

Lecture Notes

MSci Physics w/ Particle Physics and Cosmology
University of Birmingham

Year 1, Semester 1
Ash Stewart

Lectures Index

LC Classical Mechanics and Relativity 1	2
Lecture 1: Orders of Magnitude and Dimensional Analysis	2
Lecture 2: Dimensional Analysis (contd.) and Vectors	3
Lecture 3: Kinematics Introduction	6
Lecture 4: Projectile Motion	8
Lecture 6: The End of Special Relativity	10
Lecture 8	11
Lecture 17: Gravitation	12
LC Introduction to Probability and Statistics	14
Lecture 1: Introduction and Descriptive Statistics	14
Lecture 2: Population Statistics	15
Lecture 3: Error Propagation and Combinations of Variables	16
Lecture 4: Covariance and Correlation	18
Lecture 5: Distributions	19
Lecture 6: Likelihood and Log Likelihood	23
Lecture 7: Fitting a Straight Line 1	29
Lecture 8: Fitting a Straight Line 2	32
Lecture 9: Linear Regression	36
Lecture 10: Goodness of Fit	41
Lecture 12: Introduction to Probability	45
LC Mathematics for Physicists 1A	46
Lecture 1	46
Lecture 2	47
Lecture 3	48
Lecture 4	49
Lecture 5	50
Lecture 6	51
Lecture 7	52
Lecture 8: More Planes	53
Lecture 9	54
Lecture 10: More Algebraic Methods	55
LC Optics and Waves	56
Lecture 1: Intro to Waves and SHM Recap	56
Lecture 2: Wave Functions	58
Lecture 3: Generalised Wavefunctions	60
Lecture 4: Waves at Boundaries	64
Lecture 5	70
Lecture 6: Standing Waves 2 Electric Boogaloo	71
Lecture 7: Energy and Power of Waves	73
Lecture 8: EM Standing Waves and Lasers	74

Lecture 9: Optics Part Two	77
--------------------------------------	----

LC Quantum Mechanics 1	81
Lecture 1: Atomic Structure	81
Lecture 2: The Ultraviolet Catastrophe	86
Lecture 3: Particle Nature of Light	90
Lecture 4: Atomic Energy Levels and Spectra	97
Lecture 5: X-Ray Production and Diffraction	102
Lecture 6: X-Ray Spectra	106
Lecture 7: “Matter Waves”	112
Lecture 8: Wave-Particle Duality	117

Tue 30 Sep 2025 12:00

Lecture 1 - Orders of Magnitude and Dimensional Analysis

Given some equation, for example $E = mv^2$, we can decompose it into the basic physical quantities that make it up, for example in terms of Mass, Time and Length. We can denote the dimensions of some quantity by wrapping it in square brackets.

$$E = \frac{1}{2}mv^2$$

$$[E] = ML^2T^{-2}$$

Example:

$$\text{Pressure} \equiv \frac{\text{Force}}{\text{Area}}$$

Suppose we want to test whether pressure and linear momentum flux (amount of linear momentum per unit time, per unit surface) were equivalent quantities, we could do this using dimensional analysis:

$$[P] = \frac{[F]}{[A]}$$

$$[P] = \frac{M \times LT^{-2}}{L^2}$$

$$= M/LT^2$$

And for linear momentum flux ($\Phi(p)$ where lowercase p is momentum):

$$\Phi(p) = \frac{[p]}{[A]\text{[time]}}$$

$$= \frac{MLT^{-1}}{L^2T}$$

$$= \frac{M}{LT^2}$$

So yes, they seem to be (at least dimensionally) equivalent.

0.1 Challenging the LHC

We want to use orders of magnitude calculations to challenge the idea that the LHC is the “Big Bang Machine”.

The LHC operates on the order of magnitude of approx 10TeV. The age of the universe is approx 13.7Bn Years, or (in orders of magnitude) 10^{10} yrs.

What time was the big bang? The Big Bang started the universe, but we can't really say it happened at 0s, because that doesn't really make sense. What about 1sec? or 1ms? Well it's clearly less than both of those, so we want to find the smallest possible increment of time “Plank Second” and say it happened after one of them.

Thu 02 Oct 2025 15:00

Lecture 2 - Dimensional Analysis (contd.) and Vectors

0.1 Continuation of Dimensional Analysis

What if, in theory, we could build a system of units entirely from c , the speed of light, G , Newton's constant and h , the Plank Constant?

Cont. from Lec01, we can try to use this to work out the earliest possible cosmic time.

$$h = 6.6 \times 10^{-34} \text{ Js}$$

$$G = 6.67 \times 10^{-11} \text{ Nm}^2/\text{kg}^2$$

$$c = 3 \times 10^8 \text{ m/s}$$

Dimensionally:

$$[h] = \frac{ML^2}{T}$$

$$[G] = \frac{L^3}{T^2 M}$$

$$[c] = \frac{L}{T}$$

We want to use these to build out a time unit, so:

$$[h^u G^v c^z] = T$$

$$\left(\frac{ML^2}{T}\right)^u \left(\frac{L^3}{T^2 M}\right)^v \left(\frac{L}{T}\right)^z = T$$

$$M^{u-v} L^{2u+3v+z} T^{-u-2v-z} = T$$

Solving for:

$$u - v = 0$$

$$2u + 3v + z = 0$$

$$-u - 2v - z = 1$$

Gives us:

$$u = \frac{1}{2} \quad (2.1)$$

$$v = \frac{1}{2} \quad (2.2)$$

$$z = \frac{-5}{2} \quad (2.3)$$

$t_p = \sqrt{\frac{Gh}{c^5}}$ and plugging in the values for G , h , c gives us a value of time, which the earliest possible cosmic time equal to about 10^{-43} s

0.2 Plank Energy

Doing the same process for energy gives us (this time, the plank energy is the energy at which traditional theories of physics break down):

$$E_p = \frac{hc^5}{G}^{0.5} \approx 10^9 J$$

On the other hand, the LHC manages about 10TeV, which is orders of magnitude smaller than this, so the LHC cannot accurately simulate energies of this magnitude.

0.3 More Vectors

Again, vector notation will be \vec{a} . We define the x, y, z unit vectors as $\underline{\hat{e}_x}, \underline{\hat{e}_y}, \underline{\hat{e}_z}$.

We can therefore define any vector as:

$$\vec{a} = a_x \underline{\hat{e}_x} + a_y \underline{\hat{e}_y} + a_z \underline{\hat{e}_z}.$$

The length of a vector is again $|\vec{a}|$.

0.4 Vector Multiplication

Given \vec{a} and \vec{b} we can define the dot (scalar) product and the cross (vector) product

$$\vec{a} \cdot \vec{b} = |a||b| \cos \theta$$

Say we want to know the component of a vector along an axis, we can do the following (eg for x):

$$\vec{a} \cdot \underline{\hat{e}_x} = a_x$$

For the vector product, we can define:

$$\vec{a} \times \vec{b} = |a||b| \sin(\theta) \hat{j}$$

As the vector perpendicular to the plane containing a and b. It is in the direction given by the *right hand rule*, where curling four fingers into a fist, and orienting your fist so these fingers sweep from \vec{a} to \vec{b} , the new vector will point in the direction of an extended thumb. Theta is the angle between a and b, while j is the unit vector in the direction the new vector will point.

0.5 Solar Energy Example

The world yearly energy usage is about 180,000TWh, which is about $5 \times 10^{20} J$ total. Is it (theoretically) possible to get this all from solar energy? We can check using an approximate order of magnitude calculation.

The Sun's total luminosity is $L_\odot = 3.8 \times 10^{26}$. This energy is radiated in a spherically symmetric way (we assume). Therefore the energy per time, per unit surface is (using 1AU for distance):

$$\frac{L_\odot}{4\pi \times (1.5 \times 10^6)^2}$$

Which is approximately (using order of magnitude):

$$\frac{3.8 \times 10^{26} W}{10 \times 10^{22} m^2} \approx \frac{1 kW}{m^2}$$

This is true in ideal conditions, and real energy supply is lower (due to clouds, atmosphere etc).

If we totally covered the earth's surface area ($A_{\text{surface}} \approx \pi R_\odot^2$) which is approximately:

$$A_{\text{surface}} \approx \pi \times (6 \times 10^3 \times 10^3 m)^2 \approx 10^{14} m^2$$

Therefore total energy received is approximately:

$$P = \frac{1 kW}{m^2} \times 10^{14} m^2 \approx 10^{17} W$$

And to power the world:

$$E = \frac{5 \times 10^{20} J}{3 \times 10^7 s} \approx 10^{13} W$$

So, it's theoretically possible, if we could cover enough of the world in solar panels and if we could perfectly capture the sun's energy without losing some to sources such as clouds, atmosphere, areas of the ocean we cannot cover in solar panels etc.

Thu 04 Sep 2025 12:00

Lecture 3 - Kinematics Introduction

For kinematics, we'll treat all objects as points and disregard aspects like rotation/the physical size of the body etc.

Given some point, we can define its position as a function of time $\vec{r}(t)$, and velocity as the derivative wrt time of this:

$$\vec{v}(t) = \frac{d\vec{r}}{dt}$$

And acceleration:

$$\vec{a}(t) = \frac{d\vec{v}}{dt} = \frac{d^2\vec{r}}{dt^2}$$

0.1 Position from Unit Vectors

We can define:

$$\vec{r}(t) = r_x(t)\hat{e}_x + r_y(t)\hat{e}_y + r_z(t)\hat{e}_z = \sum_{j=1}^3 r_j(t)\hat{e}_j$$

So:

$$\begin{aligned} \frac{d\vec{r}}{dt} &= \frac{d}{dt} \left(\sum_{j=1}^3 r_j(t)\hat{e}_j \right) \\ &= \sum_j \frac{d}{dt} (r_j \hat{e}_j) \\ &= \sum_j \frac{dr_j}{dt} \hat{e}_j \\ \vec{v} &= \sum_j v_j \hat{e}_j \end{aligned}$$

And:

$$\vec{a} = \frac{d\vec{v}}{dt} = \sum_{j=1}^3 a_j \hat{e}_j$$

Note: Taking the derivative of a vector wrt time is looking at how the variable changes in some infinitesimal time. This can be a change in direction, and/or a change in magnitude. To differentiate a vector we can differentiate it component-wise.

0.2 Cartesian and Polar

Instead of representing a point as x and y components (in 2D), we can instead define it as a distance from the origin r and the angle this distance line forms with the positive x-axis θ .

Therefore (by basic right angle trig) $x = r\cos\theta$, $y = r\sin\theta$, and hence:

$$\vec{r} = r \cos \theta \hat{e}_x + r \sin \theta \hat{e}_y$$

So:

$$\begin{aligned}\vec{u}(t) &= \frac{d\vec{r}}{dt} = \frac{d}{dt}(r \cos \theta) \hat{e}_x + \frac{d}{dt}(r \sin \theta) \hat{e}_y \\ &= (\dot{r} \cos \theta + r(-\sin \theta)\dot{\theta}) \hat{e}_x + (\dot{r} \sin \theta + r(\cos \theta)\dot{\theta}) \hat{e}_y \\ &= r(\cos \theta \hat{e}_x + \sin \theta \hat{e}_y) + r\dot{\theta}(-\sin \theta \hat{e}_x + \cos \theta \hat{e}_y)\end{aligned}$$

0.3 Example

Lets model a particle, in a single dimension moving with constant acceleration (a_0) along a line. What is $x(t)$?

$$a = \text{constant} = a_0$$

$$a = a_0 = \frac{dv}{dt}$$

$$\int_{v_0}^v dv = \int_{t_0}^t a_0 dt = a_0 \int_{t_0}^t dt$$

$$v - v_0 = a_0(t - t_0)$$

$$v = v_0 + a_0(t - t_0)$$

And now the position function:

$$v = \frac{dx}{dt}$$

$$\int_{x_0}^x v dt = \int_{t_0}^t v_0 + a_0(t - t_0) dt$$

$$x - x_0 = v_0(t - t_0) + \frac{1}{2}a_0(t - t_0)^2$$

Thu 09 Oct 2025 15:00

Lecture 4 - Projectile Motion

Projectile Motion: The motion of a particle subject to gravitational acceleration, $g \approx 9.81\text{m/s}^2$

1 Projectile Motion

For this to hold, the height of the particle above the ground must be $m \ll R_e \approx 6 \times 10^3\text{km}$.

$$x(t) = x_0 + v_0(t - t_0) + \frac{1}{2}a_0(t - t_0)^2$$

Lets begin solely by considering motion in the vertical axis (called z here, for some strange reason). This particle is falling from height $h\text{m}$, to the ground at $h = 0$, with constant acceleration gm/s^2 . It has been dropped at time $t = t_0$

At time t_0 , $v = 0, z = h$

$$\begin{aligned} z(t) &= h + 0 - \frac{1}{2}gt^2 \\ \frac{1}{2}gt^2 &= h \\ \implies t &= \sqrt{\frac{2h}{g}} \end{aligned}$$

1.1 What about 2D?

Now we can expand our example to (rather than drop the particle from rest) give the particle some initial velocity $v_0\text{m/s}$ parallel to the ground. We now want two position functions, $x(t)$ and $z(t)$. As previously calculated:

$$z(t) = h + 0 + \frac{1}{2}(-g)t^2$$

And horizontally:

$$x(t) = 0 + v_0(t) + 0$$

So:

$$\begin{cases} z = h - \frac{1}{2}gt^2 \\ x = v_0t \end{cases}$$

Rearranging:

$$\begin{aligned} t &= \frac{x}{v_0} \\ z &= h - \frac{1}{2}g\left(\frac{x}{v_0}\right)^2 \end{aligned}$$

Since h, g, v_0 are all constants, this is an x^2 parabola.

1.2 Interplanetary Example

Lets consider some planet, with $g_{\text{planet}} = 5m/s^2$. You (denoted Y) fall into the atmosphere at some distance h from the ground, and some horizontal distance d from $O(x = 0)$. There is an alien who wants to kill you, by shooting you down. This “gun” can throw pebbles at some constant speed v_0 . The only degree of freedom the alien has to target you is change the shooting angle wrt the horizontal, θ . From the alien’s perspective, what is the required θ to hit the incoming spacecraft?

To hit you, there is some time t , when the position of the bullet B , with initial velocity v where B is in the same position as Y

Consider B

$$\begin{aligned}x_B(t) &= v_0 \cos(\theta)t \\z_B(t) &= v_0 \sin(\theta)t - \frac{1}{2}g_p t^2\end{aligned}$$

Consider Y

$$\begin{aligned}x_Y(t) &= d \\z_Y(t) &= h - \frac{1}{2}g_p t^2\end{aligned}$$

We want to find a θ where $x_B = x_Y$ and $z_B = z_Y$ at the same t :

$$v_0 \cos(\theta)t = d \quad (4.1)$$

$$v_0 \sin(\theta)t - \frac{1}{2}g_p t^2 = h - \frac{1}{2}g_p t^2 \quad (4.2)$$

From 2:

$$\begin{aligned}v_0 \sin(\theta)t &= h \\ \implies t &= \frac{h}{v_0 \sin \theta}\end{aligned}$$

And subsituting:

$$\begin{aligned}v_0 \cos(\theta) \left(\frac{h}{v_0 \sin \theta} \right) &= d \\ \frac{\cos(\theta)h}{\sin(\theta)} &= d \\ \frac{\cos \theta}{\sin \theta} &= \frac{d}{h} \\ \tan \theta &= \frac{h}{d}\end{aligned}$$

Since we have the value of θ in terms of two constants, yes, the alien can always hit the spaceship provided it correctly selects the angle corresponding to the value of these two constants. This means that the required angle does not depend on velocity, in this example.

2 Frames of Reference

“Observer” represents a frame of reference.

Thu 23 Oct 2025 15:00

Lecture 6 - The End of Special Relativity

Doppler Effect of Photons / EM Waves

Rather than considering sound waves, we consider the doppler effect for photons travelling at the speed of light.

$f_{\text{emission}} \neq f_{\text{observer}}$, if observer is moving relative to the source.

In this context, frequency is often denoted ν not f . λ keeps its original meaning of wavelength.

Thu 30 Oct 2025 15:00

Lecture 8

Tue 25 Nov 2025 12:00

Lecture 17 - Gravitation

1 Gravitation

Consider two particles, with masses m_1 and m_2 at distance \vec{r} . We know there is an attractive force between the two masses:

$$\vec{F} = -G \frac{m_1 m_2}{r^2} \hat{e}_r \quad (7.1)$$

$$G = 6.67 \times 10^{-11} \text{ m}^3/\text{kg s}^2$$

1.1 Key Properties

- “Long Range” force. A force will exist between any two masses anywhere in the universe, regardless of distance and cannot be cancelled. Negligible at large distances, as force quickly tends to zero as distance increases - but never zero.
- Weak.
- $\vec{F} \propto \frac{1}{r^2}$

1.2 Mass Caveats

While we don't practically make a distinction between them, m in gravity refers to ‘gravitational mass’, m_g , while mass in Newton's Second Law $\vec{F} = m\vec{a}$ is ‘inertial mass’, m_i .

Einstein's Equivalence Principle says that they're equal, i.e:

$$\frac{dm_g}{dm_i} = 1$$

Why ‘they could be different but they fundamentally are not’ is important is beyond me.

2 Freefall

Consider an object of mass m at height h from the ground. It has force:

$$\vec{F} = mg$$

And potential energy:

$$\Delta U(h) = mgh$$

However this first expression looks quite different to our definition of force.

2.1 Derivation

Gravitational force is conservative, and there is a potential energy at all points in the gravitational field, associated with this force.

Give two bodies, M and m . m is moved away from M from r_1 to r_2 along the radial direction (preserving angle, i.e. moving only in \hat{e}_r).

$$\vec{F} = -G \frac{Mm}{r^2} \hat{e}_r$$

This is the negative gradient of the potential (as potential is area under a force curve):

$$\vec{F} = -\nabla U = -\frac{dU}{dr}\hat{e}_r$$

$$\vec{F} dr = -dU \hat{e}_r$$

The work done from r_1 to r_2 is:

$$w_{1 \rightarrow 2} = \int_{r_1}^{r_2} \vec{F} \cdot \hat{e}_r dx$$

Wed 01 Oct 2025 12:00

Lecture 1 - Introduction and Descriptive Statistics

0.1 Course Welcome

- First half of the semester: Statistics
- Second half the of semester: Probability
- All slides and notes on Canvas.

Why Descriptive Statistics? If we want to share an interesting bit of data, sharing the whole data is going to be confusing. Instead, we can share a small number of stats which describe and summarise the data.

0.2 Sample Statistics

One of the most simple is the number of samples (N), and the sample mean:

$$\text{Sample Mean: } \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

We can also calculate the sample standard deviation as the average of mean squared error across the points in the sample:

$$\text{Sample STDev: } s_n^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

We can also use median or mode as measures of central tendency. The mode is a poor estimator however (as it massively depends on how binning is done, for a continuous measurement), while the median is more resistant to outliers.

Thu 02 Oct 2025 09:00

Lecture 2 - Population Statistics

0.1 Accuracy and Precision

We usually take measurements to determine some kind of true value. Usually, we can't actually know what this true value is, but if we could there are two bits of terminology that is particularly important:

Accuracy: Accuracy is the 'closeness' between our value and the 'true' value.

Precision: Precision is the 'closeness' between our measurements, i.e. how spread out are our various measurements.

0.2 Error

Random Error: is uncertainty related to the fact that our measurements are only a finite sample, so is not going to be immediately representative of the true value. The smaller this error, the more precise the measurement is.

Systematic Error: is related to some kind of issue with the measurement or the equipment. This shifts all values, and negatively affects accuracy (but leaves precision unchanged)

Taking many repeat measurements decreases the effects of random error, but the effects of systematic error are much harder to combat.

Ideally, we want to be both precise and accurate, however accuracy is arguably more important. This is because a value which is precise, but not accurate may lead to false conclusions around the inaccurate value.

Wed 08 Oct 2025 12:02

Lecture 3 - Error Propogation and Combinations of Variables

Office Hours: 11:00 to 13:00 Thursdays, Physics West Rm 122

1 Types of Error

Broadly two types of error: Statistical/Random Error (resulting from low precision) and Systematic Error (from Low Accuracy).

Random error widens the distribution, while systematic error shifts the whole distribution up or down, meaning no matter how many repeats you take and how precise you think you are, the value is still nonsense as all datapoints have been equally shifted (i.e. by a poor experimental setup).

For example, you are trying to measure the length of an object using a ruler that has been unknowingly stretched. You cannot get a true value no matter the number of repeats or degree of precision.

1.1 Accuracy vs Precision

High accuracy is preferable to high precision - having high precision but low accuracy can lead to false conclusions (as an incorrect value appears confidently correct). Accuracy is more difficult to improve - precision can be improved by gathering more data, while higher accuracy can only be improved by a better experimental design.

2 Error Propagation

If we take a distribution, and add a constant value to all points, the distribution is shifted up/down without changing the variance.

$$\langle x + k \rangle = \langle x \rangle + k$$

$$Var(x + k) = Var(x)$$

If we multiply by a constant value, the mean is multiplied by this value, but the distribution becomes stretched and the variance grows:

$$\langle xk \rangle = k\langle x \rangle$$

$$Var(kx) = k^2 Var(x)$$

Or taking the natural log:

$$\langle \ln x \rangle \approx \ln \langle x \rangle$$

$$Var(\ln x) \approx \frac{Var(x)}{x^2}$$

As this is a non-linear operator, these become good approximations rather than strict rules of equivalence.

And another example:

$$\langle e^x \rangle \approx e^{\langle x \rangle}$$

$$Var(e^x) \approx (e^{\langle x \rangle})^2 Var(x)$$

Note here, even though our underlying distribution is Normal and symmetric, the new distribution after e^x is neither, and these are an even worse approximation than before.

2.1 Combining Operators

We can apply some linear transformation $mx + c$, we can chain these rules together by doing the multiplicative transformation m first, then the linear scale c .

$$\langle mx + c \rangle = m\langle x \rangle + c$$

$$Var(mx + c) = m^2 Var(x)$$

2.2 Multiple Variables

What if we have multiple distributed variables we want to add?

$$\langle A + B \rangle = \langle A \rangle + \langle B \rangle$$

$$Var(A + B) = Var(A) + Var(B)$$

And multiplying them (again this are now approximations)?

$$\langle AB \rangle \approx \langle A \rangle \langle B \rangle$$

$$Var(AB) \approx \langle B \rangle^2 Var(A) + \langle A \rangle^2 Var(B)$$

$$\frac{Var(AB)}{\langle AB \rangle^2} \approx \frac{Var(A)}{\langle A \rangle^2} + \frac{Var(B)}{\langle B \rangle^2}$$

Or division?

$$\left\langle \frac{A}{B} \right\rangle \approx \frac{\langle A \rangle}{\langle B \rangle}$$

$$Var\left(\frac{A}{B}\right) = \frac{Var(A)}{\langle B \rangle^2} + \frac{Var(B)}{\langle A \rangle^2}$$

3 One Rule to Rule Them All

This single rule allows us to propagate error in any situation, assuming the two variables are uncorrelated:

$$Var(f) \approx \left(\frac{\partial f}{\partial A} \Big|_{A=\langle A \rangle, B=\langle B \rangle} \right)^2 Var(A) + \left(\frac{\partial f}{\partial B} \Big|_{A=\langle A \rangle, B=\langle B \rangle} \right)^2 Var(B)$$

Thu 09 Oct 2025 09:00

Lecture 4 - Covariance and Correlation

Office Hours: Thursday 11am - 1pm, Physics West Rm 222 (b.becsy@bham.ac.uk)

Previously, when looking at two or more variables for error propagation/combinations etc, we assumed that they were independent of one another. Today we look at how to handle multiple variables which may be correlated.

1 Covariance

Covariance is a measure that indicates how much two variables fluctuate together:

$$\text{Cov}(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

Covariance matrices represent all combinations of covariance (noting $\text{Cov}(x, y) = \text{Cov}(y, x)$ and $\text{Cov}(x, x) = \text{Var}(x)$)

$$\Sigma = \begin{pmatrix} \text{Cov}(x, x) & \text{Cov}(x, y) \\ \text{Cov}(y, x) & \text{Cov}(y, y) \end{pmatrix}$$

We can then define correlation:

$$\text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

This is bounded between -1 ($x = -y$), 0 ($x = y$) and zero for no correlation. We can again put this in a matrix, noting it is symmetrical:

$$\begin{pmatrix} 1 & \text{Corr}(x, y) \\ \text{Corr}(y, x) & 1 \end{pmatrix}$$

1.1 Variable Combinations

Now, with correlated variables, we can say:

$$\langle x + y \rangle = \langle x \rangle + \langle y \rangle$$

$$\text{Var}(x, y) = \text{Var}(x) + \text{Var}(y) + 2\text{Cov}(x, y)$$

And (noting the mean slightly increases with correlated variables):

$$\langle xy \rangle = \langle x \rangle \langle y \rangle + \text{Cov}(x, y)$$

And the one formula to rule them all, taking correlation into account:

$$\text{Var}(f) \approx \frac{\partial f}{\partial A}^2 \text{Var}(A) + \frac{\partial f}{\partial B}^2 \text{Var}(B) + 2 \frac{\partial f}{\partial A} \frac{\partial f}{\partial B} \text{Cov}(A, B)$$

Wed 15 Oct 2025 12:00

Lecture 5 - Distributions

1 Coin Flips and Probability Recap

Flipping a coin is one of the simplest distributions we can create.

Given a:

$$\begin{aligned} P(H) &= 0.5 \\ P(T) &= 0.5 \end{aligned}$$

We know that $P(HHHH) = 0.5^4$.

And $P(\text{Three Heads and One Tail}) = P(\text{HHHT or HHTH or THHH or HTHH}) = 4 \times 0.5^4$. We will take 4 coins, A, B, C, D. We denote a single result as A_{Heads} or C_{Tails} etc.

We can also say that the coins are independent, i.e. the probability of one result given another result is equal to just the probability of the first result:

$$P(A_h | B_h) = 0.5$$

The chance of A and B being heads is:

$$P(A_h \text{ and } B_h) = P(A_h) \times P(B_h) = P(HH)$$

The chance of A or B being heads is (noting or excludes the case where both are true):

$$P(A_h \text{ or } B_h) = P(A_h) + P(B_h) - P(A_h \text{ and } B_h)$$

1.1 Discrete Distribution

Lets consider flipping 4 coins and counting the number of heads. This forms a discrete distribution (where only 5 possible values are possible, 0, 1, 2, 3, 4). This distribution must be normalised (sum to 1), so:

$$\sum_r P(r) = 1$$

We can also consider the mean (expected) number of heads:

$$\langle r \rangle = \sum_r rP(r)$$

This function, $P(x)$ is called a *probability mass function*, and the sum of all values must be 1.

1.2 Continuous Distributions

Continuous distributions have similar conditions:

$$\int_{-\infty}^{\infty} P(x)dx = 1$$

$$\langle x \rangle = \int_{-\infty}^{\infty} xP(x) dx$$

And for the probability of the result lying between a and b:

$$\int_a^b P(x) dx$$

We cannot, in a continuous distribution consider the probability of an exact result, i.e. $P(x = a)$, $a \in \mathbb{R}$. As there are infinitely many possible values, the probability of any precise one is not meaningful (always zero). We therefore must always consider the probability of the result lying in some non-zero range.

$P(x)$ in this case is called a *probability density function* and the area under the PDF curve must sum to one. Note that this means that $P(x)$ at any point may exceed one, so long as the overall area is equal to 1. For some probability $a < P(x) < b$ (noting that since the $P(a)$ for any precise a is zero, the equalities being strict or not is meaningless), the probability is the area of the curve between a and b.

2 Binomial Distribution

The Binomial Distribution represents a scenario where we conduct some number of identical trials, where each trial has two possible outcomes (which we denote success and failure). For example, flipping a coin. Here:

- n - The number of trials.
- p - The probability of success.
- q - The probability of failure ($q = 1 - p$).
- r - The number of successes.

This has probability mass function:

$$P(r; n, p) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$

And has the following properties:

$$\langle r \rangle = np$$

$$Var(r) = np(1-p)$$

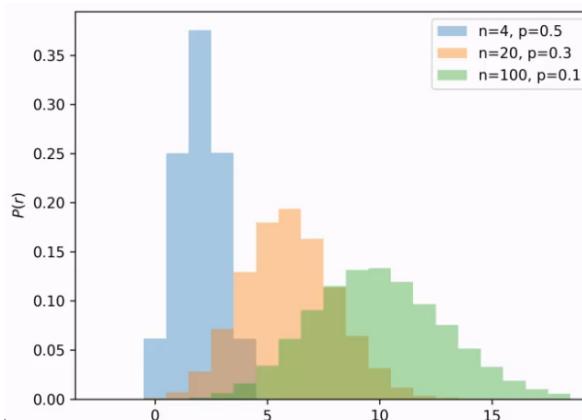


Figure 12.1: The Binomial Distribution

3 Poisson Distribution

This describes the number of events (i.e. the number of neutrinos detected by a neutrino detector) occurring in some time interval, given:

- The mean rate of events is constant.
- Each event occurs independently from the last.

This is created by taking the limit of a Binomial distribution, as:

- The number of trials tends to infinity ($n \rightarrow \infty$)
- The mean number of successes remains fixed ($np = \lambda = \text{constant}$)

Given λ as the mean number of expected events (per unit time) and r as the number of events occurring in that time, it has PMF:

$$P(r; \lambda) = \frac{\exp(-\lambda)\lambda^r}{r!}$$

And has the following properties:

$$\langle r \rangle = \lambda$$

$$\text{Var}(r) = \lambda$$

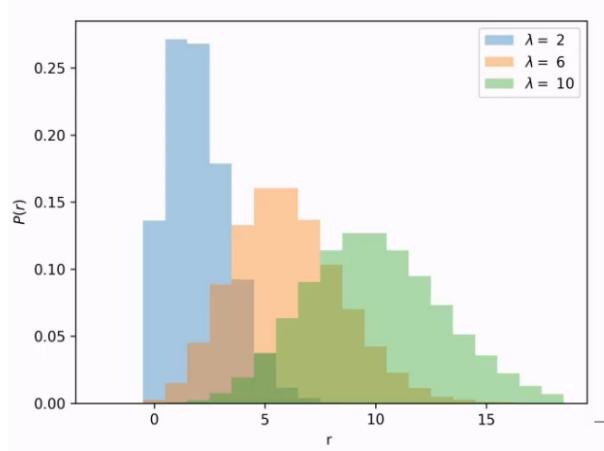


Figure 12.2: The Poisson Distribution

4 Normal Distribution

A.K.A. The Gaussian distribution. This is the most well known and most useful distribution. Given a mean μ and a standard deviation σ , the probability density function is:

$$P(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

It is a very important distribution as it arises as a result of the Central Limit Theorem, which we will cover properly in the probability section of the course. It looks like this, noting it is symmetric and forms a “bell-shaped curve”:

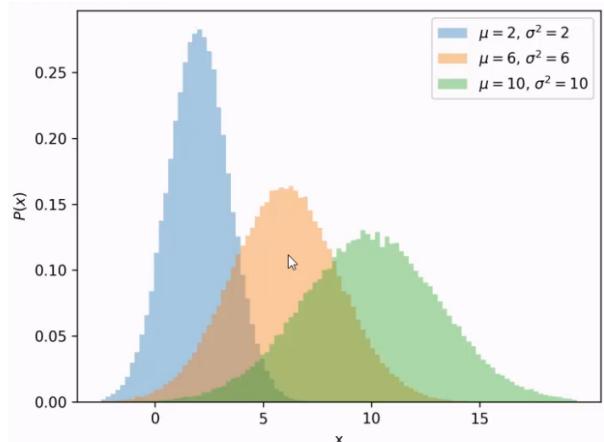


Figure 12.3: The Normal Distribution

Note that since the tails are logarithmic, they tend to zero, but never reach it truly. A Poisson distribution approaches a Normal distribution as $\lambda \rightarrow \infty$. It is generally a good approximation for $\lambda > 30$ but this depends on the application being used.

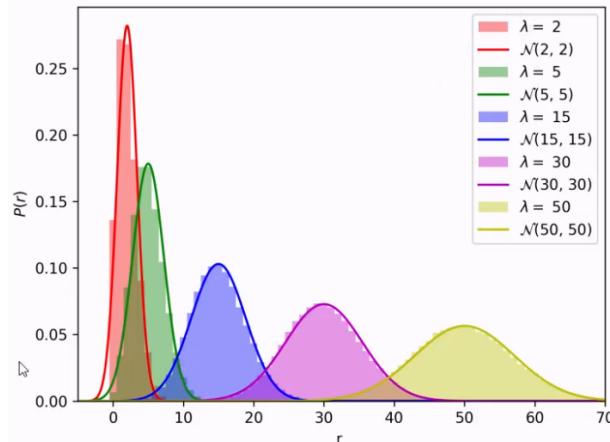


Figure 12.4: Poisson approximations to a Normal

Thu 16 Oct 2025 09:00

Lecture 6 - Likelihood and Log Likelihood

1 Likelihood

We want to fit a model to our data. We want some kind of function to specify how well this model fits the data, so that we can optimise to find the best. This is the likelihood function. There are many different ways to formulate it, but we denote it:

$$P(D \mid \theta)$$

Where D is our data, and θ is our model parameters. This is the probability of the data, given some parameters.

