L3 Astro - Section 1

Nick Kaiser

Contents

List of Figures

1 Electromagnetic Radiation

1.1 Energy density u of EM field

Consider the work done pulling two capacitor plates – with separation along the x-axis – apart as illustrated on the left side of in figure 1. Assuming, for simplicity, a separation much less than the size of the plates, the field between the plates will be $\mathbf{E} = \hat{\mathbf{x}} E_x$ with, by Gauss's law, $E_x = \Sigma/\epsilon_0$, where Σ is the charge density. The force is $F = AE_x\Sigma/2$ with A the area of the plates – the factor $1/2$ coming from the fact that the field ramps from zero to the inter-plate value passing through the plate so the mean field is $E_x/2$ – and so the work done in increasing the separation by δx is $\delta W = AF\delta x = \epsilon_0 A \delta x E_x^2/2 = \epsilon_0 \delta V E_x^2/2$. Equating this to the change in the volume times the energy density u of the field gives $u = \epsilon_0 |\mathbf{E}|^2/2$.

It is crucial to recognise that we are here attributing the energy entirely to the field. In this process there was no change in the mechanical energy of the plates as they started and ended at rest. If we release the plates, they will gain mechanical energy at the expense of the field and field plus mechanical energy will be conserved. It is somewhat arbitrary how we assign the energy; whether we say it resides in the field or whether it resides in the particles.

Similar considerations can be applied to the magnetic field. Consider a long solenoid of length L with radius r and with N turns as illustrated on the right side of in figure 1. Ampère's law says the field is $B = \mu_0 NJ/L$, where J is the current. If we ramp the current up to increase the field there will be an induced EMF (according to Faraday's law) such that $E \times 2\pi r = AdB/dt$ where $A = \pi r^2$ is the area. The power required to drive the current J against the induced electric field is $dW/dt = EJ \times 2\pi rN = \mu_0^{-1} AL \times BdB/dt$. It follows that the work done is $dW = \mu_0^{-1} V dB^2/2$, with $V = AL$ the volume, so, attributing this to the the energy density of the magnetic field gives $u = \mu_0^{-1}|\mathbf{B}|^2/2$. The sum of the electric and magnetic field densities is therefore

$$\boxed{u = (\epsilon_0|\mathbf{E}|^2 + \mu_0^{-1}|\mathbf{B}|^2)/2} \tag{1}$$

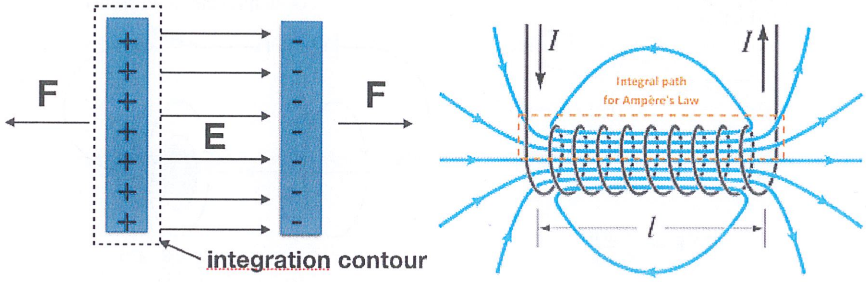


Figure 1: The energy density of an electric field can be determined by considering the work done when pulling two capacitor plates apart (see text for details) which, in the process, creates new field. The work done is proportional to the field times the charge density, but the field and charge density are linearly related – by Gauss's law – so the work done is proportional to the volume increase times the square of the field strength. The energy density of a magnetic field can be obtained by considering the work done in increasing the current flowing in a solenoid; here the increase in current causes an increase in the **B**-field, and hence in the magnetic flux through the solenoid. This induces an EMF opposing the change and the energy going into the field is the work we need to do to keep the current flowing in opposition to the induced **E**-field.

1.2 Maxwell's equations

E&M pre-Maxwell:

$$\begin{aligned} &\text{Gauss}: \int \mathbf{dA}\cdot\mathbf{E} = Q/\epsilon_0 \Leftarrow \text{charge} && \int \mathbf{dA}\cdot\mathbf{B} = 0 \Leftarrow \text{no monopoles} \\ &\text{Faraday}: \oint \mathbf{dl}\cdot\mathbf{E} = -\int \mathbf{dA}\cdot\dot{\mathbf{B}} && \text{Ampere}: \oint \mathbf{dl}\cdot\mathbf{B} = \mu_0 \int \mathbf{dA}\cdot\mathbf{j} \end{aligned} \tag{2}$$

Maxwell:

- Added 'displacement current' $\mu_0\mathbf{j} \Rightarrow \mu_0\mathbf{j} + c^{-2}\partial\mathbf{E}/\partial t$. Required for consistency (see figure 2).
- Differential formulation. Modern form of equations provided by Heaviside.

The physical content of Maxwell's equations is illustrated in figure 2. In the 'rationalised' SI system they are

$$\begin{aligned} &\nabla\cdot\mathbf{E} = \rho/\epsilon_0 && \nabla\cdot\mathbf{B} = 0 \\ &\nabla\times\mathbf{E} = -\dot{\mathbf{B}} && \nabla\times\mathbf{B} = \mu_0(\mathbf{j} + \dot{\mathbf{E}}/\epsilon_0) \end{aligned} \tag{3}$$

where

- $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic fields
- $\dot{\mathbf{E}}$ and $\dot{\mathbf{B}}$ are their partial derivatives with respect to time t
- ϵ_0 is the *permittivity of free space*
  - defined in terms of the electrostatic force between two charges
- μ_0 is the *permeability of free space*

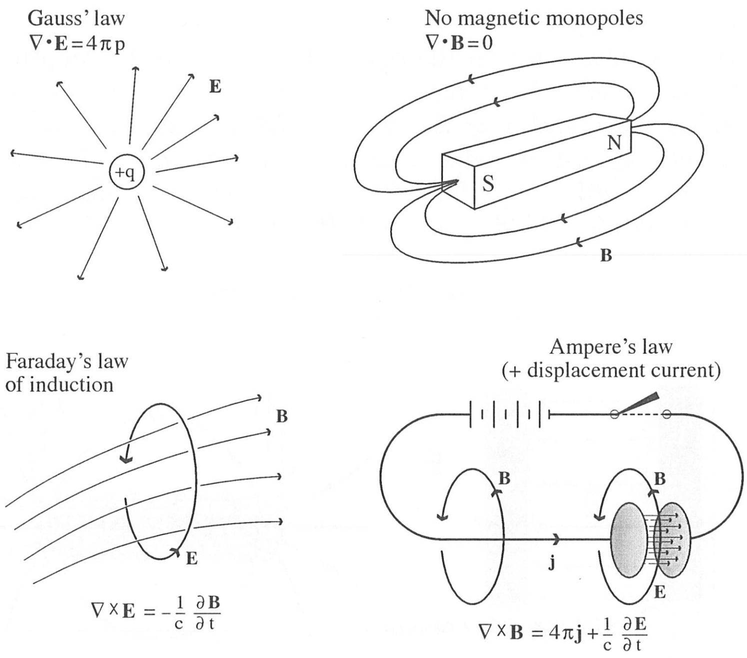


Figure 2: Pictorial representation of Maxwell's equations. The form of the equations here assume Gaussian units. We will use mostly SI units, for which Gauss's law is $\boldsymbol{\nabla} \cdot \mathbf{E} = \rho/\epsilon_0$ and the Ampère-Maxwell equation is $\boldsymbol{\nabla} \times \mathbf{B} = \mu_0 \mathbf{j} + c^{-2}\partial_t \mathbf{E}$. However, we will often assume that space and time units are such that the numerical value of the speed of light c is unity. There is a useful distinction to be drawn between the 'sourced' equations containing the charge and current (Gauss and Ampère/Maxwell) and the 'un-sourced' equations in the upper right and lower left.

  - defined in terms of the magnetostatic force between two current element (Biot & Savart)

- $c = 1/\sqrt{\epsilon_0 \mu_0}$
- ρ is the charge density
- $\mathbf{j}$ is the current density

Q: Maxwell's equations (mid 19th century) are a relativistically covariant and causal (no action at a distance) field theory for electromagnetism. They also embody 'gauge-invariance'. Why didn't Maxwell apply this approach to gravity? Hint: look for his Encyclopaedia Brittanica article 'Attraction'.

Units and systems for electromagnetism: Here we will (try to) use SI (kg, m, s) units. But cgs units are more common in astrophysics – in fact nearly ubiquitous. There are also different systems. The equations in figure 2 are in what's called Gaussian units and are not 'rationalised' as they have 4π in place of μ_0, for instance. The fields $\mathbf{E}$ and $\mathbf{B}$ there have the same units, so the Lorentz force is $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}/c)$ whereas in SI they have different units and $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$. A much nicer system of units is Kirchoff-Lorentz, where the unit of charge is changed so that $\epsilon_0 = 1$. If you also choose your units of length/time so that $c = 1$ then, since $c^2 = 1/\epsilon_0\mu_0$, $\mu_0 = 1$ also. If one is then sloppy one can write all the equations without these constants appearing. That makes the equations look a lot cleaner and perhaps clearer. But it can be confusing.

1.3 Poynting's theorem

Dotting Faraday's law $\dot{\mathbf{B}} = -\boldsymbol{\nabla} \times \mathbf{E}$ with $\mathbf{B}$ and dotting the Ampère-Maxwell equation $\dot{\mathbf{E}}/c^2 = \boldsymbol{\nabla} \times \mathbf{B} - \mu_0 \mathbf{j}$ – which we can also write as $\epsilon_0 \dot{\mathbf{E}} = \mu_0^{-1} \boldsymbol{\nabla} \times \mathbf{B} - \mathbf{j}$, since $c^2 = 1/(\epsilon_0 \mu_0)$ – with $\mathbf{E}$ and adding gives

$$\partial_t(\epsilon_0 |\mathbf{E}|^2 + \mu_0^{-1}|\mathbf{B}|^2)/2 = \mu_0^{-1}[\mathbf{E} \cdot (\boldsymbol{\nabla} \times \mathbf{B}) - \mathbf{B} \cdot (\boldsymbol{\nabla} \times \mathbf{E})] - \mathbf{j} \cdot \mathbf{E} \tag{4}$$

Using the identity $\boldsymbol{\nabla} \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\boldsymbol{\nabla} \times \mathbf{a}) - \mathbf{a} \cdot (\boldsymbol{\nabla} \times \mathbf{b})$ gives *Poynting's theorem:*

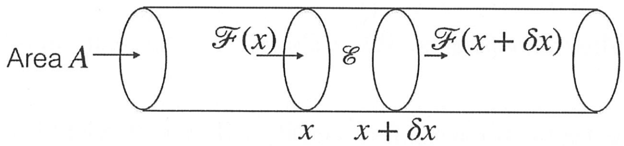
$$\boxed{\partial_t u + \boldsymbol{\nabla} \cdot \mathbf{S} = -\mathbf{j} \cdot \mathbf{E}} \tag{5}$$

where the *Poynting vector* (or *Poynting flux*) is defined as

$$\boxed{\mathbf{S} \equiv \mu_0^{-1}\mathbf{E} \times \mathbf{B}} \tag{6}$$

The right hand side of (5) is minus the rate at which the particles are gaining energy from the field. So Poynting's theorem expresses conservation of total energy in which we identify the Poynting flux with the energy flux density.

The energy continuity (or conservation) law



- The energy in the volume element $\delta V = A\delta x$ is $\delta E = \mathscr{E}\delta V$
- the change in δE in time Δt is energy in minus energy out, or
  - $\Delta\delta E = A\Delta t(\mathscr{F}(x) - \mathscr{F}(x+\delta x))$ or $\Delta\delta E = -A\Delta t \times \delta\mathscr{F}$
- where $\mathscr{F}$ is the energy flux density = energy per area per time
- but $\Delta\delta E = \delta V\Delta t\partial\mathscr{E}/\partial t = A\Delta t \times \delta x\partial\mathscr{E}/\partial t$
  - $\partial\mathscr{E}/\partial t$ being the rate of change of energy density at that x
- so $\delta x\partial\mathscr{E}/\partial t = -\delta\mathscr{F}$ or, taking the limit, $\partial\mathscr{E}/\partial t = -\partial\mathscr{F}/\partial x$

Figure 3: Illustration of the energy continuity equation (in 1D). The rate of change of the energy density is the gradient the energy flux density (energy per unit area per unit time). The generalisation of this to 3D, and with $\mathcal{E} \to u$, the electromagnetic field energy density, and $\mathcal{F} \to \mathbf{S}$, would be $\partial_t u + \boldsymbol{\nabla}\cdot\mathbf{S} = 0$: this would express conservation of energy and is valid in the absence of charges. Poynting's theorem is that the right hand side is non-zero; it is the (minus) the energy being given to the charges by the electromagnetic field.

1.4 Planar electromagnetic waves in vacuum

- Waves in vacuum:
- Guess $\begin{array}{l}\mathbf{E} = \hat{\mathbf{x}}E_0\cos(\omega t - kz)\\ \mathbf{B} = \hat{\mathbf{y}}B_0\cos(\omega t - kz)\end{array}$
- $\oint \mathbf{dl}\cdot\mathbf{B} = -A\dot{E} \to LdB = -Ldz\dot{E}$
- $\oint \mathbf{dl}\cdot\mathbf{E} = -A\dot{B} \to LdE = -Ldz\dot{B}$
- so $\dot{E} = -B'$ and $\dot{B} = -E'$
- so $\omega E_0 = -ckB_0$ and $\omega B_0 = -ckE_0$
- and the dispersion relation is $\omega^2 = c^2k^2$

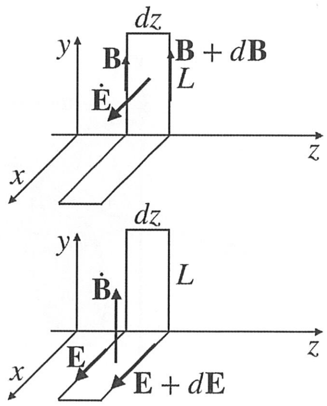


Figure 4: Linearly polarized plane EM waves from the Faraday and Ampère/Maxwell laws.

- Faraday: $\oint \mathbf{dl}\cdot\mathbf{E} = -\int \mathbf{dA}\cdot\dot{\mathbf{B}}$ Ampère/Maxwell: $c^2\oint \mathbf{dl}\cdot\mathbf{B} = \int \mathbf{dA}\cdot\dot{\mathbf{E}}$
- A solution: $\mathbf{E}(z,t) = \hat{\mathbf{x}}E_0\cos(kz - \omega t)$ $\mathbf{B}(z,t) = \hat{\mathbf{y}}B_0\cos(kz - \omega t)$
  - plane traveling waves moving at speed $v = \omega/k$
  - $\mathbf{E}$, $\mathbf{B}$ & $\mathbf{z}$ mutually perpendicular.
- Proof: Consider little area $\delta\mathbf{A} = l\hat{\mathbf{y}} \times \delta\mathbf{z}$ (parallel to $\hat{\mathbf{x}}$ and therefore to $\mathbf{E}$)
- Loop integral round boundary is
  - $\oint \mathbf{dl}\cdot\mathbf{B} = l \times \hat{\mathbf{y}}\cdot(\mathbf{B}(z+\delta z,t) - \mathbf{B}(z,t)) = lB_0k\delta z\cos(kz - \omega t)$

– while $\int \mathbf{dA} \cdot \dot{\mathbf{E}}/c^2 = c^{-2} l E_0 \omega \delta z \cos(kz - \omega t)$

– these are identical if $B_0 k = E_0 \omega$

- Similarly for $\oint \mathbf{dl} \cdot \mathbf{E}$ and $\int \mathbf{dA} \cdot \dot{\mathbf{B}}$ (using $\delta\mathbf{A} = l\hat{\mathbf{x}} \times \delta\mathbf{z}$)

  – identical if $E_0 ck = B_0 \omega$

- So $\mathbf{E}(z,t)$, $\mathbf{B}(z,t)$ above are a solution in vacuum if $E_0 = B_0$ and $v = \omega/k = c$.

  – these are 'non-dispersive' waves; the speed is independent of frequency

Q: We have only used Faraday's law and the Ampère/Maxwell equation. What about the other equations? Do they tell us anything?

Q: How does $u = (\epsilon_0 |\mathbf{E}|^2 + \mu_0^{-1} |\mathbf{B}|^2)/2$ depend on z? Is there an energy flux density? How is it related to the (mean) energy density?

1.5 Polarization

- Maxwell's equations (or Faraday + Ampère/Maxwell) are linear in $\mathbf{E}$, $\mathbf{B}$ - so we can add and subtract solutions to make new solutions.

