Lecture Notes: Physical Chemistry in Outer Space

Diethelm Johannsmann, Institute of Physical Chemistry, TUC

Last updated: 4/27/2025

1 Motivation 3
2 Philosophical remarks borrowed from Binney 3
2.1 In heaven as on earth 3
2.2 If it can happen it will 4
2.3 It must hang together 6
3 The heat death is not our fate – it is behind us (almost so) 7
3.1 Cosmic expansion and black holes create a large sinks for entropy 7
3.2 Recap: Black-body radiation 7
3.3 The microwave background: An unstable equilibrium 8
3.4 Advanced topic: Polarization of the microwave background, primordial gravitational waves 10
4 Deterministic chaos 12
4.1 The three-body problem leads into deterministic chaos 12
4.2 Robust and less robust sets of parameters, relevance concepts 13
4.3 Mean-field theories 15
4.4 Transfer of momentum between Venus and Earth as a consequence of instantaneous correlations 15
5 Statistics of self-gravitating systems 18
5.1 Recap: The entropy reaches a maximum at thermodynamic equilibrium 18
5.2 Recap: Thermal equilibrium 21
5.3 The internal energy is not an extensive parameter in self-gravitating systems 22
5.4 The virial theorem 24
5.5 No thermodynamic equilibrium for self-gravitating systems, mass segregation 26
5.6 The gravothermal catastrophe 27
5.7 The Jeans collapse 28
5.8 Local thermodynamic equilibria 29
5.9 Background: The photon gas 30
5.10 Baryonic acoustic oscillations 34
5.11 The biosphere needs the dark and cold sky 35
5.12 Cosmic expansion is an arrow of time 36
6 Rotating self-gravitating systems 37
6.1 The streaming instability in the protoplanetary disks 37
6.2 Advanced Topic: Exoplanets 39
6.3 Accretion disks 40
6.4 Advanced topic: Digital imaging with coherent detectors, the pair correlation function 41
7 Black hole thermodynamics 43
8 Degenerate matter and the Chandrasekhar limit 45
8.1 Recap: The uncertainty relation 45
8.2 Degenerate matter and its behavior under compression 47
8.3 Advanced topic: Neutron stars 51
8.4 Advanced topic: Supernovae 52
9 Space is pervaded by magnetic fields and plasma 55
9.1 Different types of plasma 55
9.2 The intergalactic medium 56
9.3 Consequences of magnetohydrodynamics 57
9.4 Strong magnetic fields around neutron stars 59
9.5 Advanced Topic: Synchrotron radiation 60
10 Neutral gas, molecular clouds 62
10.1 Gas in the galaxies 62
10.2 Advanced Topic: The 21-cm line 63
10.3 Molecules in space 65
11 Peculiarities of spectroscopy in astrophysics 66
11.1 Forbidden lines are never completely forbidden 66
11.2 Recap: Rotational spectroscopy 67

11.3 CO rotation temperature ... 67
11.4 Cosmic masers ... 68
12 Condensed matter ... 70
12.1 Cosmic dust ... 70
12.2 Meteorites ... 71
12.3 Recap: The partition sum and its relation to free energy ... 72
12.4 Enrichment of heavy isotopes in condensed phases ... 72
13 Dark matter ... 75
14 Neutrinos ... 77
14.1 Neutrino oscillations prove that neutrinos have a mass ... 77
14.2 Neutrino mixing may explain the imbalance between matter and antimatter in the universe ... 78
15 The anthropic principle in cosmology ... 80
15.1 Fine tuning of the fundamental constants ... 80
15.2 Cosmic inflation ... 82
15.3 A time dependence of the fine structure constant? ... 84
15.4 A time dependence of Λ? ... 87
16 Appendices ... 88
16.1 The role of the fine structure constant ... 88
16.2 Proof of Eq. 5.14 (the virial theorem applied to the time-averaged energy and potential) ... 89
16.3 Thermodynamic variant of the virial theorem, equipartition theorem ... 90
16.4 Proof of Eq. 12.1 (partition sum and free energy) ... 94
16.5 Topics from physical chemistry touched upon in these notes ... 95

1 Motivation

Physical chemistry is inherently cross-disciplinary, with its principles extending into various fields, including astronomy. When we look up, the same laws we see are the same as the ones governing the world around us. Actually, the cosmos a bit richer than the laboratory. The same laws apply, but the ways in which they are materialized differ. Example are the cosmic microwave background, globular clusters of stars, and black holes.

This course targets topics of particular interest from a physical-chemistry perspective. Among the questions explored are:

- To what extent must thermodynamics be modified to accommodate cosmic phenomena?
- What mechanisms drive structure formation in the universe?
- What are analogies and differences between the processes in outer space and in the laboratory?

This course is not intended as an astrophysics lecture. While there are mentions of topics such as stars and galaxies, these are brief. For a good introduction to astrophysics, see the book "Astrophysics: A Very Short Introduction" by James Binney.[1] Binney is mentioned so often that the reference to footnote 1 is occasionally omitted.

There are reminders on physical chemistry in general, but these are really brief. Looking things up in Wikipedia may be required. Also, complications often are ignored. That was a choice, which may in some instances be problematic, admitted.

Disclaimer: These notes are intended as a guide for the course, not a textbook. As lecture notes are subject to change, inaccuracies creep in here and there. Feedback is appreciated (johannsmann@pc.tu-clausthal.de).

2 Philosophical remarks borrowed from Binney

2.1 In heaven as on earth

The insight that the same laws apply everywhere is often attributed to Newton. He sat in a garden and saw an apple falling to the ground. For Newton, the apple in free fall and the planets orbiting the sun adhere to the same principles – specifically, the equation of motion defined by "force is mass times acceleration," with the force in both cases being gravity. Newton's view is widely accepted today.[2] A striking success was the discovery of helium (see Fig. 2.1).

Binney begins his book with a quote from the bible. In John 1:1 it says: "In the beginning was the Word". Binney claims that the laws of nature were above the material world. That is an opinion. The laws might be spatiotemporal patterns in our minds and sequences of symbols in our books. Undoubtedly, these patterns and sequences exist. However, we remain uncertain whether they constitute a preexisting truth or whether they are part the world we aim to describe, much like Newton's falling apple.[3]

It is widely presumed that elementary logic applies everywhere. "1 + 1 = 2" holds true on Aldebaran, even if the means to test that remain unclear. Moreover, there is little evidence to suggest that the

---

[1] Binney, J. *Astrophysics: A Very Short Introduction,* Oxford University Press 2016

[2] At the beginning of the 20th century, scientists believed that the atom resembled a miniature planetary system. This "planetary model" was later refined by Bohr. The planetary model, however, misses that electrons are waves. The behavior of the 1s electron in the hydrogen atom does not parallel that of the planet Mercury.

[3] Binney adheres to the mind-body duality. He quotes the bible but this particular quote was influence by a hellenistic frame of mind. The mind-body duality was part of Platon's philosophy.

fundamental structure of the laws of nature was different in outer space. Astrophysical processes are successfully modeled with partial differential equations, similar to what is done in weather forecast.[4] However, the values of the fundamental constants may be different far away from Earth or may have been different in a distant past (sections 15.3 and 15.4). A somewhat separate topic is the applicability of thermo-

Box 2.1: **Discovery of helium in the solar spectrum**
The discovery of helium is an example of how the universality of the laws continues to astonish people. Towards the end of the 19th century, our understanding of nature was perceived as more or less complete – with the exception of a few clouds in the sky of science that were later given the name "quantum mechanics".
Line spectra were known (and understood later in the context of quantum mechanics) and the absorption lines in the spectrum of the sun (Fig. 2.1) matched the emission lines from the flames seen exposing salts to a flame burner (a "Bunsen burner"). A perfect match? Not quite: There was a set of absorption lines in the spectrum of the sun that could not be assigned. A new element was postulated and named helium (from "helios", Greek for "sun"). Later, this element was also found on Earth. Helium was discovered on the sun, but apart from that it has just as much to do with the sun as copper and lead. The periodic table is universal.

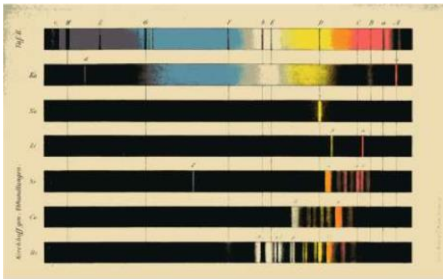


Fig. 2.1
Kirchhoff and Bunsen: "Chemische Analyse durch Spectralbeobachtungen" in: Annalen der Physik und Chemie, Vol. 110, No. 6, 1860, pp. 161-189.
Newly published by Gabriele Dörflinger, Heidelberg University Library. *www.haus-der-astronomie.de/3642696/04strahlung.pdf*

dynamics to astrophysical processes. While thermodynamics is not tied to specific physical laws, it has prerequisites that are not always fulfilled. These prerequisites also exist on Earth, but the limits of thermodynamics are more impressive in astrophysics (section 5.3).

2.2 If it can happen it will

Even if the laws of nature are the same everywhere, in the sky offers more to see than the earth. The argument goes back to Fritz Zwicky.[5] He may have thought of the neutron stars, which had been postulated before they were found (section 8.3). Here is the quote from Binney: "In the 1930s the eccentric Swiss astronomer Fritz Zwicky restored the primacy of the heavens to some extent by asserting that 'if it can happen it will'. That is, anything that is permitted by the laws of physics will happen somewhere in the Universe, and with the right instruments and a bit of luck we can see it happening. Zwicky's Principle indicates that it is profitable to think hard about what weird objects and exotic events are in principle."

The following objects and phenomena had been postulated before they were found:
- Black holes
- Neutron stars
- Neutrino flashes produced during collapse supernovae
- The cosmic microwave background (CMB)
- The 21-cm line from radio astronomy, caused by the spin flip of the proton in the H-atom
- Extremely high magnetic fields
- Gravitational waves

[4] The ordinary differential equations were developed by Newton *with the purpose of* modeling celestial mechanics.
[5] The sentence „If it can happen, it will" is attributed to Zwicky, but he never actually wrote it down. There is no source to be quoted.

The following phenomena or objects have not been seen, although the standard model is open to them:

- Magnetic monopoles
- Worm holes, i.e. solutions to Einstein's field equations that would allow a journey into the past.[6]

The following phenomena were not on anyone's radar before they were found:

- Cosmic expansion following Hubble’s law
- Dark matter and dark energy
- Quasars
- Gamma ray bursts.

Astrophysics is special in the following regards:

- In interstellar space, the collision rates between particles are lower than on Earth. The particles have up to 100 years before they collide for the first time.[7] If a molecule is in an excited state and if radiative deexcitation occurs across a “forbidden” transition,[8] collisional deexcitation will occur on Earth, so that the very slow optical emission is not observed. In space, such lines become visible because the optical emission does eventually occur (section 11.1).
- It is much hotter inside stars than on Earth (center of the sun: $15 \cdot 10^6$ K). Nucleosynthesis takes place in the center of the sun, but that does not include elements heavier than iron and nickel. Temperatures of 10^{10} K are reached in the shock fronts of supernovae. The heavy elements are produced there.[9]
- The cosmic rays occasionally contain particles with an energy of 10^{20} eV. This is 10^8 times more than what the Large Hadron Collider (LHC) in Geneva achieves. If those were thermal particles, the temperature would have to 10^{24} K. There is talk of moving „magnetic bottles“ at shock fronts. In this scenario, particles repeatedly collide with the walls of these bottles, gaining energy with each impact due to the movement. What accelerates them is an interplay between electric and magnetic fields, similar to what is done in the particle accelerators.
- The force of gravity on neutron stars is 10^9 times stronger than on Earth. Inside a neutron star, the density is 10^{14} times larger than the density of rock in the earth's crust. The pressure in the center is larger than 10^{25} atmospheres.
- The magnetic field in the vicinity of a neutron star is between 10^4 and 10^{14} Tesla (section 9.4). For comparison: NMR magnets reach several Tesla.
- There are large-scale magnetic fields in interstellar space.
- There are gravitationally bound, rotating disks (section 5.12) in several variants. Saturn's rings are rotating gravitating disks, but the interactions between the particles is weaker than in other – and more interesting – disks of that kind. In the "accretion disks", friction heats up the material.

---

[6] It is argued that worm holes should be unstable for thermodynamic reasons. The problem with this argument is that there is no straight-forward way to apply thermodynamics to space-time curvature. See the end of section 3.1.

[7] A density of 1 particle per cm^3 and a temperature of 100 K were assumed, *www.astro.princeton.edu/~burrows/classes/204/ISM.pdf*
The density of the intergalactic medium (a plasma) is about 1 particle/m^3.

[8] "Forbidden" here means that the electric dipole transition is forbidden. The other types of transitions (such as the magnetic dipole transition) are less efficient.

[9] Temperatures even higher than what is reached in supernovae are reached on the earth, when heavy nuclei are made to collide in particle accelerators. Nuclei smash into each other and form a loose aggregate for about 1 ns. The energy is so high that the protons and neutrons disintegrate. They turn into a quark-gluon plasma. Even though this state is short-lived, it is said to be close to an equilibrium with a temperature of 10^{12} K. Nowhere in today’s universe is the temperature that high. Nowhere in the universe do protons and neutrons disintegrate to form a quark-gluon plasma.

– The kinetic energy of material in an accretion disk around a stellar-size black can amount to sizeable fraction of mc^2 with m the rest mass. This happens shortly before the material crosses the event horizon. Nowhere else do macroscopic bodies have such energies.[10]

Structure formation is of some importance to our understanding of the world because structure is a prerequisite for complexity. Paradigmatic of spontaneous structure formation is the biomembrane, formed by a process called "self-organization". Structure formation processes in space include[11]

– the gravothermal catastrophe (section 5.6)
– the Jeans collapse (section 5.7)
– the baryonic acoustic oscillations (section 5.9)
– the streaming instability (section 6.1).
– the spontaneous generation of magnetic fields (section 9.1)
– magnetic flux tube concentration (section 9.1).

The following phenomena are well known from earth but have no counterpart outer space:

– Nowhere in the sky is it much colder than 2.7 K because this is the temperature of the microwave background.[12] There is no superconductivity of electrons (but see footnote 68).
– There are no high-purity materials anywhere (such as the high-purity silicon needed for computer chips, defect density is about 10^{11} cm^{-3}).
– Nowhere is there as much complexity as on Earth.

Of course, this list ignores extraterrestrial civilizations.

2.3 It must hang together

In a chapter entitled "It must hang together", Binney emphasized that the laws of physics not only describe astrophysics and the atoms equally well, but that they also do so in a transparent and – in all due caution – a simple way. Magnetism and electricity describe a large range of phenomena and they are parts of a common theory of electromagnetism, the basis of which can be condensed into four equations (Maxwell's equations). The search for a simple model guided Einstein when he formulated general relativity. The standard model of physics is not simple in the narrow sense, but it can still be can be formulated in compact form. In particular, there is a common formalism (the "gauge theories") for quantum electrodynamics, the weak interaction, and the strong interaction.

Still: The "Weltformel", which Einstein had hoped for and dreamed of, has remained a dream. Also, the standard model of physics is incomplete. There must be a theory of quantum gravity, but it is unclear what it should look like. The big picture exists, but there are gaps, open ends, and peculiarities whose simplicity and beauty are not obvious (an opinion). Furthermore, some fundamental constants (all of them?) may have taken their values by chance and may be different outside our small part of the universe (section 15.2).

---

[10] Elementary particles in accelerators have energies much larger than mc^2.

[11] With the exception of magnetic flux tube concentration, these mechanisms are linked to a non-equilibrium. Within the framework of non-equilibrium thermodynamics, they can be referred to as "dissipative structures". Self-organized structures, on the other hand, exist in equilibrium.

[12] There are regions in the Boomerang nebula, where the "CO rotation temperature" (section 11.3) is 1 K, which less than the temperature of the cosmic microwave background. The Boomerang nebula has seen adiabatic expansion and adiabatic cooling. It is a cosmic refrigerator. Much better refrigerators have been created in the laboratory. Temperatures below 1 mK have been reached.

The principle that simple theories are good theories is „Ockham's razor" (after William of Ockham, 1287–1347). Ockham's razor has often guided the search for fundamental laws, but there is no a priori reason, why the fundamental laws should be simple. Recently, the rules found with artificial intelligence have provided numerous counter-examples to Ockham's razor. These are not at all simple but still have tremendous predictive power. Another example of less-than-perfect simplicity are the elliptical orbits of the planets. When Kepler discovered those, Galilei rejected them because they appeared unnecessarily complicated to him. The underlying, more abstract and more generally applicable model – Newton's laws – restored simplicity. Newton's laws, however, allow for elliptical orbits.

3 The heat death is not our fate – it is behind us (almost so)

Sections 3 to 5 discuss thermodynamics in astrophysics from various perspectives. The title of section 3 is intended to sound optimistic. Clausius assumed that the sad final state of this universe would be a uniformly warm soup in thermodynamic equilibrium.[13] (This expectation is believed to have contributed to the depression that plagued Boltzmann in his final years.) However, the universe was in a state close to the heat death 380 000 years after the Big Bang. It consisted of a largely homogeneous plasma with a temperature of around 3000 K. The universe started from the heat death and became alive. Clausius' view is outdated.

3.1 Cosmic expansion and black holes create large sinks for entropy

A prime reason for the universe departing from the state, which appears to have been close to the heat death, is its continued decrease in temperature. Structure formation upon cooling is not surprising. Every time water is cooled to below 0 °C, crystals form.

The sky expands and as new space is continually created, entropy can be dispersed into it. Less well known is the role of black holes (section 7). Following Eq. 7.1, the entropy of a black hole is $S_{BH} = k_B A/(4l_p^2)$. A is the surface of the black hole (the "event horizon") and l_p is the Planck length (Box 15.2). For a black hole with a Schwarzschild radius of 30 kilometers (typical of a stellar-mass black hole) the entropy is about 10^{79} k_B. This corresponds to about 2×10^{78} photons, using that each photon carries an entropy of 3.6 k_B (see the text next to Fig. 5.9). Throughout its lifespan, the sun emits about 4×10^{61} photons. That does not necessarily mean that the same amount of entropy indeed disappears from the outside into the black, but it might, in principle.

<u>Remember</u>

- Structure formation in a cooling universe is not surprising.
- Reasons for global cooling are the expansion of the universe and how the black holes swallow entropy.

3.2 Recap: Black-body radiation

Black-body radiation and its explanation by quantization ignited the development of quantum mechanics at the beginning of the last century. We repeat some essentials.

A "black body" absorbs and emits all wavelengths so efficiently, that is does not have a color of any kind. It has no optical properties, which would depend on the material it is made of. Black bodies glow

---

[13] As things stand today, the universe will not end in the heat death, but rather in the "Big Freeze". The time scales involved requires some critical reflection. Extrapolation of the laws of physics to cosmic dimensions works amazingly well. Whether these laws may be extrapolated to time scales of $10(\wedge10^{130})$ years, may be doubted. Those are playful exercises (an opinion). *en.wikipedia.org/wiki/Future_of_an_expanding_universe*

in the dark and because the material does not matter, the spectrum of the emitted light is universal. It is a function of temperature only. Universal laws of this kind always intrigue researchers.

At some point, Rayleigh and Jeans succeeded in explaining the long-wavelength part of the black-body spectrum with thermodynamic arguments, alone. The arguments are of the most elementary kind and they match experiment – on the long-wavelength tail. Rayleigh and Jeans took a somewhat daring approach insofar, as they claimed that electromagnetic waves should obey the laws of thermodynamics. In thermodynamics, there is the equipartition theorem (section 16.3), which predicts that the time-averaged energy contained in any excitation, which contributes quadratically to the overall energy, is equal to 1/2 $k_B T$. Among the known excitations of this kind were the translation, the rotation and the vibration of molecules. Electromagnetic waves also carry an energy. They can be decomposed into "modes" (similar to how an arbitrary vibration of a molecule can be described as a superposition of modes) and each mode should carry an energy of $k_B T$. It is $k_B T$ rather than 1/2 $k_B T$ because the electromagnetic modes contribute to the energy with an electric field as well as a magnetic field. The modes are the standing waves. They can be labeled and counted based on the number of nodal planes. The mathematics is a bit involved, but at the end the agreement with experiment was just marvelous – on the long-wavelength tail. That the short wavelengths behaved differently was later explained with the quantization of electromagnetic fields in the form of photons ($E = nh\nu$). For now: Bodies, which heavily absorb light, emit a characteristic spectrum, which only depends on temperature.