2 An Example

Lets say we have this model:

$$T(t) = T_{\text{env}} + (T_0 - T_{\text{env}}) \exp(-t/\tau)$$

Which represents the cooling of an object, where τ is a constant of cooling. We think we will observe some additive, normally distributed noise on these measurements, giving us:

$$T_{\text{obs}}(t) = T(t) + \epsilon$$

We may observe something like this, where the blue dots are the model-predicted values and the observed data with error noise is in black:

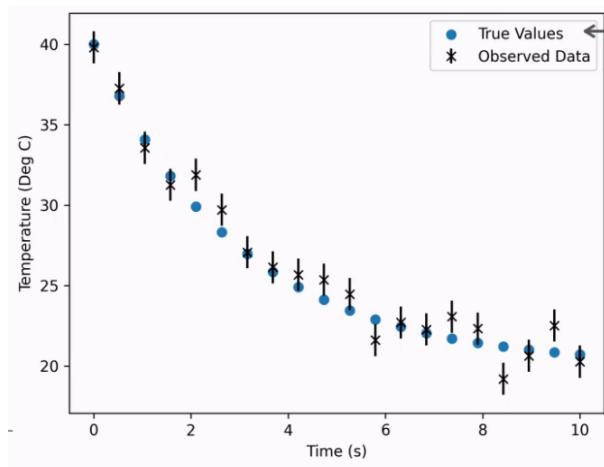


Figure 13.1: Simulated Data

Given the noise is normally distributed, we would expect the true values to lie within the error bars of our observation about 68% of the time. We see this approximately here. How can we then fit a model to this data? We need to:

1. Formulate a model.
2. Estimate a probability that the model is correct.

Given we now have data, and some model we would like to try to fit the data to (we want to fit it to Newton's Law of Cooling, the model previously, and determine an appropriate value of parameters and τ). We therefore want to find a 'merit function' to describe how good a fit any model we might create is. We start from the probability of getting some value of the noise.

As a reminder, our model is, noting we are treating time as a discrete set of times, indexed by i :

$$M(t_i, \theta) = T_{\text{env}} + (T_0 - T_{\text{env}}) \exp(-t_i/\tau)$$

And the probability of getting some value of the noise on the i th measurement is (note the first equality, where we can also write it ignoring theta, because noise is independent of the parameters):

$$P(\epsilon_i | \theta) = P(\epsilon_i) = \frac{1}{\sigma_{D_i} \sqrt{2\pi}} \exp\left(\frac{-\epsilon_i^2}{2\sigma_{D_i}^2}\right)$$

This is a Normal distribution with a mean of zero, and a standard deviation of σ_{D_i}

We cannot directly measure ϵ_i , but we know it is the difference between the measured value in the data and the 'true' value predicted by our model:

$$\epsilon_i = D_i - M(t_i, \theta)$$

Since the noise is additive and Normal, we can combine these two equations to get our merit function - the probability of a single observed data point given the parameters as:

$$P(D_i | \theta) = \frac{1}{\sigma_{D_i} \sqrt{2\pi}} \exp\left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2}\right)$$

Where θ is our list of parameters $\theta = [T_0, T_{\text{env}}, \tau]$. Crucially, this is a Normal distribution where the mean is our model's prediction given the parameters, and the standard deviation is the uncertainty on the error point. Assuming we have multiple uncorrelated data points, the total likelihood function is:

$$P(D | \theta) = \prod_{i=1}^n P(D_i | \theta)$$

Considering τ as the variable we actually change, we get:

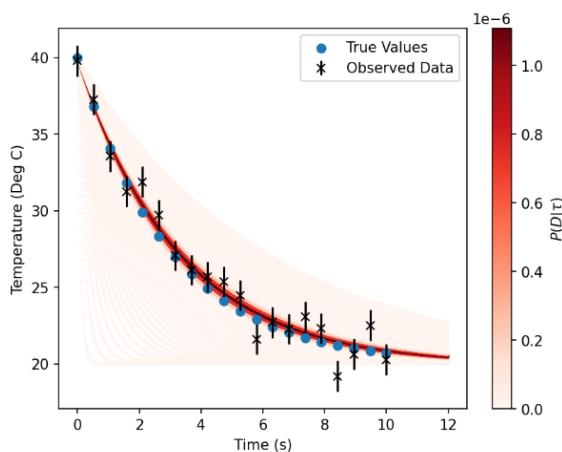


Figure 13.2: A heatmap of the likelihood function overlaid on the data.

The high likelihood values of $P(D | \tau)$ are the models which are most likely to generate the observed data, given the parameters. This, therefore, means that they are the models which best fit the data.

If we plot $P(D, \tau)$ against τ , we can see that the likelihood does a reasonable job of giving us a value which is close to true:

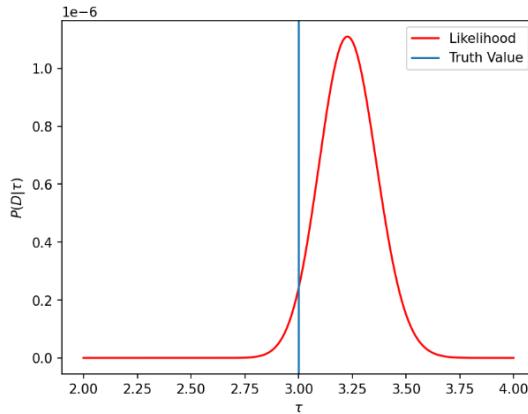


Figure 13.3

We can use Maximum Likelihood Estimation (note that it is just an estimate, the value of tau given by the maximum likelihood and the true value are **not** the same) to estimate the best value of tau for the model. We chose the value of tau that gives the maximum likelihood:

$$\hat{\tau} = \arg \max_{\tau} P(D | \tau)$$

In general, given a set of multiple parameters, we say:

$$\hat{\theta} = \arg \max_{\theta} P(D | \theta)$$

3 Log Likelihood

We still need to estimate the uncertainty on this predicted best value of tau. It turns out that a good way to do this is by taking the log likelihood instead of just the likelihood. We take the natural log of the normal probability density function for a single data point:

$$\begin{aligned} P(D_i | \theta) &= \frac{1}{\sigma_{D_i} \sqrt{2\pi}} \exp \left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2} \right) \\ \ln P(D_i | \theta) &= \ln \left(\frac{1}{\sigma_{D_i} \sqrt{2\pi}} \exp \left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2} \right) \right) \\ \ln P(D_i | \theta) &= \ln \left(\frac{1}{\sigma_{D_i} \sqrt{2\pi}} \right) + \ln \left(\exp \left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2} \right) \right) \\ \ln P(D_i | \theta) &= \left(\exp \left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2} \right) \right) - \ln(\sigma_{D_i}) - \ln(\sqrt{2\pi}) \end{aligned}$$

If we ignore the constant term, as it does not change the results (as we care about the results comparative to each other to find the maximum), and if we assume that the uncertainty is the same on each data point (which we must be careful about, in case uncertainties are variables too):

$$\ln P(D_i | \theta) = \left(\exp \left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2} \right) \right) + \text{constant}$$

This gives the final log likelihood as:

$$\mathcal{L} = \ln P(D | \theta) = \sum_{i=1}^n P(D_i | \theta) = \sum_{i=1}^n \left(\exp \left(\frac{-(D_i - M(t_i, \theta))^2}{2\sigma_{D_i}^2} \right) \right)$$

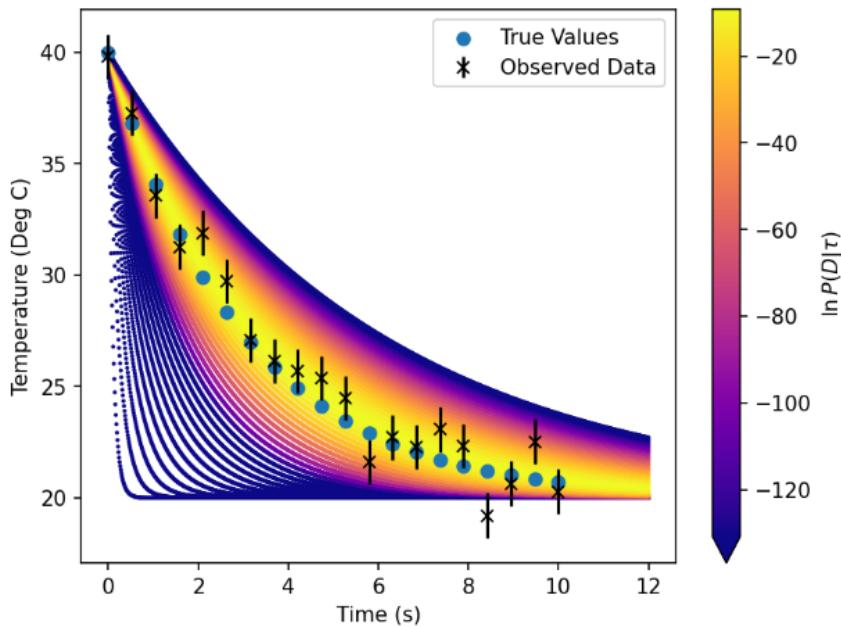


Figure 13.4: The likelihood plot, repeated with log likelihood

We can see this generates similar values to the standard likelihood, but with much friendlier values (-20 to -120, rather than very small numbers). The equation is also nicer to calculate as we're able to get rid of the constant terms. If we again consider this as a function of τ (the parameter we're actually changing to produce a fit):

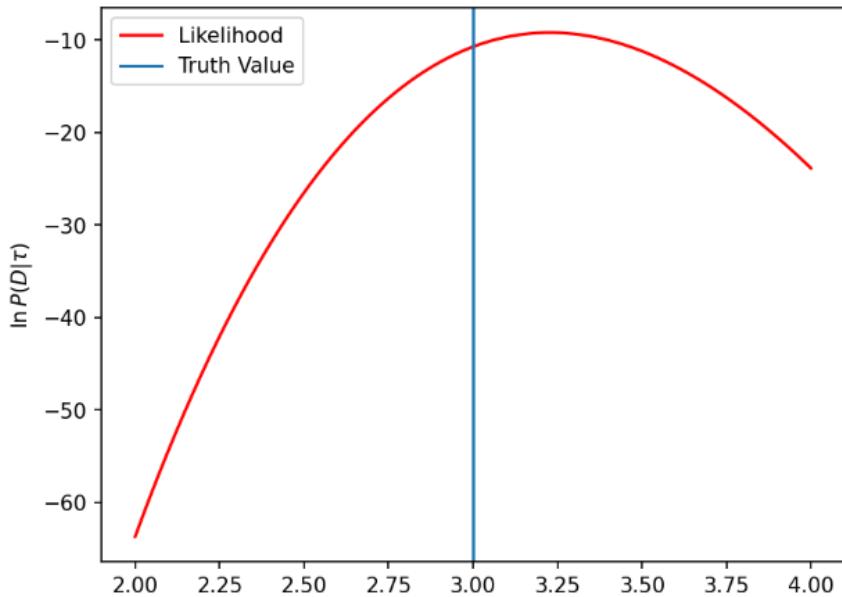


Figure 13.5

This is similar again, but with a very different order of magnitude. We can calculate the predicted optimum τ by determining the maximum point where:

$$\frac{\partial}{\partial \tau} \ln P(D | \tau) = 0$$

Note the swap from theta to tau, this is because tau is the parameter we're actually using to fit, while the other parameters bundled into theta are constant. This gives $\tau \approx 3.23$. Yes we could have done this with traditional likelihood, but the differentiation is nicer in log form and it becomes relevant later.

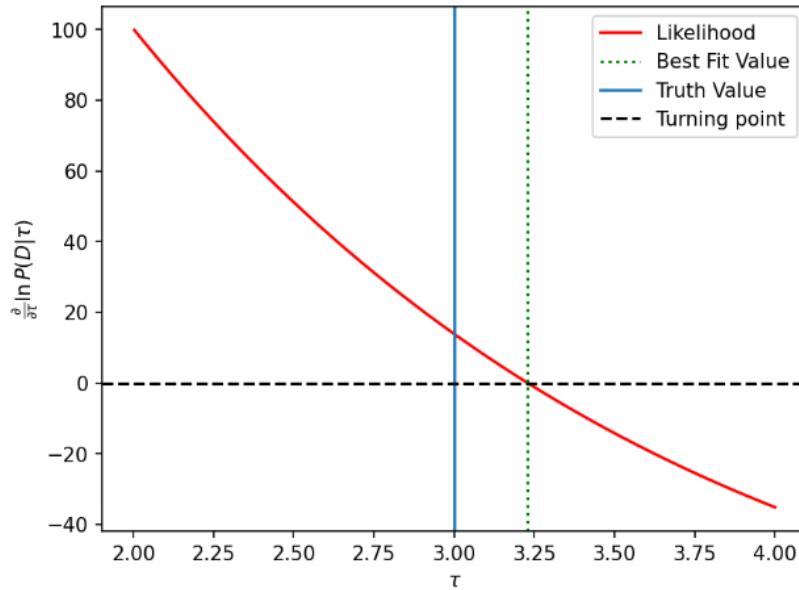


Figure 13.6

4 Uncertainties on Log Likelihood

In our single dimension problem, we'll create a new value called the Hessian (H), we define this as:

$$H = \left. \frac{\partial^2 \mathcal{L}}{\partial \tau^2} \right|_{\tau=\hat{\tau}}$$

Why is this (and the log likelihood) actually useful? It turns out that the curvature of the log likelihood around the maximum point tells us the uncertainty. We can use the Hessian and log likelihood to estimate the uncertainty on the parameter τ , using the inverse of H , H^{-1} where $HH^{-1} = 1$.

$$\sigma_{\hat{\tau}} \approx \sqrt{|H^{-1}|}$$

Since we're in 1D, this value of the Hessian is a scalar, and we can estimate it by either differentiating twice or estimating using two points:

$$f''(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}$$

And calculating:

$$\sigma_{\hat{\tau}} = \sqrt{\frac{1}{|H|}}$$

But note that this works differently in higher dimensions (more parameters) has H becomes a matrix. This gives us $\hat{\tau} \approx 3.23$ from the previous result, and now $\sigma_{\hat{\tau}} \approx 0.133$. Note that the true value is 1.8 standard deviations away from the best estimate, which is okay - anything larger than 3 sigma away would start to become worrying.

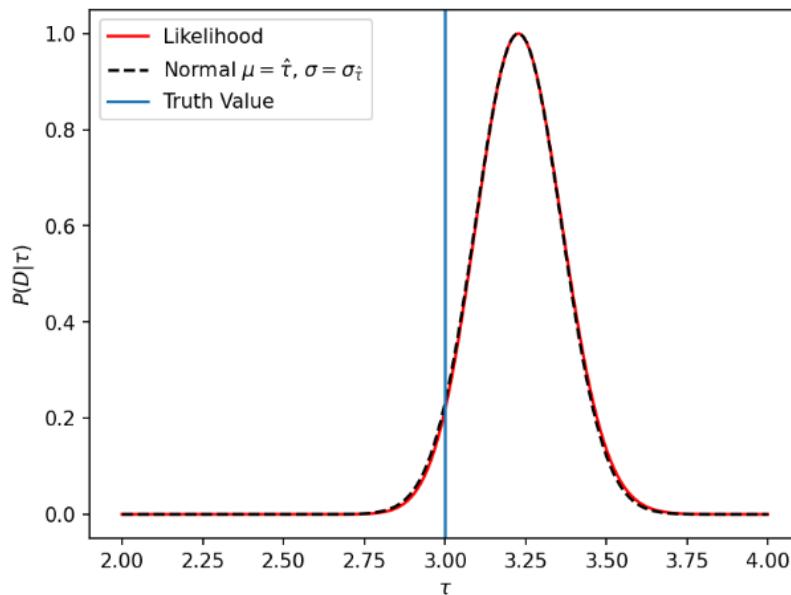


Figure 13.7

Using the log likelihood has:

- Been a nicer calculation, which is computationally easier, as we can strip out all the constant terms.
- Allowed us to calculate the uncertainties on our predicted values.
- Allowed us to get back to the standard likelihood (as the black and red curves above are approximately the same) anyways.

4.1 Multidimensional Generalisation

This works in 1D, but we can generalise to a higher number of parameters. We assume here however that $P(D | \theta)$ is going to be a normal distribution. If this is not the case, we cannot use the approximations for uncertainties, and the problem becomes too complex for first year stats. It is generally the case that it either will approximate a normal as many datapoints are taken. If this isn't true, it becomes a problem for Y4 Bayesian Stats.

We find the maximum likelihood (given multiple parameters $\hat{\theta}$) with:

$$\frac{\partial \mathcal{L}}{\partial \theta_1} = \frac{\partial \mathcal{L}}{\partial \theta_2} = \dots = 0$$

And the Hessian is given by:

$$\mathbf{H}(\hat{\theta}) = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \theta_1^2} \Big|_{\theta=\hat{\theta}} & \frac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_2} \Big|_{\theta=\hat{\theta}} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_k} \Big|_{\theta=\hat{\theta}} \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_2 \partial \theta_1} \Big|_{\theta=\hat{\theta}} & \frac{\partial^2 \mathcal{L}}{\partial \theta_2^2} \Big|_{\theta=\hat{\theta}} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial \theta_2 \partial \theta_k} \Big|_{\theta=\hat{\theta}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_k \partial \theta_1} \Big|_{\theta=\hat{\theta}} & \frac{\partial^2 \mathcal{L}}{\partial \theta_k \partial \theta_2} \Big|_{\theta=\hat{\theta}} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial \theta_k^2} \Big|_{\theta=\hat{\theta}} \end{bmatrix}.$$

Wed 22 Oct 2025 12:00

Lecture 7 - Fitting a Straight Line 1

We want to create a model for a straight line:

$$M(x, \theta) = mx + c$$

Where are datapoints are given by this model and some additive noise:

$$D = M(x, \theta) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma_i)$$

The general recipe for line fitting is given by:

1. A generative model for the data, with knowledge of how the noise is distributed.
2. Likelihood function.
3. A method for finding the maximum likelihood.
4. Method for finding the uncertainties on best fit parameters.
5. A method for checking how good the fit is.

We can write down the likelihood function for this model as:

$$\begin{aligned} P(D_i | \theta) &= \frac{1}{\sigma_{D_i} \sqrt{2\pi}} \exp\left(\frac{-(D_i - M(x_i, \theta))^2}{2\sigma_{D_i}^2}\right) \\ P(D | \theta) &= \prod_{i=1}^n P(D_i | \theta) \end{aligned}$$

And again:

$$\mathcal{L} = \ln P(D | \theta) = \sum_{i=1}^n \ln P(D_i | \theta) \propto \sum_{i=1}^n \left(\frac{-(D_i - M(x_i, \theta))^2}{2\sigma_{D_i}^2} \right)$$

We want to find the parameters of distribution that maximise the (log)likelihood:

$$\hat{\theta} = \arg \max_{\theta} P(D | \theta)$$

Or:

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}$$

There are a number of different approaches to do this:

- Find where all first derivatives equal zero (as last lecture, various clever algorithms to do so).
- Brute force on a grid.
- Iterative or stochastic methods.
- Analytic maximisation for a simple linear model - see next lecture.

1 Finding Maximum Likelihood

The most crude way to do this is to build a grid of all values of m and c , and iterate through over all points (with some resolution) to find the maximum likelihood generated by them. Plotting M (as colour):

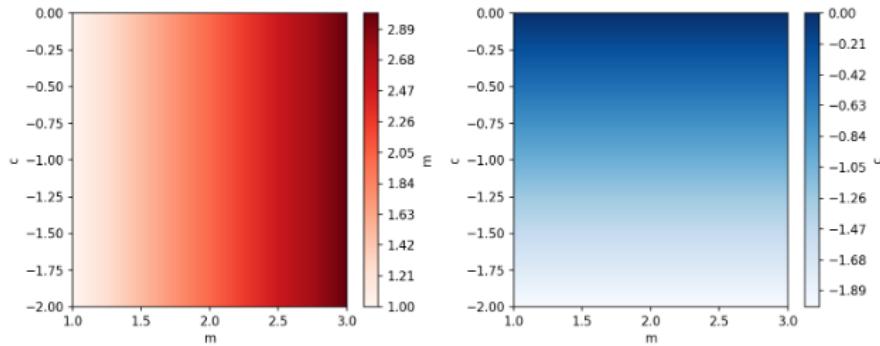


Figure 14.1

And the likelihood:

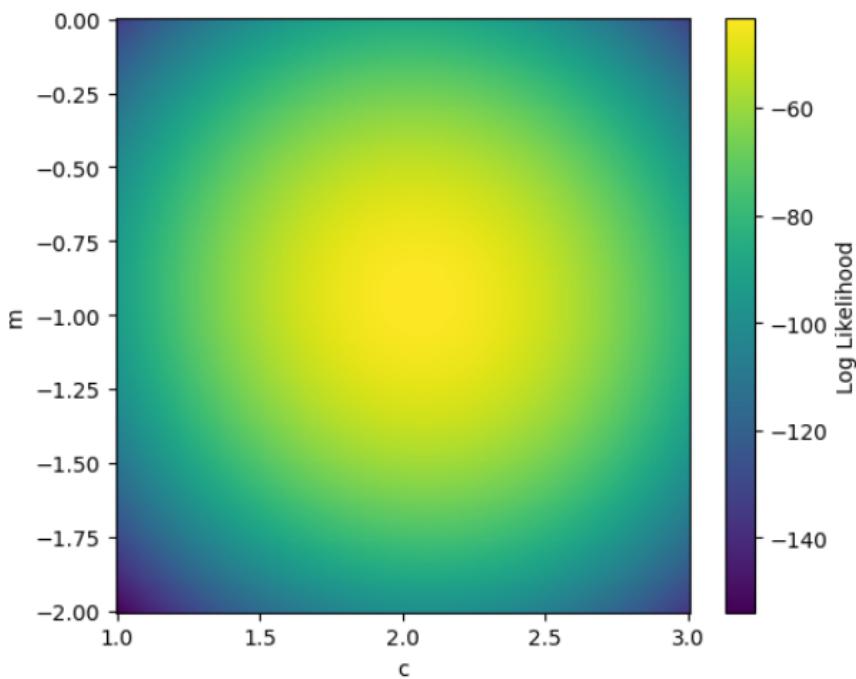


Figure 14.2

We assume that the grid point with the highest likelihood and the true point with the highest likelihood are the same. In this case, the grid resolution is small enough that this is true, but it may not always be. This gives:

$$m = 2.0552763819095476$$

$$c = -0.9447236180904524$$

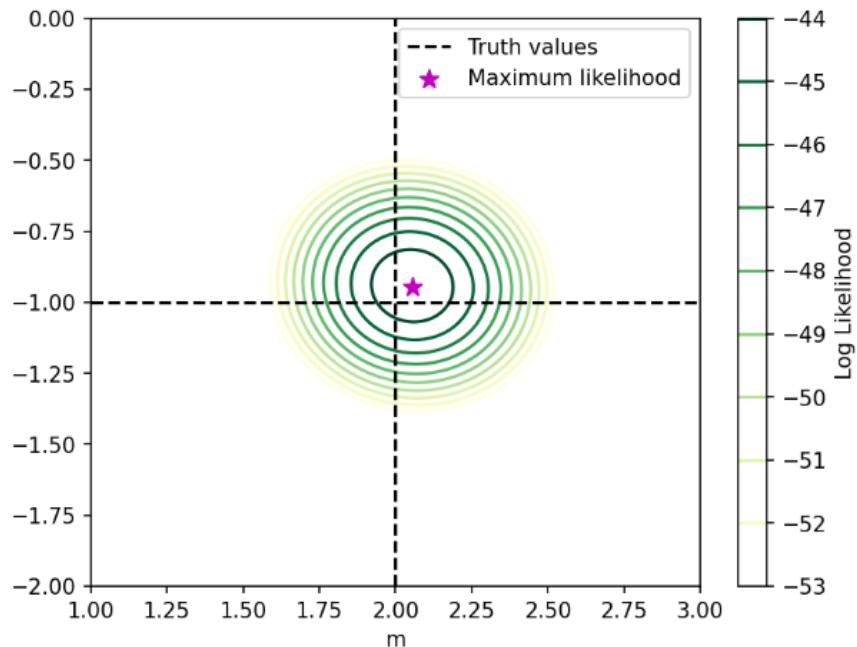


Figure 14.3

We see that this does a good, but not perfect job, of fitting the data. This is due to noise in the data, and is acceptable, provided it's within a reasonable uncertainty.

Thu 23 Oct 2025 09:00

Lecture 8 - Fitting a Straight Line 2

1 Uncertainties on Best Fit Parameters

1.1 What do these uncertainties actually mean?

Previously, we found that the likelihood of a single value could be described by a Normal distribution.

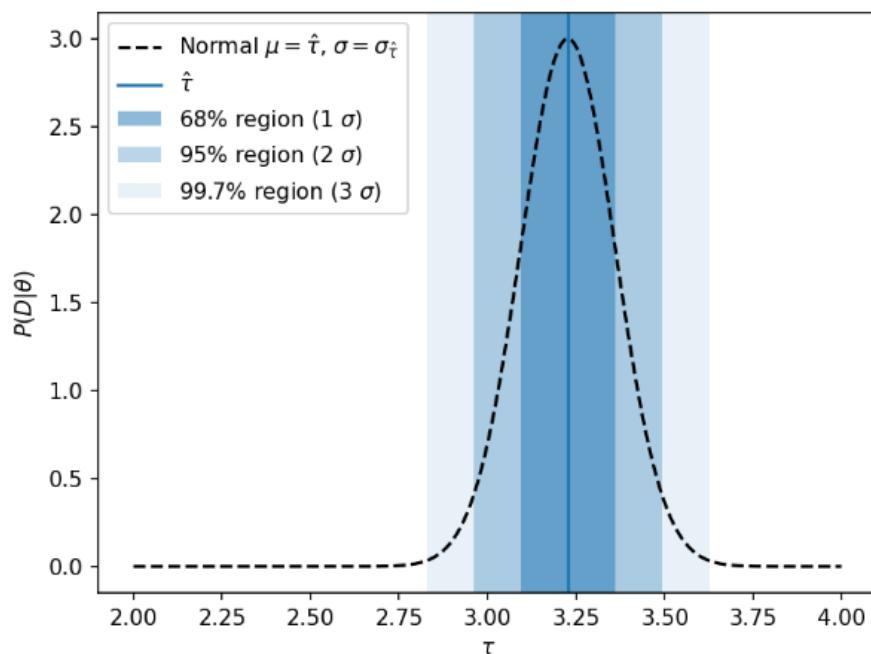


Figure 15.1

We can therefore quote the best fit value as $\hat{\tau} = 3.227 \pm 0.133$. What we're effectively saying is that our 1 sigma uncertainty encompasses 68% of the probability density such that:

$$\int_{\hat{\tau}-\sigma_{\hat{\tau}}}^{\hat{\tau}+\sigma_{\hat{\tau}}} P(D | \tau) d\tau \approx 0.68$$

And the same for 2 sigma uncertainty with 0.95, and 3 sigma uncertainty with 0.997. We can therefore say that while it may be common for the value to lie outside the 1 sigma uncertainty, it is rare for it to lie outside the 3 sigma uncertainty and if this happens (including error), something probably went wrong with our measurement.

Here, where we're attempting to determine an uncertainty, we instead increase the value of $\sigma_{\hat{\tau}}$ until this first integral is satisfied.

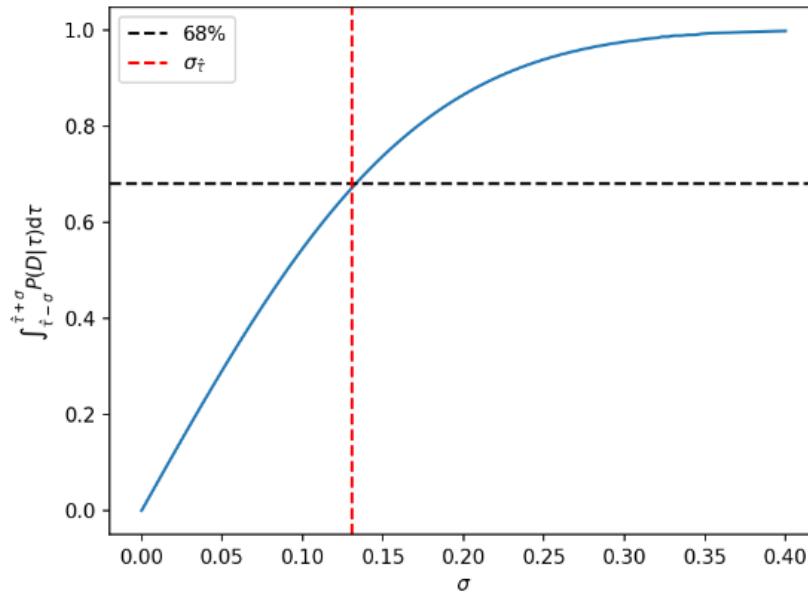


Figure 15.2

However, with a line of best fit, we want to consider uncertainties in two dimensions. This makes life a little bit more difficult, we want some boundary on the parameter space that is centred on the best fit parameters and encompasses 68% of the whole probability space.

Plotting the same distribution again, but including the 1D likelihood for each parameter individually:

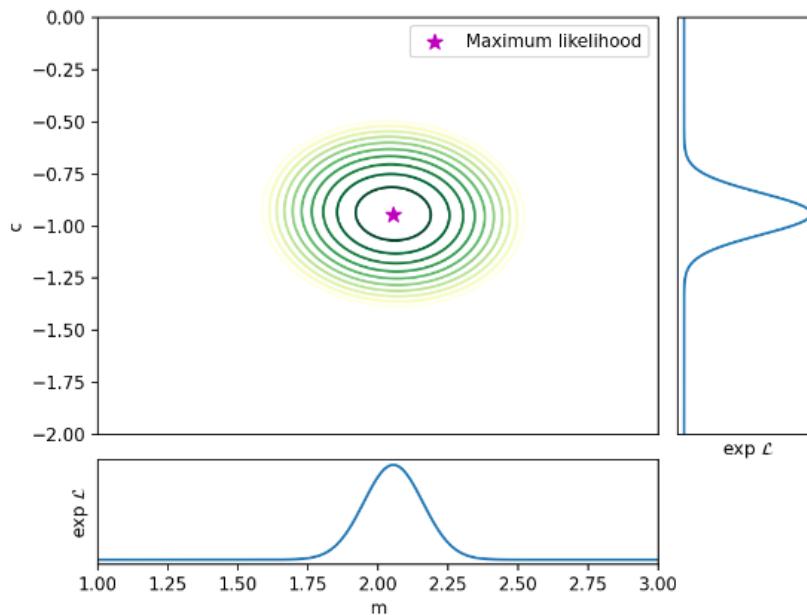


Figure 15.3

In general, we shouldn't just separately calculate the 1 dimensional approach per parameter and combine them (as the parameters may be correlated), but it's a useful starting point. Swapping to log likelihood:

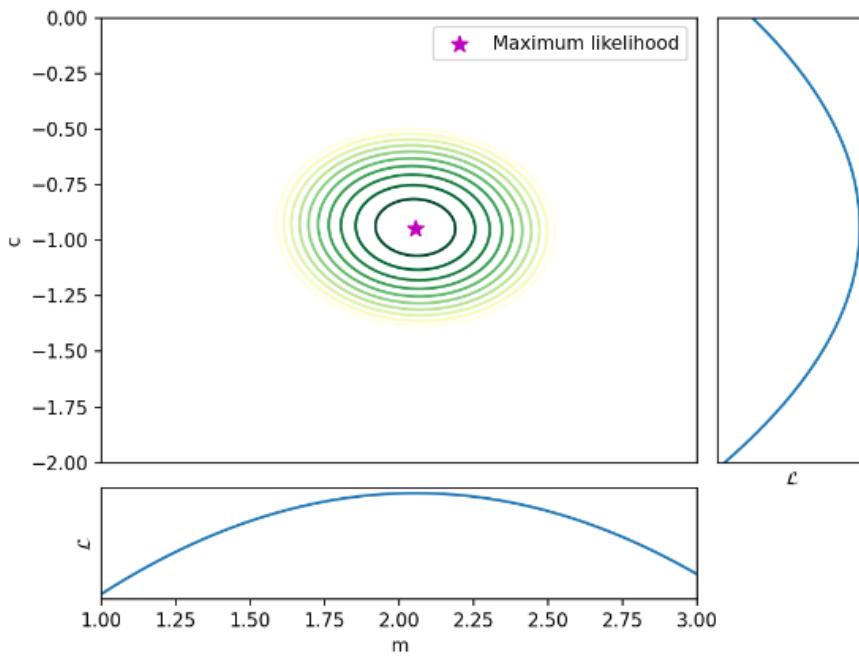


Figure 15.4

We want to estimate:

$$\left. \frac{\partial^2 \mathcal{L}}{\partial m^2} \right|_{\theta=\hat{\theta}}$$

$$\left. \frac{\partial^2 \mathcal{L}}{\partial c^2} \right|_{\theta=\hat{\theta}}$$

While we shouldn't use these 1D distributions to estimate the best fit parameters, this problem has deliberately been created to minimise the correlation between m and c and creating something with them that describes $P(D | \theta)$ is still instructive. We can however create a grid of the values of m and c , calculate the likelihood for all points with some resolution and pick the maximum. This is slow, but crucially does find the maximum point of the 3D distribution surface and is not the same as finding the max for each variable individually.