- We can rotate $\mathbf{E}$ and $\mathbf{B}$ (together) around the $\mathbf{z}$-axis.

  – These are *linearly polarized planar waves*.

- We can also make a '90-degree lagged' and rotated – so the $\mathbf{E}$ field points along the y-axis – version with $\omega t \rightarrow \omega t - \pi/2$ and add it to the original.

  – Now $\mathbf{E}$, $\mathbf{B}$ (at fixed $\mathbf{z}$) rotate – in the plane perpendicular to $\mathbf{z}$ – with time.

  – These are *circularly polarized waves.*

  – The sign of the lag determines the *helicity*.

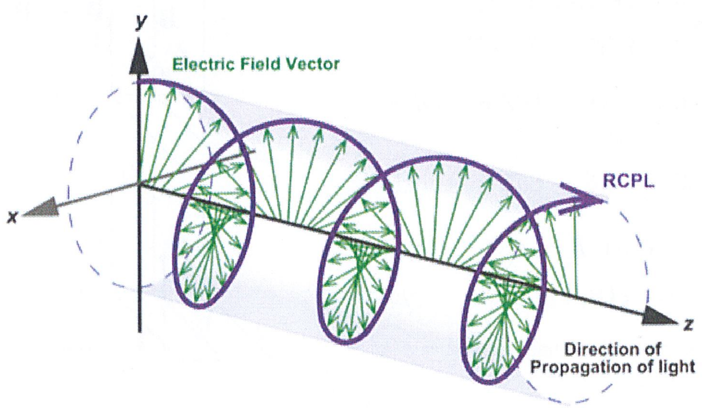


Figure 5: A circularly polarized wave is the superposition of two linearly polarised waves; one being obtained from the other by rotating the plane of polarisation by 90 degrees and by 'lagging' by 90 degrees in phase. The electric and magnetic fields at a fixed position both rotate with time. The magnetic field is always perpendicular to the electric field.

Q: Can one linearly combine such waves to make a beam of which *isn't* polarised? I.e. – 'natural' – radiation?

1.6 Linear and angular momentum of a classical EM beam

Q: How can we understand physically the momentum of e.g. a classical beam of radiation?
A: Beautiful problem in Rybicki and Lightman (see TD1 and figure 6):

- Shine a linearly polarized beam of light on a cell containing a charged ball in a viscous fluid.

- Oscillating **E** field will cause the charge to oscillate up and down
  - charge moves at 'terminal velocity' $\mathbf{v}(t) \propto \mathbf{E}(t)$ if 'viscous damping time' $\ll$ period of radiation
  - magnetic field has relatively small effect on motion of charge if $|\mathbf{v}| \ll c$
- **E** field does work on the charge (heats the fluid)
- The 'Lorentz force' $\mathbf{F} = q\mathbf{B} \times \mathbf{v}$ on the charge points along beam direction $\mathbf{k}$.
  - can show that if the beam transfers energy dW the momentum transferred is $dp = dW/c$
  - just as if the cell has absorbed $N = dW/\hbar\omega$ photons
    - with each photon having $E = \hbar\omega$ and momentum $p = \hbar k = E/c$
    - and therefore satisfying $E^2 = p^2c^2 + m^2c^4$ with $m = 0$.
  - and it follows from this that the energy flux density (Poynting vector) $P = \mathbf{E} \times \mathbf{B}/\mu_0$ is equal to the momentum density.
    - something that can be inferred using the Maxwell stress tensor for EM
- A similar calculation shows that a circularly polarised wave drives the charge around in a circular path
  - so there is a torque on the cell which transfers angular momentum
  - just as for absorption of photons with spin angular momentum $\pm\hbar$

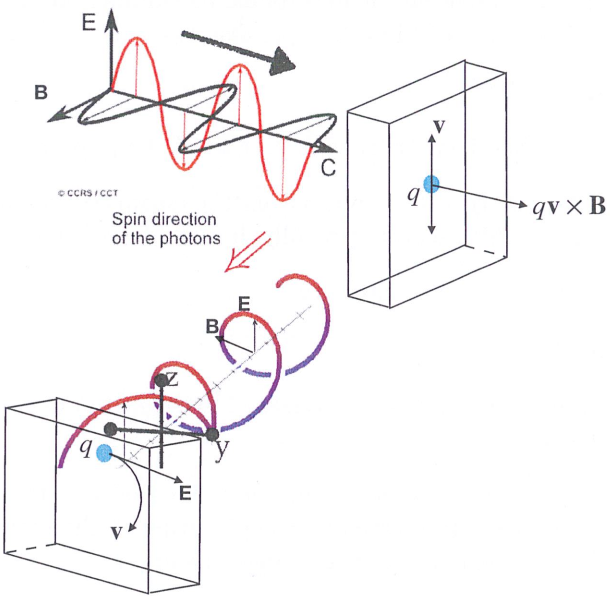


Figure 6: Linear and angular momentum of EM waves (from Rybicki & Lightman). In the upper part of the figure is shown a linearly polarized wave incident on a charge in a cell containing a viscous fluid. The main force on the charge – if it is slowly moving, as we shall assume – is the electric force. The charge moves up and down at the instantaneous 'terminal velocity', which is proportional to the **E** field. The viscous drag force will do work on (and heat) the fluid. But there will also be a small $q\mathbf{v} \times \mathbf{B}$ force that will point in the direction of the wave. This will give momentum to the charge & cell system. In the TD you will show that for each $\hbar\omega$ of energy deposited there is $\hbar\omega/c$ momentum transferred. In the lower part of the figure is shown a circularly polarized wave impinging on the same cell. Now the charge is driven in a circular path (as the **E** field rotates. There is again work done on the fluid, but now the viscous force exerts a torque on the cell which transfers angular momentum. In the TD you will show that for each $\hbar\omega$ of energy deposited an amount $\hbar$ of angular momentum is transferred.

Q: A puzzle (from Martin Harwitt): The result for linear polarisation could be understood in terms of the *Poynting flux* $\mathbf{E} \times \mathbf{B}/\mu_0$ which is the flux density of energy. This, it turns out, is also quite generally, equal to the the momentum density and, for a travelling wave, that is just c times the momentum flux density).

$\mathbf{E}$ and $\mathbf{B}$ are both perpendicular to $\mathbf{k}$, so this points along the wave vector $\mathbf{k}$-direction. But what about the angular momentum for a circularly polarized beam? The momentum density is is still pointing along the direction of the beam? So there doesn't obviously appear to be any helicity of the momentum. Yet the problem above shows convincingly that, if we put an absorber in the path of the beam, it will gain angular momentum. What gives? Hint: You might want to think about a somewhat analogous problem of how a windmill can gain momentum from a smoothly flowing wind.

2 Telescopes and Detectors

2.1 Optics - the 'point-spread function' (PSF) and image formation

2.1.1 PSF from Kirchoff-Fresnel's diffraction theory

- ideally we'd solve Maxwell's equations with boundary conditions for e.g. mirrors – too complicated
- instead we consider the (complex) field amplitude on a 'pupil plane' at the entrance to the telescope – constant for a distant point source
  - We are using here the *para-axial approximation* that the $\mathbf{E}$-field at the entrance is
    - $\mathbf{E}(x, y, z, t) = \mathrm{Re}(\mathbf{E}(x, y)e^{i(kz-\omega t)})$

    which is valid for radiation from a smallish source passed through a narrowish band filter with the complex $\mathbf{E}(x, y)$ encoding the amplitude and the phase.
- and compute the amplitude at a point on the focal plane by summing over elements on the pupil plane with factor $e^{2\pi i L/\lambda}$ where L is the 'optical path' (true distance times $n \leftarrow$ refractive index). This gives
  - $E(\mathbf{r}) \propto \int d^2q\ E(\mathbf{q})e^{i2\pi\mathbf{q}\cdot\mathbf{r}/L\lambda}$
- so the field amplitude $E(\mathbf{r})$ is the Fourier transform (FT) of the pupil plane amplitude $E(\mathbf{q})$
  - if there is no atmosphere $E(\mathbf{q})$ is constant within the entrance aperture so we get the FT of the aperture function, whose width scales, generally, inversely with the size of the telescope aperture
- then we square the amplitude to get the intensity
- this is illustrated in figure 7
- and sum the result (the *monochromatic* PSF), weighted by spectrum of source, to get the PSF appropriate for *'broad-band'* imaging.
- So a telescope can be considered to be an analogue computer that calculates the square of the Fourier transform of the input wave-front. Interestingly, the evolution of the complex electric field field amplitude as one moves down the beam obeys Schrödinger's equation.
  - it is obtained by convolving the input field with a kernel that is the same as the kernel that Feynman obtained as the Green's function for the S-equation.
  - so we can also think of a telescope as being an analogue computer that solves the S-equation.

2.1.2 Diffraction limited point source imaging

- The width Δr of the PSF for *'diffraction limited'* imaging (i.e. where the PSF width is determined solely by the telescope aperture diameter D) is such that for points within $|\mathbf{r}| \sim \Delta r$ the phase-error to all pupil plane points (i.e. $|\mathbf{q}| \lesssim D$) is $\delta\phi \lesssim 1$ radian.
- This gives $\Delta r \sim \lambda D/L$, corresponding to an apparent angular size $\Delta\theta = \Delta r/L \sim \lambda/D$. This is the *'resolving power'* of a diffraction limited telescope.
- this is relevant for telescopes in space and for small terrestrial telescopes (see TD)

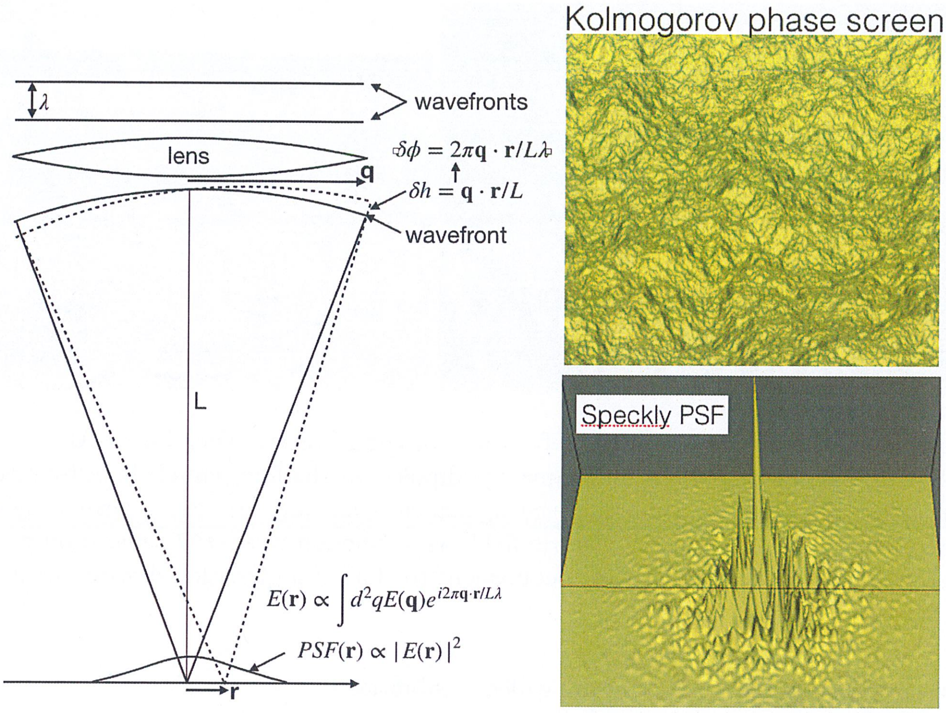


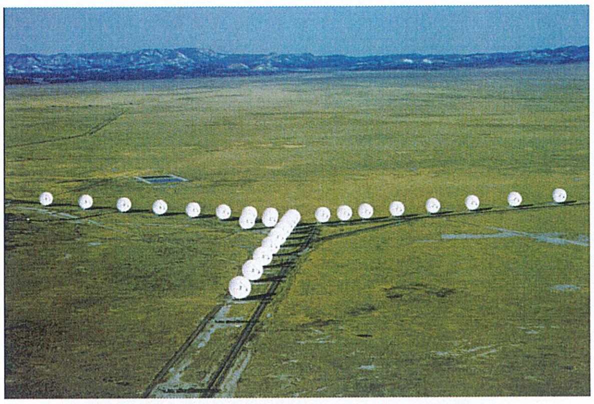
Figure 7: Left: Point spread function $PSF(\mathbf{r})$ for a telescope as calculated using Fresnel diffraction theory. We express the incoming electric field from a distant on-axis source as the real part of a complex *field amplitude* $E(\mathbf{q})$ times $\exp i\omega t$. The electric field amplitude $E(\mathbf{r})$ at some position $\mathbf{r}$ on the *focal plane* is given as an integral over the 'pupil plane' of $E(\mathbf{q})$ times $\exp i\delta\phi$ where $\delta\phi(\mathbf{q};\mathbf{r})$ is the *'phase-error'*. This is found by simple geometry. The on-axis point $\mathbf{r} = 0$ is equidistant from, and therefore in phase with, all points on the solid curved wavefront. The off-axis point $\mathbf{r}$ is equidistant from all points on the dashed wavefront, which is tilted with respect to the solid wavefront surface by an angle $\theta = r/L$, where L is the focal length. It follows that the distance between the two surfaces is $\delta h = \mathbf{q}\cdot\mathbf{r}/L$ and so $\delta\phi(\mathbf{q};\mathbf{r}) = 2\pi\delta h/\lambda = 2\pi\mathbf{q}\cdot\mathbf{r}/L\lambda$. The PSF is the energy density which is proportional to the squared modulus of $E(\mathbf{r})$. Right panel shows the form of the incoming wavefront (top) from a distant point source viewed through the atmosphere and (bottom) the form of the resulting *'speckly'* PSF. See §2.1.6 below.

2.1.3 Diffraction limited extended source imaging

- *Extended sources*: The observed image is true 'scene' convolved with the PSF.

2.1.4 Radio interferometry and aperture synthesis

- the image $|E(\mathbf{r})|^2 \propto |\int d^2q E(\mathbf{q}) e^{i2\pi\mathbf{q}\cdot\mathbf{r}/L\lambda}|^2$ on the focal plane of a conventional telescope is the square of the Fourier transform of the complex electric field amplitude on the pupil plane.
- so $|E(\mathbf{r})|^2 \propto \int d^2q \int d^2q' E(\mathbf{q}) E^\star(\mathbf{q}') e^{i2\pi(\mathbf{q}-\mathbf{q}')\cdot\mathbf{r}/L\lambda} = \int d^2z C(\mathbf{z}) e^{i2\pi\mathbf{z}\cdot\mathbf{r}/L\lambda}$ where $C(\mathbf{z}) \equiv \int d^2q E(\mathbf{q} + \mathbf{z}) E^\star(\mathbf{q})$ is the auto-correlation function of the incoming electric field.
  - for a long integration $\boxed{C(\mathbf{z}) \rightarrow \langle E(\mathbf{q}+\mathbf{z}) E^\star(\mathbf{q})\rangle}$
  - this 2-point statistic of the complex field amplitude encodes in a very clean way the angular structure of the source
- The image is obtained by taking the Fourier transform of $C(\mathbf{z})$. The information about the small angular scale structure of a source is encoded in the form of $C(\mathbf{z})$ at spatial 'lags' $\mathbf{z} \sim \lambda/\Delta\theta$. Thus, rather than build a giant dish $D > \lambda/\Delta\theta$ capable of resolving the source, an alternative is to observe the source with an array of lower resolution telescopes and form the auto-correlation function of the electric field for a set of 'base-lines' that sample the so-called '$u - v$' plane at the desired scale and then Fourier transform the result to 'synthesize' the image.



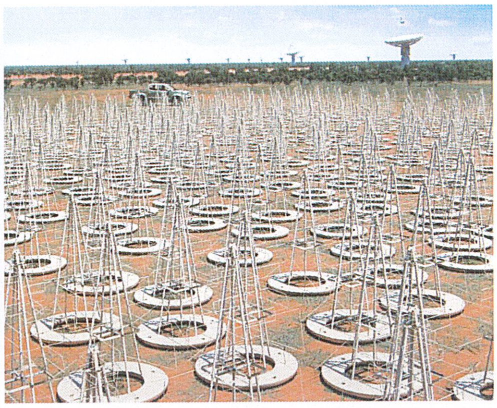


Figure 8: Aperture synthesis arrays: on the left is a photograph of the Very Large Array. On the right is part of the Square Kilometer Array (low frequency 'dipoles' in the foreground; high frequency dishes in the rear). Such telescopes work by combining the signals electronically and cross-correlating to determine $C(\mathbf{z})$, the autocorrelation function of the electric field, as a function of spatial separation $\mathbf{z}$. This gives a resolving power – though not the sensitivity – equivalent to that of a gigantic telescope without having to make massive dishes.