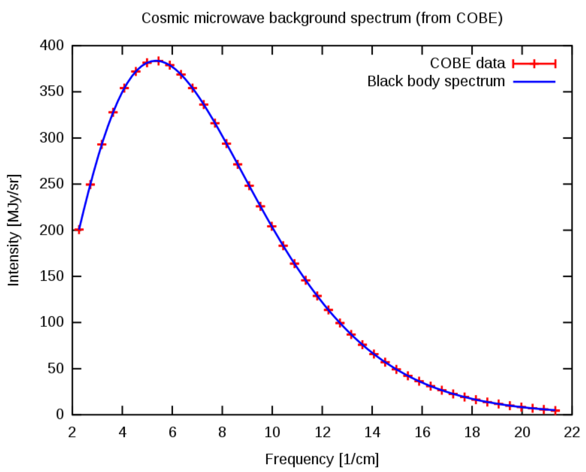


Fig. 3.1
Spectrum of the cosmic microwave background (CMB)
en.wikipedia.org/wiki/Cosmic_microwave_background

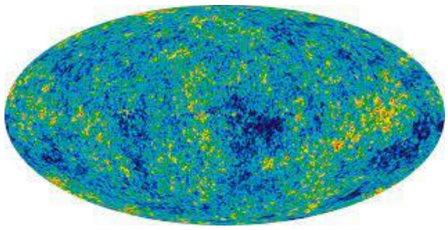


Fig. 3.2
A map of the cosmic microwave background. Color encodes temperature.
www.esa.int/Science_Exploration/Space_Science/Planck/Planck_reveals_an_almost_perfect_Universe

3.3 The microwave background: An unstable equilibrium

The apparent (and unstable) state of equilibrium at the beginning of the universe is visible to us in the form of the cosmic microwave background (CMB). The CMB is the thermal radiation of a plasma. In a plasma, charged particles strongly interact with electromagnetic waves. In equilibrium, the temperature of the photons (calculated from the black-body spectrum) is the same as the temperature of the particles (calculated from the Maxwell distribution of their velocities).[14]

[14] In technical plasmas the different components (photons, electrons, ions) often have different temperatures. They are not in a thermal equilibrium. The CMB was (almost).

As ions and electrons recombined to form neutral atoms, the universe became transparent. The photons we detect today were emitted when the plasma stopped glowing. Since then, these photons have traveled through space largely undisturbed. The radiation temperature seen by us is $T = 2.725$ K. This low value results from the plasma's receding motion, leading to a large Doppler shift of all wavelengths. The redshift is $\Delta\lambda/\lambda = z = 1089$. Without the redshift, the peak in spectrum at $\tilde{\nu} = 5\ \text{cm}^{-1}$ in Fig. 3.1 would be a peak at $\tilde{\nu} = 5000\ \text{cm}^{-1}$, which corresponds to $\lambda = 2.5$ µm The radiation temperature would be ~3000 K.

The CMB is among of the astrophysical phenomena that were predicted before they were discovered. Penzias and Wilson were unaware of the prediction. They had built a highly sensitive microwave antenna and found a persistent noise, independent of the antenna's orientation. It is said that Penzias and Wilson never quite understood implications of their finding, even after having been awarded the Nobel Prize.

Figure 3.1 shows the spectrum of the CMB with a fit to the Planck curve. Until the early 1990s, the CMB was considered isotropic. It took much effort to measure the small variations, but this small anisotropy today is a cornerstone of cosmology. The radiation temperature fluctuates by about ±18 µK, the latter being the standard deviation. Fig. 3.2 shows the famous map. The yellow islands are not all the same size, but the sizes are not randomly distributed, either. This impression is quantified by plotting the spectral power density against the multipole order (Fig. 3.3). A complication: Because this temperature distribution lives on a sphere rather than a plane, the temperature distribution is not Fourier decomposed (i.e. not represented as a sum of sines and cosines), but is decomposed into spherical harmonics $Y_{l,m}(\theta,\varphi)$ with l multipole order.[15] Fig. 3.3 shows the squared absolute values of the weights in units of µKelvin^2. On the right in Fig. 3.3 are the small-scale structures (the high multipole components, the analog of the high Fourier components). On the left are the large-scale (more precisely: large-angle) structures.

The power spectrum of the CMB shows characteristic peaks corresponding to the preferred sizes of the yellow islands in Fig. 3.2. These are caused by the baryonic acoustic oscillations (BAOs, section 5.9).

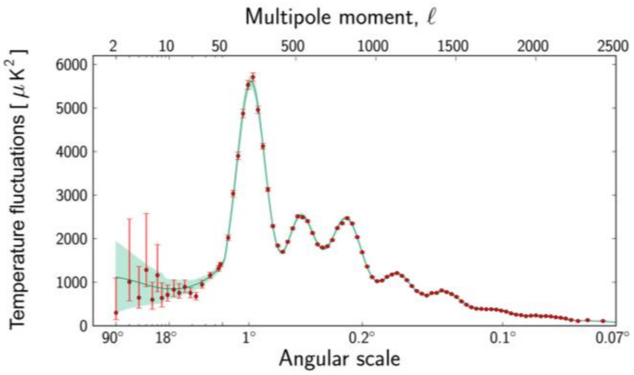


Fig. 3.3
Power spectrum of the map from Fig. 3.2. The entries are the squares of the multipole moments (the analog of the Fourier components for angles)
https://www.mpg.de/7044245/Planck_cmb_universe

Remember

– The CMB is the radiation from a plasma that was close to thermodynamic equilibrium. The radiation temperature was about 3000 K.

[15] Some readers may remember the spherical harmonics as the eigenfunctions to the L^2-operator in quantum mechanics. The angle-dependence of the p-orbitals, the d-orbitals, and the f-orbitals of the hydrogen atom is given by the spherical harmonics with $l = 1$, $l = 2$, and $l = 3$, respectively. That has nothing to do with the CMB. Here, the spherical harmonics are just a convenient set of basis functions for the decomposition of the CMB into components, similar to the Fourier decomposition.

– There are small fluctuations in temperature, caused by the baryonic acoustic oscillations (BAOs).

3.4 Advanced topic: Polarization of the microwave background, primordial gravitational waves

It is believed that the fluctuations in the CMB originate from “vacuum fluctuations” in the early universe. These may have been scalar fields (like the Higgs field), vector fields (like light) or tensor fields (like gravitational waves, in this case called "primordial gravitational waves"). The tensor fields can be recognized as such because they may violate parity symmetry (the symmetry under a point reflection at the coordinate origin). As a consequence, a map of the polarization of the CMB may contain vortices. Vortices have a handedness (clockwise or counterclockwise).

How does a polarization of the CMB come about? It may come about by scattering of the light by the electrons and ions ("Thompson scattering"). If scattering occurs under an angle of 90°, the scattered light is polarized perpendicular to the primary beam. For the same reason, the blue sky is polarized. The sky is blue because photons from the sun experience Rayleigh scattering. If the intensity of the CMB fluctuates, the CMB from regions next to the brighter regions is slightly polarized. According to this explanation, the source of polarization is a scalar field (a fluctuation of density, pressure or temperature). Such scalar fields can never generate vortex fields (of whatever) for reasons of symmetry.

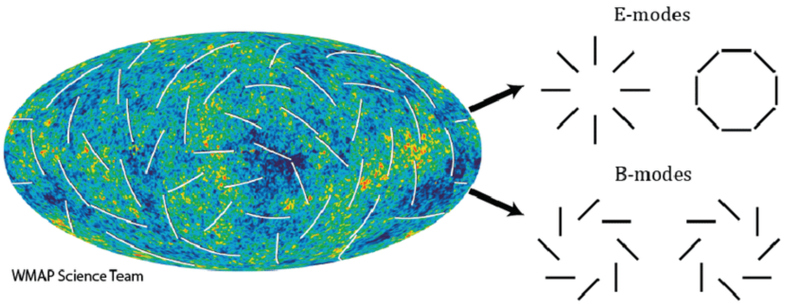


Fig. 3.4
The CMB polarization map may be decomposed into curl-free even-parity E-modes and divergence-free odd-parity B-modes. Primordial B-modes are only created by tensor perturbations (inflationary gravitational waves).
www.researchgate.net/figure/The-CMB-anisotropy-polarization-map-may-be-decomposed-into-curl-free-even-parity-E-modes_fig1_263811882

Gravitational waves are described by a 2nd-rank tensor (by the metric tensor $g_{\mu\nu}$). In mathematics, a distinction is made between vectors (such as the electric field or the velocity) and "axial vectors" (such as the magnetic field or the vortex density of a velocity field). Axial vectors have three components and can be represented as arrows. However, unlike normal vectors, they do not change direction when space is reflected at the origin. The axial vectors result from a cross product. Write the magnetic field as the rotation of the vector potential ($\vec{B} = \vec{\nabla}\times\vec{A}$) or write the vortex density (the "vorticity", ξ) as $\vec{\xi} = \vec{\nabla}\times\vec{v}$. The x-component of $\vec{B}$ is $B_x = \partial_y A_z - \partial_z A_y$.[16] Under reflection at the coordinate origin, both $\vec{\nabla}$ and $\vec{A}$ change sign. $\vec{B}$ then does not. Neither does the vorticity.

To avoid the strange behavior of axial vectors, consider the 2nd-rank tensor $\vec{\nabla}\otimes\vec{A}$ (with two indices, given as $(\vec{\nabla}\otimes\vec{A})_{ij} = \partial_i A_j$). This tensor has 9 entries. Such tensors can be split into a symmetric and an antisymmetric part:

$$(\vec{\nabla}\otimes\vec{A})_{ij} = (\vec{\nabla}\otimes\vec{A})_{sym,ij} + (\vec{\nabla}\otimes\vec{A})_{anti,ij} = \frac{(\vec{\nabla}\otimes\vec{A})_{ij} + (\vec{\nabla}\otimes\vec{A})_{ji}}{2} + \frac{(\vec{\nabla}\otimes\vec{A})_{ij} - (\vec{\nabla}\otimes\vec{A})_{ji}}{2}$$

---

[16] ∂_x is short for ∂/∂_x.

The antisymmetric part contains zeros on the diagonals and 6 elements off the diagonals, but these are antisymmetric when indices are swapped ($(\vec{\nabla}\otimes\vec{A})_{anti,ij} = -(\vec{\nabla}\otimes\vec{A})_{anti,ji}$). Three independent parameters remain, which can be written as a vector (i.e. as the cross product) if desired. $\vec{B}$ and $(\vec{\nabla}\otimes\vec{A})_{anti}$ are related as $B_x = (\vec{\nabla}\otimes\vec{A})_{anti,yz}$, $B_y = (\vec{\nabla}\otimes\vec{A})_{anti,zx}$ and $B_z = (\vec{\nabla}\otimes\vec{A})_{anti,xy}$.

Because $g_{\mu\nu}$ is a tensor, it can have an antisymmetric part. The antisymmetric part can create vortices in the CMB polarization map. These vortices also go by the name of "B-modes", to be distinguished from "E-modes" (Fig. 3.4). So far, no one has been able to nail down the B-modes.

4 Deterministic chaos

The molecules in a gas move randomly (it seems), although the underlying dynamic equations are deterministic. The dynamic laws of microscopic physics all are deterministic. "Deterministic chaos" was discovered in celestial mechanics, just like Newtonian mechanics. The hero was Poincaré.

4.1 The three-body problem leads into deterministic chaos

The solution to the two-body problem (Kepler, Newton) was a transformative success. If you want to solidify a success, you have to move on and apply the principles to other problems. A natural next step in this case leads from the two-body problem to the three-body problem. This turned out to be difficult, where "difficult" is a misleading term from today's perspective. That the problem is so difficult is far more important than the ways in which clever people have in some cases tamed it.

We encounter the three-body problem in the movement of asteroids. These are attracted by all other bodies of the planetary system, but the sun and Jupiter play the largest role, hence the name "three-body problem". The asteroids are the good-natured cases in that there is a clear hierarchy (Sun >> Jupiter >> asteroid). Certain approximations can be made and these lead to good approximate solutions. Some non-experts may be familiar with the 2^{nd} Lagrange point (L_2). The 2^{nd} Lagrange point lies behind the earth on the axis formed by the sun and the earth. The distance from the Earth is just right for a body located at the 2^{nd} Lagrange point to remain there. The James Webb space telescope lives and works at the 2^{nd} Lagrange point of the Earth-Sun system.

Again, there are good-natured and less good-natured cases. The theory of the three-body problem stumbled along somewhat joylessly. In 1885, the editors of the journal *Acta Mathematica* wanted to tick the matter off and announced an award for its comprehensive and final treatment. It was foreseeable that the award would go to Poincaré, as Poincaré was the leading figure in the field. Poincaré sent something in and received the award. Before his contribution went to print, the editors had a few questions. There were a few gaps and as Poincaré tried to fill those, the hole got deeper and deeper. At this point Poincaré deserves credit. He did not quietly turn away; busying himself with more transparent problems, but rather kept chewing on this one. Years later, it became clear that the three-body problem *cannot* always be solved. Sometimes, the three-body problem leads into chaos.[17] Chaos is more important for our view of the world than the 2^{nd} Lagrange point. If we know the positions and momenta of three bodies at a certain time, we can always calculate the orbit for the next year – or even for the next 40 million years[18,19] – numerically. The question is what happens in the long term. In the case of the three-body problem, this question sometimes cannot be answered.

The fundamental dynamic laws of physics are deterministic. This applies to $F = ma$, it applies to Maxwell's equations (to electrodynamics), to the Schrödinger equation (to quantum mechanics), and to the standard model of particle physics. Anyone who – first – knows the positions and velocities of the sun, Jupiter and Ceres with infinite precision and who – second – can calculate with infinite precision, can also calculate the orbit of Ceres for all future (and for all past as well). Neither of these two conditions are met. A small inaccuracy in our knowledge of the current positions and velocities will turn into a

---

[17] Because of the chaotic behavior there are many binary stars, but few triple stars. If three bodies with similar masses orbit each other, one of them will be ejected from the group sooner or later, leaving behind a (stable) two-body system.

[18] According to Binney, the current accuracy in the positions and momenta of the planets suffices to predict their motion for the next 40 million years. At this time the calculations hit a wall. If the time stretch is supposed to be 60 million years, the measurement accuracy must improve by a factor of 10. If it is to be 80 million years, the accuracy must improve by another factor of 10. The relation between the prediction span and the required accuracy is exponential.

[19] The time, after which the predictions become unreliable in qualitative terms, is called Lyapunov time.

large uncertainty in 40 million years. "Large uncertainty" means that the orbits may be highly elliptic or that a planet may have left the solar system.[20] In the words of Edward Lorenz: "[Chaos is realized when] the present determines the future, but the approximate present does not approximately determine the future." Random behavior does not mean that the deterministic dynamic laws would not hold. Random behavior means that thinking beings cannot anticipate the future. "Thinking beings" here includes robots of the most advanced kind. A demon capable of foreseeing the entire future of the universe would need a representation of the entire universe inside its computational machinery.

There is chaos in molecular physics, as well. Molecules in liquids move around chaotically. Even single molecules in the gas phase can vibrate in chaotic ways if they are highly excited. In particular, the vibrational energy may wander back and forth between the different modes. The vibrational energy then is distributed over many modes. It is said to be "thermalized".

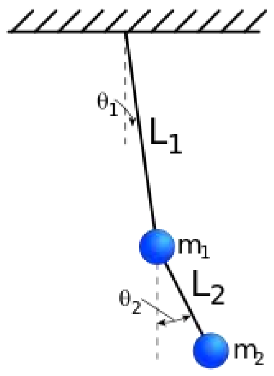


Fig. 4.1
The double pendulum exhibits chaotic motion
en.wikipedia.org/wiki/Double_pendulum

The analogy has limits:

- Molecules are subject to quantum mechanics. Quantum chaos differs from classical chaos.
- Molecules behave less chaotically when the excitation energy is low. The potentials then are approximately parabolic. The restoring force is largely proportional to the excursion from the state of minimum energy. Such linear systems do not go into chaos. In gravitation there is no such parabolic low-energy limit. The force of gravitation never is proportional to distance.

  The transition to chaos at high excitation energies, only, can be illustrated with the double pendulum (Fig. 4.1). The double pendulum has only two degrees of freedom (the two angles). At small angles, the restoring torques are almost proportional to the angle because of $mg \sin(\theta) \approx mg\,\theta$. The term mg is the force of gravity. The double pendulum then oscillates regularly. There is a symmetric and an antisymmetric mode (same or opposite sign of both angles). Chaos only occurs at larger amplitudes.

Remember:

- For chaotic systems in the sense of chaos theory, the trajectories critically depend on the initial conditions and are unpredictable in the long-time limit.
- In deterministic chaos, the present determines the future, but the approximate present does not approximately determine the future.

4.2 Robust and less robust sets of parameters, relevance concepts

At this point different cases must be distinguished. First, open systems differ from closed systems. Open systems often approach a fixed point or a limit cycle (left and center in Fig. 4.2). An example is the double pendulum, which dissipates energy in the form of heat and comes to rest. "Openness" here is created by dissipation. Energy leaves the system in the form of heat. Another example are small clusters of

[20] A side note from Binney: In the far future of our planetary system, there is a possibility that Mercury may go into resonance with Jupiter. Its orbit will then become elliptical and the elliptical orbit will destabilize the orbits of Venus and Earth. The planetary system then is in great peril. Earth may well be ejected into interstellar space. Interestingly, the probability of this happening is reduced if general relativity (GR) is included in the calculation. The effects of GR are tiny, but tiny effects do play a role in the three-body problem. Binney suspects a general wisdom here. GR makes the entrance to resonances somewhat narrower. Binney concludes: Without GR, we would not exist.

stars ("open clusters"). The pleiades are an open cluster. Open clusters lose members by evaporation until a binary star is left. These two stars steadily orbit each other forever.

A second distinction concerns whether or not the long-time limit allows to apply a "relevance concept".[21] A relevance concept identifies "macroparameters" or "macrostates", which are robust in the sense that their evolution can be predicted. The evolution of the "microparameters" or "microstates", on the other hand, cannot be predicted and is viewed as uninteresting. In thermodynamics, the positions and the momenta of the individual particles often are microparameters (right in Fig. 4.2). Volume and pressure are macroparameters. There is some freedom in what to consider "relevant". (Are sound waves in a gas relevant?)

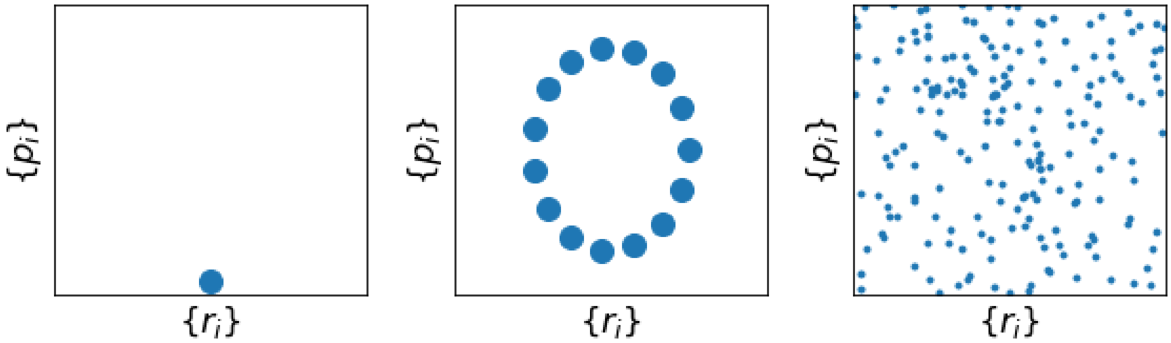


Fig. 4.2
The long-time limit of an open dynamical system can be a fixed point (left). That is the case for the pendulum, which dissipates energy to the microscopic degrees of freedom (which count as the "outside").
The long-time limit can also be a limit cycle (center). That occurs in small clusters of stars. Stars evaporate until a binary system on a fixed orbit is left over. Limiting cycles can be more complicated than what is drawn in the center. The technical term is „attractor".
In ergodic systems (right), the system visits all possible microstates (more precisely: comes close to all possible microstates) and the microstates occur with equal probability.
The dots denote microstates of the entire system at different times (as opposed to different particles). The box is the 6N-dimenional phase space, spanned by the positions and momenta of all particles.

Globular clusters of stars are an interesting model system in this regard.[22] If there are frequent "collisions"[23] between stars and if these collisions are uncorrelated between each other, a probability density of stars as a function position and momentum, $\rho_{star}(\mathbf{r},\mathbf{p})$, can be defined. $\rho_{star}(\mathbf{r},\mathbf{p})$ is "relevant". It might be called a "macroparameter", although it is a continuous function. Dealing with $\rho_{star}(\mathbf{r},\mathbf{p})$ is easier than numerically solving the N-body problem. Calling the motion of the individual stars "microparameters" is at odds with the size of stars, but otherwise the term is to the point. The evolution of $\rho_{star}(\mathbf{r},\mathbf{p})$ is governed by a partial differential equation (PDE). There are different PDEs serving this purpose. These go by the names "Langevin equation ", "Fokker-Planck equation", or "collisionless Boltzmann equation". It turns out that $\rho_{star}(\mathbf{r},\mathbf{p})$ slowly changes over time. One should write $\rho_{star}(\mathbf{r},\mathbf{p},t)$. The micro-macro distinction does not necessarily imply a thermodynamic equilibrium or a stationary state.

[21] H.D. Zeh, The physical basis of the direction of time, Springer 2007
[22] We take a somewhat idealized view of globular clusters. Real-world globular clusters feel the influence of the host galaxy. Some clusters have seen mergers in the past. In some clusters, there is cooperative dynamics. Some clusters rotate, which has consequences. We ignore all of that.
[23] The concept of collisions is made more precise in section 4.4.

Remember:

– If the collisions in globular clusters are numerous and uncorrelated between each other, a relevance concept can be applied. Robust in the sense of the relevance concept is the density of stars as a function of position and momentum, $\rho_{star}(\mathbf{r},\mathbf{p},t)$. This function is not stationary in time, though. It does not describe an equilibrium.

4.3 Mean-field theories

In principle, the preferred way to model globular clusters are the large N-body simulations. However, these are not always feasible. A first attempt to approach the problems with less computational load are the "mean-field theories". Fig. 4.3ketches the concept for globular clusters. Globular clusters contain several 100 000 stars, all of which interact gravitationally. Since 100 000 coupled differential equations are difficult to handle, an averaged potential is calculated from the time-averaged positions of all stars other than the one star, which was chosen as the test particle. The orbit of the star can be calculated from the potential. All members of cluster can be treated that way. From orbits of all stars, a time-averaged density of stars is calculated, which is input to the calculation of a new, better potential. These steps are iterated until the orbits stay unchanged within some chosen limit. Then solution then is “self-consistent”. Mean-field theory is also called "self-consistent field theory".

