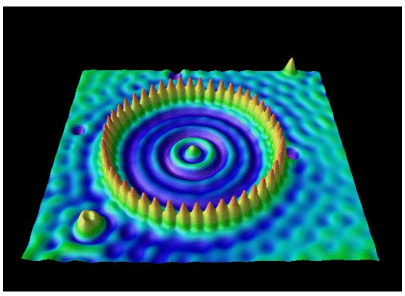
Waves are central to our understanding of the Universe.

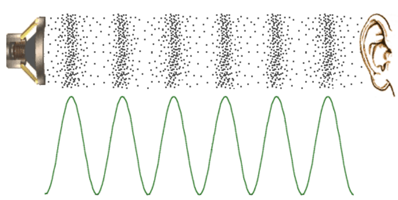












A wave moves energy, and information, from one point to another without bulk translation of the medium which it disturbs.The medium maybe a solid, a liquid, or a gas; or the vacuum in the case of electromagnetic or quantum waves.

Transverse String Waves

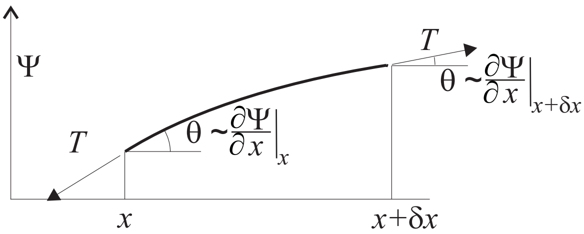


Figure 25:

Transverse waves on a string are a good model: we can derive some general results.
The transverse displacement is $\Psi(x,t)$; the tension is T and the mass per unit length is ρ.
If string's gradient $\psi' \ll 1$, then the tension is approximately uniform — because the extension caused by the wave is

$$\left(\int_0^L \sqrt{1+\psi'^2}\mathrm{d}x\right) - L = \mathcal{O}(\psi'^2).$$

A segment of length δx feels a restoring force towards the axis of

$$T\left(\left.\frac{\partial\Psi}{\partial x}\right|_x - \left.\frac{\partial\Psi}{\partial x}\right|_{x+\delta x}\right) = T\left(\left.\frac{\partial\Psi}{\partial x}\right|_x - \left\{\left.\frac{\partial\Psi}{\partial x}\right|_x + \left.\frac{\partial^2\Psi}{\partial x^2}\right|_x \delta x\right\}\right) = -T\left.\frac{\partial^2\Psi}{\partial x^2}\right|_x \delta x$$

Newton's second law yields:

$$\boxed{\frac{\partial^2\Psi}{\partial t^2} = v^2\frac{\partial^2\Psi}{\partial x^2}} \tag{33}$$

where $v = \sqrt{T/\rho}$.

Solutions To The 1-D Wave Equation

The (non-dispersive) wave equation occurs in many areas of physics.

$$\frac{\partial^2\Psi}{\partial t^2} = v^2\frac{\partial^2\Psi}{\partial x^2} \tag{34}$$

We will see that waves of all frequencies travel at the same speed v; and so does a general disturbance. This is what we mean by non-dispersive. Suppose that a disturbance of shape $f(x)$ propagates unchanged at speed v on a string:

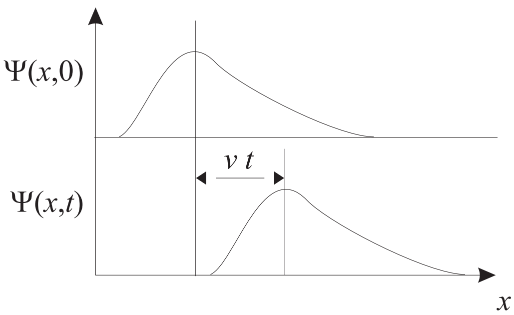


Figure 26: Wave motion.

We can write

$$\Psi(x,t) = \Psi(x \pm vt, 0) = f(x \pm vt)$$

The sign indicates direction of motion, with the negative sign implying motion to $+x$.
We now show that this waveform obeys the wave equation.

Propagation of a General Disturbance

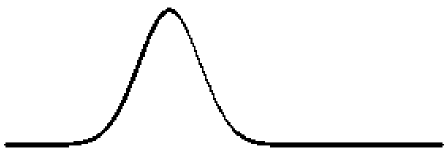


Figure 27:

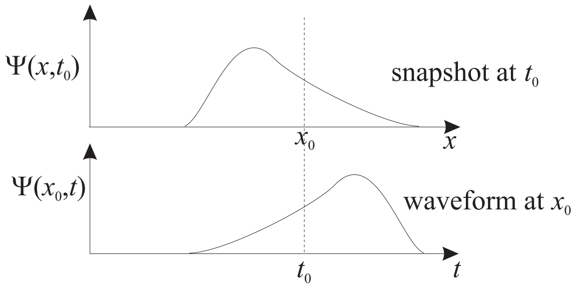


Figure 28: Waveforms and snapshots.

Let $u = x - vt$, so that $\Psi(x,t) = f(x - vt) = f(u)$. Then

$$\frac{\partial \Psi}{\partial x} = \frac{\mathrm{d}f}{\mathrm{d}u}\frac{\partial u}{\partial x} = \frac{\mathrm{d}f}{\mathrm{d}u}; \quad \frac{\partial^2 \Psi}{\partial x^2} = \frac{d^2 f}{du^2}\frac{\partial u}{\partial x} = \frac{d^2 f}{du^2}.$$

and

$$\frac{\partial \Psi}{\partial t} = \frac{df}{du}\frac{\partial u}{\partial t} = -v\frac{df}{du}; \quad \frac{\partial^2 \Psi}{\partial t^2} = -v\frac{d^2 f}{du^2}\frac{\partial u}{\partial t} = v^2\frac{d^2 f}{du^2}$$

v is the wave speed or the phase speed, and is determined by the physical properties of the wave medium. The same equation will result if the wave is travelling to the left. So we have proved the result that $f(x \pm vt)$ is a solution to the wave equation.

The Wave Equation

$$\boxed{\frac{\partial^2 \Psi}{\partial t^2} = v^2\frac{\partial^2 \Psi}{\partial x^2}}$$

Important features of this equation.

- It is general — applies to any waveform.
- Wave form will propagate without change of shape if v is constant
- It is linear in Ψ, i.e. all the terms involving Ψ are raised to the first power only. Just as we have seen previously for simple oscillating systems, this leads to the principle of superposition. If Ψ_1 and Ψ_2 each satisfy the wave equation, then any combination $\Psi = \alpha\Psi_1 + \beta\Psi_2$ is also a solution. The linearity of the wave equation thus means we can use Fourier analysis to examine the properties of waves.

Harmonic Waves

A special case is a continuous sinusoidal wave – a harmonic wave.
In a harmonic wave, the displacement $\Psi(x,t)$ varies sinusoidally with time at any point x. The displacement at $x=0$ will be given by

$$\Psi(0,t) = \Re\left(A\mathrm{e}^{\mathrm{i}\omega t}\right)$$

where A is a complex number. The wave (travelling in the positive x direction) will have the general form $f(x-vt)$, hence elsewhere it must be given by

$$\Psi(x,t) = \Re\left(A\mathrm{e}^{\mathrm{i}\omega(t-x/v)}\right) = \Re\left(A\mathrm{e}^{\mathrm{i}(\omega t-kx)}\right)$$

with $k=\omega/v$. Here, v is the wave or phase velocity, and $k=2\pi/\lambda$ is the wavenumber or wavevector.
The phase of the wave is $(\omega t-kx)$. We can use various representations as we did for oscillators; in each case we have two free parameters to specify the harmonic wave, and a third piece of information is the direction of travel.

$$\begin{aligned}
\psi(x,t) &= A\cos(\omega t-kx+\phi)\\
\psi(x,t) &= B_P\cos(\omega t-kx)+B_q\sin(\omega t-kx)\\
\psi(x,t) &= C\mathrm{e}^{\mathrm{i}\omega t-\mathrm{i}kx}+C^*\mathrm{e}^{-\mathrm{i}\omega t+\mathrm{i}kx}\\
\psi(x,t) &= \Re\Big(D\mathrm{e}^{\mathrm{i}\omega t-\mathrm{i}kx}\Big)
\end{aligned}$$

Waves in 2 and 3 dimensions

In two dimensions, the string becomes a stretched drumskin (assume horizontal). If the mass per unit area is σ and the tension per unit length in the drumskin is γ, then the vertical displacement Ψ of a small rectangle $(\mathrm{d}x,\mathrm{d}y)$ obeys

$$\sigma\mathrm{d}x\mathrm{d}y\frac{\partial^2\Psi}{\partial t^2} = \gamma\mathrm{d}y\,\frac{\partial^2\Psi}{\partial x^2}\mathrm{d}x+\gamma\mathrm{d}x\,\frac{\partial^2\Psi}{\partial y^2}\mathrm{d}y$$

Hence the wave speed is $\sqrt{\gamma/\sigma}$ and

$$\frac{\partial^2\Psi}{\partial t^2} = \frac{\gamma}{\sigma}\left(\frac{\partial^2\Psi}{\partial x^2}+\frac{\partial^2\Psi}{\partial y^2}\right) = v^2\left(\frac{\partial^2\Psi}{\partial x^2}+\frac{\partial^2\Psi}{\partial y^2}\right)$$

In general, the wave equation has the form

$$v^2\boldsymbol{\nabla}^2\Psi = \frac{\partial^2\Psi}{\partial t^2}$$

Plane Waves

A plane is defined by the geometric constraint $\hat{\boldsymbol{n}} \cdot \boldsymbol{r} = d$ for some constant d, the distance from the origin to the plane. $\hat{\boldsymbol{n}}$ is the unit vector normal to the plane.

Hence if we define a wavevector $\boldsymbol{k} = k\hat{\boldsymbol{n}}$, where $k = 2\pi/\lambda$, then it is clear that $\boldsymbol{k} \cdot \boldsymbol{r}$ is constant on this plane.

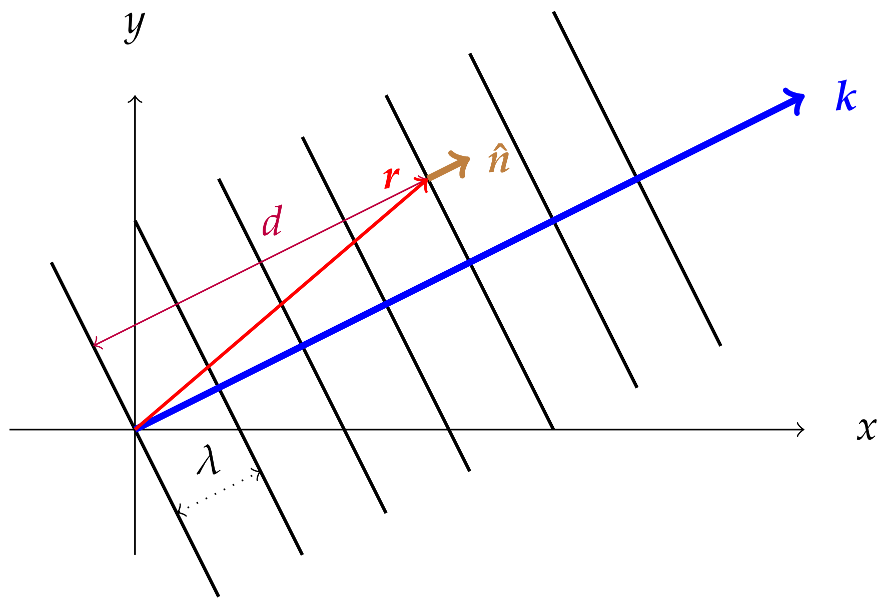


Figure 29: Plane wave front and wavevector

Plane Waves

A plane wave propagating in the direction $\hat{\boldsymbol{k}}$ can thus be written

$$\psi(\boldsymbol{r}, t) = \Re\Big(A\mathrm{e}^{\mathrm{i}\omega t - \mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\Big) \tag{35}$$

and we can get wave travelling in the opposite direction $-\hat{\boldsymbol{k}}$ by

$$\psi(\boldsymbol{r}, t) = \Re\Big(A\mathrm{e}^{\mathrm{i}\omega t + \mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\Big) \tag{36}$$

Note that to differentiate the first expression wrt time, multiply by $\mathrm{i}\omega$.
And to take its gradient, $\boldsymbol{\nabla}\psi$, multiply by $-\mathrm{i}\boldsymbol{k}$.
So for example,

$$\boldsymbol{\nabla}^2\psi = (-\mathrm{i}\boldsymbol{k}) \cdot (-\mathrm{i}\boldsymbol{k})\psi = -(k_x^2 + k_y^2 + k_z^2)\psi = -k^2\psi;$$
$$\ddot{\psi} = (\mathrm{i}\omega)(\mathrm{i}\omega)\psi = -\omega^2\psi$$

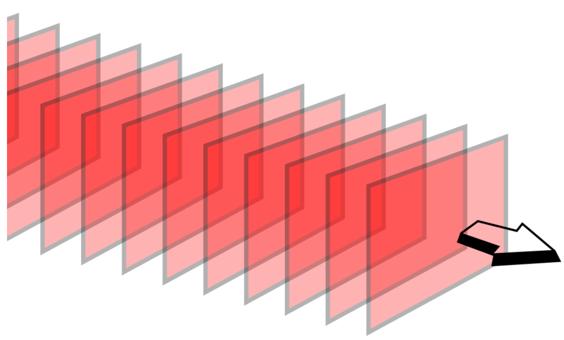


Figure 30: Wavefronts for a plane wave

Spherical Waves

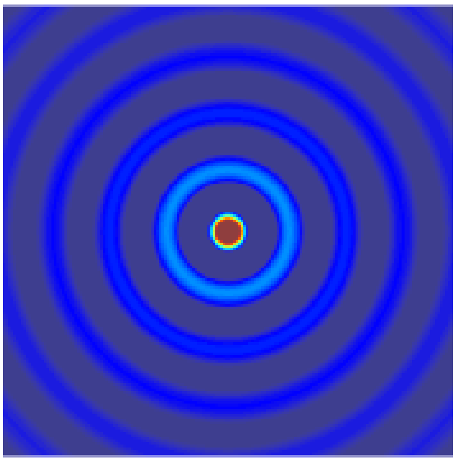


Figure 31:

Suppose a wave propagates from a point with spherical symmetry. The wave equation is now (ignoring the θ, ϕ parts which do not vary):

$$v^2 \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) = \frac{\partial^2 \psi}{\partial t^2}$$

Consider a function

$$\psi(r,t) = \frac{f(r \pm vt)}{r} \tag{37}$$

where the function f describes a disturbance propagating outwards/inwards.

Spherical Waves

Then,

$$\frac{\partial \psi}{\partial r} = -\frac{f}{r^2} + \frac{f'}{r}$$

The LHS of the wave equation is then equal to

$$v^2 \frac{\partial^2 \psi}{\partial r^2} = \frac{v^2}{r^2} \left(-f' + f' + r f'' \right) = \frac{v^2 f''}{r}$$

The RHS of the wave equation is

$$\frac{\partial^2 \psi}{\partial t^2} = \frac{v^2 f''}{r}$$

Hence equ. (37) represents a valid general solution to the wave equation in 3-dimensions. In particular, we have a harmonic solution for outward propagating waves:

$$\psi(r,t) = \Re \left(\frac{A e^{(i\omega t - ikr)}}{r} \right)$$

The $1/r$ amplitude dependence can be seen to conserve energy (i.e. this is the inverse square law for power).

Cylindrical Waves

If we launch a wave from a line source instead of a point source, we can look for solutions in cylindrical symmetry. The wave equation in cylindrical polar coordinates, with no angular dependence, reads

$$v^2 \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \psi}{\partial r}\right) = \frac{\partial^2 \psi}{\partial t^2}$$

Let's guess a possible solution

$$\psi(r,t) = \frac{f(r \pm vt)}{\sqrt{r}} \tag{38}$$

where the function f describes a disturbance propagating outwards/inwards. The $\sqrt{r}$ will conserve energy flow. The wave equation yields

$$v^2 \frac{f''}{\sqrt{r}} = v^2 \frac{f''}{\sqrt{r}} + \frac{v^2 f}{4r^{5/2}}$$

So the function is not a solution. However, if the second term on the RHS is small compared to the first, the solution is approximately correct. We require

$$\left|\frac{f}{4f''r^2}\right| \ll 1 \quad \rightarrow (kr) \gg 1$$

where we have used $f'' = -k^2 f$ (The full solution involves Bessel functions for small r.) Hence equ. (38) represents a valid (approximate) solution to the wave equation for $r \gg \lambda$.

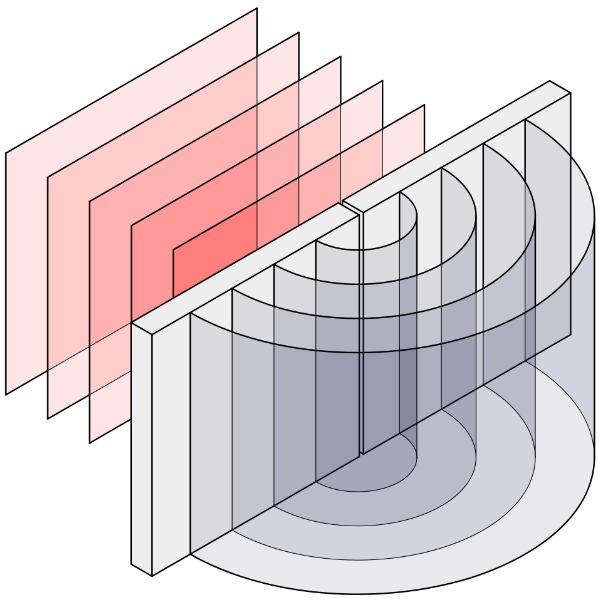


Figure 32: A cylindrical wave created by diffraction at a long slit.

$$\psi(r,t) \approx \Re\left(\frac{f(r \pm vt)}{\sqrt{r}}\right)$$

Polarisation

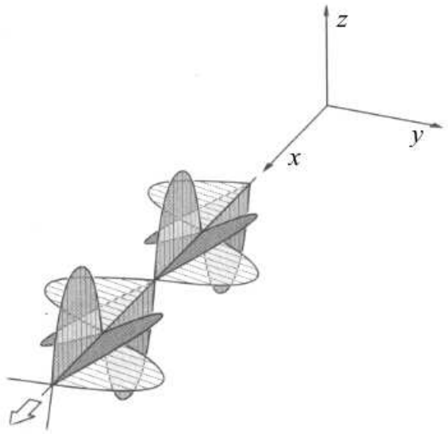


Figure 33: Linearly polarised wave (from Hecht, Optics).

We can displace the string in a transverse direction either vertically (ψ_z) or horizontally (ψ_y). Transverse means orthogonal to the direction of wave propagation.
The waves in these directions are independent but have the same speed:

$$v^2 \frac{\partial^2 \psi_y}{\partial x^2} = \frac{\partial^2 \psi_y}{\partial t^2}; \quad v^2 \frac{\partial^2 \psi_z}{\partial x^2} = \frac{\partial^2 \psi_z}{\partial t^2} \; .$$

The amplitude and relative phase of the displacement along these two directions define the *polarisation* of the wave. The two oscillations must be coherent wrt one another to create a polarised wave.

Linear Polarisation

Achieved by oscillating the end of the string along a straight line — each point on will oscillate along a parallel line. Representing the y and z components of the displacement as

$$\begin{aligned} \Psi_y &= A_y \cos(\omega t - kx) \\ \Psi_z &= A_z \cos(\omega t - kx + \phi) \end{aligned}$$

we find that linear polarisation arises where $\phi = 0$ (or integer multiples of π). Clearly a wave polarised along the y (z) axis can be described by $A_z = 0$ ($A_y = 0$). Where both A_y and A_z are non-zero, the wave will be polarised along an intermediate direction, with an amplitude $A = \sqrt{A_y^2 + A_z^2}$ and an angle of polarisation $\theta = \tan^{-1}(A_z/A_y)$ to the y axis. Thus any linearly polarised wave can be resolved into two orthogonal linearly polarised components with the same phase.

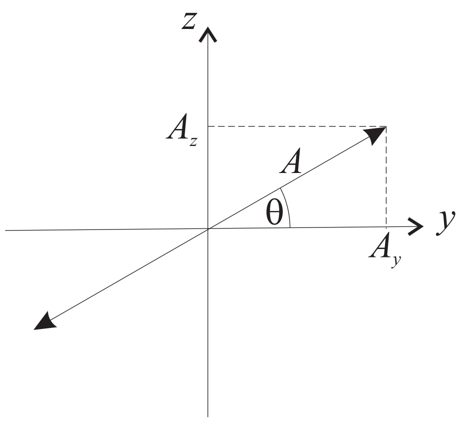


Figure 34: Displacement vector at fixed position for a linearly polarised wave.

Circular Polarisation

If the two components are equal in magnitude, but in phase quadrature, ($\phi = (m + \frac{1}{2})\pi$) we get circular polarisation.
The displacement vector traces a corkscrew.

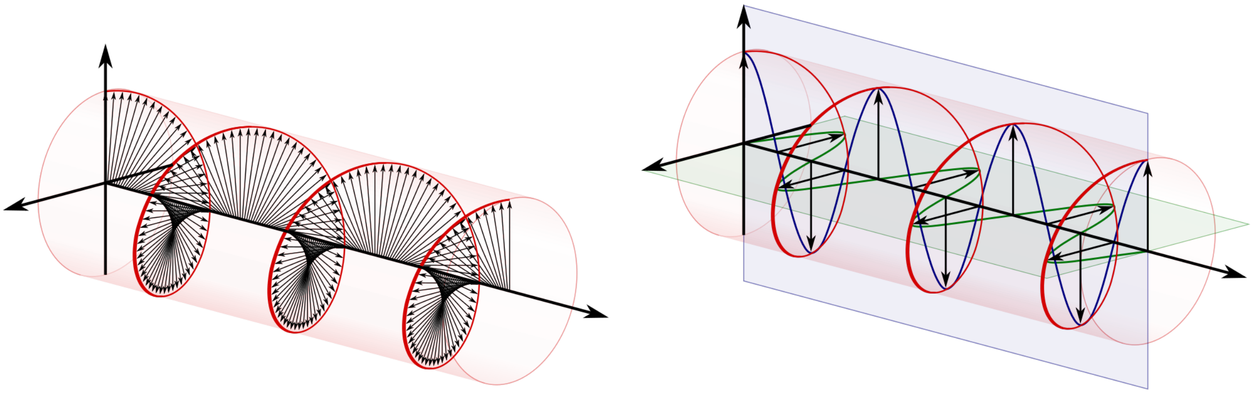


Figure 35: Left-handed circular polarisation

Looking towards the wave source, right-circular polarisation corresponds to clockwise motion of the displacement (electric field) vector. And vice versa for left-handed. This is the radio convention (IEEE is opposite!).

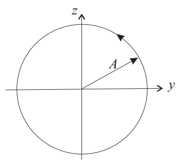


Figure 36: Displacement vector at fixed position for a circularly polarised wave.

Left and Right Circular Polarisation States

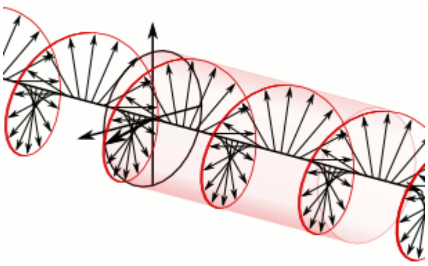


Figure 37: Left-handed circularly polarised wave (radio convention).

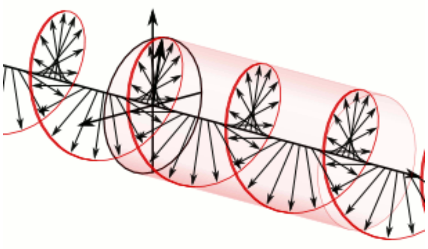


Figure 38: Right-handed circularly polarised wave (radio convention).

Elliptical Polarisation

General case: amplitudes and relative phase of the two components are arbitrary, and the displacement at any position follows an ellipse.

$$\Psi_y = A_y \cos(\omega t - kx)$$

$$\Psi_z = A_z \cos(\omega t - kx + \phi) = A_z \left(\cos(\omega t - kx)\cos\phi - \sin(\omega t - kx)\sin\phi\right)$$

Eliminating $\omega t - kx$ gives

$$\Psi_z = A_z \left(\frac{\Psi_y}{A_y} \cos\phi - \sqrt{1 - \frac{\Psi_y^2}{A_y^2}} \sin\phi \right).$$

We can rearrange this into the standard equation for an ellipse (Hecht, Optics, Ch.8)

$$\frac{\Psi_y^2}{A_y^2} + \frac{\Psi_z^2}{A_z^2} - 2\frac{\Psi_y \Psi_z}{A_y A_z}\cos\phi = \sin^2\phi; \quad \tan 2\alpha = \frac{2A_y A_z \cos\phi}{A_y^2 - A_z^2}$$

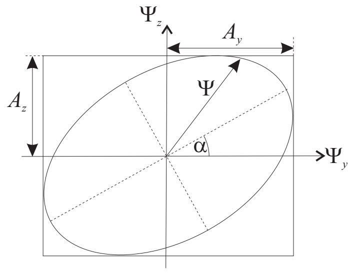


Figure 39: Elliptical Polarisation

Polarisation: some comments

- Circular and linear polarisations are special cases of elliptical polarisation.
- We can regard circular polarisation as the sum of two equal-magnitude, coherent, orthogonal linear polarisations with $\pi/2$ relative phase.
- Equally, we can regard linear polarisation as the sum of equal amplitude, coherent, Left- and Right-handed circularly polarised waves.

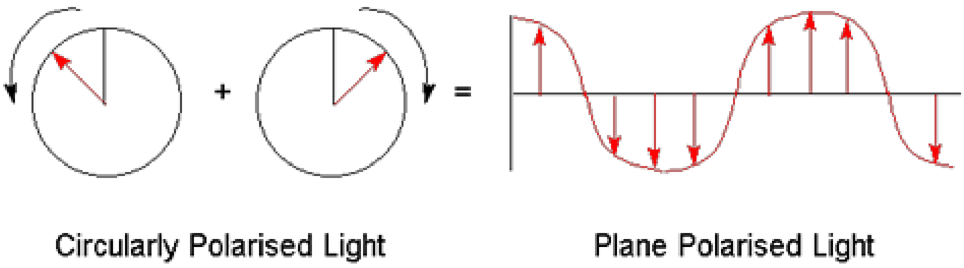


Figure 40: Linear as sum of two circular beams

- You need three numbers to specify the elliptically polarised state: two amplitudes, and an angle.
- Waves are often partially polarised. A single extra value defines the power in this unpolarised component. You can think of unpolarised light arising from two orthogonal, incoherent oscillations with equal mean square magnitudes.

Representation of Polarisation States

It is common to represent a polarised wave by expressing the magnitudes and relative phase of the y and z oscillations (for waves moving in the $+x$ direction) as a 2-component vector (Ψ_y, Ψ_z).

$$\text{Linearly polarised along y}: \begin{pmatrix} A \\ 0 \end{pmatrix} \mathrm{e}^{\mathrm{i}\omega t}$$

$$\text{Linearly polarised along z}: \begin{pmatrix} 0 \\ A \end{pmatrix} \mathrm{e}^{\mathrm{i}\omega t}$$

Usually we drop the $\mathrm{e}^{\mathrm{i}\omega t}$ term.

$$\text{left circular: } \begin{pmatrix} A\mathrm{i} \\ A \end{pmatrix} \quad \text{right circular: } \begin{pmatrix} -A\mathrm{i} \\ A \end{pmatrix}$$

$$\text{elliptical aligned with } y/z \text{ axes: } \begin{pmatrix} A\mathrm{i} \\ B \end{pmatrix}$$

$\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ represents a wave linearly polarised at an angle $\arctan(\beta/\alpha)$ to the y axis.

For example we note that the sum of equal magnitude RCP and LCP waves yields a linearly polarised one.

Wave Impedance

Wave impedance is used to define the relationship between the force and the wave response. It is important in calculating power transfer in a wave, and governs what happens when waves cross boundaries. For our string,

$$Z = \text{Impedance} = \frac{\text{driving force}}{\text{velocity response}}$$

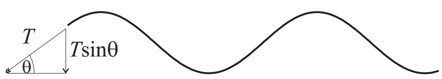


Figure 41: Transverse force in a wave.

Note this velocity is the transverse response $\dot{\psi}$, not the wave speed ω/k. Consider a free end in the string. The transverse driving force F is $F = -T\sin\theta \approx -T\frac{\partial\Psi}{\partial x}$. Hence

$$Z = \frac{\text{transverse driving force}}{\text{transverse velocity}} = \frac{-T(\partial\Psi/\partial x)}{(\partial\Psi/\partial t)}$$

But for a waveform $\Psi = f(u) = f(x - vt)$ moving in direction $+x$, we have

$$\frac{\partial\Psi}{\partial x} = \frac{\mathrm{d}f}{\mathrm{d}u} \quad \text{and} \quad \frac{\partial\Psi}{\partial t} = -v\frac{\mathrm{d}f}{\mathrm{d}u}$$

Hence

$$Z = \frac{T}{v} = \sqrt{T\rho} = \rho v. \tag{39}$$

...and Z is negative for wave travelling to $-x$.

Power in String Wave

Energy is present in a string wave in the form of kinetic and potential energy. The power input into a wave is given by

$$\text{power input} = \text{transverse force} \times \text{transverse velocity} = F\,u.$$

where u is the transverse velocity. The mean power input is $\langle P \rangle = \frac{1}{2}\Re[\mathbf{F}\mathbf{u}^*]$ (see Eq. 21). Using $\mathbf{F} = Z\mathbf{u}$, we get

$$\langle P \rangle = \frac{1}{2}\Re[\mathbf{F}\mathbf{u}^*] = \frac{1}{2}\Re[Z\mathbf{u}\mathbf{u}^*] = \frac{1}{2}\Re[Z]\,|\mathbf{u}|^2\,.$$

Assuming Z is real and that $\mathbf{u} = \dot{\mathbf{\Psi}} = i\omega A_0 \mathrm{e}^{\mathrm{i}(\omega t - kx)}$, then

$$\boxed{\text{mean power} = \frac{1}{2} Z\omega^2 A_0^2.} \tag{40}$$

This must also equal the energy per unit length of string times the wave speed. The kinetic energy and potential energy per unit length are

$$\text{KE} = \frac{1}{2}\rho\left(\frac{\partial\Psi}{\partial t}\right)^2; \quad \text{PE} = \frac{1}{2}T\left(\frac{\partial\Psi}{\partial x}\right)^2$$

These terms are equal because $\rho\omega^2 = Tk^2$, so adding and averaging over a complete number of wavelengths yields an average energy per unit length of $\frac{1}{2}\rho\omega^2 A_0^2$, hence the power flow is

$$\text{mean power} = \frac{1}{2}\rho\omega^2 A_0^2 \times v = \frac{1}{2} Z\omega^2 A_0^2$$

because $\rho v = Z$. (i.e. the same result by different method).

Wave Reflection at Boundaries 1

Let us use harmonic waves on a string to derive what will be general results. If two strings of differing densities are connected, a wave will be (partially) reflected at the boundary. Let us derive the steady-state solution for harmonic waves — so that each piece of string contains infinitely long sinusoidal waves, one forward and one backward (in general). The frequency of the wave will be unchanged. But the wavespeed, and therefore wavelength, will change.

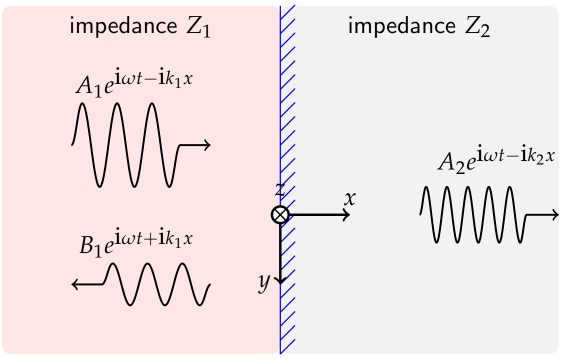


Figure 42: Reflection geometry

The boundary conditions at $x = 0$ are

- continuity of displacement Ψ
- continuity of transverse force on string $(-T\partial\Psi/\partial x)$.

Wave Reflection at Boundaries 2

Continuity of Ψ at $x = 0$:

$$A_1 e^{i\omega t} + B_1 e^{i\omega t} = A_2 e^{i\omega t}$$

$$\boxed{A_1 + B_1 = A_2.} \tag{41}$$

Continuity of transverse force $-T\Psi'$:

$$T(-ik_1)A_1 + T(ik_1)B_1 = T(-ik_2)A_2.$$

$$k_1 = \omega/v_1 \quad \text{and} \quad k_2 = \omega/v_2$$

$$\therefore -\frac{T}{v_1}\omega A_1 + \frac{T}{v_1}\omega B_1 = -\frac{T}{v_2}\omega A_2.$$

Use

$$T/v_1 = Z_1 \quad \text{and} \quad T/v_2 = Z_2$$

Hence

$$\boxed{Z_1(A_1 - B_1) = Z_2 A_2.} \tag{42}$$

We can now solve equ. (42) and equ. (41) for the amplitude reflection and transmission coefficients r and τ (the latter not to be confused with time t).

Wave Reflection at Boundaries 3

We obtain:

$$\boxed{r = \frac{B_1}{A_1} = \frac{Z_1 - Z_2}{Z_1 + Z_2}} \tag{43}$$

$$\boxed{\tau = \frac{A_2}{A_1} = 1 + r = \frac{2Z_1}{Z_1 + Z_2}} \tag{44}$$

- If $Z_2 = \infty$ (clamped string): $r = -1$ and $\tau = 0$. Complete anti-phase reflection — phase change of π on reflection.

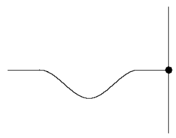


Figure 43:

- If $Z_2 = 0$, i.e. the second string is massless (same as a free end if tension could be maintained): $r = 1$, $\tau = 2$ — complete in-phase reflection. No energy transmitted.

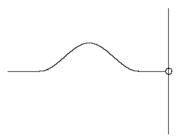


Figure 44:

- $Z_2 = Z_1$: $r = 0$ and $\tau = 1$. Strings are matched.

Notes on Reflection and Transmission

- Results are general and apply to a wide variety of wave systems.
- The General Recipe: at interfaces need to consider (i) continuity of "displacement", Ψ and (ii) continuity of "force" $\propto Z\Psi'$ for harmonic waves. Remember to use negative Z for negative-going waves.
- Depending on the precise problem, we may end up with Z_1 and Z_2 interchanged...e.g. in an electromagnetic wave, E and $H = E/Z$ are continuous at the boundary. And on a transmission line, we have continuity of voltage V and current $I = V/Z$ (see the electromagnetism course).
- Remember that the impedance is in general complex, so we can get arbitrary phase differences between the incident, reflected and transmitted waves.
- We have only found the steady-state response.

Reflection and transmission of energy

If Z is real, the power transmitted by a harmonic wave is $P = \frac{1}{2}Z\omega^2A^2$ (from equ. 40). So the incident energy flow is $\frac{1}{2}Z_1\omega^2A_1^2$, with $\frac{1}{2}Z_1\omega^2B_1^2$ reflected and $\frac{1}{2}Z_2\omega^2A_2^2$ transmitted. The Power Reflection Coefficient R is thus

$$R = \frac{Z_1B_1^2}{Z_1A_1^2} = \left(\frac{Z_1 - Z_2}{Z_1 + Z_2}\right)^2 = R$$

and the Power Transmission Coefficient T is

$$T = \frac{Z_2A_2^2}{Z_1A_1^2} = \frac{4Z_1Z_2}{(Z_1 + Z_2)^2} = T.$$

Energy is conserved:

$$R + T = 1$$

If Z_1 and/or Z_2 are complex, recall the average power input is $\frac{1}{2}\Re(Z)\omega^2|A|^2$, so

$$R = \frac{\frac{1}{2}\Re(Z_1)\omega^2|B_1|^2}{\frac{1}{2}\Re(Z_1)\omega^2|A_1|^2} = \frac{B_1}{A_1}\frac{B_1^*}{A_1^*} = rr^* = \left|\frac{Z_1 - Z_2}{Z_1 + Z_2}\right|^2.$$

Impedance Matching 1

Often useful to match two media to minimise reflections – many applications in optics, acoustics, electrical transmission... Consider a wave incident on a layer of thickness l, impedance Z_2, filling $0 \le x \le l$, on top of an infinite substrate Z_3:

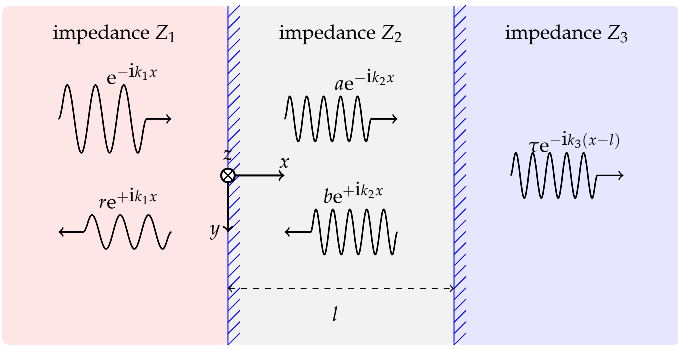


Figure 45: Reflection from an intermediate layer

To simplify algebra: drop the implicit $\exp(i\omega t)$; set the incident amplitude to unity (so that the reflected amplitude is r); also added a phase shift $\exp(ik_3l)$ to the transmission coefficient. Note the three wavevectors k_1, k_2 and k_3. It is useful to define

$$\gamma = e^{-ik_2l}$$

There are four unknowns: r, τ, a, b in the steady-state solution. We apply four boundary conditions to find them.

Impedance Matching 2

Ψ continuity at $x = 0$:

$$\boxed{1 + r = a + b}$$

$-T\Psi'$ continuity at $x = 0$:

$$\boxed{Z_1(1 - r) = Z_2(a - b)}$$

Ψ continuity at $x = l$ with $\gamma = e^{-ik_2l}$:

$$\boxed{\gamma a + \frac{b}{\gamma} = \tau}$$

$-T\Psi'$ continuity at $x = l$:

$$\boxed{Z_2\left(\gamma a - \frac{b}{\gamma}\right) = Z_3\tau}$$

Solve these 4 equations for (a, b, r, τ) — generally messy. But if we choose a quarter wavelength of material in the layer, $k_2l = (m + \frac{1}{2})\pi$ for integer m, then $\gamma = e^{-i\pi/2} = -i$ and they simplify to:

$$1 + r = a + b; \quad Z_1(1 - r) = Z_2(a - b)$$

$$a - b = \tau i; \quad Z_2(a + b) = Z_3\tau i$$

$$\therefore r = \frac{Z_1 - \frac{Z_2^2}{Z_3}}{Z_1 + \frac{Z_2^2}{Z_3}} = \frac{Z_1 - Z_{\text{eff}}}{Z_1 + Z_{\text{eff}}}$$

where $Z_{\text{eff}} = Z_2^2/Z_3$. So a quarter-wave layer of Z_2 material on a deep substrate Z_3 has an effective impedance of Z_2^2/Z_3. Hence by choosing l and Z_2 we can match two media and stop reflections. We use a quarter wave $\lambda_2/4$ section of impedance $Z_2 = \sqrt{Z_1Z_3}$.

Impedance Matching 3

- Quarter-wave matching works because the waves reflected at the first and second boundary are out of phase if $l = \lambda_2/4$.
- We have again solved the steady-state problem — a transient will occur as wave hits boundary.
- Common application: anti-reflection lens- and screen coatings

Figure 46: Quarter-Wavelength coated lens, smartphone screen glass

Need to use result that $Z = Z_0/n$. If we have a quarter wave coating of Magnesium Fluoride ($n = 1.38$) on crown glass ($n = 1.52$), it has effective refractive index $n_e = 1.38^2/1.52 = 1.25$, yielding a power reflection coefficient of 1% (compared to 4% without).

- Can also match impedances with a gradual impedance change.

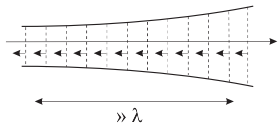


Figure 47: Gradual impedance change.

Longitudinal Waves

- Longitudinal Waves have displacement of the medium parallel to the wavevector: sound waves are the most common example
- Corresponds to a wave of rarefaction and compression
- No polarisation possible
- In a sound wave, the pressure wave and displacement wave are $\pi/2$ out of phase: maximum pressure occurs at minimum displacement.

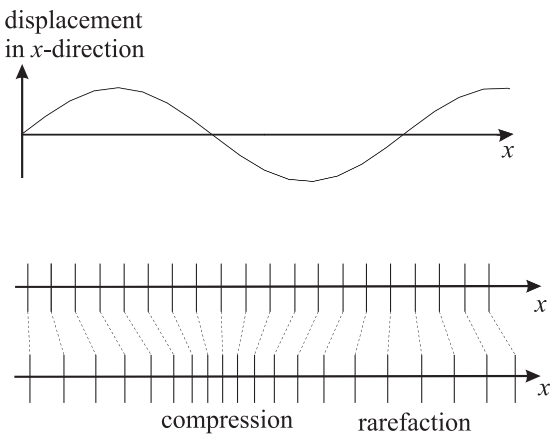


Figure 48: Displacement of gas in a longitudinal wave.