   - Some jargon: samples of $C(u, v)$ are called 'visibilities'.

- an example is the 'very large array' (VLA) which uses a 'Y'-shaped array of dishes. This might seem to give rather poor sampling of the $\mathbf{z} = (u, v)$ plane, but, as the Earth rotates, this pattern sweeps around and the result is good sampling from which high-quality images can be recovered.
- Modern applications of this are LOFAR and the low-frequency segment of the planned 'square kilometer array' (SKA), which do the 'beam-forming' digitally rather than making dishes to focus the radiation on a detector.

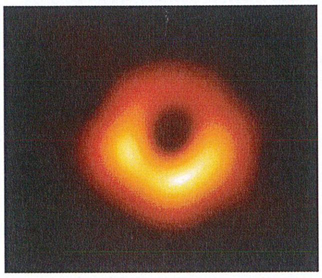


Figure 9: Image of the accretion disk around the black-hole at the centre of the elliptical galaxy M87 in the Virgo cluster (the centre of the 'local supercluster'). This was made using the 'Event Horizon Telescope' (EHT): very long baseline interferometry (VLBI) at mm wavelengths. Maintaining 'phase-stability' over the $\sim 10,000$km baselines at the short wavelengths required to provide the required angular resolution is extremely challenging.

2.1.5 Intensity interferometry

- Aperture synthesis requires that we correlate the field from widely separated antennae. That requires maintaining phase coherence. Hanbury-Brown & Twiss showed that there is an alternative that does not require this, which is called 'intensity interferometry'.
- The principle is illustrated in figure 10 and can be understood as follows. We imagine two elements of the source with separation L_s and a receiver at distance D and we ask: how does the difference in optical path length (real distance divided by wavelength – so phase) change if we move the detector around. For the phase-difference to change by more than about a radian requires that we have to move the detector by a distance $L \sim \Lambda D_s / L_s$.

2.1.6 Atmospheric seeing

- The width of the PSF for large ground based telescopes is much poorer than the diffraction limit.

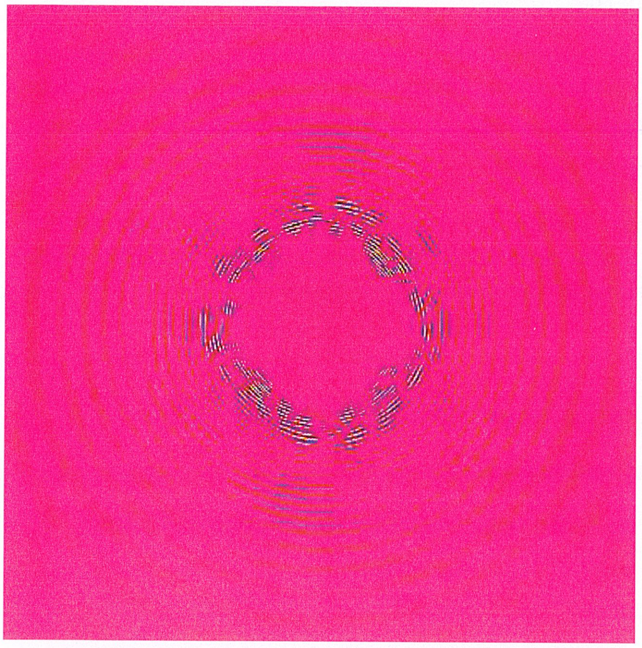


Figure 10: This figure shows waves that have propagated from a localised disturbance. These are actually deep-ocean waves which are dispersive; which is why the long waves have out-run the shorter waves. It shows that the intensity of the radiation is modulated on a scale much larger than the wavelength. This explains why, as surfers know, but physicists often deny, waves in the 'swell' from a distance localised storm come in 'sets'. The same thing happens with EM radiation from astronomical sources (though in that case the waves are not dispersive). The scale L of the modulation is, it turns out, just the size of a telescope that would be able to resolve the source: i.e. $\lambda/L \simeq L_s/D_s \simeq \theta_s$ where L_s and D_s are the source size and distance respectively, so $L \simeq \lambda/\theta_s$. By correlating the *intensity* at different separations it is possible to infer the angular – and with the distance the physical – size of the source. This is the principle of intensity mapping pioneered by Hanbury-Brown and Twiss.

- This is because the wavefronts from distant sources become 'wrinkled' or 'corrugated' in passing through the atmosphere.
- More precisely, one typically finds layers in the atmosphere where there is turbulent mixing of air with different entropies. This involves strongly sub-sonic motions, so the air is effectively *isobaric*, and the entropy fluctuations give rise to density fluctuations and therefore with fluctuations in the refractive index.
- The form of the resulting wavefronts is that of a 2-dimensional *gaussian random field* with a very 'red spectrum' (the perturbation to the tilt of the wavefront averaged over an area of size l scales as $l^{-1/6}$).
- An important characteristic scale is the *'Fried length'*. This is the distance over which the change in the phase-error is on the order of one radian. It depends on wavelength. In the optical, and at very good sites like Chile and Hawaii it is typically $\sim$ 20cm. This severely limits the resolving power of terrestrial optical telescopes.
- the PSF and its statistical properties can be calculated using Fresnel theory
  - the *instantaneous* PSF is a cloud of diffraction-limit sized ($\Delta\theta \sim \lambda/D$) 'speckles' that dance about rapidly as the turbulent layer moves across the 'beam'
  - this is the result of interference between the radiation coming from different 'Fried-length' scale patches on the pupil plane.
  - for long exposures this gets averaged out and the result is a blurred PSF with a width comparable to that of a diffraction limited telescope with aperture diameter $D \sim l_{\rm Fried}$.
  - *adaptive optics* involves taking some of the radiation (with a 'beam splitter') and measuring the wavefront corrugation and then correcting the other radiation with a deformable 'rubber mirror'.
  - very similar effects are seen in radio observations with large arrays like LOFAR, where the phase errors come from variation of the refractive index of the ionosphere.

Q: The corrugations of the wavefronts $\delta h(\mathbf{q})$ produced by the atmosphere are essentially independent of the wavelength – so the *phase* fluctuations *are* wavelength dependent. The 'structure function' $S_{\delta h}(z)$ is the mean squared difference in wavefront deformation $\langle(\delta h_1 - \delta h_2)^2\rangle$ between two points with separation $|\mathbf{q}_1 - \mathbf{q}_2| = z$. According to Kolmogorov turbulence theory $C(z) \propto z^{5/3}$. Use this to show how the 'Fried length' and the resolving power of a telescope scales with the wavelength. This shows the perhaps surprising result that the image quality is *better* at longer wavelengths.

Q: The phase errors $\phi(\mathbf{q})$ from entropy fluctuations in the turbulent atmosphere have a 'structure function' $\langle(\phi(\mathbf{q}'+\mathbf{q})-\phi(\mathbf{q}'))^2\rangle \propto |\mathbf{q}|^{5/3}$ and the have Gaussian statistics: the probability for the field at two points $\mathbf{q}_1$ and $\mathbf{q}_2$ is $p(\phi_1,\phi_2)d\phi_1 d\phi_2 \propto \exp\left(-\phi_i C_{ij}^{-1}\phi_j/2\right)$ where $C_{ij} \equiv \langle\phi_i\phi_j\rangle$. Use these to calculate the auto-correlation of the complex field amplitude $C(\mathbf{z}) = \langle E(\mathbf{q})E^*(\mathbf{q}+\mathbf{z})\rangle$.

Q: Consider a plane wave incident on a sinusoidal 'phase' screen that imprints a sinusoidal phase error of small amplitude $\delta\phi \ll 1$. Use Fresnel theory to compute the complex field at distances downstream. Show that the initial phase-errors 'rotate' to become fluctuations in the amplitude of the field.

2.2 Detectors

Various classes - (nearly) all involve some sort of rectification (squaring and averaging) of the signal

- 'Bolometers': radiation falls on a detector element and heats it.
  - temperature change detected by change of resistance (e.g. transition from superconducting)
  - measures the energy integrated (over 'band-pass' of telescope or filter)
  - used in e.g. microwave background measurements
- Photon counting devices.
  - Originally photographic plates - these had low 'quantum efficiency' - or 'image intensifiers' (electron released from photocathode triggers shower)
  - Now mostly 'charge-coupled devices' (CCD).
    - originally proposed as solid-state digital storage devices
      - but were found to be rather sensitive to radiation
    - large planar semiconductor diode
    - photon above 'band-gap' energy ($\lambda \simeq 1\mu$m for silicon) creates electron-hole pair (extra energy $\rightarrow$ phonons)
    - the holes go to ground, while the electrons remain in the diode layer
      - the electrons are confined to a 2-dimensional grid of potential wells – like an egg-carton – created by a layer of electrodes ('gates') lying under the slab of silicon
    - once a sufficient integration time has elapsed, the image is read out by changing the voltages on the underlying gates beneath the photo-active layer so as to translate the periodic potential well pattern – and the accumulated photo-electrons – row by row into the 'serial array' at the edge of the device, from which the charge packets are read out one by one. This is illustrated in figure 11.
  - often used with broad band-filters
    - count of photoelectrons measure a different 'moment' of the spectrum of the source from bolometer
  - or with spectrographs - prism and/or grating
  - widely used in optical and X-ray
  - highly efficient and *linear*
  - 'noise' in image is combination of photon counting - 'root-N' Poisson statistics - plus 'read-out noise'
- Polarimeters
  - radio detector elements (e.g. dipole antennas) naturally polarization sensitive
  - polarization specific absorbers in beam (e.g. grid of wires) or birefringent crystals
  - can measure *linear* polarization

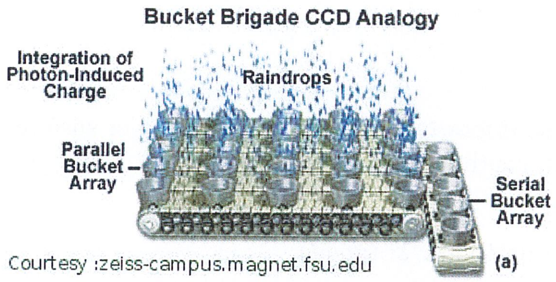


Figure 11: The 'bucket brigade' analogy for a CCD detector with radiation being likened to raindrops falling on an array of buckets. The raindrops represent photons impinging on the detector. These create electron-hole pairs that separate with the holes 'going to ground' and the electrons being trapped at the PN junction. The material where the photo-electrons accumulate is actually spatially homogeneous. The 'pixels' (buckets here) are potential wells defined by underlying electrodes (see below). These stop the accumulating image from smearing out.

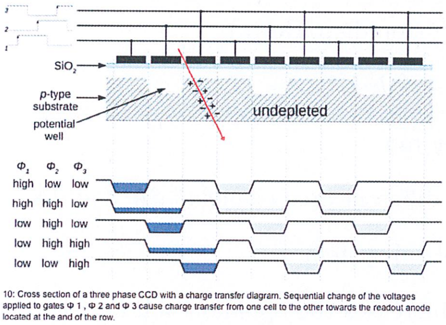


Figure 12: Charge 'clocking' in a CCD. Once the the pattern of charge has has accumulated the shutter of the telescope is closed and the voltages on the electrodes are driven so as to move the charge pattern over by one pixel, the end row then being read out of the 'serial register' into an amplifier and a 'analog-to-digital' (ADC) converter.

- circular polarization measurement can be done either with 'dichroic' materials that preferentially absorb one or other of the circular polarisation states, or with a beam splitter plus 'quarter-wave' plate and recombination
- circular polarisation detectors are not commonly used in astronomy

All of the above effectively measure, in one way or another, energy. A counter-example is LIGO, where one can see rather directly the field – gravitational rather than electromagnetic in that case.

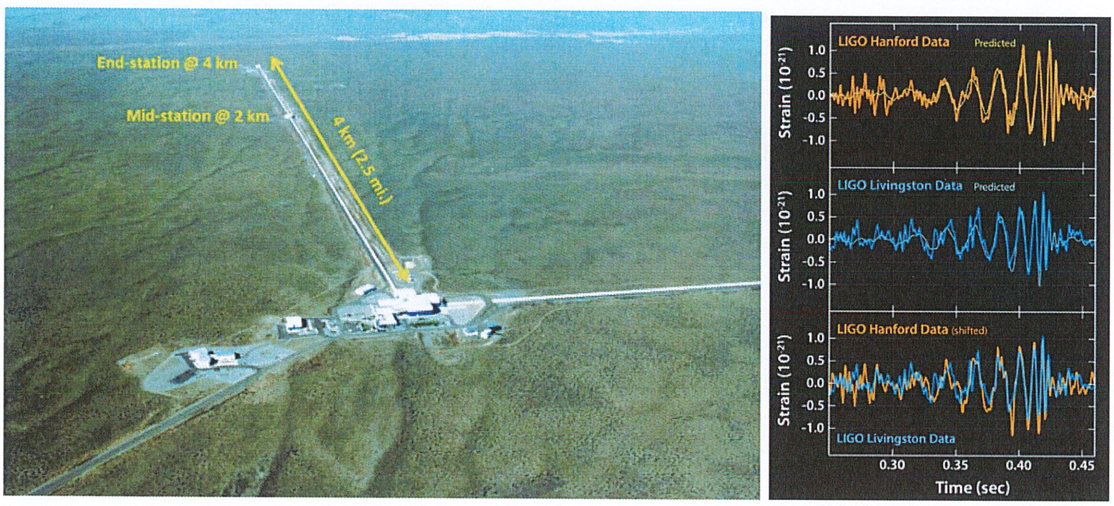


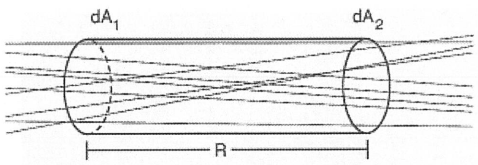
Figure 13: Left: aerial photograph of the LIGO detector at Hanford. Right: detected waveforms from GW0140915. This is different from most types of astronomical measurement not just in that it uses gravitational rather than electromagnetic radiation but also because the wave *'amplitude'* is measured and recorded rather than the energy deposited in some kind of non-linear detector.

3 Description of radiation - intensity, flux density, magnitudes and distance moduli

3.1 Intensity

Unpolarized radiation energy is most generally described by the *intensity* I_ν (or specific intensity or surface brightness). Can be thought of as superposition of waves with continuous range of directions $\mathbf{k}$ or rays (in geometric optics limit) with direction vectors $\hat{\mathbf{n}}$

- I_ν is a function of frequency $\nu = c/\lambda$ & direction $\hat{\mathbf{n}}$ (and position $\mathbf{r}$ & time t)
- defined such that the energy flowing through a surface $d\mathbf{A}$ in time dt in infinitesimal range of frequency $d\nu$ with ray directions in range of solid angle $d\Omega$ about $\hat{\mathbf{n}}$ is $dW = I_\nu d\nu d\Omega \hat{\mathbf{n}} \cdot d\mathbf{A} dt$
- units: energy/time/area/solid angle/frequency (e.g. Joule/sec/m^2/steradian/Hertz)
- if we integrate over frequency we get the 'bolometric' intensity $I \equiv \int d\nu I_\nu$
- *for freely propagating radiation the intensity is constant along any ray*
  - to prove this we consider all the rays that pass through two different area elements $d\mathbf{A}$ and $d\mathbf{A}'$
- but intensity is also constant for radiation passing through an optical system
  - a telescope does not increase the *intensity* of light – it just increases the apparent solid angle of sources
- this is a special case of *Liouville's theorem*, which states that the phase-space density $f(\mathbf{x}, \mathbf{p})$ along a beam of particles (or photons) is constant. This follows from Hamilton's equations $\dot{\mathbf{p}} = -\nabla_{\mathbf{x}} H(\mathbf{x}, \mathbf{p})$ and $\dot{\mathbf{x}} = \nabla_{\mathbf{p}} H(\mathbf{x}, \mathbf{p})$ which imply that the 6-dimensional volume occupied by a cloud of particles $d^3p d^3x$ is constant in time. Since the number of particles is $f(\mathbf{x}, \mathbf{p}) d^3p d^3x$ this means that $f(\mathbf{x}, \mathbf{p})$ is constant also. For photons the momentum is proportional to ν and this implies $\boxed{I_\nu / \nu^3 = \text{constant}}$.



$$dE_1 = I_{\nu_1} dA_1 dt \; d\Omega_1 d\nu_1 = dE_2 = I_{\nu_2} dA_2 dt \; d\Omega_2 d\nu_2$$

$$I_{\nu_1} = I_{\nu_2}$$

Figure 14: Conservation of intensity: In the absence of scattering the specific intensity along a ray is constant. To see this, consider all of the rays that pass through two small elements of area as shown. The solid angle of the rays at dA_1 that pass through dA_2 is $d\Omega_1 = dA_2/R^2$. Similarly $d\Omega_2 = dA_1/R^2$ Hence $dA_2 d\Omega_2 = dA_1 dA_2/R^2 = dA_1 d\Omega_1$. Invoking the definition of the intensity: $dE = I_\nu dA dt d\Omega d\nu$ and using $dt_1 = dt_2$ and $d\nu_1 = d\nu_2$ shows that I_ν must be constant.