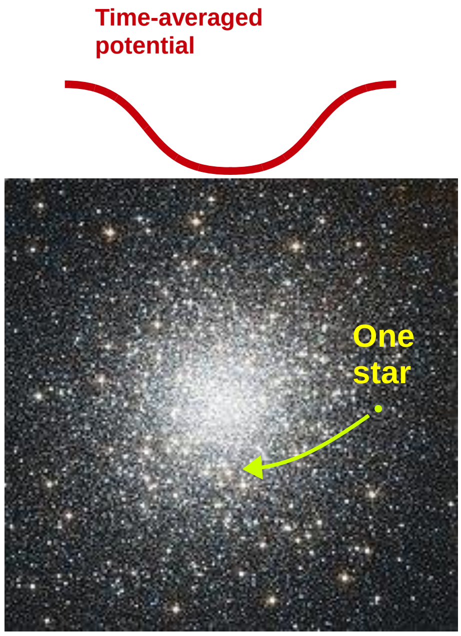


Fig. 4.3
The globular cluster 47 Tucunae is home to several million stars. In order to calculate the orbital motion of a single star within the framework of a mean-field theory, a potential is calculated from the time-averaged positions of all other stars.
en.wikipedia.org/wiki/ 47_Tucanae

The procedure is similar in the quantum mechanical treatment of many electrons: Choose one electron in the potential of the nucleus. At this point, the orbitals of the other electrons are not known and can therefore not be used to calculate a time-averaged potential accounting for the repulsion between electrons. An initial calculation of all orbitals is carried out, neglecting this repulsion. A 0^{th}-order approximation is obtained. A 1^{st}-order potential is then derived from the 0^{th}-order orbitals and the orbitals are recalculated using the 1^{st}-order potential. As before, these steps are iterated until the result has converged.

The simplest model of electrolyte solution, the Debye-Hückel theory, also is a mean-field theory. The ions are replaced by continuous time-averaged densities of positive and negative ions, ρ^+ and ρ^-. The “counter-ion cloud“ is not a cloud of many counter-ions. The counter-ion cloud is a region with increased probability to find a counter-ion.

Remember:

– In a mean-field theory, all particles are treated as single particles, interacting with some kind of field (often a potential), which represents all other particles. The field is iteratively computed from the orbits of all particles, until self-consistency is reached.

4.4 Transfer of momentum between Venus and Earth as a consequence of instantaneous correlations

A mean-field theory might also be used to estimate the influence, which Venus take on the motion of Earth. For a thought experiment, distribute the mass of Venus evenly over its orbit, as if it were a belt of dust. Replace the mass by the time-averaged mass. That makes the calculation easier, but misses an essential aspect. When Venus passes Earth, it first slows down the movement of the earth as long as it is behind Earth. Later, when it has passed, Venus accelerates Earth. These two effects do not cancel because the acceleration lasts slightly longer than the deceleration. After passing, Venus has transferred a small part of its angular momentum to Earth. This correlation effect is only noticed if the instantaneous (rather than the time-averaged) mass-distributions of the interacting bodies are considered.

There is a twist here, related to the orbits being elliptical rather than circular. After Earth got a small forward kick by Venus, its velocity is no longer equal to the Kepler velocity of its current circular orbit. Venus creates a tiny amount of ellipticity. Earth not only is kicked by Venus, occasionally, but also by the other planets. If these kicks occur randomly, the ellipticity is "damped". Earth returns to a circular orbit with a slightly increased distance from the sun. By the same mechanism, the orbit of Venus shifts inward.

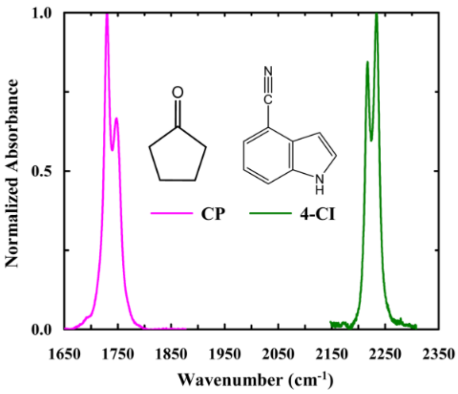
Box 4.1

"Orbital resonances" in the motion of the planets and Fermi resonances in vibrational spectroscopy

The text in this column was adapted from Wikipedia.

In celestial mechanics, orbital resonances occurs when orbiting bodies exert a regular, periodic gravitational influence on each other, usually because their orbital periods are related by a ratio of small integers. Most commonly, this relationship is found between a pair of objects (binary resonance). The physical principle behind the orbital resonance is similar in concept to pushing a child on a swing, whereby the orbit and the swing both have a natural frequency, and the body doing the "pushing" will act in periodic repetition to have a cumulative effect on the motion. Orbital resonances greatly enhance the mutual gravitational influence of the bodies (i.e., their ability to alter or constrain each other's orbits). In most cases, this results in an *unstable* interaction, in which the bodies exchange momentum and shift orbits until the resonance no longer exists. Under some circumstances, a resonant system can be self-correcting and thus stable.

Similar complications occur, when a molecular vibration has the same frequency as some overtone or some combination tone of the same molecule. This leads to what is called a "Fermi resonance". The peaks in the figure below pertain to the carbonyl stretching mode and the C≡N stretching mode. Some combination tones happens to have the same frequencies, which leads to a splitting of these peaks. The two branches correspond to synchronous and anti-synchronous motion of the main resonance and the combination tone.



https://pmc.ncbi.nlm.nih.gov/articles/PMC5538721/

Accept some small ellipticity. If a planet is always kicked when it is closest to the sun, these kicks do not occur randomly. The movement of two bodies kicking each other becomes more complicated. They enter an "orbital resonance". Orbital resonances occur, when the orbital periods of two planets are related by two small integers (e.g., T_1:T_2 = 2:3 or T_1:T_2 = 3:4). That may turn into a stable coupled motion. However, it may also result in unstable motion in the sense that one of the planets eventually leaves the planetary system, possibly destabilizing other planets on its way. That problem is of some importance when calculating the likelihood to find other civilizations somewhere in the Milky Way. Our planetary system contains eight planets. One of them is in the "habitable zone". It might be that systems with so many planets are the exception rather than the rule. We might be lucky to have – so far – not seen that kind of instability. Other planetary systems may have been less lucky. Chapter 5 in Binney's book is well written this regard. It feels like we are part of a jolly nice adventure.

Box 4.1 elaborates on an analogy between the orbital resonances and the "Fermi resonances" in vibrational spectroscopy.

The encounter between Venus and Earth discussed above would be called as “collision” in the stellar dynamics.[24] A collision must not be misunderstood as physical collision. These are extremely rare. A collision deflects the orbits of the two bodies.

There is a problem of terminology. What is called “collision” in stellar dynamics is called a “correlation” in quantum chemistry. Instantaneous correlations are what the Hartree-Fock method misses. “Correlations” have to be accounted for in a second step. A rather powerful method to do so is density functional theory (DFT). “Collisions” are called “ion correlations” in the theory of electrolytes. Debye-Hückel theory does not account for ion correlations. (Debye-Hückel theory has numerous other shortcomings.)

In globular clusters, the numerous collisions are uncorrelated between each other, which simplifies the statistics. Stars forget their original orbits similar to how particles in Brownian motion forget their previous positions and momenta. Stars undergo a diffusion in the space spanned by positions and momenta (a 6-dimensional space). The time, after which most stars have forgotten their original orbits, is the “relaxation time.”

Remember:

- Collisions – and more generally, instantaneous correlations – limit the applicability of mean-field theory.

Box 4.2

Short-range correlations

Sometimes, correlations can be accounted for in ways, which leave mean-field theory intact. These mostly are short-range correlations An example is the orientation of the segments of a polymer chain. Neighboring segments often have correlated orientations, which spoil the random walk. The random walk can be recovered by defining larger effective segments such that neighboring effective segments have uncorrelated orientation.

This procedure eliminates the consequences of local chain stiffness, but it does not eliminate the consequences of self-avoidance. Two segments, that are not adjacent on the chain, still are never allowed to occupy the exact same position in space. Self-avoidance causes a "long-range correlation". Long-range correlation are not easily modeled. The consequences of self-avoidance are most efficiently studied with simulations. These simulations lead to the Flory exponent (see Wikipedia). “N-body simulations” are the comprehensive (and expensive) answer to correlations in globular clusters.

There also are short-range correlations in the Brownian movement of particles in liquids. A test particle’s path is not a perfect random walk because it collides with particles, that have themselves collided with the test particle shortly before and have not forgotten this collision. This leads to the "long-time tails" in the velocity autocorrelation functions.

[24] Collisions here are portrayed as discrete events. That can be questioned. See also Box 16.1.

5 Statistics of self-gravitating systems

5.1 Recap: The entropy reaches a maximum at thermodynamic equilibrium

The function $\rho_{star}(\mathbf{r},\mathbf{p},t)$ introduced in section 4.2 is not stationary. Globular clusters are not in equilibrium. We digress to a discussion of the thermodynamic equilibrium anyway. Globular clusters are contrasted against equilibrium.

Thermodynamic equilibrium is an "ergodic" state (to the right in Fig. 4.2). In the ergodic state, the system comes arbitrarily close to all microstates. The time-average is equal to the ensemble average. Most importantly, all microstates are equally probable. That is a very strong statement. It is the axiom underlying statistical mechanics and equilibrium thermodynamics.

At the start of most courses on chemical thermodynamics, temperature is defined to be what thermometers measure. That cannot be the ultimate definition and is not the ultimate definition. Temperature governs the volume-pressure relation of the ideal gas. From there, the discussion meanders, until it comes to the Carnot process. Neither heat nor work are conserved in the Carnot process. Heat and work are "process variables", as opposed to state variables.[25] However, two state variables can be constructed from heat and work, which are the internal energy, U, and the entropy, S. S is defined differentially by the relation

Eq. 5.1

$$dS = \frac{\delta Q}{T}$$

δQ is a small in change in heat. (It is δQ instead of dQ because Q is not a state variable.) The entropy is conserved in the Carnot process. S is a state variable, as long as all processes are reversible.

As long as the definition of entropy is differential, entropy does not have absolute values. That is not a problem because only differences in entropy matter in chemical thermodynamics. Processes occur spontaneously if the overall entropy of the system and the environment increases (if $\Delta S_{tot} > 0$). An equilibrium constant of a chemical reaction obeys $K = \exp(-\Delta G/RT)$, where $\Delta G = \Delta U + p\Delta V - T\Delta S$ is the difference in free enthalpy.

We no come back to the distinction between "microstates" and "macrostates", introduced in section 4.2. A microstate is a set of parameters, which fully describe a system's state. It answers all questions, which might possibly be asked. (It does not answer questions, which are forbidden by quantum mechanics.) In classical mechanics, these parameters typically are the positions and the velocities of all particles. A macrostate is a much smaller set of parameters, which are "robust" in the sense that they can be predicted by reasoning. There is some freedom in the

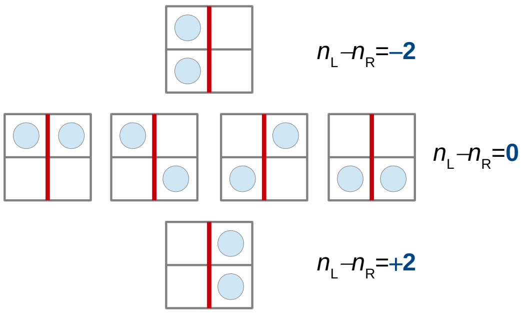


Fig. 5.1
A macrostate in thermodynamic equilibrium is characterized by maximum number of microstates being compatible with this macrostate.
$n_L - n_R = 0$ is the most probable macrostate in the above diagram.

[25] State variables pertain to equilibrium states and take values which do not depend on the system's history.

choice of macroparameters (in the choice of a "relevance concept"), but some macroparameters like volume and pressure are familiar to us.

At this point, quantum mechanics makes things much easier.[26] The number of microstates, which conform to a chosen macrostate, can be counted (Fig. 8.2). (In classical mechanics, the microstates form a continuum. These states cannot be counted.) Call this number Ω_s. The subscript s denotes the macrostate. The central postulate underlying statistical mechanics is that all microstates are equally probable. That is not always so. If it is so, equal probability leads to "equilibrium thermodynamics" with many profound consequences. In thermodynamic equilibrium, the system has reached the one macrostate, which has the largest number of microstates associated to it.[27] It is a matter of counting. There is no force pulling the system towards equilibration. Equilibrium simply is the most probable state. In Fig. 5.1, two particles on a checker board were chosen as the system. There is as total of 6 ways to arrange those on the board. The difference in occupation of the two left squares and the two right squares was chosen as the macroparameter. In more physical terms, this sketch represents two chambers connected by a small hole. The chosen macroparameter is the difference in gas pressure, Δp. Δp will equilibrate to $\Delta p = 0$. The number of microstates is counted in Fig. 5.1. There are 4 ways to arrange particles such that $n_L - n_R = 0$, while there is only 1 way to let $n_L - n_R$ be equal to +2 or to –2. (L and R stand for "left" and "right".) The even distribution is more likely.

Box 5.1

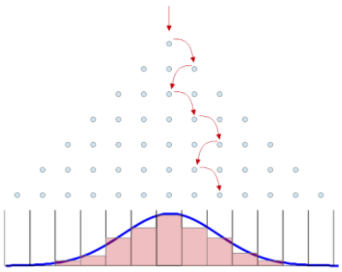
The law of large numbers

A good illustration of how macroparameters might be defined such that they are robust is Galton's board. Galton's board is a physical realization of a random walk in one dimension. Galton's board consists of an arrangement of nails as in the figure on the right. Sphere are dropped onto the board form the top. Whenever a sphere hits a nail, it may jump to the left or to the right. The set of directions of all jumps constitutes a microstate. Each sphere constitutes a system. At the bottom of the board, each sphere will have some distance from the center, Δx. Δx is the sum of the displacements from all jumps. Expressed in more abstract terms, Δx is the sum of N uncorrelated random numbers. The "central limit theorem" predicts that the probabiltiy for finding a sphere at position Δx is a Gaussian:

$$\rho(\Delta x) = \rho_0 \exp\left(-\frac{\Delta x^2}{2\sigma^2}\right)$$

The width of the Gaussian, σ, is given as $\sigma = N^{1/2}/2$ with N the number of rows of the board. The normalized width, σ/N, is proportional the $N^{-1/2}$. The larger the number of rows, the closer is $\Delta x/N$ to zero. This is the "law of large numbers". If N is very larger, most spheres will land close to the center. Δx then is a "robust" parameter. Its value can be predicted and the uncertainty on that value is small.

In a similar way the pressure exerted by a gas onto a wall results from the sum of the momenta transferred to the wall when particles collide with the wall. Evaluated on a small area and over a short time, the pressure does fluctuate. When evaluated on large areas and over long times, the fluctuations in pressure can no longer be noticed, experimentally.



Galton's board is mathematically equivalent to the situation depicted in Fig. 5.1 with Δx being replaced by $n_L - n_R$.

---

[26] Even before the advent of quantum mechanics, Nernst proposed set the entropy of ideal crystals at $T = 0$ to zero, thereby defining entropy in absolute terms.
The differential definition of entropy in Eq. 5.1 is one of the cases, where one would wish things to be a bit simpler than they are. The definition of entropy becomes simply *after* one has accepted quantum mechanics.

[27] Strictly speaking, there is a family of macrostates close to thermodynamic equilibrium. The macroparameters slightly fluctuate.

We have worked our way forward to the second law of thermodynamics. Systems evolve (meaning: systems change their macroparameters) such that Ω_s increases. You know the second law as a statement about increasing entropy, *S*, rather than increasing Ω_s. *S* is defined as[28]

Eq. 5.2

$$S = k_B \ln \Omega_s$$

Why the logarithm? The logarithm is needed because the temperature, *T*, is supposed to be an intrinsic parameter (independent of amount, staying constant when two systems with the same temperature are joined). *U* is known to us as an extensive parameter. In order to let the temperature be intensive, the numerator on the right-hand side in Eq. 5.3 must be extensive, as well. Ω, however, is not additive when systems are joined. Ω is multiplicative. 8 bits (equivalent to 1 Byte) encode $2^8 = 256$ different states (as opposed to 8 states). That Ω should be multiplicative is an experience, but it still warrants some reflection. Quantum entanglements spoil that assumption. Practical considerations may spoil it. A memory card with 1 GByte storage space can encode for 2^(10^9) states, in principle… but not in practice. Anyway: Accepting that Ω is multiplicative, *S* is extensive because of $S = k_B \ln\Omega$. Many occurrences of the

Box 5.2
The Boltzmann factor
For a system in thermal equilibrium with an environment, the ratio of the probabilities of two microstates is the ratio of the number of corresponding microstates of the environment:

$$\frac{p_2}{p_1} = \frac{\Omega_{env,2}}{\Omega_{env,1}}$$

Take the logarithm:

$$\ln\frac{p_2}{p_1} = \ln\frac{\Omega_{env,2}}{\Omega_{env,1}} = \ln\Omega_{env,2} - \ln\Omega_{env,1} = \frac{1}{k_B}\Delta S_{env}$$

Use the definition of temperature and use that the system transitions to the higher state by extracting a small amount of energy from the environment

$$\Delta S_{env} \approx \frac{\Delta S_{env}}{\Delta U_{env}}\Delta U_{env} = \frac{1}{T}(-\Delta\varepsilon)$$

Revert the logarithm:

$$\frac{p_2}{p_1} = \exp\left(-\frac{\Delta\varepsilon}{k_B T}\right)$$

The term on the right-hand side is the Boltzmann factor.

logarithm in science and nature are rooted in Ω being multiplicative. That includes – for instance – the Weber-Fechner law.[29]

Starting on the absolute definition of entropy, Boltzmann has put thermodynamics on solid ground. He defines temperature by the relation

Eq. 5.3

$$\frac{1}{T} = \frac{\mathrm{d}S}{\mathrm{d}U}$$

[28] Eq. 5.2 links heat to probability. It is a hallmark of how things „hang together“, section 2.3.

[29] Weber and Fechner discovered that many sensory organs in the living world respond on a logarithmic sensitivity scale. The log-dependence goes back to a logarithmic dependence of the membrane potential in nerve cells on the concentration of messenger molecules, as predicted by the Nernst equation. The Nernst equation itself follows from the Boltzmann distribution, which contains an exponential relationship between probability and energy, rooted in entropy being related to the log of probability (Box 5.2).

However, that does not work for globular clusters (section 5.3).

Remember:

- In quantum mechanics, the number of microstates conforming to a set of macroparameters labeled as "s" can be counted.
- The entropy is proportional to the logarithm of that number: $S = k_B \ln \Omega_s$.
- Because of the logarithm, the entropy is additive when systems are joined. It is an extensive parameter.
- Temperature is defined by $1/T = dS/dU$.

5.2 Recap: Thermal equilibrium

Consider a system in a thermal equilibrium with an environment.[30] An environment is an infinite reservoir in the sense that its temperature does not change appreciably when heat is added or removed. Maximum total entropy implies

Eq. 5.4

$$0 = \frac{dS_{tot}}{dU_{sys}} = \frac{dS_{sys}}{dU_{sys}} + \frac{dS_{env}}{dU_{sys}} = \frac{dS_{sys}}{dU_{sys}} - \frac{dS_{env}}{dU_{env}} = \frac{1}{T_{sys}} - \frac{1}{T_{env}}$$
$$\Rightarrow T_{sys} = T_{env}$$

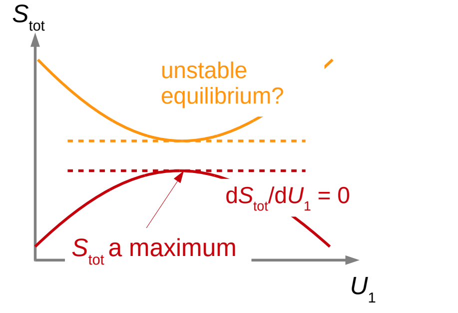


Fig. 5.2
In a stable thermal equilibrium, the total entropy is at its maximum. The tangent to the maximum is horizontal ($dS_{tot}/dU_1 = 0$).

dU is a small amount of heat, that is exchanged back and forth between the system and the environment. It was used that $1/T = dS/dU$.[31] In thermal equilibrium, both temperatures are the same. At some point, the temperature will be the same everywhere. This is Clausius' heat death.

However, the thermal equilibrium might be unstable (Fig. 5.2). On earth, it is stable, meaning that the entropy indeed is at a maximum ($d^2S/dU^2 < 0$). So see that write

Eq. 5.5

$$\frac{d^2S_{tot}}{dU_1^2} = \frac{d^2S_1}{dU_1^2} + \frac{d^2S_2}{dU_1^2}$$

The first term can be transformed as

Eq. 5.6

$$\frac{d^2S_1}{dU_1^2} = \frac{d(1/T_1)}{dU_1} = \left(\frac{d(1/T_1)}{dT_1}\right)\frac{dT_1}{dU_1} = \left(-\frac{1}{T_1^2}\right)\frac{dT_1}{dU_1} = -\frac{1}{T^2}\frac{1}{C_{V,1}}$$

The same transformation applies to the second term. C_V is the heat capacity at constant volume. We are used to C_V begin positive. The thermal equilibrium is stable as long as C_V is positive.