We assume that $P(D | \theta)$ can be described by a two dimensional normal distribution, and we can build this from two discrete normal distributions of two independent variables (with the caveats above). We say that the mean value of $P(D | m)$ is \hat{m} and the uncertainty is therefore given by:

$$\sigma_m^2 = \left(-\left. \frac{\partial^2 \mathcal{L}}{\partial m^2} \right|_{\theta=\hat{\theta}} \right)^{-1}$$

And likewise for c , \hat{c} , σ_c^2

$$\sigma_c^2 = \left(-\left. \frac{\partial^2 \mathcal{L}}{\partial c^2} \right|_{\theta=\hat{\theta}} \right)^{-1}$$

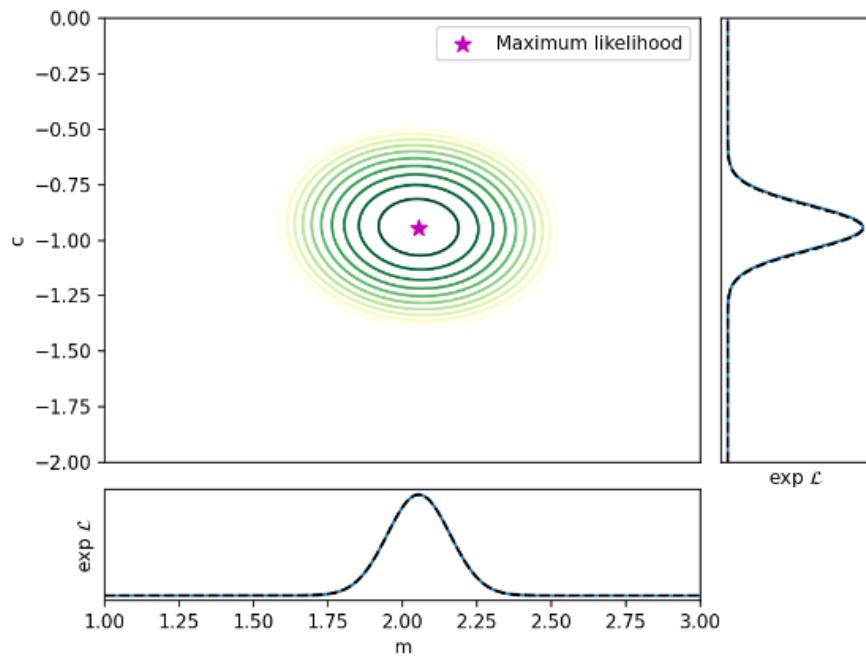


Figure 15.5

Calculating our summary stats and plotting them:

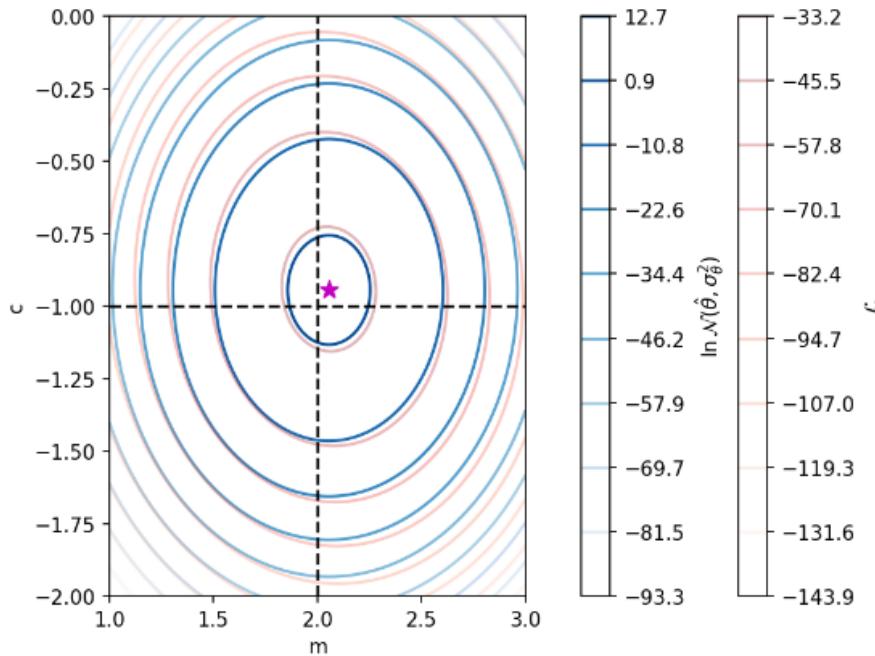


Figure 15.6

We can see that this is pretty good agreement between our estimate and the actual log likelihood function. We now have everything we need to quote the best fit parameters and (crucially) their uncertainty.

Wed 29 Oct 2025 12:00

Lecture 9 - Linear Regression

Previously, we've used numerical methods to determine the best fit parameters for a line of best fit to some data. This is good, because it easily generalises to more complex problems, however lines of best fit have a more specific (and easier to implement) algebraic method.

1 Method

And, we have to perform the same broad steps:

1. A generative model for the data, with knowledge of how the noise is distributed.
2. Likelihood function.
3. A method for finding the maximum likelihood.
4. Method for finding the uncertainties on best fit parameters.
5. A method for checking how good the fit is.

1.1 Generative Model

We use the same generative model as before, with a straight line fit with some additively generated noise (with a standard deviation that may differ from point to point, σ_{D_i}):

$$M(x, \theta) = mx + c$$

1.2 Likelihood Function

Using the same likelihood and log likelihood formulae has before, we can take this one step further by defining (as the extra factor of -2 does not matter when calculating the maxima/minima and ignoring it will make the algebra nicer):

$$\chi^2 = -2\mathcal{L} = \sum_{i=1}^n \left(\frac{D_i - M(x_i, \theta)}{\sigma_{D_i}} \right)^2$$

1.3 Finding Maximum Likelihood (Minimum χ^2)

Since our model is linear, there will only be one turning point for χ^2 , so we can be sure that the minimum of χ^2 will be at the point where the first derivatives (wrt m and c) are zero.

$$\frac{\partial(\chi^2)}{\partial m} = \frac{\partial(\chi^2)}{\partial c} = 0$$

Rather than doing this numerically, as last lecture, we can do it algebraically:

$$\chi^2 = \sum_{i=1}^n \left(\frac{D_i - M(x_i, \theta)}{\sigma_{D_i}} \right)^2 = \sum_{i=1}^n \left(\frac{D_i - mx_i - c}{\sigma_{D_i}} \right)^2$$

w.r.t m :

$$\begin{aligned}\frac{\partial(\chi^2)}{\partial m} &= \frac{\partial}{\partial m} \sum_{i=1}^n \left(\frac{D_i - mx_i - c}{\sigma_{D_i}} \right)^2 = 0 \\ &= \sum_{i=1}^n \frac{\partial}{\partial m} \left(\frac{D_i - mx_i - c}{\sigma_{D_i}} \right)^2\end{aligned}$$

Let $u_i = (D_i - mx_i - c)/\sigma_{D_i}$, so $\partial u_i / \partial m = -x_i / \sigma_{D_i}$:

$$\begin{aligned}&= \sum_{i=1}^n \frac{\partial u_i}{\partial m} \frac{\partial}{\partial u_i} u_i^2 \\ &= \sum_{i=1}^n \frac{\partial u_i}{\partial m} (2u_i) \\ &= \sum_{i=1}^n \left(\frac{-x_i}{\sigma_{D_i}} \right) (2u_i) \\ &\quad -2 \sum_{i=1}^n \frac{x_i u_i}{\sigma_{D_i}}\end{aligned}$$

So:

$$\frac{\partial(\chi^2)}{\partial m} = -2 \sum_{i=1}^n \frac{x_i}{\sigma_{D_i}} \frac{D_i - mx_i - c}{\sigma_{D_i}}$$

$$\boxed{\frac{\partial(\chi^2)}{\partial m} = -2 \sum_{i=1}^n \left(\frac{x_i(D_i - mx_i - c)}{\sigma_{D_i}^2} \right) = 0}$$

w.r.t c :

$$\begin{aligned}\frac{\partial(\chi^2)}{\partial c} &= \frac{\partial}{\partial c} \sum_{i=1}^n \left(\frac{D_i - mx_i - c}{\sigma_{D_i}} \right)^2 = 0 \\ &= \sum_{i=1}^n \frac{\partial}{\partial c} \left(\frac{D_i - mx_i - c}{\sigma_{D_i}} \right)^2\end{aligned}$$

Let $u_i = (D_i - mx_i - c)/\sigma_{D_i}$, so $\partial u_i / \partial c = -1 / \sigma_{D_i}$:

$$\begin{aligned}&= \sum_{i=1}^n \frac{\partial u_i}{\partial c} \frac{\partial}{\partial u_i} u_i^2 \\ &= \sum_{i=1}^n \frac{\partial u_i}{\partial c} (2u_i) \\ &= \sum_{i=1}^n \left(\frac{-1}{\sigma_{D_i}} \right) (2u_i) \\ &= -2 \sum_{i=1}^n \frac{u_i}{\sigma_{D_i}}\end{aligned}$$

So:

$$\boxed{\frac{\partial(\chi^2)}{\partial c} = -2 \sum_{i=1}^n \frac{(D_i - mx_i - c)}{\sigma_{D_i}^2} = 0}$$

2 Putting it All Together

We now have two simultaneous equations with two variables (m, c), so we can solve for the optimum values:

$$\frac{\partial(\chi^2)}{\partial c} = -2 \sum_{i=1}^n \frac{(D_i - mx_i - c)}{\sigma_{D_i}^2} = 0$$

$$\frac{\partial(\chi^2)}{\partial m} = -2 \sum_{i=1}^n \frac{x_i(D_i - mx_i - c)}{\sigma_{D_i}^2} = 0$$

We can make some substitutions for ease:

$$S = \sum_{i=1}^n \frac{1}{\sigma_{D_i}^2}$$

$$S_x = \sum_{i=1}^n \frac{x_i}{\sigma_{D_i}^2}$$

$$S_{xx} = \sum_{i=1}^n \frac{x_i^2}{\sigma_{D_i}^2}$$

$$S_D = \sum_{i=1}^n \frac{D_i}{\sigma_{D_i}^2}$$

$$S_{Dx} = \sum_{i=1}^n \frac{x_i D_i}{\sigma_{D_i}^2}$$

2.1 Subbing into $\partial(\chi^2)/\partial c$

$$-2 \sum_{i=1}^n \frac{(D_i - mx_i - c)}{\sigma_{D_i}^2} = 0$$

$$\sum_{i=1}^n \frac{(D_i - mx_i - c)}{\sigma_{D_i}^2} = 0$$

$$S_D - mS_x - cS = 0 \implies S_D = mS_x + cS$$

2.2 Subbing into $\partial(\chi^2)/\partial m$

$$-2 \sum_{i=1}^n \frac{x_i(D_i - mx_i - c)}{\sigma_{D_i}^2} = 0$$

$$\sum_{i=1}^n \frac{x_i(D_i - mx_i - c)}{\sigma_{D_i}^2} = 0$$

$$S_{Dx} - mS_{xx} - cS_x = 0 \implies S_{Dx} = mS_{xx} + cS_x$$

2.3 Combining

We now have two simultaneous equations to solve:

$$\begin{cases} S_D = mS_x + cS & (1) \\ S_{Dx} = mS_{xx} + cS_x & (2) \end{cases}$$

Rearranging (1) gives:

$$c = \frac{S_d - mS_x}{S} \quad (3)$$

And (3) into (2):

$$S_{Dx} = mS_{xx} + \frac{S_d - mS_x}{S} S_x$$

$$\begin{aligned}
 S_{Dx} &= mS_{xx} + \frac{S_D S_x}{S} - \frac{m(S_x)^2}{S} \\
 S_{Dx} - \frac{S_D S_x}{S} &= mS_{xx} - \frac{m(S_x)^2}{S} \\
 S_{Dx} - \frac{S_D S_x}{S} &= m \left(S_{xx} - \frac{(S_x)^2}{S} \right) \\
 m &= \left(S_{Dx} - \frac{S_D S_x}{S} \right) / \left(S_{xx} - \frac{(S_x)^2}{S} \right)
 \end{aligned}$$

Simplifying to:

$$m = \frac{SS_{Dx} - S_D S_x}{SS_{xx} - S_x^2}$$

And for c , rearranging (2) gives:

$$m = \frac{S_D - cS}{S_x} \quad (4)$$

(4) into (2)

$$\begin{aligned}
 S_{Dx} &= \frac{S_D - cS}{S_x} S_{xx} + cS_x \\
 S_{Dx} &= \frac{S_D S_{xx}}{S_x} - \frac{cSS_{xx}}{S_x} + cS_x \\
 S_{Dx} - \frac{S_D S_{xx}}{S_x} &= cS_x - \frac{cSS_{xx}}{S_x} \\
 S_{Dx} - \frac{S_D S_{xx}}{S_x} &= c \left(S_x - \frac{SS_{xx}}{S_x} \right) \\
 c &= \left(S_{Dx} - \frac{S_D S_{xx}}{S_x} \right) / \left(S_x - \frac{SS_{xx}}{S_x} \right)
 \end{aligned}$$

Simplifying to:

$$c = \frac{S_D S_{xx} - S_x S_{Dx}}{SS_{xx} - S_x^2}$$

2.4 And Finally...

To simplify, let $\Delta = SS_{xx} - S_x^2$:

$$\begin{aligned}
 \hat{c} &\equiv \langle c \rangle = \frac{S_D S_{xx} - S_x S_{Dx}}{\Delta} \\
 \hat{m} &\equiv \langle m \rangle = \frac{SS_{Dx} - S_D S_x}{\Delta}
 \end{aligned}$$

We have therefore managed to calculate the best fit parameters \hat{m} and \hat{c} in closed form without any numerical methods.

3 Uncertainties on Best Fit Parameters

Given we're now in 2D, the Hessian matrix is given as:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \theta_1^2} \Big|_{\theta=\hat{\theta}} & \frac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_2} \Big|_{\theta=\hat{\theta}} \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_2 \partial \theta_1} \Big|_{\theta=\hat{\theta}} & \frac{\partial^2 \mathcal{L}}{\partial \theta_2^2} \Big|_{\theta=\hat{\theta}} \end{bmatrix}$$

And the covariance matrix Σ is given as:

$$\Sigma = -\mathbf{H}^{-1}$$

We therefore need to calculate the relevant second derivatives. Note that while we could ignore the -2 term before, as it did not matter for finding the location of the maximum, it does matter for errors and cannot be left off. We therefore go back to working in \mathcal{L} and not χ^2

We know the first derivatives of the log likelihood are:

$$\frac{\partial(\chi^2)}{\partial m} = -2(S_{Dx} - mS_{xx} - cS_x) \implies \frac{\partial\mathcal{L}}{\partial m} = S_{Dx} - mS_{xx} - cS_x$$

$$\frac{\partial(\chi^2)}{\partial c} = -2(S_D - mS_x - cS) \implies \frac{\partial\mathcal{L}}{\partial c} = S_D - mS_x - cS$$

Taking second derivatives:

$$\frac{\partial^2\mathcal{L}}{\partial m^2} = \frac{\partial}{\partial m}(S_{Dx} - mS_{xx} - cS_x) = -S_{xx}$$

$$\frac{\partial^2\mathcal{L}}{\partial c^2} = \frac{\partial}{\partial c}(S_D - mS_x - cS) = -S$$

And for the term wrt both variables:

$$\frac{\partial^2\mathcal{L}}{\partial m\partial c} = \frac{\partial^2\mathcal{L}}{\partial c\partial m} = \frac{\partial}{\partial c}(S_{Dx} - mS_{xx} - cS_x) - S_x$$

Hence (taking $\theta_1 = m$, $\theta_2 = c$):

$$\mathbf{H} = \begin{bmatrix} -S_{xx} & -S_x \\ -S_x & -S \end{bmatrix}$$

And finally:

$$\Sigma = -\mathbf{H}^{-1} = \frac{1}{SS_{xx} - S_x^2} \begin{bmatrix} S & -S_x \\ -S_x & S_{xx} \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} S & -S_x \\ -S_x & S_{xx} \end{bmatrix},$$

Therefore (from the definitions of the covariance matrix):

$$Var(m) = \Sigma_{11} = \frac{S}{\Delta}$$

$$Var(c) = \Sigma_{22} = \frac{S_{xx}}{\Delta}$$

$$Cov(m, c) = \Sigma_{12} = \Sigma_{21} = -\frac{S_x}{\Delta}$$

And using the definition of correlation:

$$Cor(m, c) = \frac{Cov(m, c)}{\sqrt{Var(m)Var(c)}} = \frac{-S_x}{\sqrt{SS_{xx}}}$$

Thu 30 Oct 2025 09:00

Lecture 10 - Goodness of Fit

Using likelihood, we can quantify how close our model is to the data, but how do we know if we've got the right model in the first place? We can use a *goodness of fit statistic* to quantify this probabilistically.

1 Straight Line Example

Lets start by taking some data and fitting a straight line to it, using last lectures content:

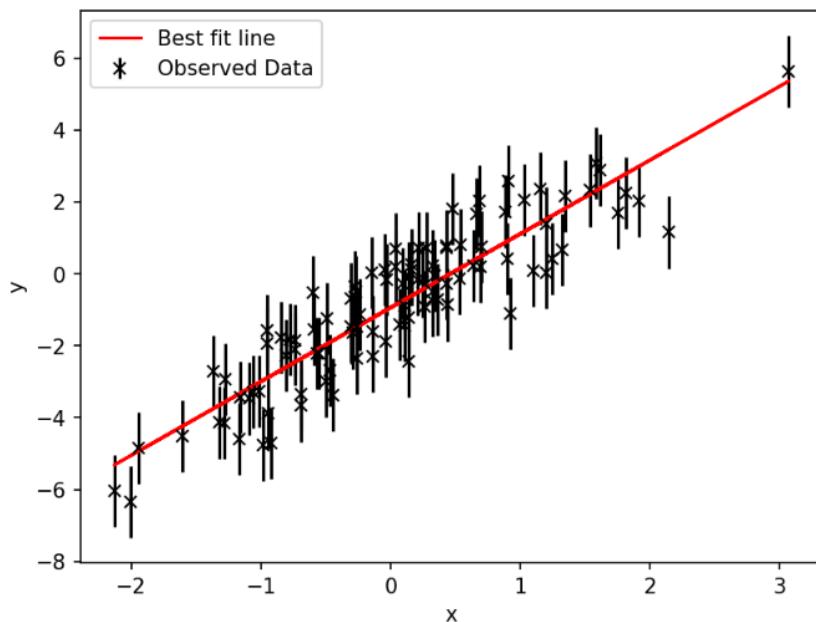


Figure 17.1

We can calculate the χ^2 value for this fit with:

$$\chi^2 = \sum_{i=1}^n \left(\frac{D_i - M(x_i, \theta)}{\sigma_{D_i}} \right)^2$$

Where $M(x_i, \theta) = \hat{m}x_i + \hat{c}$. This gives $\chi^2 \approx 86.37$. We can generate many different datasets, and create lines of best fit for them, and calculate each fit's χ^2 . If we plot these values as a histogram we get:

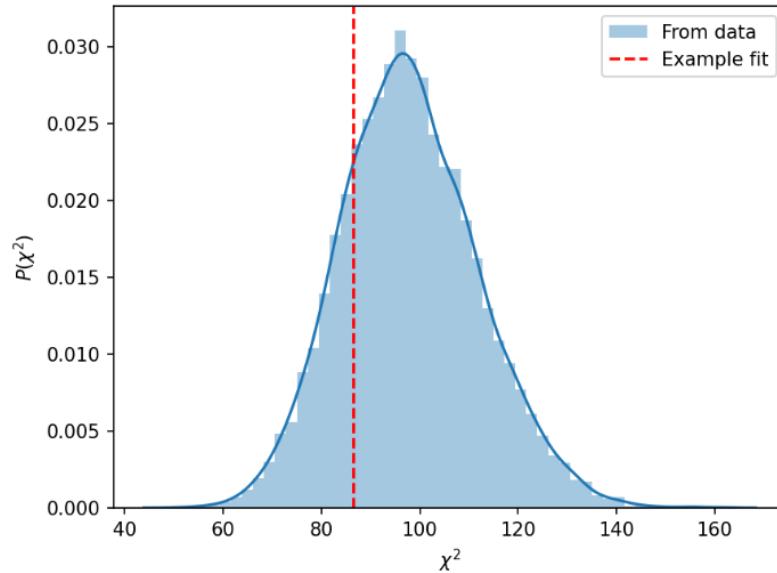


Figure 17.2

This is called a “ χ^2 distribution” and is generated by the frequencies of χ^2 values across many realisations of the data, if the model fitted to the data is the same model actually used to generate the data. A chi squared distribution can be described using the number of “degrees of freedom”, k , and we denote a chi squared distribution with k degrees as χ_k^2 . The PDF is given by:

$$P(x; k) = \frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-x/2}$$

Where $\Gamma(x)$ is the gamma function, an interpolation of the factorial function across all reals. The number of degrees of freedom is given by the number of data points minus the number of fitted parameters, i.e. $k = N - M$. Comparing our calculated distribution to the theoretical distribution given by the PDF gives:

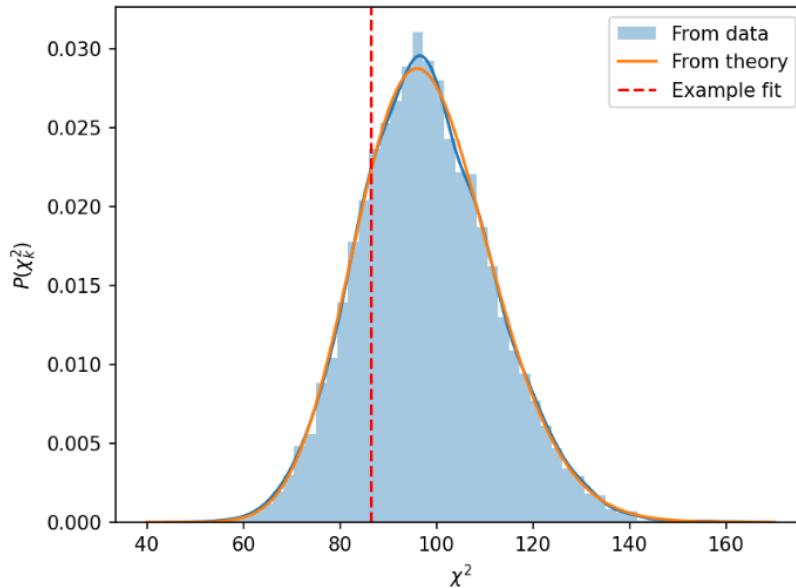


Figure 17.3

Which is a pretty good fit. We can also look at this distribution for different numbers of data points, where a larger number of data points gives a larger number of degrees of freedom:

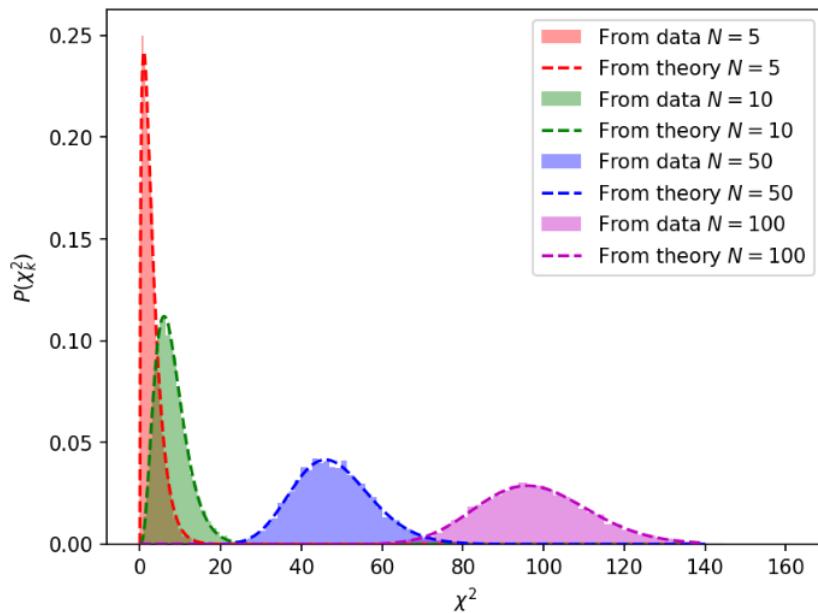


Figure 17.4

If we are fitting the correct model to our data, then we know the distribution of expected χ^2 values is (a χ_k^2 distribution). This provides a goodness of fit statistic, by checking to see if the χ^2 value from the best fit parameters is consistent with what we'd expect to see if our model was correctly chosen.

If assume our model is correct, we can ask how likely it is that we'd get the data χ^2 value from the χ_k^2 likelihood. If this is reasonably likely, we say the goodness of fit is acceptable. If not, then we have a problem. This problem could be many things, including:

- The incorrect model is being fitted - *larger χ^2 than expected*.
- The uncertainties on the data are too small to account for the observed noise - *larger χ^2 than expected*.
- The uncertainties on the data are too large so account for more than the actual noise - *smaller χ^2 than expected*.
- Something else has gone wrong...

The test itself cannot tell us exactly which of these is true, we have to use scientific judgement.

2 Quantifying the Likelihood

Since the χ_k^2 distribution is continuous, the likelihood of getting a specific χ^2 value is zero. Instead, we reframe and look at "what is the chance of getting this value of χ^2 or larger?". This is done by:

$$P(\chi^2 \geq a) = \int_a^{\infty} P(\chi^2; k) d\chi^2$$

Crucially, **the probability returned is only valid on the assumption that the model fitted to the data is correct**. If this value is above a certain threshold, we say this is evidence of a sensible fit. This is **not the chance that the model is correct, as the model is just that, a model, and is almost certainly never entirely correct**. It is the chance of getting this χ^2 value or larger *if the model is correct*.

Taking the previous example of $\chi^2 = 86.37$, with $k = 100 - 2 = 98$. This gives:

$$P(\chi^2 \geq 86.37) = \int_{86.37}^{\infty} P(\chi^2; 98) d\chi^2$$

This needs to be evaluated numerically, as it becomes unpleasant for non-trivial values of k . This gives a probability of 0.793. This is a reasonable value, not worryingly high (i.e. better than 99%) or worryingly low (i.e. less than 1%), so we say the fit is adequate.

To counterexample, say we have a poor fit. The data is from $y = 0.85x^2 + 2x - 1$, and we try to fit a linear model to it:

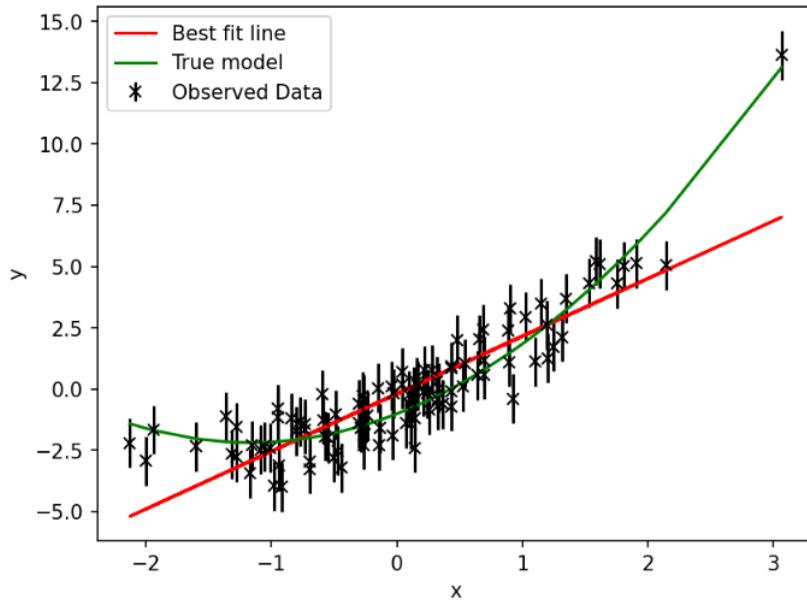
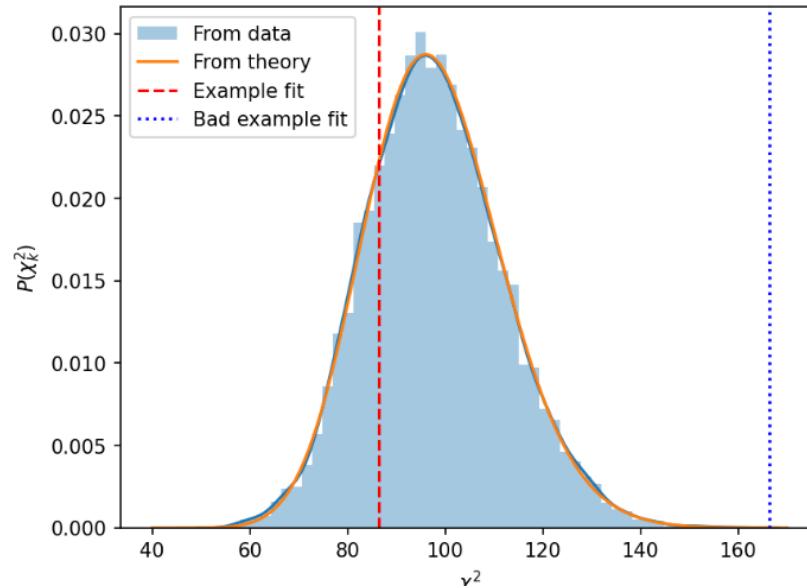


Figure 17.5

This gives a χ^2 value of 166.407. This is much larger than before, lets plot it on the distribution of the expected χ^2 :

Figure 17.6: Probability is $1.98e - 05$

So we can confidently say that the fit is not an acceptable quality fit. This is despite it not being a terrible fit by eye.

Fri 07 Nov 2025 11:00

Lecture 12 - Introduction to Probability

What is probability? Probability is the pure mathematical description of randomness.

1 Set Theory

Say we want to group trees into four sets:

- Tall, or not.
- Variegated (has a lighter coloured leaf border) or not.

In a park of 142 trees, we observe:

Thu 02 Oct 2025 15:53

Lecture 1

Thu 02 Oct 2025 15:53

Lecture 2

Thu 02 Oct 2025 16:00

Lecture 3

Thu 02 Oct 2025 15:53

Lecture 4

Thu 02 Oct 2025 15:53

Lecture 5

Thu 02 Oct 2025 15:53

Lecture 6

Thu 02 Oct 2025 15:53

Lecture 7

Mon 13 Oct 2025 12:00

Lecture 8 - More Planes

1 Recap

Given the origin O , a point on the plane O' and a vector \vec{a} between them, we can take two vectors \vec{b} and \vec{c} from this point (which are not parallel). Using some combination of these two vectors, we can reach any point on the plane:

$$\vec{r}(s, t) = \vec{a} + s\vec{b} + t\vec{c}$$

This is the parametric equation of a plane, and is very robust. We can describe a flat plane in any dimensional space using this.

We can also define the scalar equation of a plane. Given these same two vectors, we can define a normal vector \vec{n} which is perpendicular to any vector that sits within the plane. We can construct this by using the cross product:

$$\vec{n} = \vec{b} \times \vec{c}$$

Given some generic point P :

$$\vec{OP} = \vec{a} + \vec{O'P}$$

And:

$$\underline{r}(s, t) = \underline{a} + s\underline{b} + t\underline{c}$$

We have:

$$(\underline{b} \times \underline{c}) \cdot \underline{r} = (\underline{b} \times \underline{c}) \cdot \underline{a} + s(\underline{b} \times \underline{c}) \cdot \underline{b} + t(\underline{b} \times \underline{c}) \cdot \underline{c}$$

Which (as a vector dotted with itself is 0) simplifies to (using $\underline{b} \times \underline{c} = \underline{n}$):

$$\underline{n} \cdot (\underline{r} - \underline{a}) = 0$$

Mon 20 Oct 2025 12:01

Lecture 9

Mon 20 Oct 2025 12:00

Lecture 10 - More Algebraic Methods

Goal:

$$\text{Show That: } \lim_{\theta \rightarrow \infty} \frac{\sin \theta}{\theta} = 1$$

As this result is needed to prove the derivative of $\sin \theta$

Wed 01 Oct 2025 11:00

Lecture 1 - Intro to Waves and SHM Recap

Course Objectives

- Have a sound understanding of basic wave properties
- Have a basic understanding of interference effects, inc diffraction
- Be able to use simple geometric optics and understand the fundamentals of optical instruments.

Recommended Textbooks

1. University Physics, Young and Freedman (Ch 15, 16 for Waves, Ch 33-36 for Optics)
2. Physics for Scientists and Engineers (Ch 20, 21 for Waves, Ch 22-24 for Optics)
3. 5e, Tipler and Mosca, (Ch 15, 16 for Waves, 31-33 for Optics)
4. Fundamentals of Optics, Jenkins and White
5. Optics, Hecht and Zajac

What is a wave? Waves occur when a system is disturbed from equilibrium and the disturbance can travel from one region to another region. Waves carry energy, but do not move mass. The course aim is to derive basic equations for describing waves, and learn their physical properties.

Periodic Motion

Waves are very linked to periodic motion. Therefore we recap periodic motion first.

It has these characteristics:

- A period, T (the time for one cycle)
- A frequency, f , the number of cycles per unit time ($f = \frac{1}{T}$)
- An amplitude, A , the maximum displacement from equilibrium.

Periodic motion continues due to the restoring force. When an object is displaced from equilibrium, the restoring force acts back towards the equi point. The object reaches equi with a non-zero speed, so the motion continues past the equi point and continues forever.

Energy

Periodic motion is an exchange between potential and kinetic energy, with no energy loss. Energy is conserved.