Compression of an Ideal Gas

How does a gas react to compression/rarefaction? Two limits:

- Isothermal process: heat Q flows rapidly out of/into gas to keep temperature constant. The gas law applies: $pV = nRT =$ constant.
- Adiabatic process: the compression/rarefaction is so fast that heat cannot flow into/out of the gas before the next phase of the pressure wave passes through. The work done on the gas is stored as extra internal energy U:

$$-p\,\mathrm{d}V = \mathrm{d}U$$

A perfect gas has an internal energy $U = nC_vT$, where C_v is its molar specific heat at constant volume.

$$\mathrm{d}U = nC_v\,\mathrm{d}T = -p\,\mathrm{d}V.$$

Using $pV = nRT$, we obtain

$$nC_v\,\mathrm{d}T + \frac{nRT}{V}\,\mathrm{d}V = 0; \quad \therefore \frac{dT}{T} + \frac{R}{C_v}\frac{\mathrm{d}V}{V} = 0$$

$$\therefore T\,V^{\frac{R}{C_v}} = \text{constant}$$

$$\boxed{\therefore p\,V^{\frac{C_v+R}{C_v}} = pV^{\gamma} = \text{constant}}$$

where the heat capacity at constant pressure $C_p = C_v + R$, and $\gamma = C_p/C_v$. For diatomic molecules at room temperature (air) $\gamma \approx 7/5$. For atomic gas, $\gamma \approx 5/3$.

We will see that sound waves are usually adiabatic.

Wave Equation for Sound in a Gas: the Strain

Consider dynamics of a parcel of gas lying (when undisturbed) between x and $x + \Delta x$. It has cross section ΔS and moves in the x direction only.

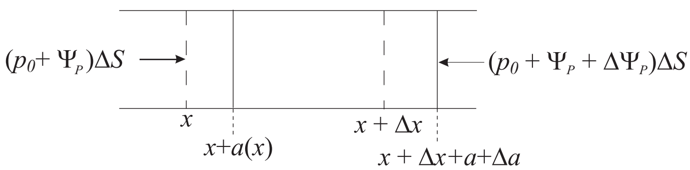


Figure 49:

The wave displaces gas originally at x to $x + a(x,t)$, and creates an extra pressure $\Psi_p(x,t)$, so the total gas pressure locally is $p = p_0 + \Psi_p$, where p_0 is the atmospheric pressure. When the wave is present, the parcel extends from $x + a(x)$ to $x + \Delta x + a(x) + \Delta a = x + \Delta x + a(x) + \frac{\partial a}{\partial x}\Delta x$. The volume change of the element is given by

$$\Delta S \Delta a = \Delta S \frac{\partial a}{\partial x}\Delta x.$$

The fractional change in volume (the volume strain) is

$$\frac{\Delta V}{V} = \frac{\Delta S\,\frac{\partial a}{\partial x}\,\Delta x}{\Delta S \Delta x} = \frac{\partial a}{\partial x} \tag{45}$$

Wave Equation for Sound in a Gas: the Stress

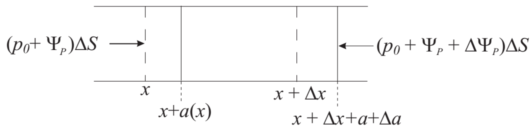


Figure 50:

In a gas, the stress is equal to the pressure, and is isotropic. The force imbalance on the ends of the parcel is

$$F_{\text{net}} = (p + \Psi_p)\Delta S - \left(p + \Psi_p + \frac{\partial \Psi_p}{\partial x}\Delta x\right)\Delta S = -\frac{\partial \Psi_p}{\partial x}\Delta x \Delta S$$

i.e. the pressure gradient times the volume.
This force accelerates the parcel of gas:

$$-\frac{\partial \Psi_p}{\partial x} = \rho \frac{\partial^2 a}{\partial t^2} \tag{46}$$

Wave Equation for Sound in a Gas

We now need to relate the stress to the strain. Assume the waves are adiabatic, so that heat cannot flow quickly in/out of the compression regions: pV^γ is constant. (We can check this by experiment.) Differentiating:

$$\mathrm{d}p\, V^\gamma + p\gamma V^{\gamma-1}\mathrm{d}V = 0$$

$$\therefore\ \mathrm{d}p = \Psi_p = -\gamma p \frac{\Delta V}{V} = -\gamma p \frac{\partial a}{\partial x}. \tag{47}$$

using equ. (45). Note that $\mathrm{d}p$ is the pressure change associated with the wave: $\mathrm{d}p \equiv \Psi_p$.
We need the spatial derivative of Ψ_p to form the equation of motion:

$$\frac{\partial \Psi_p}{\partial x} = -\gamma p \frac{\partial^2 a}{\partial x^2} - \gamma \frac{\partial p}{\partial x}\frac{\partial a}{\partial x}.$$

The ratio of the 2nd term to the first term on the RHS is $\sim \Psi_p/p \sim \gamma k a \sim a/\lambda$ and is negligible in most practical situations i.e. we usually have $a \ll \lambda$. Using equ. (46), we obtain

$$\gamma p \frac{\partial^2 a}{\partial x^2}\Delta x \Delta S = \rho \Delta x \Delta S \frac{\partial^2 a}{\partial t^2}.$$

Wave Equation for Sound in a Gas

- Finally, we have the wave equation for (small amplitude) sound waves:

$$\frac{\partial^2 a}{\partial x^2} = \frac{\rho}{\gamma p} \frac{\partial^2 a}{\partial t^2} \tag{48}$$

Note the speed is independent of wavelength/frequency i.e. there is no dispersion.

- This is the wave equation, with wave speed v given by

$$v = \sqrt{\frac{\gamma p}{\rho}} = \sqrt{\frac{\gamma nRTV}{Vm}} = \sqrt{\frac{\gamma RT}{M}} \tag{49}$$

where M is the molar mass. Note the dependences on temperature, mass and γ.

- The mean squared speed of molecules of mass m in an ideal gas is given by the equipartition theorem:

$$\frac{1}{2} m \langle v^2 \rangle = \frac{3}{2} k_B T; \quad \therefore v_{\rm rms} = \sqrt{\langle v^2 \rangle >} = \sqrt{\frac{3RT}{M}}.$$

Hence the r.m.s. molecular speeds will be close to, but slightly larger than the wave speed. Remember, though, that the passage of a sound wave does not require the net movement of gas molecules from one point to another.

Sound speed: examples

- Speed of sound in dry air at 20 °C and atmospheric pressure (101.325 kPa) is

$$v = \sqrt{\frac{1.4 \times 101.325 \times 10^3}{1.204}} = 343\,\mathrm{m\,s^{-1}}$$

Accurately confirmed by experiment — so adiabatic assumption is good.
(γ is effectively 1 for isothermal change, so speed difference would be about 20%).

- In fact, Newton was the first to estimate the speed of sound; he assumed isothermal changes. His estimate was $\sim 15\%$ too low because of this.
- Speed rises by about 0.6 $\mathrm{m\,s^{-1}}$ per °C of temperature rise.
- Speed of sound in Helium at same pressure and temperature is

$$v = \sqrt{\frac{1.667 \times 8.314 \times 293}{0.004}} = 1004\,\mathrm{m\,s^{-1}}$$

- Solar corona: $\gamma = 5/3$, ionised hydrogen ($m = 1$amu), $T = 1 \times 10^6$ K, and $v = 1.18 \times 10^5\,\mathrm{m\,s^{-1}}$
- In cold interstellar gas clouds: $T \approx 8$ K, and $v = 75\,\mathrm{m\,s^{-1}}$

The Displacement Wave and Impedance

Our wave equation was expressed in terms of the molecular displacements a but this is very hard to measure. We normally measure the pressure fluctuation with e.g. a microphone. We relate the two via

$$\Psi_p = -\gamma p \frac{\partial a}{\partial x}.$$

If we assume a harmonic wave traveling to $+x$ of the form

$$a = a_0 \mathrm{e}^{(\mathrm{i}\omega t - \mathrm{i}kx)}$$

then

$$\Psi_p = \mathrm{i}\gamma pka$$

i.e. the pressure leads the displacement by $\pi/2$ and has amplitude γpka_0. The characteristic impedance Z can be defined as the driving force divided by response speed, in the usual way. For this harmonic wave we have

$$Z = \frac{\text{force}}{\text{velocity}} = \frac{\Delta S \Psi_p}{\dot{a}} = \frac{\Delta S \mathrm{i}\gamma pka}{i\omega a} = \Delta S \frac{\gamma p}{v} = \Delta S \sqrt{\gamma p \rho}$$

Often, we use the impedance per unit area $\mathcal{L}$ (often just called the acoustic impedance):

$$\mathcal{L} = \sqrt{\gamma p \rho} = v\rho = \frac{\gamma p}{v}$$

In dry air at 20 °C, 101.325 kPa: $\mathcal{L} \approx 413\,\mathrm{kg\,m^{-2}\,s^{-1}}$.

Power in an acoustic wave

The power transferred by the wave is force times velocity in the usual way. For a harmonic sound wave, the particle speed and pressure are

$$a(x,t) = a_0 \exp(\mathrm{i}\omega t - \mathrm{i}kx)$$

$$\dot{a}(x,t) = \mathrm{i}\omega a_0 \exp(\mathrm{i}\omega t - \mathrm{i}kx)$$

$$\Psi_p = A_0 \exp(\mathrm{i}\omega t - \mathrm{i}kx) = \mathrm{i}\gamma pka_0 \exp(\mathrm{i}\omega t - \mathrm{i}kx)$$

where $A_0 = i\gamma pka_0$ is the pressure amplitude of the wave. Note that the speed and force are in phase.

The mean power per unit area, or intensity I of the wave, is thus

$$I = \frac{1}{2}\Re(\Psi_p \dot{a}^*) = \frac{1}{2}\gamma pk\omega a_0^2 = \frac{|A|^2}{2\mathcal{L}} = \frac{A_{\mathrm{rms}}^2}{\mathcal{L}}$$

(c.f. usual electrical expression for mean power V_{rms}^2/Z). Note the intensity is independent of wavelength for a given pressure amplitude.

An alternate formulation for the intensity is

$$I = \frac{1}{2}\mathcal{L}\omega^2 |a_0|^2$$

The Decibel Scale for Sound

We can specify the sound pressure level, SPL of a sound in decibels with respect to some reference level; remember the decibel scale is a relative scale only.

$$\text{SPL in decibels} = 10\log_{10}\left(\frac{p_{\text{rms}}^2}{p_{\text{ref}}^2}\right) = 20\log_{10}\left(\frac{p_{\text{rms}}}{p_{\text{ref}}}\right) \tag{50}$$

We usually use a reference level of $p_{\text{ref}} = 20\,\mu\text{Pa}$: note this is a root-mean-square (rms) value. p_{ref} is approximately the threshold of human hearing at 1 kHz In a young adult. This is only 2×10^{-10} atmospheres!.
Equivalently, we can express the sound as a power level in decibels:

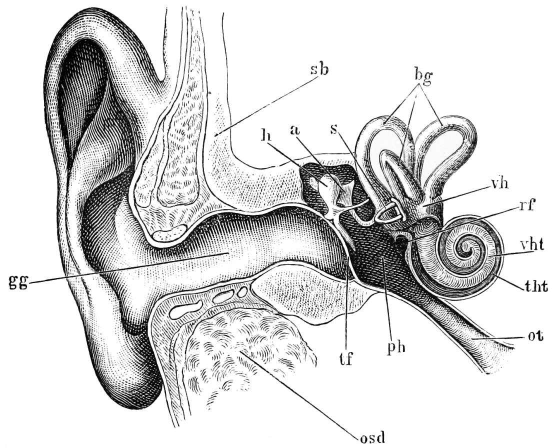
$$\text{dBA level} = 10\log_{10}\left(\frac{I}{I_{\text{ref}}}\right) \tag{51}$$

with the reference intensity

$$I_{\text{ref}} = p_{\text{ref}}^2/\mathcal{L} = 1 \times 10^{-12}\,\text{W m}^{-2}$$

where we used for air $\mathcal{L} = 400\,\text{kg m}^{-2}\,\text{s}^{-1}$.

Human Hearing



The human ear can respond to pressure wave amplitudes from $1\,\text{pW m}^{-2}$ to $1\,\text{W m}^{-2}$, and frequencies from 30 Hz to 20 kHz.

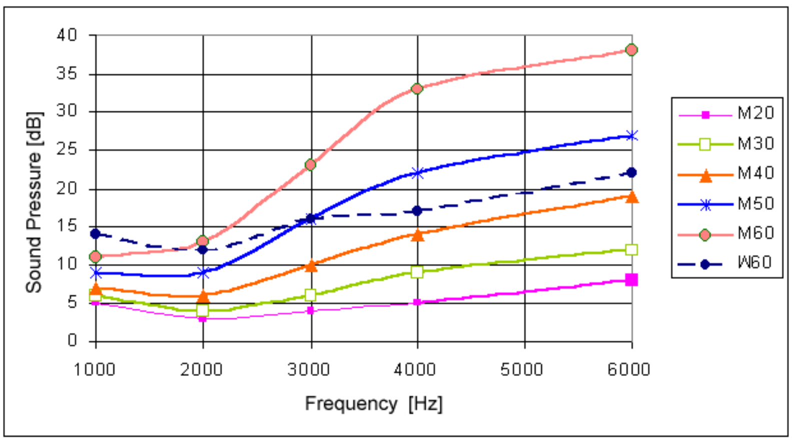


Figure 51: Human Hearing Response with Age

Typical Sound Characteristics

- Middle C has a frequency (in some music systems) of 256 Hz, so has a wavelength of 1.34 m in air at 20 °C.
- How powerful is the human voice? 60dB at 1 m seems a reasonable average. Now 60dB (relative to 1 pW m^{-2}) corresponds to a flux of 1 μW m^{-2}, so a total power of $4\pi R^2$ times the flux, yielding $\sim$ 10 μW
- What is the pressure level you are nearing now in the lecture theatre? What is the molecular displacement?
- The Sound Pressure Level for 60dB is $1000 p_{\text{ref}} =$ 20 mPa i.e. only 2×10^{-7} of atmsospheric pressure.
- The displacement comes from
$$\Psi_p = -\gamma p \frac{\partial a}{\partial x}$$
so $|a| = |\Psi_p|/(\gamma p k) \sim$ 30 nm i.e. 300 atomic diameters.
- Barely audible sounds, 10dB say, have molecular amplitudes of order 1×10^{-10} m!

Sound waves in solids and liquids 1

Longitudinal waves also exist in liquids and solids. The properties of the medium are described by a generalisation of equ. (47), the relationship between the pressure due to the wave and the strain in the medium

$$\Psi_p = -K \frac{\partial a}{\partial x}$$

where K is the relevant elastic modulus. This gives the general result

$$v = \sqrt{\frac{K}{\rho}}$$

where ρ is the density.
For gases and liquids, the pressure is isotropic, but the expansion and rarefaction takes place only in the direction of passage of the wave. Hence the fractional volume change is directly proportional to the fractional change in length of the element ($\Delta V/V = \Delta a/\Delta x = \partial a/\partial x$). The relevant modulus is the bulk modulus B:

$$\mathrm{d}p = -B \frac{\mathrm{d}V}{V}$$

where $B = \gamma p$ for a gas undergoing adiabatic changes (and $B = p$ for isothermal ones). Solids are more complicated as they can support shear stresses. Shear waves travel more slowly than compression waves and are important in earthquakes.

Sound waves in solids and liquids 2

When a thin bar is compressed longitudinally it can expand in the transverse direction (the ratio of the two strains is known as Poisson's ratio). However, in a bulk solid, the medium cannot expand sideways, hence more longitudinal pressure is required to produce a longitudinal strain, and the modulus in therefore larger. (See Physics B: Dynamics). We can ignore these effects if the solid is a thin bar and we look at compression waves travelling along the bar. The longitudinal stress τ (force/area) in the solid is then related to the strain $\frac{\partial a}{\partial x}$ by

$$\tau = Y\frac{\partial a}{\partial x}$$

where Y is Young's modulus for the material.
The equation of motion then becomes

$$Y\frac{\partial^2 a}{\partial x^2} = \rho\frac{\partial^2 a}{\partial t^2}$$

with a wave speed

$$v = \sqrt{\frac{Y}{\rho}}$$

where ρ is the density.

Longitudinal Waves Usually Travel Faster than Transverse Waves

Figure 52:

On a stretched string, transverse waves have

$$v_T = \sqrt{\frac{T}{\mu}} = \sqrt{\frac{220\,\mathrm{N}}{0.6\,\mathrm{g\,m^{-1}}}} \approx 600\,\mathrm{m\,s^{-1}}$$

whereas longitudinal waves have

$$v_L = \sqrt{\frac{Y}{\rho}} = \sqrt{\frac{200\,\mathrm{GPa}}{7700\,\mathrm{kg\,m^{-3}}}} \approx 5000\,\mathrm{m\,s^{-1}}$$

The fundamental mode oscillations both have a wavelength of about 1m for a violin (twice the violin length), so the frequencies are about 600 Hz and 5 kHz.

Note also the acoustic impedance $\mathcal{L} = Y/v = \sqrt{Y\rho}$ for waves in a solid will be much larger than in a gas typically. For steel, $\rho = 7700\,\mathrm{kg\,m^{-3}}$, so $\mathcal{L} \approx 4 \times 10^7\,\mathrm{kg\,m^{-2}\,s^{-1}}$, which is 10^5 time larger than for air. So an open air-steel interface is essentially a free end with a pressure node, and displacement antinode occurring there.

Standing Waves in One Dimension

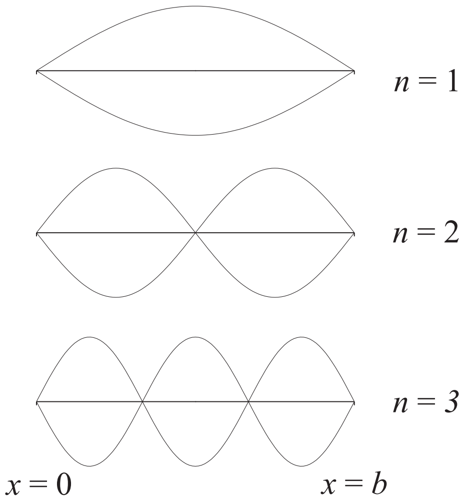


Figure 53: Standing waves on a string.

- A standing wave is a wave confined within a region of space by applying boundary conditions, e.g. a wave on a violin string
- The standing wave can be found as a superposition of forward and backward traveling waves. Simplest case to analyse is where perfect reflection occurs at the boundary.
- On a string of length b, with boundary conditions are $\Psi(0,t) = \Psi(b,t) = 0$:

$$\begin{aligned}\Psi &= A\cos(\omega t - kx) - A\cos(\omega t + kx)\\ &= 2A\sin\omega t\sin kx.\end{aligned}$$

The boundary conditions are satisfied if $kb = n\pi$, where $n = 1, 2, 3\ldots$, i.e. $k = \frac{n\pi}{b}$. For a wave speed v, the allowed frequencies are given by

$$\omega = \frac{nv\pi}{b}.$$

Standing Waves in 2 and 3 Dimensions

Easiest to look for solutions to the wave equation

$$c^2\boldsymbol{\nabla}^2\Psi = \frac{\partial^2\Psi}{\partial t^2}$$

which satisfy the boundary conditions. For waves confined in a 3-d box, for example, the wavefunctions are given by

$$\Psi = A\sin k_x x\sin k_y y\sin k_z z\cos\omega t.$$

For a box with a corner at the origin, of dimensions a, b, c, the boundary conditions are satisfied by a wavevector

$$\boldsymbol{k} = \left(\frac{l\pi}{a}, \frac{m\pi}{b}, \frac{n\pi}{c}\right)$$

where $l,\ m,\ n = 1, 2, 3\ldots$. In two dimensions, this gives the patterns of oscillation of a rectangular drum.

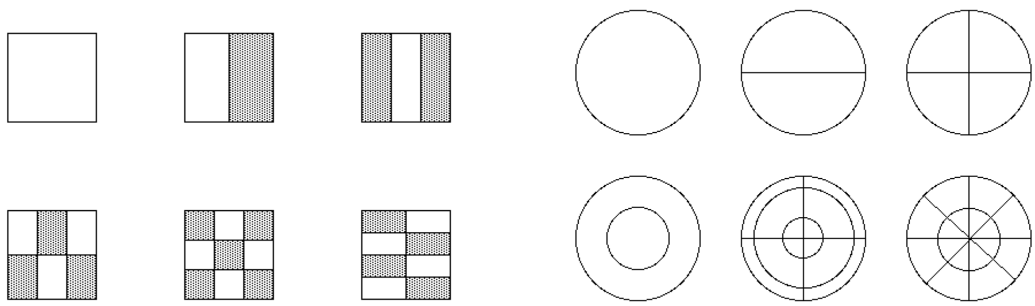


Figure 54: Standing Waves on a rectangular and a round drum

Damped Waves

Like oscillators, real waves usually lose energy through damping. A string wave will lose energy via air resistance: there will be an extra transverse force on an element $\mathrm{d}x$ of string. Assume it is proportional to the transverse speed, so has form $-\beta\dot{\psi}\mathrm{d}x$. The revised wave equation is now

$$\frac{\partial^2\Psi}{\partial t^2} = v^2\frac{\partial^2\Psi}{\partial x^2} - \frac{\beta}{\rho}\frac{\partial\Psi}{\partial t} \tag{52}$$

which we can rewrite as

$$\boxed{\frac{\partial^2\Psi}{\partial t^2} + \Gamma\frac{\partial\Psi}{\partial t} = v^2\frac{\partial^2\Psi}{\partial x^2}} \tag{53}$$

where $\Gamma = \beta/\rho$ expresses the damping. To find solutions, try a harmonic wave $\psi = D\exp(\mathrm{i}\omega t - \mathrm{i}kx)$: we find

$$\omega^2 - \mathrm{i}\Gamma\omega = v^2k^2 \tag{54}$$

so that k must be complex in general (ω is real). Thus we must have exponential and oscillating parts in the solution for Ψ. Splitting k into real and imaginary parts: $k = k_r - \mathrm{i}k_i$, we find that

$$k_r^2 - k_i^2 = \frac{\omega^2}{v^2}; \quad 2k_rk_i = \frac{\Gamma\omega}{v^2} \tag{55}$$

Lightly Damped Waves

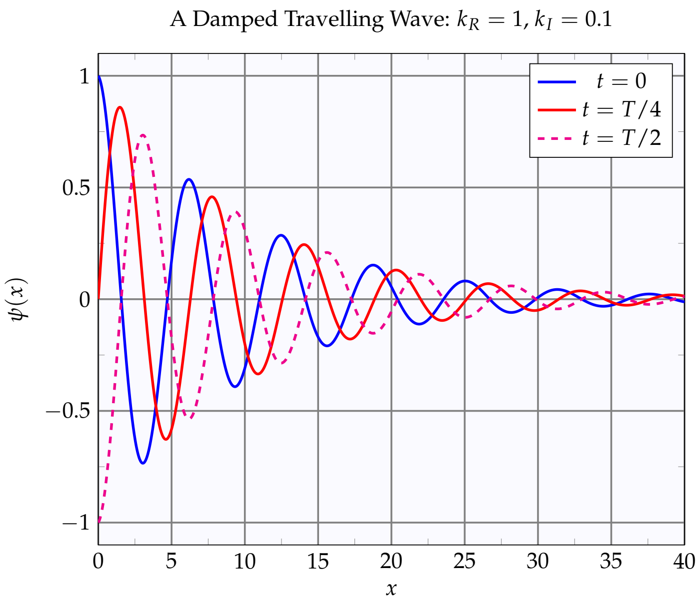


Figure 55: A lightly-damped wave

If we assume light damping ($\Gamma \ll \omega$), we can see from equ. (55) that $k_r \approx \omega/v$, and thus $k_i \approx \Gamma/(2v)$: so the solution is

$$\Psi = \exp(-k_ix)\Re[D\exp(\mathrm{i}\omega t - \mathrm{i}k_rx)]$$

which is a decaying traveling wave. The imaginary component of k sets the damping length; the real part sets the wavelength and thus speed.
Note the damping length is independent of the wavelength.

Heavy Damping of String Waves

If we now assume heavy damping ($\Gamma \gg \omega$), and again write $k \equiv k_r - \mathrm{i}k_i$, we can see from equ. (54) that

$$k' = k_r \approx k_i \approx \pm \left| \left(\frac{\Gamma\omega}{2v^2} \right)^{1/2} \right|$$

i.e. the wavevector k has equal real and imaginary parts.
The wave now decays over a short distance, and the decay length varies as $\omega^{-0.5}$.

Heavily Damped Waves

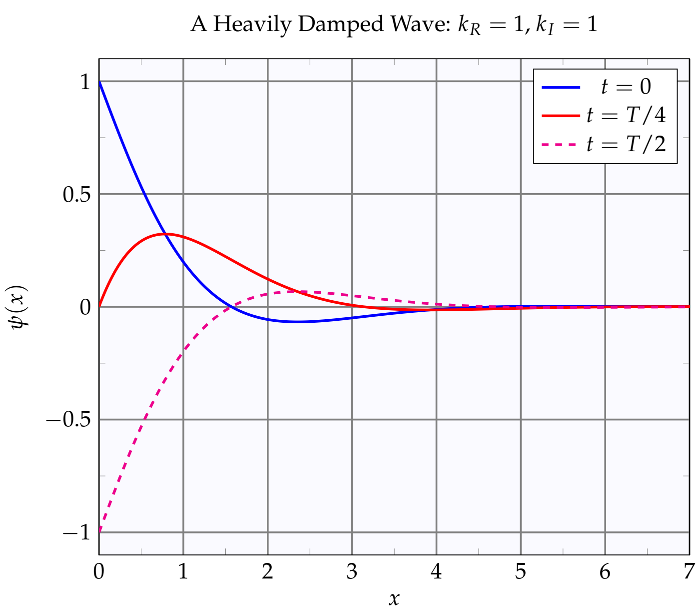


Figure 56: A heavily-damped wave

$$\Psi = \exp(-k'x)\Re[D\exp(\mathrm{i}\omega t - \mathrm{i}k'x)]$$

$$k' = k_r \approx k_i \approx \pm \left| \left(\frac{\Gamma\omega}{2v^2} \right)^{1/2} \right|$$

The wave is rapidly attenuated. (You will see such waves when an electromagnetic wave tries to enter a good conductor: the skin effect).

Impedance and reflection from damped wave system

The characteristic impedance of the damped string waves is force/transverse speed in the usual way:

$$Z = \frac{-T\Psi'}{\dot{\Psi}} = \frac{Tk}{\omega} = \frac{T}{\omega}\left(k_r - \mathrm{i}k_i\right)$$

For light damping, $\Gamma \ll \omega$:

$$Z(\omega) = \sqrt{T\rho}\left(1 - \frac{\mathrm{i}\Gamma}{2\omega}\right) = Z_0\left(1 - \frac{\mathrm{i}\Gamma}{2\omega}\right)$$

where Z_0 is the impedance if the damping is zero. For heavy damping $\Gamma \gg \omega$:

$$Z(\omega) \approx Z_0\left(1 - \mathrm{i}\right)\sqrt{\frac{\Gamma}{2\omega}}$$

Reflection

Consider now an undamped wave system (perfect string, impedance Z_0) meeting a damped one (string with lateral resistance, impedance Z).
We will get reflections at the boundary, and can use the usual relation for the reflection coefficient:

$$r = \frac{Z_1 - Z_2}{Z_1 + Z_2}$$

- Light Damping, $\Gamma \ll \omega$: we obtain

  $$r(\omega) = \frac{\mathrm{i}\Gamma}{4\omega}$$

  so $|r|$ is small, and has phase $\pi/2$: most power is transmitted, little is reflected.
- Heavy Damping ($\Gamma \gg \omega$): we obtain

  $$r(\omega) = \frac{1 - (1-\mathrm{i})\sqrt{\frac{\Gamma}{2\omega}}}{1 + (1-\mathrm{i})\sqrt{\frac{\Gamma}{2\omega}}} \approx -1$$

  so the wave is almost all reflected in anti-phase. (You will see in Electromagnetism that this is a good model for reflection of light from a good conductor.)

Phase Speed of Damped String Waves

In the lightly damped case, we saw that the wave propagates with an exponentially decaying amplitude:

$$\Psi = \exp(-k_i x)\Re[D\exp(\mathrm{i}\omega t - \mathrm{i}k_r x)]$$

The phase of the propagating wave is $(\omega t - k_r x)$ and the phase speed v_ϕ is therefore

$$v_\phi = \frac{\omega}{k_r}$$

In equ. (55), we can eliminate k_i and express ω in terms of k_r: we find

$$v_\phi = \frac{\omega}{k_r} = v\left(1 + \frac{\Gamma^2}{4v^2k_r^2}\right)^{-1/2}$$

Hence the speed of the waves depends on their wavelength. This is an example of wave dispersion.

The dispersion relation is the relationship between ω and the wavevector, and is controls many aspects of wave propagation. For a non-dispersive system, we simply have $\omega = vk$ with speed v being a constant.

In this case, we see that the phase speed increases as the frequency/wavevector increases; and it approaches the undamped speed v in the high frequency limit.

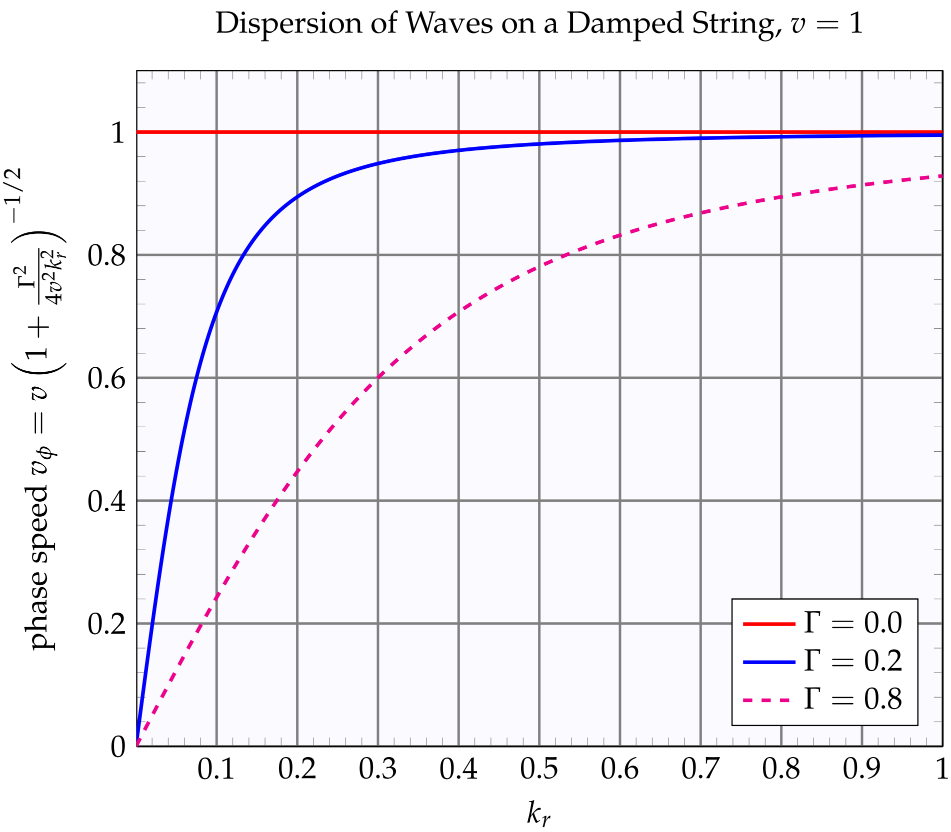


Figure 57: Speed of waves on a damped string

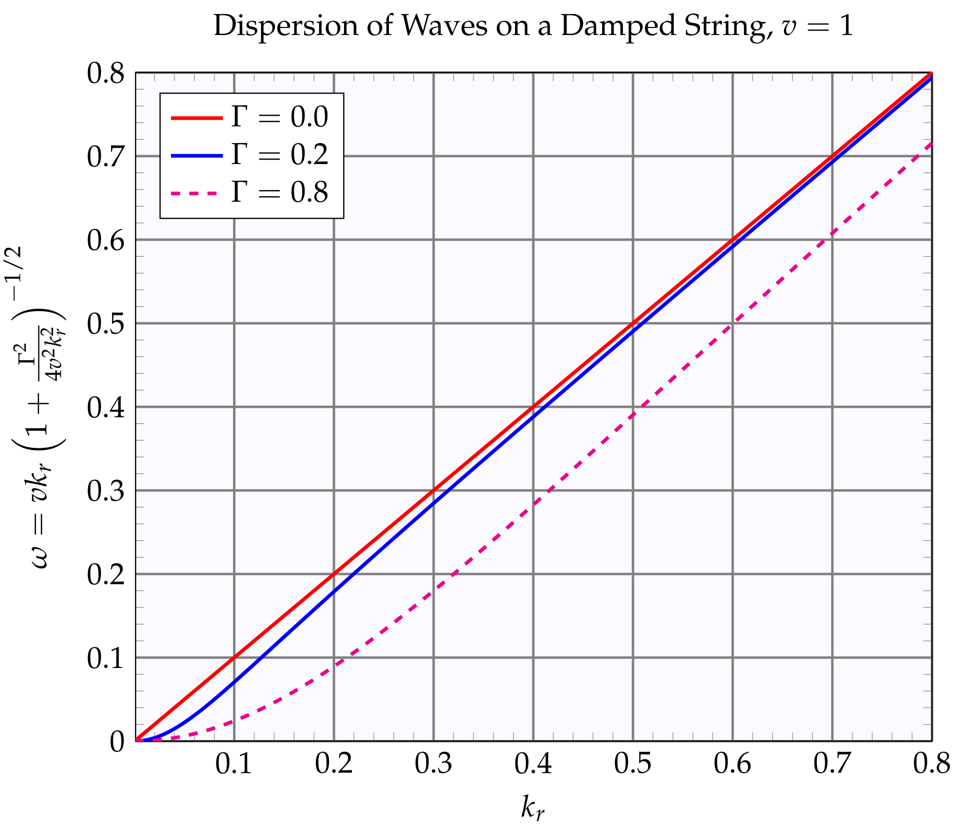


Figure 58: Dispersion Relation for Waves on a damped string

Second Dispersion Example: A Stiff String

As a second example of dispersion, consider a string which resists bending but has no external resistance (no damping). There exists a torque on any element of string when it is curved. This introduces a second term into the wave equation (details of derivation not important here — see dynamics course):

$$\frac{\partial^2\Psi}{\partial t^2} = \frac{T}{\rho}\left(\frac{\partial^2\Psi}{\partial x^2} - \alpha\frac{\partial^4\Psi}{\partial x^4}\right) = v^2\left(\frac{\partial^2\Psi}{\partial x^2} - \alpha\frac{\partial^4\Psi}{\partial x^4}\right) \tag{56}$$

Trying a harmonic wave solution $\Psi \propto \Re[\exp(\mathrm{i}\omega t - \mathrm{i}kx)]$, we find

$$\omega = \pm vk\sqrt{(1 + \alpha k^2)}$$

In contrast to the damped waves, there is no attenuation (because there is no loss of energy), but the phase speed of the wave $v_\phi = \omega/k$ is no longer v: in fact the wave travels at phase speed

$$|v_\phi| = \omega/k = v\sqrt{(1 + \alpha k^2)}$$

If k is small, the wavelength is long, the curvature is small, so $v_\phi \approx v$ — the string stiffness has little effect. If k is large, the string becomes more curved, so the stiffness becomes important.

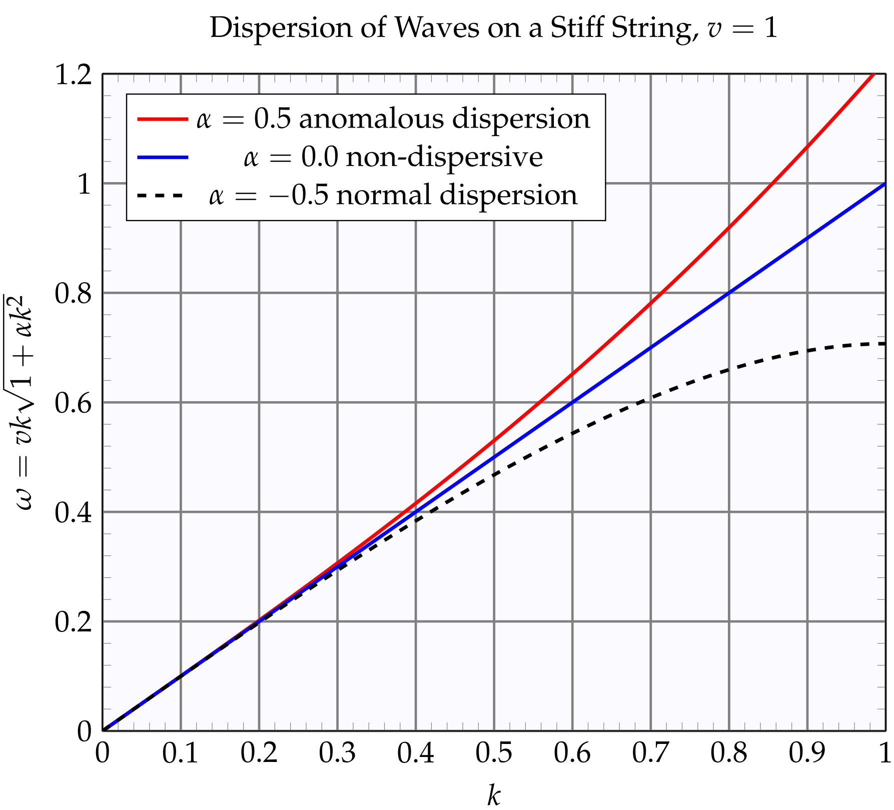


Figure 59: Dispersion Relation for a stiff string

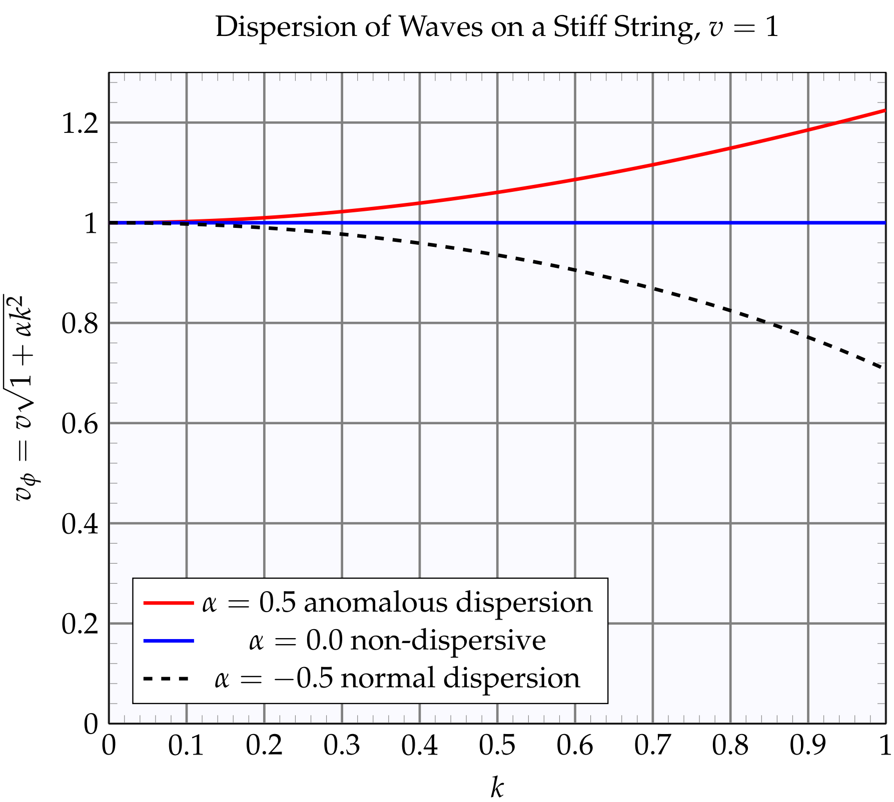


Figure 60: Phase Speed on a stiff string

Implications for Tuning a Piano

- A piano string of length l has fixed ends. Its fundamental mode (middle C, 256 Hz say) is a standing wave of wavelength $\lambda = 2l$; its first harmonic has wavelength $\lambda = l$; etc.
- If the phase speed as a function of wavelength is $v_\phi(\lambda)$, the frequencies of these are $f_{\text{fund}} = v_\phi(2l)/2l =$ and $f_{\text{harm}} = v_\phi(l)/l$.
- We want to tune the next C string one octave higher, i.e. at twice the frequency. But do we tune to twice f_{fund} or to f_{harm}? If v_ϕ is constant, these are the same and piano tuning is easy.
- But piano strings are stiff ($\alpha > 0$) so that v_ϕ increases with frequency, and so $f_{\text{harm}}/f_{\text{fund}} > 2$. Suppose $f_{\text{harm}} =$ 542 Hz So we can either
  - tune the second string's fundamental have twice the frequency of the first's fundamental (i.e. to 512 Hz); or
  - tune the second string to match the frequency of the first harmonic of the first string. (i.e. to 542 Hz)
- In fact we choose the second so you won't hear any beats (in this case, at 30 Hz) between these two modes when the strings are played simultaneously, so things will sound nice.