Eq. 5.5 can be explained in prose. If a small fluctuation lets heat flow from left to right despite thermal equilibrium, this will raise the temperature on the right. In consequence, the heat will flow back.

---

[30] The argument applies more generally to any two systems in thermal contact.
[31] More precise would be: $1/T = (dS/dU)_V$ because $dU = -p dV + T dS$.

Should, however, the temperature on the right decrease upon an influx of heat, this will amplify the fluctuation. Increased heat flow from hot to cold (from left to right) results, potentially leading to a runaway heat transfer (an instability).

<u>Remember:</u>

– If two systems are in thermal equilibrium, they have the same temperature.
– The thermal equilibrium is stable as long as the heat capacity is positive.

5.3 The internal energy is not an extensive parameter in self-gravitating systems

In globular clusters, gravitation is the only interaction between the numerous stars. There is no thermodynamic equilibrium because core contraction and shell expansion continue forever (section 5.5). On the other hand, core contraction and shell expansion are slow. They can be ignored for some purposes, but even then a temperature cannot be defined in a meaningful way.

The long-range nature of the gravitational interaction makes familiar ideas from thermodynamics inapplicable to globular clusters.[32] For the $1/r$-potential, the internal energy, U, is no longer an extensive quantity. Extensive quantities are additive when two systems are joined. Extensive quantities in this sense are the amount of substance, the volume, and – often, but not always – the internal energy. (Examples of intensive quantities are temperature and pressure.) The internal energy U is extensive if interactions only occur between neighbors. Then $U \approx N\varepsilon z/2$ with N the number of molecules, ε the pair interaction energy and z the number of neighbors ($4 < z < 14$). The factor 1/2 avoids interactions between two bodies being counted twice. If $U \approx N\varepsilon z/2$, the internal energy is proportional to the amount of substance and therefore is extensive.

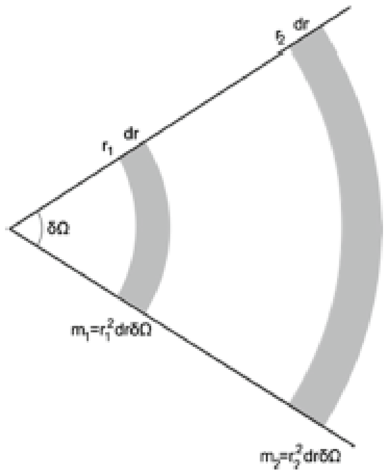


Fig. 5.3
Consider the star on the left-hand side of the diagram. Let the mean density of the other stars, ρ, be constant. The stars on the shell on the far right then exert the same force onto the star to the left as the stars on the smaller shell further in. The force exerted by each star scales as $1/r^2$ but the number of stars on the shells scales as r^2. The two factors cancel. The sum of all forces depends on how many shells there are. Put differently, the gravitational energy of each single star depends on the system size.
https://doi.org/10.1007/s11229-018-02032-5.

We repeat this argument, assuming that the pair interaction, w, scales as some power of the distance, r. First, assume $w = -C/r^6$ as in the van-der-Waals potential. C is related to the Hamaker constant, A, as $C = A/(\pi\rho)^2$ with ρ the density. Consider a particle in the center of a cloud. Its interaction energy with all other members of the system is

Eq. 5.7

$$\varepsilon_{\text{body}} = \int_{r\_\min}^{r\_\max} \frac{-C}{r^6}\rho 4\pi r^2 \mathrm{d}r = -4\pi\rho C \int_{r\_\min}^{r\_\max} \frac{1}{r^4}\mathrm{d}r = -4\pi\rho C\left[\frac{1}{3}\frac{1}{r_{\max}^3} - \frac{1}{3}\frac{1}{r_{\min}^3}\right] \approx 4\pi\rho C\left(\frac{1}{3}\frac{1}{r_{\min}^3}\right)$$

[32] Mukamel, D., Notes on the Statistical Mechanics of Systems with Long-Range Interactions **2009** *arxiv.org/pdf/0905.1457.pdf*.

r_{min} is the minimum distance between two bodies, r_{max} is the system size. If r_{max} is large, $1/r_{max}^3$ disappears from the final result. The interaction energy then only depends on the distance to the nearest neighbor. This result is equivalent to writing $U \approx N\varepsilon_{body} = N\varepsilon z/2$, where ε and z must be redefined to match Eq. 5.7.

Now switch to gravitation with $w = -Gm^2/r$. G is the gravitational constant, m is the mass (same for all bodies). The interaction energy of the particle in the center with all other particles now is

Eq. 5.8

$$\varepsilon_{body} = \int_{r\_min}^{r\_max} \frac{-Gm^2}{r} \rho 4\pi r^2 \mathrm{d}r$$

The term r_{max} does no longer disappear:

Eq. 5.9

$$\varepsilon_{body} \approx \int_{r\_min}^{r\_max} -Gm^2 \rho 4\pi r \mathrm{d}r = -Gm^2 \rho 4\pi \left[r_{max}^2 - r_{min}^2 \right]$$

ε_{body} depends on the system size. Fig. 5.3 illustrates the problem based on force rather than energy. Any member of the ensemble interacts with all other members of the ensemble. ε_{body} itself depends on N and the sum over all interactions energies no longer is extensive. The temperature, defined by $1/T = \mathrm{d}S/\mathrm{d}U$, no longer is intensive. Usually, S and U are both extensive. This is the case if all interactions are short-ranged. The ratio of dS and dU then is intensive. However, it is not the case for self-gravitating systems. A side remark: In self gravitating systems, S is not extensive either because the particles are not distributed randomly and because the distribution changes when two systems are joined (Fig. 5.4).

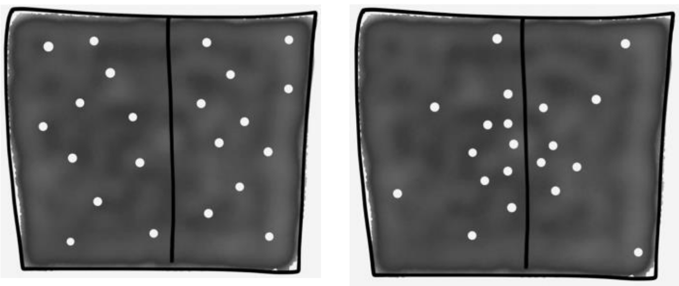


Fig. 5.4
If two systems with long-range interactions are combined, the particles are distributed as shown in the figure on the right (short-range interaction: left). The internal energy and the entropy then are no longer additive.
https://link.springer.com/article/10.1007/s11229-018-02032-5

Short-range interactions also are a prerequisite for the "model of missing neighbors", which is the simplest explanation of surface energy. If only the interactions between neighbors are to be significant, the interaction must be short-ranged. Hydrogen bonds, covalent bonds, the hydrophobic interaction, and the van-der-Waals interaction all are short-ranged in that sense. We avoid a discussion of what – exactly – is meant by "short-ranged" in the sense of the model of missing neighbors. This is a complicated question.

Are there long-range interactions ($1/r$-interactions) other than gravity? Not many. Of course, the Coulomb potential is of long range, but the Coulomb potential is screened by countercharges at distances larger than the "Debye length". Diverging integrals (as in Eq. 5.9 for the case $r_{max} \to \infty$) always entail complications. If those would occur often in physical chemistry, physical chemistry would be more complicated than it actually is.

Remember:

– In systems held together by long-range attractive interactions, the internal energy is not an extensive quantity. A temperature cannot be defined.

5.4 The virial theorem

While a global temperature is not defined for globular clusters, a *local* temperature can still defined for the self-gravitating gas. It will turn out that the self-gravitating gas has a negative heat capacity in a certain sense. To discuss this in more detail requires the virial theorem. Because the virial theorem has applications in all of physical chemistry, we discuss it in some depth.

The virial theorem starts from an energy function, H, which depends on some variables $\{q\}$. In the following, $\{q\}$ shall include the positions and the momenta of the particles. To keep the mathematics manageable, we exclude angles and angular momenta and we also exclude the amplitudes of electromagnetic waves. For any given energy and given set $\{q\}$, the parameter $<q_i\,\mathrm{d}H/\mathrm{d}q_i>$ has interesting properties. Angular brackets denote a time average. Consider momenta first. The system's kinetic energy is

Eq. 5.10

$$E_{kin} = \sum_{particles,i} \frac{1}{2} m_i \left(v_{x,i}^2 + v_{y,i}^2 + v_{z,i}^2 \right) = \sum_{particles,i} \frac{1}{2m_i} \left(p_{x,i}^2 + p_{y,i}^2 + p_{z,i}^2 \right)$$

The parameter $<q_i\,\mathrm{d}H/\mathrm{d}q_i>$ with $q_i = p_{x,i}$ is

Eq. 5.11

$$\left\langle p_{x,i} \frac{\mathrm{d}E_{kin,x,i}}{p_{x,i}} \right\rangle = \frac{1}{m_i} \left\langle p_{x,i}^2 \right\rangle = 2 \left\langle E_{kin,x,i} \right\rangle$$

Now consider positions, r_i, and a potential $V(r_i)$. The analog of Eq. 5.11 is

Eq. 5.12

$$\left\langle r_{x,i} \frac{\mathrm{d}V}{r_{x,i}} \right\rangle = -\left\langle r_{x,i} F_{x,i} \right\rangle$$

Consider all particles and consider all three dimensions. Add a factor of 1/2 for later convenience:

Eq. 5.13

$$\frac{1}{2} \left\langle \sum_{particles,i} \sum_{\alpha=x,y,z} r_{\alpha,i} \frac{\mathrm{d}V}{r_{\alpha,i}} \right\rangle = -\frac{1}{2} \left\langle \sum_{particles,i} \sum_{\alpha=x,y,z} r_{\alpha,i} F_{\alpha,i} \right\rangle = -\frac{1}{2} \left\langle \sum_{particles,i} \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle = -\frac{1}{2} \left\langle \sum_{particle\ pairs,ij} \mathbf{r}_{ij} \cdot \mathbf{F}_{ij} \right\rangle$$

The last identity is proven in section 16.2. The last term on the right-hand side is the "inner virial". The non-thermodynamic version of the virial theorem states that

Eq. 5.14

$$\left\langle E_{kin} \right\rangle = -\frac{1}{2} \left\langle \sum_{ij} \mathbf{r}_{ij} \cdot \mathbf{F}_{ij} \right\rangle$$

Eq. 5.14 is also proven in section 16.2. Again: Thermodynamic equilibrium is not required. <…> is the time average, as opposed to the Boltzmann-weighted ensemble average in Eq. 5.18.

For an application of Eq. 5.14, consider two spheres linked by a Hookean spring. This two-body problem turns into a one-body problem, if the mass is replaced by the reduced mass and if the dynamic variable is the distance between the two particles. The dynamic variable may also be the excursion from equilibrium, Δr. We find

Eq. 5.15

$$\left\langle E_{kin,HO}\right\rangle = -\frac{1}{2}\left\langle \Delta r\cdot\frac{\mathrm{d}}{\mathrm{d}\Delta r}\left(-\frac{1}{2}\kappa\Delta r^2\right)\right\rangle = \left\langle \frac{1}{2}\kappa\Delta r^2\right\rangle = \left\langle V_{HO}\right\rangle$$

The subscript HO stands for "harmonic oscillator". For the harmonic oscillator (and also for the pendulum), the time-averaged kinetic energy is the same as the time-averaged potential energy. This result is well known.

Consider two spheres interacting by gravitation:

Eq. 5.16

$$\left\langle E_{kin}\right\rangle = -\frac{1}{2}\left\langle r\cdot\left(-\frac{\mathrm{d}V}{\mathrm{d}r}\right)\right\rangle = \frac{1}{2}\left\langle r\cdot\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{Gm_1m_2}{r}\right)\right\rangle = -\frac{1}{2}\left\langle r\cdot\frac{Gm_1m_2}{r^2}\right\rangle = -\frac{1}{2}\left\langle\frac{Gm_1m_2}{r}\right\rangle = -\frac{1}{2}\left\langle V\right\rangle$$

Again, the result in brief

Eq. 5.17

$$\left\langle E_{kin}\right\rangle = -\frac{1}{2}\left\langle V\right\rangle$$

Wikipedia calls Eq. 5.17 "the virial theorem". It was derived for a two particles system here, but it also holds for many-particle systems. The virial theorem has far-reaching consequences. Both the planets and the electrons in atoms obey Eq. 5.17. Mercury moves faster than Jupiter because the Kepler velocity in the planetary system scales as $r^{-1/2}$. (The Kepler velocity is the velocity on a circular orbits which lets the centripetal force be equal to the centrifugal force.) The kinetic energy of the 1s-electron in hydrogen is larger than the kinetic energy of the 2p-electron. That is easily proven in the frame of the Bohr model, but also holds when the electrons are modeled as waves. Stars in globular clusters obey the virial theorem. Self-gravitating clouds of gas obey the virial theorem.

A second version of the virial theorem connects the virial with temperature, as discussed in section 16.3. The central equation is

Eq. 5.18

$$\left\langle q_i\frac{\mathrm{d}H}{\mathrm{d}q_i}\right\rangle = k_B T$$

The angular brackets not only are a time average, they also are an ensemble average because the system is in thermodynamic equilibrium. If H depends quadratically on some variable q_i (as in $E_{\mathrm{kin,i}} = p_\mathrm{i}^2/(2m)$), this part of the energy – call it H_i – obeys $\langle H_\mathrm{i}\rangle = 1/2\ k_\mathrm{B}T$. This is the famous *equipartition theorem*. Eq. 5.18 also leads to the pressure-volume relation of the real gas (Eq. 16.31). The pressure-volume relation can be Taylor-expanded in density, leading to what is called the "*virial expansion*".

Remember:

– In systems, where the interaction is mediated by a $1/r$-potential, the time-averaged kinetic energy equals the negative half of the time-averaged potential energy: $\langle E_\mathrm{kin}\rangle = -1/2\ \langle V\rangle$.

– The thermodynamic version of the virial theorem leads to the equipartition theorem and to the virial expansion.

5.5 No thermodynamic equilibrium for self-gravitating systems, mass segregation

I was stated in section 4.2 that self-gravitating systems do not reach a thermodynamic equilibrium. That can be understood, making use of the virial theorem. When the core contracts, the bodies gain kinetic energy. The increased energy allows for an increased entropy contained in the momenta. When the shell expands, there is increased entropy contained in the positions. Core contraction together with shell expansion thereby increases the overall entropy. There is no limit to this process (Fig. 5.5).

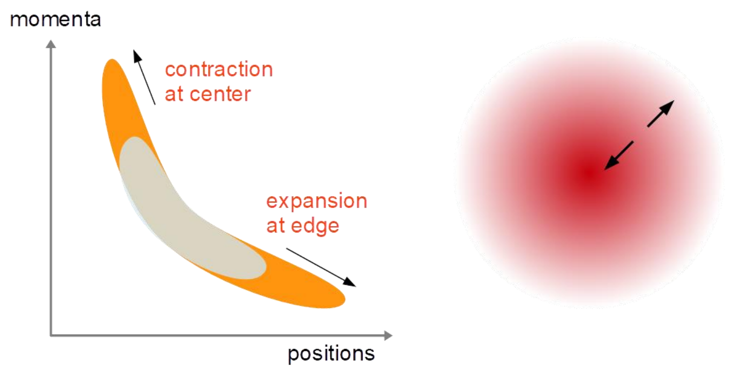


Fig. 5.5
Self-gravitating systems can always increase their entropy by contracting the core and expanding the shell. Core contraction elevates the temperature, thereby increasing the entropy contained in the momenta. Meanwhile, the shell's expansion increases the entropy contained in the positions. The expansion may also result in bodies leaving the system, a process referred to as "evaporation".

In self-gravitating clouds of gas, there may still be local thermodynamic equilibria. Those many particles frequently collide, physically (a physical collision to be distinguished from the "collisions" of stars, which only deflect the orbits). An internal energy may defined, which does not include gravity. Call it U_{ng}, where the subscript "ng" stands for non-gravitational. This local internal energy contains the kinetic energy, possibly the Coulomb interaction, and possibly the van-der-Waals attraction. Temperature is defined by $1/T = dS/dU_{ng}$. Such local equilibria are known from "non-equilibrium thermodynamics" (section 5.8).

There is a twist to the slow evolution of $\rho_n(\mathbf{r},\mathbf{p},t)$, which is that core compaction is preferentially carried by heavy bodies, while shell expansion and evaporation are preferentially carried by light bodies. The process is called "mass segregation". When two bodies collide, there is a tendency towards an even distribution of kinetic energy between the collision partners, related to the equipartition theorem. After the collision, the heavier body no longer has the velocity matching the virial theorem. Should it be on a circular orbit, it's velocity is less than the Kepler velocity. The heavier body then moves inwards. The reverse happens for the lighter body.

Core contraction and shell expansion is known from the red giants. At old age, stars expand by large amounts. The core, contained burnt fuel, contracts. When the sun turns into a red giant, it will swallow the inner planets, including the earth.

In the protoplanetary disks, dust drifts inwards while gas moves outwards (section 6.1). This is a form of mass segregation, although it is not called that in protoplanetary disks.

In clouds of dust and gas, core contraction ends in the formation of stars. The stars form "open clusters". Open clusters contain fewer stars than globular clusters. Open clusters loose stars by evaporation until a binary star is left, eventually. Evaporation here is equivalent to shell expansion. Binary stars are stable. (Triple stars are not.) In the case of globular clusters (> 10^5 stars), core contraction and shell expansion are very slow.

In the large clouds of gas in the early universe, core contraction ended in supermassive black holes. The formation of the supermassive black holes is poorly understood.

Remember:

- In self-gravitating systems, the core contracts while the shell expands. Shell expansion can amount to evaporation.
- Self-gravitating systems with no interactions other than gravity are never in a global thermodynamic equilibrium.
- For self-gravitating clouds of gas, there may be local thermodynamic equilibria.

5.6 *The gravothermal catastrophe*

Self-gravitating clouds of gas, for which there is a local thermodynamic equilibrium and a local temperature, have a negative heat capacity. (Recall that gravitation does not enter the definition of the temperature.) From the virial theorem (Eq. 5.17) and the equipartition theorem (section 16.3 leading to $\langle E_{kin}\rangle = 3/2\, k_B T$), it follows that

Eq. 5.19

$$\frac{3}{2}k_B T = -\frac{\langle V\rangle}{2}$$

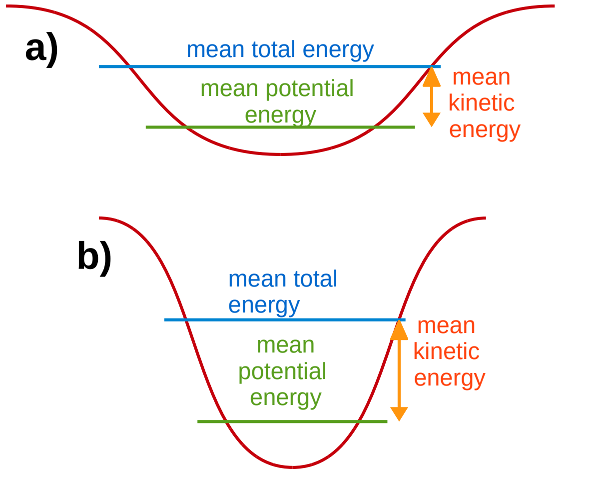


Fig. 5.6
For self-gravitating systems, the mean kinetic energy increases when particles evaporate or when energy is radiated into the dark outer space.

The total energy, $\langle E_{tot}\rangle = \langle V\rangle + \langle E_{kin}\rangle$, equals $-\langle V\rangle/2$. The heat capacity is negative:

Eq. 5.20

$$C := \frac{\mathrm{d}\langle E_{tot}\rangle}{\mathrm{d}T} = \frac{\mathrm{d}\langle -V/2\rangle}{\mathrm{d}T} = \frac{\mathrm{d}\langle -E_{kin}\rangle}{\mathrm{d}T} = -\frac{3}{2}Nk_B$$

Contracting regions heat up, they radiate away the energy contained in the heat, they thereby loose energy, they continue to contract, and they continue to heat up (Fig. 5.6). This “gravothermal catastrophe” ends with the stars.

A few more remarks:

- When it was said that the contracting regions radiate away the energy contained in their heat, it was assumed that they radiate away more energy than they absorb from the outside. This process requires a cold sky, meaning: requires an expanding universe.
- The shell will also eject material into the environment by evaporation. The process is opposite of evaporative cooling, which is familiar to us.
- This treatment is a bit relaxed about formal correctness. For a more accurate treatment see: Wallace, David. "Gravity, Entropy, and Cosmology: In Search of Clarity". *The British Journal for the Philosophy of Science* 61 (July 3, 2009).

Remember:

- Self-gravitating systems heat up when they release energy.

5.7 The Jeans collapse

There also is a hydrostatic instability, which, however, requires a certain minimum size. The Jeans collapse does not require heat to flow from the core to the shell of the collapsing region, differing from the gravothermal catastrophe. The Jeans collapse is modeled as isothermal. Whether the real-world cases are purely isothermal always is in question.

If the pressure produced by gravity is larger than the pressure of the gas to be compressed, the equilibrium is unstable. This happens when the mass of the cloud is larger than the mass in Eq. 5.26 (larger than the "Jeans mass").

The following text is adapted from https://en.wikipedia.org/wiki/Jeans_instability.