Simple Harmonic Motion

If the restoring force is directly proportional to the displacement $F = -kx$, then the periodic motion becomes Simple Harmonic Motion and the object is called a harmonic oscillator.

In a single dimension, displacement is given by:

$$x = A \cos(\omega t + \phi)$$

Where $\omega = 2\pi f$ is the angular velocity, and ϕ is the phase angle. In cases like this, where the phase angle is 90 deg we can simplify to $x = -A \sin(\omega t)$

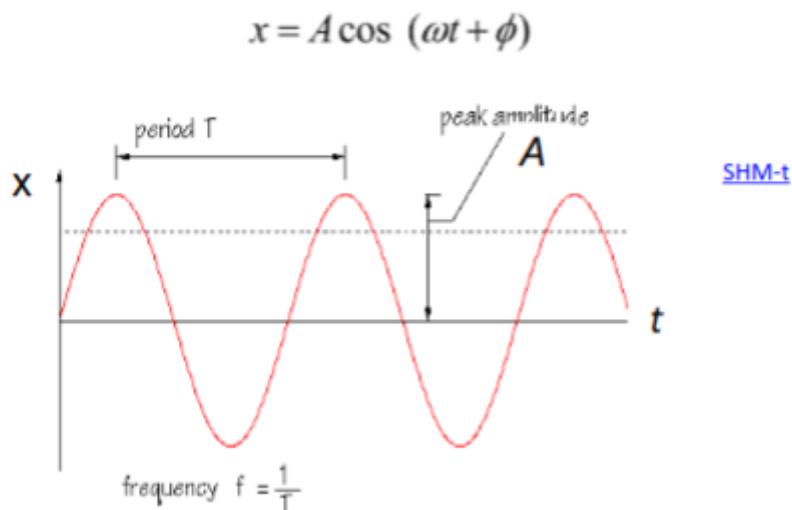


Figure 29.1: A Phase Angle of 90

More SHM Equations

Velocity

$$v_x = \frac{dx}{dt} = -\omega A \sin(\omega t + \phi)$$

Acceleration

$$a_x = \frac{dv_x}{dt} = \frac{d^2x}{dt^2} = -\omega^2 A \cos(\omega t + \phi)$$

Both properties are signed to indicate direction, as they are both vectors.

Thu 02 Oct 2025 13:00

Lecture 2 - Wave Functions

Sine Waves

Mechanical Waves A mechanical wave is a disturbance through a medium. It's formed of a single wave pulse or a periodic wave.

Mechanical Waves have the following properties:

1. Transverse: Where displacement of the medium is perpendicular to the direction of propagation.
2. Longitudinal, displacement of the medium is in the same direction as propagation.
3. Propagation depends on the medium the wave moves through (i.e. density, rigidity)
4. The medium does not travel with the wave.
5. Waves have a magnitude and a direction.
6. The disturbance travels with a known exact speed.
7. Waves transport energy but not matter throughout the medium.

Wave Functions

We want to define a wave function in terms of two variables, x and t . In any given moment, if we consider a single point on the wave (i.e. $t = 0$), and wait a short while, the wave will have travelled to some $t = t_1 > 0$.

In order to quantify displacement, we therefore want to specify both the time, and the displacement. This will let us find the wave speed, acceleration and the (new) wave number.

We are also able to talk about the velocity and acceleration of individual particles on the wave.

Wave Function for a Sine Wave

Consider a sine wave. We want to find a wave function in the form $y(x, t)$. Consider the particle at $x = 0$.

We can express the wave function at this point as $y(x = 0, t) = A \cos \omega t$. However we want to expand this to any general point. Now consider a point (2) which is one wavelength away. We know the behaviour of particle 1 is mirrored by particle 2 (with a time lag).

Since the string is initially at rest, it takes on period (T) for the propagation of the wave to reach point 2, therefore point 2 is lagging behind the motion of point 1. The wave equation is therefore (if particle two has $x = \lambda$) $y(x = \lambda, t) = A \cos(\omega t - 2\pi)$.

For arbitrary x , $y(x, t) = A \cos(\omega t - \frac{x}{\lambda} \cdot 2\pi)$ to account for this delay. This quantity is called the wave number:

$$\text{Wave Number: } k = \frac{2\pi}{\lambda}$$

So:

$$\begin{aligned} y(x, t) &= A \cos(\omega t - kx) \\ &= A \cos(kx - \omega t) \end{aligned}$$

Note the second step is possible as cos is an even function. k can also be signed to indicate direction: if $k > 0$, the wave travels in the positive x . If $k < 0$, the wave travels in the negative x direction. Again, $\omega = 2\pi f$

Displacement Stuff

Considering a point (starting at equi), the time taken for the particle on the sin wave to reach maximum displacement, minimum displacement and back takes the time period T . The speed of the wave is distance travelled over the time taken. We take the distance to be the wavelength λ , as we know the time by definition this takes is one time period T . Therefore wave speed v is:

$$v = \frac{\lambda}{T} = \lambda f$$

Since $\lambda = \frac{2\pi}{k}$ and $f = \frac{\omega}{2\pi}$ (as ω is defined as $\frac{2\pi}{T}$), we can also write:

$$v = \frac{2\pi}{k} \cdot \frac{\omega}{2\pi} = \frac{\omega}{k}$$

Particle Velocity

We can also determine the velocity of individual particles in the medium. We can use this to determine the acceleration.

We know that

$$y(x, t) = A \cos(kx - \omega t)$$

The vertical velocity v_y is therefore given by:

$$v_y = \frac{dy(x, t)}{dt}$$

Which is unhelpful (as we can't differentiate two variables at once), we can slightly cheat this by looking at purely a certain value of x , and therefore treating x as constant (to get a single variable derivative).

$$v_y = \left. \frac{dy(x, t)}{dt} \right|_{x=\text{const.}}$$

However this is notationally yucky, so we therefore use the notation:

$$\frac{\partial y(x, t)}{\partial t}$$

To represent the same thing. Finally (carrying out the partial derivative):

$$v_y(x, t) = \frac{\partial y(x, t)}{\partial t} = \omega A \sin(kx - \omega t)$$

Particle Acceleration

We can work out particle acceleration (transverse acceleration) by differentiating in the same manner again:

$$a_y(x, t) = \frac{\partial^2 y(x, t)}{\partial t^2} = \frac{\partial}{\partial t} \left(\frac{\partial y(x, t)}{\partial t} \right) = -\omega^2 A \cos(kx - \omega t) = -\omega^2 y(x, t).$$

Wed 08 Oct 2025 11:00

Lecture 3 - Generalised Wavefunctions

Recap: For a sine wave, the wavefunction is:

$$y(x, t) = A \cos(kx - \omega t)$$

Where k is the wave number, $k = \frac{2\pi}{\lambda}$ and omega is $\frac{2\pi}{T} = 2\pi f$

More Wavefunctions

What is the general form of a triangular shaped wave? What about a square stepped wave? TODO: Diagram
We know that each particle (i.e. along a string) copies the motion of its immediate left-hand neighbour (for a particle moving in the positive x), with a time delay proportional to their distance. Every wave is describable $y(x, t)$ wave function, and every mechanical wave relies on a medium (i.e. a piece of string or water, concrete etc) to travel.

Ideally, we want to be able to find a general form of a wave function.

In a single fixed instant of time, the moving wave pulse is stationary, so purely a function of $y = f(x)$. We want a wave function where we can input any value of t , so we need a moving frame of reference. We define this frame of reference as O' (for the origin) and x', y' axes. This frame of reference moves with the wave pulse and at the same speed, therefore y' is a function of x' only, independent of speed.

New System

x is the distance from the origin O to the relevant point, while x' is the distance from O' .

$$x = x' + vt$$

$$x' = x - vt$$

$$y' = f(x') = f(x - vt)$$

However, as the wave is moving purely in one direction (along x), $y = y'$, so:

$$y = f(x - vt)$$

Back to Basics

Going back to:

$$\begin{aligned} y(x, t) &= A \cos(kx - \omega t) \\ &= A \cos \left[k \left(x - \frac{\omega}{k} t \right) \right] \\ &= A \cos [k(x - vt)] \end{aligned}$$

(Note, this is true for a wave in the positive x , for a wave moving in the negative x this would be $x + vt$)

Equivalent Representations

There are some equivalent representations for a sine wave:

$$y(x, t) = A \cos\left(2\pi f \frac{x}{v} - \omega t\right) = A \cos\left(2\pi \left[\frac{x}{\lambda} - \frac{t}{T}\right]\right) = TODO$$

More Differentiation

We've already looked at differentiating with respect to t , but what about x ? This would give us the slope of the string at that point:

$$\frac{\partial y(x, t)}{\partial x}$$

And the curvature of the string:

$$\frac{\partial^2 y(x, t)}{\partial x^2} = -k^2 A \cos(kx - \omega t) = -k^2 y(x, t)$$

Note the similarities here with the equation for transverse acceleration:

$$(1) \frac{\partial^2 y(x, t)}{\partial t^2} = -\omega^2 y(x, t)$$

$$(2) \frac{\partial^2 y(x, t)}{\partial x^2} = -k^2 y(x, t)$$

Dividing (1) by (2):

$$\frac{\partial^2 y(x, t) / \partial t^2}{\partial^2 y(x, t) / \partial x^2} = \frac{\omega^2}{k^2} = v^2$$

Therefore:

$$\frac{\partial^2 y(x, t)}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y(x, t)}{\partial t^2}$$

We call this the 'wave equation' as every wave function $y(x, t)$ must satisfy it, regardless of whether or not it is periodic or its direction of travel. If $y(x, t)$ does not satisfy this, it is not a wave function.

An Example

$$y(x, t) = \frac{x^3 - vt^2}{e^t}$$

TODO, the example in our own time

Wave Equation for a String

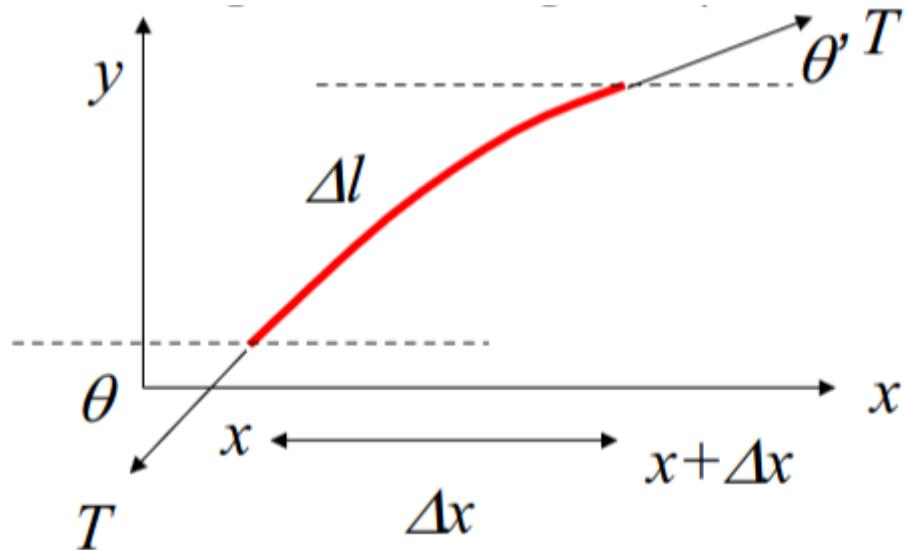


Figure 31.1: A snapshot of the wave.

Lets say we have some string, suspended horizontally under tension. We generate a single wave pulse and allow it to propagate down the string.

We assume the string is 1D, under tension T (constant throughout) and has mass per unit length μ .

Consider a small segment of string of length ΔL from x to $x + \Delta x$. This string makes angle θ with the horizontal at the bottom of the string, and angle θ' with the horizontal at the top of the string.

Net force (transverse in y) is:

$$F_y = T \sin \theta' - T \sin \theta$$

And using the small angle approx $\sin \theta \approx \tan \theta = \frac{dy}{dx}$:

$$F_y = T \left(\frac{dy}{dx} \Big|_{x+\Delta x} - \frac{dy}{dx} \Big|_x \right)$$

Using differentiation by first principles:

$$\frac{dy}{dx} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

We can say:

$$\frac{d^2y}{dx^2} = \frac{d}{dx} \left(\frac{dy}{dx} \right) = \lim_{\Delta x \rightarrow 0} \frac{\frac{dy}{dx} \Big|_{x+\Delta x} - \frac{dy}{dx} \Big|_x}{\Delta x}$$

Therefore:

$$F_y = T \frac{d^2y}{dx^2} \Delta x$$

The mass of this section of string is $\mu \Delta x$, and considering the acceleration in the y direction we can plug into $F = ma$ to get:

$$F = ma$$

$$T \frac{d^2y}{dx^2} \Delta x = \mu \Delta x \frac{d^2y}{dt^2}$$

$$\frac{d^2y}{dx^2} = \frac{\mu}{T} \frac{d^2y}{dt^2}$$

The LHS is the rate of change of the string's gradient, which as mentioned is the curvature of the string. The RHS includes the transverse acceleration of the string, therefore acceleration is proportional to curvature. To evaluate this at a fixed time/position we should write:

$$\frac{\partial^2 y}{\partial x^2} = \frac{\mu}{T} \frac{\partial^2 y}{\partial t^2}$$

And comparing to the wave equation we get:

$$\frac{1}{v^2} \frac{\partial^2 y}{\partial t^2} = \frac{\mu}{T} \frac{\partial^2 y}{\partial t^2}$$

$$\frac{1}{v^2} = \frac{\mu}{T}$$

$$v = \sqrt{\frac{T}{\mu}}$$

So a wave travels faster under a higher tension with a lower mass per unit length.

Thu 09 Oct 2025 13:00

Lecture 4 - Waves at Boundaries

Recap

We previously derived:

$$v = \sqrt{\frac{T}{\mu}}$$

And we also have:

$$v = \frac{\omega}{k}$$

The former is useful explicitly for a wave travelling over a string, while the latter is applicable to the movement of any wave. On a string, a higher tension yields a higher restoring force and therefore a higher speed, while a higher mass per unit length gives a higher mass for some arbitrary length of string, therefore a lower acceleration and lower speed.

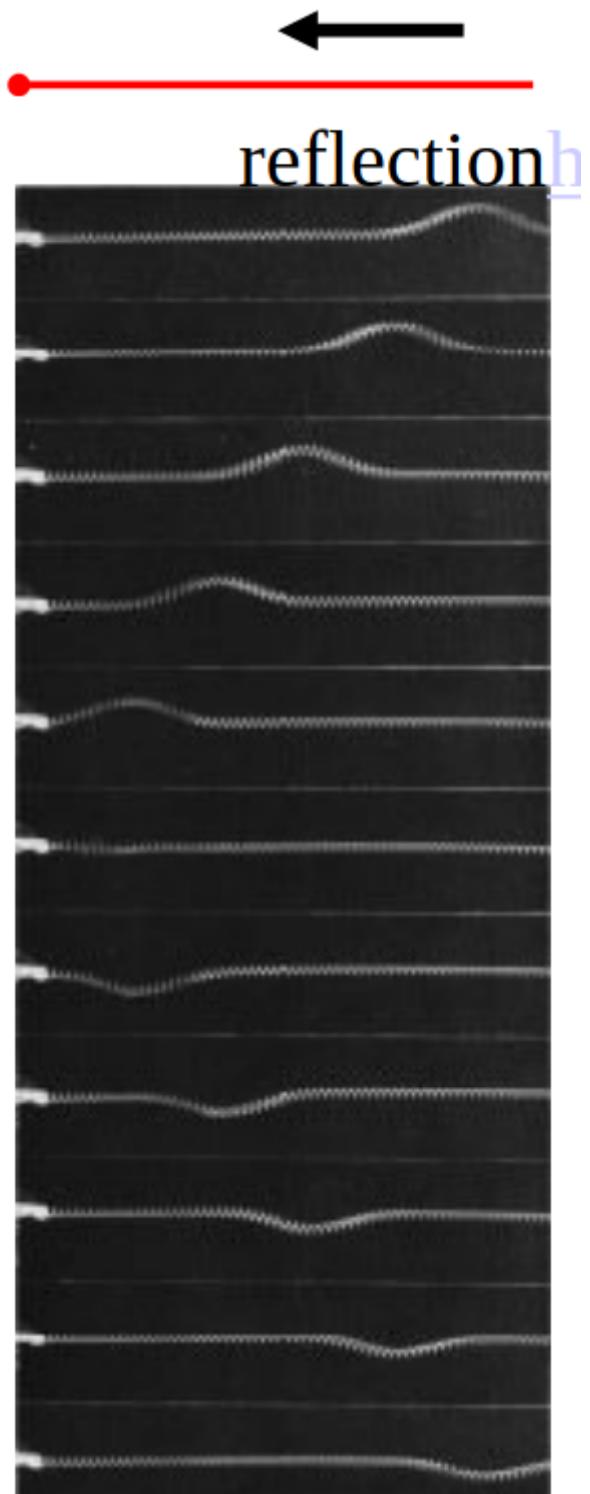
A Quick Interlude

The speed of a mechanical wave has the general form:

$$v = \sqrt{\frac{\text{Restoring force returning to equilibrium}}{\text{Inertia resisting return to equilibrium}}}$$

Reflection

When a wave hits a fixed boundary, it is reflected and inverted. Lets consider a case where a string is fixed on the LHS and is reflected back:



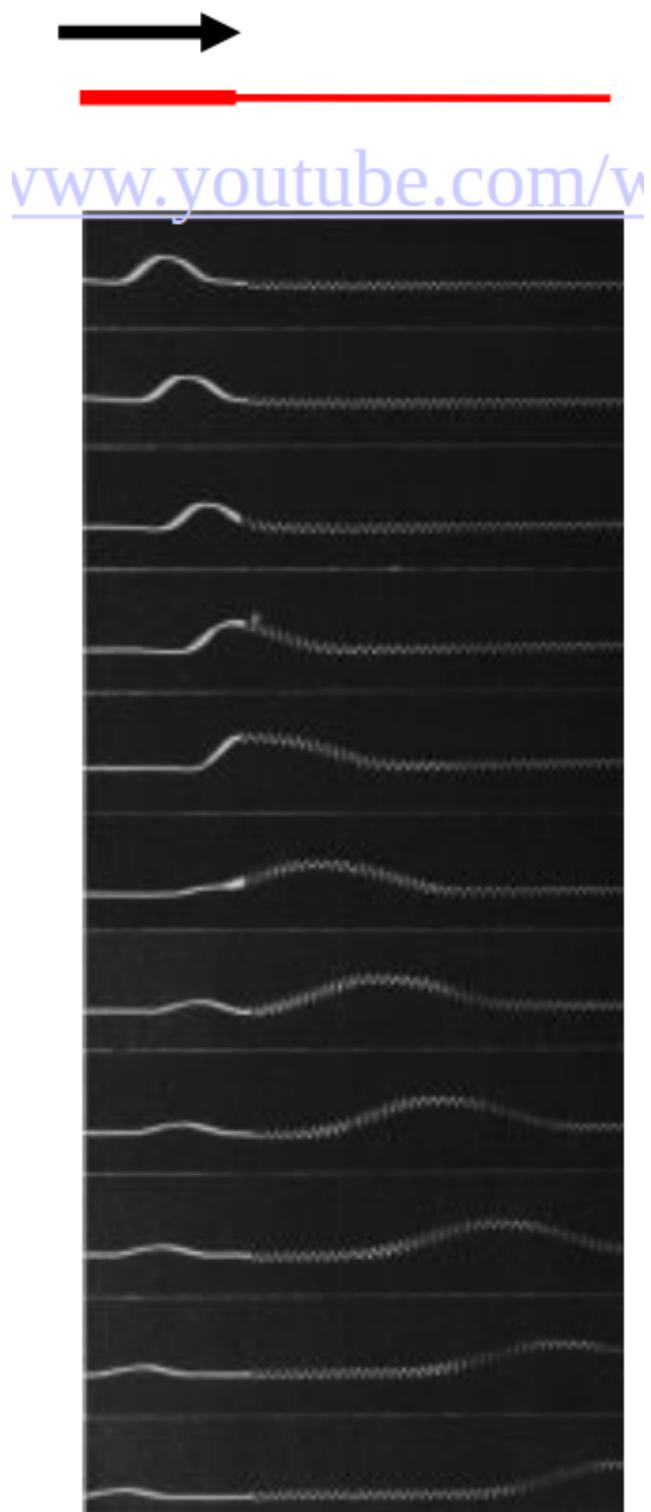


Figure 32.1: Or with two strings, going from thick to thin:

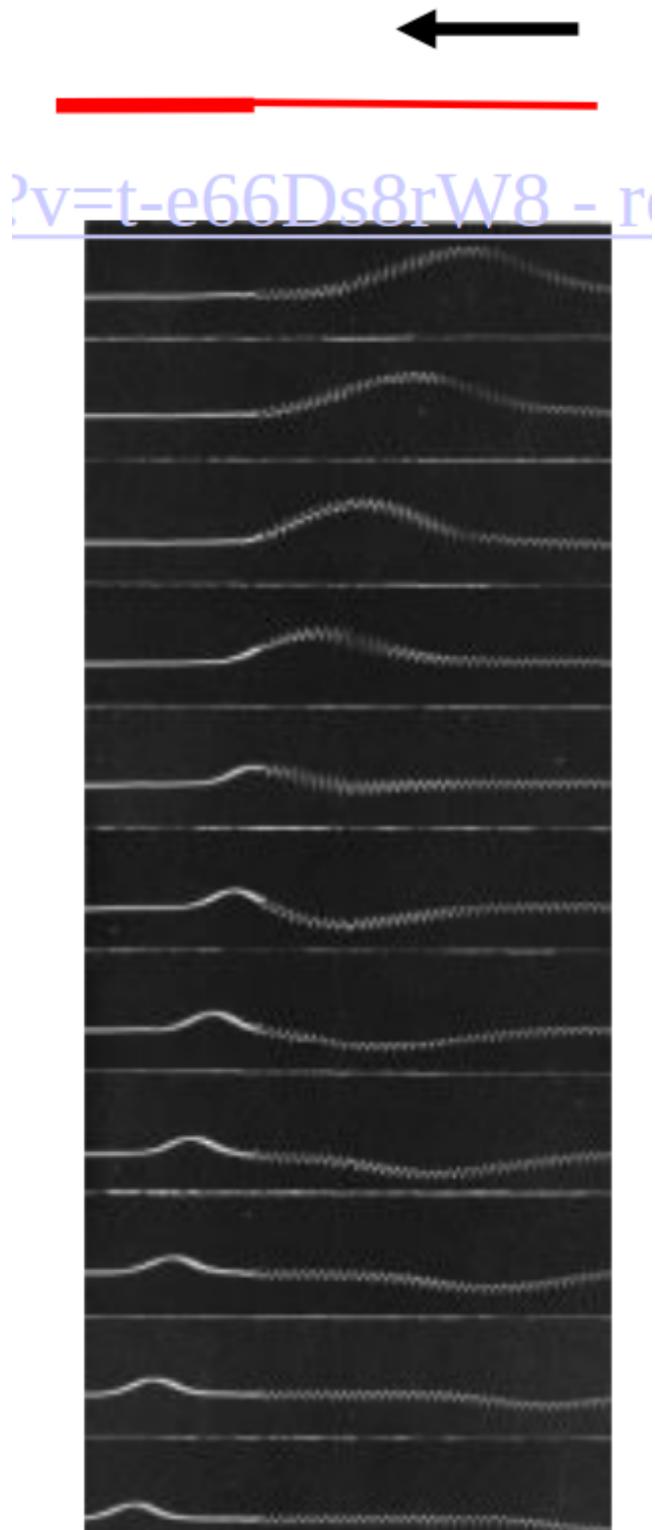


Figure 32.2: Or from thin to thick.

Waves Interacting at Boundaries

Say we have two pieces of string connected to each other, one thin string with mass per unit length μ_1 and a thicker string with μ_2 (both under the same tension, T). If a wave pulse is passed along from the thin string to the thicker string, what happens at the point of connection, P ?

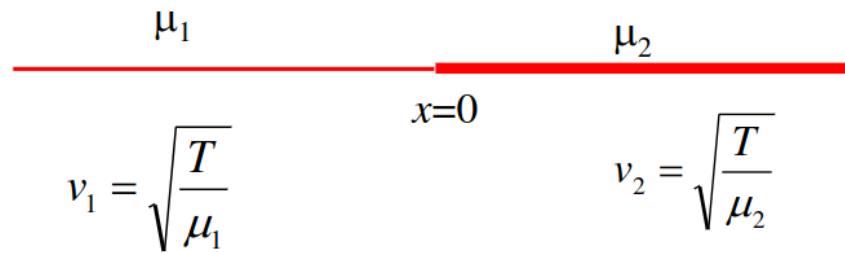


Figure 32.3: The two connected strings.

Consider a travelling wave coming from the left:

$$y_1 = A \cos(k_1 x - \omega_1 t)$$

$$y_2 = B \cos(k_2 x - \omega_2 t)$$

At the point of connection, we'll define this as $x = 0$. Here (as the string is not broken, so it must be connected):

$$y_1 = y_2 \quad (32.1)$$

Also, force is finite, therefore curvature must be finite (as force is proportional to curvature). Therefore we cannot have any discontinuities in curvature.

$$\frac{\partial y_1}{\partial x_1} = \frac{\partial y_2}{\partial x_2} \quad (32.2)$$

Disregarding non-linear effects, so assuming that the frequency with which the waves travel down the string is the same for both parts of the string, $\omega_1 = \omega_2 = \omega$.

From (1) and noting $x = 0$:

$$A \cos(-\omega t) = B \cos(\omega t) \quad (32.3)$$

And from (2) with the same note:

$$-k_1 A \sin(-\omega t) = -k_2 B \sin(-\omega t) \quad (32.4)$$

However this suggests that $A = B = 0$. This is technically a solution, but not really - it doesn't represent an actual wave (just two flat lines with no amplitude ever)

Including Reflection

The previous case did not work as we disregarded reflection at the wave boundary, P . Lets add an extra wave (the C term) to represent the reflection back into the light string:

$$y_1 = A \cos(k_1 x - \omega t) + C \cos(k_1 x + \omega t) = B \cos(k_2 x - \omega t)$$

From $y_1 = y_2$ at $x = 0$ we get:

$$y_1 = A \cos(-\omega t) + C \cos(-\omega t) = B \cos(-\omega t)$$

So: $A + C = B$

From (2) we get:

$$-k_1 A \sin(-\omega t) - k_1 C \sin(\omega t) = -k_2 B \sin(-\omega t)$$

Which has solutions:

$$B = \frac{2k_1}{k_1 + k_2} A$$

$$C = \frac{k_1 - k_2}{k_1 + k_2} A$$

We have the incident wave:

$$y_1 = A \cos(k_1 x - \omega t)$$

The transmitted wave:

$$y_2 = B \cos(k_2 x - \omega t) = \frac{2k_1}{k_1 + k_2} A \cos(k_2 x - \omega t)$$

And lastly the reflected wave:

$$y_3 = C \cos(k_1 x + \omega t) = \frac{k_1 - k_2}{k_1 + k_2} A \cos(k_1 x - \omega t)$$

We note:

k_1 is the wave number in the medium where the incident wave comes from.

k_2 is the wave number in the medium where the transmitted wave goes into.

Example

If the wave comes in from the left, then $\mu_2 > \mu_1$ per example diagram, then $k_2 > k_1$, and C is negative.

Note that if the single pulse has a positive y amplitude, then the transmitted pulse in the heavier string will also have a positive y amplitude, but will have a smaller magnitude. The reflected wave will have a negative y -amplitude as the reflection inverts it.

If the wave comes from the right (from thick to thin), then $\mu_1 > \mu_2$, $k_1 > k_2$ and C is positive.

Fri 17 Oct 2025 12:56

Lecture 5

Thu 16 Oct 2025 13:00

Lecture 6 - Standing Waves 2 Electric Boogaloo

Recap

$$\frac{\partial y}{\partial x} = 2Ak \cos(kx) \sin(\omega t)$$

$$\frac{\partial^2 y}{\partial x^2} = 2A(-k^2) \sin(kx) \sin(\omega t)$$

$$\frac{\partial y}{\partial t} = 2A\omega \sin(kx) \cos(\omega t)$$

$$\frac{\partial^2 y}{\partial t^2} = 2A(-\omega^2) \sin(kx) \sin(\omega t)$$

Therefore:

$$\frac{\frac{\partial^2 y}{\partial t^2}}{\frac{\partial^2 y}{\partial x^2}} = \frac{2A(-\omega^2) \sin(kx) \sin(\omega t)}{2A(-k^2) \sin(kx) \sin(\omega t)} = \frac{\omega^2}{k^2} = v^2$$

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y}{\partial t^2}$$

Therefore a standing wave still obeys the wave equation, as it must.

Standing Wave Properties

Wavelength

Consider a horizontal string from $x = 0$ to $x = L$, with both ends fixed. We generate a sinusoidal wave pulse, which must satisfy:

$$y(x, t) = 2A \sin(kx) \sin(\omega t)$$

We know at $x = L$ and $x = 0$, $y = 0$ at all times as fixed at this point. Therefore:

$$kL = n\pi, (n \in \mathbb{N})$$

For $n = 1$, we have half a wavelength on the string:

$$\lambda = \frac{2L}{1} = 2L$$

And this has a general form: $\lambda = \frac{2L}{n}$.

Frequency

$$f_n = \frac{v}{\lambda_n} = \frac{v}{\left(\frac{2L}{n}\right)} = \frac{nv}{2L}$$

Crucially:

$$f_1 = \frac{v}{2L}$$

Is the first harmonic (or fundamental). $f_2 = 2f_1$ is the second harmonic, or first overtone, etc. All of these, f_n where $n \in \mathbb{N}$ are called “normal modes”. For each normal mode, the corresponding frequency is called the resonant frequency (natural frequency of the system).

What happens if we try to create a standing wave where $\lambda \neq \frac{2L}{n}$? In short, we cannot. The system will reject any attempts to do so.

Energy

Energy is proportional to ω^2 . Energy can only take certain discrete values (corresponding to f_1, f_2, \dots, f_n), we find that the system has quantised possible values for energy.

To generate a wave with a higher frequency we either have to use a shorter L, or a higher v. A higher v is achieved by using a lighter string or placing the system under higher tension.

Sound Waves

Notation

Displacement of a sound wave is denoted:

$$s(x, t) = S_m \cos(kx - \omega t)$$

And pressure is given by:

$$\Delta P(x, t) = \Delta P_m \sin(kx - \omega t)$$

Different Boundary Conditions

The equations for standing waves given is only true for the boundary conditions of both ends fixed. If we vary these (for example left end fixed, right end not, wave initially travelling left) we get a different solution. For example, the first harmonic:

$$L = \frac{1}{4\lambda_1}$$

$$f_1 = \frac{v}{4L}$$

Where the left end forms a node (as required by boundary conditions) and the right end forms an antinode, as it is free to move. For the third harmonic:

$$L = \frac{3}{4\lambda_3}$$

$$f_3 = \frac{3v}{4L} = 3f_1$$

And fifth:

$$L = \frac{5}{4\lambda_5}$$

$$f_5 = \frac{5v}{4L} = 5f_1$$

Notably, this system cannot support even harmonics.

Thu 04 Sep 2025 01:00

Lecture 7 - Energy and Power of Waves

Kinetic Energy for a Sine Wave

Assuming a sine wave created by a harmonic oscillator.

$$\begin{aligned} KE &= \frac{1}{2}mv^2 \\ (KE)_{\max} &= \frac{1}{2}mv_{\max}^2 \\ (KE)_{\max} &= \frac{1}{2}(\mu dx)(\omega A)^2 \end{aligned}$$

And then to get the total wave energy, we can integrate over lambda: TODO

And Power

$$\begin{aligned} \text{Power} &= \frac{\text{Energy}}{\text{Time}} \\ P &= \frac{\frac{1}{2}(\omega A)^2 \mu \lambda}{T} \\ &= \frac{1}{2}(\omega A)^2 \mu v \end{aligned}$$

Example

Considering a standing wave with $\lambda = 2L$:

$$\begin{aligned} y_t &= y_1 + y_2 \\ \implies (KE)_t &= (KE)_1 + (KE)_2 \end{aligned}$$

Where y_1 and y_2 are identical except for their direction, but they carry the same kinetic energy. The KE of a standing wave is the sum of the two waves that make it up.

Note: Normally, A is the amplitude of the travelling wave (hence $2A$ is the amplitude of the standing wave created by them), however it is sometimes ambiguous what is being referred to by A.

Interference

Superposition: When waves overlap in the same region, the resulting wave is the algebraic sum of waves (they interfere)

Consider two waves:

Thu 23 Oct 2025 13:00

Lecture 8 - EM Standing Waves and Lasers

Lasers

A traditional laser has two mirrors on either side of a cavity. One is (ideally) 100% reflecting, while another is almost perfectly reflecting (say 99% reflective), but allows some transmission. This causes EM waves to reflect back and forth (with some transmitted to actually cause the visible laser).

These reflecting waves cause a standing wave inside the laser cavity. This standing wave oscillates with the equation given in Lec 07.

A Quick Audio Interlude

Listening to various sounds from a frequency generator, we notice two things:

- The human ear is very sensitive to changes in frequency, even in the single Hz range.
- A square wave has a higher perceived pitch than a sine wave of the same frequency.
 - This is because a square wave can be decomposed into the sum of many sine waves, and some of these sine wave components have a higher frequency compared to the square wave itself, which our ears can detect.