Group Velocity (Take One)

With no dispersion, a wavepacket or wave group propagates without change of shape at the speed given by the wave equation: $\ddot{\Psi} = v^2 \Psi''$. We have proved this earlier.
But if v depends on wavelength, the wave group may (a) travel at a different speed, and (b) may change its shape. This is the phenomenon of dispersion.
Water Wave Dispersion video
Consider two equal-amplitude waves propagating together, with slightly different frequencies, $\omega_1 \approx \omega_2$ (and corresponding wavevectors $k_1 \approx k_2$):

$$\Psi = \cos(\omega_1 t - k_1 x) + \cos(\omega_2 t - k_2 x)$$

Define the average and half-difference frequencies and wavevectors:

$$\omega_+ = \frac{1}{2}(\omega_1 + \omega_2), \quad \omega_- = \frac{1}{2}(\omega_1 - \omega_2), \quad k_+ = \frac{1}{2}(k_1 + k_2), \quad k_- = \frac{1}{2}(k_1 - k_2).$$

then simple algebra shows the usual beat phenomenon:

$$\Psi = 2\cos(\omega_+ t - k_+ x)\cos(\omega_- t - k_- x)$$

There is a high-frequency wave with speed $v_\phi = \omega_+/k_+$, which is very close to either of the phase velocities ω_1/k_1 or ω_2/k_2. This is modulated by a lower-frequency travelling envelope with speed $v_g = \omega_-/k_-$. For very close frequencies, we have

$$v_\phi = \frac{\omega}{k} \tag{57}$$

$$v_g = \frac{\omega_-}{k_-} = \frac{\omega_1 - \omega_2}{k_1 - k_2} \approx \left.\frac{\mathrm{d}\omega}{\mathrm{d}k}\right|_{\omega_+} \tag{58}$$

Group Velocity: Take Two

A more general argument is as follows. The group velocity v_g is the speed at which a 'bump' in the wave travels. We can regard the wave being composed of many individual harmonic waves of different frequencies (Fourier components); each has its own phase $(\omega t - kx + \phi)$.
If these phases are equal for the constituent waves, they will add up in phase to make a 'bump'.
Thus we can find the bump where $\omega t - kx + \phi$ is equal for all the constituent waves: this occurs when

$$\left[\frac{\mathrm{d}}{\mathrm{d}\omega}(\omega t - kx + \phi)\right]_{\omega_0} = 0$$

where we evaluate at ω_0, the angular frequency of the typical wave in the group. Hence

$$t - x\left(\frac{\mathrm{d}k}{\mathrm{d}\omega}\right)_{\omega_0} = 0$$

which shows the speed of the 'bump' is the group velocity:

$$\boxed{v_g = \frac{\mathrm{d}\omega}{\mathrm{d}k}}$$

Group Velocity: Take Three

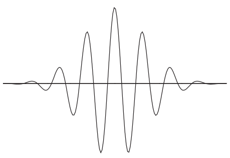


Figure 61: A wave group.

Consider a wave packet consisting of a carrier wave with wavenumber k_0 multiplied by an envelope $f(x)$ (in this example, f is a Gaussian). At $t = 0$:

$$\Psi(x, 0) = \Re\left[f(x)\mathrm{e}^{\mathrm{i}k_0 x}\right].$$

Using the Fourier transform, write the envelope in terms of its components:

$$\Psi(x, 0) = \Re\left[\left(\int_{-\infty}^{\infty} F(k_1)\mathrm{e}^{\mathrm{i}k_1 x}\mathrm{d}k_1\right)\mathrm{e}^{\mathrm{i}k_0 x}\right] = \Re\left[\int_{-\infty}^{\infty} F(k_1)\mathrm{e}^{\mathrm{i}(k_0+k_1)x}\mathrm{d}k_1\right]$$

$F(k_1)$ is the Fourier transform of $f(x)$ (ignore factors of e.g. $\sqrt{2\pi}$)
Each sinusoid with wavenumber $k_0 + k_1$ will propagate at its own phase velocity, so we can write the waveform at some later time t as

$$\Psi(x, t) = \Re\left[\int_{-\infty}^{\infty} F(k_1)\mathrm{e}^{\mathrm{i}[(k_0+k_1)x-(\omega_0+\omega_1)t]}\mathrm{d}k_1\right] = \Re\left[\int_{-\infty}^{\infty} F(k_1)\mathrm{e}^{\mathrm{i}(k_1 x-\omega_1 t)}\mathrm{d}k_1\mathrm{e}^{\mathrm{i}(k_0 x-\omega_0 t)}\right] \quad (59)$$

where $\omega_0 = \omega(k_0)$ and $\omega_0 + \omega_1 = \omega(k_0 + k_1)$ for a dispersion relation $\omega = \omega(k)$.

Now assume that $F(k_1)$ is non-zero over a small wavenumber range $\pm\Delta k$ ($\Delta k \ll k_0$), so only need consider waves in the band $k_0 \pm \Delta k$. (This is same as assuming $f(x)$ is many carrier wavelengths across). Now expand the dispersion relation $\omega(k)$ about k_0:

$$\omega = \omega_0 + \left.\frac{\partial \omega}{\partial k}\right|_{k_0} k_1 + \frac{1}{2} \left.\frac{\partial^2 \omega}{\partial k^2}\right|_{k_0} k_1^2 + \cdots \tag{60}$$

Working to first order in k_1:

$$\omega_1 \approx v_g k_1 \tag{61}$$

where the group velocity is $v_g = \left.\frac{\partial \omega}{\partial k}\right|_{k_0}$. Substituting equ. (61) into equ. (59) we have

$$\Psi(x,t) = \Re\left[\int_{-\infty}^{\infty} F(k_1) \mathrm{e}^{\mathrm{i}k_1(x-v_g t)} \mathrm{d}k_1 \mathrm{e}^{\mathrm{i}(k_0 x - \omega_0 t)}\right] = \Re\left[\int_{-\infty}^{\infty} F(k_1) \mathrm{e}^{-\mathrm{i}k_1 v_g t} \mathrm{e}^{\mathrm{i}k_1 x} \mathrm{d}k_1 \mathrm{e}^{\mathrm{i}(k_0 x - \omega_0 t)}\right]$$

Recognising the k_1 integral as an inverse Fourier transform, we can make use of the convolution theorem and the fact that the Fourier transform of a delta-function is a complex exponential to yield

$$\Psi(x,t) = \Re\left[f(x) * \delta(x - v_g t) \mathrm{e}^{\mathrm{i}(k_0 x - \omega_0 t)}\right] = \Re\left[f(x - v_g t) \mathrm{e}^{\mathrm{i}k_0(x - v_p t)}\right] \tag{62}$$

where $v_p = \omega_0 / k_0$.

Notes on Group Velocity

- equ. (62) says that after a time t the carrier wave $\mathrm{e}^{\mathrm{i}k_0 x}$ has propagated a distance $v_p t$ while the modulating envelope $f(x)$ has propagated a distance $v_g t$. For a non-dispersive wave $\omega/k = \frac{\partial \omega}{\partial k}$ for all frequencies, and so the envelope will propagate at the same speed as the carrier wave, but for a dispersive wave the two velocities can be different, and so the wave crests will move relative to the envelope.
- Typically we use the envelope as a tracer of the information carried by the wave packet, so the group velocity will be the speed of interest when looking at the propagation of information.
- The range of wavevectors Δk in a wave group is inversely related to the spatial extent of the group Δx_0 at $t = 0$:

$$\Delta k \Delta x \approx 1$$

- Note that the above result is restricted to wave groups consisting of sinusoidal components with a very narrow spread of frequencies. A broad-band signal can be thought of as consisting of a number of narrow-band wave groups, and if the group velocity is different between these groups (i.e. the second-order term in the expansion in equ. (60) is non-negligible) then the different wave groups will begin to spread apart as time progresses and the envelope will become distorted.

Estimate of rate of spreading of wave packet

By approximating the dispersion relation as a linear function near the carrier wavevector k_0 (i.e $\omega = \omega_0 + v_g k_1$), we have shown that a wave group propagates at $v_g = \frac{\partial \omega}{\partial k}$.
If we include a higher order (quadratic) term in the Taylor expansion, we also find that the envelope elongates and changes shape — it disperses.
A simple way to estimate this is to note that if the group contain wavevectors in a band of size $2\Delta k$ about k_0, then there exists a range of group velocities in the group, with typical extreme values being

$$v_{g1} = \left.\frac{\partial \omega}{\partial k}\right|_{k_0+\Delta k}, \quad v_{g2} = \left.\frac{\partial \omega}{\partial k}\right|_{k_0-\Delta k}$$

After a time t, the packet will thus spread by an amount equal to t times this difference between these two speeds; expanding the group velocity close to k_0 we predict that the wavepacket will have a length given (roughly) by

$$\Delta x \sim \Delta x_0 + (v_{g1} - v_{g2})t \sim \Delta x_0 + 2|\beta|\Delta k\, t \sim \Delta x_0 + 2|\beta| \frac{t}{\Delta x_0}$$

where Δx_0 is the initial length of the group, and

$$\beta = \left.\frac{\partial^2 \omega}{\partial k^2}\right|_{k_0}$$

and we used $\Delta x_0 \Delta k \approx 1$.

Water Waves 1

Water waves are surprisingly complex (see Main: Ch 13), but it's worth knowing their basic properties (but not how to derive them). In general, the water moves in both longitudinal and transverse directions, with these oscillations in quadrature – making ellipses.
In deep water (that is, many wavelengths deep) water has a dispersion relation:

$$\omega^2 = gk + \frac{\sigma k^3}{\rho}$$

where we recognise the first is due to gravitational forces, and the second is caused by the surface tension σ of the water. The two terms are equal when $k^2 = \rho g/\sigma$, which corresponds for water ($\sigma = 0.073\,\mathrm{N\,m^{-1}}$) to a wavelength of $\lambda_0 = 17\,\mathrm{mm}$ (corresponding to a velocity of $0.23\,\mathrm{m\,s^{-1}}$).

Water Waves: Ripples on Deep Water

Figure 62: Ripples

Ripples (or capillary waves) have shorter wavelengths than λ_0 and are surface tension driven: the dispersion relation is

$$\omega \approx \sqrt{\frac{\sigma k^3}{\rho}}, \quad v_\phi \approx \sqrt{\frac{\sigma k}{\rho}}, \quad v_g \approx \frac{3}{2}\sqrt{\frac{\sigma k}{\rho}} \approx \frac{3}{2} v_\phi$$

so that their phase speed and group speed decrease as wavelength increases. This is a case of 'anomalous' dispersion. Here, $v_g > v_\phi$, and the ripples appear to move backwards through the wavepacket. At 1 mm wavelength, the phase speed is 0.68 m s^{-1}.
ripples from a drop video

Water Waves: Gravity Waves in Deep Water

Figure 63: Deep ocean waves

Gravity waves have longer wavelengths than λ_0: propagation is controlled by the gravitational force. The dispersion relation is

$$\omega \approx \sqrt{gk}, \quad v_\phi \approx \sqrt{\frac{g}{k}}, \quad v_g \approx \frac{1}{2}\sqrt{\frac{g}{k}}$$

so that their phase speed and group speed increase as the wavelength increases. This is a case of 'normal' dispersion. In this case, $v_g = v_\phi/2$, and the individual waves appear to move forwards through the wavepacket. At 100 m wavelength, the phase speed is 12.5 m s^{-1} (about 28mph).
Ripples on a lake, Pond Ripples, Water Wave Dispersion video (speedboat)

Shallow Water Gravity Waves

- If the wavelength λ exceeds the water depth h, gravity waves have a dispersion relationship
$$\omega^2 = ghk^2\left(1 - \frac{h^2k^2}{3}\right)$$
and the wave motion is mainly longitudinal.
- For very shallow water, we can neglect the second term and $\omega \approx k\sqrt{gh}$, and we have approximately non-dispersive waves with phase/group speed $\sqrt{gh}$.
- Water waves approaching the shore slow down as h decreases, and so must increase in amplitude (height) to conserve water. (They often then break).
- A tsunami is caused by an earthquake which creates a wave of huge wavelength ($\sim$ 100 km); so the 'shallow water' approximation holds even in the deep ocean. For a typical Pacific Ocean depth of $h = 4$ km, the wavespeed is about 200 m s^{-1} (450 mph). The amplitude in deep ocean is $\sim$ 1 m, and the period is ~ 10 minutes, so ships hardly notice.
- There is very little dispersion, so the wavegroup arrives almost simultaneously at the shore line even after a long (say 12 hour) journey, where it slows and grows in amplitude — hence causing much damage.
`https://www.youtube.com/watch?v=3xKMFzKOIfQ` Japanese Tsunami footage from the air

Guided Waves

It is useful to confine waves to transmit energy or information:

- Electromagnetic Waveguides
- Optical fibres
- Stethoscopes

Note this list list does not include the national grid, or ethernet cables, both of which are transmission lines and have different characteristics.
The boundaries of the guide set conditions which affect the wave's propagation.

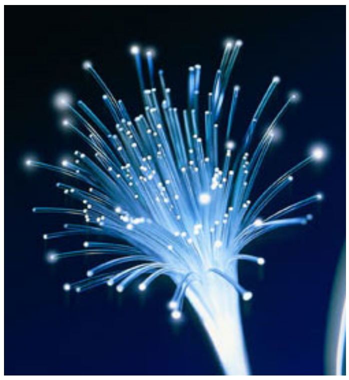


Figure 64: Waveguide Examples

Simple Model: Guided waves on a membrane

Consider waves on a large two-dimensional elastic membrane in the xy plane, with mass σ per unit area, and under tension γ per unit length. Its motion obeys

$$\frac{\gamma}{\sigma}\boldsymbol{\nabla}^2\Psi = v^2\boldsymbol{\nabla}^2\Psi = \frac{\partial^2\Psi}{\partial t^2}$$

So the membrane waves are non-dispersive, have speed v, and obey $\omega = vk$.

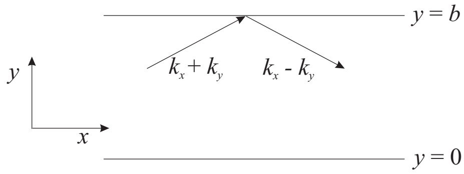


Figure 65: Two-dimensional waveguide.

Now let us clamp it along two parallel edges at $y = 0$ and $y = b$. Consider two travelling waves with wavevector k on the membrane, moving at equal angles $\pm\theta$ to x:

$$\Psi_A = +A\exp\mathrm{i}(\omega t - k_x x - k_y y), \quad \Psi_B = -A\exp\mathrm{i}(\omega t - k_x x + k_y y).$$

These waves have $\boldsymbol{k} = (k_x, \pm k_y) = (k\cos\theta, \pm k\sin\theta)$, and the total displacement is

$$\Psi = A\exp\mathrm{i}(\omega t - k_x x)\left[\exp\left(-\mathrm{i}k_y y\right) - \exp(\mathrm{i}k_y y)\right] = -2\mathrm{i}A\sin\left(k_y y\right)\exp\mathrm{i}(\omega t - k_x x).$$

This is a wave travelling in the x direction of wavelength $2\pi/k_x$, with amplitude modulated by a standing wave in the y direction.

Dispersion Relation for Guided Waves

You can think of the constituent waves being multiply reflected at the boundaries with a phase change of π at each as they move down the guide. The boundary conditions require that $\Psi = 0$ at $y = 0$ and $y = b$, i.e. that $\sin k_y b = 0$. This implies

$$k_y = \frac{n\pi}{b}$$

for integer n. Only discrete values of k_y are allowed. To determine the dispersion relation we substitute Ψ into the wave equation:

$$v^2\left(k_x^2 + k_y^2\right) = v^2k^2 = \omega^2$$

But k_y is quantised, so we obtain the dispersion relation

$$\omega^2 = v^2\left(k_x^2 + \frac{n^2\pi^2}{b^2}\right). \tag{63}$$

And the phase velocity of the waves travelling along the guide is

$$v_p = \frac{\omega}{k_x} = \frac{\omega}{\sqrt{\left(\frac{\omega^2}{v^2} - \frac{n^2\pi^2}{b^2}\right)}} \tag{64}$$

Dispersion

The guided waves are dispersive. Differentiating equ. (63), we find

$$2\omega\frac{\mathrm{d}\omega}{\mathrm{d}k_x} = 2v^2 k_x$$

$$\therefore \left(\frac{\omega}{k_x}\right)\frac{\mathrm{d}\omega}{\mathrm{d}k_x} = v^2$$

The product of the phase and group velocities is v^2 for this dispersion relation: $v_p v_g = v^2$, and

$$v_g = \frac{v^2}{\omega}\sqrt{\left(\frac{\omega^2}{v^2} - \frac{n^2\pi^2}{b^2}\right)}.$$

which has several branches fixed by the integer n.

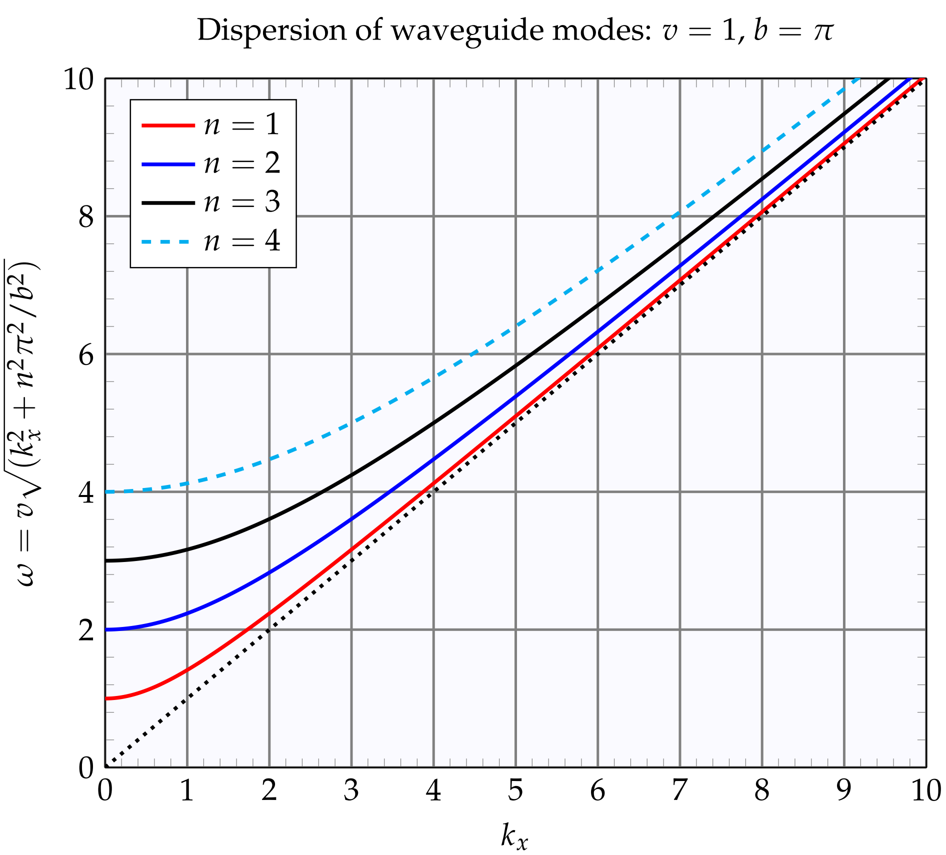


Figure 66: Dispersion Relation in Waveguides

Note the disallowed range of frequencies below $\omega_c = \pi v/b$ (which is unity in this example).

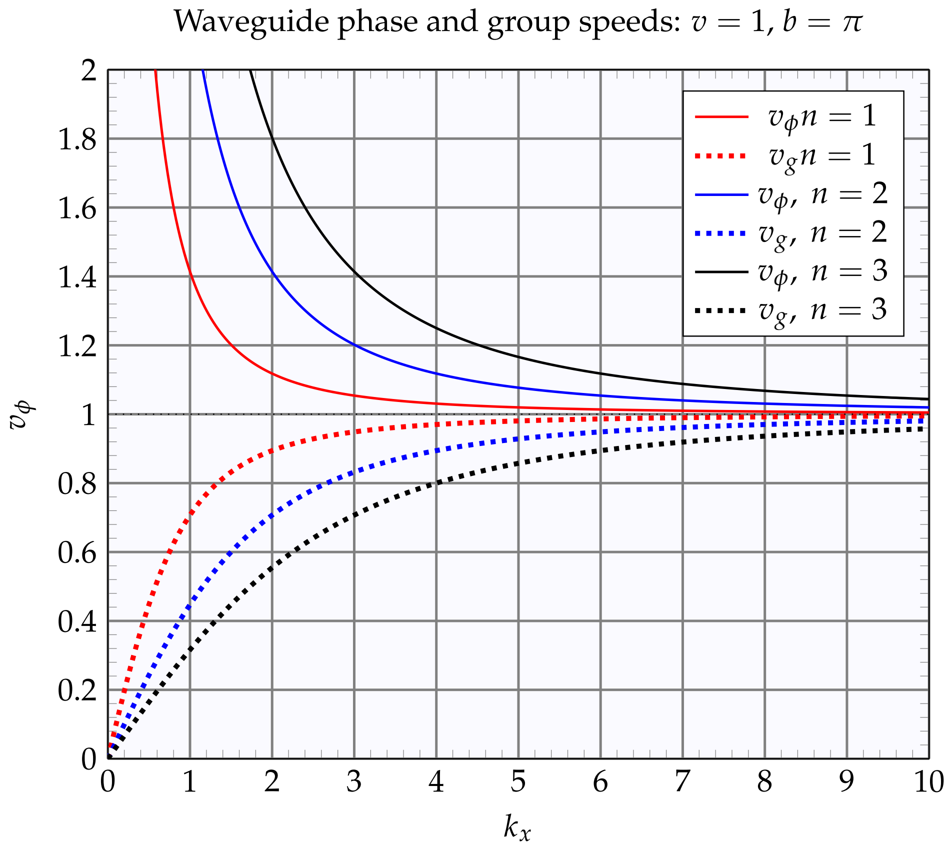


Figure 67: Phase and group Speed in Waveguides

Note that $v_\phi v_g = 1$ for this dispersion relation.

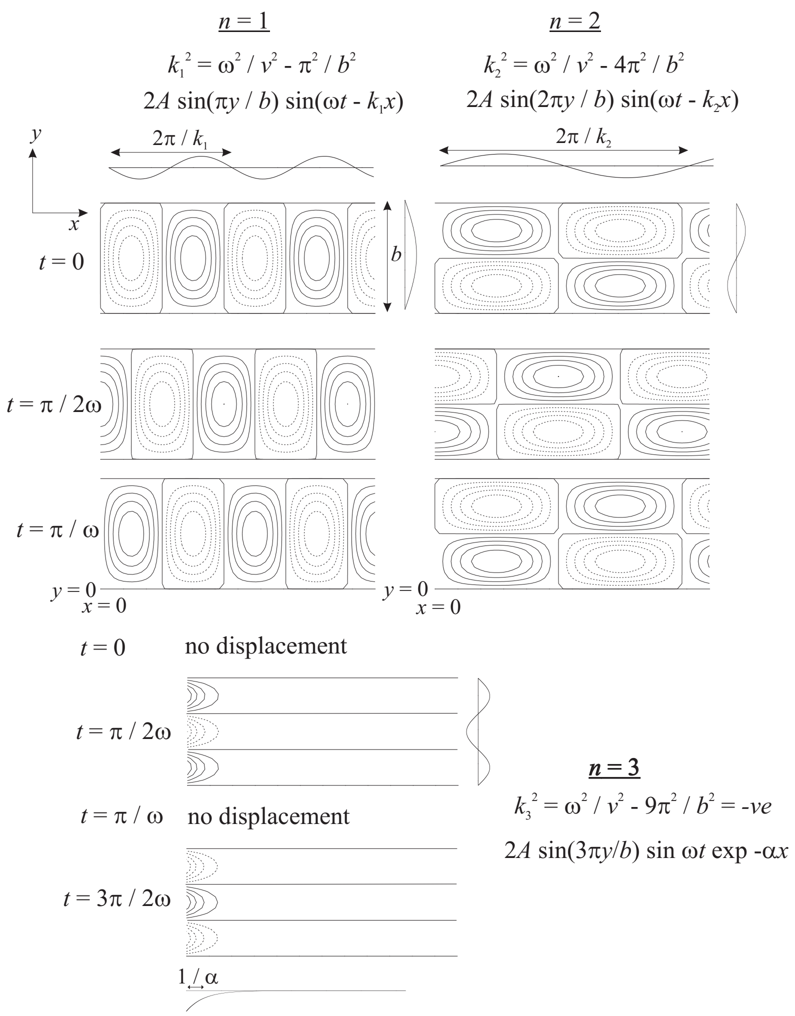


Figure 68: Modes in a waveguide.

Guided Wave Properties 1

- There is a series of waveguide modes specified by n, with different dispersion relations and displacement patterns.
- The wavelength of the guided wave is $\lambda_x = 2\pi/k_x$, which exceeds λ.

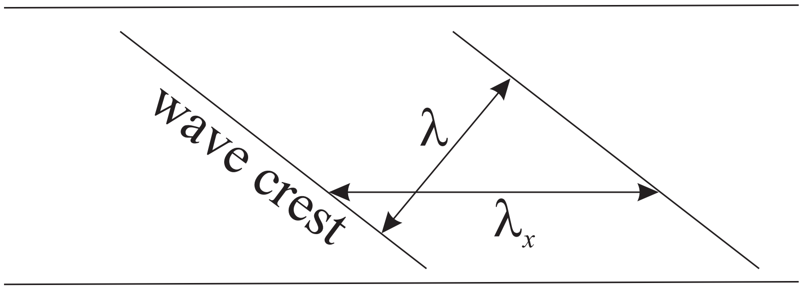


Figure 69: Wavelength in a guided wave

- Hence the phase velocity, $v_p = v/\cos\theta$ is greater than v, and the group velocity v_g is less than v.
- The group velocity can be expressed $v\cos\theta$ where $\tan\theta = k_y/k_x$. This makes sense — the constituent plane waves travel at angle θ to x.

Guided Wave Properties 2

- As $k_x \to 0$, $v_p \to \infty$ and $v_g \to 0$ (i.e. dω/d$k_x = 0$ at $k_x = 0$). This does not violate special relativity, since energy and information are transmitted at v_g.
- As $k_x \to \infty$, the behaviour approaches that of an unguided wave, with v_p and $v_g \to v$.
- Below a cutoff angular frequency $\omega_c = \frac{n\pi v}{b}$, k_x^2 becomes negative and propagation is not possible. The lowest cut-off frequency occurs for $n = 1$; for $\omega < (\pi v/b)$, propagation is not possible.
- If a spread of frequencies exists, multiple modes can be can excited. These have different speeds, and the signal will distort (disperse) as it travels. To avoid this problem, ensure only one mode $n = 1$ is possible by choosing a frequency below the cutoff frequency of mode $n = 2$:

$$\frac{\pi v}{\omega_0} < b < \frac{2\pi v}{\omega_0}$$

The guide is then single-moded for the given frequency ω_0.

Optical Fibres

- Optical fibres are an extremely important example of a waveguide: a cylindrical silica core (typical diameter 9 μm), is clad by a lower refractive index glass. Usually, infrared waves are guided down the core of the fibre.
- The data is sent as a series of pulses, and the data rate which can be achieved over a given length of fibre is determined by the dispersion of these pulses. The aim is to choose a wavelength where the dispersion is a minimum (but also need low loss).

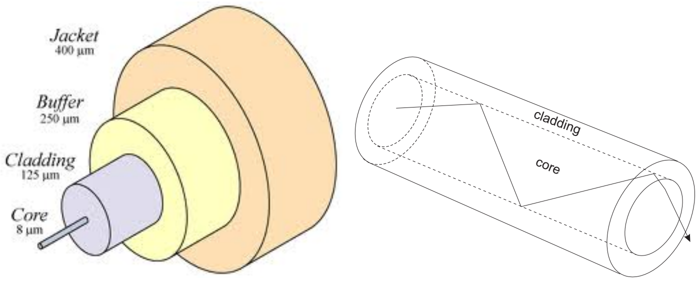


Figure 70: Schematic representation of an optical fibre.

- Different propagation modes exist, with different group velocities. Most fibres are designed so only one propagating mode exists. There is dispersion (a) from the waveguide dispersion relation, and (b) due to the refractive index of glass depending on wavelength. Try to design a fibre so these two effects have opposite signs and therefore cancel. Current optical fibre systems can carry data at rates of 100 Gbits/s by modulating a single-wavelength laser source operating at 1550 nm. With Wavelength Division Multiplexing (WDM), using e.g. 16 wavelengths, can get 1.6 Terabits/s.

Evanescent Waves: some notes

We showed that the dispersion relation for guided waves is

$$k_x^2 = k^2 - k_y^2 = \frac{\omega^2}{v^2} - \frac{n^2\pi^2}{b^2}. \tag{65}$$

- Below the lowest cut-off frequency, $\omega_c = nv\pi/b$ with $n = 1$, we see that k_x^2 becomes negative, so that $k_x = \pm i\alpha$ for some real α. Hence

$$\Psi = A \sin k_y y \exp(\mathrm{i}\omega t) \exp(\pm|k_x|x)$$

- This is not a propagating wave — there no phase change with x; all parts of the membrane move up and down in phase together, with the amplitude decaying exponentially with distance along the wave. This is an evanescent wave.
- The positive exponential solution is usually excluded if the membrane extends over many times $1/|k_x|$ — otherwise energy would not be conserved.
- Since the phase is now independent of position (as in standing waves), there is no net power flow along the guide. As a consequence, a travelling wave with $\omega < \omega_c$ which is incident on the guide will undergo perfect reflection (with a phase shift) and a standing wave will be set up outside the guide.

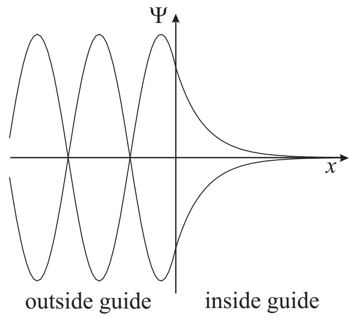


Figure 71: Travelling wave with $\omega < \omega_c$ incident on a waveguide.

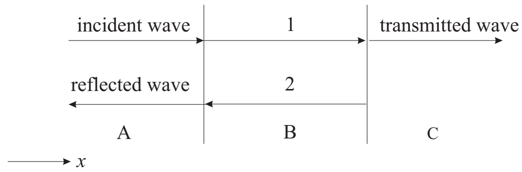


Figure 72: Evanescent wave between two discontinuities.

If a system has a second discontinuity beyond which it is again possible to have a travelling wave, some of the incident wave would be transmitted into this region, (C), and there will only be partial reflection at the boundary AB. There are two evanescent waves in region B; wave 1 decays as $\exp(-\alpha x)$, and wave 2 increases as $\exp(+\alpha x)$. The combination of these two gives a position dependence of the phase angle for the total disturbance in region B and, as a result, a net flow of energy from region A to region C is now possible.
This can be applied to quantum tunnelling.

Superposition

H134(270921)

The equation of motion for the damped driven oscillator

$$F(t) = m\ddot{x}(t) + b\dot{x}(t) + kx(t)$$

is linear.

if	$F_1(t)$	gives displacement	$x_1(t)$
and	$F_2(t)$	gives displacement	$x_2(t)$
then	$c_1F_1(t) + c_2F_2(t)$	gives displacement	$c_1x_1(t) + c_2x_2(t)$.

Consider a force which is the sum of two sinusoidal driving forces at different frequencies. We can calculate the individual responses to the two forces (using the response function $R(\omega)$), and add the results.

$$\begin{array}{ccc}
\mathbf{F}_1 e^{i\omega_1 t} & \rightarrow & R(\omega_1)\mathbf{F}_1 e^{i\omega_1 t} \\
+ & & + \\
\mathbf{F}_2 e^{i\omega_2 t} & \rightarrow & R(\omega_2)\mathbf{F}_2 e^{i\omega_2 t} \\
\| & & \| \\
\mathbf{F}_1 e^{i\omega_1 t} + \mathbf{F}_2 e^{i\omega_2 t} & \rightarrow & R(\omega_1)\mathbf{F}_1 e^{i\omega_1 t} + R(\omega_2)\mathbf{F}_2 e^{i\omega_2 t}
\end{array}$$

This leads to the idea that if we can decompose a force into sinusoids, we can determine the response to that force in terms of the sum of the responses to each of the sinusoids.

Fourier Series

What is the response of an oscillator to e.g. a square wave driving force? A periodic force can be represented as a sum of discrete sinusoidal components. The Fourier series of a periodic function $f(t)$, period T, angular frequency $\omega_0 = 2\pi/T$ is

$$f(t) = \frac{1}{2}A_0 + \sum_{n=1}^{\infty}\left[A_n \cos\left(\frac{2\pi nt}{T}\right) + B_n \sin\left(\frac{2\pi nt}{T}\right)\right] \tag{66}$$

i.e. a superposition of sinusoidal waves with frequencies $\omega_0, 2\omega_0, 3\omega_0 \ldots$.
To find the coefficients A_n, B_n, we note these orthogonality properties:

$$\begin{aligned}
\int_{-T/2}^{T/2} \cos\left(\tfrac{2\pi mt}{T}\right)\cos\left(\tfrac{2\pi nt}{T}\right)\mathrm{d}t &= 0 && \text{for } m \neq n \\
&= T/2 && \text{for } m = n \\
\int_{-T/2}^{T/2} \cos\left(\tfrac{2\pi mt}{T}\right)\sin\left(\tfrac{2\pi nt}{T}\right)\mathrm{d}t &= 0 && \text{for all } m,\ n \\
\int_{-T/2}^{T/2} \sin\left(\tfrac{2\pi mt}{T}\right)\sin\left(\tfrac{2\pi nt}{T}\right)\mathrm{d}t &= 0 && \text{for } m \neq n \\
&= T/2 && \text{for } m = n.
\end{aligned}$$

Fourier Series 2

Hence, to find the value of a particular coefficient A_m or B_m, we multiply the defining equation equ. (66) by $\cos\left(\frac{2\pi mt}{T}\right)$ or $\sin\left(\frac{2\pi mt}{T}\right)$ and integrate from $-T/2$ to $T/2$ to give

$$\boxed{A_n = \frac{2}{T}\int_{-T/2}^{T/2} f(t)\cos\left(\frac{2\pi nt}{T}\right)\mathrm{d}t} \tag{67}$$

$$\boxed{B_n = \frac{2}{T}\int_{-T/2}^{T/2} f(t)\sin\left(\frac{2\pi nt}{T}\right)\mathrm{d}t} \tag{68}$$

These integrals 'project out' the coefficient we need — all the other sinusoidal components in $f(t)$ are orthogonal to the component we are interested in, so make no contribution to the integral.

Fourier Series for a Square Wave

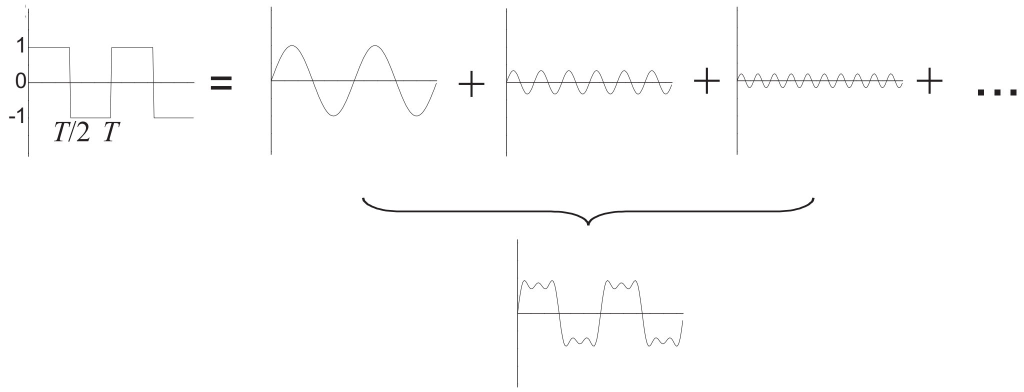


Figure 73: Fourier analysis of a square wave.

Symmetry considerations are often helpful: since the square wave drawn is an odd function ($f(x) = -f(-x)$), $A_n = 0$ for all n. And

$$\begin{aligned} B_n &= \frac{4}{\pi n} \text{ for } n \text{ odd} \\ B_n &= 0 \text{ for } n \text{ even,} \end{aligned}$$

$$\therefore f(t) = \frac{4}{\pi}\left(\sin(\omega_0 t) + \frac{1}{3}\sin(3\omega_0 t) + \frac{1}{5}\sin(5\omega_0 t) + \ldots\right)$$

Square wave driving of a damped oscillator

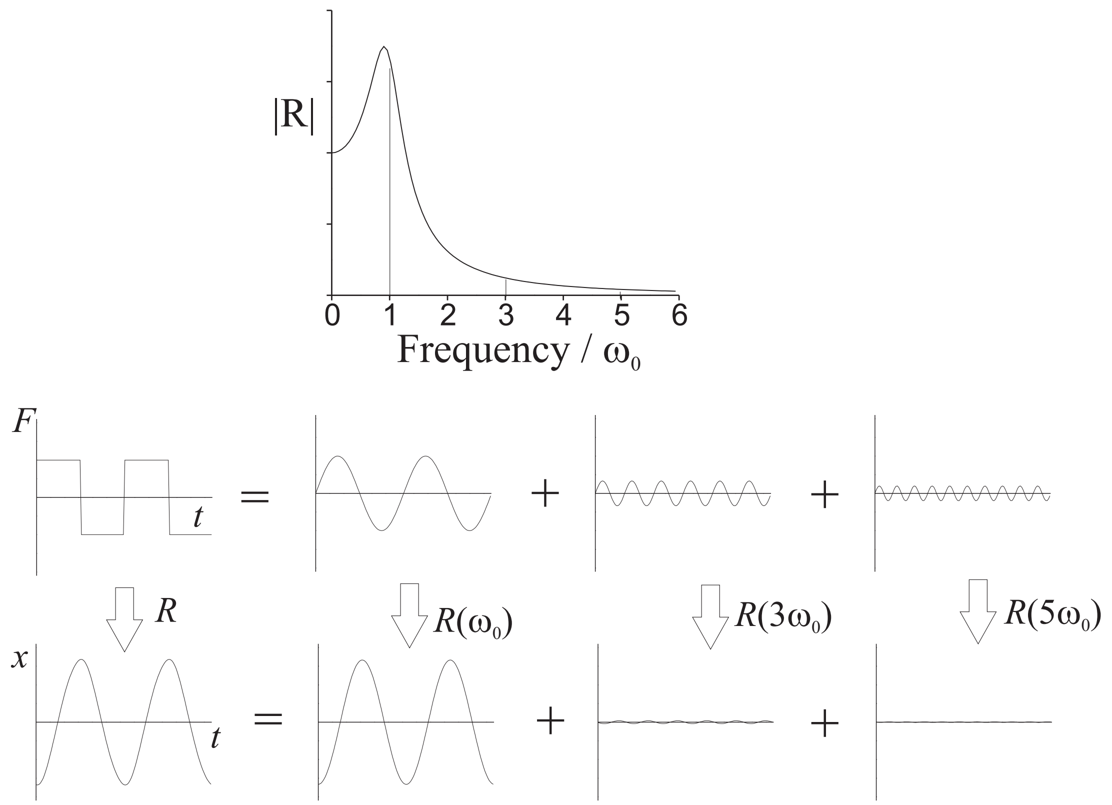


Figure 74: Response function for a damped oscillator with $Q = 1.7$, and the response of this oscillator to a square wave at frequency ω_0

Since $|R|$ tends to zero at high frequencies, first few Fourier components often give a good approximation to the response.

Complex coefficients

Rather than write the series as a sum of sine and cosine terms, it often convenient to describe the Fourier components with complex coefficients which represent the amplitude and phase of each component.

$$f(t) = \sum_{n=-\infty}^{\infty} C_n \mathrm{e}^{\mathrm{i}2\pi nt/T} = \sum_{n=-\infty}^{\infty} C_n \mathrm{e}^{\mathrm{i}n\omega_0 t}. \tag{69}$$

The coefficients C_n are given by multiplying by $\mathrm{e}^{-\mathrm{i}n\omega_0 t}$ and integrating from $-T/2$ to $T/2$,

$$C_n = \frac{1}{T}\int_{-T/2}^{T/2} f(t)\mathrm{e}^{-\mathrm{i}n\omega_0 t}\mathrm{d}t.$$

This results relies on the fact that

$$\begin{aligned} \int_{-T/2}^{T/2} \mathrm{e}^{\mathrm{i}n\omega_0 t}\mathrm{e}^{-\mathrm{i}m\omega_0 t}\mathrm{d}t &= 0 \quad \text{for } m \neq n \\ &= T \quad \text{for } m = n. \end{aligned} \tag{70}$$

We are interested in functions $f(t)$ which are real, and in this case we find that $C_{-m} = C_m^*$. This ensures that the imaginary parts of the terms in equ. (69) cancel to zero when summing over positive and negative n.

Example Fourier Series

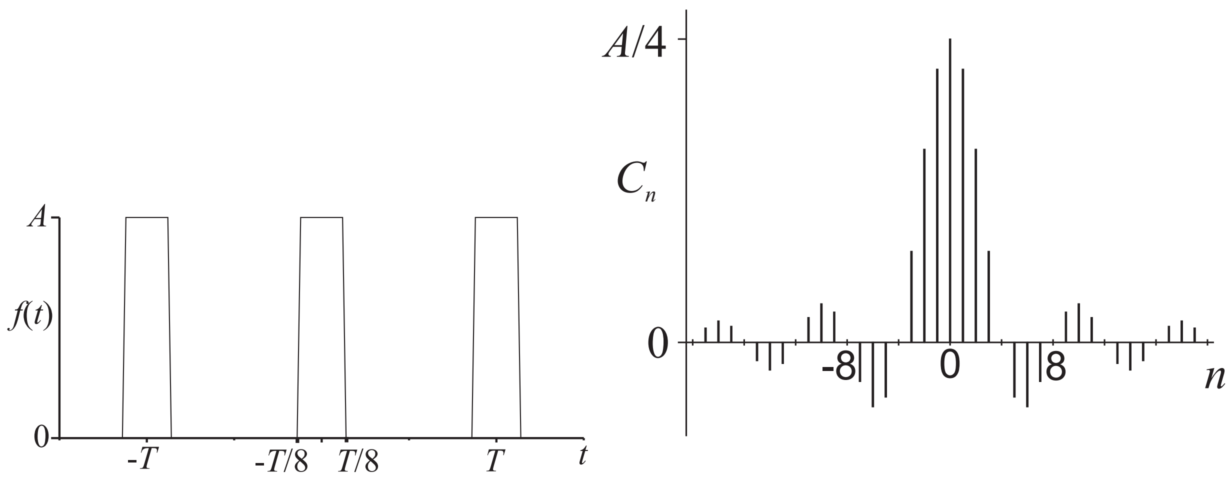


Figure 75: Fourier coefficients for a periodic function.