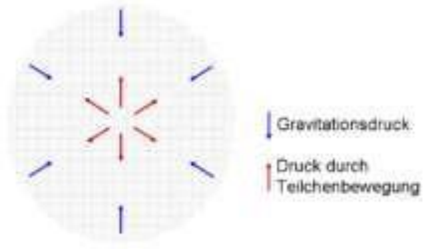


Fig. 5.7
In self-gravitating clouds, the gas pressure acts against the gravitational pressure.
en.wikipedia.org/wiki/jeans-criterion

Consider a homogenous spherical gas cloud with radius R. In order to compress this sphere to a radius $R - \mathrm{d}R$, work must be done against the gas pressure. During compression, gravitational energy is released. When this energy equals the amount of work to be done on the gas, the critical mass is reached. Let M be the mass of the cloud, T the (absolute) temperature, N/V the particle density, and p the gas pressure. The work to be done equals $p\mathrm{d}V$. Using the ideal gas law, according to which $p = N/V\, T = \rho_N T$,[33] and further using $\mathrm{d}V = 4\pi R^2 \mathrm{d}R \propto R^2 \mathrm{d}R$, one arrives at:

Eq. 5.21

$$\mathrm{d}W \propto \rho_n T R^2\, \mathrm{d}R$$

The gravitational potential energy of a sphere with mass M and radius R is proportional to M^2/R:

Eq. 5.22

$$U \propto -\frac{M^2}{R}$$

The amount of energy released when the sphere contracts from radius R to radius $R - \mathrm{d}R$ is obtained by differentiation with respect to R:

Eq. 5.23

$$\mathrm{d}U \propto \frac{M^2}{R^2}\mathrm{d}R$$

The critical mass is reached when the released gravitational energy is equal to the work done on the gas:

Eq. 5.24

$$\frac{M^2}{R^2} \propto \rho_n T R^2 \quad \Rightarrow M^2 \propto \rho_n T R^4$$

---

[33] The gas constant (usually called R, as well) was absorbed into the definition of temperature.

Next, the radius R must be expressed in terms of the particle density $\rho_n = N/V$ and the mass M. This can be done using the relation

Eq. 5.25

$$M \propto \rho_n R^3$$

Eliminating R from Eq. 5.24 leads to :

Eq. 5.26

$$M^2 \propto \rho_n T R^4 \propto \rho_n T \left(\frac{M}{\rho_n}\right)^{4/3} \Rightarrow M^{2/3} \propto \frac{T}{\rho_n^{1/3}} \Rightarrow \quad M \propto \left(\frac{T^3}{\rho_n}\right)^{1/2}$$

The last equation links the minimum mass for the Jeans collapse to temperature and density. The higher the temperature, the larger is the gas pressure that must be overcome by the Jeans collapse.

1/*r*-potentials are rare on earth. There is a playful experiment which emulates a "Jeans collapse" in the laboratory. Kavokine et al. found a 1/*r*-potential in thermophoresis.[34] Colloidal particles sometimes move along a temperature gradient. The particles from ref. 34 were absorbent and were heated with light. They then heated the surrounding fluid. Because these particular particles move towards the hot regions, there was a net attraction between the hot particles. A lengthy calculation with some assumptions leads to an effective 1/*r*-potential. The implosion seen in these experiment indeed shared some peculiarities with the Jeans collapse.

The original argument by Jeans is said to be flawed because it does not treat the surrounding medium correctly. The improved analysis leads the same result because of a fortuitous cancellation of terms.

Remember:

– Self-gravitating systems collapse if the mass is above the Jeans mass.

5.8 Local thermodynamic equilibria

When there are local equilibria in self-gravitating clouds, non-equilibrium thermodynamics applies. Non-equilibrium thermodynamics was popular in the 1970s and 1980s. The claim was that non-equilibrium thermodynamics would explain the formation of dissipative structures. Dissipative structures only exist far away from equilibrium. Examples are the roles of thermal convection, embryonic development, and turbulence. Since then, there has been a shift in emphasis. Today, non-equilibrium thermodynamics primarily describes flows along gradients (diffusion, heat flow, etc.). As long as local equilibria exist, there is an extremum principle, similar to how the entropy is at its maximum for closed systems in equilibrium thermodynamics. In non-equilibrium thermodynamics, all flows result from the requirement that the entropy production rate, $\dot{\sigma}$, tends to a maximum ($\dot{\sigma} \rightarrow \dot{\sigma}_{\max}$).[35] A few decades ago, it was a widespread opinion that dissipative structure formation should also follow from this principle. Rayleigh-Benard convection was cited as an example. When heating a hotplate just a little, the heat is transported upwards via heat conduction. Upon heating a little more, convection rolls form, also known from the

[34] Kavokine, N.; Zou, S. Y.; Liu, R. B.; Nigues, A.; Zou, B. S.; Bocquet, L., Ultrafast photomechanical transduction through thermophoretic implosion. *Nature Communications* **2020,** 11, (1).

[35] The argument can be turned around. The principle $\dot{\sigma} \rightarrow \dot{\sigma}_{\max}$ follows from the linear dependence of the fluxes on the generalized forces.

weather. These "dissipative structures" disappear once the heat flow stops. In contrast, the self-organized structures are stable in equilibrium.

Today, it is believed that there is no general extremum principle for these structure formation processes. The principle "$\dot{\sigma} \rightarrow \dot{\sigma}_{max}$" only applies to weak non-equilibria in which the fluxes are proportional to the corresponding generalized forces.

There are many structure formation mechanism in astrophysics, but these are most easily discussed in terms of the relevant dynamical laws (often non-linear partial differential equations). Entropy always increases, but the entropy production rate does not necessarily take on its maximum value. This being said: The instabilities in the accretion disks and the protoplanetary disks are processes of dissipative structure formation. We only discuss the streaming instability in section 6.1.

Remember:

- Non-equilibrium thermodynamics applies to the local equilibria in self-gravitation clouds.
- There are numerous instabilities in accretions disks and protoplanetary disks. These lead to dissipative structures, similar to the convection roles in Rayleigh-Benard convection.

5.9 Background: The photon gas

This section is based on the Ref. 36. There are characteristic differences between the photon gas and to the particle gas:

- Photons can be generated and annihilated. The number of photons is a function of temperature and volume. Because amount is not an independent parameter, the definition of the chemical potential as $\mu := (\mathrm{d}G/\mathrm{d}n)_{p,T}$ does not make sense for the photon gas.
- Photons do not scatter from each other in vacuum. A thermal equilibrium can only be reached if it is mediated by sufficiently opaque matter, which absorbs and re-emits photons. This may include walls and a plasma. Neutral gases usually are too transparent.
- Photons do not have a rest mass, but they still carry an energy and a momentum. Photons can exert pressure on a wall, as is known from light mills (Fig. 5.8).[37] Energy and momentum are related as $E = cp = \hbar k$ instead of $E = p^2/(2m)$, the latter being the energy of the non-relativistic monoatomic gas.

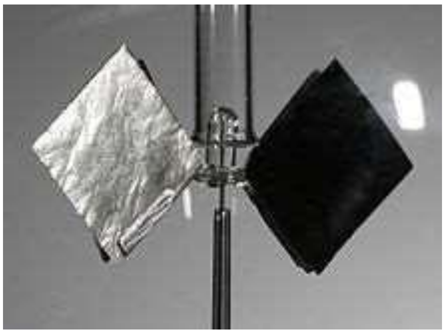


Fig. 5.8
The impeller of a light mill. The left shows the bare side of a mica plate, the right shows the blackened side.
en.wikipedia.org/wiki/Lightmill

[36] Leff, H. S., Teaching the photon gas in introductory physics. *American Journal of Physics* **2002,** 70, 792

[37] Note that a vacuum is needed to let photon pressure drive the light mill. Otherwise, a competing mechanism is stronger: The air in front of the black surfaces heats up, expands, and exerts a force on the surface, pushing it back. In vacuum, photons push back the shiny surface more strongly than the black surface because they transfer twice their momentum to these upon reflection. When absorbed by the black surface, they only transfer their momentum to it.

Quantum mechanics predicts:

Eq. 5.27

$$U(V,T) = bVT^4 \qquad a)$$
$$p(T) = \frac{b}{3}T^4 \qquad b)$$
$$b = \frac{8\pi^5 k_B^4}{15h^3c^3} = 7.56\times10^{-16}\ \frac{\mathrm{J}}{\mathrm{K^4m^3}}$$

In the following we prove the T^4-dependence of energy and pressure. The energy density of the photon gas (the Planck curve, Fig. 3.1) as a function of frequency, ν, is

Eq. 5.28

$$U(\nu,T)\,\mathrm{d}\nu = 2\ 4\pi k^2 h\nu \frac{1}{\exp\left(\frac{h\nu}{k_BT}\right)-1}\mathrm{d}\nu = 8\pi\frac{h\nu^3}{c^3}\frac{1}{\exp\left(\frac{h\nu}{k_BT}\right)-1}\mathrm{d}\nu$$

We do not prove Eq. 5.28, but make the terms plausible. The factor 2 on the left results from the two polarizations. The factor $4\pi k^2$ is the surface area of a sphere in k-space. $h\nu$ is the energy of each photon. The factor $1/(\exp(h\nu/(k_BT))-1)$ results from the Bose-Einstein distribution. Note that this factor is different from the Boltzmann factor. The latter would be $\exp(-h\nu/(k_BT)) = 1/\exp(h\nu/(k_BT))$.

A side remark: The Planck distribution also applies to phonons (sound waves), as long as the phonon wavelength is large compared to the distance between atoms. (There is a complication because longitudinal and transverse sound have different velocities.) Because the energy contained in the phonons scales as T^4, the specific heat capacity scales as T^3. This is the Debye law.[38]

The total energy is proportional to T^4. This is the content of the Stefan-Boltzmann law. The T^4 dependence can be understood from Eq. 5.28 by replacing the variable ν with $\nu' = h\nu/(k_BT)$. With $\nu = k_BT\nu'/h$, one obtains

Eq. 5.29

$$E_{tot} = \int_0^\infty 8\pi\frac{h\nu^3}{c^3}\frac{1}{\exp\left(\frac{h\nu}{k_BT}\right)-1}\mathrm{d}\nu = \int_0^\infty 8\pi\frac{h}{c^3}\left(\frac{k_BT\nu'}{h}\right)^3\frac{1}{\exp(\nu')-1}\mathrm{d}\left(\frac{k_BT\nu'}{h}\right)$$
$$= 8\pi\frac{h}{c^3}\left(\frac{k_BT}{h}\right)^4\int_0^\infty \nu'^3\frac{1}{\exp(\nu')-1}\mathrm{d}\nu'$$

Since the integral does not depend on temperature, the total energy scales as T^4 (see Eq. 5.27a). The value of the integral is $\pi^4/15$. Leaving constants aside, that proves Eq. 5.27a).

Differing from the particle gas, the pressure cannot be calculated as $p = -(\mathrm{d}U/\mathrm{d}V)_{S,N}$ because the number of photons is not conserved. The pressure has to be calculated from quantum mechanics, similar to how the energy is calculated in Eq. 5.29. We allow ourselves a bit of hand-waving. First, pressure is an intensive parameter. The volume in Eq. 5.27a disappears for the pressure as predicted by Eq. 5.27b.

[38] The Debye law does apply to metals. In metals, the heat contained in the electrons close the Fermi edge leads to linear dependence of the heat capacity on temperature.

Second, Eq. 5.29 contains a term $h\nu$ for the energy of each photon. This term must be replaced by the momentum per photon, which is $h\nu/c$ because of $E = cp$ for photons. However the factor c comes back in when counting the number of photons hitting the wall per unit time. This number is proportional the velocity of the photons, which is c. The factor of 1/3 results from the directions of the photons being distributed isotropically. If some photon hits the surface under an angle θ (where θ = 0 at vertical incidence), the momentum transfer is proportional to cos(θ). The rate at which the photons hit the surface is proportional to cos(θ), as well. The pressure therefore contains a term <$\cos^2(\theta)$>, which is 1/3 for isotropic distributions.

The Planck distribution itself is not needed in the following. We continue with some rather elementary thermodynamic considerations. Based on Eq. 5.27b, an expansion by ΔV requires a work, W, given as

Eq. 5.30

$$W = \int_{V}^{V+\Delta\varsigma} -p\mathrm{d}V = -\frac{b}{3}T^4\Delta V$$

The internal energy changes by

Eq. 5.31

$$\Delta U = bT^4\Delta V$$

Because $\Delta U = Q + W$ and $Q = \Delta U - W$, the heat is given as

Eq. 5.32

$$Q = \frac{4}{3}bT^4V$$

$\Delta S = Q/T$ leads to

Eq. 5.33

$$\Delta S = \frac{4}{3}bT^3\Delta V$$

If zero entropy is assigned to zero volume, the entropy can be defined in absolute terms:

Eq. 5.34

$$S = \frac{4}{3}bT^3V$$

If the entropy remains constant (as in adiabatic processes), it follows that

Eq. 5.35

$$T^3V = const$$

Using $p \propto T^4$ leads to

$$pV^{4/3} = const \qquad \text{Eq. 5.36}$$

The photon gas therefore has an adiabatic index of 4/3, similar to relativistic degenerate matter (section 5.7). Note than an equation of the form $\gamma = c_p/c_V$ (which exists for the particle gas) would not make sense for the photon gas because c_p does not exist for the photon gas. Temperature cannot be varied at constant pressure for the photon gas.

It is instructive to insert numbers for the photon gas and the particle gas Fig. 5.9. This comparison requires the number density of photons, which is $60.4(k_B T)^3/h^3c^2$. The calculation proceeds similar to Eq. 5.29.

– The average energy per photon of 2.7 $k_B T$ (ideal monatomic gas: 1.5 $k_B T$)
– The pressure of 0.9 $k_B TN/V$ (ideal gas: $k_B TN/V$)
– The entropy per photon of 3.6 k_B ($3k_B$ per particle for solid crystals, following the rule of Dulong and Petit[39]).

Because the pressure of the photon gas is proportional to T^4, it plays no role at room temperature compared to the pressure of the particle gas.[40] Otherwise, the ideal gas law would have to be modified accordingly. Even in the center of the sun ($T \approx 15{\cdot}10^6$ K), the photon pressure is insignificant.

Table III. Numerical comparison of classical ideal and photon gas functions. Here the ideal gas is 1.00 mol of monatomic argon at $P = 1.01\times10^5$ Pa, $V = 2.47\times10^{-2}$ m^3, and $T = 300$ K.

Function	Classical ideal gas	Photon gas
N	6.02×10^{23} atoms	1.35×10^{13} photons
U	3.74×10^3 J	1.51×10^{-7} J
P	1.01×10^5 Pa	2.04×10^{-6} Pa
S	155 J/K	6.71×10^{-10} J/K

Fig. 5.9
Some numerical values for comparing photon gas and ideal gas. The pressure of the photon gas at room temperature is negligible compared to the pressure of the molecular gas. This is the case even in the center of the sun ($T \approx 15{\cdot}10^6$ K). Only in the center of large and bright stars does the photon gas make a significant contribution to the pressure. The temperature there is around 10^9 K.
Source: Footnote 36

The pressure of the photon gas is important in the star-forming regions (where it removes the gas from cloud). It is also important in the baryonic acoustic oscillations, BAOs, where it lets the baryonic matter expand (section 5.9).

<u>Remember:</u>

– Even though photons are massless, they carry an energy and a momentum.
– Because photons do not scatter from other photons, a thermal equilibrium (resulting in a black-body spectrum) only forms in contact with matter that has a similar temperature.
– In contrast to the particle gas, the number of photons is a function of volume and temperature.
– The pressure and the energy of the photon gas scale as T^4. In stars, the pressure of the photon gas is only significant at high temperatures (only in the cores of heavy stars).
– The photon gas has an adiabatic index of 4/3.
– The entropy per photon is 3.6 k_B.
– The pressure of the photon gas is important in the star-forming regions and the in the BAOs.

[39] The entropy per particle of the ideal gas depends on density and temperature, following the Sackur-Tetrode equation.
[40] The number of photons in a liter of gas at room temperature is comparable to the number of atoms. The momentum of the photons is much smaller and they therefore hardly contribute to the pressure.

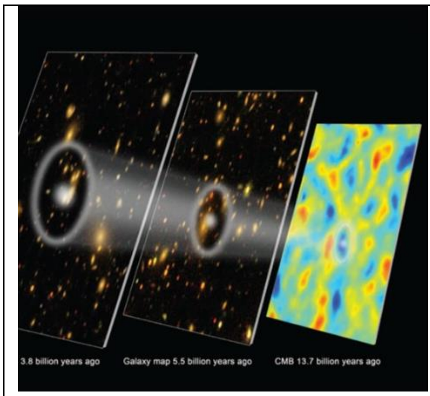


Fig. 5.10
The bright areas on the map of the CMB are the shells produced by the baryonic acoustic oscillations.
cerncourier.com/a/baryon-oscillation-spectra-for-all

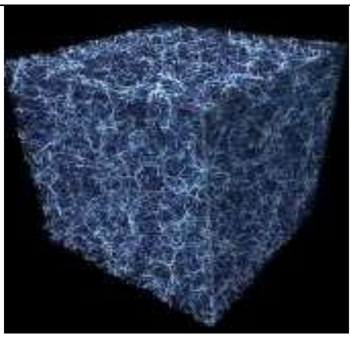


Fig. 5.11
On a large scale, the universe consists of "voids" and "filaments".
en.wikipedia.org/wiki/Filament_(cosmos)

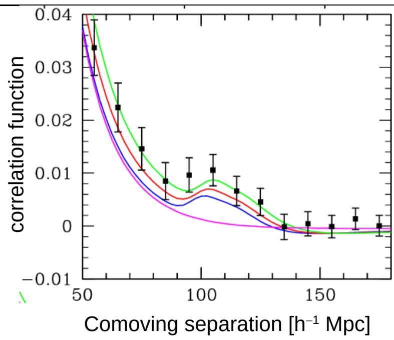


Fig. 5.12
Maps of the distribution of galaxies can be analyzed with regard to the two-point correlation function. The correlation function shows a bump at a distance matching the first maximum in the power spectral density of the CMB from Fig. 3.3.
Adapted from *Eisenstein, Daniel J. et al. The Astrophysical Journal 633, (**2005**): 560. doi.org/10.1086/466512.*

5.10 Baryonic acoustic oscillations

For the earliest self-gravitating clouds, the pressure of the photon gas modifies the picture. We see those clouds as patterns in the cosmic microwave background, but the brighter regions are not the centers of these clouds, as one might think. They are the precursors of the “cosmic web”, which consists of voids, shells, and filaments. The cores of the clouds, which later formed the cosmic web, are dark in the CMB because they contain dark matter. Those clouds were rather large (> 1000 light years), they were not very dense (a few particles per m^3), and the baryonic fraction (the non-dark matter) consisted of plasma. Under these conditions, the pressure of the photon gas (section 5.9) gains an importance not familiar to us. At ambient conditions on earth, the pressure exerted by the thermal photons is 2×10^{-11} bar (Fig. 5.9). The pressure of the photon gas in these early clouds was larger (10^{-7} bar) because the clouds were hot (hotter than 3000 K). Also, the pressure of the particle gas was lower because of the low density (< 10^{-15} bar). In this case, gravity is not counteracted by gas pressure but rather by photon pressure. There is a twist here insofar, has photons only couple to ordinary (“baryonic”) matter. The larger part of the matter was dark matter and therefore did not feel the photon pressure. The outcome of this situation are the baryonic acoustic oscillations (BAOs). BAOs amount to a peculiar form of sound waves. The BAOs continued until the matter became transparent because of recombination at a temperature of less than 3000 K. That stopped the oscillations. Shells of baryonic matter kept expanding until different shells merged, forming the cosmic web.

<u>Remember:</u>
- The BAOs produced the large-scale structure of the universe, namely the “cosmic web”.
- In the early clouds of plasma, the pressure of the photon gas was larger than the pressure of the particle gas.

5.11 The biosphere needs the dark and cold sky

In the course of the gravothermal catastrophe, a loss of energy to the outside heats a self-gravitating cloud. The loss of energy requires an outside, from which the energy does not return. It requires an expanding and cold sky.

Similarly, the biosphere needs the cold sky as an entropy sink. There are two versions of this argument. The first is based on the Carnot process, while the second does not take recourse to a heat flow.

The first argument applies to convection in the atmosphere, in the oceans, and in the crust of the earth. In all these case, there is a net flow of heat either between the sun and the sky or between the earth's core and the sky.[41] The weather, the oceans, and plate tectonics are tremendously powerful drivers of structure formation. They create order.

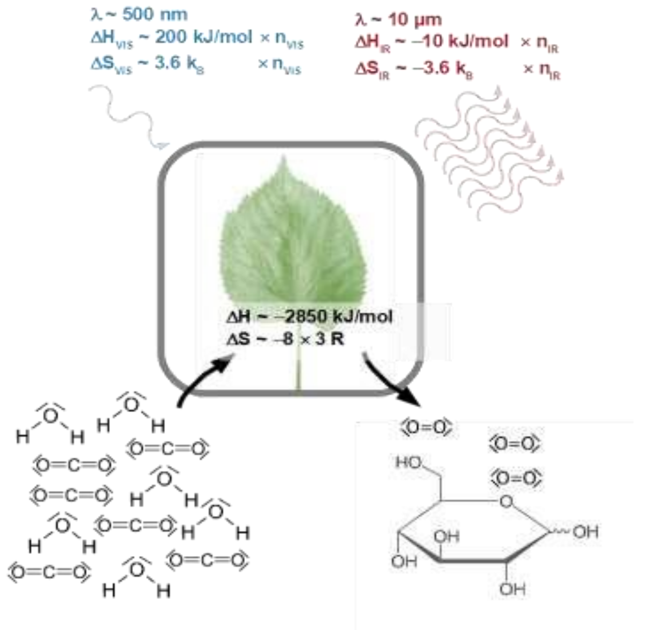


Fig. 5.13
Photosynthesis lowers the entropy. The excess entropy is dumped into the sky in the form of thermal photons..