And Back to Superposition

Given two travelling waves:

$$y_1 = A \cos(k_1 x - \omega_1 t)$$

$$y_2 = A \cos(k_2 x - \omega_2 t)$$

$$y = y_1 + y_2$$

Hence:

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

And using a double angle formula:

$$y = 2A \cos\left(\frac{k_1 - k_2}{2}x - \frac{\omega_1 - \omega_2}{2}t\right) \cos\left(\frac{k_1 + k_2}{2}x - \frac{\omega_1 + \omega_2}{2}t\right)$$

Or briefly:

$$y = 2A \cos(\Delta k x - \Delta \omega t) \cos(k_{avg} x - \omega_{avg} t)$$

Noting that we define the delta portions as including the /2.

We define the blue portion as the group and the red portion as the carrier. These are important and will be covered properly later.

We define the phase velocity as:

$$v_{phase} = \frac{\omega_{avg}}{k_{avg}}$$

And the group velocity as:

$$v_{group} = \frac{\Delta \omega}{\Delta k}$$

Carriers and Groups

What actually is a carrier and a group? Lets take radio as an example. Radio waves travel for potentially hundreds of kilometers, and it's not possible to send sound waves remotely close to that distance because of high attenuation.

The idea therefore arose of using a radio frequency wave to carry sound. Lets consider an example EM RF wave at 500kHz . We then approximate our sound wave as a sine wave of pressure variations, say at 1000Hz .

The first thing we can do is to modulate the amplitude of the RF wave by the audio wave. The RF wave keeps the same frequency, but changes amplitude depending on the amplitude of the audio wave.

Upon receipt of the wave, we filter out the high frequency oscillations of the radio wave itself, but we extract a lower frequency wave defined by a wave that passes through the maximum points of oscillation. This effectively restores the sound wave.

This is known as AM radio. The carrier is the high frequency wave, and the group is the lower frequency constructed wave.

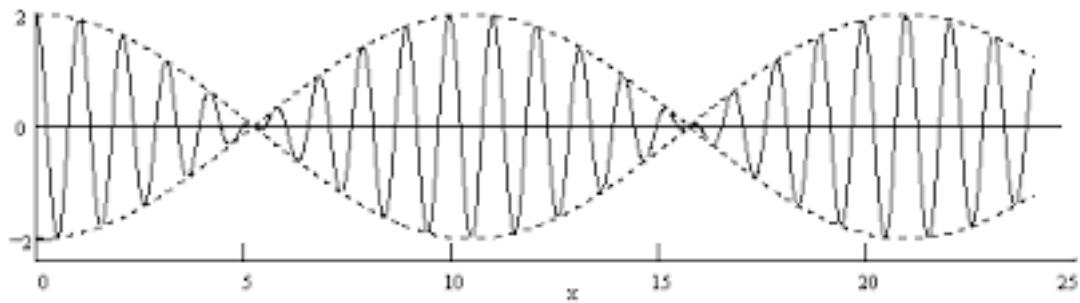


Figure 36.1: Carrier and Group waves. The lower freq. dotted line is the extracted wave (group), the higher frequency wave is the carrier.

Velocities

Which of the two waves have a higher speed?

[TODO]

In a medium where the wave speed depends on frequency, we call this medium dispersive.

Air is non-dispersive (i.e. sound), and in a vacuum the same applies (i.e. speed of light).

What about the speed of light in glass? Does this depend on frequency? We'll investigate later in the optics section.

A Few More Definitions

Coherence: Two (or more) sources that are in phase, or have a constant phase difference (Δ) are said to be coherent.

Adding Coherent Waves

Constant phase difference means we can nicely add amplitudes, so total intensity:

$$I = |A|^2 = (A_1 + A_2 + \dots, A_n)$$

Adding Incoherent Waves

TODO

The Decibel Scale

$$\beta = 10 \log \left(\frac{I}{I_0} \right)$$

Where I is the intensity of the source, and I_0 is a reference level approximately at threshold of hearing equal to $10^{-12} W/m^2$

At the threshold of hearing:

$$\beta = 10 \log \left(\frac{10^{-12}}{10^{-12}} \right) = 10 \log(1) = 0dB$$

Crucially, 0dB is not zero sound intensity.

At the upper pain threshold:

$$\beta = 10 \log \left(\frac{1}{10^{-12}} \right) = 120dB$$

But, how do we calculate I here?

$$I = \frac{\text{Power}}{\text{Area}}$$

Power is related to energy, so:

$$I = \frac{\text{energy}}{\text{time} \times \text{area}}$$

And to get intensity for an actual volume:

$$I = \frac{\text{energy} \times \text{length}}{\text{time} \times \text{volume}}$$

$$I = \frac{\text{energy}}{\text{volume}} \times \text{wavespeed}$$

What displacement on the eardrum do we actually have? For 120dB, $10^{-5} m$, or 10^{-11} for 0dB. The latter measurement is smaller than an atom. The ear is a very sensitive instrument.

Human Senses

Interestingly, we see, hear and smell all on a log scale. Physical feeling is not very scientific and can't easily be quantified. Emotional feeling is also not quantifiable...

Thu 30 Oct 2025 13:00

Lecture 9 - Optics Part Two

Key Principles

There are two main principles that we'll use:

1. Hugen's Principle

- Each point on a wavefront serves as a source of spherical secondary *wavelets* that advance with speed and frequency identical to the primary wave.
- If we consider each point on a wave as a spherical wavelet source, the wavefront is given by a line tangent to all of these.

2. Fermat's Principle

- The actual path taken by a beam of light is the one which takes the least time to traverse.
- i.e. $dt/dl = 0$
- Light could theoretically take an infinite number of possible routes between any two points, but in practice (in the same material) it will travel in a straight line as this is the fastest route.
- When light is travelling through an interface (i.e. an air to glass boundary), the fastest point is no longer a straight line, as refraction appears.

Reflection

Lets consider some mirror as a reflecting surface, with light travelling from point A to point B. There are multiple different theoretical paths that the light could take, so Fermat's principle requires the shortest possible time taken.

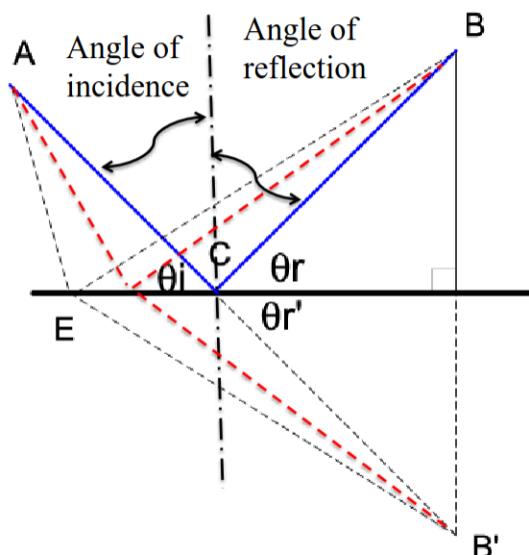


Figure 37.1

We require, by Fermat's Principle (as same material, so shortest time is simply shortest path):

$$AC + CB = \min$$

By considering congruent triangles (let O be the midpoint of B and B'), therefore:

$$AC + CB = AC + CB'$$

Provided ACB' is a straight line, $\theta_i = \theta_r (= \theta_{r'})$. Therefore the angles of incidence and reflection must be equal.

Refraction

When light is travelling through a material, the speed of light in each is different. Therefore, a straight line is no longer the fastest path.

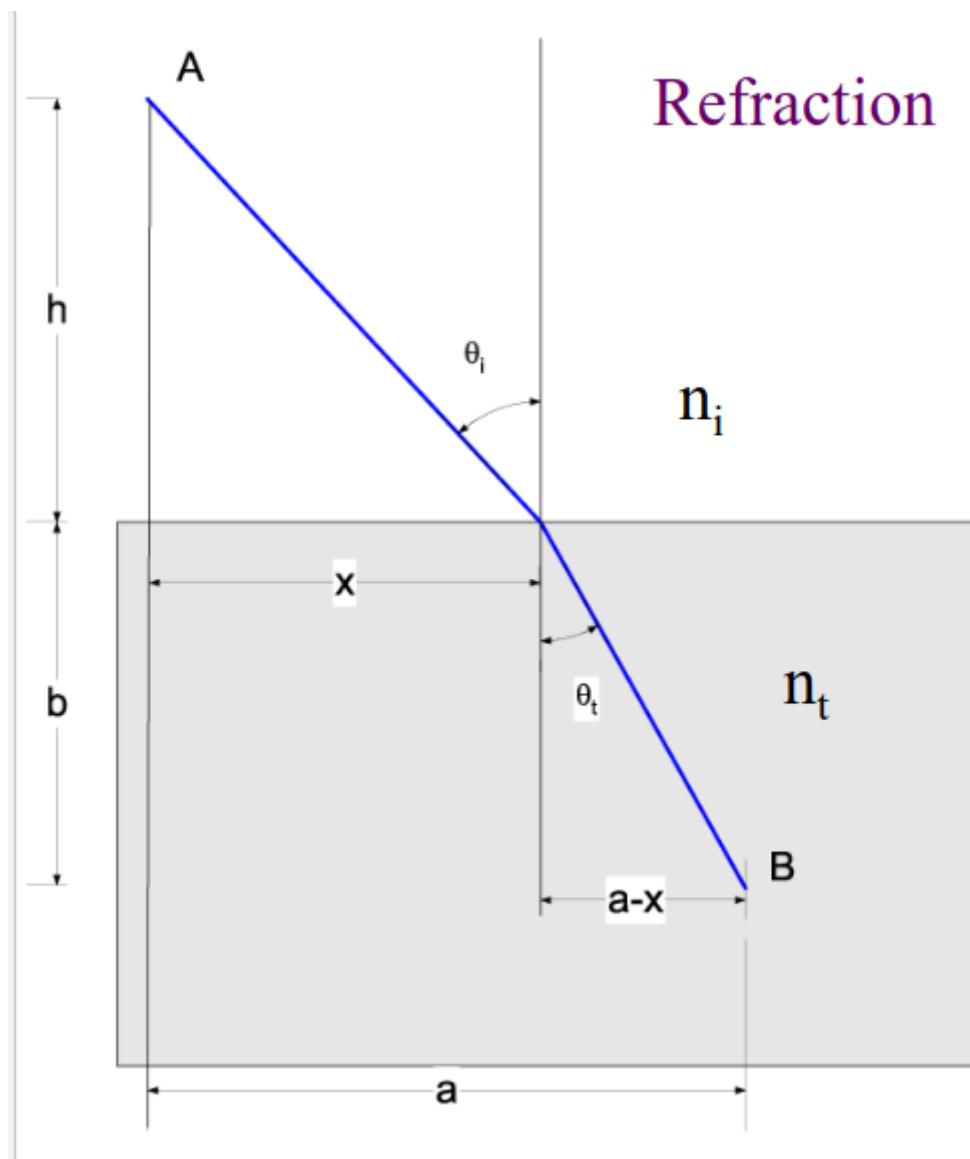


Figure 37.2

Where:

- n_i is the incident index.
- n_t is the post-refraction index.

- θ_i is the angle of incidence.
- θ_t is the angle of refraction.
- Light is travelling from point A to point B.
- Distances are given per diagram.

The aim is to be able to find the true trajectory (per Fermat) that takes the shortest amount of time. By $s = d/t$, for given distance values:

$$t = \frac{\sqrt{h^2 + x^2}}{v_i} + \frac{\sqrt{b^2 + (a-x)^2}}{v_t}$$

Where v_i, v_t are speeds of light in the respective media. t is a function of x , with constants, so to find a min value we differentiate wrt x and solve for minima:

$$\frac{dt}{dx} = \frac{x}{v_i \sqrt{h^2 + x^2}} + \frac{-(a-x)}{v_t \sqrt{b^2 + (a-x)^2}} = 0$$

Subsituting values for distances, we get:

$$\frac{\sin \theta_i}{v_i} = \frac{\sin \theta_t}{v_t}$$

And since $n = c/v$, we get:

$$n_i \sin \theta_i = n_t \sin \theta_t \quad (37.1)$$

The shortest (wrt time) path must have angles which satisfies this. This is called Snell's Law.

Connecting Fermat and Huygen

Speed of light in a medium is less than in a vacuum, this is categorised by the index of refraction, n :

$$n = \frac{c}{v}$$

Where c is speed of light in a vacuum, and v is speed of light in the medium. i.e. for water, $n = 1.333$, for air, $n = 1.0003$.

Suppose we have a wavefront travelling as a series of wavelets. As soon as each wavelet hits the glass, they slow down. This leads to an uneven distribution of speeds across the wavelength, as some have slowed down and some have not. Therefore, the wave bends around and we get a new wavefront at a different angle to the previous wavefront.

Note the condition where the light travels perpendicular to the interface, i.e. $\theta_i = \theta_t = 0$. In this case, all wavelets hit the medium and change speed at exactly the same time, therefore there is no difference in velocity across the wavefront, so no direction change.

But how?

How does light actually "choose" which path to take? Apparently it just does and we have to wait to find out...

Combining Reflection and Refraction

When light strikes a boundary surface, there is two components - both reflected and transmitted (just like on a string). To determine this, we have to use the wave equation:

$$E = E_0 \cos(kx - \omega t)$$

Noting that instead of traditional mechanical amplitude we use the amplitude of the electromagnetic field at this point, E_0 . For now, we ignore polarisation as a potential scenario.

More Index

We note that since $v = f\lambda$, the index of refraction will change depending on the frequency of light:

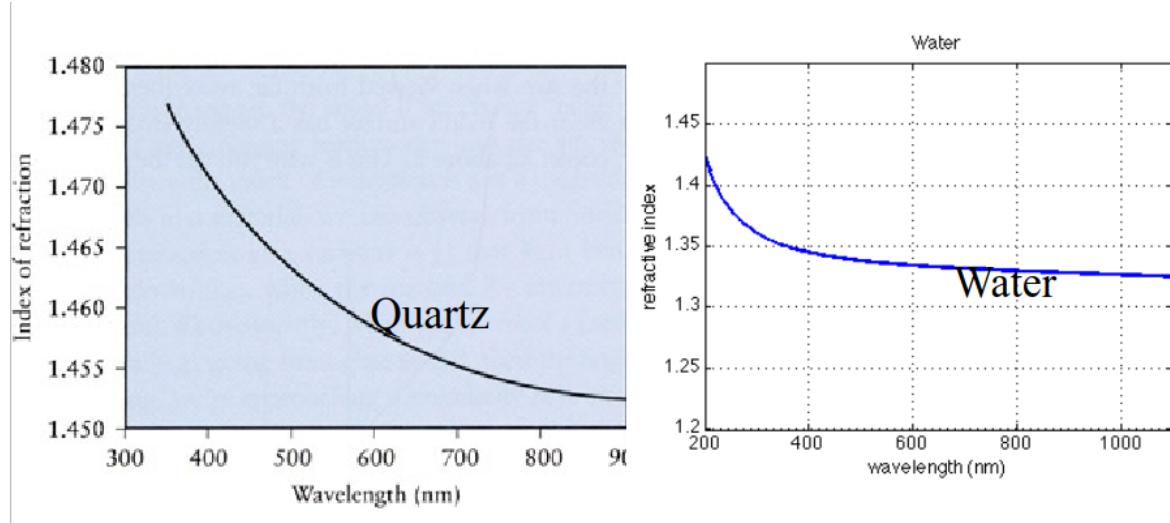


Figure 37.3

When we shine light of multiple frequencies (i.e. white light), the higher frequency light bends further than the lower frequency light. This causes dispersion of light, i.e. the formation of a rainbow out of a prism. This arises because each frequency has a different speed and therefore a different change in θ .

If we shine pulses of light of multiple frequencies down a fibre optic line, the higher frequencies will have a higher n , hence a lower speed and will arrive later.

This allows us to explain the formation of sunsets and rainbows etc. In the case of a sunset, when we see the sun just above the horizon, the sun has actually set just below the horizon. We cannot therefore see the sun via direct line of sight, and yet we can still see it as if we could?

This is because the light from the sun is refracted and bends towards us. The atmosphere is higher density at the bottom, and lower density as altitude increases. The amount of refraction depends on density, hence changing the index of refraction. We can say that n is a function of y , where y is height.

We imagine it as being actually present precisely where we see it, because our brain does not account for this refraction and extrapolates the light as a straight line. We therefore see the sun has just about to set, when in reality the sun physically has set below the horizon.

Fri 03 Oct 2025 12:00

Lecture 1 - Atomic Structure

1 Atomic Structure

1.1 What is the course?

- Quantum mech is weird and unintuitive, we will build up a case in the course for why this weird theory was necessary and why we're confident it works.
- Each week will be a self-contained concept and/or historical experiment, working up to the Shroedinger Equation and wave-particle duality.
- Names and dates do not need to be memorised.
- Recommended text: University Physics (Young and Freedman).
- Office hours: 13:00 – 13:50 Fridays (immediately post-lecture), Physics East Rm 207.

1.2 Atomic Structure

What actually is an atom? What does it actually look like inside?

Early Clues

- Periodic Table (Mendelev, 1869), periodic patterns in elements properties.
- Radioactivity (Becquerel, 1896, Curie 1898)
- Atoms emit and absorb specific discrete wavelengths, (Balmer, 1884)
- Discovery of the Electron (Thompson 1897). Cathode rays - heating metal in a vacuum with an electric field above it, to strip away electrons from the metal.
 - This showed electrons were negatively charged and extremely light (1/2000th of the atomic mass).

Atoms emit/absorb light at discrete wavelengths

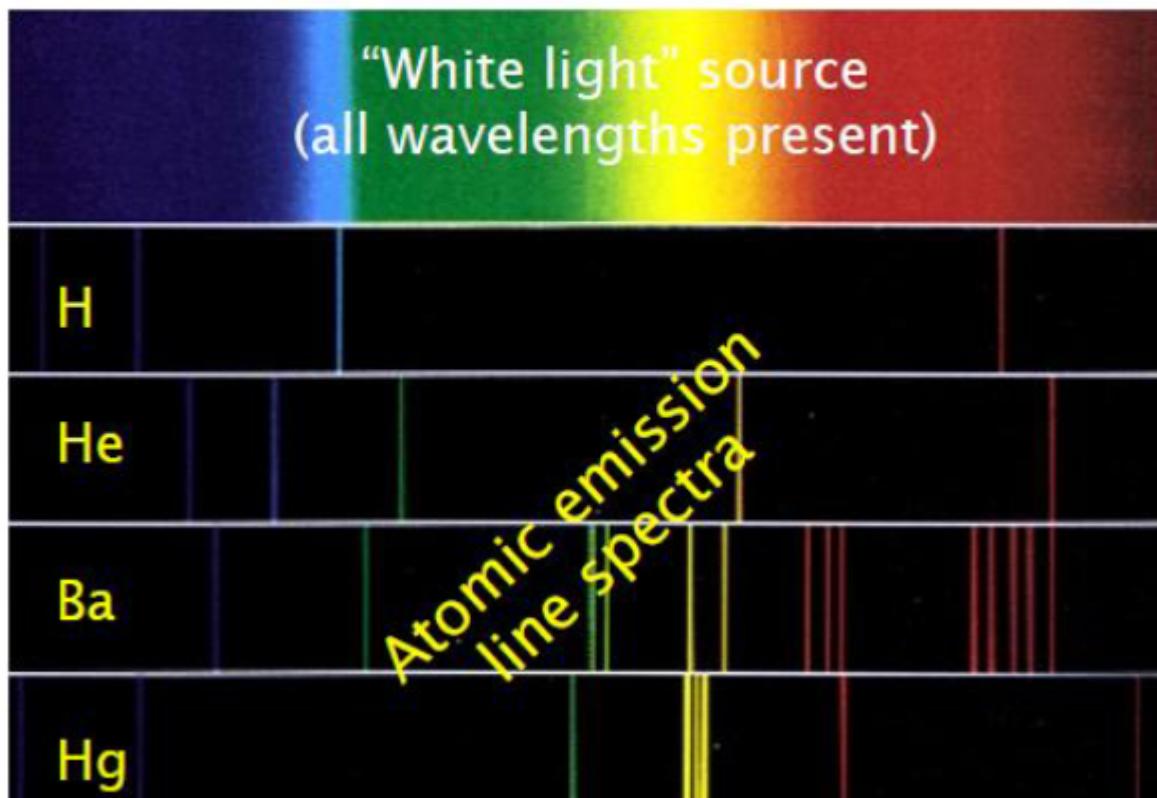


Figure 38.1: Absorption Spectra

1.3 Plum Pudding Model

A solid, uniform lump of positively charged matter, approximately 10^{-10}m across. This had evenly distributed negative charges (electrons) scattered throughout.

1.4 Discovery of the Nucleus

Geiger and Marsden (1908-1913), fired alpha particles (He nuclei) at thin gold foil and measured the deflection / scattering.

The alpha particles had a mass of $4u$, a charge of $+2e$ and an energy of approximately 5MeV .

They found that most α were scattered only by small angles, but (surprisingly) a small number were scattered right back towards to emitter (through $\theta > 90\text{ deg}$). The distribution of the angles is approximately Normally distributed, with a mean of 0. Only approximately 1 in 8,000 fired α s were scattered by $\theta > 90$ ("back-scattering").

Can this be explained with the Plum Pudding Model? No, it cannot. This was used to demonstrate that atoms cannot be evenly distributed.

2 Demonstrating by Calculation

Lets work out the work done to take an α from infinity to the pudding centre. If the electrostatic repulsion is not enough to overcome this, we cannot stop the α and cannot back scatter.

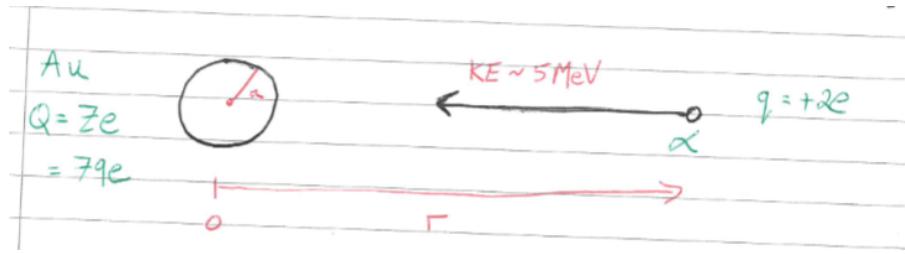


Figure 38.2: The experiment

2.1 Assumptions

- The atom stays still.
- Ignore the gold electrons (this is fine, as they would cancel some positive charge and make repulsion weaker, which would be even worse. If we can't do it without them, it would be equally impossible to do it with.)

2.2 Recap and Eqns

Coulomb Potential Energy is:

$$u(r) = \frac{qQ}{4\pi\epsilon_0 r}$$

Force is:

$$F(r) = -\frac{du}{dr} = \frac{qQ}{4\pi\epsilon_0 r^2}$$

Change in potential energy ($u_2 - u_1$) is work done:

$$\int_{u_1}^{u_2} du = - \int_{r_1}^{r_2} F(r) dr$$

From outside the atomic radius, we treat the atomic pudding as a point charge of charge Q. From inside the atomic radius, we treat it as a smaller point charge $Q'(r)$, where we only consider the charge inside the portion of the pudding where $a < r$, where a is the current position inside the sphere and r is the atomic radius.

If charge is spread uniformly, the total charge is proportional to the volume of the sphere. So:

$$\frac{Q'}{Q} = \frac{\frac{4}{3}\pi r^3}{\frac{4}{3}\pi a^3}$$

$$Q' = Q \frac{r^3}{a^3}$$

Inside the Pudding

$$F = \frac{qQ'}{4\pi\epsilon_0 r^2}$$

$$F = \frac{qQr^3}{4\pi\epsilon_0 r^2 a^3}$$

$$F = \frac{qQr}{4\pi\epsilon_0 a^3}$$

$$F = \frac{qQ}{4\pi\epsilon_0 a^3} \times r$$

Hence inside, $F \propto r$

Outside the Pudding

$$F = \frac{Qq}{4\pi\epsilon_0 r^2}$$

$$F = \frac{Qq}{4\pi\epsilon_0} \times \frac{1}{r^2}$$

Hence outside, $F \propto \frac{1}{r^2}$

To integrate, we are therefore integrating the area under this (almost) triangle:

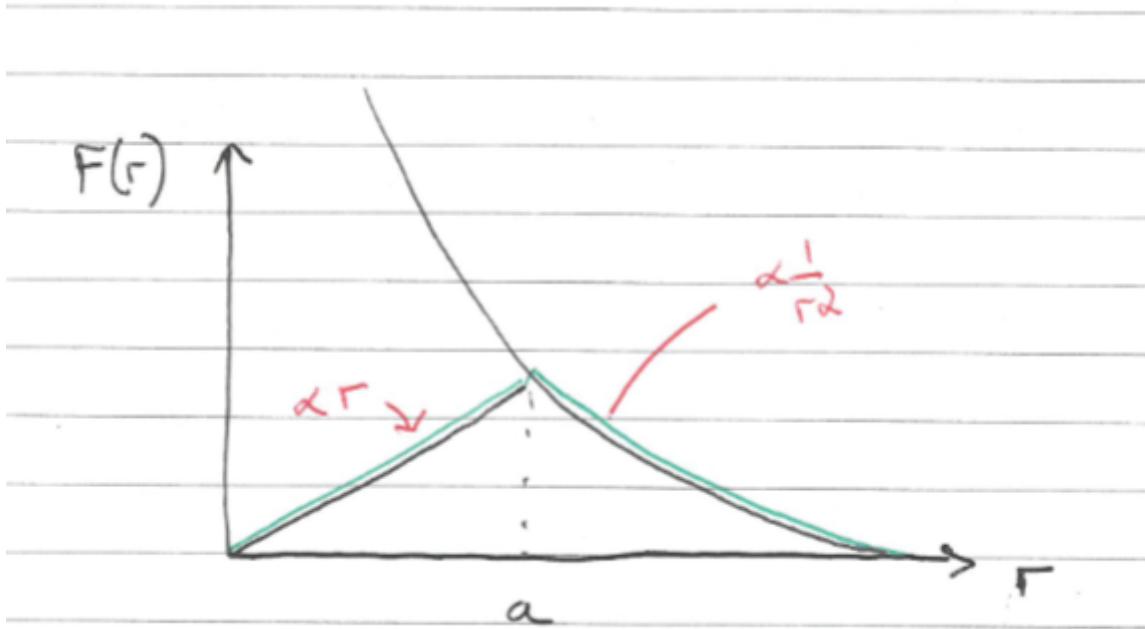


Figure 38.3: Radius vs electrostatic repulsion force

$$\Delta u = - \int_{r_1}^{r_2} F(r) dx$$

Splitting into two sections (prior to, and after the point where $r = \text{atomic radius, } r = a$), and integrating across all r (as we are attempting to work out the work done to bring an α from infinity to the charge, at which point distance is 0):

$$\begin{aligned} &= - \int_{\infty}^a \frac{qQ}{4\pi\epsilon_0 r^2} dx - \int_a^0 \frac{qQr}{4\pi\epsilon_0 a^3} dx \\ &= - \frac{qQ}{4\pi\epsilon_0} \int_{\infty}^a \frac{1}{r^2} dx - \frac{qQ}{4\pi\epsilon_0 a^3} \int_a^0 r dx \\ &= - \frac{qQ}{4\pi\epsilon_0} \lim_{x \rightarrow \infty} \int_x^a \frac{1}{r^2} dx - \frac{qQ}{4\pi\epsilon_0 a^3} \int_a^0 r dx \\ &= - \frac{qQ}{4\pi\epsilon_0} \left[-\frac{1}{r} \right]_x^a - \frac{qQ}{4\pi\epsilon_0 a^3} \left[\frac{1}{2} r^2 \right]_a^0 \\ &= - \frac{qQ}{4\pi\epsilon_0} \left(-\frac{1}{a} - 0 \right) - \frac{qQ}{4\pi\epsilon_0 a^3} \left(\frac{1}{2} 0^2 - \frac{1}{2} a^2 \right) \\ &= - \frac{qQ}{4\pi\epsilon_0} \left(-\frac{1}{a} \right) - \frac{qQ}{4\pi\epsilon_0 a^3} \left(-\frac{1}{2} a^2 \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{qQ}{4\pi\epsilon_0 a} + \frac{qQa^2}{8\pi\epsilon_0 a^3} \\
&= \frac{qQ}{4\pi\epsilon_0 a} + \frac{qQ}{8\pi\epsilon_0 a} \\
&= \frac{qQ}{4\pi\epsilon_0 a} + \frac{1}{2} \frac{qQ}{4\pi\epsilon_0 a} \\
&= \frac{3}{2} \frac{qQ}{4\pi\epsilon_0 a}
\end{aligned}$$

As required! Plugging in values gives us:

$$\begin{aligned}
\Delta u &= \frac{3}{2} \frac{(2e)(79e)}{4\pi(8.854 \times 10^{-12}) \times 10^{-10}} \\
&= 5.45 \times 10^{-16} J = 3.41 KeV
\end{aligned}$$

This is much less than the kinetic energy of the 5MeV alpha particle, therefore (as this value is maximum work done against the repulsive force) a plum pudding could not backscatter a 5MeV alpha particle. However, since $\Delta u \propto 1/a$, a smaller volume of charge could. How small, however?

$$\begin{aligned}
\Delta u = 5\text{MeV} &= - \int_{\infty}^{r_{max}} \frac{qQ}{4\pi\epsilon_0 r^2} dr \\
-5\text{MeV} &= \frac{qQ}{4\pi\epsilon_0} \int_{\infty}^{r_{max}} \frac{1}{r^2} dr \\
-5\text{MeV} &= \frac{qQ}{4\pi\epsilon_0} \lim_{x \rightarrow \infty} \left[-\frac{1}{r} \right]_{x}^{r_{max}} \\
-5\text{MeV} &= \frac{qQ}{4\pi\epsilon_0} \left[-\frac{1}{r_{max}} - \lim_{x \rightarrow \infty} \frac{1}{x} \right] \\
-5\text{MeV} &= \frac{qQ}{4\pi\epsilon_0} \left[-\frac{1}{r_{max}} \right] \\
5\text{MeV} &= \frac{qQ}{4\pi\epsilon_0 r_{max}}
\end{aligned}$$

Substitution and rearrangement gives $r_{max} = 4.5 \times 10^{-14} m = 45 fm$. We accept the $10^{-10} m = 100,000 fm$ figure as the total width of the atom, but this demonstrates that there must exist a nucleus of no larger than 45fm.

2.3 Next Idea: The Solar System Model

Therefore, the next idea was an orbiting solar system model, where electrons orbit in fixed paths around a central nucleus. However, accelerating charges (i.e. a charge in circular motion) radiates energy, so this orbiting electron would be on a decaying path to crash into the nucleus. We can observe this does not happen, so need another idea...

Bohr made two postulates:

- The electron in hydrogen moves in a set non-radiating circular orbit.
- Radiation is only emitted or absorbed when an electron moves from one orbit to another.

This works (at least for hydrogen) and explains the absorption spectra, but for now lacks a physical grounding.

Fri 10 Oct 2025 12:00

Lecture 2 - The Ultraviolet Catastrophe

In this lecture:

- How classical theories fail to explain black body radiation ("The Ultraviolet Catastrophe").
- How quantising light into photons gives predictions that fit this observation.

1 Black Body Radiation

A 'black body' is an idealised perfect object, that does not reflect, and absorbs internally all light (regardless of wavelength) incident upon it. No light is transmitted, so nothing shines out the other side. The object is perfectly black.

All bodies emit electromagnetic energy, usually outside the visible portion of the spectrum. For example, Paul Hollywood (and other humans) emit at about 300 Kelvin, which is infrared (at the temperature which night vision goggles are tuned to).

For the black body, emission spectrum is **only** from this thermal emission (no reflection, no fluorescence, etc). Hotter objects are brighter and bluer (hotter means higher energy, and therefore a shorter wavelength)

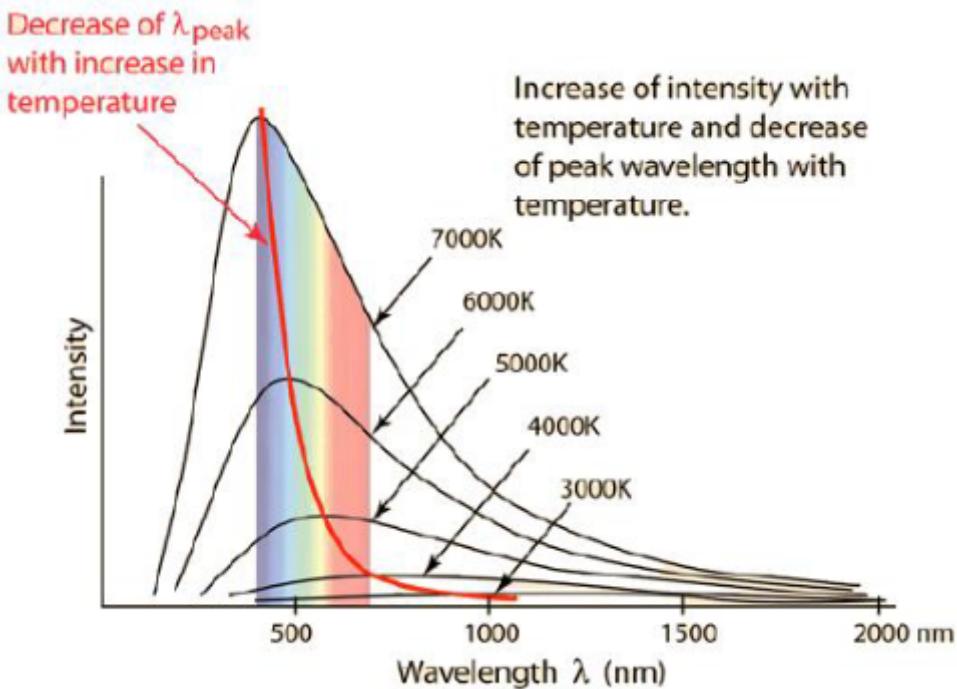


Figure 39.1: Observed Emission Spectra

However, we run into a problem. If we plot the spectra predicted by classical thermodynamics, vs the observed spectra for a given temperature object, the classical prediction gets it totally wrong, especially at shorter wavelengths.