This function repeats with period $T = 2\pi/\omega_0$, and is non-zero for $-T/8 < t < T/8$. Its Fourier coefficients are

$$C_n = \frac{1}{T}\int_{-T/8}^{T/8} A\mathrm{e}^{-\mathrm{i}n\omega_0 t}\,\mathrm{d}t = \frac{A}{T}\left[\frac{\mathrm{e}^{-\mathrm{i}n\omega_0 t}}{-\mathrm{i}n\omega_0}\right]_{-T/8}^{T/8} = \frac{A}{\pi n}\sin(n\pi/4)$$

The coefficient is zero whenever n is a multiple of 4.

Fourier Transforms

If a function is not-periodic, it is still useful to be able to decompose it into sinusoidal waves.
Taking the limit of a Fourier series as $T \to \infty$, the fundamental frequency tends to zero, and the frequency components become infinitesimally close together. The sum over discrete frequency components then becomes an integral over a continuous spectrum of frequency components.
We write the function as

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(\omega) \mathrm{e}^{\mathrm{i}\omega t} \mathrm{d}\omega \tag{71}$$

where

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}t. \tag{72}$$

These equations define the Fourier transform; $g(\omega)$ is the Fourier transform of $f(t)$ and $f(t)$ is the inverse Fourier transform of $g(\omega)$. There is some arbitrariness in the choice of constants before the integrals: their product must be $(2\pi)^{-1}$, but you may see various conventions for how this is split up between the two integrals. Also, you may see these written using $f = \omega/2\pi$.
It is important to realise that $f(t)$ and $g(\omega)$ contain the same information in different ways.
Both are complete descriptions of the signal.

Fourier Space

We denote the Fourier transform with the operator $\mathcal{F}$ (or sometimes "F.T.") which transforms the function f into the function g (i.e. separating a function into its component sinusoids)

$$\mathcal{F}[f(t)] = g(\omega),$$

and its inverse $\mathcal{F}^{-1}$ which finds $f(t)$ given $g(\omega)$ (adding the sinusoids back together)

$$\mathcal{F}^{-1}[g(\omega)] = f(t).$$

Fourier transforms are not just applied to functions of time: they can equally be applied to functions of any variable. For example we can Fourier transform a function of a spatial variable x to derive a function of wavenumber k, i.e.

$$\mathcal{F}[f(x)] = g(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) \mathrm{e}^{-\mathrm{i}kx} \mathrm{d}x.$$

The spaces represented by ω and k are "reciprocal spaces" complementary to t and x respectively, since they have units of s^{-1} and m^{-1} respectively. The idea of a reciprocal space will be familiar from crystallography, where Fourier transforms abound.
k is often called a spatial frequency.

General Response of a Linear System

We know how a linear system (e.g. an oscillator) responds to a sine wave: the response is $R(\omega)$ times the input.
Using Fourier theory, we can find the general response, because the Fourier transform $g(\omega)$ represents the splitting up of our arbitrary function $f(t)$ into its component sinusoids.
For each frequency, $|g(\omega)|$ tells us the amplitude of the component sinusoid at that frequency and $\arg[g(\omega)]$ tells us the phase.
The response to each sinusoid is another sinusoid at the same frequency, with (complex) amplitude $R(\omega)$ times the input. The total response is found by adding back together the output sinusoids, using the inverse Fourier transform. Thus for an arbitrary input force $F(t)$, the response is

$$x(t) = \mathcal{F}^{-1}\left[R(\omega)\mathcal{F}\{F(t)\}\right]. \tag{73}$$

This kind of approach is very general, and we will use Fourier theory many times in this course, and in other courses this year.

Example Fourier transforms

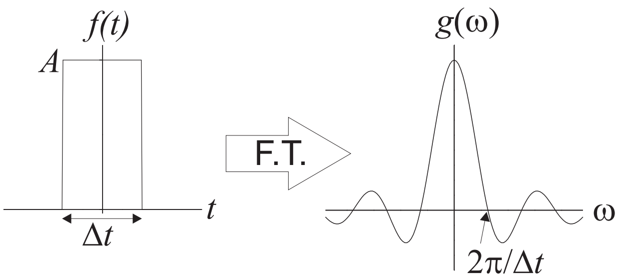


Figure 76: Fourier transform of a top-hat function.

The Fourier transform can be evaluated analytically in simple cases. Consider the top hat-function

$$\begin{aligned} f(t) &= A \quad -\Delta t/2 < t < \Delta t/2 \\ &= 0 \quad |t| > \Delta t/2. \end{aligned}$$

The Fourier transform, $g(\omega)$ is given by

$$\begin{aligned} g(\omega) &= \frac{1}{\sqrt{2\pi}}\int_{-\Delta t/2}^{\Delta t/2} A\mathrm{e}^{-\mathrm{i}\omega t}\mathrm{d}t \\ &= \frac{A}{\sqrt{2\pi}}\left[\frac{\mathrm{e}^{-\mathrm{i}\omega t}}{-\mathrm{i}\omega}\right]_{-\Delta t/2}^{\Delta t/2} \\ &= \frac{A}{\sqrt{2\pi}}\left(\frac{\mathrm{e}^{\mathrm{i}\omega\Delta t/2}-\mathrm{e}^{-\mathrm{i}\omega\Delta t/2}}{i\omega}\right) \\ &= \frac{2A}{\sqrt{2\pi}\,\omega}\sin\left(\omega\Delta t/2\right) \\ &= \frac{A\Delta t}{\sqrt{2\pi}}\,\mathrm{sinc}\left(\omega\Delta t/2\right). \end{aligned}$$

This demonstrates the more general result that the width of a signal in the frequency domain is inversely proportional to the width in the time domain. Here the width is Δt in the time domain, and the first zero in the frequency domain occurs when $\omega\Delta t = 2\pi$

Power Spectrum

The spectrum of a signal, often called the power spectrum, is simply the modulus squared of its Fourier transform, i.e. $|g(\omega)|^2$. This particular form occurs in many experimental situations where the phase of $g(\omega)$ cannot be measured or is not meaningful. For example, a prism splits light into sinusoids of different frequencies i.e. a Fourier transform. The light wave is oscillating at very high frequencies, so we typically cannot measure the phase of the light in each spectral channel and instead we record the intensity of that light as a function of frequency, i.e. we take a value proportional to the modulus squared of the light amplitude, either by looking at the spectrum with our eye or electronically in a spectrometer.

Fourier Transform of Arbitrary Function

The Fourier transform of a function can be computed rapidly using a "Fast Fourier Transform" (FFT). Here is the time-domain waveform of the spoken syllable "ah", and its frequency spectrum.

The fundamental frequency is easily visible in the time domain – this is the frequency of oscillation of the vocal chords which gives the perceived "pitch" of the vowel. Other frequency components are hard to distinguish in the waveform, but are easy to see in the frequency spectrum. These "formants" are determined by resonances in the vocal tract, and give the characteristic timbre which allow us to distinguish the vowel as "ah".

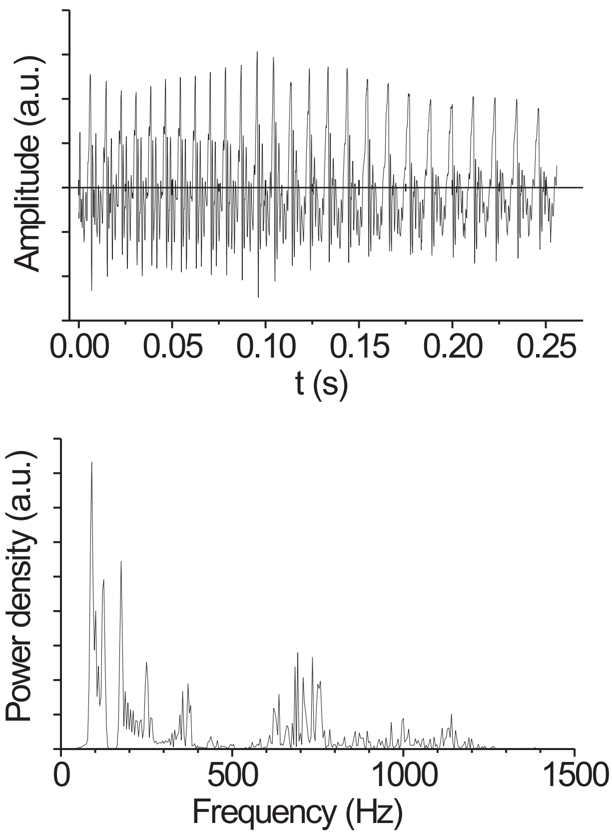


Figure 77: Waveform of "ah" and its Fourier transform (shown as a power spectrum).

The Delta Function

The Dirac delta function is used to denote a sharp "spike" which occurs over an infinitesimal time, but with finite area. An example of where this might be used is the idea of an "impulse" in Newtonian mechanics, where a sharp "kick" imparts finite momentum even though it occurs over an infinitesimally short time Δt. The momentum change is $F\Delta t$, so as Δt tends to zero, then F must tend to infinity during the kick. We would represent the force as a function of time in this limiting case as a Dirac delta function. Formally, the Dirac delta function $\delta(t)$ is defined by the property

$$\boxed{\int_{-\infty}^{\infty} \delta(t) f(t)\,\mathrm{d}t = f(0)}$$

for any arbitrary function $f(t)$. This means that $\delta(t)$ is a unit-area "spike" at $t = 0$ and is zero everywhere else. The function $\delta(t - t_0)$ is the same "spike" offset from the origin by an amount t_0 so

$$\int_{-\infty}^{\infty} \delta(t - t_0) f(t)\,\mathrm{d}t = f(t_0)\,.$$

Multiplying a function by $\delta(t - t_0)$ and integrating returns a "sample" of the value of the function at t_0.

Fourier transform of a delta function

- We can use this previous property to show that the Fourier transform of a delta function is a complex exponential:

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \delta(t - t_0) \mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t = \frac{\mathrm{e}^{-\mathrm{i}\omega t_0}}{\sqrt{2\pi}}\,.$$

- Note that this transform has constant magnitude, independent of frequency.
- And that the Fourier transform of a δ-function at the origin is a constant $1/\sqrt{2\pi}$. A delta function contains all frequencies in equal amounts. For the function $\delta(x)$, its component sinusoids add up in phase at $x = 0$, and out of phase everywhere else!

Convolution

The convolution of two functions $f(y)$ and $g(y)$, is denoted by the operator "$*$".

$$f(y) * g(y) = \int_{-\infty}^{\infty} f(u)g(y-u)\,\mathrm{d}u.$$

This is easy to visualise when one of the functions is a δ-function:

$$f(y) * \delta(y-y_0) = \int_{-\infty}^{\infty} f(u)\delta(y-y_0-u)\,\mathrm{d}u = f(y-y_0).$$

In other words, the function $f(y)$ is replicated, but centred at the location of the δ-function spike rather than around zero.
A general function $g(y)$ is made up of the sum of an infinite number of δ-functions with different "heights".

$$g(y) = \int_{-\infty}^{\infty} g(u)\delta(y-u)\,\mathrm{d}u$$

Hence, when we take the convolution of f and g, each of the δ-functions making up g is replaced by a copy of f. The function g is therefore "smeared out" by the function f (and vice versa).

Convolution has an important application in optics where imaging systems such as microscopes or telescopes are concerned. These instruments are imperfect, so a point source object (a δ-function) is smeared out in the image. The image produced from a delta function object is described by the resolution function (or point spread function) of the instrument ($f(\mathbf{r})$). For a general object, the image produced is the convolution of the object with the resolution function. The operation of trying to remove the effects of the resolution function is known as deconvolution.

Convolution Becomes Multiplication in Fourier Space

Convolutions are particularly useful in the context of Fourier transforms, and hence in Fraunhofer diffraction. It can be shown that the Fourier transform of the product of two functions is the convolution of the Fourier transforms of the individual functions, i.e. if

$$F(q) = \mathcal{F}[f(x)]$$

and

$$G(q) = \mathcal{F}[g(x)]$$

then

$$\mathcal{F}[f(x)g(x)] = \frac{1}{\sqrt{2\pi}} F(q) * G(q).$$

Similarly, the Fourier transform of the convolution of two functions is the product of the Fourier transforms of the individual functions.

$$\mathcal{F}[f(x) * g(x)] = \sqrt{2\pi}\, F(q)G(q).$$

Impulse response functions 1

Suppose we strike a linear system (e.g. an oscillator) with a sharp impulse $F(t) = \delta(t)$. Its response is its impulse response function $R'(t)$.
For a damped oscillator, the impulse response function is simply the transient after giving the system an initial velocity $v_0 = 1/m$ at $t = 0$.
Suppose, though, that we do not know the transient response of our system, only the frequency response function $R(\omega)$. We can then use Fourier theory to find $R'(t)$.
equ. (73) tells us to take the Fourier transform of $F(t)$, multiply by $R(\omega)$ to calculate the response in the frequency domain, then inverse Fourier transform the result to get the response in the time domain. The Fourier transform of $\delta(t)$ is a constant, hence the impulse contains all frequencies with equal amplitude

$$\delta(t) \xrightarrow{\text{F.T.}} F_\omega(\omega) = \frac{1}{\sqrt{2\pi}}.$$

The response is given by

$$x_\omega(\omega) = R(\omega)F_\omega(\omega) = \frac{1}{\sqrt{2\pi}} R(\omega).$$

To find the response in the time domain, $R'(t)$, we take the inverse Fourier transform

$$R'(t) = \frac{1}{\sqrt{2\pi}} \mathcal{F}^{-1}\left\{R(\omega)\right\}.$$

Impulse response functions 2

Hence we can see that $R'(t)$ and $R(\omega)$ are directly related by the Fourier transform: if we know the frequency response we can calculate the impulse response, and vice versa.
If we know $R'(t)$, we can calculate the response to any driving force $F(t)$. We can write $F(t)$ as the sum of an infinite number of δ-functions with different weights:

$$F(t) = \int_{-\infty}^{\infty} F(t')\delta(t - t')\,\mathrm{d}t'.$$

Using the principle of superposition, we can then find the responses of the system to each of these δ-functions, and add them up:

$$x(t) = \int_{-\infty}^{\infty} F(t')R'(t - t')\,\mathrm{d}t'.$$

We note that this is just the convolution of F with R': this is to be expected from the fact that equ. (73) contains the product of their Fourier transforms. Hence, knowing either $R'(t)$ or $R(\omega)$ gives us all the information we need about a well-behaved linear system in order to calculate its response to any driving force.

Some Building Block Transforms

$f(t)$	$g(\omega) = \int_{\infty}^{\infty} f(t)\exp(-\mathrm{i}\omega t)\mathrm{d}t/\sqrt{2\pi}$
delta-function $\delta(t - t_0)$	complex exponential $\dfrac{\mathrm{e}^{-\mathrm{i}\omega t_0}}{\sqrt{2\pi}}$
top-hat $f(t) = \begin{cases} 0 & \text{if } \lvert t\rvert > \frac{1}{2} \\ 1 & \text{if } \lvert t\rvert \le \frac{1}{2} \end{cases}$	sinc $\dfrac{\sin(\omega/2)}{\sqrt{2\pi}(\omega/2)}$
Gaussian $\mathrm{e}^{-t^2/2}$	Gaussian $\mathrm{e}^{-\omega^2/2}$
decaying exponential $f(t) = \begin{cases} 0 & \text{if } t < 0 \\ \mathrm{e}^{-t} & \text{if } t \ge 0 \end{cases}$	low-pass filter $\dfrac{1}{\sqrt{2\pi}(1 + i\omega)}$
comb function $\sum_{n=-\infty}^{\infty} \delta(t - nT)$	comb function $\dfrac{\sqrt{2\pi}}{T} \sum_{n=-\infty}^{\infty} \delta(\omega - \dfrac{2\pi n}{T})$

Table 1: Fourier transforms of some useful functions

Evaluating the Fourier transform or inverse Fourier transform of a function by performing the integrals in equ. (71) and equ. (72) can be time-consuming. It is often possible to use a few "building-block" Fourier transforms and combine them using known mathematical properties of the Fourier transform.

Reciprocity: If you know the forward transform, you obtain the inverse transform by "flipping" the result about the $t = 0$ axis: if the F.T. of $f(t)$ is $g(\omega)$, then the inverse F.T. of $f(\omega)$ is $g(-t)$. Hence all following results apply when you replace $\mathcal{F}$ with $\mathcal{F}^{-1}$.

Scaling law: A function "stretched" horizontally by a has its transform compressed by the same factor: $\mathcal{F}[f(t/a)] = |a|g(a\omega)$. This reciprocal scaling is the origin of Heisenberg's uncertainty principle, since wavefunctions of quantum variables such as position and momentum form a Fourier transform pair. Note that the vertical dimension of the transformed function is stretched by $|a|$.

Linearity: If a function is the superposition of two other functions, its F.T. is the superposition of the respective F.T.'s:
$\mathcal{F}[a_1 f_1(t) + a_2 f_2(t)] = a_1 \mathcal{F}[f_1(t)] + a_2 \mathcal{F}[f_2(t)]$ where a_1 and a_2 are arbitrary constants.

Convolution:

$$\mathcal{F}[f_1(t)f_2(t)] = \frac{1}{\sqrt{2\pi}}\mathcal{F}[f_1(t)] * \mathcal{F}[f_2(t)]$$

$$\mathcal{F}[f_1(t) * f_2(t)] = \sqrt{2\pi}\,\mathcal{F}[f_1(t)]\mathcal{F}[f_2(t)].$$

Fourier Examples: Cosine Function

A cosine can be decomposed into a sum of exponentials

$$\cos(\omega_0 t) = \frac{1}{2}(\mathrm{e}^{\mathrm{i}\omega_0 t} + \mathrm{e}^{-\mathrm{i}\omega_0 t})$$

Using the first line of Table 1 and reciprocity, the F.T. of $\mathrm{e}^{\mathrm{i}\omega_0 t}$ is $\sqrt{2\pi}\delta(\omega - \omega_0)$ and the F.T. of $\mathrm{e}^{-\mathrm{i}\omega_0 t}$ is $\sqrt{2\pi}\delta(\omega + \omega_0)$. Using linearity we have

$$\mathcal{F}[\cos(\omega_0 t)] = \frac{\sqrt{2\pi}}{2}[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)] \tag{74}$$

So the F.T. of a cosine function is a pair of delta-functions, one on either side of the origin.

Gaussian wavepacket 1

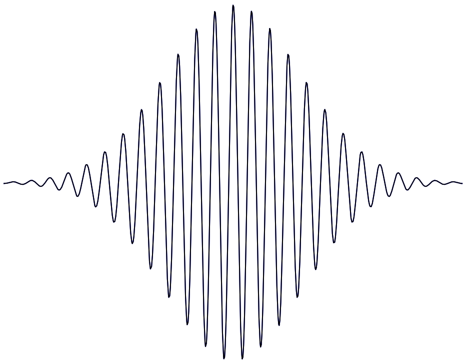


Figure 78: A Gaussian wavepacket

A Gaussian wavepacket is a sinusoid multiplied by a Gaussian envelope.

$$f(t) = A\mathrm{e}^{-t^2/2\sigma^2}\cos\omega_0 t$$

where σ is the width of the Gaussian, A is the peak amplitude and ω_0 is the "carrier frequency" of the packet. We decompose this function into the product of two functions we know the transforms of, namely a Gaussian and a cosine. For the Gaussian, we can use the scaling law and linearity to get the transform:

$$\mathcal{F}\left\{\mathrm{e}^{-t^2/2}\right\} = \mathrm{e}^{-\omega^2/2} \Rightarrow \mathcal{F}\left\{A\mathrm{e}^{-t^2/2\sigma^2}\right\} = A\sigma\mathrm{e}^{-\omega^2\sigma^2/2}$$

Gaussian Wavepacket 2

The convolution theorem then allows us to combine the F.T. of the Gaussian with the F.T. of the cosine (equ. (74)) to give

$$\begin{aligned} g(\omega) &= \frac{1}{\sqrt{2\pi}}\mathcal{F}\left\{A\mathrm{e}^{-t^2/2\sigma^2}\right\} * \mathcal{F}\left\{\cos\omega_0 t\right\} \\ &= \tfrac{1}{2}A\sigma\left([\delta(\omega-\omega_0)+\delta(\omega+\omega_0)] * \mathrm{e}^{-\omega^2\sigma^2/2}\right) \\ &= \tfrac{1}{2}A\sigma\left(\mathrm{e}^{-(\omega-\omega_0)^2\sigma^2/2} + \mathrm{e}^{-(\omega+\omega_0)^2\sigma^2/2}\right). \end{aligned} \quad (75)$$

In other words we get two Gaussians centred around $\pm\omega_0$.
The width of the Gaussians in the frequency domain is inversely proportional to the width of the Gaussian in the time domain, so we have the result that a short packet occupies more bandwidth than a longer packet. If, for example, we were using such packets to transmit digital pulses over a radio link at a central frequency of ω_0, then it is clear that if we want to transmit shorter pulses we need a larger frequency allocation to avoid overlapping nearby frequency channels.

Gaussian Wavepackets and Their Transforms

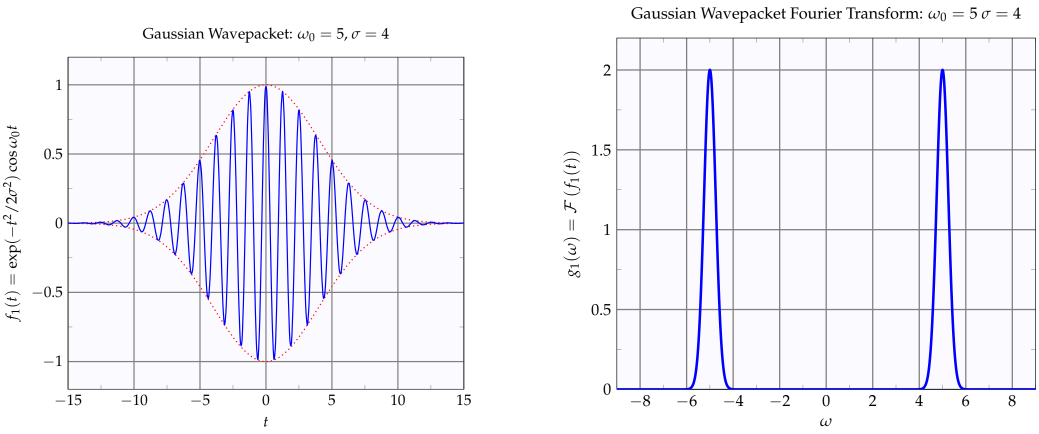


Figure 79: Gaussian wavepacket and its transform

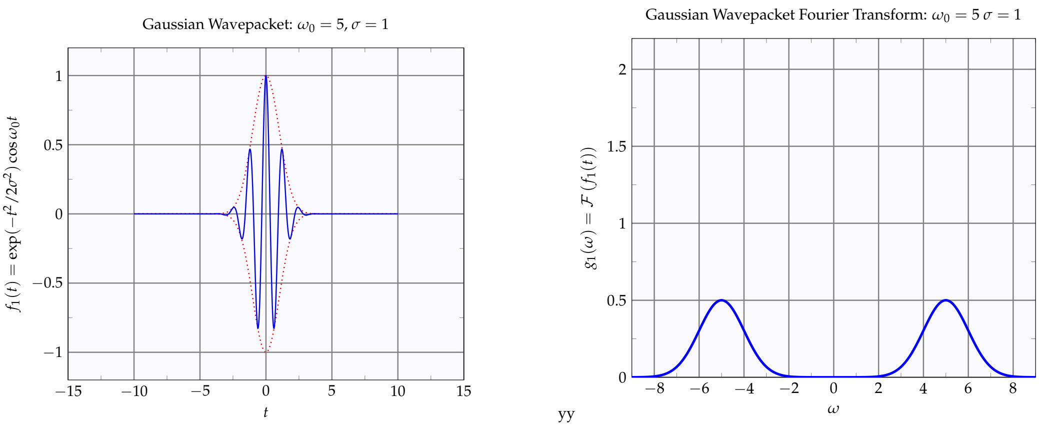


Figure 80: Gaussian wavepacket and its transform

Symmetry

It is often helpful to know the symmetry properties of Fourier transforms.
It is straightforward to show from equ. (72) that the Fourier transform of a real function $f(t)$ (something which is true of most of the physical variables we will be taking the Fourier transform of) has so-called Hermitian symmetry, i.e. that $g(-\omega) = g(\omega)^*$. This means that the properties of the function can be determined purely from the positive frequency components of the Fourier transform, so typically we only plot the positive half of the F.T. in these cases.
Similarly, we can show that if a function is both real and symmetric (i.e. $f(-t) = f(t)$) then its Fourier transform is both real and symmetric, while if the function is both real and antisymmetric (i.e. $f(-t) = -f(t)$), then its transform is purely imaginary and antisymmetric.

Electromagnetic Waves

The detailed physics of EM waves is covered in Physics B. We only need to know the key results here. Maxwell's equations in a linear, isotropic medium:

$$\nabla \cdot (\epsilon \boldsymbol{E}) = \rho \tag{76}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{77}$$

$$\nabla \times (\boldsymbol{B}/\mu) = \boldsymbol{J} + \epsilon \frac{\partial \boldsymbol{E}}{\partial t} \tag{78}$$

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{79}$$

- ϵ_r is the relative permittivity of the material, and ϵ_0 is the permittivity of free space, equal to $\approx 8.85 \times 10^{-12}\,\mathrm{F\,m^{-1}}$, and $\epsilon = \epsilon_r \epsilon_0$.
- A material that polarises easily has a high permittivity: $\boldsymbol{P} = \chi_e \epsilon_0 \boldsymbol{E}$ where χ_e is the electric susceptibility, and $\epsilon_r = (1 + \chi_e)$.
- μ_r is the relative permeability of the material, and μ_0 is the permeability of free space, equal to $4\pi \times 10^{-7}\,\mathrm{H\,m^{-1}}$, and $\mu = \mu_r \mu_0$.
- A material that can be easily polarised magnetically has a high permeability: $\boldsymbol{M} = \chi_m \boldsymbol{H}$ where χ_m is the magnetic susceptibility, $\mu_r = (1 + \chi_m)$, and $\boldsymbol{B} = \mu_r \mu_0 \boldsymbol{H}$.

Wave Equation

Taking the curl of equ. (79), and using equ. (78),

$$\nabla \times (\nabla \times \boldsymbol{E}) = -\nabla \times \dot{\boldsymbol{B}}; \quad \therefore \nabla(\nabla \cdot \boldsymbol{E}) - \nabla^2 \boldsymbol{E} = -\nabla^2 \boldsymbol{E} = -\epsilon\mu \frac{\partial^2 \boldsymbol{E}}{\partial t^2}$$

Thus $\boldsymbol{E}$ satisfies a wave equation

$$\boxed{\nabla^2 \boldsymbol{E} = \mu\epsilon \frac{\partial^2 \boldsymbol{E}}{\partial t^2} = \frac{1}{c^2} \frac{\partial^2 \boldsymbol{E}}{\partial t^2}} \tag{80}$$

with a wavespeed $c = (\epsilon\mu)^{-1/2}$.
In free space, the speed is exactly (by definition)

$$c_0 = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = 299\,792\,458\,\mathrm{m\,s^{-1}}$$

The refractive index n is defined by the speed of light in a material relative to free space:

$$n = \frac{c_0}{c} = \sqrt{\epsilon_r \mu_r}$$

Hence the polarisation properties of the material, which are expressed via μ_r and ϵ_r, determine the speed of light and any dispersion effects in the material. For many materials, μ_r is very close to unity, and we have $n \approx \sqrt{\epsilon_r}$.

Plane EM Wave

In a plane wave in free space or a simple dielectric, the electric and magnetic fields are transverse to the propagation direction, orthogonal to one another, and oscillate up and down in phase.

$$\boldsymbol{E} = \boldsymbol{E}_0 \exp(\mathrm{i}\omega t - \mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r})$$
$$\boldsymbol{B} = \boldsymbol{B}_0 \exp(\mathrm{i}\omega t - \mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r})$$

Taking the curl of the electric field, we have

$$\boldsymbol{\nabla} \times \boldsymbol{E} = -\dot{\boldsymbol{B}} \quad \therefore -i\boldsymbol{k} \times \boldsymbol{E}_0 = -i\omega \boldsymbol{B}_0$$

This gives the relative orientations of $\boldsymbol{E}$, $\boldsymbol{B}$ and $\boldsymbol{k}$; in addition, the magnitudes are related via

$$\frac{|\boldsymbol{E}_0|}{|\boldsymbol{B}_0|} = \frac{\omega}{k} = c = \frac{c_0}{n}$$

We define the impedance as

$$Z = \frac{|\boldsymbol{E}|}{|\boldsymbol{H}|} = \frac{|\boldsymbol{E}|}{|\boldsymbol{B}/\mu|} = c\mu = \sqrt{\frac{\mu}{\epsilon}} = \sqrt{\frac{\mu_r \mu_0}{\epsilon_r \epsilon_0}} = Z_0 \sqrt{\frac{\mu_r}{\epsilon_r}}$$

If we assume $\mu_r = 1$, then

$$Z = \frac{Z_0}{n}, \quad \text{where } Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} \approx 376.730\,\Omega$$

Z_0 is the impedance of free space.

Reflection at Normal Incidence

At a boundary, the components of $\boldsymbol{E}$, and of $\boldsymbol{B}/\mu = \boldsymbol{H} = \boldsymbol{E}/Z = n\boldsymbol{E}/Z_0$, parallel to the boundary are continuous. Consider an EM wave of unit electric field amplitude, in a medium with refractive index n_1 (filling $x < 0$), striking a medium of refractive index n_2 filling $x > 0$, at normal incidence. The electric fields in the two regions, accounting for the reflected and transmitted waves, and omitting a common $\exp(\mathrm{i}\omega t)$ factor, are

$$E_1 = \exp(-\mathrm{i}k_1 x) + r\exp(\mathrm{i}k_1 x)$$

$$E_2 = t\exp(-\mathrm{i}k_2 x)$$

Here r and t are the amplitude reflection and transmission coefficients for the electric field. The continuity of E and $H = nE/Z_0$ at the boundary imply

$$1 + r = t; \quad n_1(1 - r) = n_2 t$$

remembering that Z is negative for a wave traveling to $-x$. Solving, we obtain

$$r = \frac{n_1 - n_2}{n_1 + n_2} = \frac{Z_2 - Z_1}{Z_1 + Z_2}$$

For example, if $n_1 = 1$ and $n_2 = \infty$, then $Z_1 = Z_0$ and $Z_2 = 0$, and we predict $r = -1$: this is like a string wave meeting a clamped end.
At optical wavelengths, reflection and transmission formulae are usually quoted in terms of the refractive indices of the media, instead of the wave impedances.
So for example, the quarter wave matching condition becomes $n_2 = \sqrt{n_1 n_3}$.

Physical optics

There are many ways of treating optical phenomena, each of them emphasising different aspects and approximating other aspects so as to allow a tractable analysis. We can arrange these as a hierarchy as to the type of approximation which is involved:

Quantum Electrodynamics : the full theory of E.M. interaction with matter, but too complicated for everything but simple systems

Maxwell's Equations : Valid when the energy of individual photons is negligible in comparison to the light intensities being examined, but the boundary conditions are complicated to compute for all but the simplest cases (see work of Sommerfeld).

Physical Optics : This involves so-called scalar wave theory, i.e. we ignore polarisation (mostly), and simplify the boundary conditions. We use Huygens' construction of secondary waves to derive most results.

Ray Optics : Here we ignore wave properties (i.e. assume λ is much smaller than any scale of interest). This gives so-called geometric optics and is equivalent to Newton's "corpuscular theory".

We will be using physical optics approximations for most of this course. This allows us to tackle a range of problems where the system may include complex shaped apertures but where the significant features are typically larger than a wavelength.

Diffraction

Diffraction concerns the passage of a wave past some obstruction. The phenomenon is general to all waves, but is particularly important in optics. You have already been introduced to the diffraction grating last year; here we will develop a more general treatment which will show the relationship to Fourier transforms, will generalise into two dimensions, and will allow us to study diffraction closer to the aperture.

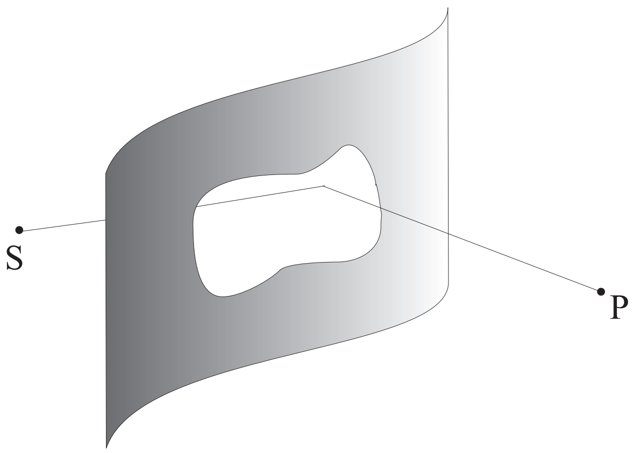


Figure 81: Diffraction from an aperture.

A source of waves, S, is placed behind an aperture, and we wish to calculate the distribution of intensity as a function of position on the other side of the aperture. We will assume that the illumination produced by S is spatially coherent, i.e. there is a well-defined phase relationship between the parts of the wave arriving at different points on the aperture.

Huygens' Principle

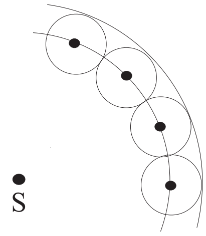


Figure 82: Huygens' Principle: secondary wavelets emerging from the primary wavefront originating from the source point S.

Huygens' principle: each point on a wavefront acts as a source of secondary wavelets which propagate, overlap, interfere, and thus carry the wavefront forward.
Can use this to find angles of reflection and refraction at an interface, e.g.

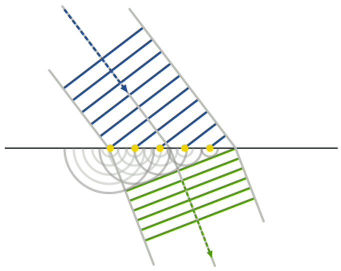


Figure 83: Huygens' Principle: the refraction problem.

$$n_1 \sin\theta_i = n_2 \sin\theta_r$$

Huygens-Fresnel Theory

The simple idea of spherical wavelets would lead to a backward-propagating wavefront as well as a forward-propagating one. This is not observed experimentally. To fix Huygens' principle, we have to introduce an obliquity or inclination factor, $K(\theta)$, which describes the fall-off in intensity of the wavelets with angle, θ, away from the forward direction.
In its simplest form, $K = 1$ for $|\theta| < \pi/2$ and zero elsewhere, as assumed by Huygens.
Fresnel tried to find the form of $K(\theta)$ which agreed with experiment results. He proposed

$$K(\theta) = \frac{1 + \cos\theta}{2}.$$

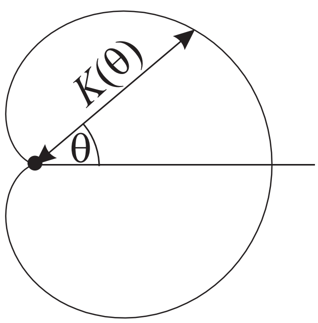


Figure 84:

In addition, he found that to get the correct diffraction patterns, the secondary wavelets have an amplitude $(-\mathrm{i}/\lambda)$ relative to the primary wave.
This extension to Huygens' principle is known as Huygens-Fresnel theory.

The Diffraction Integral

Consider a planar aperture, Σ (although the approach extends naturally to apertures of any shape). Consider an element of aperture $(\mathrm{d}x\,\mathrm{d}y)$ at position (x, y):

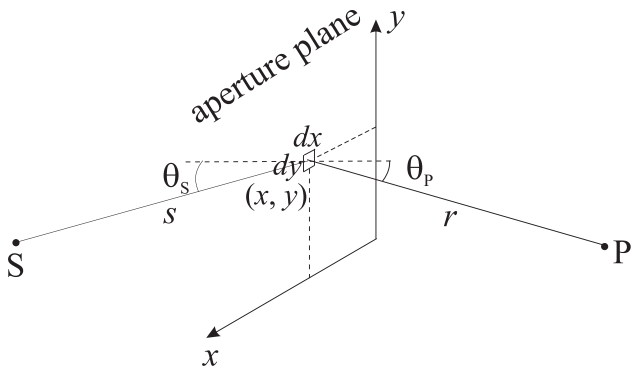


Figure 85: Diffraction from a planar aperture.

We will consider monochromatic 3-dimensional waves:

$$\Psi(\boldsymbol{r}, t) = \Re\{\psi(\boldsymbol{r})\mathrm{e}^{-\mathrm{i}\omega t}\}$$

Since the wave is monochromatic with a known value of ω everywhere, then we can trace the propagation of the wave $\Psi(\boldsymbol{r}, t)$ through the propagation of the complex spatial phasor $\psi(\boldsymbol{r})$.

The source S produces spherical waves with a "strength" a_S, and is at a distance s from the element of aperture. The wave arriving at the aperture element is therefore

$$\psi_1(\boldsymbol{r}) = \frac{a_S \mathrm{e}^{\mathrm{i}ks}}{s}$$

The aperture (in a coordinate plane denoted by Σ) can change the amplitude and phase of the incident radiation, and for a planar aperture we can describe its transmission properties by the (generally) complex aperture function, $h(x, y)$. Usually $h = 0$ or 1 for obstructing or open areas respectively. The element of aperture can now be considered to act as a source of secondary wavelets with a strength and phase given by

$$a_\Sigma = A\,\psi_1(x, y)\,h(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

The amplitude of the wavelets relative to the incoming wave is $A = -\mathrm{i}/\lambda$ (see later). The secondary wave generates a disturbance at P, which is a distance r away from the aperture element:

$$\begin{aligned} \mathrm{d}\psi_P &= -\frac{\mathrm{i}}{\lambda}\,\frac{a_S\,\mathrm{e}^{\mathrm{i}ks}}{s}\,h(x, y)\,\mathrm{d}x\,\mathrm{d}y\,K(\theta)\,\frac{\mathrm{e}^{\mathrm{i}kr}}{r} \\ &= -\frac{\mathrm{i}}{\lambda}\,h(x, y)\,K(\theta)\,\frac{a_S\,\mathrm{e}^{\mathrm{i}k(s+r)}}{s\;r}\,\mathrm{d}x\,\mathrm{d}y. \end{aligned} \tag{81}$$

$K(\theta)$ here is the obliquity factor, which turns out to depend on the angles θ_S and θ_P associated with the directions of the vectors from the aperture element to points S and P, relative to the aperture normal (derivation: Goodman, Introd. to Fourier Optics, Ch.3):

$$K = \frac{\cos\theta_S + \cos\theta_P}{2}.$$

Often, the angles are assumed small and we set $K \approx 1$.

The Fresnel-Kirchhoff Diffraction Integral

- To calculate the total amplitude at point P, we finally sum over all the elements of the aperture

$$\psi_P = \iint_\Sigma -\frac{\mathrm{i}}{\lambda}\, h(x,y)\, K(\theta_S,\theta_P)\, \frac{a_S\, \mathrm{e}^{\mathrm{i}k(s+r)}}{s\ r}\, \mathrm{d}x\, \mathrm{d}y. \tag{82}$$

- Note that both s and r are functions of x and y.
- This is often known as the Fresnel-Kirchhoff diffraction integral.
- This is a general result, and allows the amplitude of diffracted light to be calculated at any point P, although it breaks down for for points very close ($< \lambda$) to the edge of an aperture.
- At these locations, we need to take into account the vector nature of the electromagnetic field. We will not consider these very near-field cases any further, and will concentrate on distances $> \lambda$.

"The sole virtue of Kirchhoff's theory of diffraction lies in its correct predictions and not in its false assumptions" (Andrews, 1947)

Justification for the $(-\mathrm{i}/\lambda)$ factor*

Consider a plane wave propagating along z, meeting an aperture Σ in the xy plane at $z = 0$. The plane wave has amplitude ψ_s at the aperture. The diffraction integral tells us that the amplitude at a point P on axis a distance L from the aperture is

$$\psi_P = \iint_\Sigma K(\theta_S,\theta_P)\, \frac{A\psi_s \mathrm{e}^{\mathrm{i}kr}}{r}\, \rho \mathrm{d}\rho\, \mathrm{d}\phi$$

where A is the (complex) amplitude of the secondary wavelets, and (ρ,ϕ) are polar coordinates in the xy plane. Now

$$L^2 + \rho^2 = r^2; \quad \therefore \rho \mathrm{d}\rho = r \mathrm{d}r$$

The upper limit for the r integral is some function of the aperture geometry — call it $R(\phi)$. Assume the aperture is small and L is large, $K = 1$. Hence

$$\psi_P = \int_0^{2\pi} \int_L^{R(\phi)} \frac{A\psi_s \mathrm{e}^{\mathrm{i}kr}}{r}\, r\, \mathrm{d}r\, \mathrm{d}\phi = \frac{-2\pi A}{ik} \psi_S \mathrm{e}^{\mathrm{i}kL} + \frac{A\psi_S}{ik} \int_0^{2\pi} \mathrm{e}^{\mathrm{i}kR(\phi)} \mathrm{d}\phi$$

The integrand in the second term will average to zero for most apertures, as it oscillates rapidly as $R(\phi)$ varies. This is perhaps more obvious if we move a small distance off axis. (Strictly speaking, if $R(\phi)$ is constant, i.e. we have a perfect circular aperture, we will get an extra non zero term on axis) But for a large aperture, we expect on axis to get an amplitude $\psi_S \mathrm{e}^{\mathrm{i}kL}$, so

$$\boxed{A = -\mathrm{i}/\lambda}$$

Fraunhofer Diffraction

We can always evaluate the diffraction integral equ. (82) numerically. But it's good to develop physical insight with a few cases where the geometry is simple and the integral is easy to evaluate.
Consider the diffraction pattern in a plane at a distance L from the aperture. We denote the coordinates of point P in this plane by (x_0, y_0). Assume that the source S is a large distance behind the aperture (and centred on the aperture) so that $s \to \infty$, and the aperture is illuminated with a plane wave at normal incidence.