Photosynthesis creates order by a chemical mechanism not connected to heat flow. Entropy decreases, when small molecules turn into large molecules. At the same time, few solar photons ($\lambda \approx 500$ nm) turn into many more thermal photons ($\lambda \approx 10$ µm, Fig. 5.13). Each photon carries away an entropy of 3.6 k_B (section 5.9). The global increase in entropy upholds the second law of thermodynamics.

The starry sky above him filled Immanuel Kant's mind with admiration and awe.[42] Arguably, the dark space between the stars is just as mind-boggling as the stars.

Remember:

– In the biosphere, entropy decreases locally when small molecules are turned into large biomolecules. Globally, the entropy increases because numerous thermal photons are radiated away into the cold sky.

---

[41] The heat in the earth's core partly originates from radioactivity and partly from a sustained contraction, continuing the gravothermal catastrophe, which formed the earth.
[42] Two things filled his mind with admiration and awe: the starry sky above him and the moral law within him.

5.12 Cosmic expansion is an arrow of time

Among the question addressed in physical chemistry is what distinguishes the future from the past. The mechanism doing that are the "arrows of time". The microscopic dynamical laws are invariant under time reversal.[43] They are not arrows to time. An easy answer to the question of why the future differs from the past is that time-asymmetric processes start from special, unlikely initial conditions. Any solution to a system of ordinary differential equations (ODEs) or partial differential equations (PDEs) requires such initial conditions. The solutions often require boundary conditions, as well.

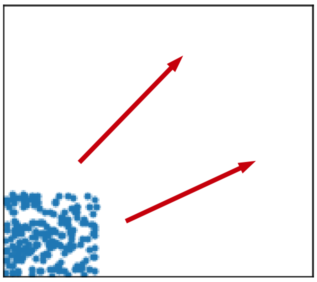


Fig. 5.14
Arrows of time mostly often are the consequence of unlikely initial conditions. Molecules, which are confined to some space, initially, and are released from that space, will just about never return to that configuration, although the microscopic dynamic laws would allow for that. The initial state is exceedingly unlikely.

In the case of cosmology, the distinction between the initial conditions and the boundary conditions is blurred because a look into the distant universe always amounts to a look into the early universe. Still: If the universe would switch form expansion to contraction, that would not change the initial condition (the Big Bang).

Consider the "radiation arrow of time", where radiation here stands for causality. A radio antenna emits waves. These waves are solutions to Maxwell's equations. Maxwell's equations allow those same waves to run backwards and to be absorbed by the antenna. That, however, never happens. Never ever do concentric waves encoding music converge onto an antenna, being completely absorbed there.[44] More technically, the antenna amounts to a source term of a heterogeneous partial differential equation. Sources in that sense always are in the past.

Arguably, the waves never return, because they are dispersed into the ever expanding universe. If the universe would switch to contraction, the sky would turn bright and hot, for one. Also, all those waves, which previously disappeared in the vast space, might come back. To what extent a contraction of the universe would affect causality is not easily answered.

<u>Remember:</u>

- Arrows of time mostly are caused by unlikely initial conditions.
- In the case of the universe, the Big Band is such a condition. It has also created a special boundary condition, which is the dark sky.
- Cosmic expansion contributes to the difference between the future and the past.

[43] Section 14.2 discusses an exception, which does not usually count as an arrow of time.
[44] A similar argument applies to black holes. The field equations of general relativity also allow for „white holes“, where „hole“ is a misleading term because nothing disappears into the „hole“. White holes are just as unlikely as concentric radio waves propagating towards an antenna.

6 Rotating self-gravitating systems

The gravitational collapse of rotating clouds of gas and dust may lead to a disk. The motions along the axis of rotation are damped. After some time, the bodies are enriched at the disk. The circular motion in the plane, on the other hand, is stabilized by centrifugal force. Examples are the spiral galaxies, planetary systems, the rings of Saturn, protoplanetary disks (section 6.1), and accretion disks (section 6.3).

The bodies mostly rotate at the Kepler velocity, given as $(GM/r)^{1/2}$. The Kepler velocity results from the requirement that the centrifugal force is equal to the centripetal force (Eq. 6.1). The orbits are almost circular, because ellipticity is damped out by the interactions. Disks with stronger interactions usually consist of many small bodies, which may or may not include gas. Sometimes the interactions are strong enough to heat the disk. This happens in the accretion disks. These exist near some compact objects, where the field of gravitation and its gradient are strong.

<u>Remember:</u>

- Rotating self-gravitating systems may form disks.
- If the interactions between bodies are strong enough, they dampen elliptic motion so that the orbits are circular.
- In the accretion disks, interactions lead to frictional heating.

6.1 The streaming instability in the protoplanetary disks

Today, protoplanetary disks can be observed better than a few years ago. A famous image is shown in Fig. 6.1. Protoplanetary disks are an active field of research. There are numerous instabilities other than the streaming instability.

The streaming instability fills a gap in our understanding of how the planets formed from dust. The problem is that colliding particles do not stick together, when their size is above a few millimeters. The adhesion energy does no suffice to dissipate the kinetic energy contained in the impact. The adhesion energy scales as the surface area of the bodies, while the kinetic energy scales as the volume. The surface-to-volume ratio decreases as the size of the bodies increases. Collisions cannot explain the formation of large, solid bodies.

This is where the streaming instability comes into play. The streaming instability occurs if the disk contains both dust and gas. The force equilibrium for the dust is

Fig. 6.1
This image of a protoplanetary disk (HL Tauri) from 2014 (acquired with a microwave telescope, named ALMA) has become famous. More recent images of other protoplanetary disks also show spiral arms.
https://technobyte.org/hl-tauri-alma-birth-solar-system-/

$$\frac{mv^2}{r} = \frac{GmM}{r^2} \qquad \text{Eq. 6.1}$$

m is the mass of the test body, M is the mass of the star. Solving this equation for the velocity leads to the Kepler velocity:

The force equilibrium for the gas is different insofar, as it involves a pressure gradient:

Eq. 6.2

$$\frac{mv^2}{r} = \frac{GmM}{r^2} - \nabla p$$

The pressure gradients lets the overall centripetal force decrease, thereby decreasing the velocity of the gas. The dust particles also feel the pressure gradient, but it plays no role compared to inertia and gravity. Because the gas orbits the star slower than the dust, the dust feels a headwind. Friction with the gas slows down the motion of the dust. Dust drifts inwards, pulled by the imbalance between centripetal and centrifugal force. This is the "radial drift".

As the gas slows down the dust, the dust conversely accelerates the gas. Assume that some fluctuation has increased the density of the dust at some distance from the star. The dust then more effectively accelerates the gas and the increased velocity of the gas decreases the radial drift of the dust at this particular distance from the star. The dust orbiting the star at a slightly large distance bumps into the dust with decreased drift, which further increases the density of the dust and lets the drift decrease further. There is positive feedback, leading to the "streaming instability". The dust forms filamentous structures, which eventually collapse into spheres under the influence of gravity.

<u>Remember:</u>

- Particles with a size larger than a few millimeters do not clump together upon collision because the adhesion energy does not suffice to dissipate the kinetic energy.
- The formation of planets still is possible because of the streaming instability.
- The dust in protoplanetary disks moves inwards (“radial drift”), while the gas moves outwards.

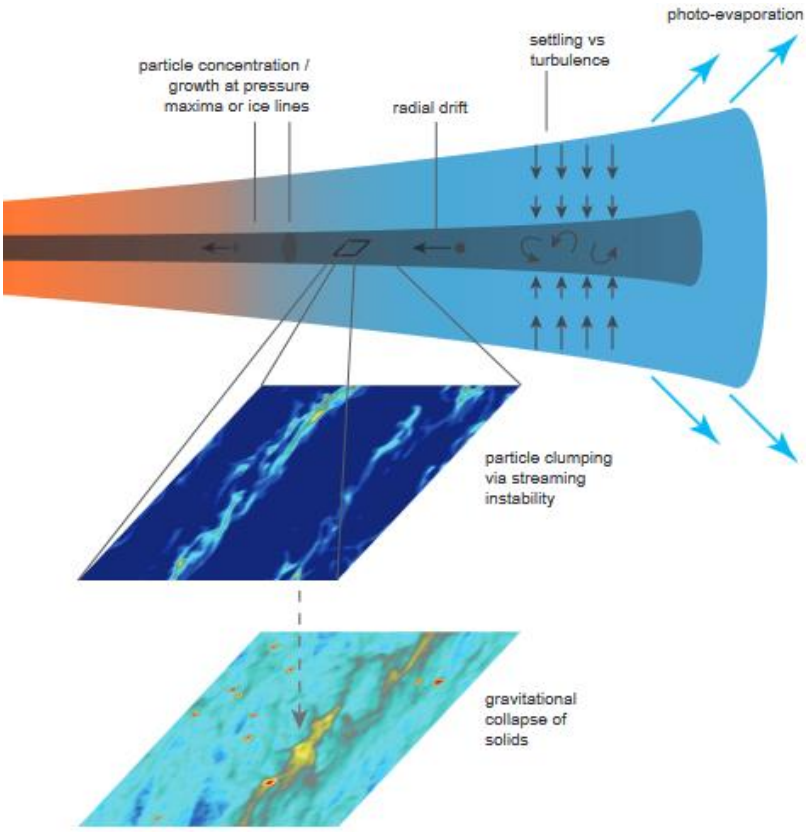


Fig. 6.2
The mechanism behind the streaming instability.
https://jila.colorado.edu/~pja/saas-fee/streaming.pdf

6.2 *Advanced Topic: Exoplanets*

The public takes an enormous interest in planets outside the solar system. About 5000 exoplanets have "confirmed" status. Most have been detected indirectly, based on a periodic variation of the Doppler shift in the spectrum of the star or on a periodic small change in brightness following an occultation of the star by the planet. Spectra of planets can also be obtained indirectly if the planet passes behind the star and the spectrum of the entire system then changes slightly.

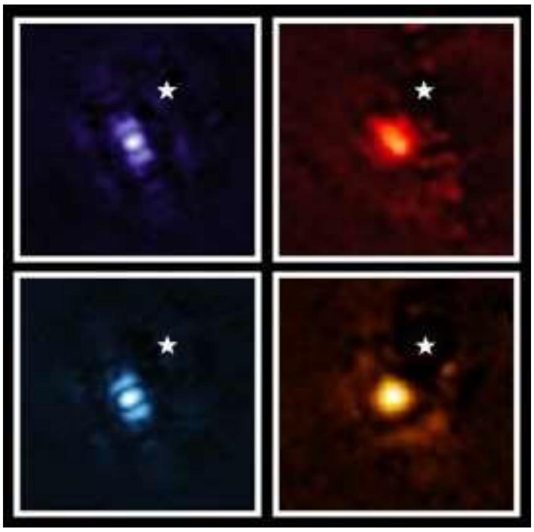


Fig. 6.3
Images of exoplanet HIP 65426b obtained with the James Webb Space Telescope.
Purple: λ = 3 µm, blue: λ = 4.44 µm,
yellow: λ = 11.4 µm, red: λ = 15.5 µm.
The planet is not a circle because of limits in the telescope's optics. The white asterisk indicates the position of the star, covered by the coronograph.
Text from *www.sciencenews.org/article/james-webb-spacetelescope-first-exoplanet-imag*

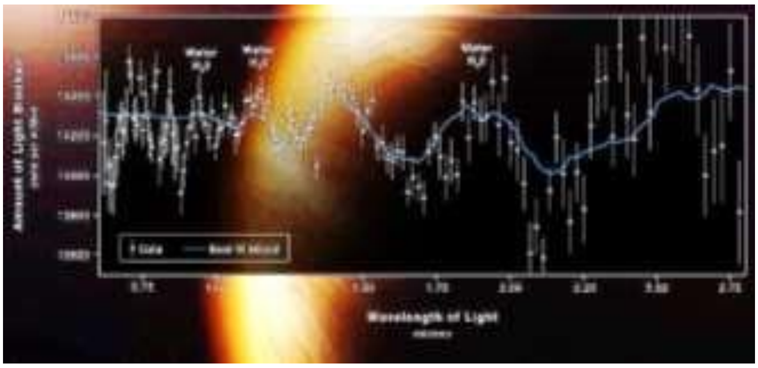


Fig. 6.4
NASA's James Webb Space Telescope has captured the signature of water, along with evidence for clouds and haze, in the atmosphere surrounding a hot gas giant planet orbiting a distant Sun-like star. WASP-96 b is one of more than 5 000 confirmed exoplanets in the Milky Way. Located roughly 1,150 light-years away in the southern-sky constellation Phoenix, it represents a type of gas giant that has no direct analog in our solar system. With a mass less than half that of Jupiter and a diameter 1.2 times greater, WASP-96 b is much puffier than any planet orbiting our Sun. And with a temperature greater than 1000°F, it is significantly hotter.
www.nasa.gov/image-feature/goddard/2022/nasa-s-webb-reveals-steamy-atmosphere-of-distant-planet-in-detail
These are near-infrared spectra (NIR spectra, Box 6.1). The bands correspond to overtones and combination tones.

The information acquired in this indirect way includes the orbit, the mass, the size, and – to some extent – the chemical composition as inferred from the spectra. Those studies reveal a large variability of properties. Naturally, large planets (much larger than Jupiter) are more easily seen and studied than small planets (similar to Earth).

The James-Web space telescope is equipped with a coronograph. A coronograph is a disk that blocks the light from the central star. Planets have been seen directly with this device (Fig. 6.3). Near-infrared spectra have been obtained (directly in emission, Fig. 6.4). In the foreseeable future, such direct studies will be limited to a few dozen examples, simply because those planets have to be close by. Along similar lines: The search for extraterrestrial intelligence (SETI) is a waste of time (an opinion). It is hard to see how civilizations so far away from us would be able to talk to us.

<u>Remember:</u>

- There are a few thousand confirmed exoplanets.
- Spectroscopy (direct or indirect, often in the NIR range) in a few cases has allowed to study their composition.

6.3 Accretion disks

When matter quickly orbits a compact object, the tidal forces rip all large bodies apart.[45] The matter is composed of a hot plasma. There is difference between accretion disks around neutron stars and stellar-size black holes, on the one hand, and accretion disks around supermassive black holes, on the other. Accretion disks around stellar-size black holes are among the most energetic objects in the universe. Shortly before disappearing into the black hole, the gravitational energy of whatever is bound to disappear is a sizeable fraction of its rest mass. The kinetic energy is half of this energy, following the virial theorem (Eq. 5.17). Some of the most violent events we know occur in those accretion disks. Many x-ray sources are accretion disks. The accretion disks around the supermassive black holes are brighter because of their larger size. They form the "quasars" (see also footnote 102). Quasars are not quite as hot as the accretion disks around stellar-size black holes because the distance to the center is larger and the gradients in the velocity of matter orbiting the center are less strong. The fast and highly energetic events (including the gamma-ray bursts) occur in the more compact accretion disks.

<u>Remember:</u>

- In accretions disks, friction heats the material orbiting the compact object.
- Some of the most energetic events known to us happen in the accretion disks around stellar-size black holes.
- The accretion disks around supermassive black holes are among the brightest objects in the sky. These are the quasars.

---

[45] "Tidal forces" come about when a body of finite size orbits another body and when the centripetal and the centrifugal force equal each at the center of the body, but not elsewhere.

Box 6.1
NIR spectroscopy

Near infrared spectroscopy (NIR spectroscopy) is a compromise. UV-Vis spectroscopy is cheaper. LEDs, photodiode arrays and light guides cost close to nothing. (The gratings needed to disperse the colors are a little more expensive.) But UV-Vis spectroscopy is not very useful for chemical analysis because all lines are vibronically broadened, forming bands.
NIR spectroscopy can still rely on LEDs, diode arrays, and light guides. Data analysis is no fun, because the adsorption mostly is caused by combination tones (frequencies being sums and differences of other vibration frequencies). Because the combinations are numerous due to the combinational explosion, a detailed understanding of NIR spectra mostly is out of reach. But if there is a hypothesis for what to expect, NIR is useful and more helpful than UV-Vis spectroscopy. It is the method of choice to study the atmospheres of exoplanets.
IR spectroscopy is even more useful for chemical analysis, but IR spectroscopy requires cooled detectors, purging the chamber with nitrogen, and an interferometric scheme of measurement (Fourier transform IR spectroscopy).

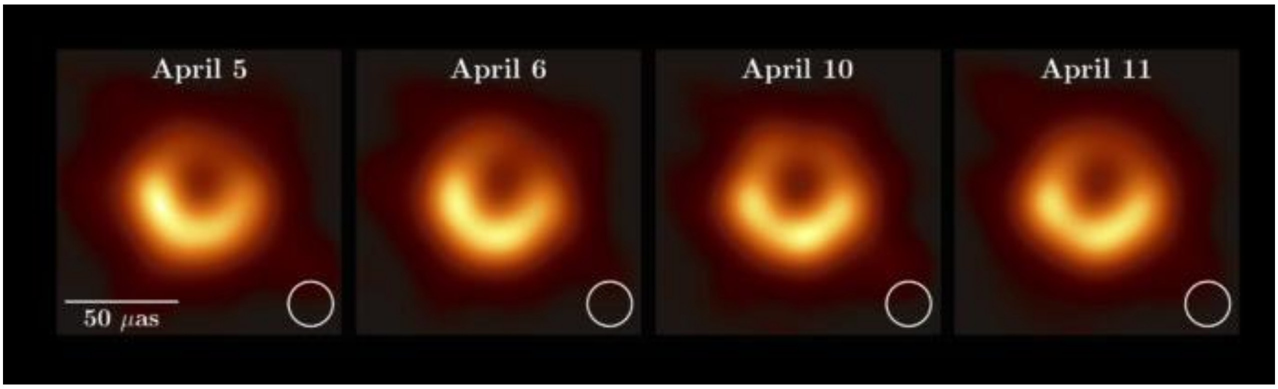


Fig. 6.5
The first images of an accretion disk around a supermassive black hole (at the core of the galaxy M87) date from In 2019. Advanced methods of radio astronomy were used. Until then, the models were largely based on spectroscopy, the variability of the objects, and the jets that are sometimes emitted vertical to such disks. The variability can also be seen in this series of images.
aasnova.org/2019/04/10/first-images-of-a-black-hole-from-the-event-horizon-telescope/

6.4 Advanced topic: Digital imaging with coherent detectors, the pair correlation function

In the discussion of how scattering experiments lead to the pair correlation function, we encountered the difference between coherent and incoherent detectors. Coherent detectors (such a radio antenna) can detect the phase in addition to the intensity, while incoherent detectors (such as photodiodes or x-ray detectors) only detect the intensity. In scattering, the sample amounts to a distribution of a “scattering length density”. The detectors sees a superposition of waves, which all have been scattered under the same angle. A coherent detector measures the complex amplitude of this superposition, which amounts to a Fourier-transform. (That requires weak scattering and a large distance of detector from the sample.) If an array of coherent detectors is employed, the scattering length distribution representing the sample can be calculated from the complex amplitudes at all detectors, using an inverse Fourier transform. That is a calculation on a computer, leading to digitally computed image. A similar result can be achieved by optical means, replacing the detector array with an optical lens. When focusing all beams to the points constituting the image, the lens creates an inverse Fourier transform of the amplitude distribution at its center plane. The lens also is a coherent device. It does not scramble the phases of the optical waves and it performs an inverse Fourier transform, by which the image is reconstructed.

The image can be obtained physically (with a lens) or digitally. Should the phase have been lost, image reconstruction is no longer possible. All that can be obtained is a pair correlation function, but that is far more than nothing (Fig. 6.6). The pair correlation is the starting point for study of local order in amorphous materials. For instance, it shows a maximum at the next-neighbor distance. It may also display an amorphous halo.

The resolution of any image is proportional to λ/D, where D is the diameter of the lens (of the "pupil"). The larger the mirror of a telescope, the better the angular resolution. Modern telescopes reach a resolution if about 1 milliarcseconds (1 mas). When creating the image digitally, the diameter of the pupil is replaced by the distance between the antennae. In the case of the ALMA observatory, the maximum distance is 16 km. With a wavelength of 1 mm, that also leads to a resolution of about 1 mas. The image from Fig. 6.1 was taken by the ALMA observatory. Other radio-telescopes are distributed over the entire planet. The diameter of the pupil then is the diameter of the earth. The technique is called "very long baseline interferometry" (VLBI). The resolution can be as low as 12 µas (Fig. 6.5). That is best resolution achieved in astronomy so far. Doing even better would require radio telescopes in space. These are conceivable, but there are no plans to build those any time soon.