3.2 The relation between the intensity and the energy density

- The photons in a cylinder of length l with directions in range $\delta\Omega$ about the axis of the cylinder will pass out of the cylinder in time $t = l/c$.
- These have energy $dW = I_\nu d\nu \delta\Omega dA l/c = (I_\nu/c) d\nu \delta\Omega dV \leftarrow$ volume
- So the energy density (for rays in this infinitesimal range of directions) and per unit frequency is $\delta u_\nu = dW/dV d\nu = (I_\nu/c)\delta\Omega$
- Integrating over direction gives the total energy density (per unit frequency) $u_\nu = \int d\Omega I_\nu / c$

u

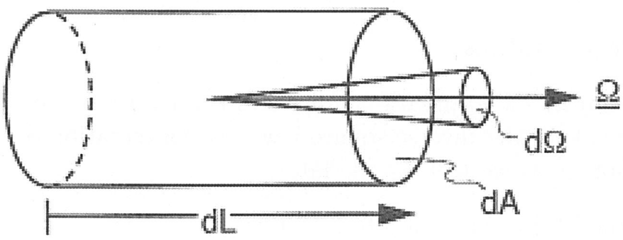


Figure 15: Illustration of quantities involved in the definition of intensity and its relation to the energy density. We define $u_\nu(\hat{\mathbf{n}})$ to be the energy per unit volume per unit frequency per unit solid angle in the direction $\hat{\mathbf{n}}$. The amount of energy in the cylinder is $dE = u_\nu dA dL d\nu d\Omega$ and this passes through the end of the cylinder in a time $dt = dL/c$. Hence $dE = (cu_\nu) dA dt d\nu d\Omega$ but this is $I_\nu dA dt d\nu d\Omega$, so $u_\nu = I_\nu / c$.

3.3 Flux density

Astronomical sources are usually of small solid angle. It is then most interesting to set the element of area – which is the area of the aperture of your telescope in practice – to be perpendicular to the direction to the source (so $\hat{\mathbf{n}} \cdot d\mathbf{A} \to dA$). The energy you will receive, per unit collecting area and per unit frequency, from a small part of the source subtending a small solid angle $\delta\Omega$ is $\delta F \equiv \delta\Omega I_\nu(\hat{\mathbf{n}})$, and integrating that over the whole source gives what we call the *specific flux density* of the source:

- $F_\nu \equiv \int d\Omega I_\nu(\hat{\mathbf{n}})$
- the energy crossing surface per unit time per unit frequency per unit area is $dW = F_\nu dA d\nu dt$
- so the units of F_ν are Joule/sec/m^2/Hz
  - or 'Janskys': 1Jy $\equiv 10^{-26}$J/sec/m^2/Hz
- and it is F_ν that obeys the expected *inverse square law*:
  - $F_\nu \propto 1/r^2$ because $d\Omega \sim 1/r^2$
  - and energy is conserved

3.4 Magnitudes and distance moduli

Astronomers measure flux densities, but often express them in terms of log-flux density

- The *apparent magnitude* m of a source is $m = -2.5 \log_{10}(F) + \text{constant}$
- So, for historical reasons, a factor 10 decrease in flux density is an *increase* of 2.5 magnitudes.
  - One magnitude is quite close to one e-folding - so magnitudes are close to natural logs
- The flux density is usually observed integrated over some range of frequency – defined by a filter, which will generally not have a simple transmission function. Photon counting *vs.* bolometric detectors further muddy the waters.
- Historically, and for convenience, the star Vega was used as a reference (the constant above was defined so that Vega has magnitude zero for all filters).
  - I.e. $m = -2.5 \log_{10}(F/F_{\rm Vega})$.
- The magnitudes (in different passbands) will depend on the filters used. So one might quote the magnitude of a star or galaxy as being $m_{\rm B}$, $m_{\rm R}$, $m_{\rm V}$ (in blue, red and 'visual' bands) in e.g. the 'Johnson-Cousins system'.
- It is now more common to see 'AB' magnitudes. These are defined such that a source with a flat spectrum (F_ν = constant) has $m_{\rm AB} = -2.5 \log_{10}(F_\nu / 3631{\rm Jy})$ where 1 Jy $= 10^{-26}$ W/Hz/m^2.

The *absolute magnitude* is (minus) the logarithm of the intrinsic luminosity of the source.

- Usually denoted by M.
- Defined such that a source of absolute magnitude M at distance $D = 10$pc has $m = M$.

- 1 pc = 1 AU / 1" $\simeq 3 \times 10^{16}$m

- The difference DM $= m - M$ is known as the *distance modulus*.
  - If you know the absolute luminosity of some object - a 'standard candle' - you can work out its distance from its apparent magnitude. This exploits the inverse square law – or conservation of energy. Things are more complicated in cosmology as we will discuss later.

The sun has $M = 4.83$. A typical ($L \sim L_\star$) galaxy like the Milky Way has $M \simeq -20$
Q: what is $L_\star$ in units of the solar luminosity $L_\odot$?

3.5 Polarized radiation and Stokes's parameters

The intensity I_ν is sufficient to describe 'natural' radiation; that is to say unpolarised radiation. To describe polarized radiation we need a more complete description. These are the 4 'Stokes parameters' I, Q, U and V (which, like I are generally functions of frequency).

- Q and U quantify the amount *linear polarization* (parallel to and at 45 degrees to the axes respectively).
- V measures circular polarisation.
- any of these can be positive or negative

3.5.1 Polarization of a nearly monochromatic beam: the complex field amplitude

- consider radiation from a distance source (so it has a small range of direction about the z-axis - and so the field will be very nearly perpendicular to the z-axis) that has a small range of temporal frequencies (perhaps it has been passed through a narrow band filter) around some central frequency ω.
- at fixed z we can write the x- and y-components of the electric field as $\mathbf{E}(t) = (x(t), y(t))$ where
  - $x = \mathrm{Re}(Xe^{i\omega t}) = \frac{1}{2}(Xe^{i\omega t} + X^\star e^{-i\omega t}) = \frac{1}{2}((X + X^\star)\cos\omega t - i(X - X^\star)\sin\omega t)$
  - $y = \mathrm{Re}(Ye^{i\omega t}) = \frac{1}{2}(Ye^{i\omega t} + Y^\star e^{-i\omega t}) = \frac{1}{2}((Y + Y^\star)\cos\omega t - i(Y - Y^\star)\sin\omega t)$
- so the 2-vector with complex components (X, Y) – 4 real degrees of freedom – encodes the amplitude and phase of the x and y components of the field.
- As the beam is nearly monochromatic, the amplitudes $(X(t), Y(t))$ will be slowly varying
  - timescale for variation $\tau \sim 1/\Delta\omega$ where $\Delta\omega$ is the range of frequencies (or bandwidth)
- If measured over a time-scale $\omega^{-1} \ll t \ll (\Delta\omega)^{-1}$ the field will be found to describe an ellipse in (x, y) space.
- But over a time $\tau \sim 1/\Delta\omega$ the parameters (orientation and semi-major and semi-minor axes) will change.

3.5.2 Measurement of polarisation

- In nearly all astronomical applications we do not measure the rapidly oscillating electric (or magnetic) field directly.
  - we may measure the energy deposited (in a bolometer or a CCD) and thus measure the time average of the square of the field
  - or we might send the output of an antenna to a non-linear device such as a diode and accumulate the charge
  - this again would measure the squared field
- A crude device for measuring linear polarisation, for example, might be a bolometer made of a narrow resistor

- an electric field parallel to the resistor will drive a current and raise the temperature (which we might measure through a change in the resistance)
- but an electric field in the perpendicular direction would not drive a current
- so the difference between the signals for two resistors – one aligned with the x-axis and the other aligned with the y-axis – would tell us if the radiation has a linear polarisation in the x- or y-direction
  * this would measure $\langle E_x^2 - E_y^2 \rangle$, which could be positive or negative, and is called the '+' component of polarisation
  * and the average would tell us the total intensity $\langle E_x^2 + E_y^2 \rangle$
- Another pair of resistors aligned along the diagonals would measure the '×' component of the polarisation
- the *phase* of the radiation, however, would not be measurable by such a device
- while this might seem crude, there exist, in reality, extremely sensitive detectors that work in this manner and which exploit the fact that a tiny amount of energy deposited in a superconductor that is close to the critical temperature can cause a drastic change in the resistance. These are called 'transition edge' sensors.

• What about a beam of circularly polarised radiation?

- here the field is constant in amplitude but rotating with time
- e.g. $(x, y) \propto (\cos \omega t, \sin \omega t)$
- to sense the circularity of the polarisation requires correlating the field amplitudes at different times – this can be done by introducing path delay in a waveguide. Or by correlating e.g. x and $\dot{y}$.
- in principle one could measure circular polarisation by measuring the angular momentum deposited in the detector as the radiation is absorbed (see TD1).

3.5.3 The Stokes parameters

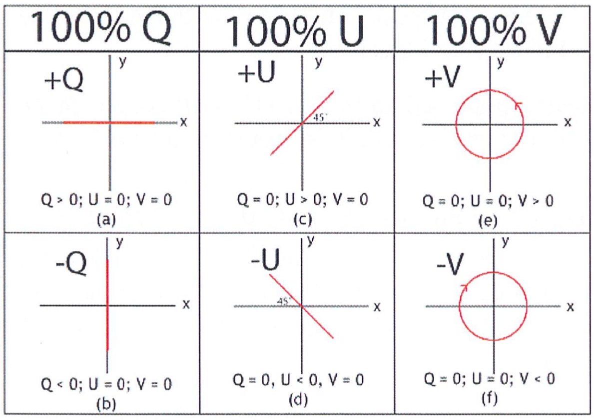


Figure 16: The Stokes parameters. The components of the vector (x, y) here are the components of the electric field for a monochromatic wave propagating along the z-direction with frequency ω. These are purely polarised waves where, for Q and U the field oscillates along the red line and for V the field rotates in field space. We can add these components with arbitrary phases to generate a general monochromatic wave. For 'narrow band' radiation – radiation that has been passed through a narrow band filter of width $\Delta\omega$ – the field is nearly monochromatic, but the combination of components will change over a timescale $\tau \sim 1/\Delta\omega$.

• the results of the kind of measurements described above can be expressed in terms of 4-intensities that are called the *Stokes parameters*.

• in general – as is the case for the simple case of a linearly polarised beam – there will be fluctuations about the mean intensity: the instantaneous intensity varying sinusoidally about the mean value with frequency 2ω

• the Stokes parameters all describe intensities averaged over (at least) a period of the radiation. They are as follows:

$$
\begin{array}{rl}
I \equiv \langle x(t)^2 + y(t)^2 \rangle & \text{the total intensity} \\
Q \equiv \langle x(t)^2 - y(t)^2 \rangle & \text{the `}+\text{' polarisation} \\
U \equiv \langle 2x(t)y(t) \rangle & \text{the `}\times\text{' polarisation} \\
V \equiv \langle x(t)y(t+T/4) - y(t)x(t+T/4) \rangle & \text{the circular polarisation}
\end{array}
$$

- where T is the period

- the total intensity is always positive, but Q, U and V can have either sign
- the circular polarisation is also given by $V = \langle x\dot{y} - y\dot{x} \rangle / \omega$
- the Stokes parameters are easily shown to be related to the real and imaginary parts of the field amplitude by:

$$
\begin{aligned}
I &= \tfrac{1}{2}\langle XX^\star + YY^\star \rangle \\
Q &= \tfrac{1}{2}\langle XX^\star - YY^\star \rangle \\
U &= \tfrac{1}{2}\langle XY^\star + YX^\star \rangle \\
V &= \tfrac{i}{2}\langle XY^\star - YX^\star \rangle
\end{aligned} \qquad (7)
$$

  - where $\langle \ldots \rangle$ denotes average over a period of the radiation
  - let's verify, as an example, $Q = \frac{1}{2}\langle XX^\star - YY^\star \rangle$
  - use $x = \text{Re}(Xe^{i\omega t}) = \frac{1}{2}(Xe^{i\omega t} + X^\star e^{-i\omega t})$
  - so $x^2 = \frac{1}{4}(X^2 e^{i2\omega t} + 2XX^\star + X^{\star 2} e^{-i2\omega t})$
  - taking the time average, the first and last terms disappear and we have $x^2 = XX^\star/2$
  - similarly, $y^2 = YY^\star/2$, so $Q = \langle x^2 - y^2 \rangle = \frac{1}{2}\langle XX^\star - YY^\star \rangle$

- Using (7) above one can readily verify the identity $Q^2 + U^2 + V^2 = I^2$.
  - so instantaneously (or over a timescale $t \ll (\Delta\omega)^{-1}$) the beam intensity in the various components is fully determined by the three parameters Q, U and V
  - these determine the *shape* of the ellipse $x(t)$, $y(t)$ in field space but not the *temporal phase*

- At any instant of time, the radiation is therefore always polarised (the only way for Q, U and V to vanish is for I to be zero). But when integrated over a longer timescale what one measures is a time average of the Stokes parameters, for which the above identity no longer holds.
  - for example one might have instantaneously linearly polarised radiation ($V = 0$) whose plane of polarisation rotates so the time average of Q and U vanish while the total intensity I has a non-vanishing time-average
  - it is in this way that 'natural' radiation (that given off by an incandescent source) is realised
  - and in general all 4 Stokes parameters are required to describe the time averaged properties of a beam of radiation, with Q, U and V telling us whether there is a *tendency* for the electric field to be fluctuating along one or other of the axes (Q) or along one or other of the diagonals (U) or having a tendency to rotate one way or the other (V)

- one particularly important mechanism for generating linear polarisation is scattering of radiation
  - if electrons are bathed in radiation for which the intensity has a quadrupole ($\cos 2\theta$) variation then the scattered radiation will be linearly polarised
  - polarisation measurements therefore provide a means of 'remote sensing' of the environment of the scatterers.

- another application is that dust particles in the Milky Way are aligned by the magnetic field and emit radiation that is linearly polarised.

PolarisationFromThomsonScattering

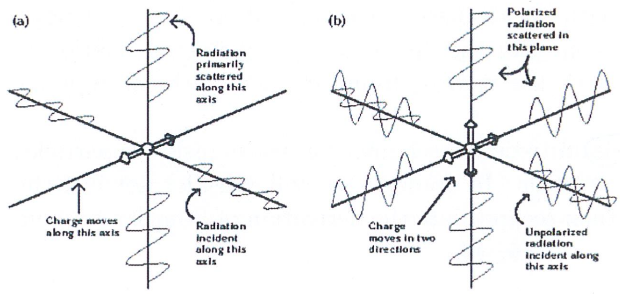


Figure 17: Illustration of how scattering of radiation by, for example, electrons in Thomson scattering, produces linear polarisation. On the left is shown scattering of linearly polarised radiation and unpolarised radiation on the right.