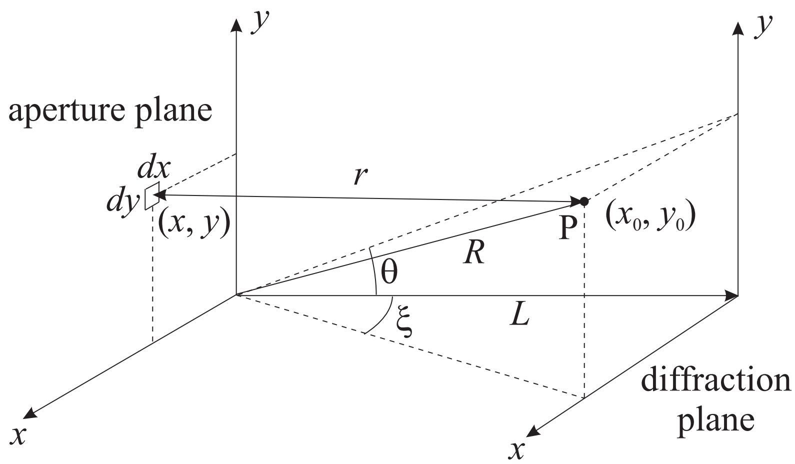


Figure 86: Geometry for Fraunhofer diffraction.

We will consider points P close to the axis, where (x_0/L) and (y_0/L) are sufficiently small that we can assume that the obliquity factor $K = 1$.

The distance r from the element of aperture $\mathrm{d}x\,\mathrm{d}y$ to P is given by

$$\begin{aligned} r^2 &= L^2 + (x_0 - x)^2 + (y_0 - y)^2 && (83)\\ &= L^2 + x_0^2 + y_0^2 - 2(x_0 x + y_0 y) + x^2 + y^2 && (84)\\ &= R^2 \left(1 - 2\frac{x_0 x + y_0 y}{R^2} + \frac{x^2 + y^2}{R^2}\right) && (85) \end{aligned}$$

where $R^2 = L^2 + x_0^2 + y_0^2$. Using binomial expansion, we obtain

$$r \approx R - \frac{x_0 x + y_0 y}{R} + \frac{x^2 + y^2}{2R}$$

The phase of each wavelet $\propto kr$, so the phase will in general have terms which vary linearly and quadratically with the position (x, y) in the aperture. If we make L (and hence R) large enough, we can arrange that

$$\frac{k(x^2 + y^2)}{2R} \ll \pi \qquad (86)$$

for all elements in the aperture. This is the condition for Fraunhofer diffraction. The wave amplitude at P is then given by the Fraunhofer integral

$$\psi_P \propto \iint_\Sigma \psi_\Sigma\, h(x, y)\, \exp\left[-\mathrm{i}k\left(\frac{x_0 x + y_0 y}{R}\right)\right] \mathrm{d}x\,\mathrm{d}y$$

with $\psi_\Sigma =$ constant if we illuminate the aperture with a plane wave at normal incidence. We have dropped the prefactor in the integral since we are mostly interested in the shape of the diffraction pattern rather than its absolute intensity.

Condition for Fraunhofer Diffraction

- We showed that if

$$\frac{k(x^2 + y^2)}{2R} \ll \pi$$

  we get Fraunhofer Diffraction, where the phase of the contributing waves varies linearly across the aperture.
- For an aperture of maximum extent D, this condition becomes

$$R \gg \frac{D^2}{\lambda}$$

- For example, a slit of width $D = 1\,\mathrm{mm}$ illuminated by a laser light with $\lambda = 589\,\mathrm{nm}$ will show Fraunhofer diffraction if $R \gg 1.7\,\mathrm{m}$.
- A radio telescope, 12-m in diameter, operating at a wavelength of 1 mm, will produce a Fraunhofer pattern at a distance $R \gg 144\,\mathrm{km}$.

Fraunhofer Diffraction in 1D

A 1-D aperture consists of patterns which are extended in the x direction, such as slits or diffraction gratings. In this case, the integral over x just gives a multiplicative constant, and the Fraunhofer integral becomes

$$\psi_P \propto \int_\Sigma h(y)\, \mathrm{e}^{-\mathrm{i}ky_0y/R}\, \mathrm{d}y.$$

Noting $y_0 = R\sin\theta$, and defining $q = k\sin\theta$, we find

$$\psi_P \propto \int h(y)\, \mathrm{e}^{-\mathrm{i}ky\sin\theta}\, dy = \int h(y)\, \mathrm{e}^{-\mathrm{i}qy}\, \mathrm{d}y \qquad (87)$$

So, we can see that ψ_P (as a function of y_0, or of angle θ) is related to $h(y)$ by a Fourier transform. Specifically

$$\psi_P(q) \propto \mathrm{FT}\{h(y)\}.$$

y and q are *reciprocal coordinates* which play the same role as time and frequency in the Fourier transforms you have seen previously. By analogy, you can see that the spatial frequencies present in the diffraction pattern are determined by the form of the aperture function (and vice versa).

Fraunhofer Examples: Young's Slits

Consider an aperture consisting of 2 narrow (δ-function) slits a distance D apart, spaced evenly about the origin

$$h(y) = \delta(y + D/2) + \delta(y - D/2).$$

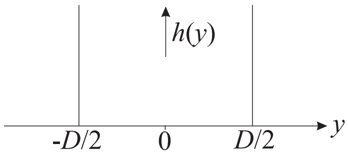


Figure 87: Two narrow slits

$$\therefore \psi_P = \int [\delta(y + D/2) + \delta(y - D/2)]\, \mathrm{e}^{-\mathrm{i}qy}\, \mathrm{d}y$$

Remembering that

$$\int \delta(x - x_0) f(x)\, dx = f(x_0)$$

we obtain

$$\psi_P(q) \propto \left(\mathrm{e}^{-\mathrm{i}qD/2} + \mathrm{e}^{\mathrm{i}qD/2}\right) = 2\cos\left(\frac{qD}{2}\right). \tag{88}$$

The intensity therefore varies as

$$I_P(q) = I_0 \cos^2(qD/2). \tag{89}$$

Note that the smaller the spacing between the slits, the larger the spacing between maxima in the diffraction pattern.

Fraunhofer Examples: 3 narrow slits

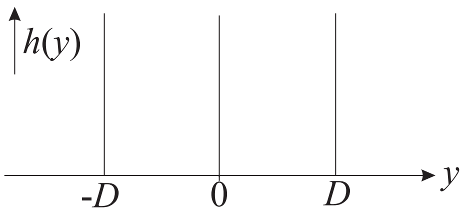


Figure 88: 3 slits

$$h(y) = \delta(y + D) + \delta(y) + \delta(y - D)$$

$$\therefore \psi_P \propto \mathrm{e}^{\mathrm{i}qD} + 1 + \mathrm{e}^{-\mathrm{i}qD} = 1 + 2\cos(qD)$$

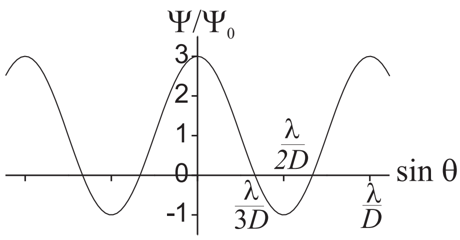


Figure 89: 3 slits: diffracted amplitude

$$I_P(q) = I_0(1 + 2\cos(qD))^2.$$

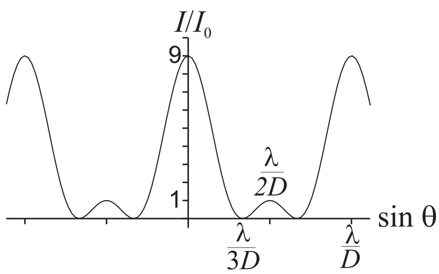


Figure 90: 3 slits: diffracted intensity.

The diffraction intensity has a regular pattern showing one subsidiary maximum between the primary maxima. We can also do this with a phasor diagram:

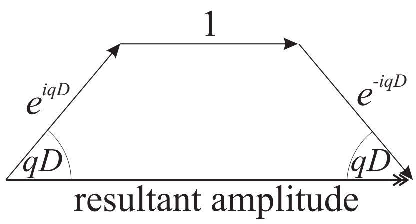


Figure 91: Phasor diagram for diffraction by 3 narrow slits.

Frauhofer Example: Narrow slits - a grating 1

The aperture function for N narrow slits with spacing D is

$$h(y) = \sum_{m=0}^{N-1} \delta(y - mD)$$

hence

$$\psi_P \propto 1 + \mathrm{e}^{-\mathrm{i}qD} + \mathrm{e}^{-\mathrm{i}2qD} + \mathrm{e}^{-\mathrm{i}3qD} + \ldots + \mathrm{e}^{-\mathrm{i}(N-1)qD}.$$

This is the sum of a simple geometric series, giving

$$\psi_P \propto \frac{1 - \mathrm{e}^{-\mathrm{i}NqD}}{1 - \mathrm{e}^{-\mathrm{i}qD}} = \mathrm{e}^{-\mathrm{i}(N-1)qD/2}\left(\frac{\mathrm{e}^{\mathrm{i}NqD/2} - \mathrm{e}^{-\mathrm{i}NqD/2}}{\mathrm{e}^{\mathrm{i}qD/2} - \mathrm{e}^{-\mathrm{i}qD/2}}\right) = \mathrm{e}^{-\mathrm{i}(N-1)qD/2}\ \frac{\sin(NqD/2)}{\sin(qD/2)}$$

so

$$I_p = I_0 \left[\frac{\sin(NqD/2)}{\sin(qD/2)}\right]^2. \tag{90}$$

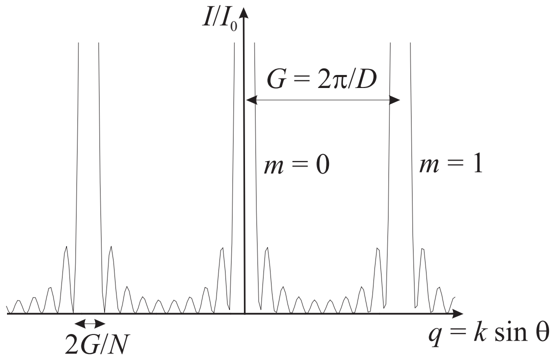


Figure 92:

Frauhofer Example: Narrow slits - a grating (2)

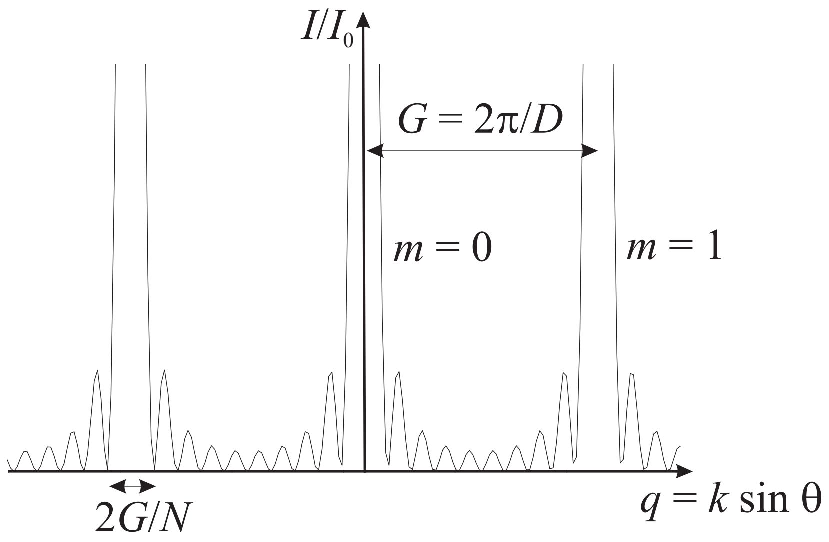


Figure 93: Diffracted intensity from a finite grating ($N = 10$). Note that the primary maximum (cut off in the graph) is a factor of ~ 20 larger than the first subsidiary maximum.

This function has primary maxima when $\sin(qD/2) = 0$, i.e. when $q = 2m\pi/D$ where m is an integer. It also has subsidiary maxima whenever $\sin(NqD/2) = \pm 1$, i.e. where $q = (2M+1)\pi/(ND)$, where M is an integer. For a grating with N slits, there will be $(N-2)$ subsidiary maxima and $(N-1)$ zeros between primary maxima. The larger the value of N, the larger the ratio between primary maxima and subsidiary maxima.

Frauhofer Example: Narrow slits - a grating (3)

When light of a single wavelength is incident on a grating with a large number of slits, diffraction will therefore occur into well-defined directions defined by

$$k \sin\theta = q = 2m\pi/D = mG.$$

where $G = 2\pi/D$.
Schematically (for $N \to \infty$)

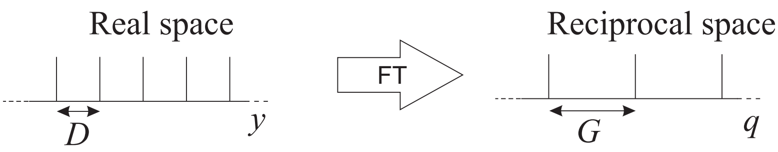


Figure 94:

For a grating with N slits, the width of the primary maxima is given by the position of the zero of $\sin(NqD/2)$, i.e. at

$$q = \frac{2\pi}{ND} = \frac{G}{N}. \tag{91}$$

Increasing the width of the grating, and hence N, will therefore give narrower diffraction peaks. Due to the finite width of the diffraction peaks, illumination at a single wavelength will produce diffraction at a range of angles. This width will limit the chromatic resolving power of the grating — its capacity to separate waves of different wavelengths.

Fraunhofer Example: Single wide aperture

The slit allows light to pass in a width a:

$$h(y) = \begin{cases} 1 & \text{for} \quad |y| < a/2 \\ 0 & \text{for} \quad |y| > a/2. \end{cases}$$

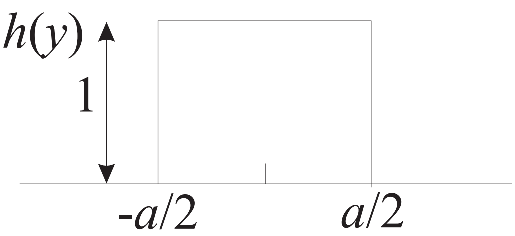


Figure 95: A single slit.

The diffracted amplitude is

$$\psi_P(q) \propto \int_{-a/2}^{a/2} \mathrm{e}^{-iqy}\mathrm{d}y = \frac{a\sin(qa/2)}{qa/2} \tag{92}$$

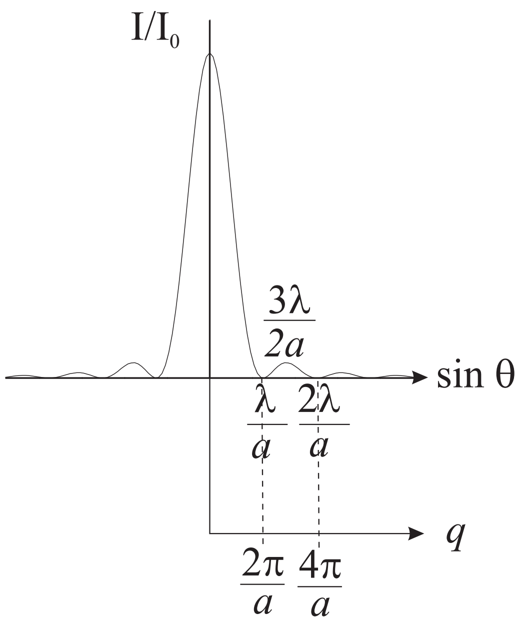


Figure 96: Diffracted intensity for a single slit.

So the intensity observed on the screen is

$$I_p(q) \propto a^2 \,\mathrm{sinc}^2\left(\frac{qa}{2}\right)$$

where, as before, $q = k\sin\theta = (2\pi/\lambda)\sin\theta$.

More complicated apertures

Since the Fraunhofer diffraction pattern is the Fourier transform of the aperture function, we can use Fourier methods to find patterns of complex apertures. One example is the diffraction pattern of two wide slits. The aperture function is the convolution of two delta functions (two narrow slits) with a top-hat function (wide aperture).

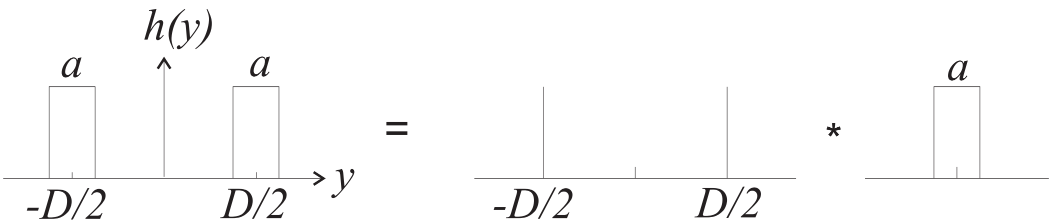


Figure 97: Two slits as convolution of a slit with two delta functions.

The diffraction pattern will simply be the product of the Fourier transforms of these two functions, each of which we have evaluated previously. For two slits of width a with their centres separated by a distance D, the diffraction pattern is the product of equ. (88) with equ. (92):

$$\psi_P \propto \cos\left(\frac{qD}{2}\right) a\,\mathrm{sinc}\left(\frac{qa}{2}\right)$$

Two Slit Pattern

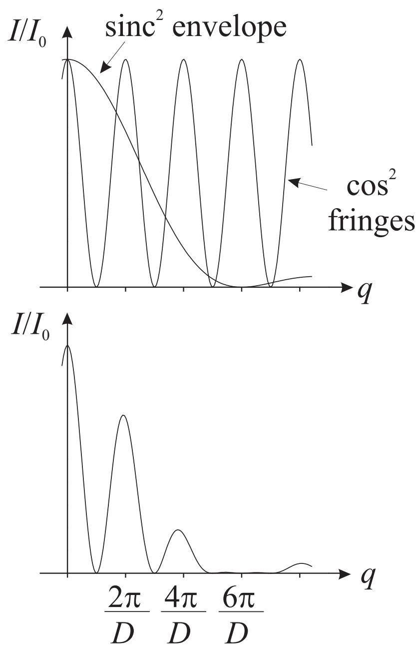


Figure 98: $\cos^2$ fringes modulated by a sinc ^{2} envelope, for $D = 3a$.

The $\cos^2$ fringes from the two slits are now modulated by a sinc ^{2} envelope function due to the finite width of the slits. In the plot, $D = 3a$, and the third fringe is almost completely eliminated by the zero in the envelope function. In general, this leads to missing orders where a maximum expected in the underlying function coincides with a minimum in the envelope function.

General Grating Diffraction Pattern

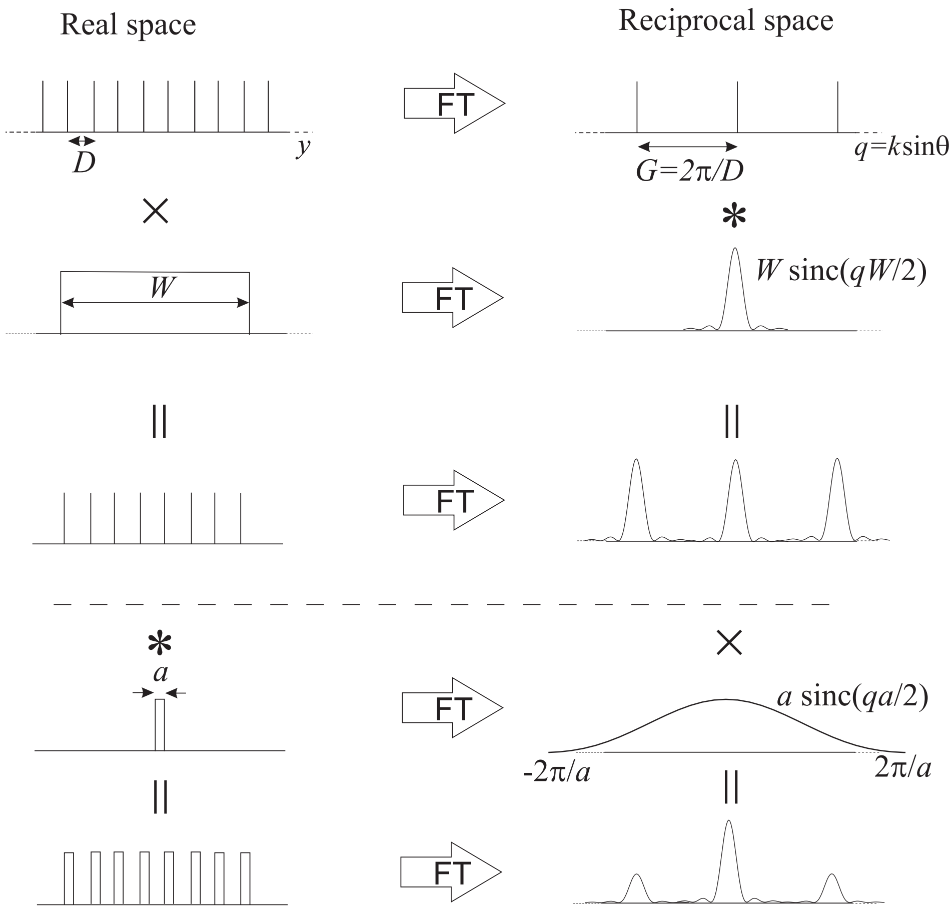


Figure 99: Diffraction from a grating of finite width with finite slit size.

Fraunhofer Diffraction: Complex Aperture Function

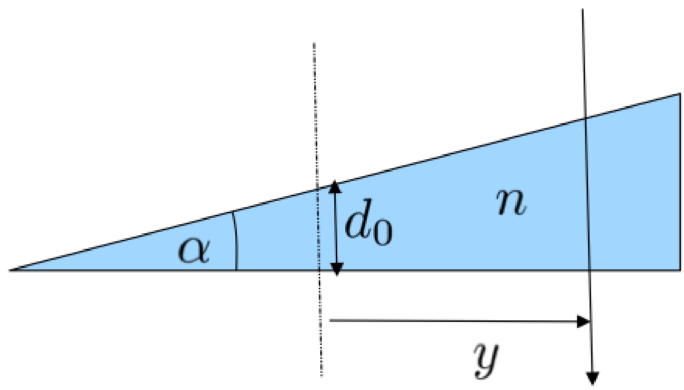


Figure 100: A Glass Wedge

Consider placing a wedge of glass in front of the aperture. What is its effect? It introduces a phase shift proportional to position y across the wedge:

$$\phi = k(n-1)\alpha y + \text{constant}$$

if α is small. So the aperture function is multiplied by a varying phase term $\exp(\mathrm{i}k(n-1)\alpha y)$.

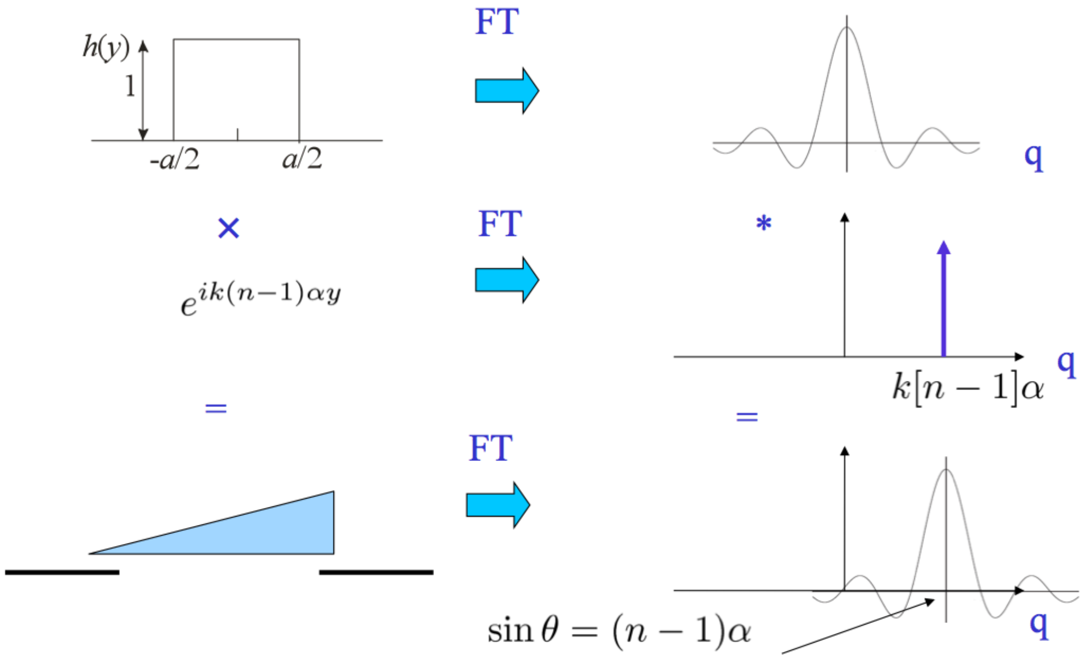


Figure 101: A Glass Wedge shifts the diffraction pattern

As expected, the pattern is shifted by an angle $(n-1)\alpha$. This is an example of the shift theorem for Fourier transforms: translation of a function in the space domain introduces a linear multiplicative phase term to its Fourier transform (and vice versa).

Conditions for observing Fraunhofer diffraction

So far we have placed the source and the observation plane at a large distance from the aperture — this means that the waves arriving at the aperture and at the observation screen can be considered to be plane waves.
This can be inconvenient, so alternative arrangements need to be made to reach the Fraunhofer limit. We can use lenses on either side of the aperture to convert the diverging beam from the source into a parallel beam, and to focus the diffracted parallel beams onto an observation screen. The source and observation screen must be placed at the focal points of the lenses.

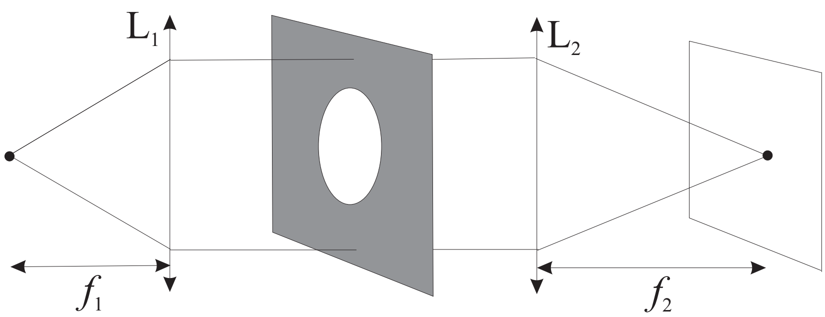
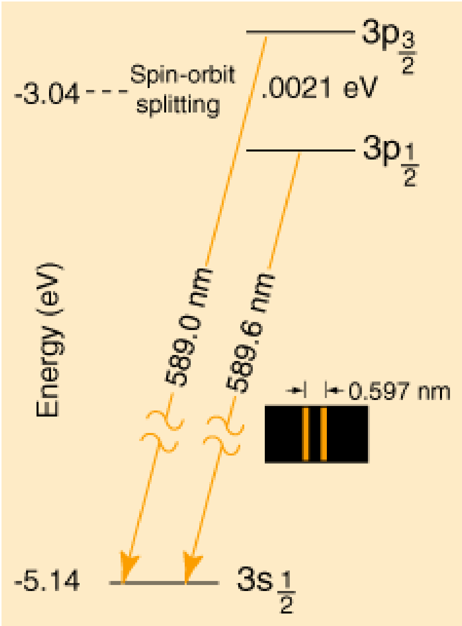


Figure 102: Use of lenses to produce Fraunhofer far-field conditions for diffraction.

Plane waves leaving the aperture at an angle θ are focused to a point on the screen at a location $f_2 \sin\theta$: so the lens turns the angles into positions on the screen.

Spectral Line Emission

Spectral lines arise from transitions between quantum states (electronic, vibrational, rotational). Optical emission lines typically arise from electronic transitions in atoms.



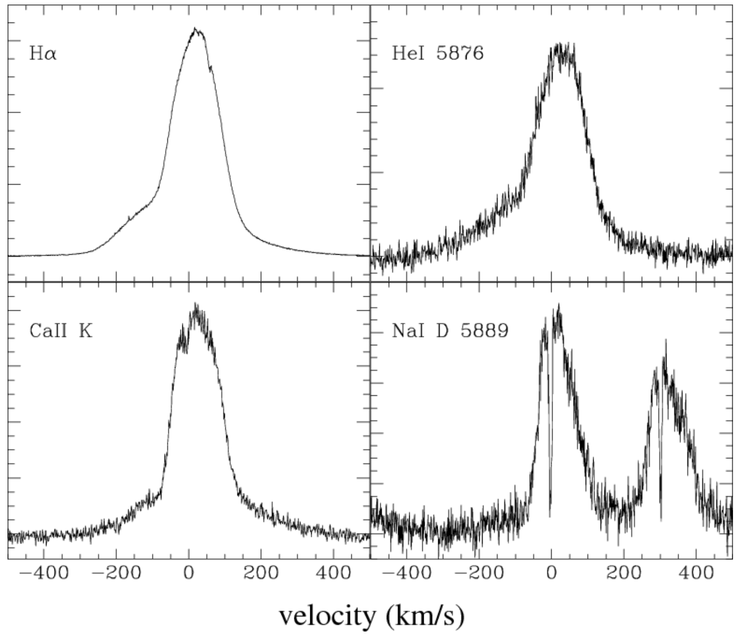


Figure 103: The Sodium emission doublet at 589nm, and spectrum of a young star. Note the x axis is labelled with velocity derived from the red/blue shift of the spectral line: $u = c(\omega - \omega_0)/\omega_0$, where ω is the observed frequency and ω_0 the rest frequency.

Natural line width

An electronic state in an atom has a finite lifetime (typically $\tau_L \sim 1 \times 10^{-8}$ s.) The uncertainty principle tells us that there is thus a (small) uncertainty in energy. The spectral line emitted thus has finite width. Typically, this contributes $\Delta\omega \sim (1/\tau_L) \sim 1 \times 10^8\,\mathrm{rad\,s^{-1}}$ to the linewidth in discharge lamps.

The electric field takes the form

$$E(t) = E_0 \exp(-\gamma t) \cos \omega_0 t$$

for $t > 0$. The Fourier transform (see table on H155) yields a Lorentzian line shape:

$$I(\omega) \propto \frac{1}{(\omega - \omega_0)^2 + \gamma^2} \qquad (93)$$

which has a half width at half maximum of $\Delta\omega = \gamma \sim (1/\tau_L)$. (Recognise this equation as the power response of a driven oscillator with γ being the damping term).

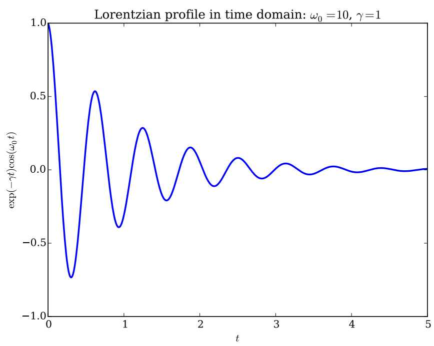


Figure 104: Lorentzian (time domain)

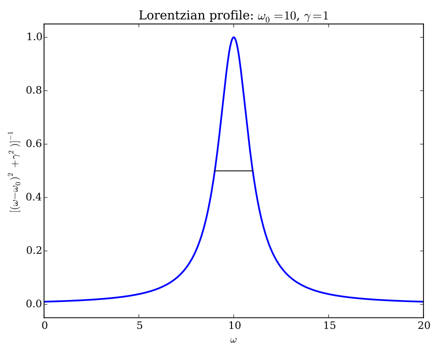
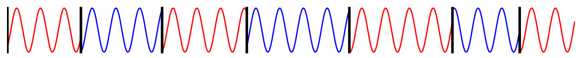


Figure 105: Lorentzian Power Spectrum

Collision (or pressure) broadening

Collisions between atoms while they are emitting limits the coherence of the emitted light waves. The effective lifetime of an excited state is reduced by the collisions.



The mean time between collisions in a gas is $\tau_c \sim 1/(n\Sigma u)$ where n is the number density of particles, Σ is the collision cross section, and u is the particle speed. (For air at ambient pressure and temperature, this evaluates to $\tau_c \sim 2 \times 10^{-10}$ s. The collision cross section can exceed the physical cross section).
Hence the range of frequencies emitted $\Delta\omega$ will be $\sim (1/\tau_c) \sim n\Sigma u$. It turns out the line shape is also Lorentzian, with $\Delta\omega$ taking the place of (or in fact adding to) γ.

$$\Delta\omega \sim n\Sigma u$$

So for air at sea level, $\Delta\omega \sim (1/\tau_c) \sim 1 \times 10^{10}$ rad s^{-1}. In most high pressure atmospheres, pressure broadening is much larger than natural broadening.

Doppler (or thermal) broadening

An atom will in general be in motion when it emits light. This causes a Doppler shift, so the observed frequency ω is related to the rest-frame frequency ω_0 by

$$\omega = \omega_0 \sqrt{\frac{1 + u_x/c}{1 - u_x/c}} \approx \omega_0 \left(1 + \frac{u_x}{c}\right)$$

where u_x is the velocity of the atom along the line of sight x, and the approximation is true for $u_x \ll c$. So $u_x \approx c(\omega - \omega_0)/\omega_0$.
The Boltzmann distribution tells us the 1-D velocity distribution is $p(u_x) \propto \exp(-mu_x^2/(2k_BT))$ where m is the atomic mass (see Thermodynamics course, Physics B). The line shape is thus given by

$$I(\omega) \propto \exp\left(-\frac{mc^2(\omega - \omega_0)^2}{2\omega_0^2 k_B T}\right) \propto \exp\left(-\frac{(\omega - \omega_0)^2}{2\sigma_\omega^2}\right) \qquad (94)$$

The line is Gaussian with standard deviation

$$\sigma_\omega = \omega_0 \sqrt{\frac{k_B T}{mc^2}} \qquad (95)$$

In air at room temperature, for a line with $\lambda = 0.5\,\mu$m, this yields $\sigma_\omega \sim 4 \times 10^9$ rad s^{-1}. This is less than pressure broadening at sea level; but at high altitudes, the Doppler broadening can exceed pressure broadening.

Gaussian vs Lorentzian Profiles

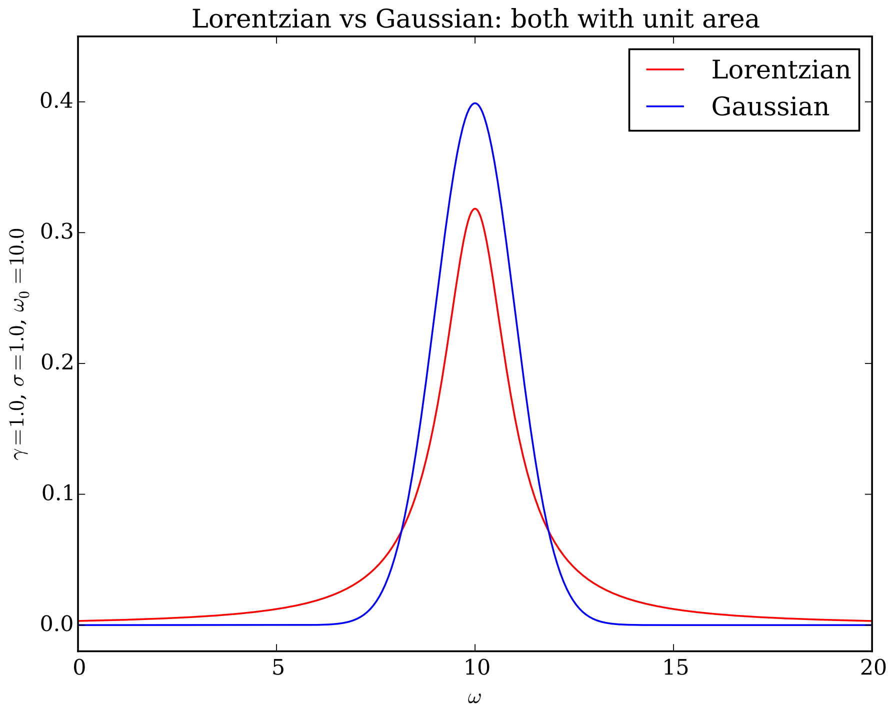


Figure 106: Comparison of Gaussian and Lorentzian profiles

Note the slow fall off ($\propto (\Delta\omega)^{-2}$) of the Lorenztian profile.
In most general case, we expect lines to have a profile equal to the convolution of the Lorentzian and a Gaussian profile.

Grating spectrometers

In a grating spectrometer, a diffraction grating is used to measure the spectrum of incident light. Concave mirrors are often used rather than lenses (since they suffer less chromatic abberation).

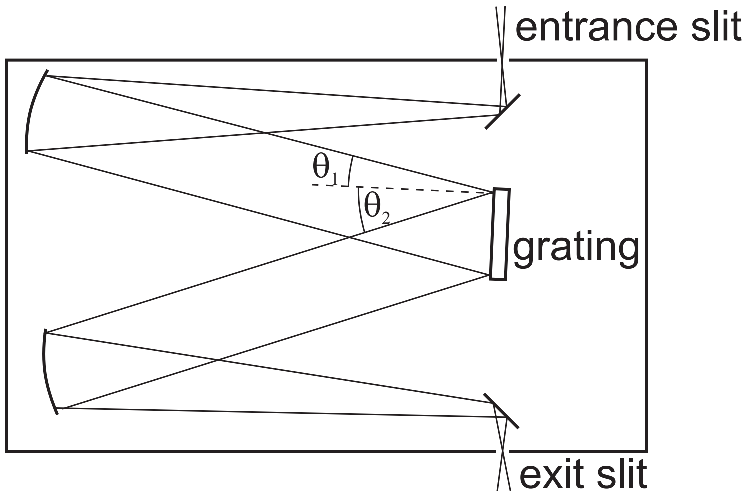


Figure 107: A grating spectrometer.