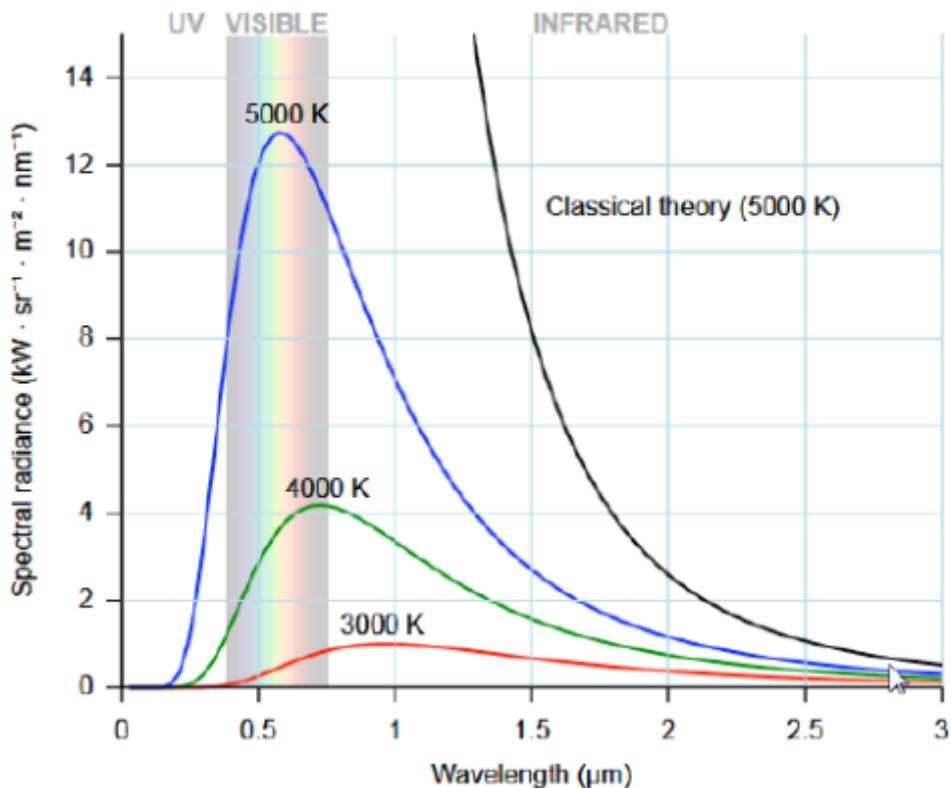


Figure 39.2: Predicted and Observed Spectra for 5000K (and Observed for 4k K and 3k K)

1.1 Notation

$I(\lambda)$ is the intensity per wavelength for an emitted wavelength λ . I is the total intensity across all wavelengths per unit time (in W/m^2 , power per unit area).

$$I = \int_0^\infty I(\lambda) d\lambda$$

I is the total area under the $I(\lambda)$ curve, i.e. the sum of intensity per wavelength, across every wavelength.

2 The Ultraviolet Catastrophe

2.1 Empirical Results

The Stefan-Boltzmann Law gives $I = \sigma T^4$, where σ is the Stefan-Boltzmann constant, $\sigma = 6.57 \times 10^{-8} Wm^{-2}K^{-4}$

Wien's Displacement Law gives $\lambda_{\text{peak}} = \frac{b}{T}$, where $b = 2.898 \times 10^{-3} Km$.

2.2 Why does classical mechanics break?

Lets model the $I(\lambda)$ spectrum by slotting standing waves into a cavity. Inside the blackbody, EM waves form standing waves (in a limited number of possible configurations).

We can simplify by considering a 1D cavity of length L . We can consider 'cavity modes' as the possible standing waves that can exist in this cavity. As we know the wave is bound at each end, the displacement at each end of the cavity must be 0. Therefore, the only possible waves must obey this, and these are cavity modes.

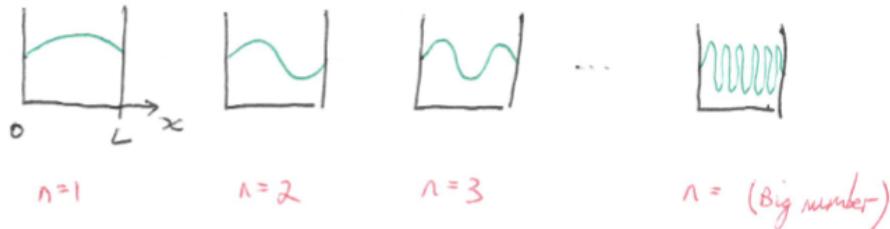


Figure 39.3: Possible cavity modes

The amplitude $a(x)$ can be given by this:

$$a(x) = \sin\left(\frac{n\pi x}{L}\right), \quad n = 1, 2, 3, \dots n \quad (\text{mode number})$$

And by inspection from the figures:

$$\lambda = \frac{2L}{n}$$

And therefore the number of nodes per wavelength is:

$$n(\lambda) = \frac{2L}{\lambda}$$

So, classically:

$$I(\lambda) \propto \frac{n(\lambda)}{\lambda} \times k_B T \propto \frac{1}{\lambda^2}$$

Where the first term is the density of nodes at lambda, and the second is the average energy of nodes. As we head to UV and $\lambda \rightarrow 0$, $I(\lambda) \rightarrow \infty \dots$ which is not accurate. This is the UV Catastrophe!

2.3 Where did it go wrong?

The issue was assuming that all cavity modes have average energy $k_B T$ - the "Equipartition Theorem" (which we'll meet in later courses).

In brief: the probability distribution of energies is a "Boltzmann Distribution":

$$p(E) = \frac{\exp\left(-\frac{E}{k_B T}\right)}{k_B T}$$

Average energy:

$$\bar{E} = \int_0^\infty E p(E) dE = k_B T$$

Which the UV Catastrophe says is incorrect... We therefore need another model to replace the idea of the Boltzmann Distribution.

2.4 Plank's Hypothesis

A rather desperate Plank hypothesised that energy was quantised, i.e. it comes in discrete packets, called quanta. The energy of these quanta is proportional to frequency. This was radical at the time, even though we accept it now.

$$\Delta E = hf = \frac{hc}{\lambda}$$

Where $h = 6.626 \times 10^{-34} \text{ Js}$ is Plank's constant.

Sticking this into the Partition Function from statistical mechanics (which we will properly encounter later on, for now don't worry!), we get an average energy:

$$\bar{E}(\lambda) = \frac{hc/\lambda}{\exp(hc/\lambda k_B T) - 1}$$

Looking at limits:

$$\text{For } \bar{E}(\lambda \rightarrow \infty) : \quad \frac{hc}{\lambda k_B T} \ll 1$$

And the Taylor Series of e^x :

$$\exp\left(\frac{hc}{\lambda k_B T}\right) \approx 1 + \frac{hc}{\lambda k_B T} + \dots$$

Yields:

$$\bar{E}(\lambda) \approx \frac{hc/\lambda}{1 + (hc/\lambda k_B T) - 1} = k_B T$$

This is a good sign, because it means that Plank's Hypothesis holds the correct classically predicted and empirically observed behaviour for higher wavelengths.

Now what about lower wavelengths, where the classical behaviour broke?

$$\text{For } \bar{E}(\lambda \rightarrow 0) : \quad \exp\left(\frac{hc}{\lambda k_B T}\right) \rightarrow \infty \quad (\text{very quickly})$$

$$\text{For } \bar{E}(\lambda \rightarrow 0) : \quad \frac{hc}{\lambda} \rightarrow \infty \quad (\text{slower})$$

Therefore, in the expression:

$$\bar{E}(\lambda) = \frac{hc/\lambda}{\exp(hc/\lambda k_B T) - 1}$$

The numerator and denominator both tend to infinity, but the denominator does so much faster. Therefore (and this can be done in a less handwavy manner via L'Hopital):

$$\bar{E}(\lambda \rightarrow 0) \rightarrow \frac{1}{\infty} \rightarrow 0$$

Which recovers the behaviour at UV wavelengths, so no Catastrophe!

3 Conclusion

- This strange quantisation hypothesis actually fits the data.
- Quantising energy means that the average energy of each cavity mode is wavelength dependant, and not fixed $k_B T$ as seen at larger wavelengths.
- This solves the UV Catastrophe!

Thu 17 Oct 2025 12:00

Lecture 3 - Particle Nature of Light

In this lecture:

- The photoelectric effect.
- Compton scattering.

Which are two examples where classical theory (light as a wave) break down.

1 The Photoelectric Effect

When shining ultraviolet light on a metal surface, electrons are emitted. This is the photoelectric effect.

Why are we not bombarded by electrons in daily life? For the electron to fly off, we must be in a vacuum. Otherwise, it'll immediately strike an air molecule and be absorbed.

Photoelectric Effect Background

- Discovered by Hertz, 1887
- Thomson (1889) went further, so did Lenard (1902) and others.
- Einstein won his Nobel Prize for explaining this, not from relativity.

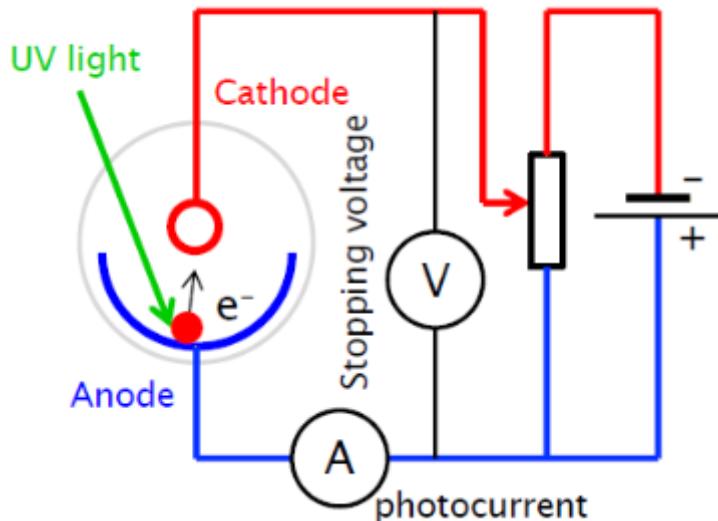


Figure 40.1: A circuit diagram for measuring the photoelectric effect.

The above setup would be encased in a glass ball (containing a vacuum), with a setup like this:

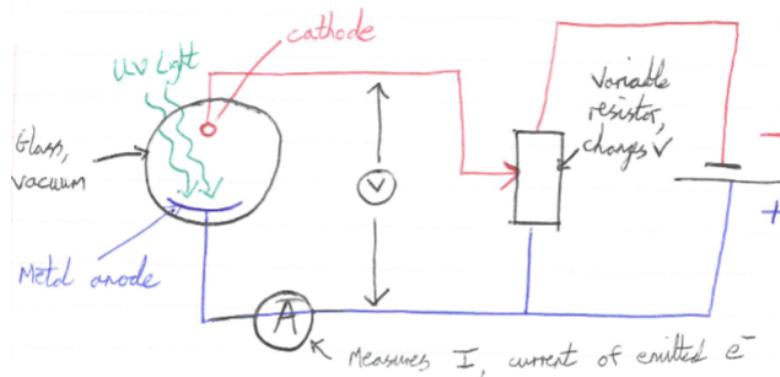


Figure 40.2: Experimental Setup

1.1 Results

Result One - Changing Intensity

For fixed UV wavelength, increasing the intensity of light increases the measured photocurrent:

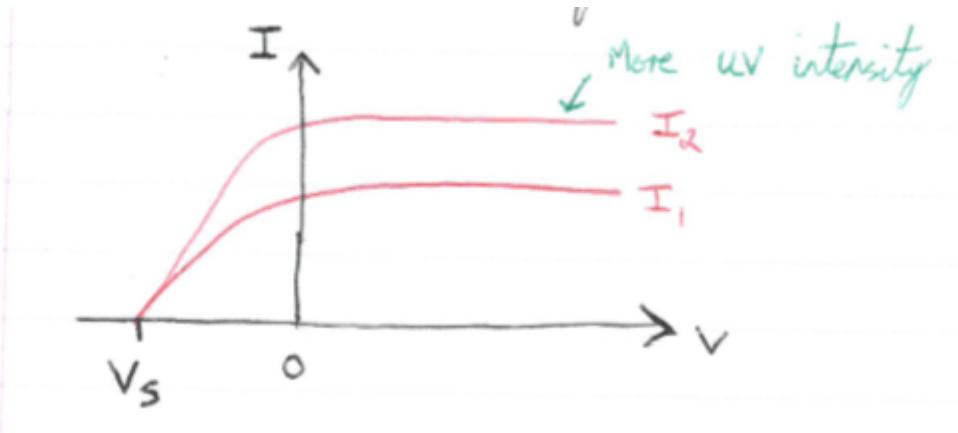


Figure 40.3

- Increasingly negative potential the cathode decreases photocurrent. At some potential v_s (the "stopping potential") this current drops to zero.
- Potential does not affect electron emission, however adding potential causes an electric field which effectively blows electrons back towards the anode. The stopping potential is when this electric field is perfectly strong to prevent electrons from reaching the cathode and causing a current.
- The fact this can happen consistently (i.e. no current means no electrons made it through) implies that there must be some maximum kinetic energy these electrons can have ($KE_{max} = eV_s$).
- The stopping potential is independent of UV intensity. More UV makes current increase, but does not change stopping potential (i.e. it does not give more energy to each electron, they each have the same energy). This does not make sense classically. Classically we would expect adding more energy to cause emitted electrons to have more energy, therefore changing the stopping potential.

Result Two - Changing Wavelength

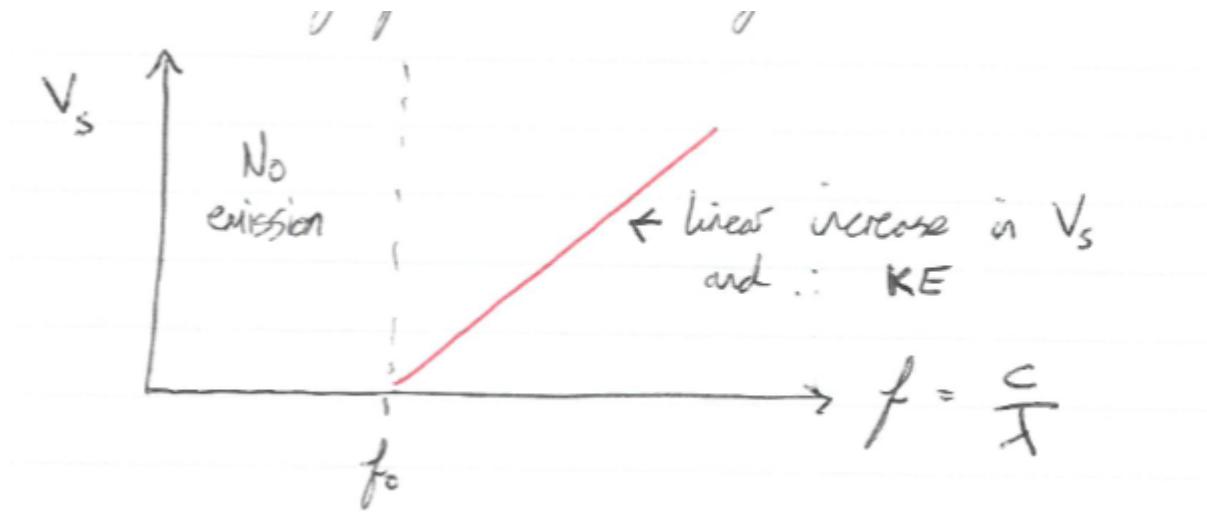


Figure 40.4

We have to reach some baseline threshold frequency f_0 before we see any photocurrent. After this, increasing wavelength increases photocurrent (and hence KE of emitted electrons) linearly.

- For a given metal, we find the threshold frequency f_0 , below which there is no emission of electrons (no current). If below the frequency f_0 , intensity is irrelevant. This contradicts classical mechanics which would suggest that turning up the light intensity would supply more (and potentially sufficient) energy.
- Above the threshold, the energy of individual emitted photons depends on UV frequency and not intensity (by result one).

1.2 Conclusions

Classically

Classically, we expect energy to be proportional to intensity, $\therefore v_s$ should increase with greater intensity. We also expect there to be no link between frequency and energy, hence no threshold frequency. We'd expect no threshold frequency, instead being a time delay as electrons "soak up" energy to reach the required threshold.

In theory, great, in practice *this is not observed*.

Einstein's Proposal

Energy in light comes from photons with energy $E = hf$. There is a minimum energy required for an electron to be able to escape from the metal. This minimum energy is called the work function ϕ .

$$KE_{\max} = hf - \phi = eV_s$$

Now:

- Higher intensity means more of the same particles (more photons), but the energy of each is unchanged.
- $E = hf$ so frequency changes energy (as observed).
- The Bohr model says that an electron can only have certain electron energy transitions when the correct energy is supplied (an electron cannot gradually soak up energy). This explains why there is a cutoff below the work function, and no observed time delay (as the "soaking up" that causes the delay does not happen). Either an incoming photon has sufficient energy, or it does not. Having more photons does not help.
- The first incoming photons immediately releases an electron (assuming the incoming light has sufficient energy), therefore there's no time delay.

1.3 In Practice

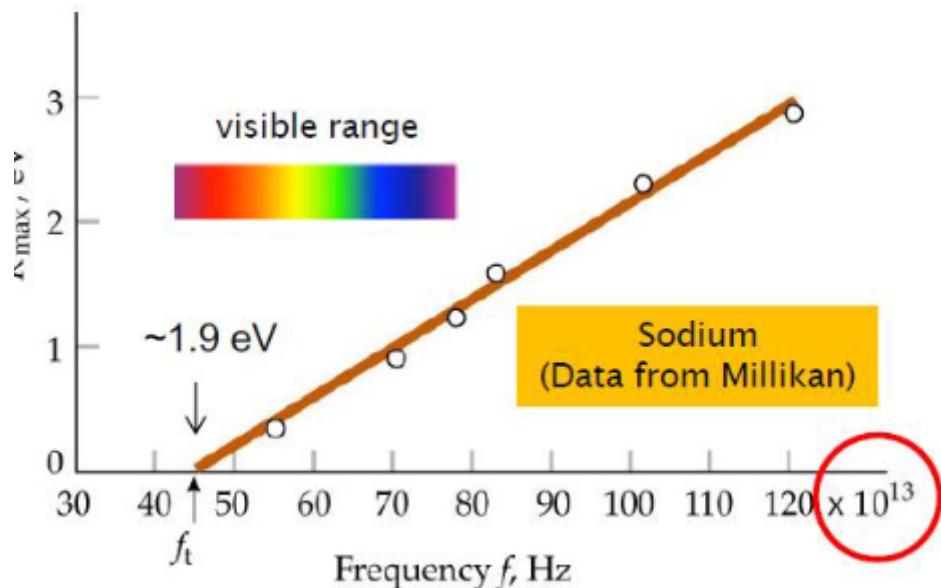


Figure 40.5: Sodium photocurrent measurements by Robert Millikan

2 Compton Scattering

Compton Scattering is the scattering of x-rays off carbon atoms.

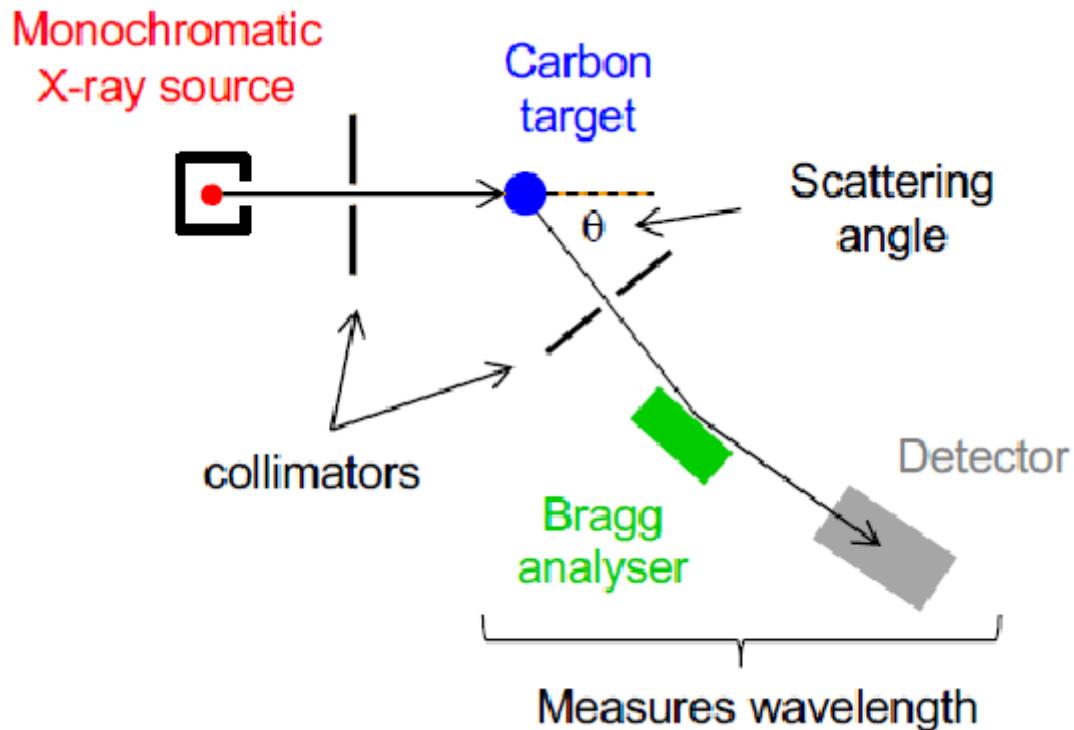


Figure 40.6: A Compton Scattering experimental setup.

The surprising result is that two wavelengths were observed (not just the original) - λ_1, λ_2 , where λ_1 is the original and λ_2 is different. Classically this is hard to explain and λ should not change.

The difference between these two wavelengths increases with scattering angle θ . This can be explained if the x-ray beam is a stream of photons, but not classically.

Possible options when a photon collides:

- Elastic: No loss of energy, hence no change in wavelength.
- Inelastic: Change of energy, so the wavelength also changes. An electron is fired out of the target, carrying energy and momentum, so the photon loses energy.

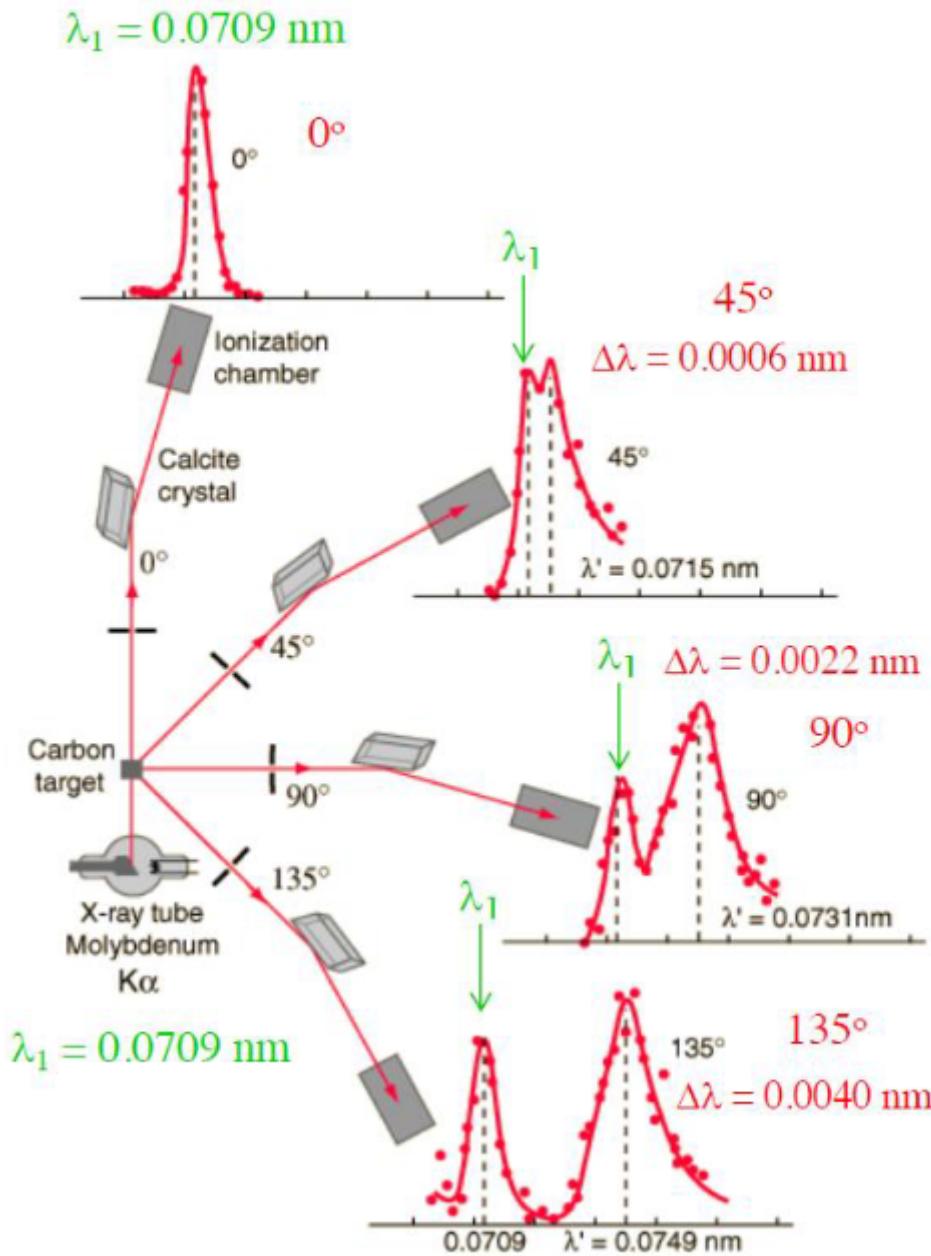


Figure 40.7: The observed results

3 Deriving Compton's Equation

Given an incoming photon with energy E_1 , wavelength λ_1 and momentum p_1 . This strikes a carbon atom and is deflected by angle θ . There is also an emitted electron at angle ϕ which must be in the opposite direction (angled up vs down) to conserve momentum. The new deflected photon has E_2 , λ_2 , p_2 .

We must consider relativistic effects here given the high speed ($E^2 = p^2c^2 + m^2c^4$)

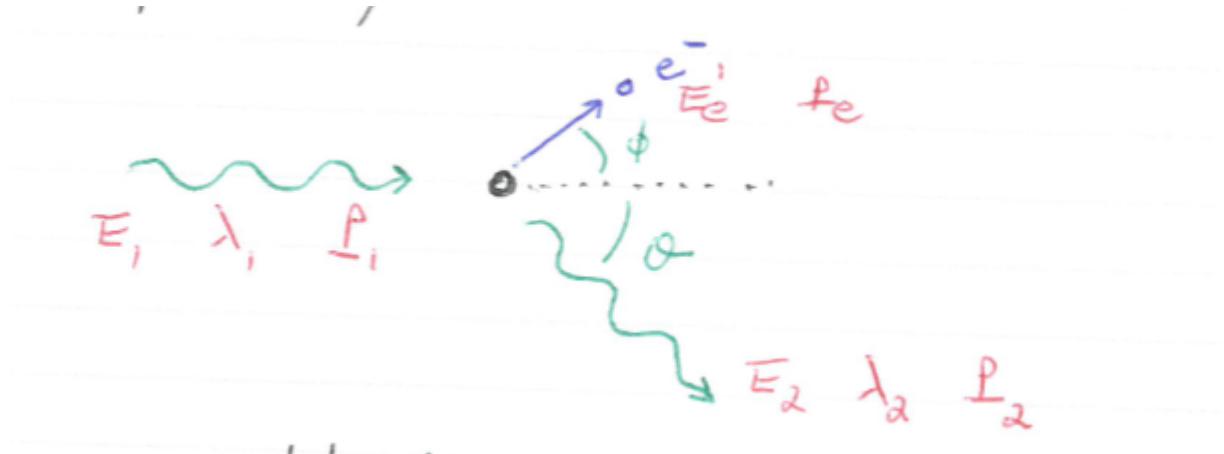


Figure 40.8

3.1 Setup

For a massless photon ($m = 0$):

$$E = pc$$

And:

$$E = \frac{hc}{\lambda}$$

So:

$$p = \frac{h}{\lambda} \quad (40.1)$$

3.2 Conservation and Relativity

Conserving momentum (underlines omitted for speed):

$$p_1 = p_e + p_2$$

$$p_e = p_1 - p_2$$

Squaring both sides:

$$\begin{aligned} p_e^2 &= p_1^2 + p_2^2 - 2p_1 \cdot p_2 \\ p_e^2 &= p_1^2 + p_2^2 - 2p_1 p_2 \cos \theta \end{aligned} \quad (40.2)$$

And then by conservation of energy:

$$E_1 + E_e = E_2 + E'_e$$

Where E_1 is the incoming photon energy, E_e is the energy of the electron at rest in atom before collision, E_2 is the deflected photon energy and finally E'_e is the deflected electron's energy.

Using this:

$$\begin{aligned} p_1 c + m_e c^2 &= p_2 c + \sqrt{p_e^2 c^2 + m_e^2 c^4} \\ \implies p_1 - p_2 + m_e c &= \sqrt{p_e^2 + m_e^2 c^2} \end{aligned}$$

And squaring both sides:

$$(p_1 - p_2)^2 + m_e^2 c^2 + 2m_e c(p_1 - p_2) = p_e^2 + m_e^2 c^2$$

Substituting in Eqn 40.2 for p_e^2

$$\begin{aligned} (p_1 - p_2)^2 + 2m_e c(p_1 - p_2) &= p_e^2 \\ (p_1 - p_2)^2 + 2m_e c(p_1 - p_2) &= p_1^2 + p_2^2 - 2p_1 p_2 \cos \theta \end{aligned}$$

And rearranging:

$$\begin{aligned}
 (p_1 - p_2)^2 + 2m_e c(p_1 - p_2) &= p_1^2 + p_2^2 - 2p_1 p_2 \cos \theta \\
 p_1^2 + p_2^2 - 2p_1 p_2 + 2m_e c(p_1 - p_2) &= p_1^2 + p_2^2 - 2p_1 p_2 \cos \theta \\
 -2p_1 p_2 + 2m_e c(p_1 - p_2) &= -2p_1 p_2 \cos \theta \\
 -p_1 p_2 + m_e c(p_1 - p_2) &= -p_1 p_2 \cos \theta \\
 m_e c(p_1 - p_2) &= -p_1 p_2 \cos \theta + p_1 p_2 \\
 m_e c(p_1 - p_2) &= p_1 p_2 (1 - \cos \theta)
 \end{aligned}$$

Substituting Eqn 40.1:

$$\begin{aligned}
 m_e c(p_1 - p_2) &= p_1 p_2 (1 - \cos \theta) \\
 m_e c \left(\frac{h}{\lambda_1} - \frac{h}{\lambda_2} \right) &= \frac{h}{\lambda_1} \frac{h}{\lambda_2} (1 - \cos \theta) \\
 m_e c \left(\frac{h}{\lambda_1} - \frac{h}{\lambda_2} \right) &= \frac{h^2}{\lambda_1 \lambda_2} (1 - \cos \theta) \\
 m_e c \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) &= \frac{h}{\lambda_1 \lambda_2} (1 - \cos \theta) \\
 m_e c \left(\frac{\lambda_1 \lambda_2}{\lambda_1} - \frac{\lambda_1 \lambda_2}{\lambda_2} \right) &= h (1 - \cos \theta) \\
 m_e c (\lambda_2 - \lambda_1) &= h (1 - \cos \theta) \\
 (\lambda_2 - \lambda_1) &= \frac{h}{m_e c} (1 - \cos \theta)
 \end{aligned}$$

Which is the Compton Equation. This shows that the change in wavelength is proportional to $1 - \cos \theta$.

4 Conclusions

The photoelectric effect and Compton scattering are two more physical phenomena that cannot be explained using traditional classical mechanics with EM waves alone. They both require assuming photons of energy $E = hf$ to be adequately explained.

Fri 24 Oct 2025 12:00

Lecture 4 - Atomic Energy Levels and Spectra

In this lecture:

- The spectra of light emitted and absorbed by electrons in:
 - Hydrogen (simple)
 - Larger atoms (not simple...)
- Electronic shells and orbits.

1 Spectra

The Bohr Model says that electrons in an atom can hold discrete values “levels” of energy. As electrons go up or down these levels they must absorb or will emit a photon. This emission causes a discrete spectra of emitted frequencies, unique to the element causing it.