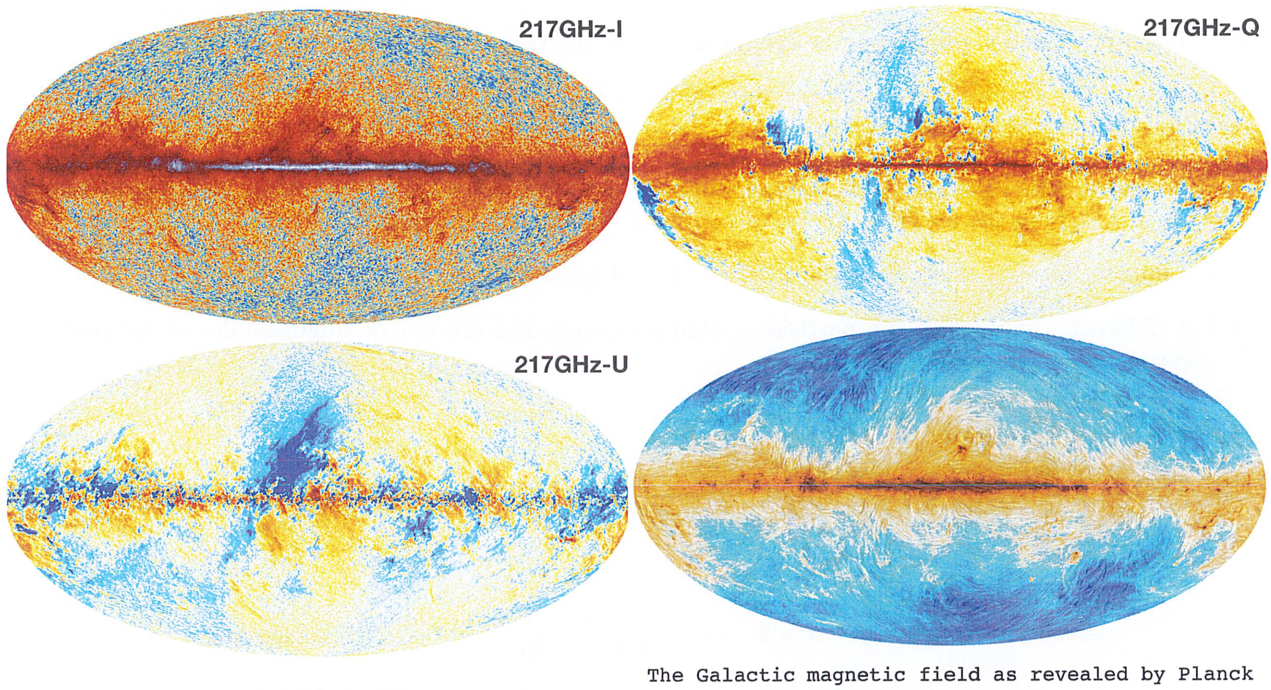


Figure 18: Maps from the ESA Planck satellite mission. Upper left is total intensity I. Upper right and lower left are the linear polarisation intensity parameters Stokes-Q and Stokes-U. All at 217GHz. Bottom right is rendering of the polarisation maps. Warning: the stripy features are not real; they are artificial modulation introduced to indicate the direction of the polarisation at each point.

4 Thermal radiation: Planck's formula

4.1 Statement of the problem; and its solution

Inject some EM energy into a box with perfectly reflecting walls. Let there be a speck of dust that can scatter, absorb and re-emit radiation. Leave it to equilibrate. What is the spectrum $u_\nu d\nu$ of the resulting thermal radiation?

According to classical 19th century physics, thermodynamics systems in equilibrium have kT of energy per degree of freedom. This works for gases where for monatomic gases the degrees of freedom are translational motion while for molecules one has, in addition, energy in the internal vibrational or rotational degrees of freedom. But applied to a classical radiation field this gave the wrong answer as there are infinitely many (Fourier) modes of the field at high frequencies. This is called the *'ultra-violet catastrophe'*.

The right answer was first given by Planck in October 1900 – in what he called an 'act of desperation':

$$\boxed{u_\nu d\nu = \frac{8\pi h\nu^3 d\nu}{c^3(e^{h\nu/kT}-1)}} \qquad (8)$$

This was the birth of quantum mechanics. But understanding it took at least another two decades. Planck's idea was that the energy of 'resonators' in the wall of the box was quantized but the radiation was considered classically. Einstein argued that the photo-electric effect showed the radiation to be quantized.

He also showed – considering radiation in equilibrium with an idealised 'two-state' atom – that Planck's formula required *stimulated* as well as spontaneous emission. But the first really satisfactory derivation of (8) was the revolutionary paper of Bose (1924) who showed that it arose from maximizing the entropy for a collection of fundamentally indistinguishable particles.

In the following we first remind ourselves of some basis statistical mechanics for distinguishable particles as considered by Boltzmann. We then review Bose's calculation. In doing so we will consider not just the spectrum but also the entropy of thermal radiation. We then recapitulate the derivation of Planck's formula (8) following Rybicki and Lightman.

4.2 Mini-review of Statistical mechanics

4.2.1 Distinguishable particles: Boltzmann statistics from a shuffling experiment

If we have particles that can have only discrete values of energy $E = 0, 1, 2, \ldots$, what is the most probable distribution function $n(E)$ assuming we have a fixed total number of particles and total energy?

- Boltzmann's proposition is that the probability is proportional to W, the number of distinguishable ways of assigning the particles to the 'energy bins' with those constraints.
- the number of ways of ordering the total number of particles is $N!$.
- but different orderings of the particles within an energy bin are not distinguishable, so we need to divide by a factor $n(E)!$ for each bin
- so $\boxed{W = N!/\prod_E n(E)!}$
  - Boltzmann coined the term *'complexion'* for this number

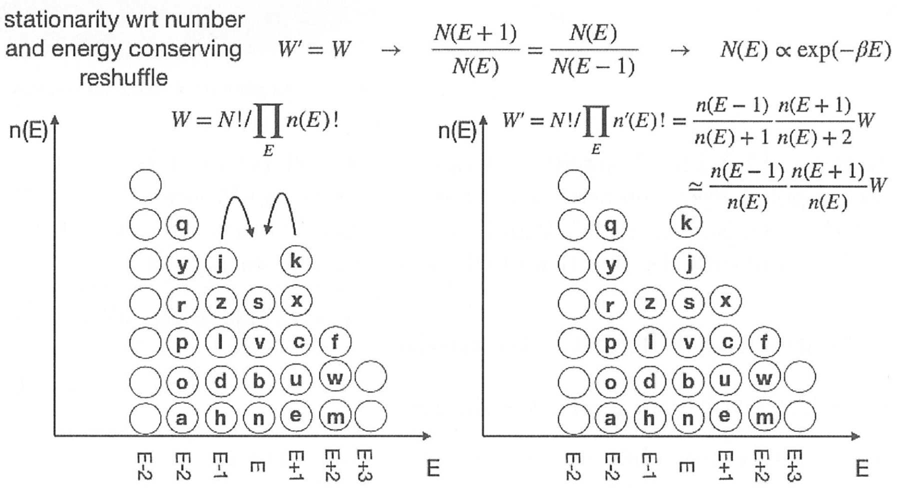


Figure 19: Boltzmann statistics: Imagine we have some particles which can have discrete values of energy (we'll choose the unit of energy to be the increment so E can be 0, 1, 2, ...). If we have a fixed total number of particles $N = \sum_E n(E)$ and fixed total amount of energy $E_{\rm tot} = \sum_E En(E)$, what is the most probable distribution function $n(E)$? Boltzmann's answer is that, if the 'occupation numbers' $n(E)$ are large, $n(E)$ must fall off exponentially with energy: $n(E) = \alpha\exp(-\beta E)$ with constants α and β determined by N and $E_{\rm tot}$.

- a sufficient way to determine the form of $n(E)$ that maximises W is to consider a simple reshuffling where we, for example, take two particles from the bin of energy E and increment by one unit the energy of one and decrement that of the other (thus conserving particles and total energy)
- or one can take one particle from each of the bins $E-1$ and $E+1$ and put them in bin E. This is illustrated in figure 19.

物理理解 why stationary ? 可细说.

- as shown in the figure, in order that W be stationary with respect to this 'reshuffling' requires, provided that we are dealing with large numbers of particles, that the ratio $n(E)/n(E-1)$ be a constant, independent of E.
- this requires that $\boxed{n(E) \propto \exp(-\beta E)}$ for some constant β.
  - this is known as *Boltzmann's law*
  - and the constant β is inversely proportional to the temperature

Q: Write a computer program (maybe in python using the Anaconda environment?) that manipulates an integer array of occupation numbers $n(E)$, initially uniformly populated with particles, by choosing bins at random and performing the kind of reshuffling described above. Plot $\log n(E)$ vs. E at various times as the occupation number distribution evolves. For large $n(E)$, what do you expect for the bin-to-bin or temporal *fluctuations* in the occupation numbers about the equilibrium relation? See what the computer says.

4.2.2 The Maxwellian distribution

An application of Boltzmann's law is the *Maxwellian distribution* for the distribution of momenta (or velocities) of gas particles in thermal equilibrium

- we divide momentum space $\mathbf{p}$ up into cubical cells of volume $(\Delta p)^3$
- let the number of atoms in the cell centred on $\mathbf{p}$ be $n(\mathbf{p})$
  - we'll call this the *occupation number*
  - and consider it to be averaged over some large volume, so n is large and the fluctuations in n can be ignored
- the Maxwellian distribution over *momenta* is $n(\mathbf{p}) \propto e^{-\beta E(\mathbf{p})}$ with $E(\mathbf{p}) = \mathbf{p}^2/2m$
  - this asserts that the probability that a randomly chosen particle be found in the cell centred on $\mathbf{p}$ obeys Boltzmann's law
- note that this gives a distribution over *energies* that is not a simple exponential as there is *degeneracy* – which we did not have in the simple shuffling experiment: here all of the cells with the same value of $|\mathbf{p}|$ have the same energy
  - to get the distribution function for energy $f(E)$ (defined such that the space density of particles with energy in the range $E : E + dE$ is $f(E)dE$) we need to multiply the occupation number $n(\mathbf{p})$ by the number of momentum cells $4\pi p^2 dp/(\Delta p)^3$ within $dp = (dE/dp)^{-1} dE$
  - so $f(E) \propto E^{1/2} \exp(-\beta E)$
- Maxwell obtained this before Boltzmann discovered his law, but his derivation is not considered to be rigorous

4.2.3 Boltzmann distribution *via* Stirling and Lagrange multipliers

A physically equivalent, but more formal, way to obtain the Boltzmann distribution is to maximise the *statistical mechanical entropy:* $\log W$ subject to the number and total energy constraints using *Lagrange multipliers*.

- Let the physical energy be $E = j\epsilon$ where j is an integer (bin label) and ϵ is the increment in physical energy and denote the occupation number by n_j.
  - we want to keep track of the 'quantum' of energy ϵ
  - though eventually we will take the limit $\epsilon \to 0$
- in terms of these, $N = \sum_j n_j$ and $E_{\rm tot} = \epsilon \sum_j j n_j$

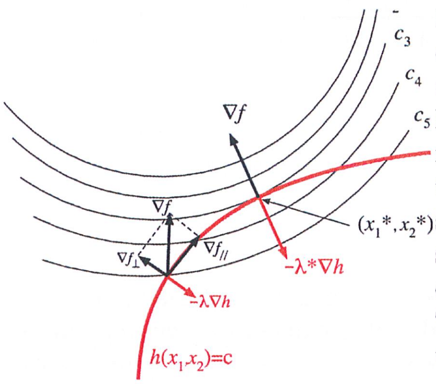


Figure 20: Lagrange multipliers are used to solve for the extremisation of a function $f(\mathbf{x})$ (here a 2-dimensional function $f(x_1, x_2)$) subject to a constraint (here that the function $h(x_1, x_2)$ is equal to c (red line)). In general, the gradient vectors ∇f and ∇h will point in different directions. The condition that ∇f and ∇h be parallel (or anti-parallel) to one another – i.e. that $\nabla(f + \lambda h) = 0$ for some 'multiplier' λ – is satisfied along a line. At any point on this line, the surfaces of constant f and constant h are 'kissing'. Solving $\nabla(f + \lambda h) = 0$ gives an expression $\mathbf{x} = \mathbf{x}(\lambda)$ for this line, parameterised by the 'multiplier' λ. The desired constrained extremum of $f(\mathbf{x})$ is at the point on this line where $h(\mathbf{x}(\lambda)) = c$. This is readily generalisable to spaces of higher dimensional and/or multiple contraints.

- For very large numbers of particles, and hence large n_j, Stirling's formula for the (natural log) of the factorial $\log n! \simeq n \log n$ gives for the *statistical mechanical entropy:*
  - $S \equiv \log W = \log\left(N!/\prod_j n_j!\right) \simeq N \log N - \sum_j n_j \log n_j$
- for a variation δn_j of the j^{th} occupation number, $\delta(n_j \log n_j) = (1 + \log n_j)\delta n_j$, while $\delta(N \log N) = (1 + \log N)\delta N = \delta n_j(1 + \log N)$ so the variation of the entropy with respect to n_j is
  - $\delta S/\delta n_j = \delta \log W/\delta n_j = (1 + \log N) - (1 + \log n_j) = -\log(n_j/N)$
- while $\delta N/\delta n_j = 1$ and $\delta E_{\text{tot}}/\delta n_j = \epsilon j$
- to obtain the distribution n_j that maximises $\log W$ subject to the constraints on the total number of particles and the total energy (see figure 20) we require that $(\delta \log W - \alpha\delta N - \beta\delta E_{\text{tot}})/\delta n_j = 0$, where α and β are Lagrange multipliers, or
  - $\log(n_j/N) + \alpha + \epsilon\beta j = 0$
- or
  - $n_j/N = e^{-\alpha-\beta\epsilon j}$.
- which, since $E = \epsilon j$, is Boltzmann's law as found before
- The constants α and β are fixed by N and E_{tot} as follows:
  - $\sum_j n_j/N = 1 = e^{-\alpha} \sum_j e^{-\epsilon\beta j} = e^{-\alpha}/(1 - e^{-\epsilon\beta})$ so $e^{-\alpha} = 1 - e^{-\epsilon\beta}$
    * so N gives a relation between α and β
  - $E_{\text{tot}} = \epsilon \sum_j j n_j = N e^{-\alpha} \sum_j \epsilon j e^{-\beta\epsilon j} = -N e^{-\alpha} \frac{d}{d\beta} \sum_j e^{-\epsilon\beta j} = -N(1 - e^{-\epsilon\beta}) \frac{d}{d\beta}(1 - e^{-\epsilon\beta})^{-1}$
  - or $E_{\text{tot}} = \epsilon N e^{-\epsilon\beta}/(1 - e^{-\epsilon\beta}) = N/\beta$
  - where in the last step we have taken the limit $\epsilon \to 0$
- so we have
  - $\boxed{\beta^{-1} = E_{\text{tot}}/N = \langle E \rangle}$
  - the average energy per particle, and
  - $\alpha = -\log\left(1 - e^{-\epsilon\beta}\right) \Rightarrow -\log \epsilon\beta$
- Key features:
  - the derivation seemed to require 'quantisation' of energy, but this is purely for mathematical convenience. The result, that the probability $p(E)$ that a particle has energy $E = \epsilon j$ is $p(E) \propto \exp(-\beta E) = \exp(-E/\langle E \rangle)$, is regular as $\epsilon \to 0$.
  - $\beta^{-1} = \langle E \rangle$ is an inverse measure of *ideal temperature*: $\beta^{-1} = k_{\text{B}} T$ where *Boltzmann's constant* k_{B} is fixed by the scale of temperature chosen.

- the *statistical mechanical entropy* – that which is maximised by the Boltzmann distribution – is
  * $$S = \text{constant} - \sum_j n_j \log n_j$$
- the *physical entropy* is related to the statistical mechanical entropy $S = \log W$ (which is a dimensionless number) by
  * $$S_{\text{phys}} = k_{\text{B}} \log W$$
  * the formula enshrined on Boltzmann's gravestone
- Boltzmann went on to show that in scattering of particles in a collisional gas, for example, the statistical mechanical entropy always tends to increase
  * this is called *'Boltzmann's H-theorem'*
  * he thereby provided a statistical mechanical explanation of thermodynamic entropy and the 2nd law of thermodynamics
- the 'complexion' $W = N!/\prod_e n_e!$ is that for *distinguishable particles*
- the fluctuations about the smooth equilibrium curve are Poissonian (see below)

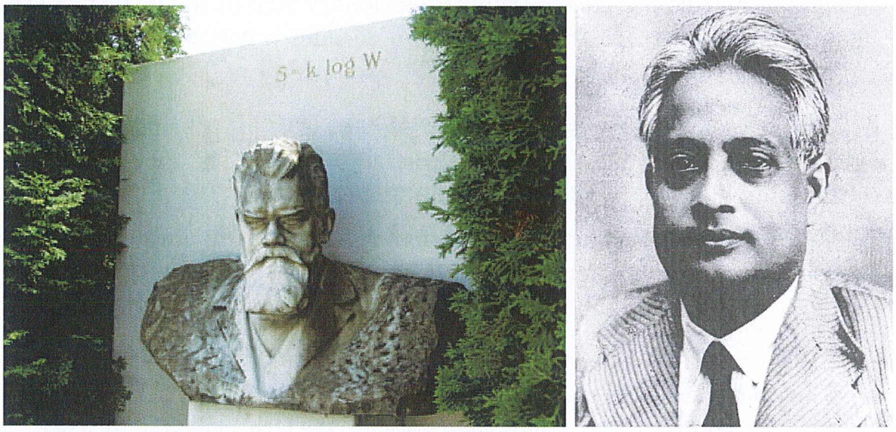


Figure 21: Left: Boltzmann's gravestone in Vienna. Right: S. N. Bose.