Incident light is focussed onto a narrow entrance slit before encountering the mirror, thus producing a well-defined angle of incidence, θ_1 on the grating. Light is diffracted through an angle which depends on its wavelength, and on the spacing of lines in the grating D, according to

$$D(\sin\theta_2 - \sin\theta_1) = m\lambda \tag{96}$$

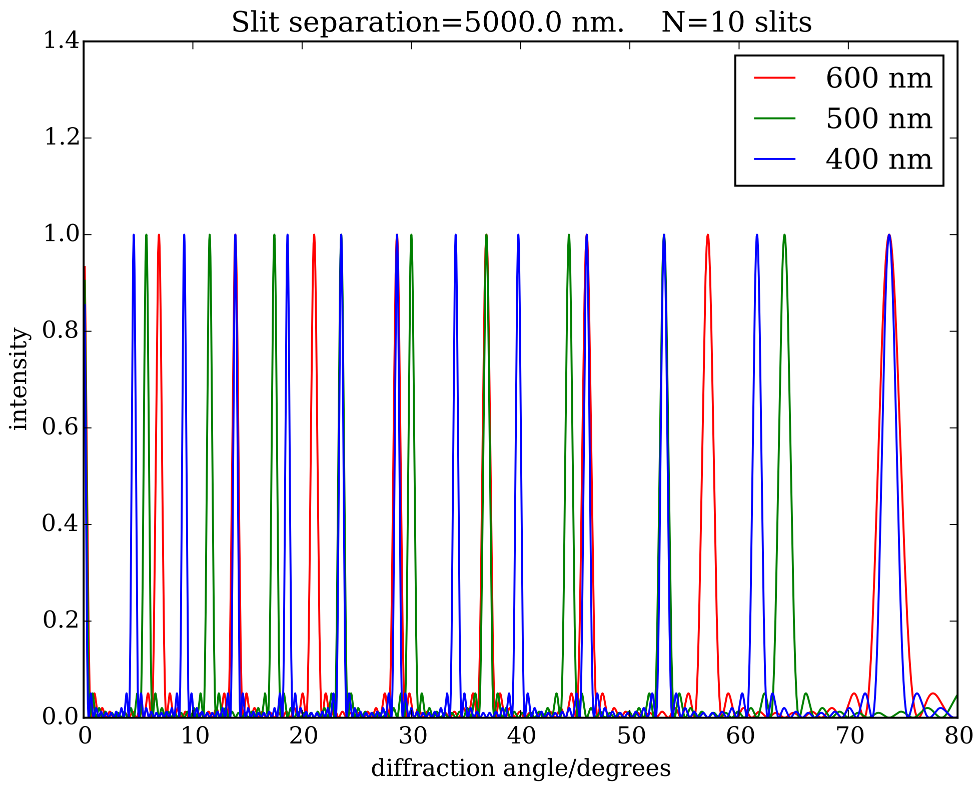


Figure 108:

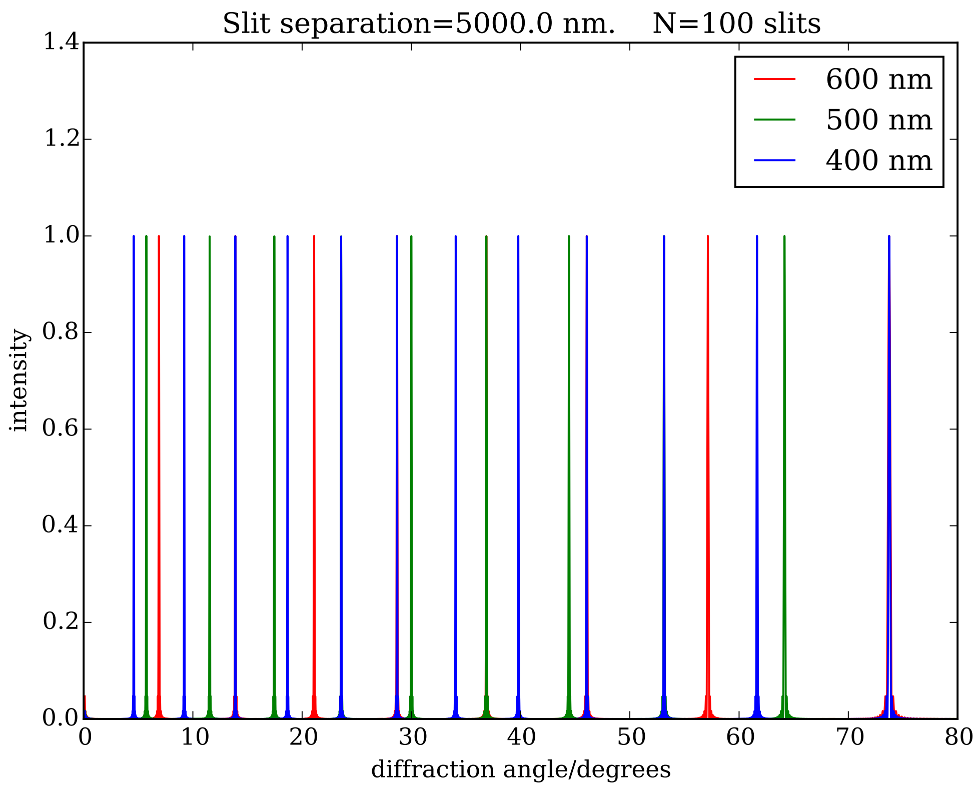


Figure 109:

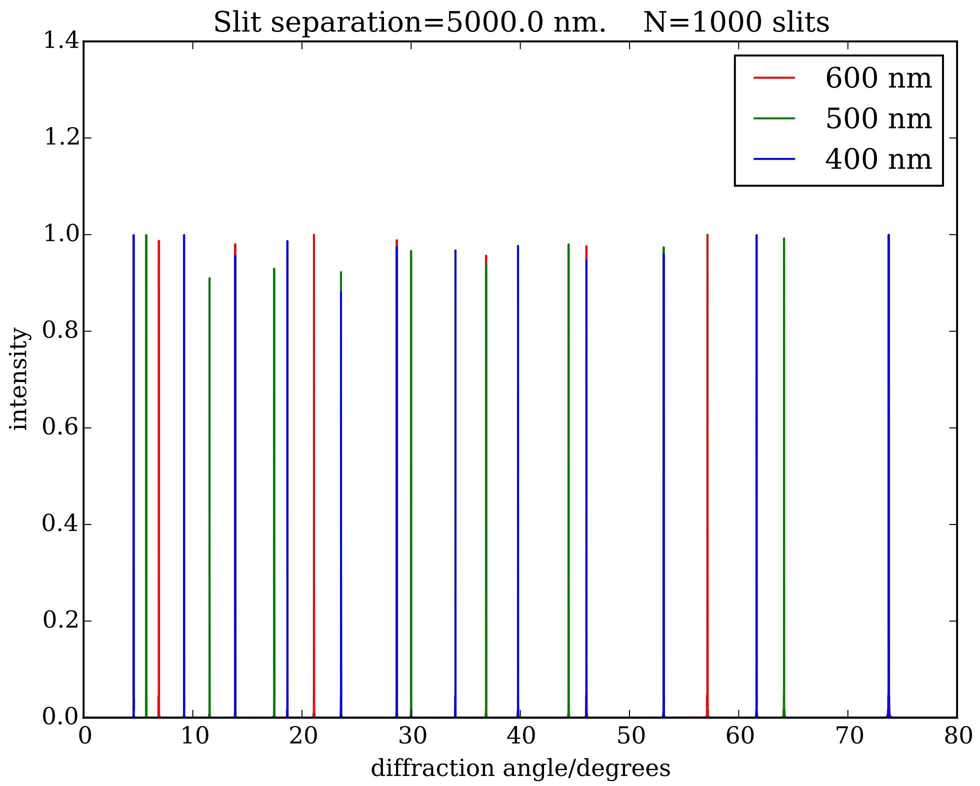


Figure 110:

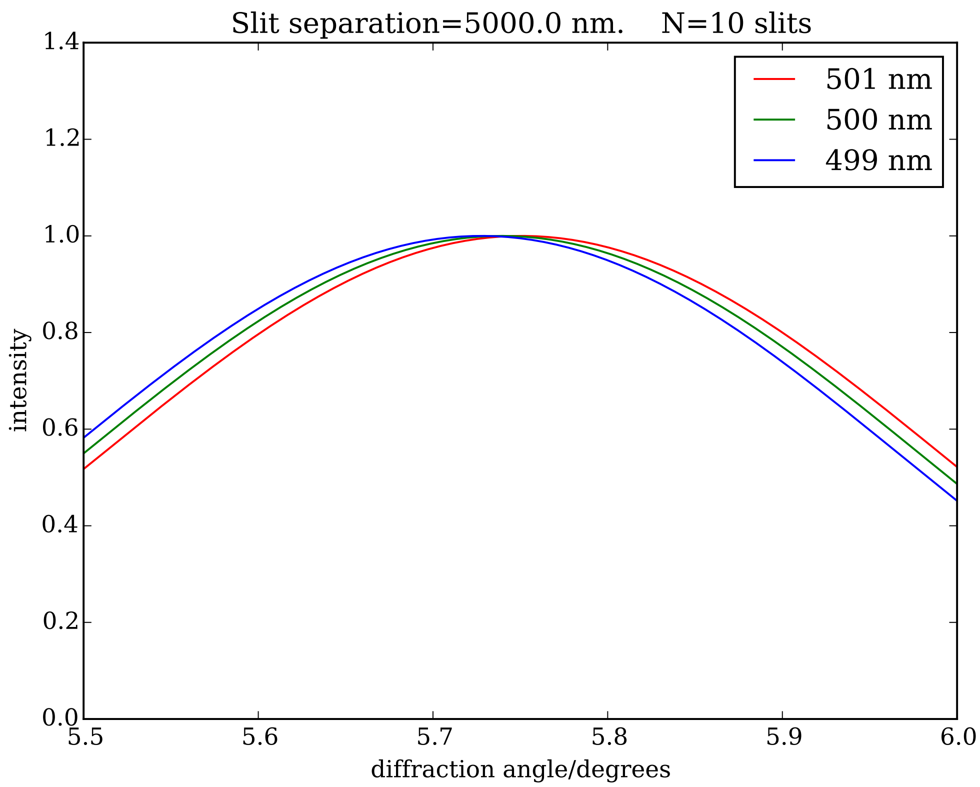


Figure 111:

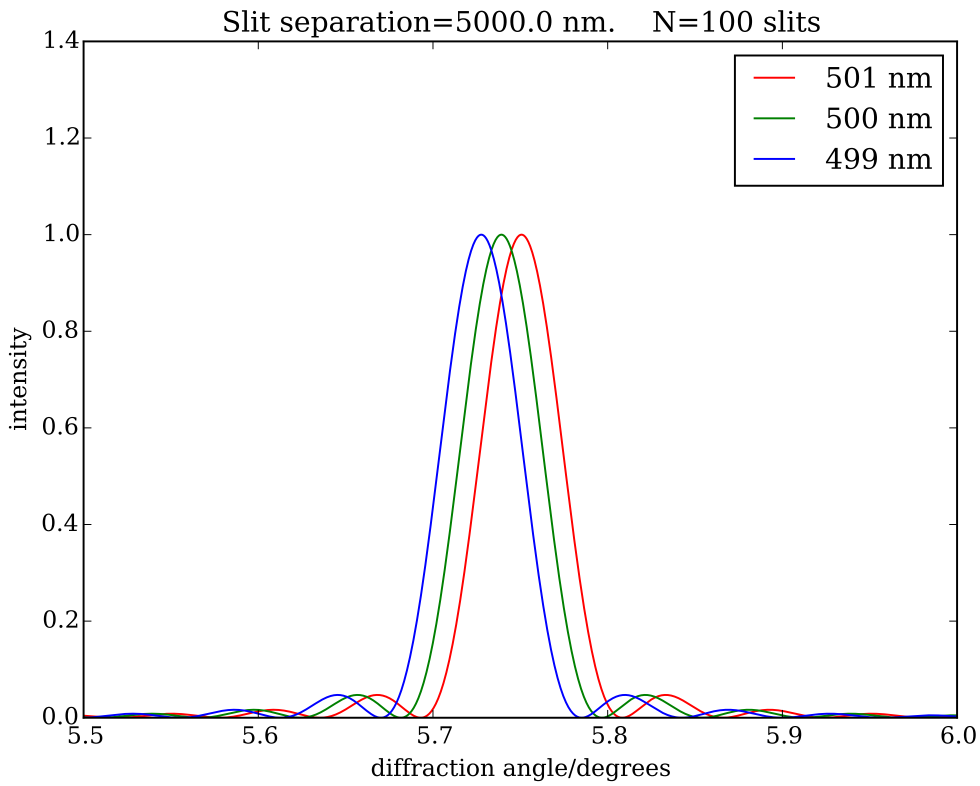


Figure 112:

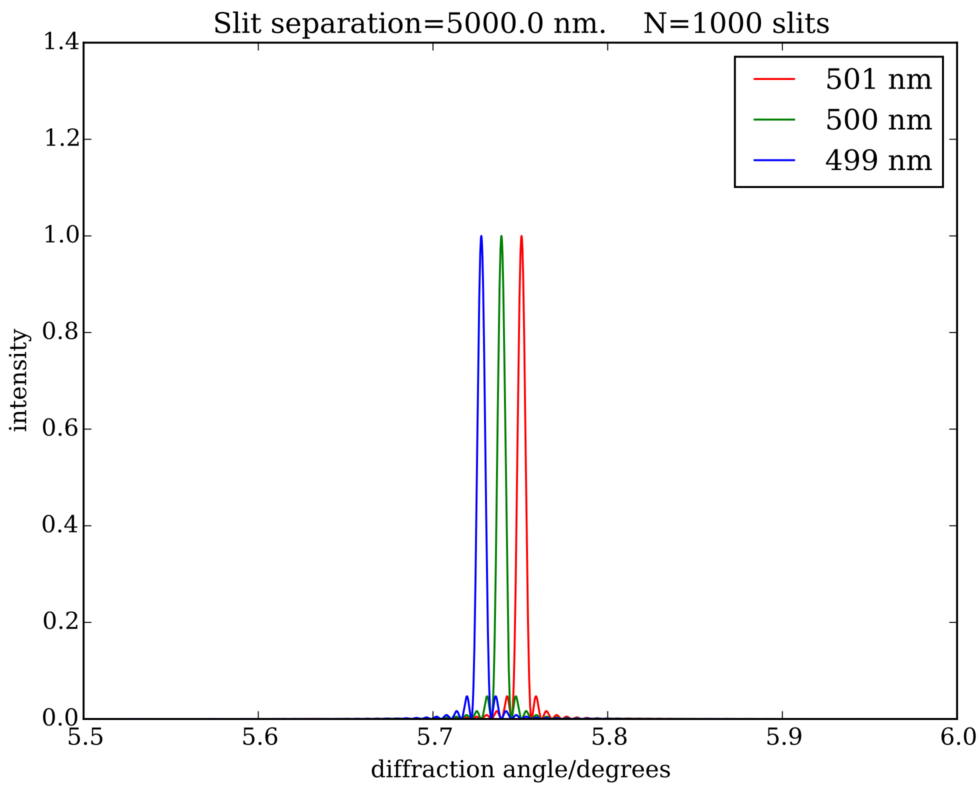


Figure 113:

Spectrometer Resolution

With sufficiently small entrance- and exit slits, the resolution of the spectrometer will eventually become limited by the diffraction properties of the grating. Light of a particular wavelength λ will now produce a well-defined angle of incidence, but for a grating of finite width the peaks in the diffraction pattern will have a finite width. If we assume for simplicity that $\theta_1 = 0$, then we have $D\sin\theta = m\lambda$, where $\theta = \theta_2$. For illumination at two wavelengths λ and $\lambda + \delta\lambda$, diffraction peaks will be produced at angles θ_λ and $\theta_{\lambda+\delta\lambda}$, where

$$D\sin\theta_\lambda = m\lambda \quad \text{and} \quad D\sin\theta_{\lambda+\delta\lambda} = m(\lambda + \delta\lambda)$$

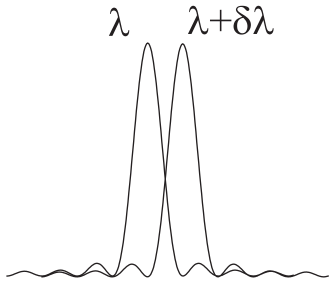


Figure 114:

Grating Resolution

For λ, the first minimum in the diffracted beam profile (from equ. (91)) occurs at

$$\sin\theta = \frac{m\lambda}{D} + \frac{\lambda}{W}$$

where $W = (N-1)D \approx ND$ is the width of the grating.
The wavelengths can be resolved if the maximum for $\lambda + \delta\lambda$ coincides with this minimum for λ — this is the Rayleigh Criterion, i.e. if

$$m\lambda + \frac{\lambda D}{W} = m(\lambda + \delta\lambda)$$

so that

$$\boxed{\frac{\lambda}{\delta\lambda} = \frac{mW}{D} = mN} \tag{97}$$

is the chromatic resolving power of the grating. This is a measure of how well the grating separates light of two wavelengths. $\delta\lambda$ is the minimum wavelength difference (near λ) which the grating can effectively distinguish.
The greater the number N of slits/lines, the higher the resolving power – wider gratings give narrower diffraction beams which are easier to distinguish.
The order m acts as a kind of "gearing" – the widths of the diffracted beams are fixed (by W), but their angular separation increases with m, so it is easier to distinguish wavelengths at higher order.

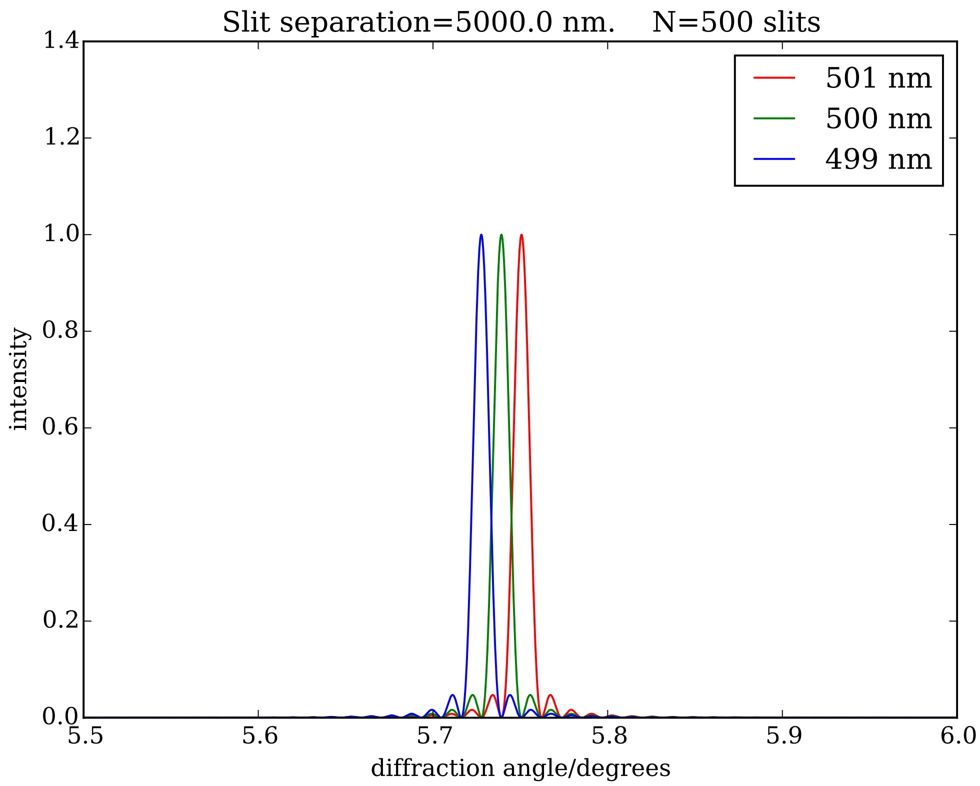


Figure 115: Rayleigh criterion

Fraunhofer Diffraction from 2D Apertures

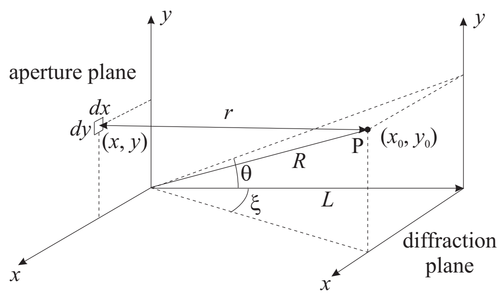


Figure 116: Geometry for Fraunhofer diffraction.

For a two-dimensional aperture $h(x, y)$, the diffraction integral reads

$$\psi_P \propto \iint_\Sigma \psi_\Sigma \, h(x, y) \, \exp\left[-\mathrm{i}k\left(\frac{x_0 x + y_0 y}{R}\right)\right] \mathrm{d}x \, \mathrm{d}y.$$

Two (small) angles θ and ξ, can be used to specify the pattern, with

$$\sin\theta \approx \theta \approx y_0/R; \quad \sin\xi \approx \xi \approx x_0/R$$

Writing $q = k\sin\theta$ and $p = k\sin\xi$, the Fraunhofer integral is

$$\boxed{\psi_P(p, q) \propto \iint_\Sigma h(x, y) \, \mathrm{e}^{-\mathrm{i}(px+qy)} \, \mathrm{d}x \, \mathrm{d}y.}$$

This is the two-dimensional Fourier transform of $h(x, y)$.

Rectangular aperture

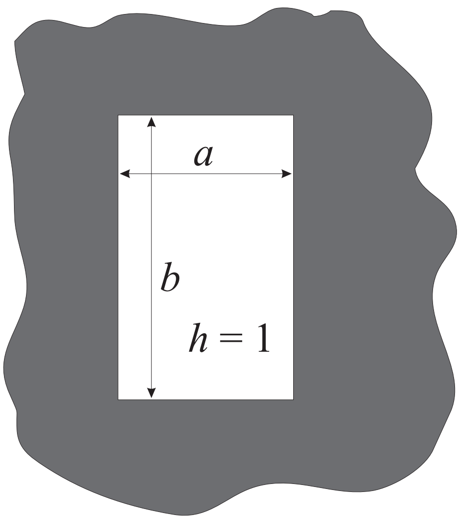


Figure 117: Rectangular aperture.

It is easier to evaluate the transform if $h(x, y)$ is separable, i.e. if $h(x, y) = f(x)g(y)$. The rectangle clearly is:

$$\begin{aligned} h(x,y) &= 1, \;\; |x| < a/2 \text{ and } |y| < b/2 \\ &= 0, \;\; \text{otherwise.} \end{aligned}$$

The diffraction pattern therefore becomes

$$\begin{aligned} \psi_P &\approx \int_{-a/2}^{a/2} \mathrm{e}^{-\mathrm{i}px}\,\mathrm{d}x \int_{-b/2}^{b/2} \mathrm{e}^{-\mathrm{i}qy}\,\mathrm{d}y \\ &= a \,\mathrm{sinc}\left(\frac{pa}{2}\right) b \,\mathrm{sinc}\left(\frac{qb}{2}\right). \end{aligned} \qquad (98)$$

Fraunhofer Pattern for a Rectangular Aperture

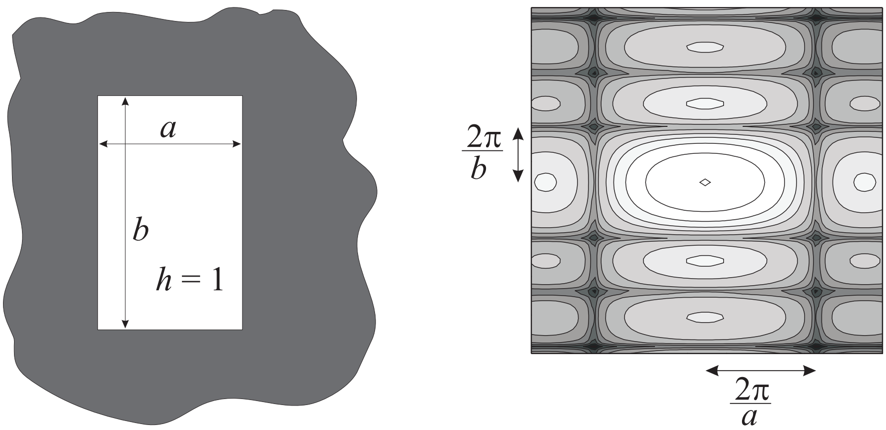


Figure 118: Fraunhofer diffraction pattern (intensity) from a rectangular aperture. Logarithmic shading is used for clarity.

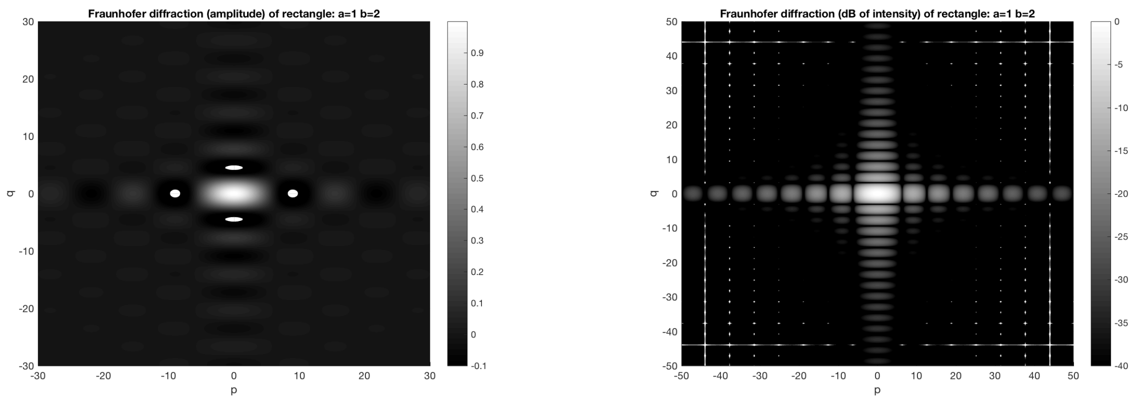


Figure 119: Fraunhofer diffraction pattern from a rectangular aperture. Amplitude on the left, log intensity on the right.

Fraunhofer Diffraction from a Circular Aperture

Many apertures in optical systems (telescopes, microscopes, cameras, etc.) are circular. Here h is a function of radius, ρ, and for an aperture of diameter d:

$$h(\rho) = 1 \text{ for } \rho < (d/2), \quad 0 \text{ otherwise}$$

h is not separable in x and y; the integral evaluates to (see eg Lipson: Optical Physics):

$$\boxed{\psi(\theta) \propto \frac{\psi_0 d^2}{2} \frac{J_1\left(\frac{kd\sin\theta}{2}\right)}{\left(\frac{kd\sin\theta}{2}\right)} = \frac{\psi_0 d^2}{2} \frac{J_1\left(\frac{\pi d\sin\theta}{\lambda}\right)}{\left(\frac{\pi d\sin\theta}{\lambda}\right)}}$$

where J_1 is the 1st order Bessel Function of the first kind. $J_1(x)$ has first zero at $x \approx 3.83$, so diffracted intensity has a zero at $\sin\theta \approx 1.22\lambda/d$. The region inside this first zero is the Airy disc; it contains 86% of the total energy flux.

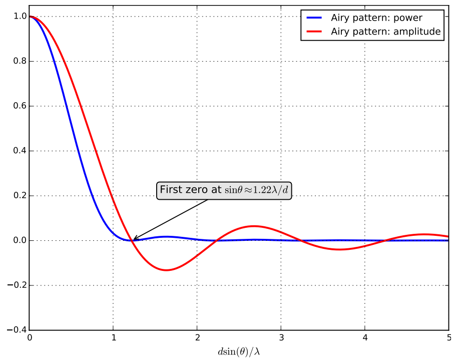


Figure 120: The diffraction pattern from a circular hole, an example of a point spread function.

Resolution of Optical Instruments

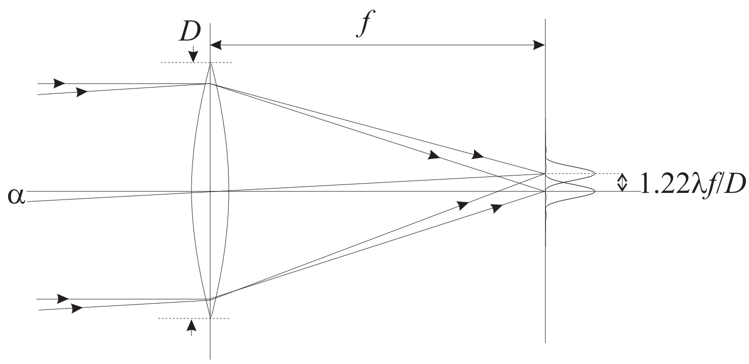


Figure 121: Resolution of a telescope.

- In geometrical optics, a perfect lens should image a point object to a point image.
- In physical optics, the finite circular extent of the lens creates diffraction.
- So an incoming plane wavefront from a distant point object will produce an Airy disc pattern in the image plane
- Its angular radius is $\alpha \approx 1.22\lambda/D$ where D is the diameter of the objective lens/mirror (assuming that there are no other apertures within the system which further obstruct the wavefronts).
- The actual radius in the image plane will be $\frac{1.22\lambda}{D} f$ where f is the focal length of the lens.

Rayleigh Criterion for Resolving Two Point Objects

- The Rayleigh criterion defines the angular resolution of the telescope.
- Two distant point objects with angular separation α will produce Airy discs in the image plane with their centres separated by a distance $f\alpha$. These objects can only be resolved if the the radius of the Airy discs is less than their separation, hence the angular resolution is $\frac{1.22\lambda}{D}$.
- This is the best you can do with a system of aperture D. If a telescope makes stellar images of this size, we say it is diffraction limited.

Large aperture $\rightarrow$ high angular resolution

Rayleigh's Criterion for Resolving Two Close Objects

Note this is an incoherent addition of the sources: add intensities not amplitudes.

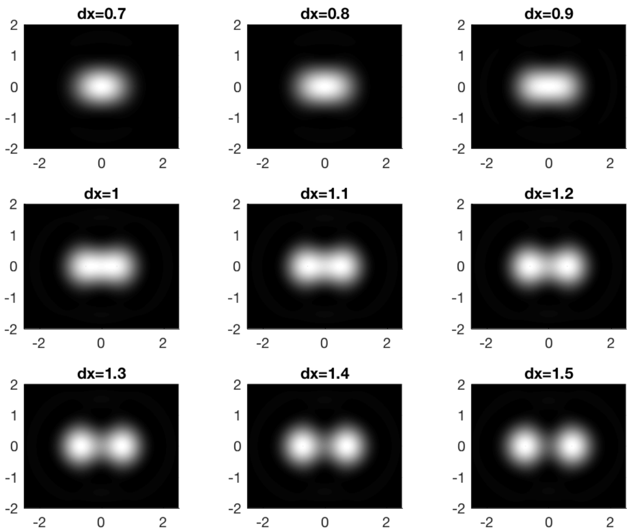


Figure 122: Rayleigh Criterion: a spacing of $1.2\lambda/D$ just resolves the objects if equally bright.

The Resolution of the Human Eye

- The human eye, iris diameter $D \approx 5\,\text{mm}$, with visible light, $\lambda \approx 400\,\text{nm}$ (in aqueous humour) has a diffraction limit of $\alpha \approx 100\,\mu\text{rad} \approx 20''$. [Note: $5\,\mu\text{rad} \approx 1''$].
- Is your eyesight diffraction limited?
- The eye's colour sensitive detectors (cones) have a peak density of $\approx 1 \times 10^{11}\,\text{m}^{-2}$, corresponding to a spacing of about $3\,\mu\text{m}$. The focal length is $f \approx 15\,\text{mm}$ so that the angular sampling by the cones is $\approx 200\,\mu\text{rad} \approx 40''$. This is about twice the diffraction limit.
- 20/20 vision corresponds to resolving two lines $60'' \approx 0.3\,\text{mrad}$ apart.

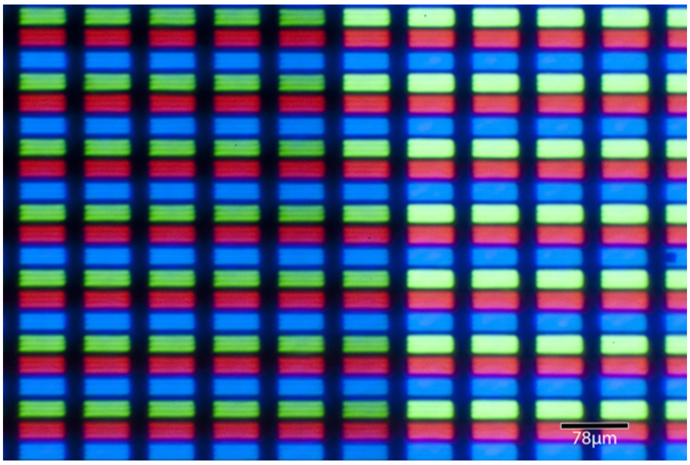


Figure 123: eye, and iPhone 4 "retina" display

The $\approx 78\,\mu\text{m}$ pixels subtend $30''$ at 50 cm distance from your eye. details here.

Examples of Diffraction-Limited Resolution: Telescopes

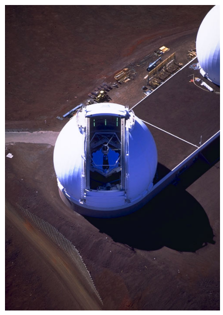


Figure 124: Lovell Telescope at Jodrell Bank; ALMA, Chile; Keck, Hawaii

- Jodrell Bank radio telescope: metal dish reflector: $D \approx 64\,\text{m}$: radio waves: $\lambda \approx 60\,\text{mm}$: $\alpha \approx 1 \times 10^{-3}\,\text{rad} \approx 225''$.
- ALMA radio telescope array: metal dish reflector: $D \approx 12\,\text{m}$: radio waves: $\lambda = 0.5\,\text{mm}$: $\alpha = 5 \times 10^{-5}\,\text{rad} \approx 10''$. But when combined with other dishes spread out over a 16 km area of desert, $D_{\text{eff}} = 16\,\text{km}$ and $\alpha = 4 \times 10^{-8}\,\text{rad} = 0.007''$.
- Keck optical telescope: optical reflector: $D = 10\,\text{m}$: visible light: $\lambda \approx 500\,\text{nm}$: $\alpha \approx 6 \times 10^{-8}\,\text{rad} = 0.01''$. But not diffraction limited in practice due to atmospheric turbulence, which limits resolution to $0.4''$ or so.

Babinet's Principle

Consider two complementary apertures a and b.

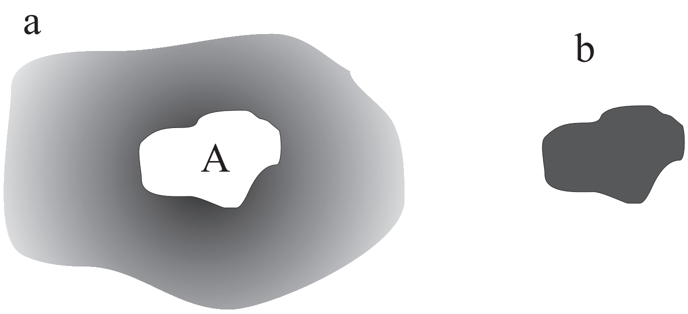


Figure 125: Complementary apertures.

The diffracted amplitudes are

$$\psi_a \propto \iint_A \mathrm{e}^{-\mathrm{i}(px+qy)}\,\mathrm{d}x\,\mathrm{d}y$$

and

$$\psi_b \propto \iint_{\text{all space}} \mathrm{e}^{-\mathrm{i}(px+qy)}\,\mathrm{d}x\,\mathrm{d}y - \iint_{\mathrm{A}} \mathrm{e}^{-\mathrm{i}(px+qy)}\,\mathrm{d}x\,\mathrm{d}y \propto \delta(p,q) - \psi_a.$$

The diffracted intensities are therefore the same, except at the origin ($p = 0$, $q = 0$), the direction corresponding to the direction of the incident beam.

Fresnel Diffraction

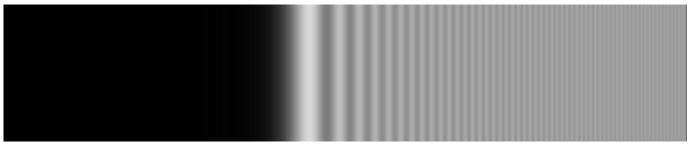


Figure 126: Fresnel Diffraction at an Edge

The diffraction integral becomes a Fourier transform to a good approximation when the pattern is observed beyond some critical distance if

$$\frac{k(x^2+y^2)}{2R} \ll \pi.$$

In this regime, the phase of the wavelets in the integral varies linearly with position x, y in the aperture. This critical distance (sometimes called the Rayleigh distance, L_R) is

$$L_R = \rho^2/\lambda$$

where $\rho^2 = x_{\max}^2 + y_{\max}^2$, i.e. ρ is a measure of the maximum dimension of the aperture. When the observing screen is much further than L_R from the aperture, the quadratic phase terms and higher in the diffraction integral can be ignored, and Fraunhofer diffraction occurs.
For distances of order L_R or smaller, we have to include the quadratic and possibly higher-order terms in the phase of the integrand in the diffraction integral.This is Fresnel Diffraction. The physics is no harder; the maths is a bit messier.

Geometry for Fresnel Diffraction

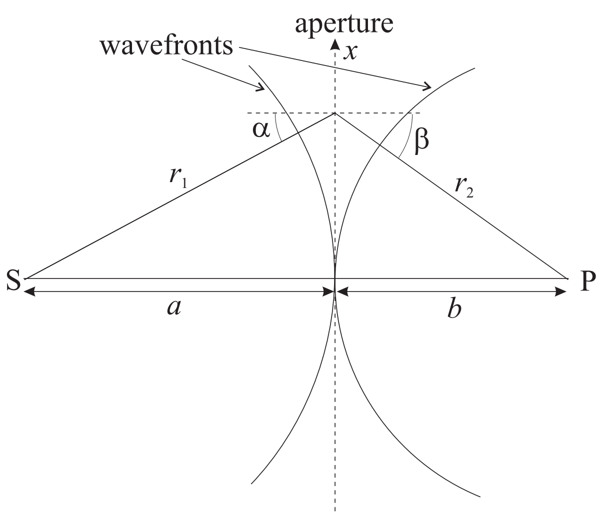


Figure 127: Geometry for Fresnel diffraction.

Let's examine the diffracted intensity at points P which are on the axis defined by the source S and the origin point O in the aperture. On axis, we have $x_0 = 0$ and $y_0 = 0$, and so the linear term in the phase is zero. By changing our choice of origin we can find the pattern at off-axis locations.

The path from S to P via the aperture element at (x, y) is

$$\begin{aligned} r_1 + r_2 &= \sqrt{a^2+x^2+y^2} + \sqrt{b^2+x^2+y^2} \\ &= a + b + \frac{x^2+y^2}{2a} + \frac{x^2+y^2}{2b} \\ &\quad + \text{higher order terms} \end{aligned}$$

Writing

$$\frac{1}{R} = \frac{1}{a} + \frac{1}{b}$$

we obtain

$$\text{optical path} = \text{const.} + \frac{x^2+y^2}{2R}$$

giving a diffracted amplitude of

$$\psi_P \propto \iint_\Sigma \frac{h(x,y)K(\theta)\exp\left(ik\frac{x^2+y^2}{2R}\right)}{r_1 r_2}\, dx\, dy.$$

Fresnel Diffraction Formula

We now assume

- that the angles to the edge of the aperture are small enough that we can neglect the obliquity factor, and thus take $K(\theta) = 1$.
- that the variations in r_1 and r_2 over the aperture are negligible as far as the denominator is concerned: $r_1 \sim a$ and $r_2 \sim b$; but note variations in $(r_1 + r_2 - a - b)$ are not negligible compared with λ, and so have a significant effect on the phase term.

Under these approximations, we have this expression for the diffracted beam on-axis under Fresnel diffraction conditions:

$$\psi_P(0,0) \propto \iint_\Sigma h(x,y) \exp\left(\mathrm{i}k\frac{x^2+y^2}{2R}\right) \mathrm{d}x\,\mathrm{d}y. \qquad (99)$$

In general this is harder to evaluate compared to the Fraunhofer diffraction integral, but we can gain some insights if the apertures are simple.

Fresnel Diffraction from a Rectangular Aperture (1)

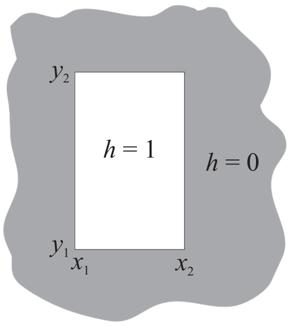


Figure 128: A rectangular aperture.

Consider the case of a uniform rectangular aperture. The aperture function is separable, and the the diffraction integral equ. (99) is

$$\begin{aligned}\psi_P &\propto \iint_\Sigma h(x,y) \exp\left(\mathrm{i}k\frac{x^2+y^2}{2R}\right) \mathrm{d}x\,\mathrm{d}y \\ &\propto \int_{x_1}^{x_2} \exp\left(\frac{\mathrm{i}kx^2}{2R}\right) \mathrm{d}x \int_{y_1}^{y_2} \exp\left(\frac{\mathrm{i}ky^2}{2R}\right) \mathrm{d}y.\end{aligned}$$

It is convenient to use the dimensionless variables u and v:

$$u = x\sqrt{\frac{2}{\lambda R}}; \quad v = y\sqrt{\frac{2}{\lambda R}} \qquad (100)$$

Fresnel Diffraction from Rectangular Aperture (2)

The previous result can then be expressed as

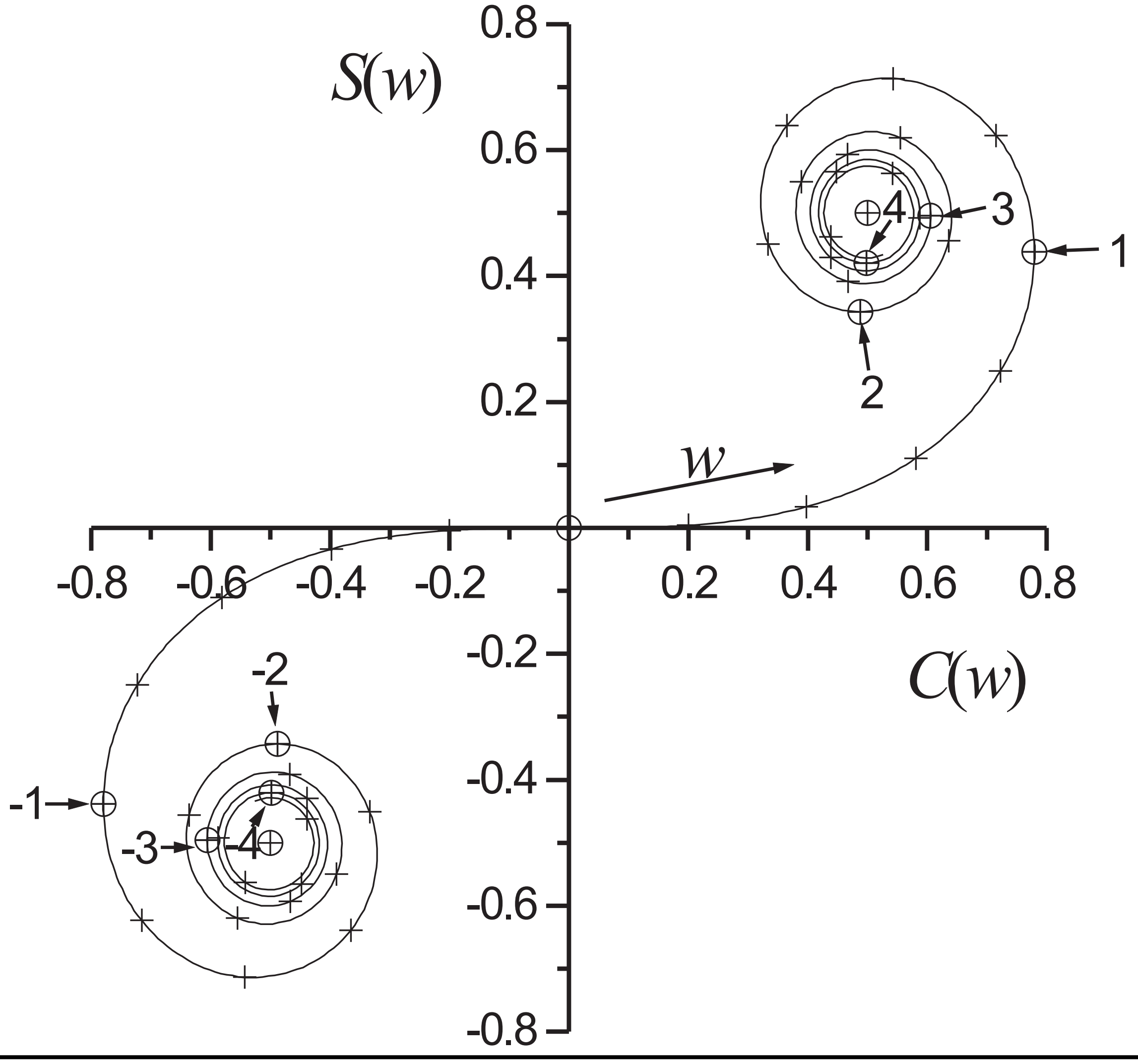
$$\psi_P \propto \int_{u_1}^{u_2} \exp\left(\frac{\mathrm{i}\pi u^2}{2}\right) \mathrm{d}u \int_{v_1}^{v_2} \exp\left(\frac{\mathrm{i}\pi v^2}{2}\right) \mathrm{d}v.$$

It is convenient to define the Fresnel Integrals

$$C(w) = \int_0^w \cos\left(\frac{\pi u^2}{2}\right) \mathrm{d}u \tag{101}$$

$$S(w) = \int_0^w \sin\left(\frac{\pi u^2}{2}\right) \mathrm{d}u \tag{102}$$

This integrals must be evaluated numerically, but can be represented graphically in the complex plane. A particular value of w determines a point $C(w) + iS(w)$ in the complex plane. The locus of these points is known as the Cornu spiral.