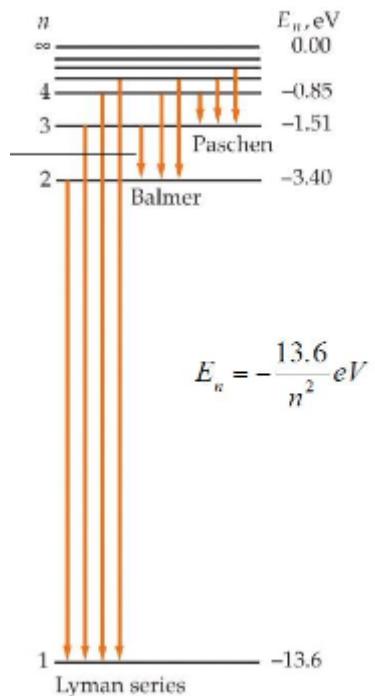


Figure 41.1: Example emission spectra.

Each of the transitions from one energy level to another have a discrete change in energy (measured in eV), therefore each transition will have a discrete wavelength of produced photons.

From the Bohr model, he postulated (from experimental observations) that light can *only* be absorbed or emitted when an electron goes up or down a discrete energy level (excitation or relaxation). The lowest energy level is known as the “Ground State” ($n = 1$).

It is important to note that the transitions in energy levels are not spatial - changing the energy level of an electron does not physically change its position (despite what the diagrams may imply).

2 Absorption and Emissions

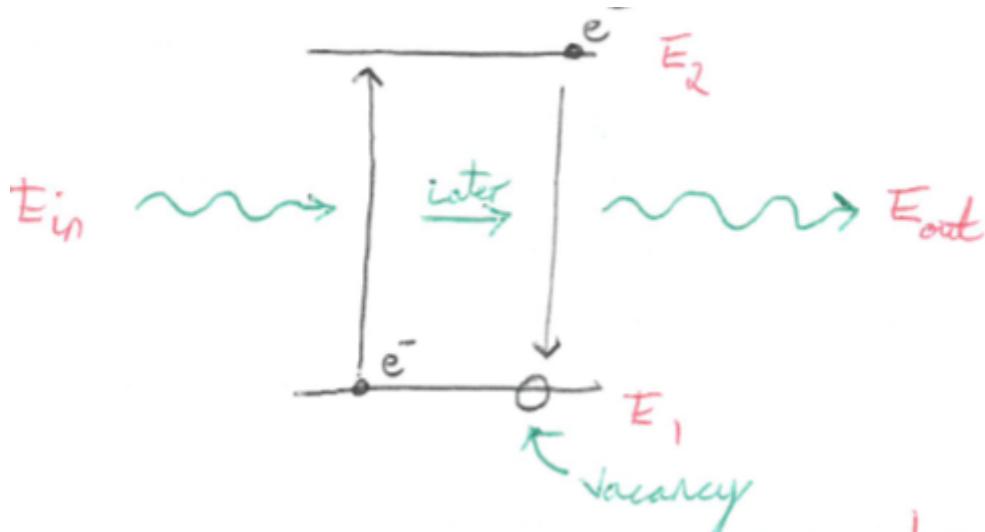


Figure 41.2

An electron is excited up an energy level, leaving a vacancy behind. Some time after, the electron drops back down into the ground state, emitting the energy in terms of a photon. By conservation of energy:

$$E_{\text{in}} = E_{\text{out}} = E_2 - E_1 = hf = \frac{hc}{\lambda}$$

Notably, the energy of the photon that initially triggered the excitation is given by $E_\gamma = E_2 - E_1$

3 Atomic Hydrogen

Atomic Hydrogen is by far the simplest example we can deal with (given the single proton and single electron). This single electron is 'orbiting'¹ the nucleus at some distance r .

The potential energy of this electron is:

$$v(r) = \frac{-e^2}{4\pi\epsilon_0 r}$$

This P.E. is negative, so the electron is trapped in a potential well and must be supplied with energy to escape, assuming no K.E. We also know (if we take into account kinetic energy too):

$$E = T + V$$

Where E is total energy, T is K.E, V is P.E. If the total is negative, the electron is bound to the atom.

We can see from the equation for potential energy that $v(r) \propto -1/r$, which gives us the green line. Of the points on this green line, only some of them are actually discrete allowed orbits. The region at the top is called the continuum, which is where an electron has positive energy and has therefore left the nucleus.

¹for the sake of argument, even if in practice it does not really

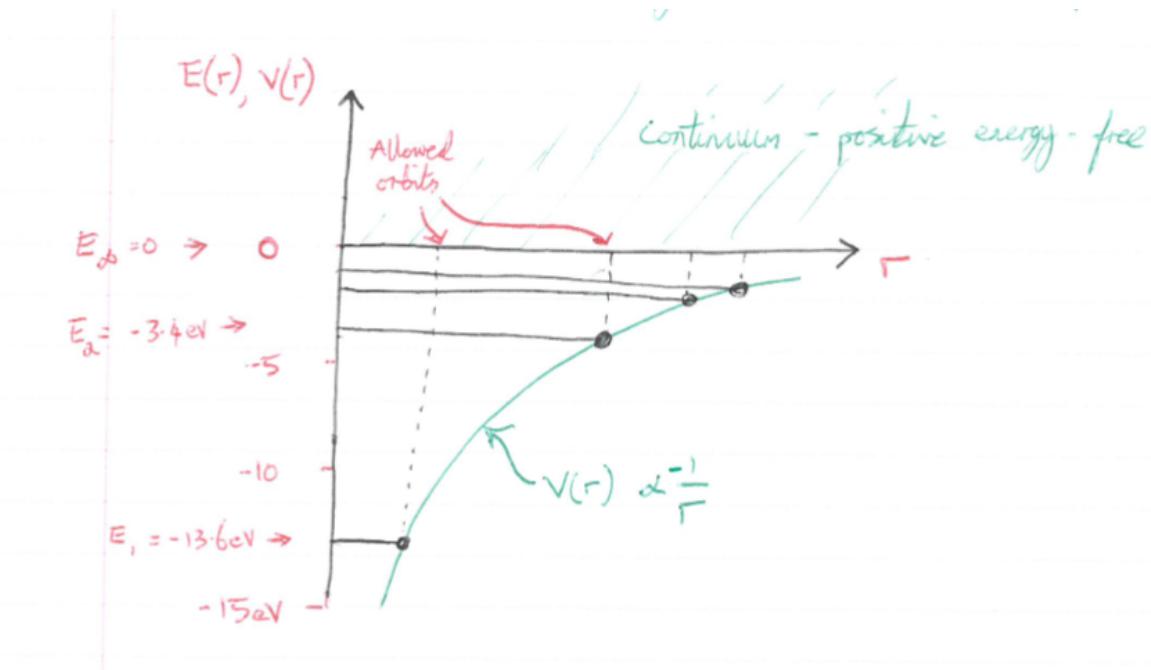


Figure 41.3

In Lec 08, we will properly derive from the Bohr model that the energy of each energy level is given by this:

$$E_n = \frac{-13.6\text{eV}}{n^2} \quad n = 1, 2, 3, \dots, \infty$$

Note that if $n \rightarrow \infty$, $E_n \rightarrow 0$. Emission and absorption happens when an electron moves from two different energy levels ($m \rightarrow n$).

$$E_\gamma = E_n - E_m = 13.6(1/n^2 - 1/m^2)\text{eV}$$

$$\frac{hc}{\lambda} = E_n - E_m = 13.6(1/n^2 - 1/m^2)\text{eV}$$

Or, finally:

$$\frac{1}{\lambda} = \frac{13.6\text{eV}}{hc} \left(\frac{1}{n^2} - \frac{1}{m^2} \right)$$

The $\frac{13.6\text{eV}}{hc}$ term is known as the Rydberg constant.

3.1 Balmer Series

The 'Balmer Series' is a portion of the hydrogen emission spectra which happens to take place at visible wavelengths. It is specifically transition which take place to $n = 2$. To find these wavelengths, we set $n = 2$:

$$\frac{1}{\lambda} = R \left(\frac{1}{2^2} - \frac{1}{n^2} \right)$$

Moving from 3- \downarrow 2 gives us the Balmer Alpha line, for example.

4 Ionisation Energy of Hydrogen

The ionisation energy is the minimum energy required to kick an electron out of the ground state and into the continuum. This is just enough energy to barely make it free (i.e. when the continuum is reached the electron has 0 KE). Setting $n = 1, m = \infty$:

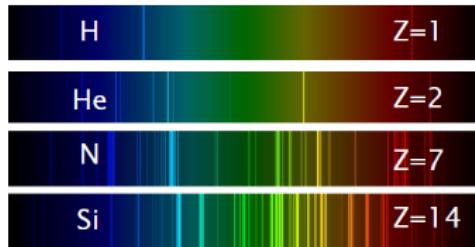
$$\frac{1}{\lambda} = R \left(\frac{1}{1} - \frac{1}{\infty} \right) = R$$

Where R is the Rydberg constant. Note: the energy levels can be very sharp, and they do not change (unless external stimuli like a strong external magnetic field are applied). We can therefore use this in e.g. atomic clocks.

5 Bigger Atoms

Bigger atoms gets more complex than we can reasonably consider now.

- Increasing complexity



- Sodium ($Z = 11$) energy levels

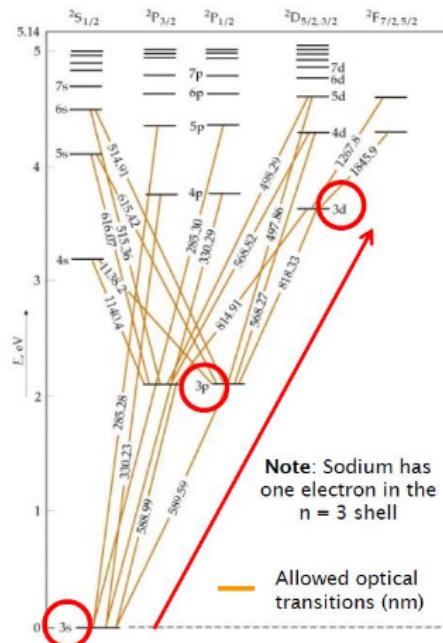


Figure 41.4

Electrons begin interacting with each other and the nucleus in weird ways, and it becomes too messy to calculate easily. If there is only a single electron, we can still use the Bohr model, just with a different charge of the nucleus, and we multiply by Z^2 when calculating. If we have multiple electrons, there is no general formula and it becomes unpleasant. The energy levels become split into multiple possible orbitals, i.e for $n = 1, 1s$. For $n = 2, 2s, 2p$. For $n = 3, 3s, 3p, 3d$ (where s, p, d are different electronic orbitals).

5.1 Spin

Electrons also have an intrinsic quantum property called 'spin', with a value $\frac{1}{2}$. They are fermions with two spin states (up, down). Note that spin is quantum, and has no relation to a physical geometric spin. The maximum occupancy of each level is $2n^2$ (i.e. the 1s state has space for one electron of each spin, i.e. one spin up and one spin down).

6 More Complexity...

- **Fine Splitting:** Electrons are moving, so have non-zero angular momentum. A moving charge creates a magnetic field, which changes the energy of the other electrons.
- **Hyperfine Splitting:** The nucleus also has a spin, they can be parallel (i.e. both have the same spin), or antiparallel, where the electron and the nucleus have opposing spins. These two configurations have slightly different energies (difference of $5.9\mu eV$). The parallel state is metastable, with a half life of 10 million years. There is so much hydrogen in the galaxy that radioastronomers can use this decay, and by detecting it they can image the location of hydrogen in the universe.

Fri 31 Oct 2025 12:00

Lecture 5 - X-Ray Production and Diffraction

In this lecture:

- Production of x-rays
- How to measure their wavelength
- Bragg scattering of x-rays by crystals and Bragg conditions.

1 Production of X-Rays

From the blackbody spectra lecture, we saw that something becoming hotter affects the wavelength of emission. Surely we, therefore, can just heat something hot enough for the peak of the emission spectra to be in the x-ray range?

Unfortunately not, this works, but would need the black body to be 1MK.

The more practical alternative is to fire high energy electrons (keV range) into a metal target in a vacuum.

2 What are they?

Electromagnetic radiation (light) outside of the visible spectrum. It sits beyond the UV portion.

Approximately 0.1keV ($\lambda = 10nm$) to 100keV ($\lambda = 0.01nm$), these are called soft and hard x-rays respectively.

3 Measuring Wavelength

How can we measure or select x-ray wavelengths? With visible light, we can use a prism to turn wavelength into an angle and measure the angle. However, a prism works for visible light because of its high refractive index. Unfortunately, for x-rays, $n \approx 1$ for all materials.

We therefore need some other way to split up wavelengths into angles, for which we can use a diffraction grating. Different wavelengths will be diffracted at different angles.

Another problem arises however... We need the slit separation d to be $d \approx \lambda$ for a diffraction grating to work. For x-rays, $\lambda \approx 10^{-10}m$, which is approximately the width of an atom. Good luck making that grating...

Luckily, nature has made these gratings for us — crystalline materials! Strontium titanate ($SrTiO_3$) has an approx 0.389nm spacing between strontium atoms. We can therefore use this crystal lattice as a diffraction grating.

4 X-Ray Diffraction from Crystals

We treat the x-rays as EM waves (and ignore weird photon stuff for now) and we assume that each atom scatters independently. We also assume that the atom absorbs the x-ray and later re-emits it, in all directions uniformly (at photon level, the emission is in a random direction, therefore we probabilistically treat it as in all).

We want to find the angles where *constructive interference* occurs. Zero intensity everywhere else due to very very high number of slots.

The Bragg Conditions determine where this interference happens. Consider a single plane of equally spaced atoms.

4.1 First Bragg Condition

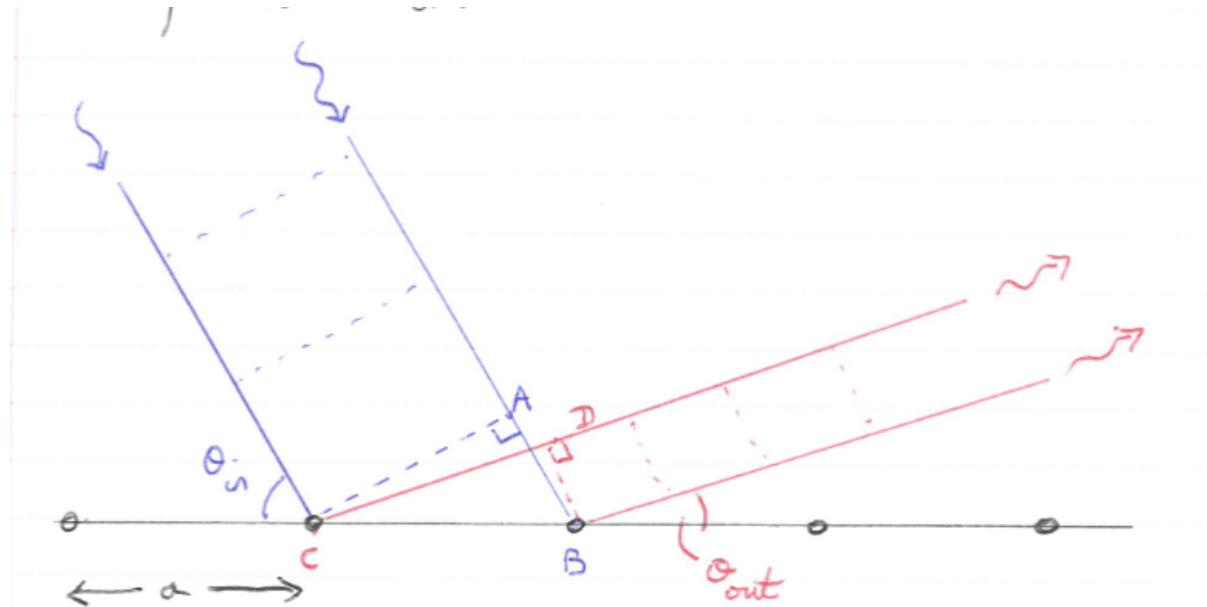


Figure 42.1

The outgoing waves (red) are all in phase when the wavefronts line up perpendicularly. This happens when path lengths AB and CD are the same. So:

$$a \cos \theta_{in} = a \cos \theta_{out}$$

$$\theta_{out} = \theta_{in}$$

We disregard higher orders where:

$$\cos(\theta_{out} + 2n\pi) = \cos(\theta_{in} + 2k\pi), \quad n, k \in \mathbb{Z}^+$$

And we only consider the case where they are explicitly equal. Note that, so far, this is independent of wavelength (i.e. there is 'no dispersion'). This means we're not acting like a diffraction grating yet, more of a mirror.

4.2 Second Bragg Condition

Now we consider the inclusion of a second plane of atoms (we have some much larger number of n planes).

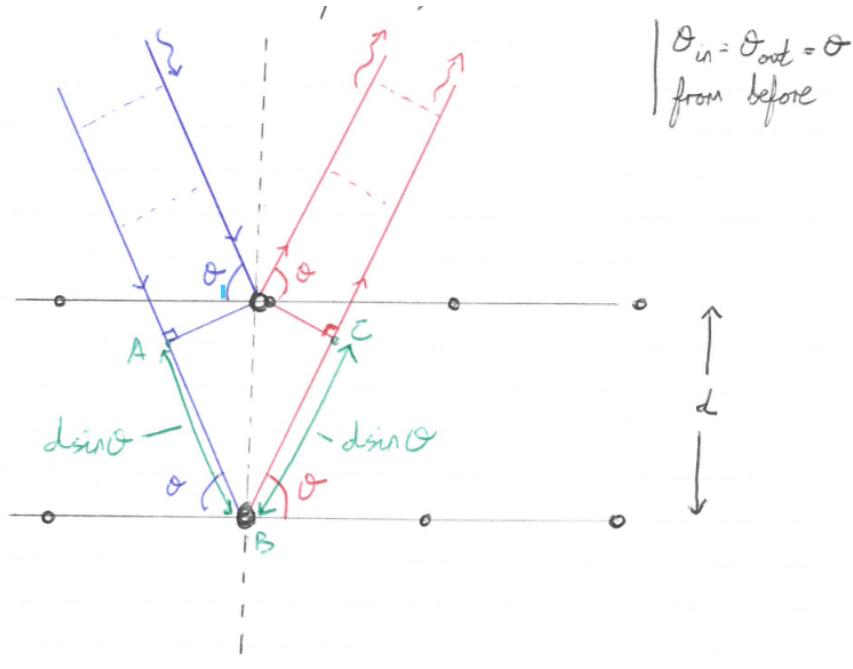


Figure 42.2

The ray which hits the lower plane travels further than the ray which hits the upper plane. For constructive interference to take place, this extra path difference must be an integer multiple of the wavelength. So:

$$AB + BC = 2d \sin \theta = n\lambda \quad n = 1, 2, 3, \dots$$

Where d is the separation of crystal planes. Note: This θ is not the beam deflection angle. It is the angle between the **horizontal** and the incoming beam.

This equation gives us different angles for different wavelengths, therefore we can select light (inc X-rays) of a certain wavelength by splitting multi-wavelength light into different angles, and physically selecting the one that we wish. This throws back to Lec 03, where a Bragg Analyser was used with a detector to measure the wavelength of the produced X-rays.

5 Uses

We can use this in two ways:

We can determine an unknown wavelength (or select a wavelength out of multi-frequency light) by varying θ and identifying the detection peak. Note that θ is here twice, the detector must also be moved, and therefore the total deflection angle is 2θ

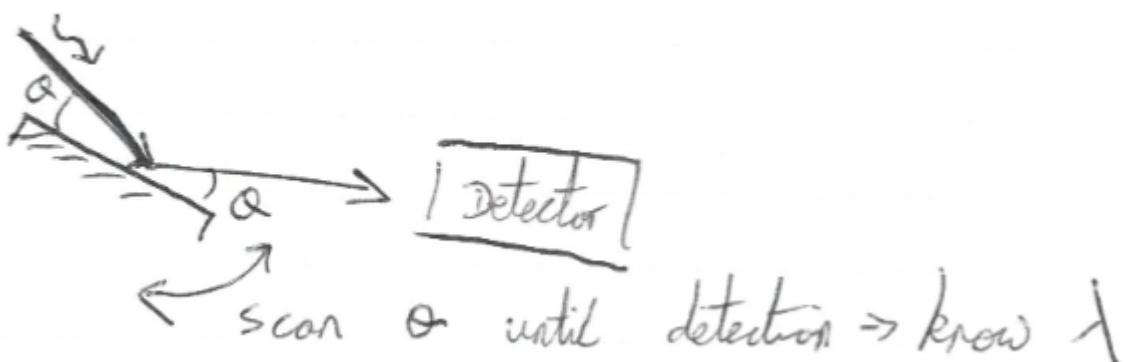


Figure 42.3

Alternatively, we can know lambda and fire at an unknown crystal to determine the atomic spacing. We can scan through 2θ and identify the angles where $n\lambda = 2d_i \sin \theta$ for various d_i . This leads us to x-ray crystallography, where we use x-rays to work out the structure of materials. Using crystals to measure or select an x-ray wavelength is called "wavelength dispersive x-ray spectroscopy".

Fri 07 Nov 2025 12:00

Lecture 6 - X-Ray Spectra

In this lecture:

- The production of X-rays - the spectrum created.
- Attenuation of X-rays in matter.
- Absorption of X-rays in materials, and 'absorption edges'.

1 Typical Tube Spectra

As a reminder, the source is metal bombarded with high-energy electrons. This causes electron energy level changes in deep electron shells in the metal. This spectrum is characteristic of the anode (the thing being hit by the electrons) material.

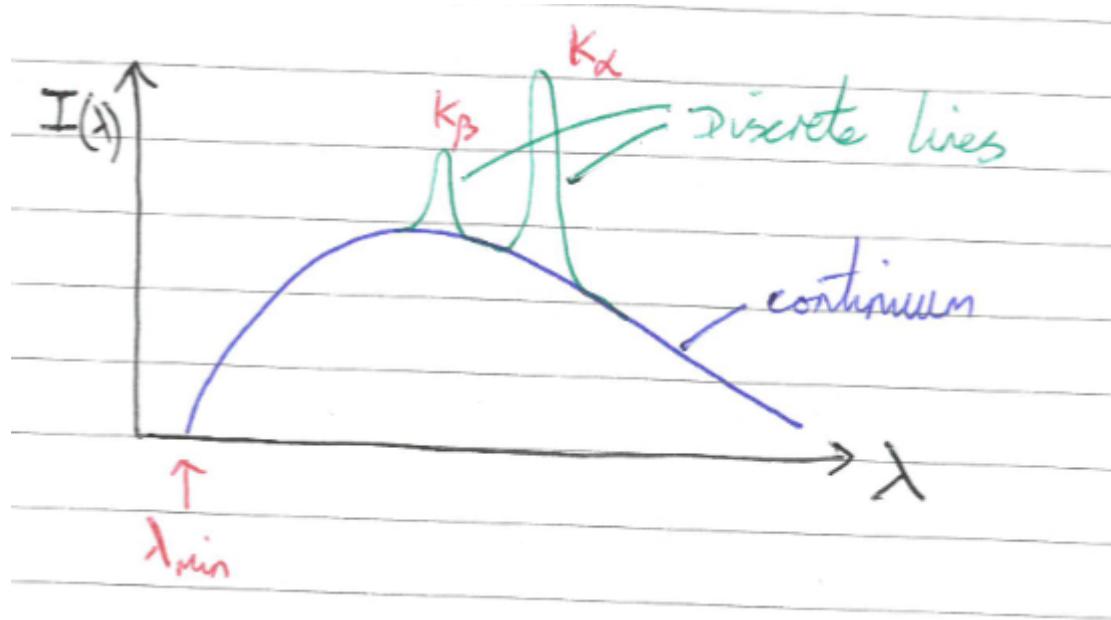


Figure 43.1: A typical x-ray spectra, with a background continuum and unique discrete peaks. Note there is some min λ

The continuous background curve is due to 'Bremsstrahlung', or 'braking radiation'.

2 Braking Radiation

Braking radiation is independent of anode materials. As an electron comes in at a high KE, it interacts with matter (charged nuclei etc) it is deflected and slows down. This lost KE is emitted as X-rays.

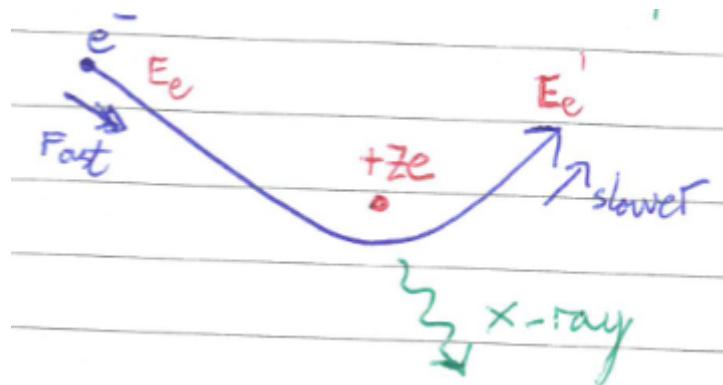


Figure 43.2

2.1 Lambda Min

By C.O.E:

$$E_{\text{x-ray}} = E_e - E'_e = \frac{hc}{\lambda}$$

If the electron came to a stop entirely, E'_e would be minimised (braking radiation maximised). The minimum wavelength λ_{\min} occurs at the highest x-ray energy, i.e. when E'_e is zero.

$$\begin{aligned} E_e - 0 &= \frac{hc}{\lambda_{\min}} = eV \\ \lambda_{\min} &= \frac{hc}{eV} \end{aligned}$$

Where V is the accelerating voltage.

3 Discrete Lines

These arise from inner shell electron transitions. These, unlike Bremsstrahlung, are characteristic of the anode material. We can use this to investigate the anode material, or we can use different materials to build different x-ray tubes with different emission spectra.

4 Jargon

Note the following stupid jargon.

- k means the $n = 1$ energy level.
- l means the $n = 2$ energy level.
- et cetera.

Lets again consider the atomic energy levels (and their associated maximum occupancy):

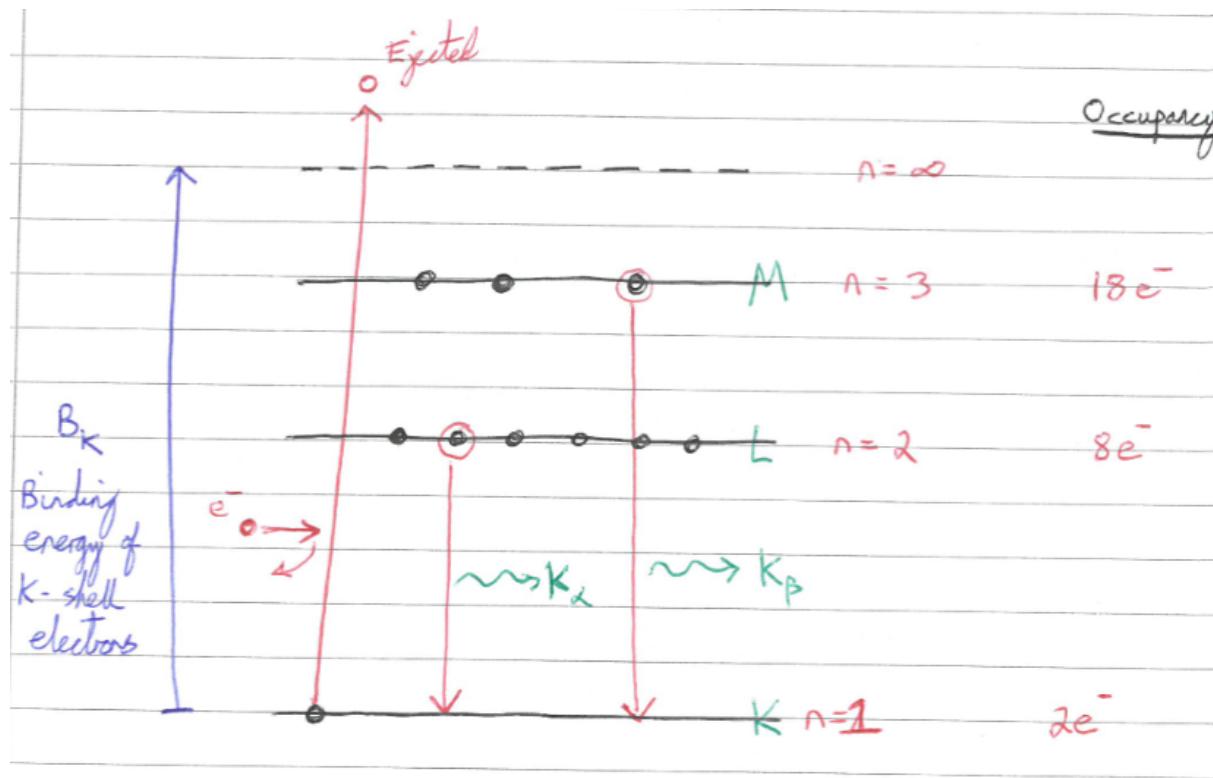


Figure 43.3

If the incoming energy eV is bigger than the binding energy B_k of the k-shell electrons, they can be ejected. This is called 'collisional ionisation'. The intention is to eject an electron from the $n=1$ shell. This leads to an open space in the $n=1$ shell (a 'hole'), and the atom will automatically settle into the lowest energy state, i.e. a higher level electron will drop down to fill the lower energy hole. This change in energy requires the emission of a photon.

This specific $n = 2$ to $n = 1$ state transition is called k_α (the k denoting 'to the 1st level'), while $n = 3$ to $n = 1$ would be k_β etc. Transitions from the third to the second would be L_α , etc etc.

k_α etc only appear if bombardment energy is sufficiently high to cause collisional ionisation. Lower energy lines will therefore appear sooner.

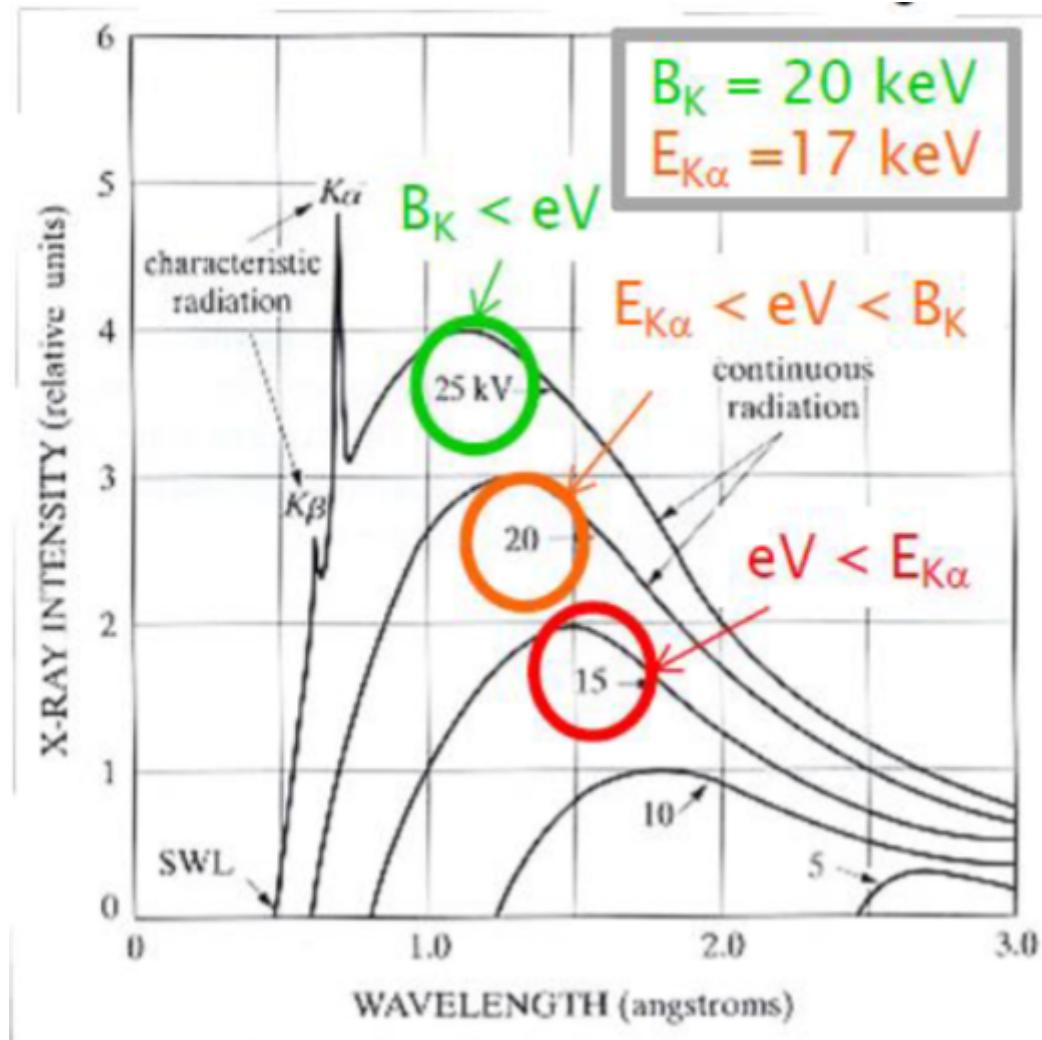


Figure 43.4: Molybdenum X-Ray Spectra (1 Angstrom = 1Å = 0.1nm)

Note that, while Bremsstrahlung is present consistently (albeit with a changing min wavelength where higher accelerating voltage leads to a higher energy x-ray and a lower min wavelength), the sharp spectra lines only appear at higher accelerating voltages where a sufficiently high threshold energy to trigger collisional ionisation is met.