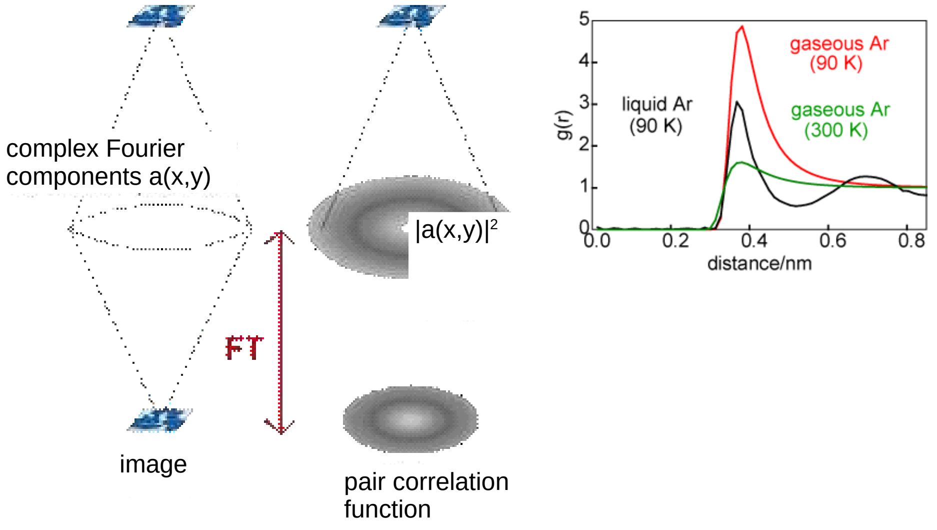


Fig. 6.6
An array of coherent detectors allows the reconstruct an image, while incoherent detectors only provide the pair correlation function. The phase information is needed to reconstruct the image.

7 Black hole thermodynamics

Karl Schwarzschild was among the first researchers to take notice of general relativity. Soon after Einstein had published the field equations named after him in 1915, Schwarzschild found a spherically symmetric solution to these equations, which looked funny at the time. Today, this solution constitutes the black holes[46] – objects from which there is no escape. To the outside observer, the black hole has a mass, an angular momentum, and possibly a charge. This list is complete. To the outside observer, the black hole has no other properties, which would be independent from these. That black holes are so simple is the content of the "no-hair theorem". Roger Penrose, who formulated the no-hair theorem, at the same time showed that black holes do not necessarily require the perfect spherical symmetry inherent to the Schwarzschild solution. At the moment of formation black holes have an irregular shape. Before Penrose's work, there were concerns that black holes would not actually exist because the slightest deviation from spherical symmetry would invalidate the entire concept.

One might think that if a black hole has so few properties, it would also have a low entropy. However, Bekenstein and Hawking showed that black holes contain all the numerous microstates, which they swallowed when they were formed. The entropy of black holes is very large. Bekenstein and Hawking predict the entropy of the black hole as

Eq. 7.1

$$S_{BH} = \frac{k_B A}{4 l_P^2}$$

The index *BH* stands for black hole (or Bekenstein-Hawking). *A* is the surface area of a black hole also called "event horizon". Note that the entropy is proportional to the surface area, as opposed to the volume of the black hole, which is actually undefined due to the singularity in space-time at its center.

A temperature of a black hole can defined based on the relation $1/T = \mathrm{d}S/\mathrm{d}U$. The result is

Eq. 7.2

$$T_H = \frac{\hbar c^3}{8\pi G m k_B}$$

The subscript *H* stands for Hawking. Inserting values, one finds the temperature of stellar-size black holes to be small. A black hole with the mass of the sun has a Hawking temperature of 60 nanoKelvin.

In principle, this temperature of a black might be determined experimentally, based on a special form of thermal radiation. Due to non-local quantum correlations across the horizon of the black hole, particle-antiparticle pairs sometimes form at the edge of the black hole, where one of the two escapes.[47] The escaped particles have a thermal energy distribution, revealing the temperature of the black hole. The "Hawking radiation" lets black holes disperse over time. The evaporation time is

Eq. 7.3

$$t_{ev} = \frac{5120\pi G^2 M^3}{\hbar c^4} = 2.1 \cdot 10^{67}\ \text{years} \left(\frac{M}{M_\odot} \right)^3$$

---

[46] Do not confuse the black hole with the black body.

[47] Quantum correlations are related to „entangled states". That is a complicated topic, which is sometimes discussed in the context of the quantum measurement.

$M_{\odot}$ is the mass of the sun. Black holes are not necessarily an ultimate final state of the universe (see also footnote 13).

A side remark: Black holes have a negative heat capacity similar to self-gravitating gas clouds. If they swallow matter, they become colder although energy is supplied to them. If they emit energy via Hawking radiation, they become lighter and hotter.

Black holes are subject to thermodynamics but there is no simple way to apply thermodynamics to all of general relativity. There is no straight-forward way to assign an entropy to the curvature of space-time. If – as Einstein claimed[48] – thermodynamics is universal, it should also be applicable to the curved space-time. There are proposals and conjectures for a "gravitational entropy", but there is no consensus. The problem is easily explained. In the calculation of the entropy of the ideal gas, a box with a certain volume is filled with the wave packages representing the particles. The wave packages might also represent photons and thermodynamics therefore applies electromagnetic waves (section 3.2). At some given temperature, the number of different ways of fitting those wave packages into the box can be counted. The log of this number essentially is the entropy. However, there is no such box in general relativity. Space itself is the dynamic variable.

How did Bekenstein then arrive at the entropy of the black hole? The black hole is embedded in flat space. The entropy of whatever was contained in the flat space around the black hole is well defined. As the black hole absorbs matter, it also absorbs the associated entropy. Bekenstein found this entropy to be proportional to the surface area.

<u>Remember:</u>

- Black holes have an entropy and a temperature.
- The entropy is large, the temperature is low.
- Due to quantum non-locality, black holes decay – albeit very slowly.

[48] The quote is: "Thermodynamics is the only physical theory of universal content which I am convinced, within the framework of applicability of its basic concepts, will never be overthrown."

8 Degenerate matter and the Chandrasekhar limit

White dwarfs and neutron stars are among the most fascinating objects in the universe. The density of white dwarfs is a million times higher than what we know from earth. For the neutron stars, the density is 10^{14} times higher. In both cases, the density is limited by the Pauli exclusion principle, applied to electrons in the case of the white dwarfs and to neutrons in the case of the neutron stars. Neither density nor pressure do depend on temperature. Quantum mechanics takes the role, which thermodynamics has for the gas.

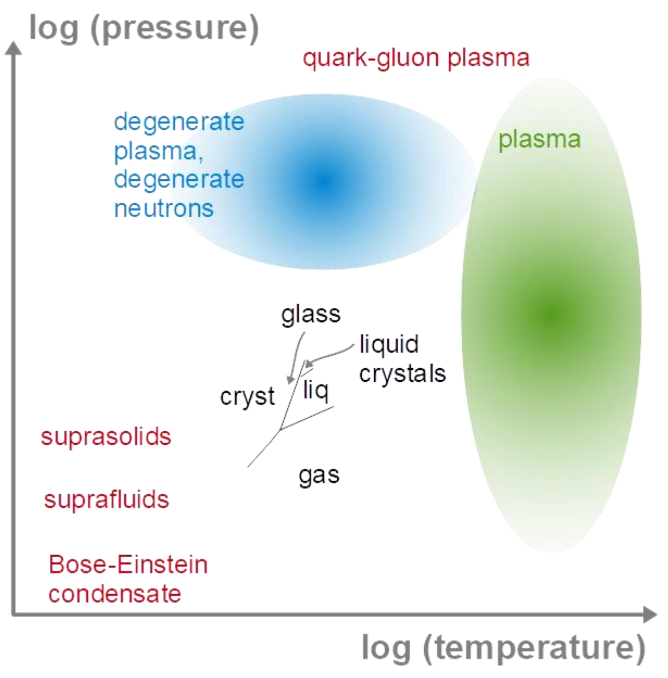


Fig. 8.1
States of matter
States in red exist on earth but not in space.

8.1 Recap: The uncertainty relation

On a mathematical level, the uncertainty relation is a property of the Fourier transform. A sharp distribution in real space turns into a broad distribution in Fourier space (in *k*-space) and vice versa. The "time-bandwidth-product" of a pulse cannot be less than some minimum value, which unfortunately varies in the literature because there are different definition this product. In quantum mechanics, a similar product applies to the widths of pulses in position space and in momentum space. Momentum is associated to the Fourier transform of a wave function in position space because of the DeBroglie relation, $p = \hbar k$. Following the uncertainty relation, the product of Δp and Δx can never be less than $\hbar/2$:

$$\Delta p \Delta x \geq \frac{\hbar}{2}$$ Eq. 8.1

Because of the uncertainty relation, any wave package occupies a region with some minimum size in the space spanned by position and momentum. This is a 6-dimensional space, not to be confused with "phase space" (Fig. 4.2, a 6*N*-dimensional space with *N* the number of particles). The position-momentum space can be thought of as a checker board.[49] There is a finite number of wave packages which can be placed into this space (if it is bounded).

In a gas, the wave packages are far apart. In degenerate matter, they are compact (Fig. 8.2b). Fig. 8.2b assumes that each particle requires its own wave package as demanded by the Pauli exclusion principle (related to the exchange interaction). The Pauli principle holds for spin-1/2-particles (for "fermions"). Electrons, protons, and neutrons are fermions (photons are not).

A second important consequences of the uncertainty relation is that a microstate in phase space (a space with 6*N* dimensions rather than 6 dimensions) has a finite volume of $(\hbar/2)^{3N}$. For that reason, the

[49] The region occupied by a wave package will not usually be quadratic, see the amoeba to the right in Fig. 8.2a.

number of possible microstates can be counted. The number of microstates is usually called Ω_s, where the subscripts s indicates that these microstates are compatible with a macrostate called “s” (see section 5.1).

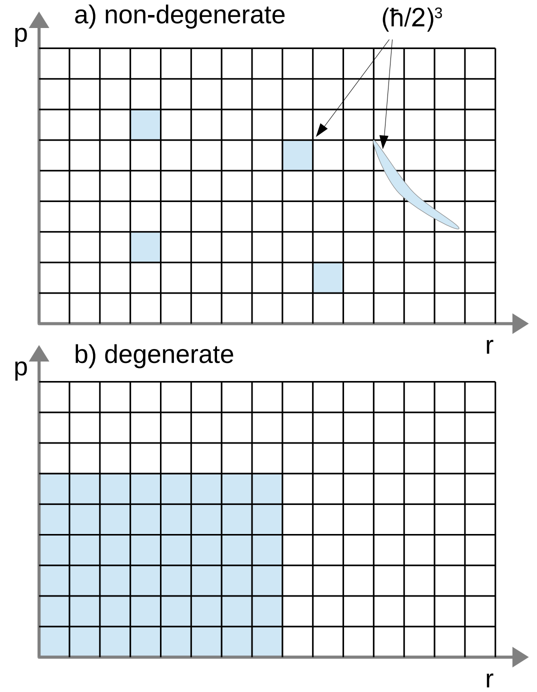


Fig. 8.2
Simplifying matters, wave functions are depicted as blue squares with sharp boundaries. What is correct about this representation, is that there is a lot of free space between wave packages for gases (a), but not for degenerate matter (b). Liquids composed of atoms or molecules do not fit into this picture because the size of the atoms and molecules is larger than the size of boxes (larger than the thermal wavelength, not discussed here). Degenerate matter is composed of electrons, protons, or neutrons, the size of which can be smaller than the diameter of the wave package.

For any given temperature, a typical size of a wave package in space can be calculated from the uncertainty relation. The following relations are input to this calculation.

Eq. 8.2

$$\Delta p \approx p$$
$$E = \frac{1}{2} m\text{v}^2 = \frac{p^2}{2m} \Rightarrow p = \sqrt{2mE}$$
$$E = \frac{3}{2} k_B T$$

Without going into the details, the characteristic size (called "thermal wavelength", Λ_{th}) comes out to be[50]

Eq. 8.3

$$\Lambda_{th} \approx \frac{h}{\sqrt{2\pi m k_B T}}$$

The thermal wavelength occurs in the Sackur-Tetrode equation for the entropy of the ideal gas, which is

Eq. 8.4

$$\frac{S}{k_B N} = \ln\left(\frac{V}{N\Lambda_{th}^3}\right) + \frac{5}{2}$$

[50] This calculation applies to the gas phase. In degenerate matter, energy and temperature are not related as $E = 3/2\ k_B T$. In degenerate matter, finite temperature smoothens the edge between the occupied and the non-occupied states close the Fermi-level.

The ratio V/Λ_{th}^3 counts the number of thermally accessible squares in Fig. 8.2a. The term N in the denominator on the right-hand side occurs because the particles are indistinguishable. If there was not such factor of N, there would be a mixing entropy between two identical gases.

Degenerate matter is characterized by particles being close-packed in position-momentum space. Note that Δx increases with decreasing mass. For that reason, the electrons in the conduction band of metals are degenerate, while the ions on the lattice are not.

The fact that two particles can never be on the same square in position-momentum space is related to the Pauli principle and, also, to what is called "exchange interaction" in chemistry. The exchange interaction prevents overlap between orbitals. It is the most important repulsive interaction in chemistry. The exchange interaction is rooted in quantum mechanics.

Remember:

- Because of the uncertainty relation, spin-1/2 particles occupy a minimum volume in position-momentum space.
- The number of microstates can be counted.
- Overlap is prohibited by Pauli exclusion principle, related to the exchange interaction.
- In degenerate matter, the particles densely occupy position-momentum space.

8.2 Degenerate matter and its behavior under compression, the Chandrasekhar limit

We aim to understand the Chandrasekar limit, which is a maximum mass of a body of degenerate matter. Beyond that mass, the body implodes under the influences of gravity. In order to understand how this limit comes about, we need the pressure-volume relations of the plasma, of the two types of degenerate matter, and of the gravitational pressure.

"Degenerate" refers to a state of matter with a density limited by the uncertainty relation and the Pauli principle.[51] Degeneracy is known from electrons in metals. The delocalized electrons form bands. The energy states are filled up from bottom to top. The energy of the highest occupied level is the "Fermi energy". In space, some cores of stars, the white dwarfs, and the neutron stars consist of degenerate matter. The constituents of white dwarfs are ions and electrons (a plasma, but a degenerate plasma). Neutron stars mostly consist of neutrons, as the name says.

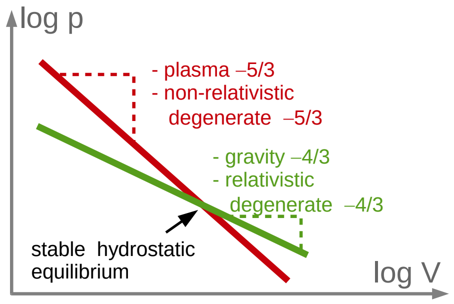


Fig. 8.3
Pressure-volume relation for the plasma, self-gravitating systems and the two forms of degenerate matter.

In contrast to metals, the degenerate matter in space builds up the pressure by gravity. (In metals, this is achieved by the ion lattice.) The pressure-density relation in degenerate matter differs from gases and plasmas. For a plasma, the ideal gas law applies as long as the kinetic energy dominates the overall energy:

Eq. 8.5

$$p = k_B T \frac{N}{V}$$

[51] The transition beween a normal plasma and a degenerate plasma is gradual. There is not sharp step.

The pressure depends on temperature. If temperature increases, the plasma expands until a new hydrostatic equilibrium is reached. If there is no heat flow involved (if the process is adiabatic), pressure and volume follow the relation:

Eq. 8.6

$$pV^{\gamma} = const \qquad p \propto \left(\frac{N}{V}\right)^{\gamma}$$

γ is the adiabatic index, given as

Eq. 8.7

$$\gamma = \frac{c_p}{c_V} = \frac{c_V + R}{c_V} = \frac{f\frac{R}{2} + R}{f\frac{R}{2}} = \frac{f+2}{f}$$

The plasma has three degrees of freedom (three for translation, none for rotation, none for vibration), which leads to $\gamma = 5/3$. Again: We are concerned with adiabatic processes, which are too fast to allow for heat exchange with the outside.

The question now arises as to whether the gravitational pressure of self-gravitating systems increases more slowly than the pressure of the plasma, when the system contracts. If this is the case, the system is stable. We estimate the pressure as the derivative of energy with respect to volume:[52]

Eq. 8.8

$$p = -\frac{\mathrm{d}E_{grav}}{\mathrm{d}V} \approx -\frac{\mathrm{d}}{\mathrm{d}V}\left(\frac{GM^2}{R}\right) \propto -\frac{\mathrm{d}}{\mathrm{d}V}\left(\frac{GM^2}{V^{1/3}}\right) \propto M^2 V^{-4/3}$$
$$pV^{4/3} = const$$

The gravitational pressure therefore increases more slowly than the plasma pressure (Fig. 8.3).[53]

The pressure of degenerate matter is also calculated as the derivative of energy with respect to volume. For each individual particle, the energy is given as

Eq. 8.9

$$E = \frac{1}{2} m \mathrm{v}^2 = \frac{\bar{p}^2}{2m}$$

The momentum is called $\bar{p}$ instead of p to avoid confusion with the pressure. Eq. 8.9 is the non-relativistic energy-momentum relation. The total energy is:

---

[52] Constant entropy was assumed: $\mathrm{d}U = -p\mathrm{d}V + T\mathrm{d}S$.

[53] Food for thought: Would a cloud of CO_2 at room temperature collapse under gravity? The specific heat capacity of CO_2 at room temperature is $c_p \approx 37.12$ J/(mol K).
The question is of some relevance in astrophysics. At some point during a collapse supernova, the temperature is high enough to let photons trigger nuclear fission ("photodisintegration"). This energy spent on the disintegration is then no longer available to counteract gravitational pressure. A new degree of freedom is created.

Eq. 8.10

$$E_{tot} \propto \sum_{particles} \frac{\bar{p}_i^2}{2m}$$

The particles successively fill all possible momentum states (equivalent to the energy states) up to the Fermi edge. Call the "Fermi momentum" $\bar{p}_F$. When counting momenta, we make use of $\Delta x^3 \Delta\bar{p}^3 \geq (\hbar/2)^3$. We replace "≥" by "=" and replace the sum by $\int \Delta x^3\, 4\pi\bar{p}^2 d\bar{p} / (\hbar/2)^3$. Every shell in momentum space of width $d\bar{p}$ contains $\Delta x^3\, 4\pi\bar{p}^2 d\bar{p} / (\hbar/2)^3$ states.

Eq. 8.11

$$E_{tot} = \sum \frac{\bar{p}_i^2}{2m} \propto \int_0^{\bar{p}_F} \frac{\bar{p}^2}{2m} \left(\frac{\Delta x}{\hbar/2}\right)^3 4\pi\bar{p}^2 d\bar{p} \;\propto \Delta x^3 \bar{p}_F^5$$

The momenta are proportional to the wave numbers. These, in turn, are proportional to $1/\Delta x$, which leads to

Eq. 8.12

$$E_{tot} \propto \Delta x^3 \frac{(\hbar/2)^5}{\Delta x^5} \propto \frac{1}{\Delta x^2} \propto \frac{1}{V^{2/3}}$$

The pressure again is the negative derivative of energy with respect to volume:

Eq. 8.13

$$p = -\frac{dE_{tot}}{dV} \propto \frac{1}{V^{5/3}} \quad \Rightarrow pV^{5/3} = const$$

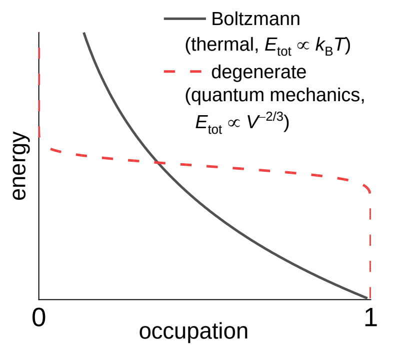


Fig. 8.4
Occupation of states in the thermal and the degenerate plasma

Following Wikipedia, the full expression is[54]

Eq. 8.14

$$p = \frac{(3\pi^2)^{2/3}\hbar^2}{5m} \left(\frac{N}{V}\right)^{5/3}$$

The adiabatic index is 5/3, similar to the plasma. White dwarfs are stable as long as the electrons are non-relativistic. Again: Degeneracy is a quantum phenomenon. The pressure does not depend on temperature as long as the diameter of the wave packets is larger than the distance between the particles.[55] Furthermore, the pressure is lower than in ideal gases. Celestial bodies made of degenerate matter are therefore more compact than ordinary stars.

[54] As shown in section 16.6, the corresponding equation for the adiabatic state of the ideal gas is

$$p = \frac{h^2}{2\pi m} \exp\left(\frac{2}{3}\frac{S}{k_B N} - \frac{5}{3}\right)\left(\frac{N}{V}\right)^{5/3}$$

[55] The wave number increases with increasing kinetic energy. If the inverse wave number becomes smaller than the distance between two particles, the diameter of the wave packets can also become smaller than this distance.

At the end of a star's life,[56] burnt fuel (carbon, oxygen, silicon, iron, nickel) accumulates at the core and forms degenerate matter. As the pressure and the density increase, the volume per particle becomes smaller and smaller. The momentum becomes larger and larger, following the uncertainty relation. At some point, the particles become relativistic. For relativistic degenerate matter, the energy-momentum relation changes to

Eq. 8.15

$$E = c\bar{p}$$

The $\bar{p}^5$ from Eq. 8.11 turns into $\bar{p}^4$. Eq. 8.11, Eq. 8.12, and Eq. 8.13 turn into

Eq. 8.16

$$E_{tot,rel} = \int_0^{\bar{p}F} c\bar{p}\left(\frac{\Delta x^3}{(\hbar/2)^3}\right)\mathrm{d}^3\bar{p} \propto \Delta x^3 \bar{p}_F^4$$

Eq. 8.17

$$E_{tot,rel} \propto \Delta x^3 \frac{(\hbar/2)^4}{\Delta x^4} \propto \frac{1}{\Delta x} \propto \frac{1}{V^{1/3}}$$

For the pressure, we find

Eq. 8.18

$$p_{rel} = -\frac{\mathrm{d}E_{tot,rel}}{\mathrm{d}V} \propto \frac{1}{V^{4/3}} \quad \Rightarrow pV^{4/3} = const$$

[56] These remarks apply to stars with a mass larger than $8M_\odot$. Smaller stars eventually blow off their shell in a process called "nova". A white dwarf remains. In a white dwarf, non-relativistic degenerate electrons dominate the pressure.