4.2.4 Indistinguishable particles: Bose statistics and Bose's complexion

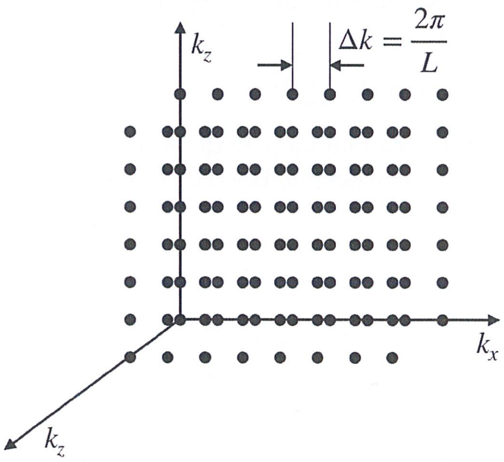


Figure 22: According to Planck, the modes of radiation in a cubical box of side L with perfectly reflecting walls are 'standing waves' with spatial frequencies $\mathbf{k}$ that live on a regular cubical lattice with spacing $\Delta k = 2\pi/L$. There are two polarisation states (2 helicity states) per $\mathbf{k}$ value. The frequencies (in Hertz = inverse period) are $\nu = c|\mathbf{k}|/2\pi$ so we can think of the modes as living on a lattice in 3-dimensional frequency space where the lattice points have spacing $\Delta\nu = c/L$. The number of modes in a shell of frequency ν and thickness $d\nu$ is $n_{\text{modes}} = 2 \times 4\pi\nu^2 d\nu/(\Delta\nu)^3 = (8\pi L^3\nu^2/c^3)d\nu$. If we let the size of the box tend to infinity the spacing of the lattice shrinks to zero and the number of modes for a given range of frequency increases in proportion to the volume.

- Following Planck, Bose argued that radiation in a box of side L has distribution of states in $\mathbf{k}$-space with 2 modes (2 polarisation states) per volume $d^3k = (2\pi/L)^3$ (see figure 22)

- so $dn_{\text{modes}} = 2 \times 4\pi k^2 dk/(2\pi/L)^3 = 8\pi\nu^2 d\nu/(c/L)^3$
- where the frequency (in Hertz – i.e. inverse period) is $\nu = c|\mathbf{k}|/2\pi$

- and, following Einstein, that these modes have integer numbers of 'light quanta' with energy $E = h\nu$
- He considers a thin shell (labelled s for shell) in $\mathbf{k}$- or frequency-space of frequency range $d\nu_s$ containing $A^s = 8\pi\nu_s^2 d\nu_s/(c/L)^3$ modes ('cells' in his terminology) and having N^s quanta.
- He defines p_r^s to be the number of cells in the s^{th} shell having r quanta and states that 'the number of ways these [N^s quanta can] be distributed among the cells belonging to $d\nu_s$' is
  - $\boxed{W^s = A^s!/\prod_r p_r^s!}$
  - from this, in a few short lines (taking the log of this, invoking Stirling's formula, and maximising the total complexion $W = \prod_s W^s$ subject to constraints on the total energy in the box just as for the derivation of Boltzmann's law), Bose obtains Planck's formula and, for good measure, an expression for the entropy of thermal radiation.
- Bose's W^s looks a lot like the Boltzmann 'complexion' for particles $W = N!/\prod_E n(E)!$ but it is actually very different. As this was such a truly revolutionary step in the history of science – one that had eluded the greatest minds of physics for over two decades – it is worth looking a little at where this comes from and its justification.
  - Figure 23 illustrates what Bose's complexion $W^s = A^s!/\prod_r p_r^s!$ means: It is the number of different sets of occupation numbers $\{r_j\}$ – where $j = 1, 2, \ldots A^s$ labels the cells in the shell – that have the same *distribution* of occupation numbers $\{p_r^s\}$.
  - It is entirely – even at a *conceptual* level – different from Boltzmann's complexion. Boltzmann's W – and hence his entropy – is defined by, and is a property of, the cell occupation numbers $\{r_j\}$.
  - Consider the set $\{r_j\} = \{1, 0, 8, 1, 0, 0\}$ (the middle row of figure 23)
    * There are $W = 10!/8! = 90$ ways to put $N = 10$ *distinguishable* particles in these cells with that $\{r_j\}$. That's because there are 10 ways to choose the particle that goes in the 1st cell and 9 ways to choose the particle that goes in the 4th cell and we then have no further choices. The statistical mechanical entropy of that $\{r_j\}$ is $\log 90 \simeq 4.499$
    * But for *indistinguishable* particles there is only *one* way to generate $\{r_j\} = \{1, 0, 8, 1, 0, 0\}$ as there is no extra information 'hiding' in the identities of the particles in the cells.
    * the entropy of $\{r_j\} = \{1, 0, 8, 1, 0, 0\}$ – or any other set – vanishes
  - According to Bose, the entropy of a shell s is a property of the histogram, or distribution function if you prefer, $\{p_r\}$ that says how many cells in the shell contain r quanta. It is the logarithm of the number of different sets $\{r_j\}$ consistent with that $\{p_r\}$.
- Next, he asserts '*The probability of the state defined by all p_r^s is clearly*' $\prod_s W^s$.
  - The implicit assumption here is that if we observe a thermal radiation bath we are somehow drawing at random a sample from all the possible sets of occupation numbers consistent with the 'macroscopic' constraint on the total energy.
  - this is called a *micro-canonical ensemble*
  - this seems to have been self-evident to Bose.
  - this provides a statistical model defining the probability of the distributions p_r^s, and it allowed Bose to calculate the most probable distribution. This gave a distribution of energy over the shells that agrees with Planck and with nature.

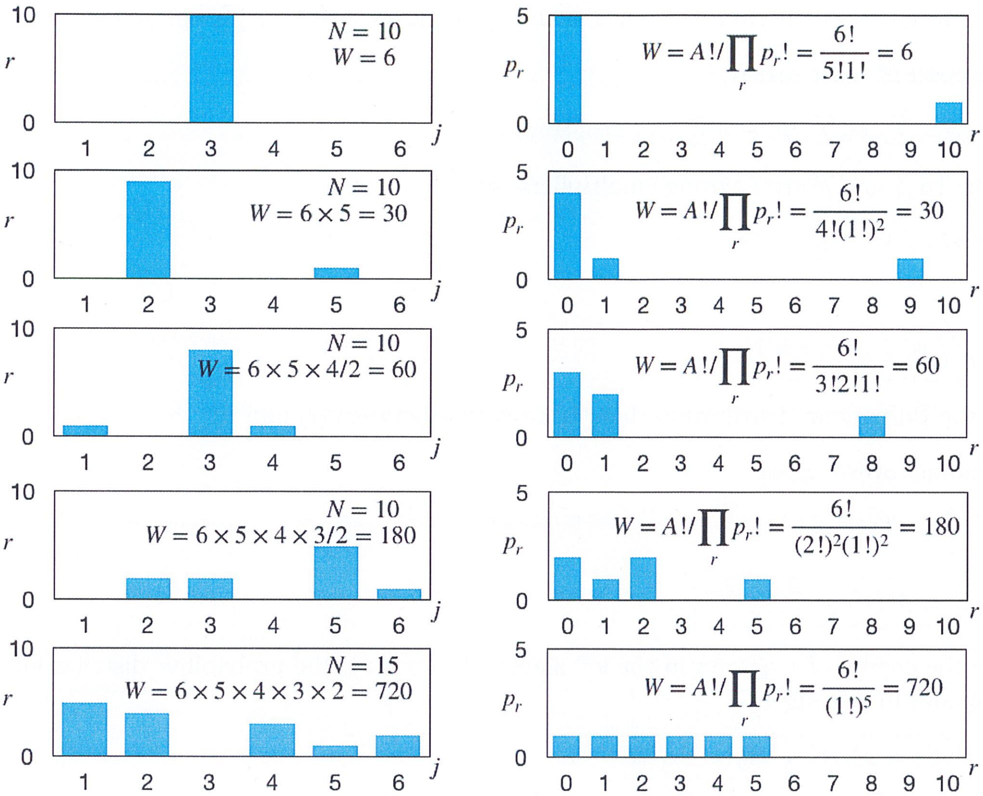


Figure 23: Panels on the left show the occupation numbers r_j for the case of a shell with $A = 6$ cells and with the total number of quanta N as indicated in the top right of each panel. On the right is shown the corresponding occupation number distribution p_r (the number of cells with occupation number r). On the left hand side is indicated the complexion W – i.e. the number of different sets of occupation numbers $\{r_j\}$ that have the same histogram $\{p_r\}$. For example in the top row, all of the 10 quanta are in the 3rd cell. There are $W = 6$ different ways the occupied cell could have been chosen. In the second row there are 9 quanta in the 2nd cell and 1 in the 5th. There are 6 ways of choosing the cell with $r = 9$ and 5 ways to choose the other occupied cell, so $W = 6 \times 5 = 30$. In the 3rd row, cell 3 has 8 quanta and cells 1 and 4 have one quantum each. There are $6 \times 5 \times 4 = 120$ ways to choose 3 occupied cells, but this generates only $6 \times 5 \times 4/2 = 60$ different sets $\{r_j\}$. In the bottom row all of the r_j are different so we get $W = A!$. On the right hand side is shown the complexion W calculated from the histogram p_r using Bose's formula. In each case this agrees with what we inferred from inspection of the $\{r_j\}$

break down

4.2.5 From Bose's complexion to the Planck spectrum

- Taking the log of W^s as defined above and summing over shells and invoking Stirling's theorem we have, for the total entropy
  - $S = \log W = \sum_s (A^s \log A^s - \sum_r p_r^s \log p_r^s)$
- the goal now is to find the distribution functions p_r^s that maximize this subject to the contraints:
  - $\sum_r p_r^s = A^s$ (which provides one constraint per shell) and
  - $\sum_s \sum_r r h\nu_s p_r^s = E_{\rm tot}$
- or, with $\nu_s = s\nu_1$ where $\nu_1 = 2\pi c/L$ is the frequency of the 'fundamental mode'
  - $\sum_s s \sum_r r p_r^s = E_{\rm tot}/h\nu_1$
- the variation of $A^s = \sum_r p_r^s$ with respect to p_r^s is $\delta A^s/\delta p_r^s = 1$, while $\delta(A^s \log A^s) = (1 + \log A^s)\delta A^s$, so the variation of the entropy is
  - $\delta S/\delta p_r^s = \log A^s - \log p_r^s = -\log p_r^s/A^s$

- while the variation of the sum representing the total energy is $\delta(\sum_s s \sum_r r p_r^s)/\delta p_r^s = sr$
- so the desired p_r^s must satisfy
  - $\delta(S - \sum_s \alpha_s \sum_r p_r^s - \beta \sum_s \sum_r s r p_r^s)/\delta p_r^s = 0$
- where the $\{\alpha_s\}$ and β are Lagrange multipliers, so
  - $\log(p_r^s / A^s) + \alpha_s + \beta r s = 0$
- or
  - $\boxed{p_r^s / A^s = e^{-(\alpha_s + \beta s r)}}$
- as with the Boltzmann distribution there is a relation between α_s and β:
  - summing over r gives
  - $1 = \sum_r p_r^s / A^s = e^{-\alpha} \sum_r (e^{-\beta s})^r = e^{-\alpha_s}/(1 - e^{-\beta s})$
  - so
  - $\boxed{p_r^s = A^s (1 - e^{-\beta s}) e^{-\beta s r}}$
- or, since the energy of r quanta in the s^{th} shell is $E = r s h \nu_1$, the probability distribution for energy for the modes in this shell is
  - $\boxed{P(E) = p_r^s / A^s \propto \exp(-E/k_{\rm B}T)}$
  - which is a Boltzmann distribution with $k_{\rm B}T = h\nu_1/\beta$
  - i.e. the same temperature, independent of shell
  - thus, despite the fact that we are dealing with fundamentally indistinguishable particles (or 'quanta') the *modes* – which *are* distinguishable – have a Boltzmann distribution for their 'occupation numbers' r and hence for their energy.
- to get Planck's formula we need the mean occupation number $f_s \equiv \langle r \rangle_s = \sum_r r p_r^s / \sum_r p_r^s = \sum_r r p_r^s / A^s$ which is straightforward to evaluate since, with $p_r^s / A^s = (1 - e^{-\beta s}) e^{-\beta s r}$, we have
  - $f_s/(1 - e^{-\beta s}) = \sum_r r e^{-\beta s r} = -\frac{1}{s}\frac{d}{d\beta} \sum_r (e^{-\beta s})^r = -\frac{1}{s}\frac{d}{d\beta}(1 - e^{-\beta s})^{-1} = e^{-\beta s}/(1 - e^{-\beta s})^2$
  - where we used $\sum\limits_{n=0}^{\infty} a^n = (1-a)^{-1}$
- from which, with $\beta s = h\nu_s / k_{\rm B}T$, we get
  - $\boxed{f_s = \dfrac{1}{e^{h\nu_s/k_{\rm B}T} - 1}}$
- this is the essence of Planck's function: to get the total energy at this frequency we multiply by the number of states $8\pi L^3 c^{-3} \nu_s^2 d\nu_s$ and by the energy of a quantum $h\nu_s$.
- finally dividing by the volume L^3 of the box gives Planck's expression
  - $\boxed{u_\nu = 8\pi c^{-3} h\nu^3 / (e^{h\nu/k_{\rm B}T} - 1)}$

4.3 Bolometric energy density, brightness and density of photons for a thermal radiation

- Integrating u_ν over frequency ν gives total energy density:
  - $\boxed{u \equiv \int d\nu u_\nu = aT^4}$
  - with $a = 8\pi k_{\rm B}^4 c^{-3} h^{-3} \int dx x^3/(e^x - 1) = 8\pi^5 k_{\rm B}^4 / 15 c^3 h^3$

- while the integral of the specific brightness $I_\nu = cu_\nu/4\pi$ (conventionally denoted by B_ν) is
  - $\boxed{B \equiv \int d\nu B_\nu = \sigma T^4}$
  - where the Stefan-Boltzmann constant is $\sigma = 2\pi^3 k_{\rm B}^4/15c^2h^3$
- and the total number density of photons is
  - $\boxed{n_\gamma = \int d\nu u_\nu/h\nu = \alpha T^3}$
  - where $\alpha = 8\pi k_{\rm B}^3 c^{-3} h^{-3} \int dx x^2/(e^x - 1)$
  - where $\int dx x^2/(e^x - 1) \simeq 2.40411$

4.4 The entropy of thermal radiation

4.4.1 The entropy for a general state

- The statistical mechanical entropy of a state is, in general,
- $$\boxed{S = \log W = \sum_s \left(A^s \log A^s - \sum_r p_r^s \log p_r^s \right)}$$
- since the A^s are constants, the interesting part of this is the second sum, which is rather reminiscent of Boltzmann's $S = -\sum n_i \log n_j$ which, for a gas, becomes $S = -\sum_{\bf p} f_{\bf p} \log f_{\bf p}$ where $f_{\bf p}$ is the number of particles in the cell around momentum $\mathbf{p}$.
- and since the p_r^s are essentially the probability to have r quanta in a shell this is also strikingly similar to Shannon's entropy $S = -\sum p \log p$ which gives e.g. the information content (per character) for messages written with a language where the p is the probability for the different letters of the alphabet (more on this later)
- what we would like to have – and will now calculate – is an expression for the entropy as a function of the mean occupation number f for the cells in shell s when the distribution functions p_r^s of the occupation numbers r take their thermal form.