5 X-Ray Absorption

As x-rays pass through a material, x-ray intensity will fall exponentially with distance travelled. Suppose we hit a material with x-rays of intensity I_0 . These x-rays travel x distance units. The final intensity is:

$$I = I_0 e^{-\mu x}$$

Where μ , the attenuation coefficient, depends on the material and incoming energy.

$$\mu \propto \frac{1}{E_{\text{x-ray}}^3}$$

This hearkens back to the photoelectric effect.

Absorption drops as energy of the x-ray falls below the binding energy of a given shell.

6 Absorption Edges

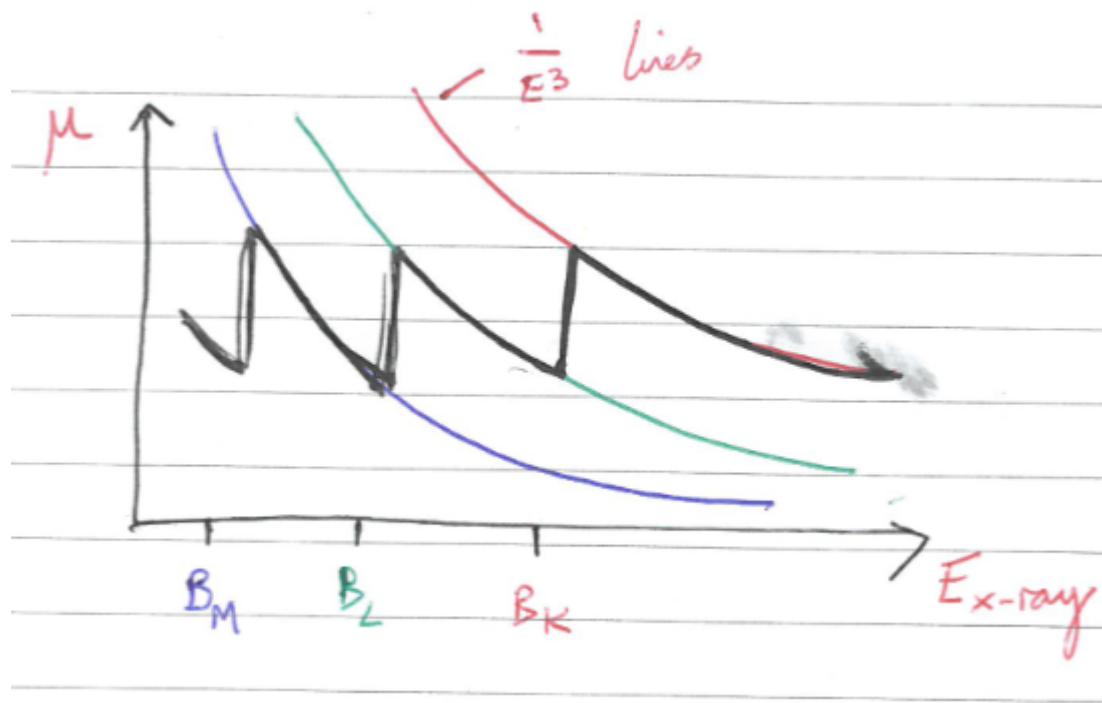


Figure 43.5: Absorption varying with x-ray energy.

Each line represents the $1/E^3$ value for a given energy shell. If we decrease energy and follow the red line along, we eventually hit the point where there is no longer sufficient energy for ejection from the k-shell. We then drop, and can now eject from everything except from the k-shell until we reach the point where we can no longer eject from the k-shell or the l-shell. We drop again, (blue line) and only have emission from all shells except for k and l.

Note:

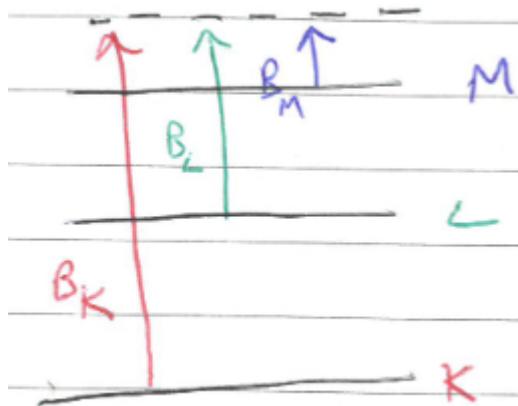


Figure 43.6

- Measuring these ‘absorption edges’ lets us determine what atom something is made of, and the associated electronic energy levels.
- Describing these as hard lines is inaccurate, there is (in reality) much more fine detail.
- This process is called x-ray absorption spectroscopy.

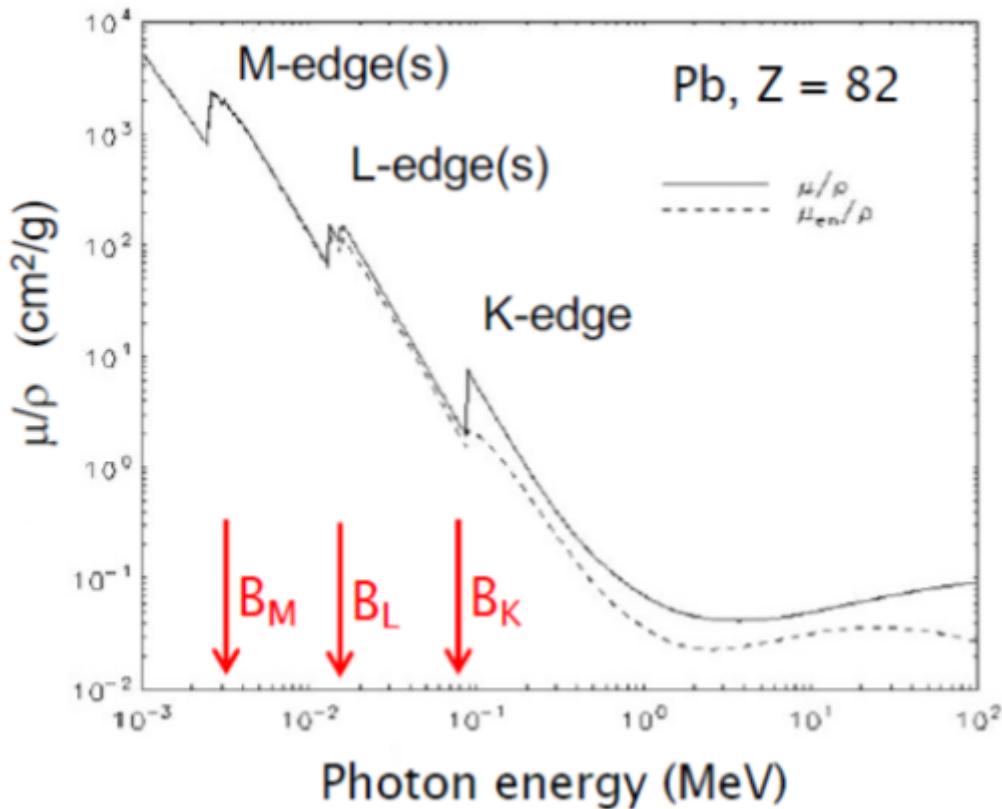


Figure 43.7: Fine detail visible in absorption edges (especially visible on L-edge).

7 Moseley's Law

If you measure the frequency of k alpha for some material, it is equal to a constant times $(Z-1)$ squared, where Z is the atomic number. This lets us identify elements:

$$f_{k\alpha} = (2.48 \times 10^{15})(Z - 1)^2$$

Note: the ' -1 ' appears because there will be a second electron 'left behind' in the k shell, which reduces effective nuclear charge by one. This is called screening or shielding.

Fri 14 Nov 2025 12:00

Lecture 7 - “Matter Waves”

In this lecture:

- Particles acting as waves: de Broglie wavelength.
- Proof: Davisson and Germer experiment.

1 de Broglie Wavelength

We've seen light waves behaving like particles - photons have an energy and momentum related to their wave frequency and wavelength. We can see however that this goes both ways - particles can also behave as waves.

De Broglie suggested that matter/particles should have wave-like properties, and indeed everything behaves as matter, and as waves, and is simultaneously both and neither.

He proposed the De Broglie wavelength:

$$\lambda = \frac{h}{p} = \frac{h}{mv} \quad (44.1)$$

A bigger momentum mv means a smaller wavelength. Taking Paul Hollywood as an example, if he has a 250kg mass and can run at 43mph (=20m/s):

$$p = 5,000 \text{ kgms}^{-1}$$

$$\lambda = \frac{6.6 \times 10^{-34}}{5000} = 10^{-37}$$

Which is so small it's practically irrelevant. Macroscopic objects have such small wavelengths we can effectively ignore them, hence why classical mechanics still holds on a larger scale.

1.1 For an Electron

For an electron accelerated by 54V, we have a KE of 54eV. First we check that this is non-relativistic:

$$m_e c^2 = (9 \times 10^{-31})(3 \times 10^8)^2 = 8 \times 10^{-14} \text{ J} = 500,000 \text{ eV}$$

Therefore KE is much less than the relativistic mass-energy, so classical mechanics are fine. Now:

$$E = \frac{1}{2}mv^2 = \frac{p^2}{2m} \implies p = \sqrt{2mE}$$

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2mE}} = \frac{h}{\sqrt{2meV}}$$

$$\lambda = 1.67 \times 10^{-10} \text{ m} = 1.67 \text{ Å}$$

This is no longer small enough that we can get away with totally ignoring it.

2 Experimental Verification - Davisson and Germer

They fired electrons into the surface of metal (with an accelerating voltage of 54V). They measured electron intensity in a scattered beam vs angle. They predicted that electrons should be diffracted, as if they were waves. They observed a constructive interference pattern, equivalent to x-rays.

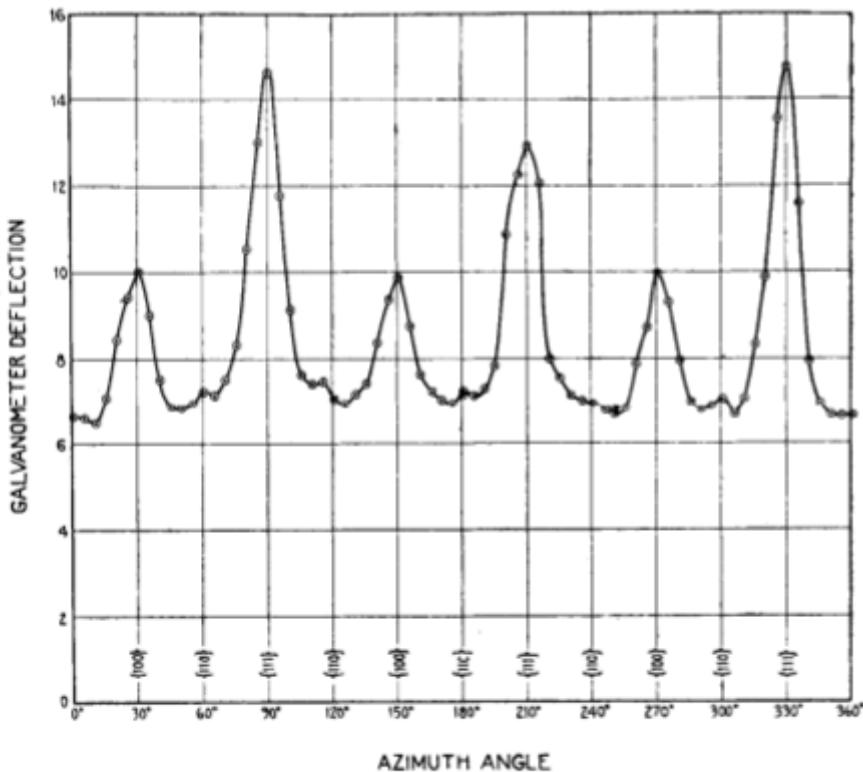


Fig. 2.—Intensity of electron scattering vs. azimuth angle—54 volts, co-latitude 50°.

Figure 44.1

3 Their Result

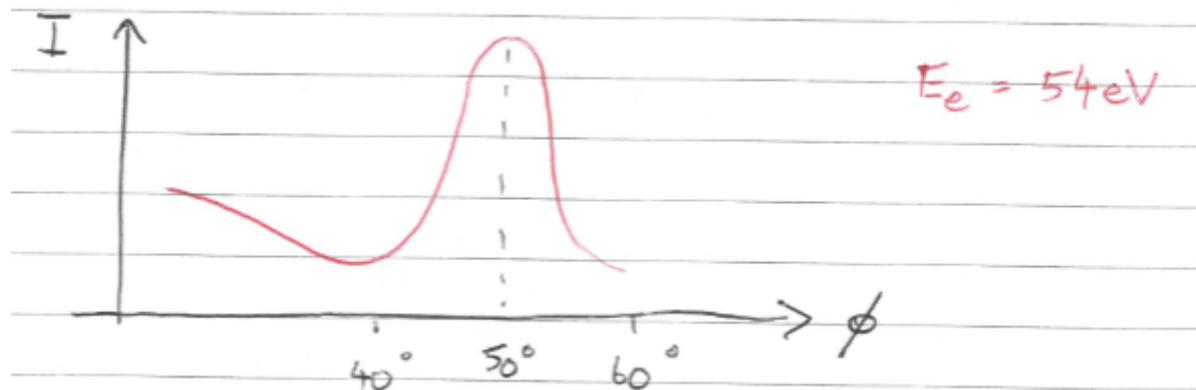


Figure 44.2

We interpret the peak at $\phi = 50$ as due to electron diffraction. Note that the 'scattering angle', ϕ measured is not the same as the Bragg angle θ .

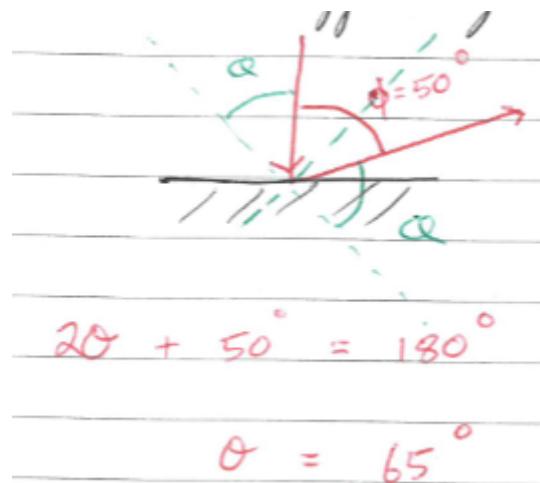


Figure 44.3

3.1 Explaining the Result

Given the Bragg condition $2d \sin \theta = n\lambda$, and $d = 0.91\text{\AA}$, we take $n = 1$ for the central peak, and substitute to get $\lambda = 1.65 \times 10^{-10}\text{m}$. This is (approximately) what we predicted earlier. Note that this is now the Bragg angle θ , not ϕ .

3.2 GP Thomson

At the same time, G.P. Thomson conducted 'powder diffraction'. Grinding a crystal into a powder creates a mess of smaller sub-crystals. This replaces the need for scanning and matching in/out angles, as there will be some correct orientation crystal for any input angle. Since the many small crystallites will be arranged at many angles, we can effectively consider getting all possible angles at once. This means that there will be some crystallites which always satisfy the Bragg conditions, *if λ makes this possible*.

There is some rotational symmetry, so we get rings produced, and if we use electrons and x-rays on the same target, we can see we get the same pattern.

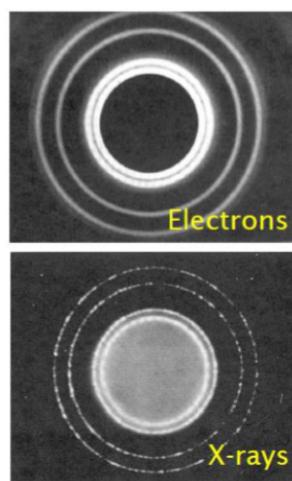


Figure 44.4

This is a little bit difficult to think about, as what does the 'amplitude' of a matter wave actually mean? This is easy to define with EM waves, in terms of the EM field, but not for matter...

4 Back to Bohr

Looking back to the Bohr model, we want to get:

$$E_n = \frac{-13.6eV}{n^2} \quad (44.2)$$

As a starting point, the electron must 'fit' into a circular orbit. Electrons are in a potential well and are bound to a positively charged nucleus:

$$V(r) = \frac{-e}{4\pi\epsilon_0 r}$$

$$u(r) = \frac{-e^2}{4\pi\epsilon_0 r}$$

The electron is in the n th orbital, with a radius r_n , velocity v_n . Plotting wave function amplitude against circumference:

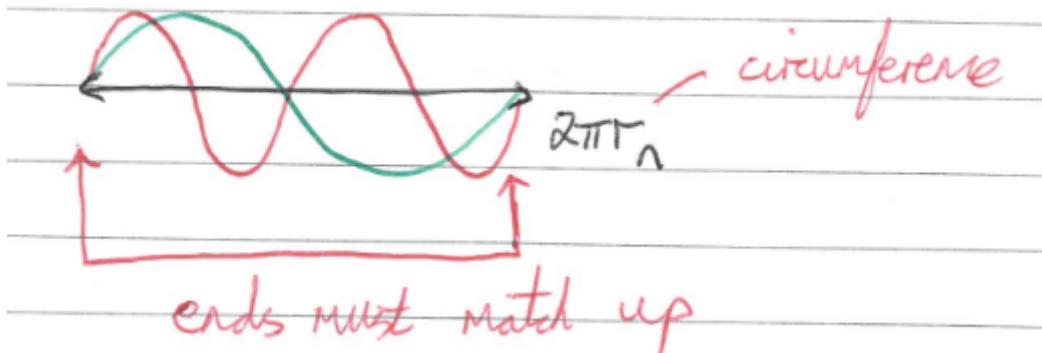


Figure 44.5

We note that this must be continuous, so the ends must join up:

$$2\pi r_n = n\lambda_n = n\frac{h}{p_n} = n\frac{h}{mv_n} \quad (44.3)$$

Also (this is non-relativistic so we can use circular motion)

$$F = ma = \frac{mv^2}{r} \quad (44.4)$$

$$\frac{e^2}{4\pi\epsilon_0 r_n} = mv_n^2 \quad (44.5)$$

We can use 44.3 and 44.5 to eliminate v_n and show:

$$r_n = n^2 a_0, \quad a_0 = \frac{\epsilon_0 h^2}{\pi m e^2} = 5.2918 \times 10^{-11} \text{ ("Bohr Radius")}$$

4.1 Energies

Using 44.3:

$$2\pi r_n = n\lambda_n = n\frac{h}{p_n} = n\frac{h}{mv_n}$$

$$p_n = \frac{nh}{2\pi r_n}$$

And using our new result:

$$p_n = \frac{nh}{2\pi r_n} = \frac{h}{2\pi n a_0}$$

Kinetic Energy

$$T_n = \frac{p_n^2}{2m} = \frac{h^2}{8\pi^2 mn^2 a_0^2} = \frac{me^4}{8\epsilon_0^2 h^2 n^2} = +\frac{13.6}{n^2}$$

Potential Energy

$$U_n = \frac{-e^2}{4\pi\epsilon_0 r_n} = \frac{-e^2}{4\pi\epsilon_0 n^2 a_0} = \frac{-me^4}{4\pi\epsilon_0^2 h^2 n^2} = -\frac{27.2}{n^2} eV$$

Total Energy

$$E_n = T_n + u_n = \frac{-13.6}{n^2} eV$$

This is the Bohr Model result, as required.

5 Conclusions

- Matter is wave-like, with a wavelength given by de Broglie wavelength.
- Electrons can therefore interfere, which we see with electron diffraction.
- A wave-like electron gives us the observed Bohr model result.

Fri 21 Nov 2025 12:00

Lecture 8 - Wave-Particle Duality

In this lecture:

- Particles acting as waves and vice-versa - depends on how we look at/measure them.
- Infinite potential wells - quantisation of energy in bound states, zero-point energy.
- The Heisenberg Uncertainty Principle.

1 Single Slit (Fraunhofer) Diffraction

We have seen particles behaving like waves (i.e. electron diffraction) and waves behaving like particles (i.e. the photoelectric effect). This seems to depend on how we look at them - different experiments give different results.

Broadly, they seem to be wave-y when in motion (i.e. interference), and particulate when detected. In a single-slit experiment, we can see both at the same time. We fire electrons through a single slit, and measure the diffraction pattern with a screen or movable detector:

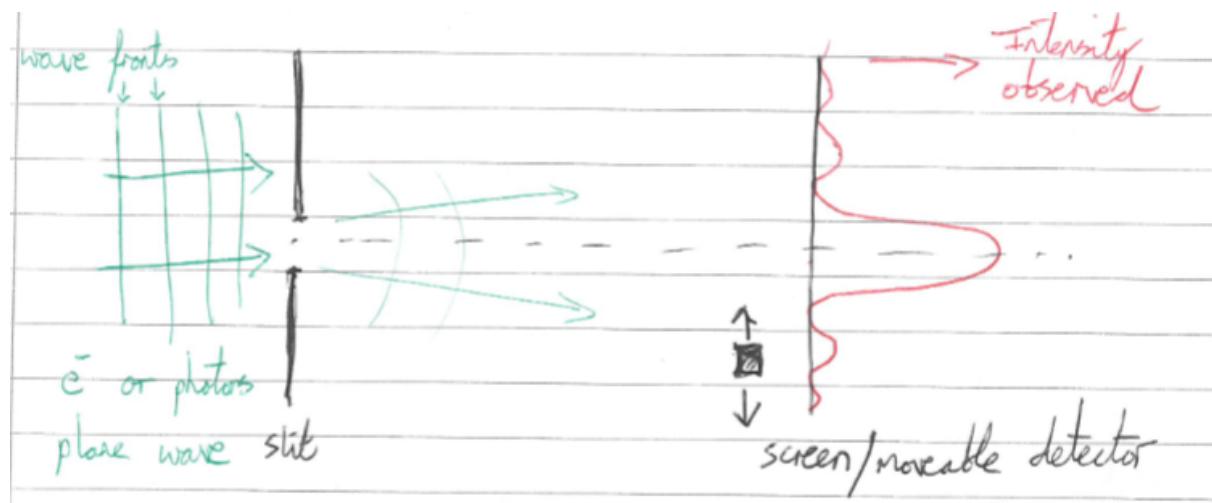


Figure 45.1

We can do this experiment with either photons, or with electrons and we observe the same result: both yield the same interference pattern due to superposition of waves cancelling each other out.

What if we turned the source way down and fired a single photon/electron at a time. We would therefore detect individual particles on our detector as they arrive one-by-one. Surely, therefore, we wouldn't get a diffraction pattern? We do, however, get a pattern... What are they interfering then, if there's no other particles to interfere with? It turns out the particle interferes with itself.

We view the intensity profile as a probability distribution of any one particle impacting here. The areas of the profile with a higher probability have a higher measured intensity, to be expanded on next lecture. The amplitude of a particle wave at a point represents the probability of the particle being found at that point.

2 Infinite Potential Wells

Given a particle in an infinite potential well, how is this impacted by wave-like behaviour of the particle?

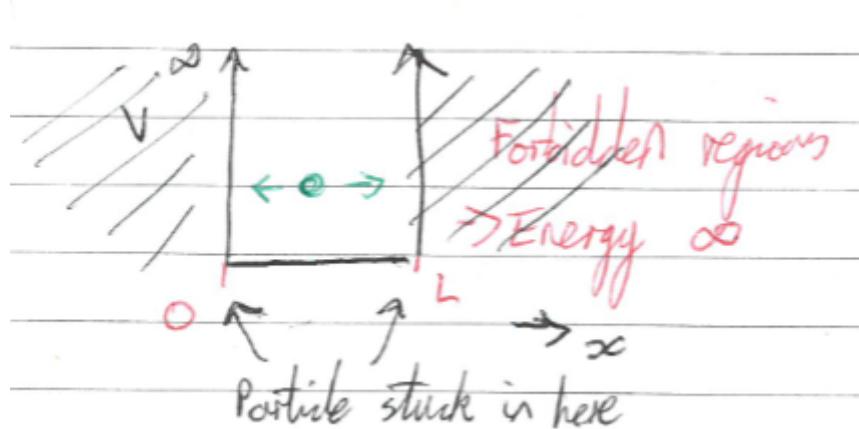


Figure 45.2

Classically, we can think of this as a squash court with a particle bouncing back and forth between the two left and right boundaries. The particle is stuck between zero and L , with potential energy measured on the y-axis. Note that since the y-axis is energy - it is not spatial. The only possible spatial movement of the particle is left-to-right.

Quantum mechanically, we have the following rules:

- The particle has finite energy. It therefore only has amplitude (probability of finding the particle) at this one spot.
- The wave function of a particle has to be continuous throughout. Discontinuities break things.

As a result of these two rules, the wavefunction $\psi(x)$ must be zero at the boundary points 0 and L . We also assume that we can define any wavefunction as a sine wave:

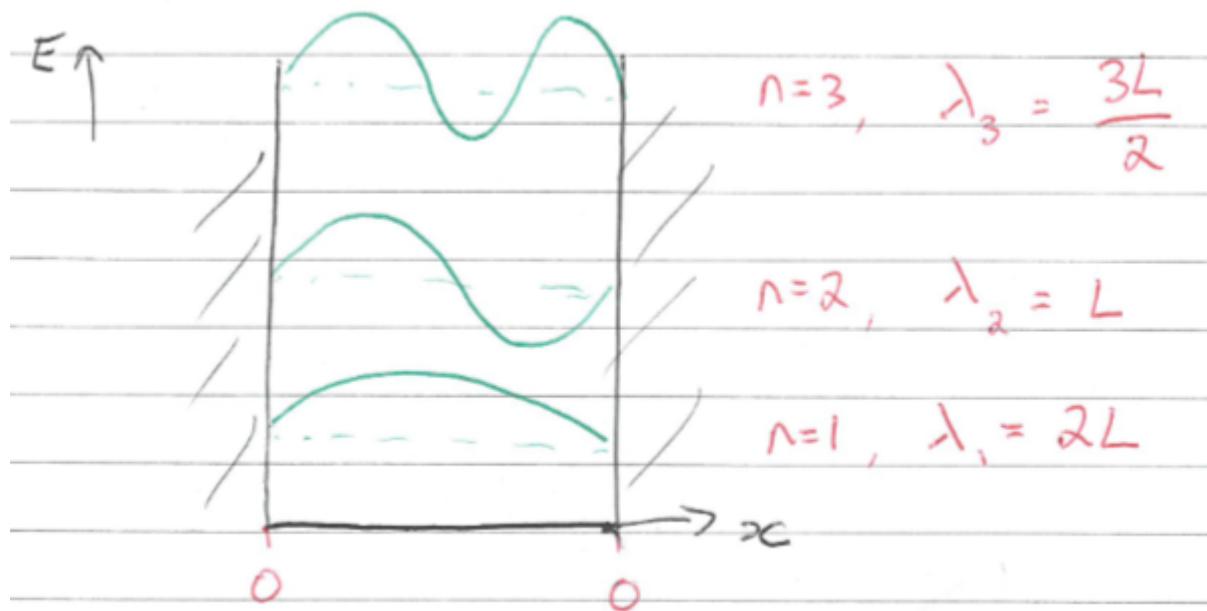
$$\psi(x) = A \sin \frac{n\pi x}{L}, \quad n = 1, 2, 3, \dots$$

Therefore:

$$\lambda_n = \frac{2L}{n}$$

We can generate more complicated shapes using Fourier Decomposition, where we define a more complex wave function as the sum of (potentially infinitely) many sine waves.

Considering this wave function at multiple different energies (infinitely many as we're in an infinite well):

Figure 45.3: Note: Bigger n is higher energy.

We know from last lecture that we can relate wavelength and momentum, so:

$$p = \frac{h}{\lambda} \implies p_n = \frac{n\hbar}{2L}$$

$$E = \frac{p^2}{2m} = \frac{n^2\hbar^2}{8mL^2}$$

Note that $E \propto n$ in an infinitely deep 1D potential well. Compare this to $E \propto \frac{1}{n^2}$ in a H atom.

Energy is uniquely determined for each quantum state $\psi_n(x)$ for the particle in the well. If we consider momentum: classically we say the particle is moving either to the left or or the right, with a known magnitude but an unknown direction (if we take a single observation as a snapshot in time).

For $n=1$:

$$p_x = \pm \frac{\hbar}{2L}$$

Which is effectively zero momentum but with an uncertainty of $\Delta p_x = \frac{\hbar}{2L}$. The expectation value (average value):

$$\langle p_x \rangle = 0$$

And for x (somewhere in the well):

$$\langle x \rangle = \frac{1}{2}L$$

With an uncertainty of:

$$\Delta x = \pm \frac{1}{2}L$$

3 Shrinking the Well

What happens if we shrink the well? Both momentum and energy are inversely proportional to cavity length:

$$p \propto \frac{1}{L}$$

$$E \propto \frac{1}{L^2}$$

Therefore they both go up. This is weird in classical mechanics. We now know less about momentum (higher uncertainty) but more about the particle's location (smaller set of possible values, lower uncertainty than before). This isn't a proof, but a hint towards the Heisenberg Uncertainty Principle.

4 Heisenberg Uncertainty Principle

$$\Delta x \Delta p_x \geq \frac{\hbar}{2} \quad (45.1)$$

Where: $\hbar = \frac{h}{2\pi}$. In QM, there is a limit to how well we can know two conjugate (effectively paired, formal word to mean a pair of variables where knowing one of them perfectly precludes knowing the other) observables simultaneously. This also applies to energy w.r.t. time:

$$\Delta E \Delta t \geq \frac{\hbar}{2}$$

Note that in a 2D system, $\Delta p_x \Delta y$ is unrestricted as they are independent and non-conjugate principles. In our example:

$$\Delta x \Delta p_x = \frac{L}{2} \frac{h}{2L} = \frac{h}{4} \geq \frac{\hbar}{4\pi}$$

The single slit diffraction is a good example of this. If we measure the x of a photon by passing it through a slit, we lose information on its momentum and it spreads out along x .

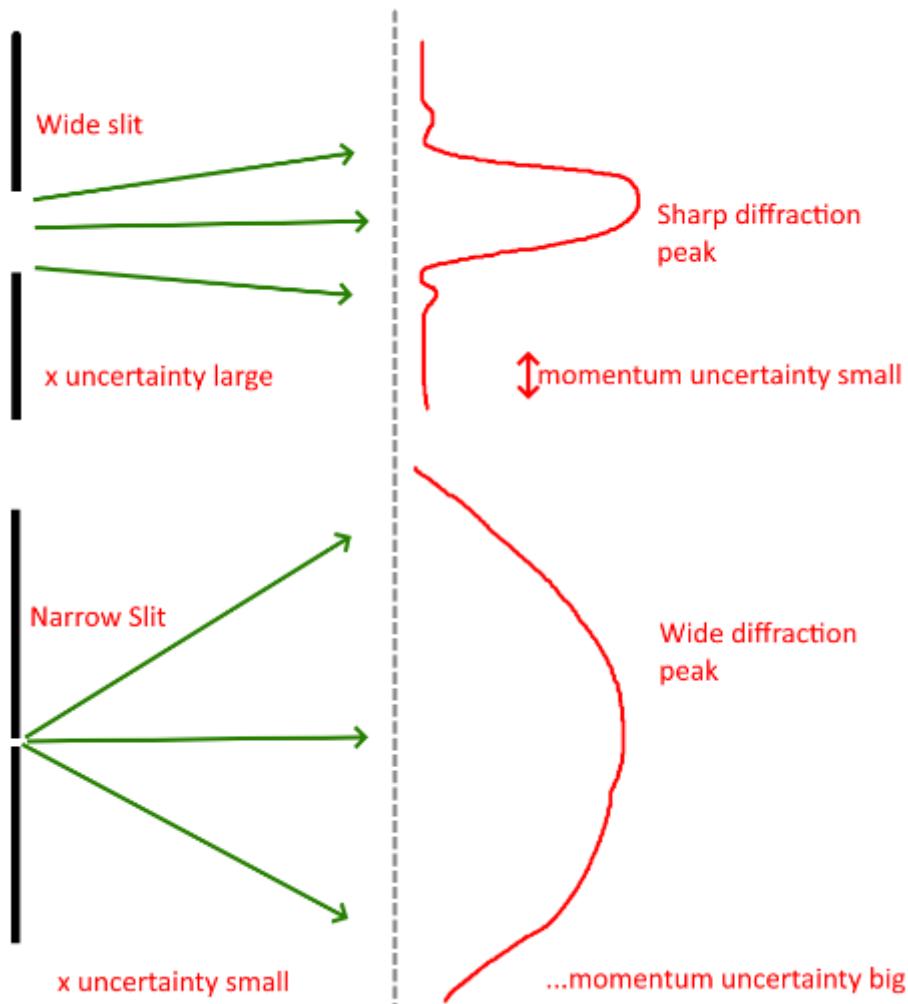


Figure 45.4

5 Conclusions

- Wave-particle duality allows particles to both diffract as if waves and to be quantised like particles.

- Wave-like particles explains why the energy in a bound system has to be quantised (as shown before) due to standing waves in a potential well.
- Energy levels depend on the shape of the potential well.
 - $E \propto n^2$ in an infinitely deep 1D well.
 - $\frac{1}{n^2}$ in a H atom.
- We cannot know a particle momentum and position simultaneously to perfect precision, we are limited that if we know one very very accurately, we cannot know the other very well.