The Cornu Spiral: Properties

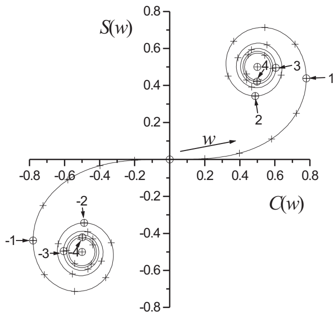


Figure 130: The Cornu spiral.

- The arc length l along the curve between two points w_1 and w_2 is equal to $w_2 - w_1$, i.e. w determines the distance from the origin measured along the curve:

$$\mathrm{d}l^2 = \mathrm{d}C^2 + \mathrm{d}S^2 = \mathrm{d}w^2[\cos^2(\pi w^2/2) + \sin^2(\pi w^2/2)] = \mathrm{d}w^2; \quad \therefore \mathrm{d}l = \mathrm{d}w$$

- The radius of curvature R_C at a point w is given by

$$\frac{1}{R_C} = \frac{|C'S'' - S'C''|}{(C' + S')^{3/2}} = \pi w$$

- The curve spirals in towards the point $(C(\infty), S(\infty)) = (0.5, 0.5)$.
- The curve is odd: $C(-w) = -C(w)$, and $S(-w) = -S(w)$.
- The curve is vertical when $w = 1, \sqrt{3}, \sqrt{5}, \ldots$, i.e. $w = \sqrt{2m+1}$; and horizontal when $w = 0, \sqrt{2}, \sqrt{4}, \ldots$ i.e. $\sqrt{2m}$
(the slope of the tangent vector is: $S'(w)/C'(w) = \tan(\pi w^2/2)$).

Fresnel Integrals

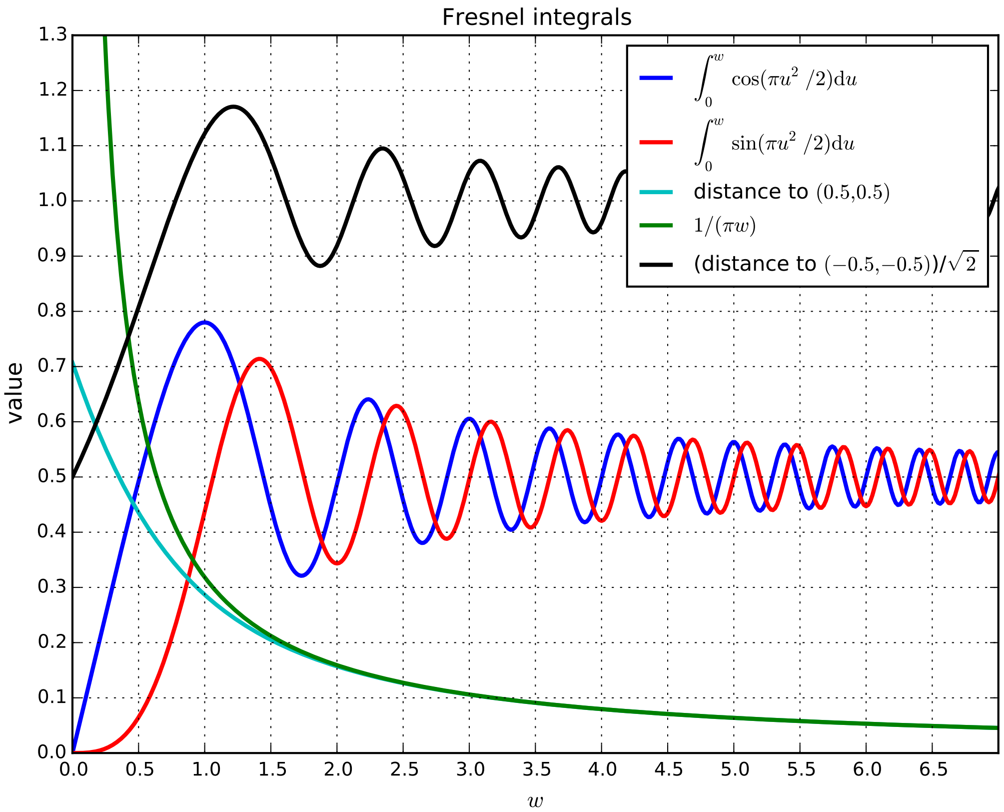


Figure 131: The Fresnel integrals

Using the Cornu spiral

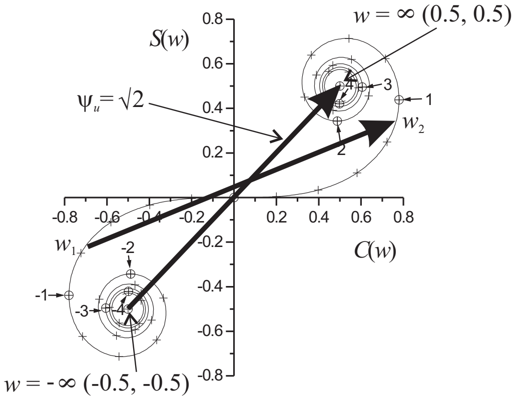


Figure 132: Diffraction amplitudes are given by the vector spanning the Cornu spiral.

We can use the Cornu spiral to determine the diffraction pattern from simple apertures. Consider a slit or edge extending in the y direction. The relevant diffraction integral is

$$\psi_P \propto \int_{w_1}^{w_2} \exp\left(\frac{\mathrm{i}\pi u^2}{2}\right) \mathrm{d}u = [C(w_2) - C(w_1)] + \mathrm{i}\,[S(w_2) - S(w_1)] \tag{103}$$

where

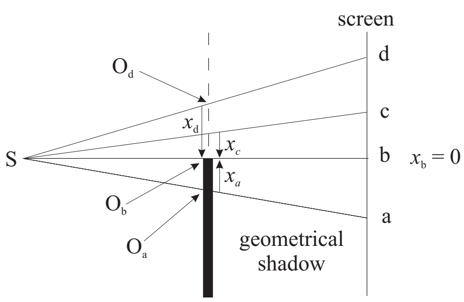
$$w_1 = x_1\sqrt{\frac{2}{\lambda R}}\ ; \quad w_2 = x_2\sqrt{\frac{2}{\lambda R}}.$$

This result is equivalent to the vector spanning the Cornu spiral from location $C(w_1) + iS(w_1)$ to location $C(w_2) + iS(w_2)$

We normalise by the amplitude resulting from an unobstructed wavefront, given by ψ_u, the spanning vector from $w = -\infty$ to $w = \infty$, which has length $\sqrt{2}$. The diffracted intensity is proportional to the square of the length of the spanning vector.

Fresnel Diffraction Pattern at an Edge (1)

Consider a wavefront obstructed by a straight edge.



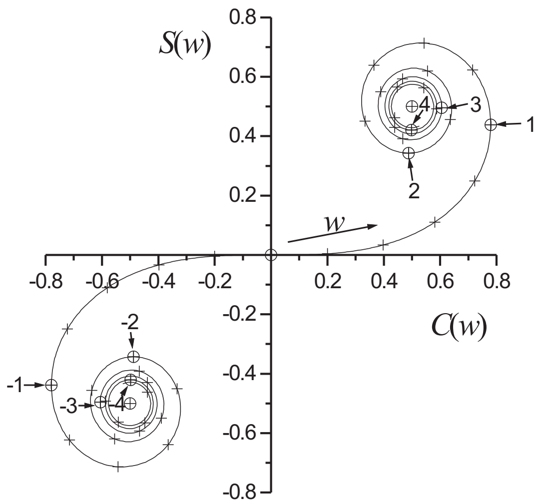


Figure 133: Wavefront obstructed by a straight edge.

Define the origin $\mathrm{O_b}$ in the aperture plane to be at the edge of the obstruction. Fresnel conditions are satisfied since S, $\mathrm{O_b}$ and b are in a straight line. The diffracted wave at b is calculated from equ. (103), integrating from $x = 0$ to $x = \infty$, i.e. $w_1 = 0$, $w_2 = \infty$.

$$\Psi_p(b) = [C(\infty) - C(0)] + \mathrm{i}\,[S(\infty) - S(0)] = (0.5, 0.5)$$

This is half the amplitude if there was no obstruction, which is $(1, 1)$; the intensity is thus a quarter of the unobstructed intensity at the edge of the geometrical shadow.

Fresnel Diffraction Pattern at an Edge (2)

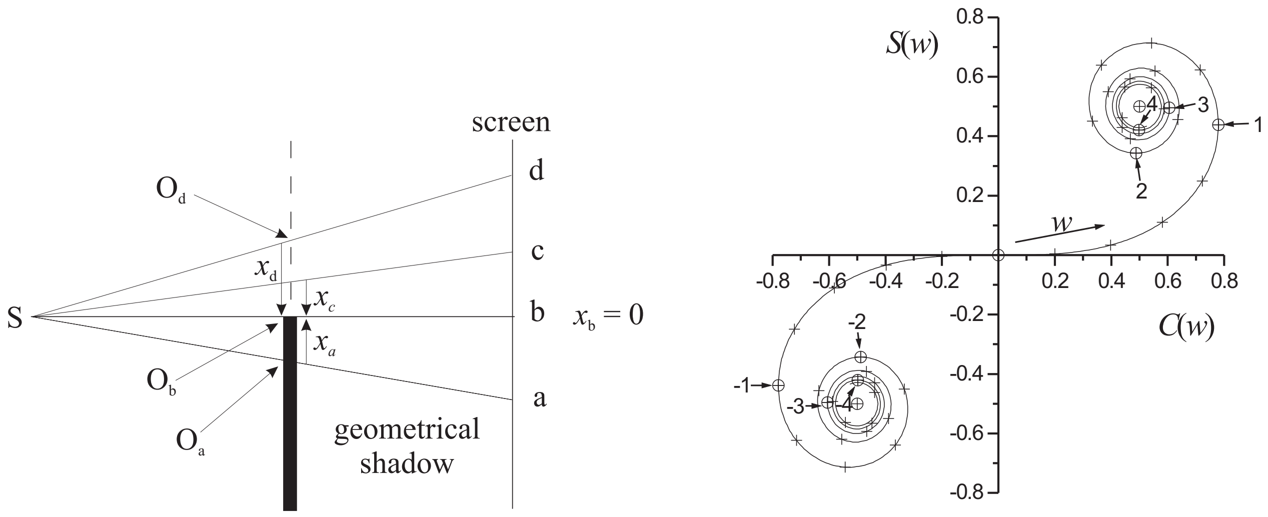


Figure 134: Wavefront obstructed by a straight edge.

To calculate the diffracted wave at other points on the screen, we move the origin so that it is still between S and our observation point and Fresnel conditions are still satisfied. For observation point c, for example, the integral over the aperture is then from $x = x_c$ (which is negative) to $x = \infty$, i.e. $x_1 = -x_c$, $x_2 = \infty$. In summary:

a	$x_2 = \infty$	$x_1 = x_a > 0$	
b	$x_2 = \infty$	$x_1 = x_b = 0$	the geometrical edge
c	$x_2 = \infty$	$x_1 = x_c < 0$	
d	$x_2 = \infty$	$x_1 = x_d$; $w_d \sim -1.22$	for maximum $\lvert\psi_P\rvert$

Fresnel Diffraction Pattern at an Edge (3)

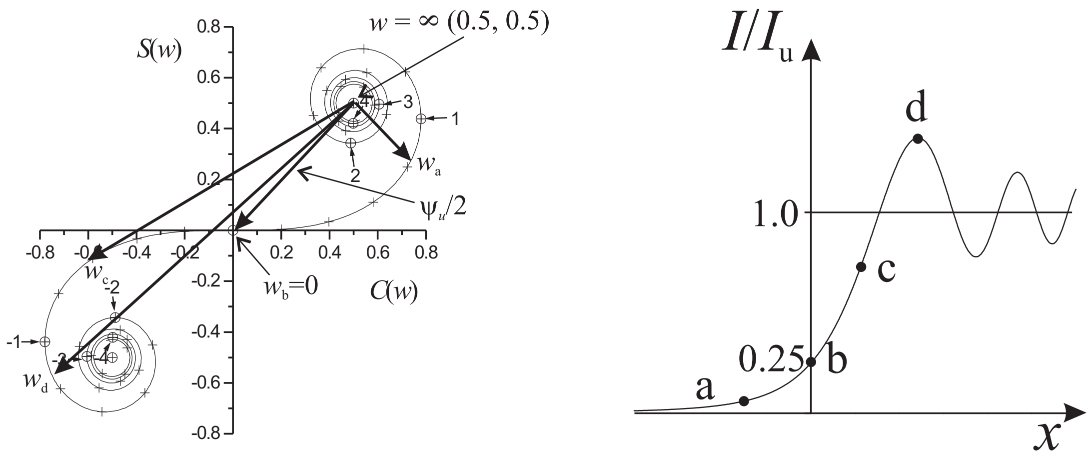


Figure 135: Diffraction amplitudes for diffraction by a straight edge.

The resulting diffracted waves are shown as vectors in the Cornu spiral, giving a diffraction pattern of the form shown. Note that

- well outside the shadow the intensity is same as for an unobstructed wavefront;
- the amplitude at the geometric edge is 50% of the unobstructed amplitude;
- the amplitude falls smoothly inside the geometric shadow, in fact roughly as $1/w$. Hence the intensity falls as $\sim 1/w^2$.
- there are fringes outside the geometric shadow, producing a maximum intensity 138% of I_u at d, just outside the edge.

Diffraction from a finite slit

Just like a single edge, but integral does not extend to infinity.

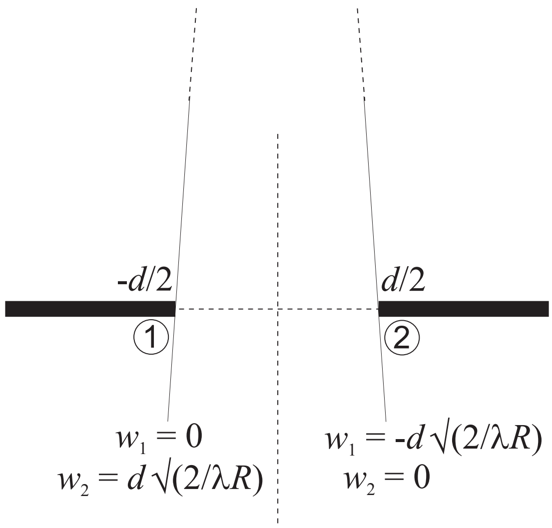


Figure 136: A slit as two straight edges.

For an observation point opposite one edge, $w_1 = 0$ and $w_2 = d\sqrt{(2/\lambda R)}$. For an observation point opposite the other edge, $w_1 = -d\sqrt{(2/\lambda R)}$, $w_2 = 0$.
For an arbitrary observation point, we integrate from w_1 to w_2, where
$w_1 - w_2 = \Delta w = d\sqrt{(2/\lambda R)}$.
So the spanning vector on the Cornu spiral is between two points separated by a fixed length along the curve.

Narrow Slit Pattern

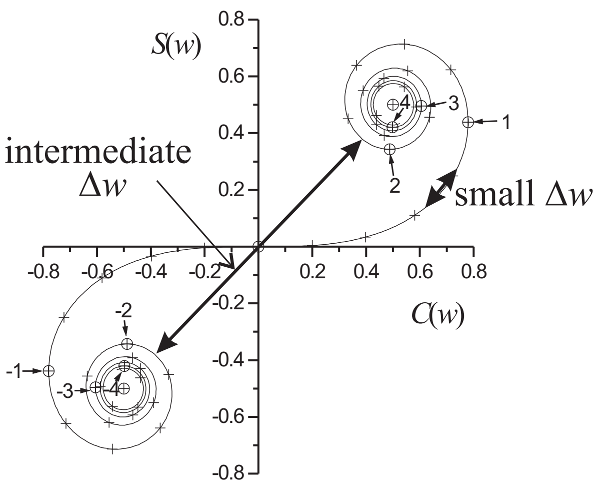


Figure 137: The Cornu spiral used to deduce the Fresnel diffraction pattern of a slit.

For small Δw, the spanning vector decreases monotonically in length as the observation point moves away from the centre of the slit, and the diffraction pattern has the form shown.

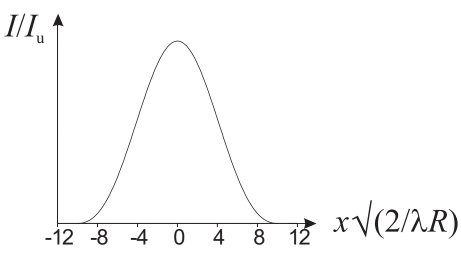


Figure 138: Fresnel diffraction for a narrow slit ($\Delta w = 0.2$).

Wide slit pattern

As $R \to 0$ or $d \to \infty$ (i.e. as the screen becomes "close" to the aperture), Δw becomes large, and we get two single edge patterns.

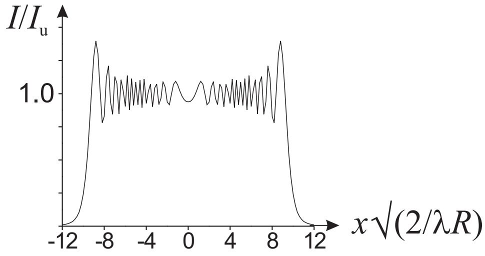


Figure 139: Fresnel diffraction for a wide slit ($\Delta w = 20$).

For intermediate situations, it may be that there is in fact a minimum on the axis of the system with fringing either side.

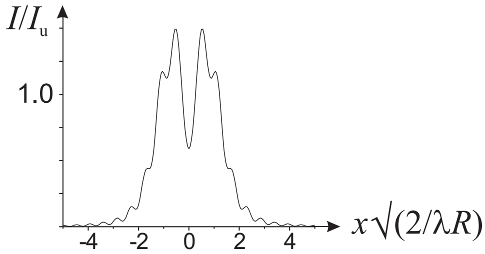


Figure 140: Fresnel diffraction for a narrow slit ($\Delta w = 3.8$).

The diffraction pattern changes as we move away from the aperture (since R changes).

Fresnel Diffraction from a Circular Aperture

Consider a circular aperture of radius r_a. The diffraction pattern is in general

$$\psi_P \propto \iint_\Sigma \frac{h(x,y)K(x,y)\exp\left(\mathrm{i}k\frac{x^2+y^2}{2R}\right)}{r_1 r_2}\,\mathrm{d}x\,\mathrm{d}y.$$

Let's retain the obliquity factor, and the variation in r_1 and r_2 across the aperture, but only consider the pattern on-axis. Dividing the aperture into annular elements of radius ρ and thickness $\mathrm{d}\rho$ we can replace $x^2 + y^2$ with ρ^2 and $\mathrm{d}x\,\mathrm{d}y$ with $2\pi\rho\,\mathrm{d}\rho$ to give

$$\psi_P \propto \int_{\rho=0}^{\rho=r_a} \frac{K}{(a^2+\rho^2)^{1/2}(b^2+\rho^2)^{1/2}} \exp\left(\frac{\mathrm{i}k\rho^2}{2R}\right) 2\pi\rho\,\mathrm{d}\rho.$$

It is convenient to make a substitution:

$$\rho^2 = s \; ; \quad 2\rho\,\mathrm{d}\rho = \mathrm{d}s$$

and to replace k by $2\pi/\lambda$, to give

$$\psi_P \propto \int_{s=0}^{s=r_a^2} \frac{K(s)}{(a^2+s)^{1/2}(b^2+s)^{1/2}} \exp\left(\frac{\mathrm{i}\pi s}{\lambda R}\right) \pi\,\mathrm{d}s. \tag{104}$$

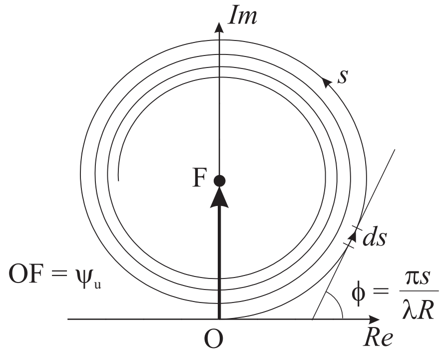


Figure 141: Phasor diagram for Fresnel diffraction and circular symmetry.

The integral can be evaluated graphically using the phasor diagram.

- the phase $\phi = \pi s/(\lambda R)$ varies linearly with s, and the elemental contributions to the integral are approximately ds, so as s increases the phasor diagram is (approximately) a circle
- the term before the exponential is a slowly decreasing function of s (the denominator obviously increases with s, and $K = \frac{1}{2}(\cos\theta_S + \cos\theta_P)$ decreases with s). Thus the radius of the circle in the phasor diagram gradually decreases with s.
- to calculate the diffracted amplitude, consider the spanning vector from point O to the point a distance s along the curve. The curve spirals inwards, so in the absence of an obstruction/aperture the integration range is from O to F, corresponding to $s = \infty$. OF is thus amplitude of the unobstructed wavefront.

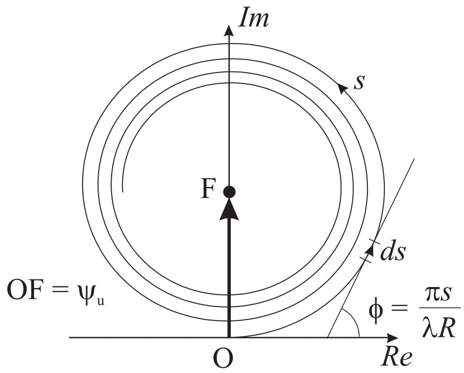


Figure 142: Phasor diagram for Fresnel diffraction and circular symmetry.

For finite apertures/obstructions the value of the diffraction integral (i.e. the spanning vector) varies considerably:

$$
\begin{array}{lll}
\phi = 2n\pi & \rho^2 = 2n\lambda R & \Rightarrow \;\; \psi \approx 0 \\
\phi = (2n+1)\pi & \rho^2 = (2n+1)\lambda R & \Rightarrow \;\; \psi \approx 2\psi_u
\end{array}
$$

Fresnel Half-period zones (1)

In the circular geometry, these are concentric circular zones in the aperture plane, over which the phase at the observation point changes by π.

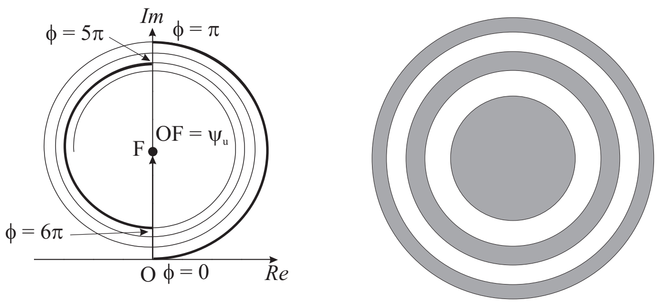


Figure 143: Fresnel half-period zones.

We define the 1st zone as the circular region in the aperture plane which satisfies

$$0 \leq \phi(\rho) \leq \pi$$

This corresponds to

$$\rho^2 \leq \lambda R. \tag{105}$$

The nth zone corresponds to the annulus satisfying

$$(n-1)\pi \leq \phi(\rho) \leq n\pi$$

$$\sqrt{(n-1)\lambda R} \leq \rho \leq \sqrt{n\lambda R}. \tag{106}$$

Fresnel Half-period zones (2)

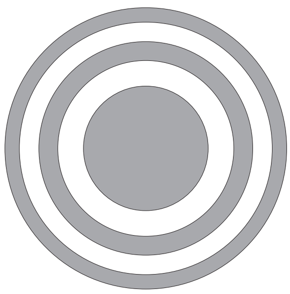


Figure 144: Fresnel half-period zones across the wavefront; circular geometry.

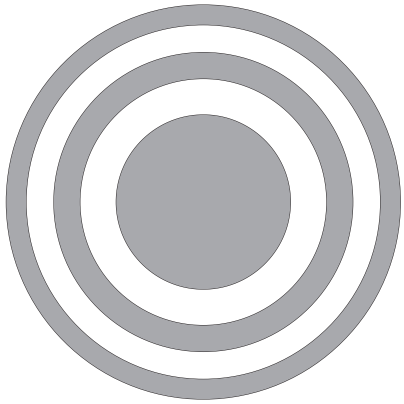
Note that the area of each zone is the same (in the quadratic approximation).

$$\pi\left(\rho_n^2 - \rho_{n-1}^2\right) = \pi\lambda R.$$

If we neglect the obliquity factor and the variation of r_1 and r_2, each zone would contribute equally to the amplitude at P; the phasor diagram would be circular. Odd numbered zones add to, and even numbered zones subtract from the overall amplitude at P. So on the optic axis of a circular aperture of radius r_a, the aperture includes N zones, given by $r_a^2 = N\lambda R$.

- if N is odd get bright spot at P; $\psi \sim 2\psi_u$; $I \sim 4I_u$
- if N is even get dark spot at P: $\psi \sim 0$; $I \sim 0$

Fresnel Half-period zones (3)



For large apertures, the small angle approximations begin to break down.

- The "spiralling in" in the phasor diagram due to the $1/r$ terms in the diffraction integral becomes noticeable
- this effect is further enhanced since $K(s) < 1$.
- the width of the outer zones also increases, as the higher-order terms become important.

Circular obstruction: Poisson's Spot

Consider a circular obstruction on the axis. For an obstacle of radius r_a, the inner zones up to $\rho = r_a$, i.e.

$$\phi_a = \frac{\pi r_a^2}{\lambda R}$$

are obscured, but all the outer zones are clear. So, the limits of integration for the diffraction integral are $\rho = r_a$ to $\rho = \infty$, and the relevant spanning vector in the phasor diagram is from F to A.

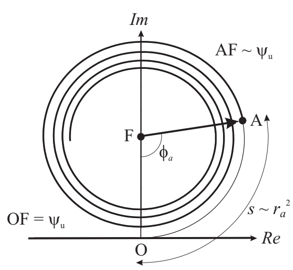


Figure 145: Phasor diagram for a circular obstacle.

Provided ϕ_a is not too large (i.e. provided r_a is not too large) the intensity is therefore close to that expected in the absence of any obstruction. This leads to a bright spot on the axis, known as Poisson's spot. Close to the obstruction, the diffraction angles also become large so that K falls, and no spot is observable.

Off-axis intensity for a circular aperture/obstacle

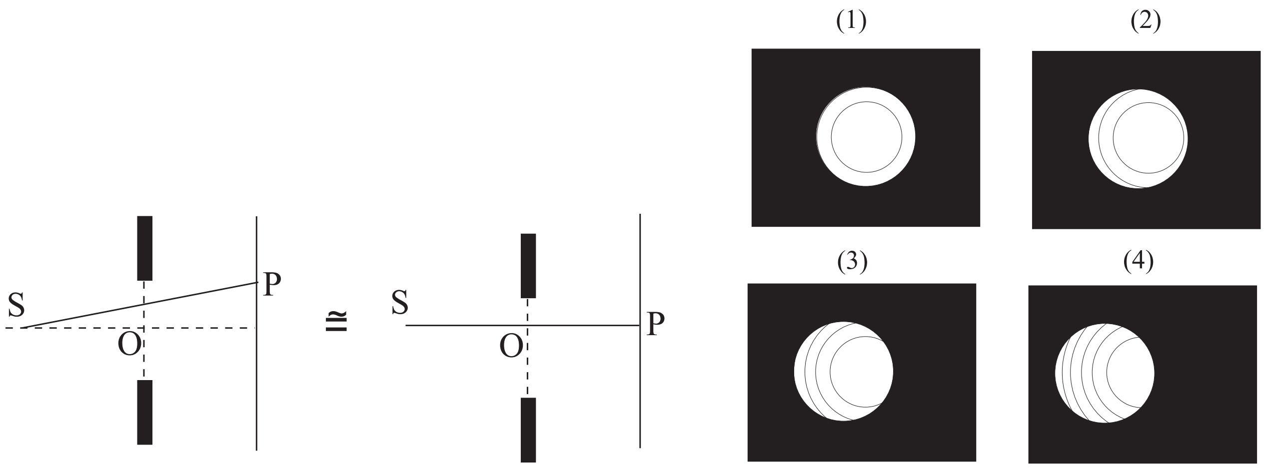


Figure 146: Off-axis diffraction determined by zone obstruction by aperture.

The Fresnel zones are centred on the source-observation point line; the zone structure moves with S. We can assume approximately that S and P are fixed, and the aperture shifts sideways across the zone structure.
Take as an example an aperture which covers the first two zones when on axis. This produces (in principle), zero intensity on the axis itself.
In (2), the 1st zone gives a full contribution, but the 2nd and 3rd are each partially obstructed, and tend to cancel each other. So there is a maximum in the intensity.

Off-axis intensity for a circular aperture/obstacle

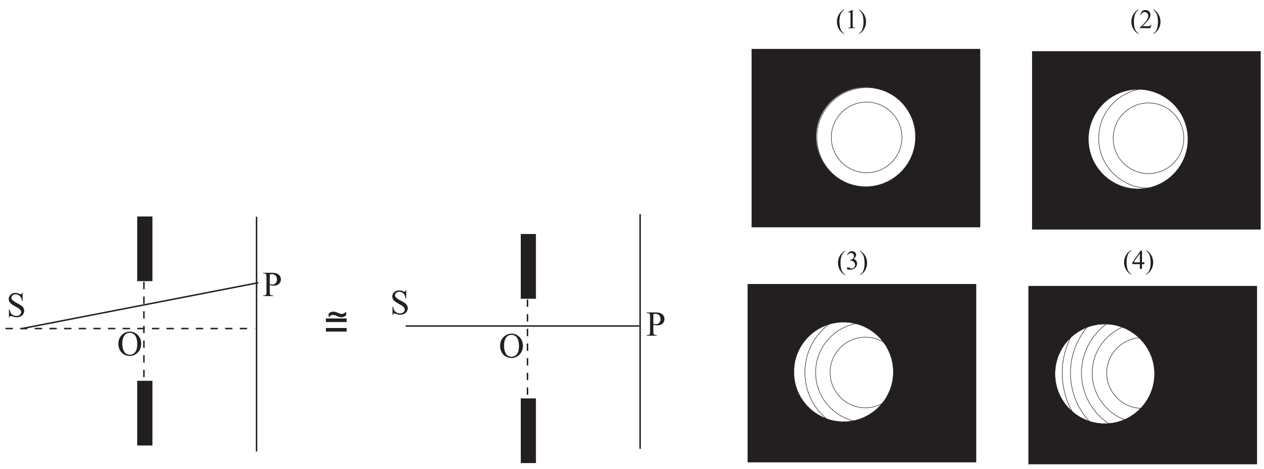


Figure 147: Off-axis diffraction determined by zone obstruction by aperture.

There is an oscillation in intensity as P moves off-axis. Circular symmetry means this must correspond to circular fringes. The fringe spacing is approximately equal to the zone width at the edge of the aperture.
For larger apertures, the zone widths are smaller at the edge, so the fringe spacing is smaller.
A long way off-axis, the aperture is clear for sectors of a large number of very narrow zones. The aperture covers most of the area of each of these zones, so they contribute rather little to ψ, decreasing as P goes further off-axis. So the intensity at P falls rapidly far from the optic axis of the aperture.

The Fresnel zone plate

A zone plate is an aperture which blocks out alternate Fresnel half-period zones.

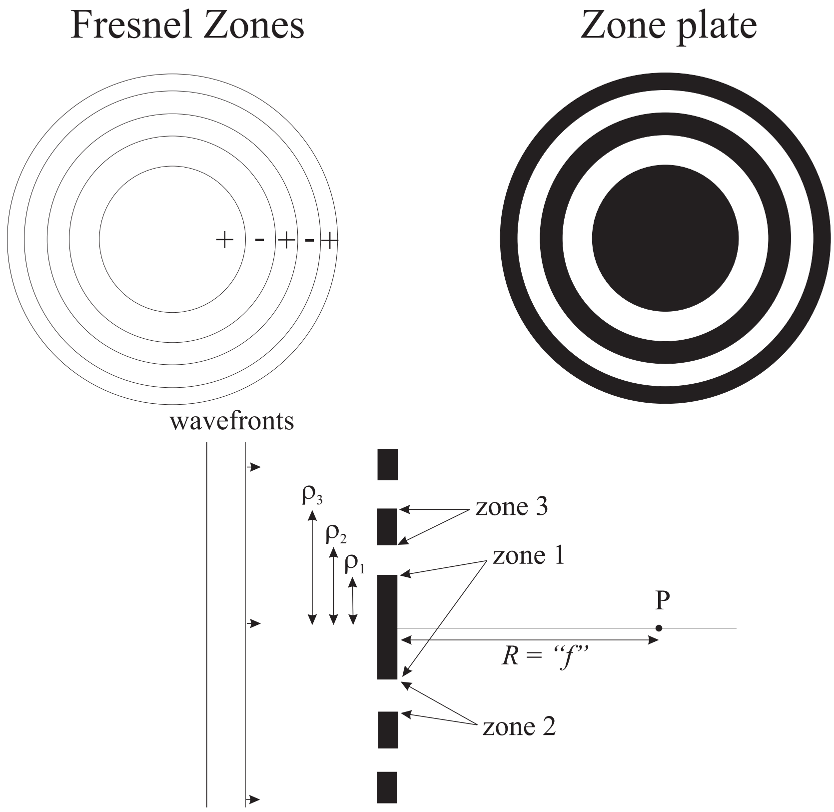


Figure 148: Fresnel zone plate; calculation of dimensions.

The aperture shown right will, for the observation point P, block out the 1st, 3rd, 5th ... zones shown on the left if designed as follows with

$$\rho_1 = \sqrt{\lambda R}, \quad \rho_2 = \sqrt{2\lambda R}, \quad \rho_3 = \sqrt{3\lambda R}, \quad \text{etc.}$$

Fresnel Zone Plate (1)

The effect on the phasor diagram is shown:

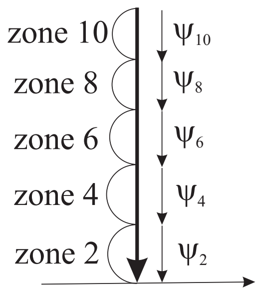


Figure 149: Fresnel zone plate; phasor diagram.

The net amplitude at P is

$$\psi_P = \psi_2 + \psi_4 + \psi_6 + \ldots \approx 2N\psi_u$$

where N is the number of open zones in the plate. Thus the intensity is

$$I_P \approx 4N^2 I_u.$$

So, an incident plane wave is "brought to a focus" at P, and the zone plate acts a lens with focal length

$$f = \frac{\rho_1^2}{\lambda} = \frac{\rho_n^2}{n\lambda}$$

Because $f \propto 1/\lambda$, this is a highly chromatic lens.

Fresnel Zone Plate (2)

Now move the observation point P along the axis towards the zone plate. The Fresnel zones for the new position R' of P become smaller in size:

$$\rho_1' = \sqrt{\lambda R'}, \rho_2' = \sqrt{2\lambda R'}, \rho_3' = \sqrt{3\lambda R'}, \text{ etc.}$$

When $R' = f/2$,

$$\begin{aligned}\rho_1' &= \sqrt{\lambda f/2} \\ \rho_2' &= \sqrt{\lambda f} = \rho_1 \\ \rho_3' &= \sqrt{3\lambda f/2} \\ \rho_4' &= \sqrt{2\lambda f} = \rho_2.\end{aligned}$$

So the open area of the plate

$$\rho_1 \leq \rho \leq \rho_2$$

now allows through two zones, 3 and 4, for the new position of P. Hence $\psi_P \sim 0$.

- when $R' = f/2m$ (for integer m), each open area of the zone plate admits an even number of Fresnel zones of opposite phase, so $\psi_P \to 0$. So there are points of zero intensity on the axis at $R' = f/2m$.
- when $R' = f/(2m+1)$, each open area of the zone plate admits an odd number of Fresnel zones, with a net contribution of one zone for each of the N open areas. So $\psi_P \to 2N\psi_u$; but this is reduced by the decreasing obliquity factor K. We expect maxima at $R = f/(2m+1)$ on axis.

Fresnel Zone Plate (3)

There are subsidiary foci and the intensity varies along the optic axis. Plot shows the on-axis intensity of a zone plate with 5 open zones (and 5 blocked), from numerical integration of

$$\sum_{m=1}^{m=5} \int_{s=(2m-1)\lambda R}^{s=2m\lambda R} \frac{1}{(a^2+s)^{1/2}} \exp\left(\frac{i\pi s}{\lambda R}\right) \pi \, ds.$$

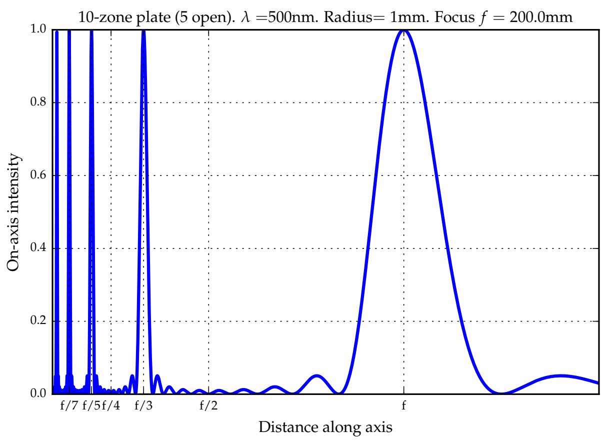


Figure 150: Subsidiary foci for a Fresnel zone plate.

This figure neglects the obliquity factor, which would reduce, eventually to zero, the intensity of the maxima at $f/(2m+1)$ as $m \to \infty$.

The lens is not high quality even at its design wavelength. But at some wavelengths (e.g. X-rays) for which the refractive index is very close to 1, it may be the only option.

Conditions for interference (1)

Diffraction is an interference effect. We calculated the patterns that emerge when we interfere, i.e. add up, light from various parts of an aperture. Consider a wave Ψ which is the superposition of two monochromatic waves at angular frequencies ω_1 and ω_2

$$\Psi = \Re\left[\psi_1 \mathrm{e}^{-\mathrm{i}\omega_1 t} + \psi_2 \mathrm{e}^{-\mathrm{i}\omega_2 t}\right].$$

Most detectors respond to the intensity I of the wave, Ψ^2. Expanding the real part of a complex quantity A as $\Re(A) = \frac{1}{2}(A + A^*)$ we can show

$$\begin{aligned} I \propto \Psi^2 =& \frac{1}{2}|\psi_1|^2 + \frac{1}{2}|\psi_2|^2 + \Re\left[\psi_1\psi_2^* \mathrm{e}^{\mathrm{i}[\omega_2-\omega_1]t}\right] \\ &+ \frac{1}{2}\Re\left[\psi_1^2 \mathrm{e}^{-2\mathrm{i}\omega_1 t}\right] + \frac{1}{2}\Re\left[\psi_2^2 \mathrm{e}^{-2\mathrm{i}\omega_2 t}\right] + \Re\left[\psi_1\psi_2 \mathrm{e}^{-\mathrm{i}[\omega_1+\omega_2]t}\right] \end{aligned}$$

The last three terms vary on a timescale which is more rapid than the response time of most detectors; if so these terms average to zero during the response time. Expanding ψ_1 and ψ_2 in terms of their amplitudes and phases $\psi = a\mathrm{e}^{\mathrm{i}\phi}$:

$$\langle I\rangle \propto \frac{1}{2}\langle a_1^2\rangle + \frac{1}{2}\langle a_2^2\rangle + \langle a_1 a_2 \Re\left[\mathrm{e}^{\mathrm{i}[\phi_1-\phi_2-(\omega_1-\omega_2)t]}\right]\rangle$$

where we $\langle$average$\rangle$ over the integration time of the detector.