4.4.2 The entropy for thermal radiation in terms of the mean occupation numbers

- If the radiation is thermal, $p_r^s = A^s(1 - e^{-\beta s})e^{-\beta sr}$, and so $\log p_r^s = \log A^s + \log(1 - e^{-\beta s}) - \beta sr$, so
  - $S = \sum_s A^s[-\log(1 - e^{-\beta s}) + (1 - e^{-\beta s}) \sum_r \beta sre^{-\beta sr}]$
- but $1 + f = (1 - e^{-\beta s})^{-1}$ so $\log(1 + f) = -\log(1 - e^{-\beta s})$ and $(1 + f)/f = e^{\beta s}$ so $\beta s = \log((1 + f)/f)$ while $(1 - e^{-\beta s}) \sum_r re^{-\beta sr} = e^{-\beta s}/(1 - e^{-\beta s}) = f$ so, putting all these together, we get $S = \sum_s A^s[\log(1 + f) + f \log((1 + f)/f)]$ or equivalently
  - $$\boxed{S = \sum_s A^s[(1 + f)\log(1 + f) - f \log f]}$$
  - where the second term is essentially the same as the Boltzmann entropy (since $\sum_s A^s \ldots$ is just summing over the possible momentum states)
  - this term dominates at high frequencies ($\beta s \gg 1$), where f becomes exponentially small and so $|\log f|$ becomes very large
  - but we have the first term in addition
  - this becomes important at low frequency ($\beta s \ll 1$) where it is essential as without it the entropy from these modes would be negative

4.4.3 Asymptotic behaviour for low and high frequencies

- for *low frequency*: $h\nu \ll k_B T \Rightarrow f \gg 1 \Rightarrow S \simeq \sum_s A^s \log f$
  - $\Rightarrow \delta S = \log f$ entropy per *mode* (or cell)
  - or $\delta S = f^{-1} \log f \ll 1$ entropy per quantum
- for *high frequency*: $h\nu \gg k_B T \Rightarrow f \ll 1 \Rightarrow S \simeq -\sum_s (A^s f) \log f$
  - $\Rightarrow \delta S = -\log f$ entropy per quantum
- entropy per quantum becomes large for high frequencies
  - but only scales *linearly* with frequency ν
  - while the occupation number f for such quanta drops *exponentially*
  - so their net contribution to the total entropy is very small
- the total entropy is dominated by the modes with $h\nu \sim k_B T$ for which entropy per quantum (or mode) is of order unity
- so the total entropy S is, aside from a dimensionless constant, just the number of quanta

4.4.4 The connection to Shannon's entropy and information content

- Shannon's formula says that the information content per character in some stream of characters is
- $S_{\text{Shannon}} = -\sum p \log p = \langle -\log p \rangle$
  - working in base-2 this is expressed as *bits* and is, for normal text, about 1.3 bits per character (less than the maximum information content for text composed from an alphabet of 26 characters, which is close to 5 bits, because letters are not equally likely
- the statistical mechanical entropy for radiation is very similar
- for low frequency modes the number of quanta r in a mode is drawn from an exponential probability distribution whose mean is f.
- so the probability to have any particular r is $p \sim 1/f$
- so the entropy per mode is, to order of magnitude, $\log f = -\log p$, just like for the information content of a message with an alphabet of f characters
- while at high frequencies, the number of quanta N – in some specified range of frequencies – is roughly $f \ll 1$ times the number of modes in that range. Choosing the range so $N \sim 1$, the number of modes available is $\sim 1/f$, and so the probability that the quantum occupies any particular one of these modes is $p \sim f$ and the Shannon entropy is then $-\log p \sim -\log f$ which is what we found for the entropy per quantum in this regime above.
- So if we think of a realisation of thermal radiation – i.e. a set of occupation numbers for all the modes of the radiation – as a message, then the Shannon information entropy – where at low frequencies we have $\sim f$ possible r values per mode and at high frequencies we have $\sim 1/f$ possible modes per quantum – is the same, to order of magnitude, as the statistical mechanical entropy

4.4.5 The entropy density for thermal radiation

- an alternative, but equivalent, expression is
  - $S = \sum_s A^s [\log(1 + f) + f\beta s]$
- or, with $A^s = 8\pi(L/c)^3 \nu^2 d\nu$, so $\sum_s A^s \ldots \to 8\pi(L/c)^3 \int d\nu \nu^2 \ldots$ and $\beta s = h\nu/k_B T$,
  - $S = 8\pi(L/c)^3 \int d\nu \nu^2 [-\log(1 + e^{-h\nu/k_B T}) + fh\nu/k_B T]$

- the second term here is just $L^3 u/k_{\rm B}T = E/k_{\rm B}T$, where E is the energy,and the first, after integrating by parts is found to be one third of this, so the physical entropy $S_{\rm phys} = k_{\rm B}S$ is
  - $\boxed{S_{\rm phys} = \frac{4}{3}\frac{E}{T}}$ and the entropy density is $\boxed{s_{\rm phys} = \frac{4}{3}\frac{u}{T}}$

Q: Verify the factor 4/3 above.

It is well worth reading Bose's 1924 paper. It is short and relatively straightforward. The history is interesting: it was first rejected, but Bose sent it to Einstein, who realised its true worth and personally translated it to German for publication (an English translation can be found on the web). Ironically, the word 'indistinguishable' does not appear in the paper. Apparently it was Ehrenfest who first realised the essential feature underlying Bose's masterstroke.

4.5 Photon statistics vs. distinguishable particles

- For thermal radiation, the mean occupation number $f_s \equiv \langle r \rangle = 1/(e^{\beta s} - 1)$ becomes very large at $\beta s \ll 1$ (the 'Rayleigh-Jeans region')
  - indeed $f_s \simeq 1/\beta s = kT/h\nu_s \gg 1$ for $h\nu_s \ll kT$
- So there are, on average, an extremely large number of quanta per mode at low frequencies
- but the probabilty for a cell in the $s^{\rm th}$ shell to have occupation number r is $P_s(r) = p_r^s/A^s = (1 - e^{-\beta s})e^{-\beta s r} \propto e^{-\beta s r}$
  - so the distribution of r values is exponential and therefore extremely broad
- this is very different from what one finds for distinguishable particles, where if $\langle r \rangle \gg 1$, the fluctuations about the mean would be Poissonian with large mean and therefore very small.
  - This is illustrated in figure 24.
- to make this more concrete, lets calculate the entropy S for distinguishable particles as a function of the distribution function p_r^s (rather like Bose did for indistinguishable particles) and obtain the Poissonian solution as the maximum entropy solution
  - let's consider a single shell, so we can drop the sub- and superscript s
  - as before, we let there be A cells and a total of N particles and let p_r be the number of cells containing r particles (so $A = \sum_r p_r$ and $N = \sum_r r p_r$)
  - we want to know, first of all: what is W the number of distinguishable ways to distribute these N particles in these A cells?
  - this is straightforward: first we need to ask what is the number of sets of occupation numbers $\{r_j\}$ $(j = 1 \ldots A)$
    * but that is just $A!/\prod_r p_r!$ as calculated by Bose
  - and then multiply this by the number of ways of distributing the particles with these $\{r_j\}$
    * which is just $N!/\prod_j r_j!$
  - but we would like the latter factor in terms of p_r, which is also straightforward since
    * $N!/\prod_j r_j! = N!/\prod_r (r!)^{p_r}$
  - so the total complexion for the distribution $\{p_r\}$ is
    * $\boxed{W = (A!/\prod_r p_r!) \times (N!/\prod_r (r!)^{p_r})}$
  - taking the log gives the total entropy
    * $S = \log W = \log A! - \sum_r \log p_r! + \log N! - \sum_r p_r \log r!$

- we may safely assume that A, p_r, and $N = \sum_r rp_r$ are all large numbers, so we can invoke Stirling's approximation to give
    - $S = \log W = A \log A - \sum_r p_r \log p_r + N \log N - \sum_r p_r \log r!$
- and using $\delta A/\delta p_r = \delta/\delta p_r(\sum_r p_r) = 1$, $\delta N/\delta p_r = \delta/\delta p_r(\sum_r rp_r) = r$ we find that the constrained maximisation of S requires
    - $0 = \delta(S + \alpha N + \lambda A)/\delta p_r = (1 + \log A) - (1 + \log p_r) + r(1 + \log N) - \log r! + \alpha r + \lambda$
- with solution
    - $p_r/A = \mu^r e^\lambda/r!$
    - where $\mu = e^{1+\alpha+\log N}$
- summing over r gives
    - $A^{-1} \sum_r p_r = 1 = e^\lambda \sum_r \mu^r/r! = e^\lambda e^\mu$
    - so $\lambda = -\mu$ and therefore
    - $\boxed{P(r) = p_r/A = \mu^r e^{-\mu}/r!}$
- which is a *Poisson distribution* with mean $\langle r \rangle = \sum_r rp_r / \sum p_r = \mu = N/A$

- thus, for distinguishable partlcles, the maximum entropy probability distribution function is a Poissonian distribution (as one might well have guessed)
    - which, for large μ, becomes very narrow
    - in fact it tends to a Gaussian distribution $P(r)dr = (2\pi\mu)^{-1/2} \exp\big(-(r-\mu)^2/2\mu\big)$
    - with $\langle(\delta r)^2\rangle/\langle r\rangle^2 = \langle(r - \langle r\rangle)^2\rangle/\langle r\rangle^2 = 1/\mu \ll 1$
    - and which is very different from the exponential distribution for indistinguishable particles which remains very broad regardless of how large $f = \langle r \rangle$ becomes
- Finally, we note that the complexion above factorises into two terms – and so the entropy is the sum of two terms
    - the first, which is the same as for indistinguishable particles, and which is the entropy of the distribution p_r
    - and the second which is the number of ways of shuffling the N distinguishable particles with some set of occupation numbers $\{r_j\}$
- Previously we considered only the second term – as we considered the entropy to be given in terms of $\{r_j\}$
    - but it is not difficult to show that for large mean occupation numbers, this is on the order of $S \sim N \log A$ and overwhelmingly dominates over the extra entropy of the distribution
    - so we were justified, in the Boltzmann case, of thinking of the entropy as being dominated by the 'shuffling entropy'
    - but for indistinguishable particles there is no shuffling entropy and the entire entropy is that which derives from the distribution $\{p_r\}$

4.6 Rybicki & Lightman derivation of Planck spectrum

This has been a long discussion, in the course of which we have derived not just the Planck spectrum, but the form for the entropy, and we have elucidated the essential difference between Boltzmann and Bose statistics. If one is solely interested in the spectrum, there is no clearer exposition than that in Rybicki and Lightman's book which we recapitulate here:

- The allowed standing modes of radiation in a large box of side L has quasi-continuous distribution of states in $\mathbf{k}$-space with 2 modes (2 spin, or polarisation, states) per volume $d^3k = (2\pi/L)^3$.
    - so $dn_{\rm modes} = 2 \times 4\pi k^2 dk/(2\pi/L)^3 = 8\pi\nu^2 d\nu/(c/L)^3$

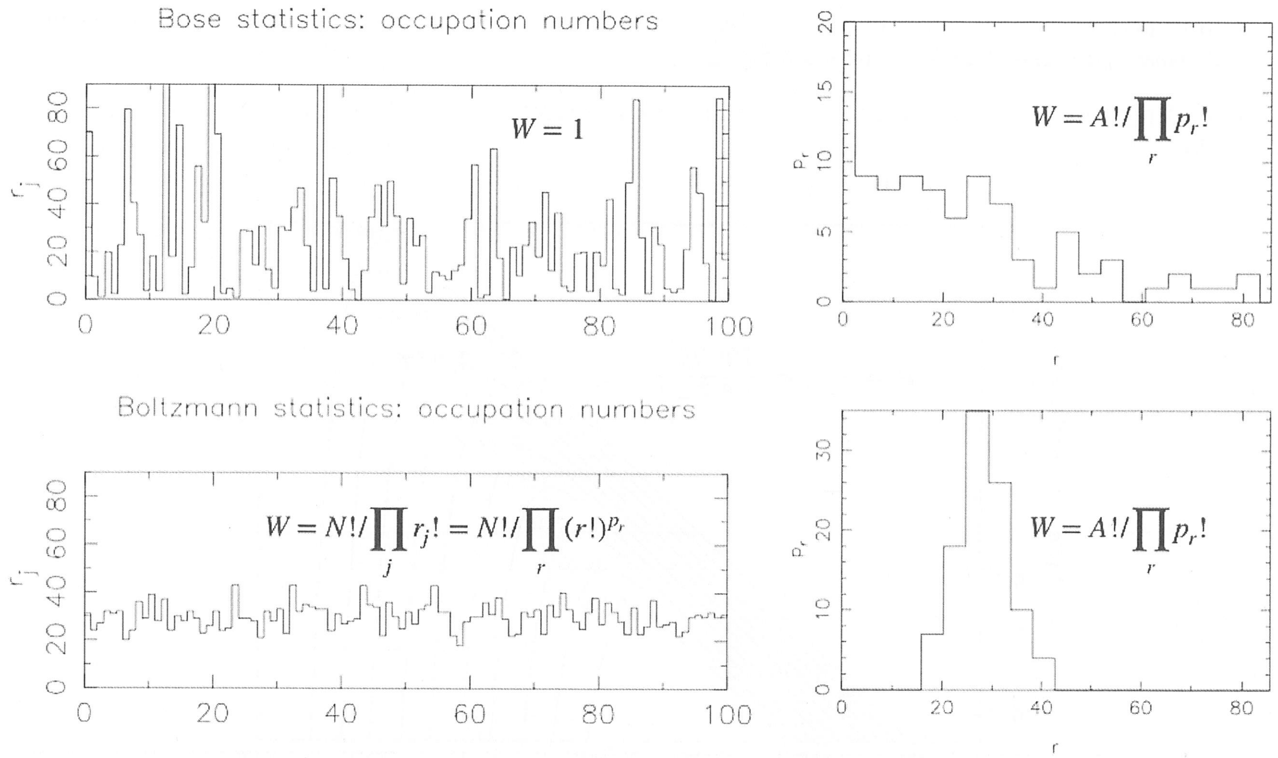


Figure 24: Illustration of Boltzmann vs Bose statistics. The top panels show, on the left, a realisation of the occupation numbers r_j for the $A = 100$ cells in one shell (all have the same energy) according to Bose statistics (i.e. for indistinguishable particles). These were drawn from an exponential distribution with mean $\mu = \langle r \rangle = 30$. As indicated there is only $W = 1$ way to generate this particular set of occupation numbers. One the right is the distribution of the occupation numbers p_r plotted vs. r with Bose's formula for W: the number of different $\{r_j\}$ like that on the left that have this $\{p_r\}$. The lower panels show the same things, and for the same mean number of particles, but with Boltzmann statistics as one would find for distinguishable particles. Here the occupation numbers $\{r_j\}$ were drawn from a Poisson distribution. The total number of ways of distributing the $N = 3000$ particles in this case is given by $W = A!/\prod_r p_r!$ – as for Bose statistics – but multiplied by the much larger factor $W = N!/\prod_j r_j!$ which is the number of distinguishable ways of shuffling the $N = 1000$ particles among the cells with these $\{r_j\}$. The entropy – being the sum of the entropies associated with $\{p_r\}$ and with $\{r_j\}$ – are, in the Boltzmann case, dominated by the 'shuffling' entropy.