Conditions for interference (2)

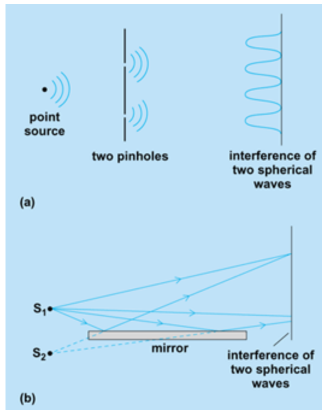
$$\langle I\rangle \propto \frac{1}{2}\langle a_1^2\rangle + \frac{1}{2}\langle a_2^2\rangle + \langle a_1 a_2 \Re\left[\mathrm{e}^{\mathrm{i}[\phi_1-\phi_2-(\omega_1-\omega_2)t]}\right]\rangle \qquad (107)$$

- We see that interference phenomena require the third term in equ. (107) to be non-zero; in other words this term is the interference term and the other two terms are present whether or not interference is occurring.
- If the detector is averaging the intensity over some finite period τ, we will not see interference if $(\omega_1 - \omega_2)\tau \gg 1$, in other words to see interference we need to have $\omega_1 = \omega_2$ to a good degree of approximation. For example, optical detectors typically have a response time $\tau > 1 \times 10^{-9}$ s whereas in the optical regime $\omega \sim 10^{15}$ rad/s, so the frequencies need to be equal to within a part per million or better.
- In practice, the phases ϕ_1 and ϕ_2 of independent sources (usually atoms) vary randomly and rapidly so interference is typically only seen when light from a single source (e.g. an atom) is split and then recombined: only in this case can $(\phi_1 - \phi_2)$ be stable for long enough that interference is seen.
- In Lasers, the phases of many emitting atoms are locked through stimulated emission, resulting in coherence over timescales of nanoseconds.

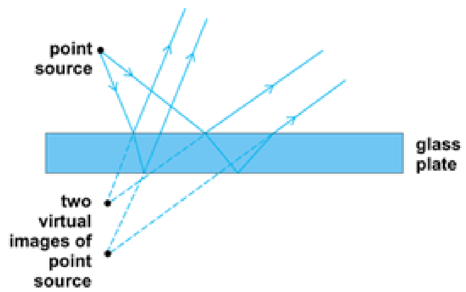
Division of Wavefront vs Division of Amplitude

Two ways to split and recombine wavefront from a light source:

Division of Wavefront The waves to be interfered are derived from different spatial points of a wavefront (which has some spatial coherence). The diffraction pattern is obtained by interfering (adding) these secondary waves at a point in space.



Division of Amplitude Waves to be interfered are derived by dividing the wavefront at a point in space in amplitude, for example by reflection and transmission at an interface. This type of interference is known as amplitude division.



The Michelson Interferometer (1)

The Michelson interferometer is perhaps the simplest amplitude division interferometer, but it has many important applications, including high-precision metrology, Fourier Transform spectroscopy and the Michelson-Morley experiment.

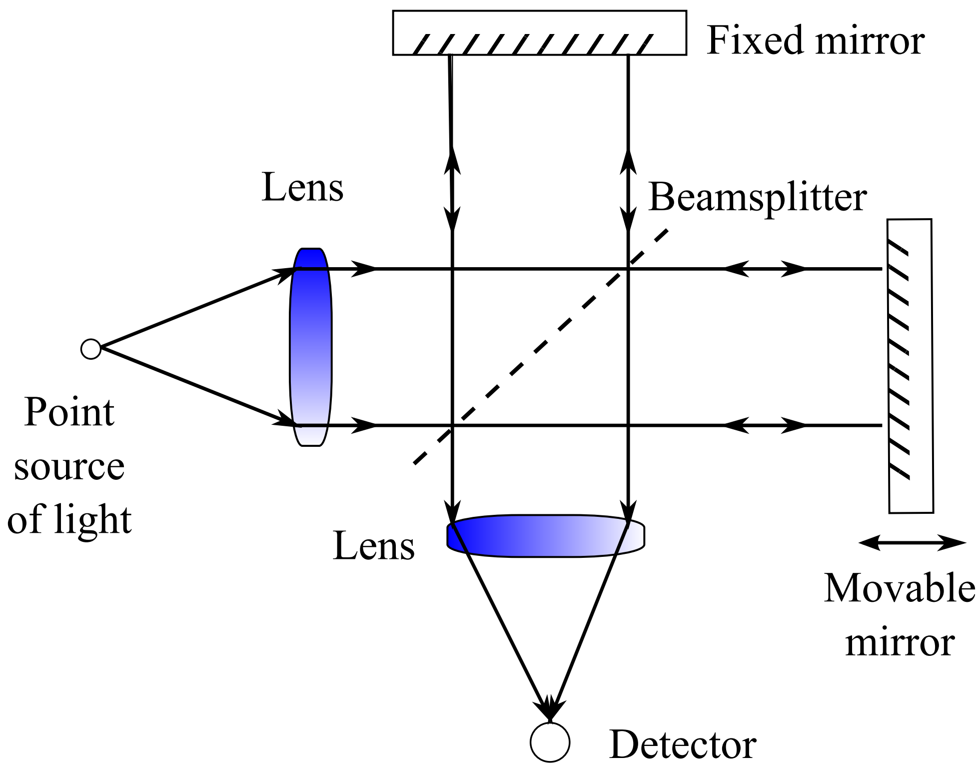
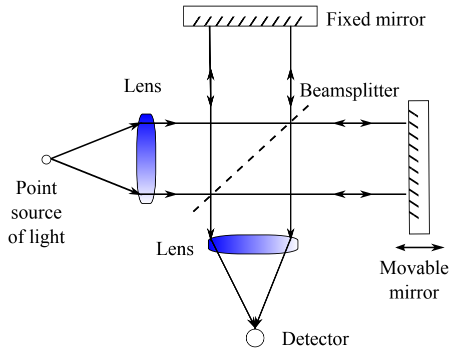


Figure 151: The Michelson interferometer

The Michelson Interferometer (2)



- The beam-splitter splits the incident beam; components travel along different paths, and recombine at the same beamsplitter.
- Constructive or destructive interference will occur depending on the difference between the two paths. Normally one path is kept fixed (the so-called reference path) and the other path is varied by moving a mirror.
- Vibrations of a wavelength will “smear” the fringes out — need rigidity.
- This arrangement focuses the beams onto a photodiode. The system is illuminated with the collimated beam; the mirrors are aligned so that the path difference is constant across the beam. The interference pattern is either all light or all dark depending on the path difference between the two arms.
- The fringe pattern is seen by varying the displacement of one of the mirrors as a function of time and recording the intensity as a function of the mirror position.

Michelson illuminated with extended light source

- Sometimes we use an extended light source. Light from different parts of the source travels at different angles through the interferometer to the eye and the differential delays will be a function of this angle. Mirrors may also be tilted to introduce different delays for beams hitting the mirrors at different points. The combination of these effects means that a spatially-varying intensity (i.e. fringes) can be seen when the mirrors are both held fixed.

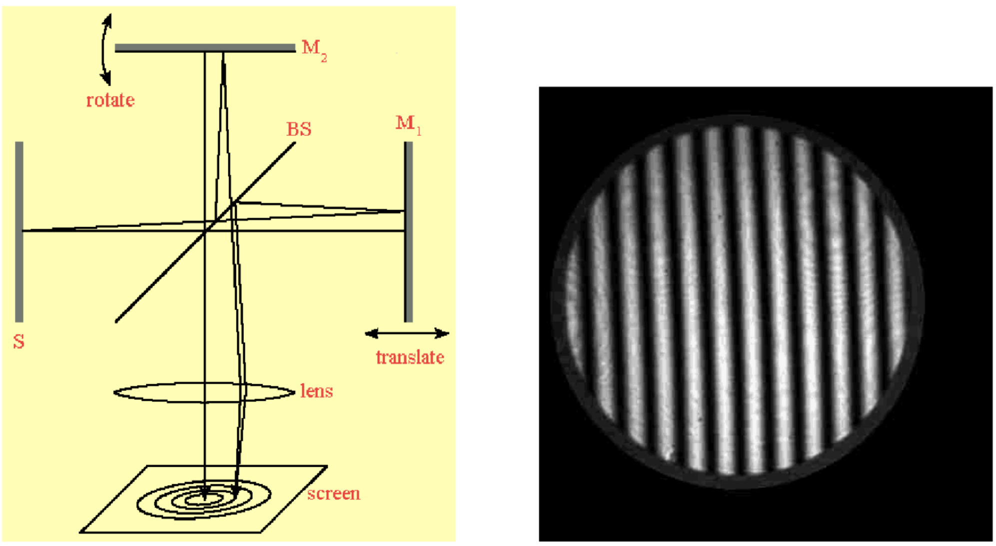


Figure 152: Fringes obtained by tilting one mirror

Monochromatic Fringes

- For light with wavenumber $k = 2\pi/\lambda$ the measured intensity can be derived from equ. (107) with $\omega_1 = \omega_2$

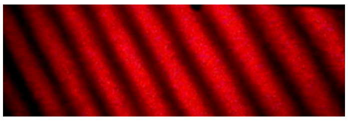
$$\langle I \rangle \propto \frac{1}{2}\langle a_1^2 \rangle + \frac{1}{2}\langle a_2^2 \rangle + \langle a_1 a_2 \Re \left[\mathrm{e}^{\mathrm{i}\delta} \right] \rangle, \qquad (108)$$

  where $\delta = (\phi_1 - \phi_2) = kx$ is the phase difference between the interfering beams. x is the difference in the paths travelled by the two beams.

- note that x is twice the difference in the distances of the two mirrors from the beamsplitter. If the interfering beams have equal intensities $I_0/2$:

$$I(x) = I_0 \left(1 + \Re \left[\mathrm{e}^{ikx} \right] \right) , \qquad (109)$$

  where we have dropped the angle brackets: averaging over the response time of the detector is implicit from now on.

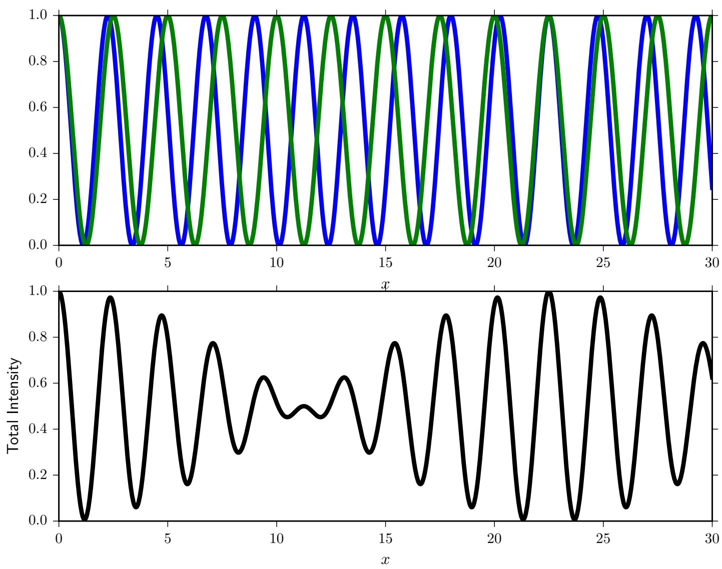


- Thus if we vary x linearly as a function of time, then the intensity seen at the output of the interferometer will vary sinusoidally, i.e. we will get a fringe pattern in time.

Measuring wavelength with a Michelson Interferometer.

Interference with two wavelengths present

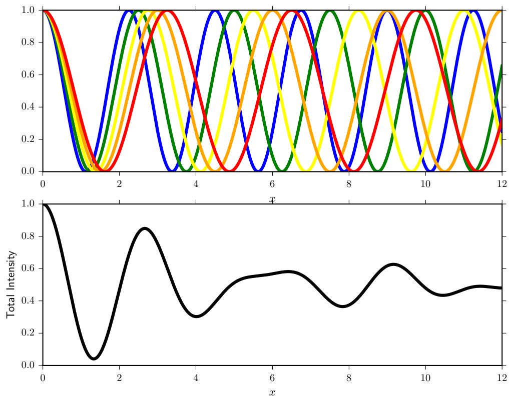
- If the light is not monochromatic, each wavelength will form its own set of fringes: there will be no fringes corresponding to interference between different wavelengths, because the condition $\omega_1 - \omega_2 \ll 1/\tau$ is not satisfied. Total intensity is now the sum of the intensities of the fringe patterns at the different wavelengths.



- As we vary x, we see fringe contrast change.

Interference with broadband light (1)

- With many frequencies present (e.g. white light):



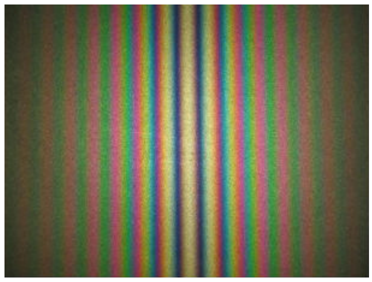


Figure 153:

- We will obtain typically a blurred, coloured pattern.
- Let intensity of light in the wavenumber range $k \to k + \mathrm{d}k$ be $2S(k)\mathrm{d}k$ (the reason for the factor of 2 will become apparent shortly) then the total intensity observed will be the sum of all the sine waves weighted by the intensity at the relevant wavelength:

$$I(x) = 2\int_0^\infty S(k)\left(1 + \Re\left[\mathrm{e}^{\mathrm{i}kx}\right]\right)\,\mathrm{d}k \tag{110}$$

Interference with broadband light (2)

- For mathematical convenience, we will also define $S(k)$ for negative k, such that $S(-k) = S^*(k)$. The fringe pattern as a function of the path difference x can then be written

$$I(x) = \int_{-\infty}^{\infty} S(k)\left(1 + \mathrm{e}^{\mathrm{i}kx}\right)\,\mathrm{d}k = I_1 + \int_{-\infty}^{\infty} S(k)\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}k \tag{111}$$

where I_1 is the total intensity of the light:

$$I_1 = \int_{-\infty}^{\infty} S(k)\,\mathrm{d}k.$$

Fourier transform spectroscopy

- The second term in equ. (111) is the inverse Fourier transform of $S(k)$. So we can find the spectrum of the source by taking the Fourier transform:

$$S(k) \propto \int_{-\infty}^{\infty} [I(x) - I_1]\, \mathrm{e}^{-\mathrm{i}kx}\, \mathrm{d}x.$$

- This method is used in the Fourier transform infrared spectrometer (FTIR) is a standard tool for characterising organic molecules by their vibrational (and rotational) frequencies, and uses a Michelson interferometer with automatic scanning of one mirror, accompanied by software to perform the Fourier transform.
- Fourier Transform Spectroscopy (FTS) is capable of higher spectral resolution than a spectrometer based on a prism or a diffraction grating. It is also easier to calibrate to high precision, because all the wavelengths of light pass through the same parts of the apparatus. The disadvantage is that the FTS relies on making many sequential measurements of $I(x)$, and this may take longer than taking a single image from a grating spectrograph.

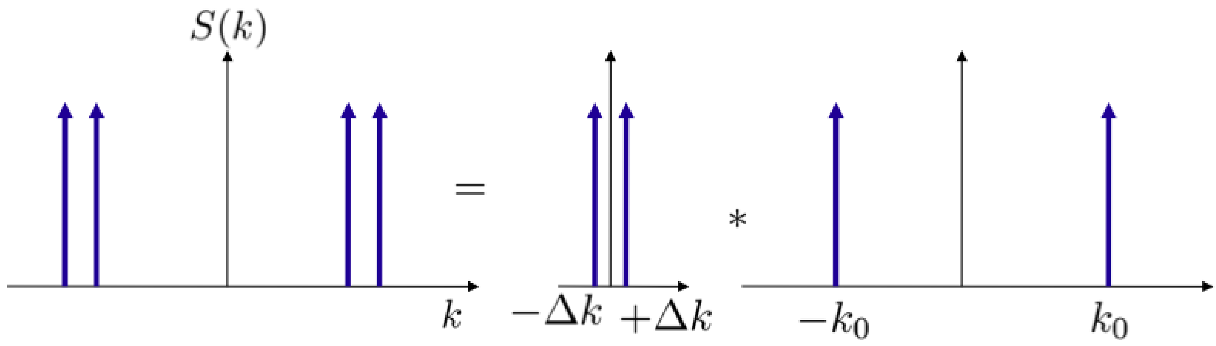
Fringe visibility

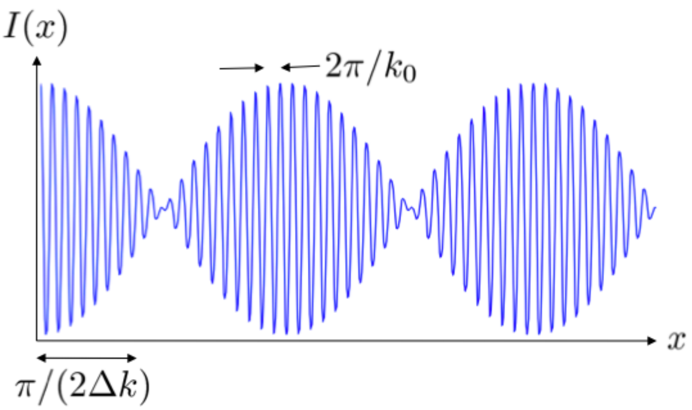
- If a light source has two closely-spaced wavevectors $k_0 \pm \Delta k$, each component will produce a fringe pattern with a slightly different fringe spacing. The observed interference pattern will be the sum of the two individual patterns.
- Around $x = 0$, the two fringe patterns approximately coincide, so the overall pattern will show well-defined maxima and minima. However, as x increases, a phase shift develops between the two patterns, and eventually (if the two components have the same intensity) the fringes become invisible since the maxima in one set of fringes overlap with the minima in the other set. The larger the spacing between the wavelength components, the more rapidly fringes disappear as x is increased.
- In our double-sided convention for a spectrum, $S(k)$ can be represented as pairs of closely-spaced delta functions spaced about $\pm k_0$. This can be written as a convolution

$$S(k) = [\delta(k - k_0) + \delta(k + k_0)] * [\delta(k - \Delta k) + \delta(k + \Delta k)]$$

Thus the intensity pattern will contain a product of two cosine functions

$$I(x) = I_1[1 + \cos(k_0 x)\cos(\Delta k x)] \tag{112}$$





- The high frequency cosine $\cos(k_0 x)$ represents the fringes while the low frequency envelope $\cos(\Delta k x)$ modulates the amplitude of these fringes. We can define the fringe contrast or fringe visibility as the ratio of the local high-frequency fringe amplitude to the mean intensity

$$\text{Visibility} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}. \tag{113}$$

- At the position of the first zero in the low-frequency envelope, the fringe contrast will be zero and this can be used to find Δk.
- Measuring the fringe visibility allows the fine structure of atomic lines to be investigated. For example in the 1890s Michelson was able to confirm that the sodium D line at 589.3 nm was in fact a doublet with $\Delta\lambda = 0.6$ nm, whereas the cadmium line at 634.8 nm was highly monochromatic, showing no minimum in fringe visibility for values of x up to 0.4 m.

Excellent set of online video demonstrations; see demo 17 for fringe pattern when the light source contains two closely spaced frequencies; and demo 10 for a basic Michelson setup; demo 11 for Michelson with circular fringes.
demo 10: Michelson interferometer with circular fringes
demo 17: fringes with two laser frequencies

Fourier Transform Spectrometry: Example

Example: A Fourier Transform Spectrograph has a maximum mirror displacement of $\Delta x = 50\,$cm. What is its spectral resolution at a wavelength of $\lambda_0 = 1\,\mu$m?

- We measure the Fourier transform $I(x)$ of the source spectrum $S(k)$ but over a limited range of x values.
- The maximum pathlength difference is 1 m. Assume that this is disposed symmetrically about the position of zero pathlength difference, i.e. x has a range of $\pm 50\,$cm. The measured intensity $I'(x)$ over this range is equal to $I(x)$ over infinite limits multiplied by a tophat function $W(x)$ of width $w = 2\Delta x = 1\,$m, i.e. $I'(x) = W(x) \times I(x)$.
- The reconstructed spectrum $S'(k)$ found by Fourier transformation of $I'(x)$ is thus $S'(k) = W^T(k) * S(k)$, where W^T is the Fourier transform of $W(x)$. That is, we obtain the true spectrum $S(k)$ convolved with the Fourier Transform of $W(x)$ (a sinc function). Spectral information at finer resolutions is lost. The width of the blurring function in spatial frequency space is $\Delta k = 2\pi/w$. Using $k = 2\pi/\lambda$, we find $\mathrm{d}k/k = -\mathrm{d}\lambda/\lambda$, so that
$$\frac{|\Delta\lambda|}{\lambda} = \frac{\Delta k}{k} = \frac{2\pi}{w}\frac{\lambda}{2\pi} = \frac{\lambda}{w}$$
- So the spectral resolving power $\lambda/\Delta\lambda$ is 10^6 in this case; we can distinguish spectral lines which are 1 pm apart when the central wavelength is 1 μm.

Resolving Power of the Fourier Transform Spectrometer

We have just shown that for an FTS, the resolution is given by
$$\frac{|\Delta\lambda|}{\lambda} = \frac{\lambda}{w}$$
where $w = 2\Delta x$ is the maximum separation of points from the wavefront which are interfered together.
Recall for a grating we had (equ. (97)):
$$\frac{|\Delta\lambda|}{\lambda} = \frac{D}{mW}$$
where D is the spacing between each slit, $W = ND$ is the total size of the grating, and m the diffraction order.
Note that in both cases it is the maximum separation between points on the wavefront (w or W) which sets the spectral resolution; the bigger the better.

Thin Film Interference (1)

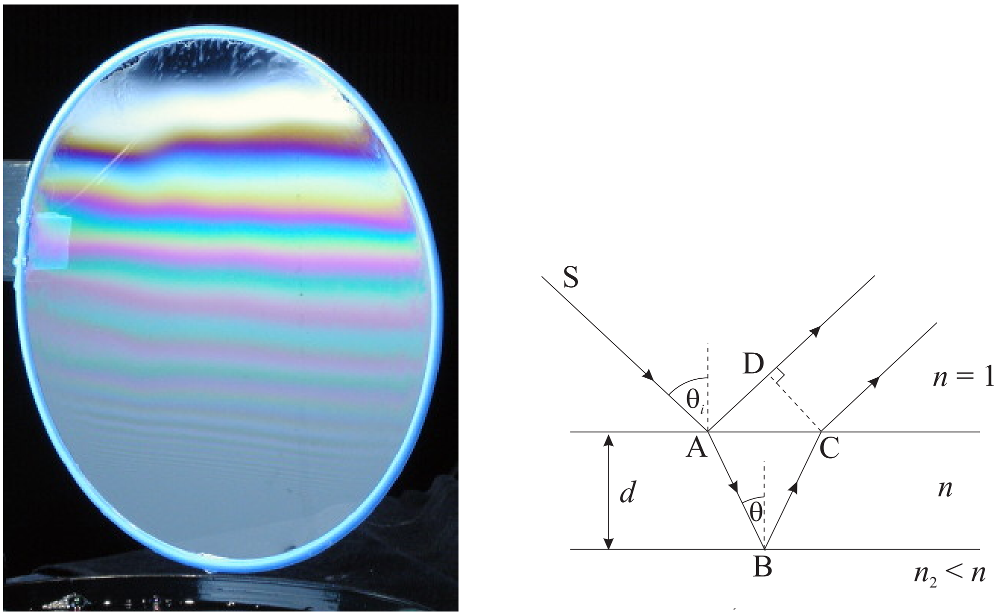


Figure 154: Interference in a thin film.

Amplitude division interference can occur when light waves are incident on a thin film of material. Interference occurs between the wave reflected at the top surface of the film and the wave which is initially transmitted, then reflected at the bottom surface and finally transmitted at the top surface.

Thin Film Interference (2)

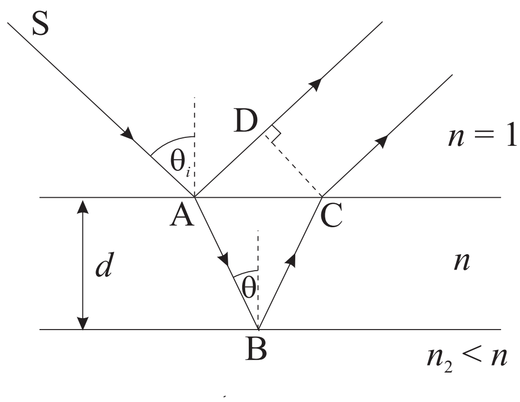


Figure 155: Interference in a thin film.

To calculate the intensity produced, we must consider the phase difference between paths AD and ABC. Taking into account the different refractive indices we have an optical path difference of

$$\begin{aligned} n(\mathrm{AB}+\mathrm{BC})-\mathrm{AD} &= 2n\,\mathrm{AB}-\mathrm{AD} \\ &= \frac{2nd}{\cos\theta}-2d\tan\theta\sin\theta_i \end{aligned}$$

Using Snell's law, $\sin\theta_i = n\sin\theta$ we obtain

$$n(\mathrm{AB}+\mathrm{BC})-\mathrm{AD} = \frac{2nd}{\cos\theta}\left(1-\sin^2\theta\right) = 2nd\cos\theta$$

Thin Film Interference (3)

This path difference introduces a phase difference

$$\delta = 2n\,d\,k\cos\theta = \frac{4\pi n d}{\lambda}\cos\theta. \tag{114}$$

The reflection at the upper surface occurs at a high-impedance to low-impedance boundary, while that at the lower surface occurs at a low-impedance to high-impedance boundary. This introduces an extra phase difference of π between the two beams. For simplicity, let us make the (unjustified) assumption that both beams have the same intensity $I_0/2$. Using the standard interference expression:

$$\langle I\rangle \propto \frac{1}{2}\langle a_1^2\rangle + \frac{1}{2}\langle a_2^2\rangle + \langle a_1 a_2 \Re\left[\mathrm{e}^{\mathrm{i}\delta}\right]\rangle,$$

we obtain a fringe system

$$I(\delta) = I_0\left(1 - \Re\left[\mathrm{e}^{\mathrm{i}\delta}\right]\right).$$

Thin Film Interference (4)

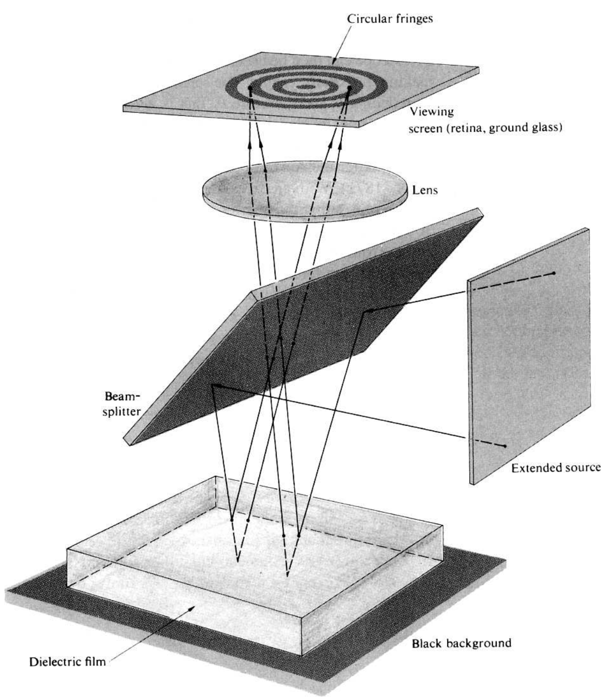
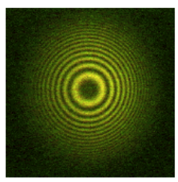


Figure 156: Haidinger fringes from a high-index film on a low-index substrate (from Hecht, Optics).

These fringes have maxima and minima as a function of angle as the two waves interfere. Where the two interfering beams differ in strength, the fringes are still present with the same period, but are less distinct since they sit on top of a constant background. These fringes are known as fringes of equal inclination, and may be observed by using a lens to focus the pattern. A particularly easy situation to visualise is the case where the film is illuminated by an extended source at close to normal incidence Here, circular fringes (Haidinger fringes) are produced, with each ring corresponding to a constant value of θ.



Thin Film Interference (5)

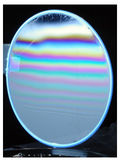
At a given value of θ, maximum reflection occurs at the wavelength corresponding to $\delta = (2m+1)\pi$, where m is an integer. This corresponds to

$$n\,d\cos\theta = \frac{(2m+1)}{4}\lambda.$$

Note that the resonance wavelength (the wavelength for constructive interference) decreases as θ increases. This is somewhat counterintuitive, since the path length inside the film increases. However, this is more than compensated for by the additional path length along AD. A more commonly observed type of interference fringes is seen for films of non-uniform thickness. These are known as fringes of equal thickness. For near-normal incidence, bright fringes are seen in the regions of the film where the thickness satisfies the condition

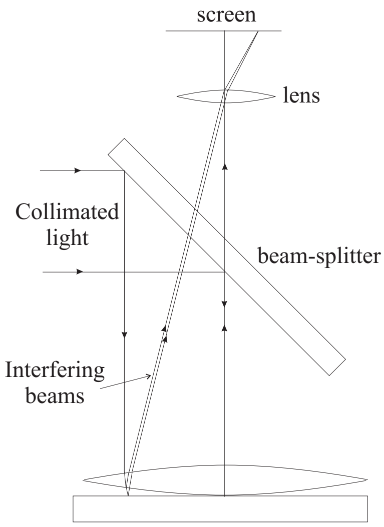
$$2nd = \left(m + \frac{1}{2}\right)\lambda.$$

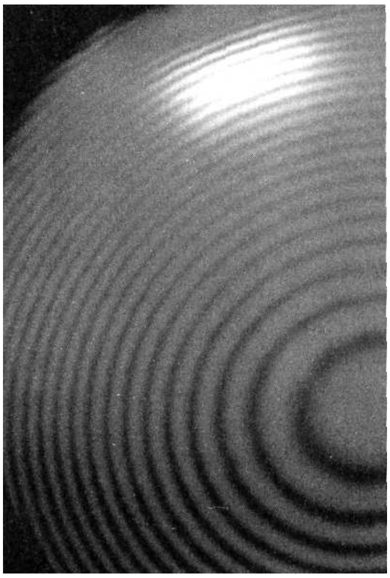
This type of fringe pattern is often seen in soap films or in thin films of oil on water. When illumination is with white light, each wavelength component produces its own set of fringes with a different period. The superposition of these fringes leads to a complex pattern of colours (interference colours) across the film.



Newton's Rings

A particular example of fringes of equal thickness occurs when an air gap is formed between two spherical surfaces (or a spherical surface and a plane surface). This produces circular fringes known as Newton's rings. This effect is widely used in quality control of optical surfaces.





The Fabry-Perot etalon: Multi-beam interference

So far, we have only considered interference of two beams. If the reflection coefficients at either side of the film are high, we must consider interference between multiple beams. An example of this type of structure is a film of air sandwiched between two mirrors, known as a Fabry-Perot etalon. This etalon produces a fringe pattern with much sharper features than a simple thin film, and is useful in high-resolution spectroscopy.

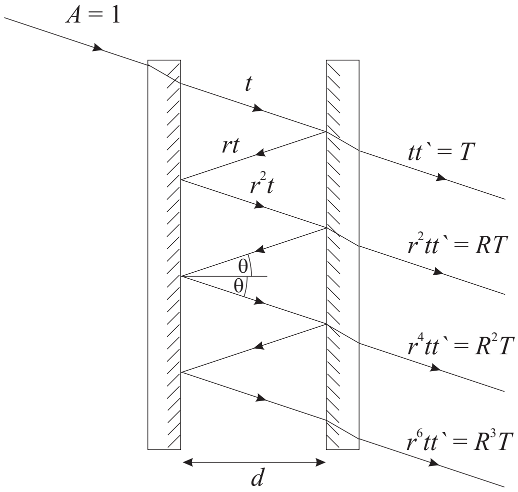


Figure 157: Multiple-beam interference in a Fabry-Perot etalon of thickness d. Amplitudes of the emerging beams are shown; the amplitude reflection coefficient of the mirrors is r, where $r^2 = R$, and the amplitude transmission coefficients at the interfaces shown are t and t', where $tt' = T$.

The Fabry-Perot etalon: Transmitted Intensity

- Consider a thin film of air of thickness d surrounded by mirrors. For light incident at angle θ the mirrors have an amplitude reflection coefficient r (which we assume is real). The amplitudes of the waves generated after successive reflections inside the cavity are shown. Each successive beam emerging from the etalon acquires a factor of $r^2 = R$ in amplitude and a phase shift δ, where

$$\delta = \frac{4\pi d}{\lambda}\cos\theta.$$

(from equ. (114) with $n = 1$). Note that we can neglect the effect of the glass supporting the mirrors, since this introduces a constant phase shift into all the beams.

- The total amplitude emerging from the etalon is the sum of a geometric progression:

$$A = T(1 + Re^{i\delta} + R^2e^{2i\delta} + R^3e^{3i\delta} + \ldots) = \frac{T}{1 - Re^{i\delta}}.$$

The total intensity transmission of the etalon, $I_{\rm out}/I_{\rm in}$ is thus

$$|A|^2 = \frac{T^2}{(1 - R\cos\delta)^2 + R^2\sin^2\delta} = \frac{T^2}{1 + R^2 - 2R\cos\delta}.$$

The Fabry-Perot etalon: Intensity (2)

- Writing $(1-R)^2 = 1 + R^2 - 2R$ we get

$$|A|^2 = \frac{T^2}{(1-R)^2 + 2R(1-\cos\delta)} = \frac{T^2}{(1-R)^2}\left(\frac{1}{1+[4R/(1-R)^2]\sin^2(\delta/2)}\right). \tag{115}$$

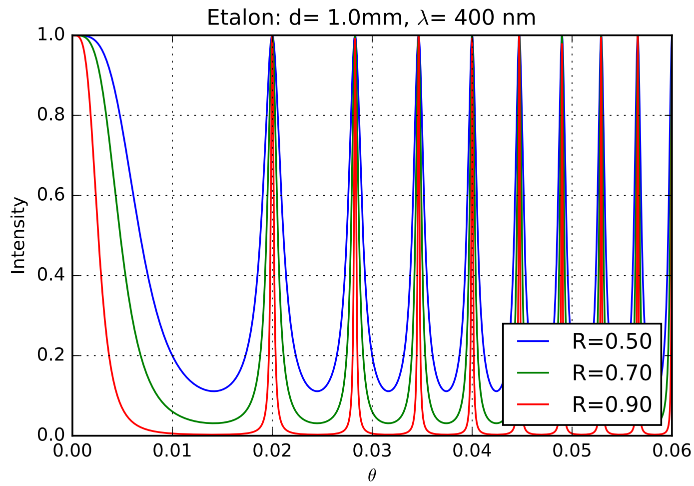


Figure 158: Transmission function for a Fabry-Perot etalon

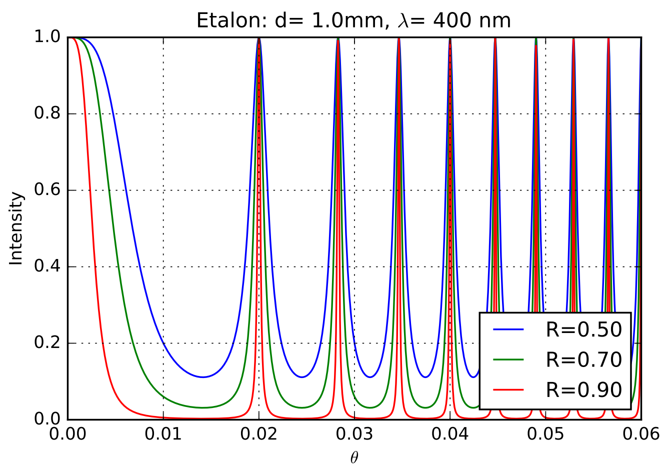


Figure 159: Transmission function for a Fabry-Perot etalon

- The fringe pattern has maxima at $\delta = \frac{4\pi d}{\lambda}\cos\theta = 2m\pi$ for integer m. At normal incidence ($\theta = 0$) we have an integer number of half wavelengths between the mirrors.
- Although a Fabry-Perot etalon will in general produce circular fringes, in practice use them at normal incidence and detect the transmitted intensity at angles very close to $\theta = 0$. By varying the spacing of the mirrors and using a photodiode to measure the transmitted intensity it is possible to measure the spectrum of the incident light, with each wavelength component producing its own set of peaks as a function of d.

The Fabry-Perot etalon: Spectral Resolution

The sharpness of the peaks as a function of δ is critical in determining the spectral resolution of the Fabry-Perot spectrometer. From equ. (115), the maximum value of I_{out}/I_{in} is $T^2/(1-R)^2$ when $\delta = 2m\pi$. We define the half-width of the peaks, $\delta_{1/2}$, as the change in δ required for the intensity to reduce to half its maximum value, hence

$$\frac{4R}{(1-R)^2}\sin^2(\delta_{1/2}/2) = 1.$$

Since $\delta_{1/2}$ is typically small, $\sin(\delta_{1/2}/2) \approx \delta_{1/2}/2$, hence

$$\delta_{1/2} = \frac{1-R}{R^{1/2}}.$$

It is useful to define the finesse, $\mathcal{F}$ as the ratio of separation of successive peaks in δ (equal to 2π) to their full-width at half maximum ($2\delta_{1/2}$).

$$\mathcal{F} = \frac{2\pi}{2\delta_{1/2}} = \frac{\pi R^{1/2}}{1-R}.$$

The Fabry-Perot etalon: Spectral Resolution (2)

The finesse is a measure of the quality of the etalon, and becomes high as R approaches 1. For $R = 0.95$, $\mathcal{F} = 61$. The chromatic resolving power of the etalon is defined as $\lambda/\Delta\lambda$, where $\Delta\lambda$ is the minimum wavelength difference between two spectral components which can just be resolved in the region of λ. Differentiating

$$\delta = \frac{4\pi d\cos\theta}{\lambda}$$

gives

$$d\delta = -\frac{4\pi d\cos\theta}{\lambda^2}d\lambda.$$

We assume that two spectral components can be distinguished if they are separated in δ by an amount $2\delta_{1/2}$, thus (forgetting the minus sign)

$$2\delta_{1/2} = \frac{4\pi d\cos\theta}{\lambda^2}\Delta\lambda; \quad \therefore \frac{\lambda}{\Delta\lambda} = \frac{2\pi d\cos\theta}{\lambda\delta_{1/2}}.$$

Recalling that, at a maximum in intensity, $2d\cos\theta = m\lambda$, the resolving power is

$$\boxed{\frac{\lambda}{\Delta\lambda} = \frac{m\pi}{\delta_{1/2}} = m\mathcal{F}.}$$

The Fabry-Perot etalon: Free Spectral Range

If $R = 0.95$, $d = 5$ mm and $\lambda = 500$ nm, the resolving power is more than 10^6, comparable with a grating spectrometer with a very large grating.
But neighbouring orders from different wavelengths will overlap. The wavelength difference at which overlapping takes place is known as the free spectral range.
At normal incidence, the intensity maxima are given by $2d = m\lambda$; differentiating:

$$-\frac{2d}{\lambda^2}\,\mathrm{d}\lambda = \mathrm{d}m.$$

Where m is large, we can use this equation to find the wavelength change $(\Delta\lambda)_{fsr}$ corresponding to $\Delta m = 1$, i.e.

$$\frac{2d}{\lambda^2}(\Delta\lambda)_{fsr} = \frac{m}{\lambda}(\Delta\lambda)_{fsr} = 1.$$

The free spectral range is thus given by

$$(\Delta\lambda)_{fsr} = \frac{\lambda}{m}.$$

For the example, the free spectral range around $\lambda = 500$ nm is only 0.025 nm. Etalons are therefore most suitable for measuring the fine structure of narrow spectral lines. The number of resolution elements across the spectrum is

$$\frac{(\Delta\lambda)_{fsr}}{(\Delta\lambda)_{res}} = \mathcal{F}$$

The Fabry-Perot etalon: Free Spectral Range (2)

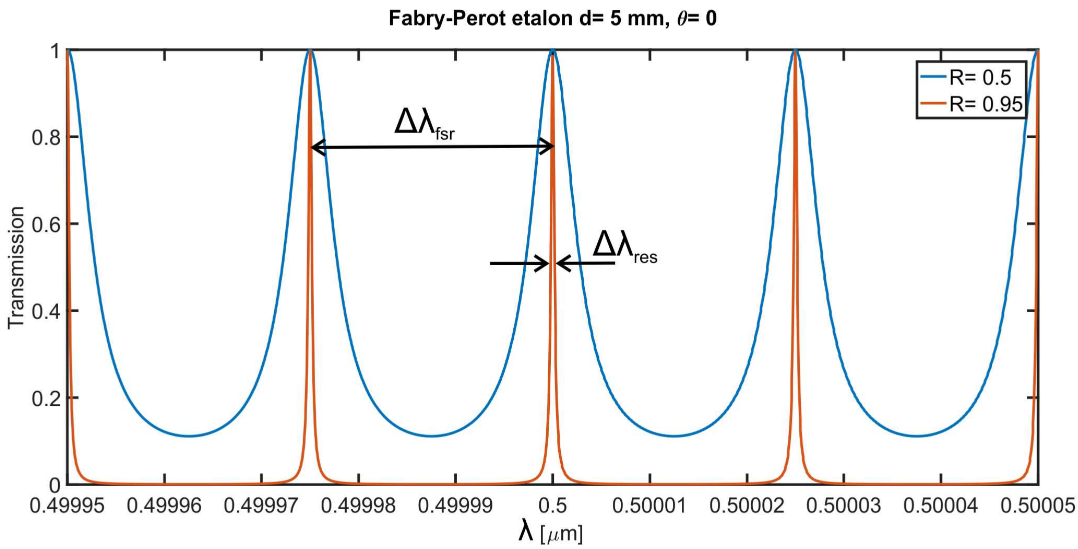


Figure 160: Free spectral range and line width of a Fabry-Perot

$(\Delta\lambda)_{fsr} = 0.025\,\text{nm}, \mathcal{F} = 61, (\Delta\lambda)_{res} = (\Delta\lambda)_{fsr}/\mathcal{F} = 0.4\,\text{pm!}$