  - since $\mathbf{p} = \hbar\mathbf{k}$, we might also say that there are 2 states per volume $(\Delta p)^3(\Delta r)^3 = \hbar^3$ in 6-dimensional *'phase space'* $(\mathbf{r}, \mathbf{p})$

- QM says the energy of a mode of frequency $\omega = c|\mathbf{k}| = 2\pi\nu$ is quantized in units of $\hbar\omega$: so $E_{\mathbf{k}} = nh\nu$
  - this ignores the constant 'zero-point' energy $h\nu/2$
- These different *modes* are distinguishable (unlike photons) so the probability to find a 'sub-system' (a mode here) to have energy E in thermal equilibrium at temperature T is given by Boltzmann's formula: $P(E) \propto \exp(-\beta E)$ where $\beta = 1/kT$.
- The average energy for a mode is then $\langle E \rangle_\nu = \sum_n E_n P(E_n) / \sum_n P(E_n)$
  - or $\langle E \rangle_\nu = \sum_n nh\nu e^{-\beta nh\nu} / \sum_n e^{-\beta nh\nu} = -Z^{-1}\partial Z/\partial\beta$
  - where the *'partition function'* is $Z \equiv \sum_n e^{-\beta nh\nu}$
  - but this is a simple geometric series: $Z = \sum_n a^n$ with $a = e^{-\beta h\nu}$ and $\sum_n a^n - a\sum_n a^n = 1 \rightarrow Z = (1-a)^{-1} = (1 - e^{-\beta h\nu})^{-1}$
  - from which $\partial Z/\partial\beta = h\nu e^{-\beta h\nu}(1 - e^{-\beta h\nu})^{-2}$
  - so $\langle E \rangle_\nu = -Z^{-1}\partial Z/\partial\beta = h\nu/(e^{h\nu/kT} - 1)$

- Multiplying $\langle E \rangle_\nu$ by the number of states in some range of frequency $d\nu$ gives the specific energy density $u_\nu d\nu$: the Planck's expression for u_ν
- Multiplying by c and dividing by 4π gives the specific brightness for a 'black-body' $I_\nu = B_\nu$ (see figure 25)

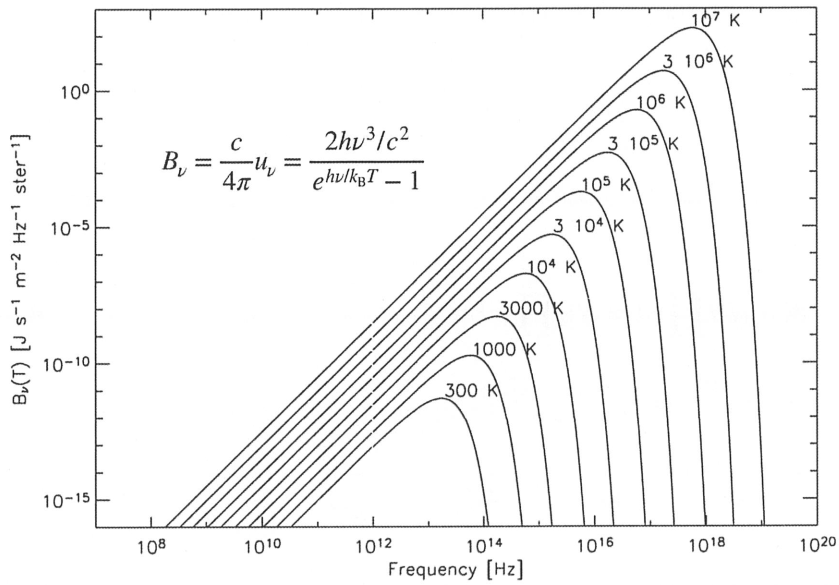


Figure 25: Planck spectrum (brightness) on a log-log scale for a range of temperatures. This shows that a) the brightness has a power-law form at low frequencies b) it is exponentially suppressed at high frequencies and c) the transition frequency scales linearly with the temperature. At all frequencies the brightness is a monotonically increasing function of temperature.

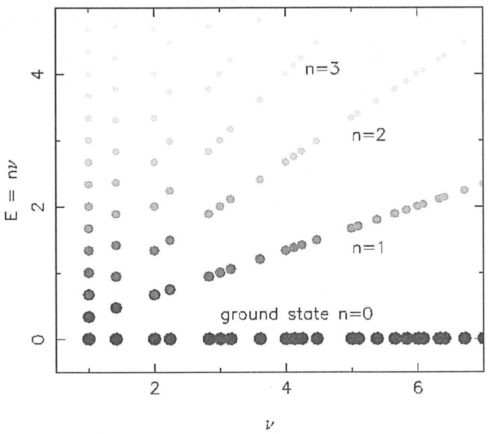


Figure 26: The size (and darkness) of the symbols is proportional to the probability as a function of the frequency ν and energy E for the thermal distribution. For each frequency, the probability obeys Boltzmann's law. This shows how, at low frequencies, there is essentially a continuous distribution of energies whereas at high frequencies the discreteness of the allowed energy levels becomes important and the probability for a mode to be excited is very small. This is what causes the 'cuf-off' in the brightness at $h\nu > k_B T$.

This formula beautifully combines electromagnetism – and therefore special relativity (characterised by c) – statistical mechanics, and quantum mechanics (through h).

Q: Boltzmann's constant also appears in the Planck function. Is k_B 'fundamental' in the way we believe h and c to be?

Q: Estimate the phase-space density (number of particles per spatial volume per momentum volume) for air molecules. How does this compare to $1/\hbar^3$.

Q: This argument invokes quantum mechanics as saying that each mode of the radiation field can only be in one of its energy eigenstates with $E = (n+1/2)h\nu$. It is true that the modes – to the extent that interactions can be neglected, and it is reasonable to assume a weak coupling – are like independent harmonic oscillators. And these have these energy eigenvalues. But does quantum mechanics really say an oscillator has to be in an eigenstate? What's to stop it being in a linear superposition of say the ground state plus a small admixture of the first excited state? That would seem to undermine the argument. Discuss.

4.7 Properties of thermal radiation

Some interesting and useful characteristics of the Planck spectrum are the following:

- For $h\nu \ll kT$ the mean energy per mode is $\langle E \rangle_\nu = kT$
  - this is what 19th century physics predicts - equipartition: kT per mode
  - this is called the 'Rayleigh-Jeans' region of the spectrum
  - and the probability distribution for the energy is Boltzmannian
    - which is what you get for waves with complex amplitude $E = a + ib$ if the real and imaginary components have a *Gaussian* distribution
    - and which is also known as the χ^2 distribution with 2 d.f.
- But at the opposite extreme $h\nu \gg kT$ the mean energy per mode is strongly suppressed
  - this is called the 'Wien' region of the spectrum
  - because the probability to be in even the lowest energy state for these modes is exponentially suppressed - a direct consequence of the quantization of energy (see figure 26)
- We can write $\langle E \rangle_\nu = h\nu f_\nu$
  - $\boxed{f_\nu = (e^{h\nu/kT} - 1)^{-1}}$ ← mean 'photon occupation number' for modes of frequency ν
  - which is $\gg 1$ and $\ll 1$ respectively in the R-J and Wien regions respectively
  - In R-J region $f_\nu \simeq kT/h\nu$ so the mean energy per mode is $h\nu f_\nu \simeq kT$
  - In the Wien region, $f_\nu \sim e^{-E_\nu/kT}$, so we can say that the probability for a photon to be found in a state of energy $E_\nu = h\nu$ is then given by Boltzmann's law.
    - this is similar to the Maxwellian distribution for energies of gas particles
    - though the energy of photons has a different dependence on the momentum (linear rather than quadratic for non-relativistic atoms or molecules)
- The total energy density is $u = aT^4$ and scales as the 4th power of temperature
- The 'characteristic energy' scales as T, so characteristic wavelength $\lambda_* \propto 1/T$
- The number density of photons n_γ scales as T^3
  - one can readily show that, to order of magnitude, there is one photon per volume λ_*^3, i.e.
  - $n_\gamma \sim 1/\lambda_*^3$
- as does the entropy

4.8 Determining the temperature of sources

Many astrophysical sources emit radiation that has a spectrum which is close to that of thermal radiation.

- If the source is unresolved, it is impossible to say what I_ν is.
  - however, for a sufficiently extended source, a measurement of I_ν at a single frequency (or with a single band-pass) determines the *brightness temperature*: the temperature of a black-body for which B_ν equals the observed brightness
- But the *ratio* of the flux densities at two different frequencies - or equivalently the '*colour*' such as $m_{\rm B} - m_{\rm R}$ in two passbands - is sufficient to determine T.
  - this becomes impossible in the RJ limit as the spectrum is, asymptotically, a power law
  - but otherwise one can determine the *colour temperature* (the temperature of a black-body with the same colour) even for a point source

Quantum State of thermal radiation !!! (warning)
coherent or not?

4.9 Thermal radiation with a non-vanishing chemical potential

- You may have noticed that in Bose's derivation of the Planck spectrum he imposed a constraint on the total energy, but not on the total number of quanta
- this is in accord with our initial statement of the problem to be solved
  - that of the equilibrium state for radiation in a container containing some 'speck of dust' that can absorb and re-emit radiation and thereby change the number of quanta
  - in such a situation the total number of quanta adjusts itself appropriately
- but in some astrophysical situations there may be scattering which can partially equilibrate the radiation but in which the number of quanta does not attain the value for a Planckian radiation field of the relevant temperature
- the spectrum obtained in such circumstances is obtained by adding an appropriate additional Lagrange multiplier
  - conventionally denoted by μ
  - and known as the *'chemical potential'*
- and the resulting spectrum has mean occupation number
  - $\boxed{f_\nu = (e^{\mu + h\nu/kT} - 1)^{-1}}$
  - note: it is more common to define the chemical potential $\mu' = kT\mu$ (i.e. as an energy) in terms of which $f_\nu = (e^{(\mu' + h\nu)/kT} - 1)^{-1}$

5 Radiation Pressure

Radiation pressure is important in cosmology and in some stars.

5.1 The particle picture

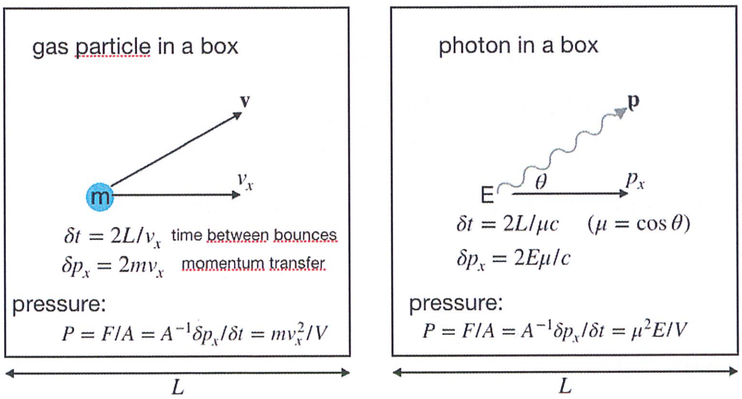


Figure 27: Kinetic pressure for a gas of particles or photons is the flux density of momentum. It may be calculated by computing the momentum transferred to the wall of a box per unit time and dividing by the area. For non-relativistic gas particles the pressure is 2/3 the energy density: $P = (2/3)u$. For photons or relativistic particles $P = (1/3)u$. These assume an isotropic distribution of momenta.

- Classical electromagnetic waves carry momentum as well as energy (momentum flux = energy flux / c). Same for 'photons' pictured as relativistic particles of zero or negligible mass.
- Pressure may be defined as the *flux density of momentum*
  - Q: is there a flux of momentum inside a pressurized balloon? Which direction is momentum flowing?
  - Q: what if I stretch a spring? Is there a flux of momentum in a stretched string? Which way is it flowing?
- Alternatively, think of photons bouncing off the walls of a reflecting box (of side L).

- Recall the kinetic pressure for a collisionless gas:
  - one molecule - it bounces of the wall at $x = +L/2$ once per time $\delta t = 2L/v_x$
  - and transfers momentum $\delta p = 2mv_x$
  - so the time-averaged force is $F = \delta p/\delta t = mv_x^2/L$
  - and therefore the pressure is $P = F/A = mv_x^2/V$
  - for N particles $P = m\langle v_x^2\rangle N/V$
  - or, defining the particle number density $n \equiv N/V$
    * $\boxed{P = mn\langle v^2\rangle/3}$
    * where $\langle v^2\rangle = \langle v_x^2 + v_y^2 + v_z^2\rangle$
    * and we have assumed an isotropic distribution of momenta: $\langle v_x^2\rangle = \langle v_y^2\rangle = \langle v_z^2\rangle$
  - but the energy density of the gas is
    * $\boxed{u = mn\langle v^2\rangle/2}$
    * the kinetic energy of a particle being $E = mv^2/2$
  - so the pressure is 2/3 of the energy density for an isotropic gas of molecules
- The situation is very similar for a gas of photons (or of highly relativistic particles). The only difference is that the particles here move at the speed of light
  - if we consider 1 photon moving at an angle θ with respect to the x-axis
    * the velocity in the x-direction is μc, where $\mu \equiv \cos\theta$
    * the time between reflections off the wall at $x = +L/2$ is $\delta t = 2L/\mu c$
    * and the transfer of momentum is $\delta p = \mu|\mathbf{p}| = \mu E/c$ (with E the photon energy)
    * so the time averaged force is $F = \delta p/\delta t = \mu^2 E/L$
    * and the time averaged pressure is $P = F/A = \mu^2 E/V$
  - for $N = nV$ photons the pressure is $P = (nV) \times \langle\mu^2 E\rangle/V$ or, assuming an isotropic distribution at any energy, so $\langle\mu^2\rangle = 1/3$,
    * $\boxed{P = n\langle E\rangle/3}$
  - while the energy density is
    * $\boxed{u = n\langle E\rangle}$
  - so the pressure is 1/3 of the energy density for isotropic 'gas' of photons

5.2 Black-body radiation as a 'thermodynamic system'

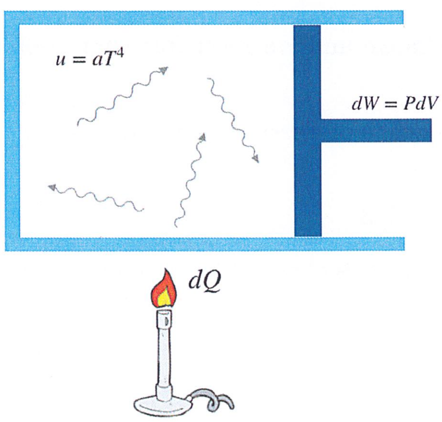


Figure 28: The entropy density of thermal radiation, which was calculated by Bose to be $s = S/V = (4/3)u/T$, can be inferred from $u = aT^4$. If we keep the piston fixed, and add heat dQ reversibly (the bunsen burner here being only figurative; more accurately we imagine heat being supplied by bringing the piston in contact with a series of a 'heat reservoirs' of progressively higher temperature à la Carnot) we have $du = 4aT^3 dT$ so $S = \int dQ/T = \int du/T = (4/3)VaT^3 = (4/3)u/T$. This is in accord with Bose's statistical mechanical entropy and establishes the connection between β and $1/k_B T$. On the other hand, if we switch off the heater, and allow the system to do some PdV work in an adiabatic manner (conserving photons) then we find $P = u/3$ and that the pressure scales with volume as $P \propto V^{-\gamma}$ with *'adiabatic index'* $\gamma = 4/3$.

- Consider a cylinder with a piston containing thermal (or 'black body') radiation

- the energy density is $u = aT^4$ and the number density of photons $n_\gamma = \alpha T^3$

- First, imagine that we add heat dQ *reversibly* at fixed V
  - the change in entropy is $dS = dQ/T = Vdu/T$
  - but $du = d(aT^4) = 4aT^3dT$ and so
  - $\Delta S = 4aV \int_{T_1}^{T_2} T^2 dT = (4/3)aV[T^3]_{T_1}^{T_2}$
  - this behaves sensibly in the limit that $T_1 \rightarrow 0$
    * in which limit the number of photons, and hence also the statistical mechanical entropy, vanish
  - so we can conclude that
    * $\boxed{S = (4/3)aVT^3}$
  - and that the entropy density is
    * $\boxed{s = (4/3)aT^3 = (4/3)u/T}$
  - which is in accord with Bose's calculation if $\beta = 1/k_{\rm B}$
  - and entropy of radiation $S \sim k_{\rm B}N_\gamma$ with N_γ the number of photons

- Next, consider what happens if we move the piston out slowly (i.e. reversibly) at constant S
  - i.e. without adding or removing heat
  - so we are considering an *'adiabatic'* transformation

- in the course of which, the radiation, exerting a pressure on the piston, will do work

- lets *assume*, first of all, that under such a change the number of photons N_γ remains constant
  - if so, the number *density* n_γ must decrease as $n_\gamma \sim V^{-1}$
  - and, since $n_\gamma \propto T^3$ for a black-body, that means that the temperature falls as $T \propto V^{-1/3}$
    * which in turn means that the characteristic wavelength $\lambda_\star$ must scale as the cube root of the volume
    * or, in cosmology, that the wavelength of light scales with the expansion of the universe
  - but $u \propto T^4$, so $u \sim V^{-4/3}$, which implies that the energy in the cylinder $E = Vu$ must scale with volume as $E = \alpha V^{-1/3}$ (where α is some constant)
  - and therefore that $dE = -(1/3)\alpha V^{-4/3}dV = -(1/3)(E/V)dV$

- on invoking the first law of thermodynamics: $dE = dQ - PdV$
  - which here (since $dQ = 0$) says $dE = -PdV$

- we conclude that the pressure is $P = ((1/3)E/V = u/3$

- But we already know that this is correct.
  - So the *assumption* – that the number of photons does not change under a reversible change of volume – is therefore valid

- Another way to express the fact that $P \propto u \propto V^{-4/3}$, using the language of classical thermodynamics, is to say that the *'adiabatic index'* for a gas of photons is $\gamma = 4/3$
  - as compared to a monatomic gas which has $P \propto V^{-\gamma}$ with $\gamma = 5/3$

- So different physical pictures ('gas' of massless particles or a thermal system coupled to environment via heat flux and physical work) are consistent with one another.

- and invoking the equivalence of mass and energy from special relativity
  - $E = mc^2 \Rightarrow u = \rho c^2$

- we have what cosmologists call the *'equation of state'* for radiation:
  - $\boxed{P = \rho c^2/3}$.