The adiabatic index turns into 4/3 (instead of 5/3). From then on, the body is no longer stable against gravitational collapse. There are two pathways for the collapse, discussed further in section 8.4.

The mass, at which this instability occurs, is the famous Chandrasekhar limit (about 1.4 $M_\odot$). For neutron stars the equivalent limiting mass is the Tolman-Oppenheimer-Volkoff limit (between 1.5 and 3 $M_\odot$).

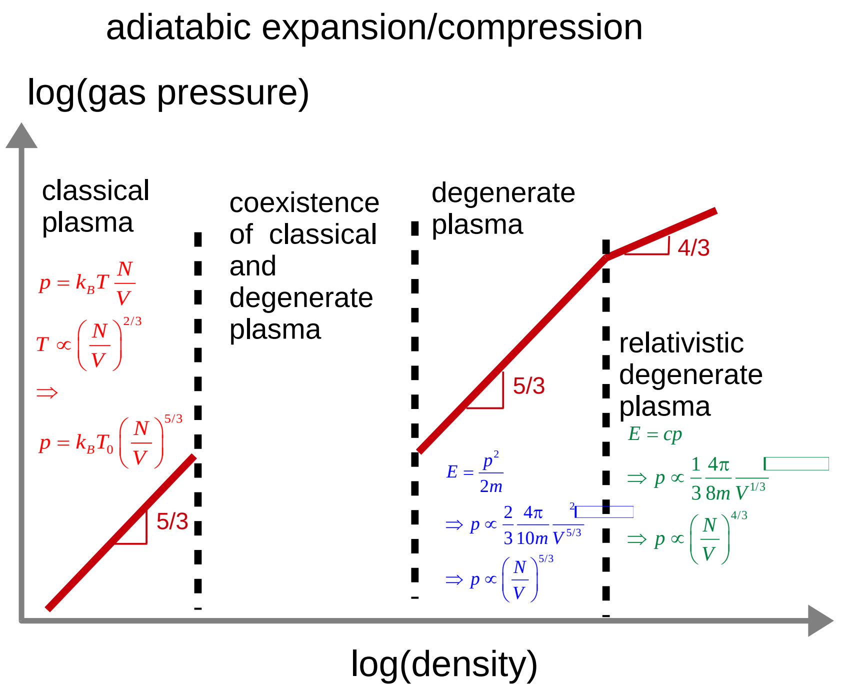


Fig. 8.5
Summary of equations from section 8.2

<u>Remember:</u>

– In the cores of stars containing burnt fuel, in the white dwarfs, and in the neutron stars, the density is governed by the Pauli exclusion principle, related to the uncertainty relation.
– For degenerate matter, pressure is not a function of temperature.
– As long as the particles are non-relativistic, the adiabatic index is 5/3, similar to ordinary atomic matter. Those balls are stable against collapse under gravity.
– At high pressure, the movement of the particles becomes relativistic. This changes the energy-momentum relation and, in consequence, the density-pressure relation. Those balls are not table against collapse under gravity. They implode.

8.3 Advanced topic: Neutron stars

Neutron stars were postulated soon after the discovery of the neutron. It was clear that these had to be compact objects with a high density. Neutron stars were first seen with radio telescopes in the form of "pulsars". The repeat rate of the pulses is about 1 second. Understandably, the pulses were initially thought to be messages from other civilizations (the "LGM hypothesis", "LGM" for "little green men"). That pulsars must be neutron stars can be inferred from a simple rule: Fast processes must originate from compact sources because the finite light travel time does not allow cooperative processes over distance beyond *ct* (*c* the speed of light and *t* the time scale of the process).[57]

Pressure, temperature, density, and magnetic fields all are extreme on neutron stars and one wonders whether the familiar laws can be extrapolated to these extreme conditions. So far, the answer is yes.

More comments:

– There are about 1 billion neutron stars in the Milky Way. As they cool down over time, they become invisible to telescopes.
– There are different types of neutrons stars. Some, but not all are pulsars. Some emit jets, some are x-ray sources. The period of rotation varies between milliseconds and 10s of seconds.

---

[57] It also follows from this argument that the fast γ-ray bursts must originate from neutron stars or black holes of the same size (possibly from collisions of such objects). The other candidates – including supermassive black holes – are too large.

- The radius is about 10 km. For the internal structure see Fig. 8.5.
- Gravity on neutron stars is 10^9 times larger than on Earth. The pressure in the center is above 10^{25} bar.
- There is a solid crust. Sometimes there are "star quakes". The star quakes can be inferred from occasional small changes in the speed of rotation,[58] because these slightly change the star's the moment of inertia (see also footnote 58).
- The moment of inertia much decreases during the collapse. Because angular momentum is conserved the speed of rotation increases correspondingly. Neutron stars rotate quickly (periods in the range of seconds).
- The millisecond-second pulsars have undergone further acceleration by an accretion disk, which formed from material falling in from a companion (Fig. 9.8).
- When an accretion disk forms around a neutron star (section 6.1), the energy in the disk is extremely high and there is a zoo of energetic events. These are the "soft x-ray binaries".

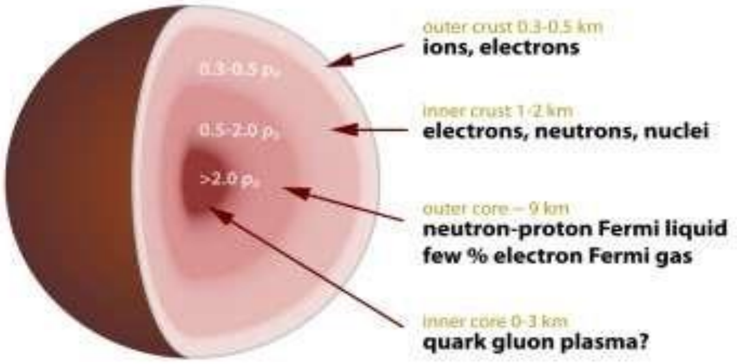


Fig. 8.6
Structure of neutron stars
en.wikipedia.org/wiki/Neutron_star

- The density in neutron stars is 2 – 5 times the density in the atomic nuclei. While the density in neutron stars is extreme, current models do not necessarily support the notion that the physics would be much different from what happens in nuclei.[59] The models say that there is no quark-gluon plasma. The models say that the temperature is not high enough to produce strange quarks. Strange quarks – and other particles with odd properties – are created when cosmic rays hit the upper atmosphere.
- Some neutron stars have magnetic fields, which have been enhanced by another factor of 1000 by a dynamo mechanism. For comments on these "magnetars" see section 9.4.

8.4 Advanced topic: Supernovae

Supernovae come in two forms. In type Ia supernovae, a compact object (mostly a white dwarf) has a red giant as a companion. This is not unusual insofar, as binary stars are common and every heavy star eventually turns into a red giant. The red giant transfers material to the compact object. This material first orbits the compact object in the form of an accretion disk (Fig. 8.6), but ultimately collapses onto it. The material is degenerate due to the high pressure. Further, this material is a fuel; it may burn. Type Ia supernovae are caused by a thermonuclear reaction. They are also referred to as "deflagration supernovae". Because of the degeneracy, a negative feedback loop which lets the fire burn steadily in normal stars (in normal ovens of nucleosynthesis) is switched off. In normal stars, increased temperature lets the

---

[58] The length of the day also varies slightly on Earth because the distribution of mass changes over time (mostly caused by plate tectonics and ocean currents). This variability on a short time scale is stronger than the slowing effect of the tides (Box 15.3)

[59] A distinction is sometimes made between "nuclear forces" and "quantum chromodynamics" (QCD). On very small scales and at very high energies, the strong interaction is mediated by gluons. It then acts between the quarks. The interaction between the nucleons (protons and neutrons) is mediated by "mesons". These are particles consisting of a quark and an antiquark. Ultimately, the nuclear force also goes back to QCD, but there is an intermediate step. Sometimes this is portrayed as an analog to the difference between the electromagnetic interaction and the van der Waals interaction. In this picture, the vdW force is mediated by short-range virtual photons.
Possibly QCD is at work in the center of neutron stars rather than the nuclear forces.

pressure increase, which lowers the density and lowers the reaction rate. The reactor is stable. In degenerate matter, pressure is not linked to temperature (section 8.2). The temperature becomes ever higher, but there is no expansion linked to it. At some point, the temperature becomes so high that it lifts the degeneracy. At this point, the fuel burns and expands. Because this instability occurs at a well-defined threshold, type Ia supernovae all have the same brightness. The apparent brightness on Earth can then be used to infer the distance. Type Ia supernovae have a characteristic spectrum, by which they can be recognized. They are among the standard candles, to which the cosmic distance ladder is anchored.

The mechanism underlying type II, type Ib, and type Ic supernovae is different. Stars with a mass larger than about 8 $M_{\odot}$ develop a heavy core of burnt material at the end of their life. The shell expands and turns the star into a red giant.[60] The core is degenerate. When the pressure increases to the extent that the particles become relativistic, the core implodes. What happens next depends on the mass. If the mass is sufficiently large (larger than the Tolman-Oppenheimer-Volkoff limit, section 8.2), a black hole is formed. With a slightly lower mass, however, the neutrons produced in the reaction $p^{+} + e^{-} \rightarrow n + \nu_e$ again become degenerate. The implosion stops. The temperature shoots up and produces an expansion, which ejects large amounts of material and leaves a neutron star behind (Section 8.3).

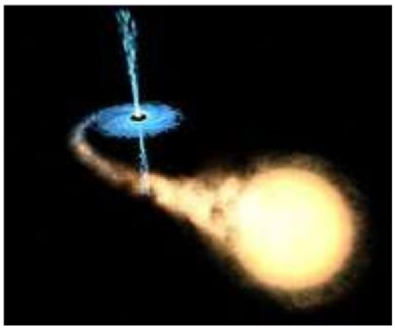


Fig. 8.7
Accretion of matter ejected by a red giant onto a neighboring compact object eventually leads to a Type Ia supernova.
simple.wikipedia.org/wiki/Accretion_disk

The neutrinos from the fusion of electrons and protons are so numerous that they carry away a significant fraction of the energy released by the supernova. They are so numerous that the supernova 1987A was detected on Earth with neutrino detectors (11 events in 13 seconds in the Kamiokande detector). Finally, they are so numerous that they exert an influence on the expanding shell despite the small collision probability. The details are complicated and models have been revised in the light of simulations with high-performance computers. In particular, the explosion does not occur in a spherically symmetric way. This can be inferred from the fact that neutron stars often move quite fast relative to the neighboring stars. Their momentum is the recoil from the unbalanced explosion.

The temperature in a supernova is high enough to produce heavy elements in a so-called “r-process” (“r” for rapid). These elements with a mass larger than the mass of iron are not produced in stellar nucleosynthesis.[61] Iron is the element with the highest binding energy.

During a supernova, an energy of 10^{44} J = 10^{51} erg is released within a few seconds. This energy is also called "foe" (“foe” for "ten to the **f**ifty-**o**ne **e**rg"). The sun generates approximately this amount of energy over its entire lifetime. For a short time, a supernova is as bright as an entire galaxy. It happens that there are other energetic events in the universe which release about 10^{51} erg. 10^{51} ergs appears to be a characteristic energy for massive explosions.

Remember:

– Type Ia supernovae are thermonuclear in nature.[62] A companion transfers fuel to a compact object. Because the material is degenerate, high temperature does not expand the material. Eventually, the

[60] The shell eventually is blown away in a process called „nova“, which leads to a planetary nebula (see Fig. 11.1)
[61] Actually, some nickel is produced in the stars, as well
[62] The classification of supernova into types is based on the absorption lines in their spectrum. .

temperature becomes high enough to lift the degeneracy, which then does lead to an expansion and ignites an explosion.

- Type II supernovae, type Ib supernovae, and type Ic supernovae follow from a collapse of the core of a heavy star, when the mass is low enough to let the ball of neutrons be stable. Otherwise the collapse ends in a black hole.
- Supernovae synthesize elements heavier than iron.

9 Space is pervaded by magnetic fields and plasma

9.1 Different types of plasma

A plasma is a gas, but differing from ordinary gases, particles are partially or fully ionized. There is no discrete transition temperature from the neutral gas to the plasma. The degree of ionization increases with increasing temperature.[63] Nevertheless, the plasma is considered a separate state of matter because charged particles interact with each other differently from neutral particles.

There is not much chemistry in a plasma and there are few reactions that would lead to complex molecules. Most plasmas are chemically aggressive.[64]

Around the ions, electric fields are strong. However the fields are screened by the counter-ions.[65] There is an analog to the Debye length, which for electrolytes is the length below which electroneutrality has no effect. This length depends on the square root of the charge density of the plasma, just like the Debye length in Debye-Hückel theory.

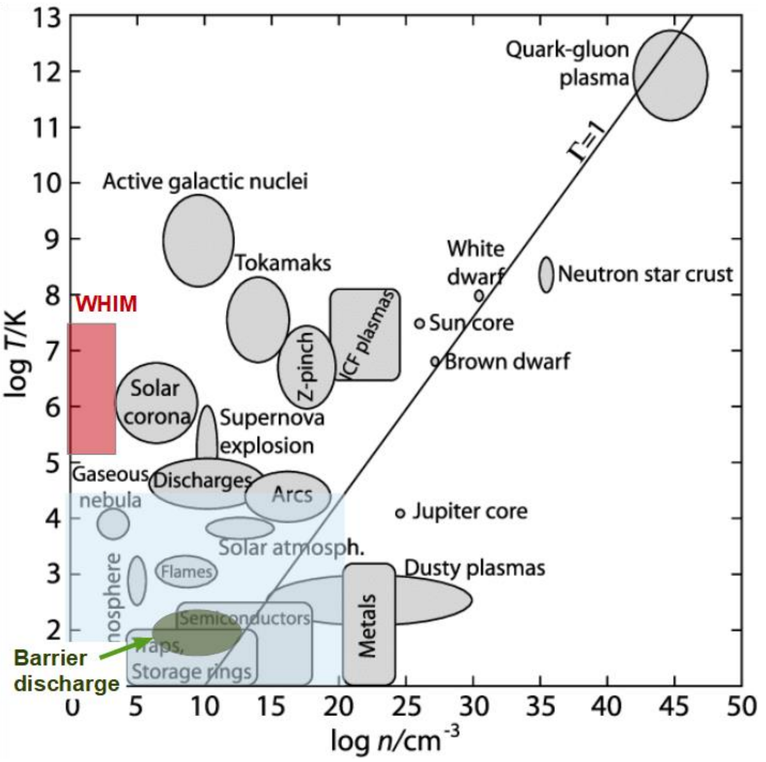


Fig. 9.1
The plasma state potentially covers a wide range of pressures, densities and temperatures. Plasma at room temperature and atmospheric pressure is unstable. It only exists as long as the source (e.g. a gas discharge in an arc or ionizing radiation) is active.
www.researchgate.net/figure/Different-types-of-plasmas-over-the-density-temperature-plane-Note-the-extremely-wide_fig1_2215721

A large part of the matter of universe is ionized.[66] Fig. 9.1 is meant to bring across that the properties of a plasma strongly depend on density and temperature – even qualitatively – and that, further, density and temperature vary over wide ranges. That includes plasma states on earth.

[63] The transition is gradual in the same way as chemical equilibria gradually shift with temperature. Cooperative behavior would be necessary for a sharp transition. The degree of ionization is predicted by the Saha equation.
[64] There is plasma polymerization, though. A polymer – often a fluorinated polymer – is created on a surface. The polymer usually is highly cross-linked.
[65] There are large-scale electric fields in the atmosphere. We know them from thunderstorms. Even when there is no thunderstorm, there is an electric field between the ground and the ionosphere. These electric fields are an exception to the rule.
[66] Neutral gas is found in the molecular clouds, in parts of the interstellar medium within a galaxy (studied using the 21-cm line, sections 10.1 and 10.2), and in protoplanetary disks (Section 6.1).

Remember:

– A large portion of the matter in the universe is ionized (that is, exists in the form of a plasma).

Box 9.1:
Freezing of field lines and magnetic pressure
Freezing of field lines:
Combine Faraday's law with Ohm's law in Maxwell's equations, to arrive at the "law of induction" [1]

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \frac{1}{\sigma_0 \mu_0} \nabla^2 \mathbf{B} = \nabla \times (\mathbf{v} \times \mathbf{B}) + D_{mag} \nabla^2 \mathbf{B}$$

σ_0 is the conductivity. Without the first term, this equation would have the structure of Fick's 2nd law.
$D_{\text{mag}} = 1/(\sigma_0\mu_0)$ is the "magnetic diffusivity". (Because the first term is nonzero, diffusion takes place relative to the movement of the plasma). With perfect conductivity, the magnetic field lines no longer diffuse. If $D_{\text{mag}} \approx 0$, the magnetic field is locked to the plasma.

Magnetic pressure:
Start from an analog of the Navier-Stokes equation for magnetic forces

$$\rho\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla\right)\mathbf{v} = \mathbf{J} \times \mathbf{B} - \nabla p$$

This equation is also called the "equation of motion". On the left is the inertial force. The first term on the right is the Lorentz force. The following general vector identity will be needed

$$\frac{1}{2}\nabla(\mathbf{B} \cdot \mathbf{B}) = (\mathbf{B} \cdot \nabla)\mathbf{B} + \mathbf{B} \times (\nabla \times \mathbf{B})$$

Following Ampere's law $\mu_0 \mathbf{J} = \nabla \times \mathbf{B}$. The above vector relation leads to $\mathbf{J} \times \mathbf{B} = \frac{(\mathbf{B} \cdot \nabla)\mathbf{B}}{\mu_0} - \nabla \frac{\mathbf{B}^2}{2\mu_0}$

$\mathbf{J}\times\mathbf{B}$ is the Lorentz force. The second term on the right has the form of a pressure gradient. It follows that

$$\rho\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla\right)\mathbf{v} = \frac{(\mathbf{B} \cdot \nabla)\mathbf{B}}{\mu_0} - \nabla\left(\frac{\mathbf{B}^2}{2\mu_0} + p\right)$$

A plasma has a tendency to avoid high magnetic fields. Conversely, the plasma sometimes concentrates the magnetic field into flux tubes. This amounts to a mechanism of structure formation.
The first term in the equation above leads to a line tension ("magnetic tension force"). This term straightens curved field lines, similar to how a tensile force stretches a thread.
[1] https://en.wikipedia.org/wiki/Magnetohydrodynamics

9.2 The intergalactic medium

The space between galaxies is almost empty, but still contains 40–50% of the baryonic matter of the universe, simply because the space is so vastly large. The medium is a plasma called WHIM, for “warm and hot intergalactic medium". Temperatures are between 10^5 and 10^7 K. Heating occurs by shock waves, generated by material collapsing onto the WHIM from the voids formed by the BAOs. Outside the galaxies, the density is between 1 and 10 particles per m^3. The average collision time is a few hundred years. While that may appear to be long, equilibrium is still reached on astronomical time scales. The particles are in equilibrium with the other particles, but not with the CMB and not with the light emitted by the stars.

The thermal x-ray radiation from the WHIM is weak because of the low density. Fig. 9.2 shows a case, where the WHIM between the members of galaxy cluster was dense enough to be imaged. The WHIM can also be detected based on absorption lines in the spectra of bright x-ray sources behind these clouds. The absorption is caused by heavy atoms that are not fully ionized. The temperature can be inferred from the degree of ionization.

Remember:

- The WHIM accounts for 40–50 % of the baryonic matter in the universe.
- In the intergalactic medium, the density is around 1 particle/m^3. The average time between collisions is several hundred years.
- The particles reach a thermodynamic equilibrium despite the long collision times.
- The temperature is between 10^5 and 10^7 K. Heating occurs by shock waves, when matter falls onto the WHIM from the voids.

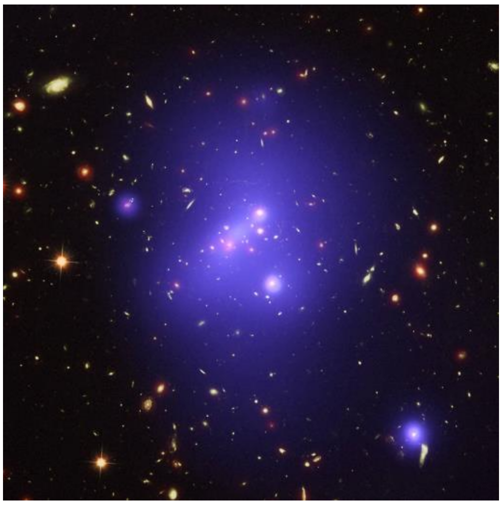


Fig. 9.2
Astronomers have made the most detailed study yet of an extremely massive young galaxy cluster using three space telescopes. This multi-wavelength image shows this galaxy cluster, called IDCS J1426.5+3508 (IDCS 1426 for short), in x-rays from the NASA Chandra x-ray Observatory in blue, visible light from the NASA/ESA Hubble Space Telescope in green, and infrared light from the NASA Spitzer Space Telescope in red. There is a region of bright x-ray emission (seen as blue-white) near the middle of the cluster, but not exactly at the center.
www.nasa.gov/mission_pages/chandra/images/galaxy-cluster-idcs-1426.html
The eROSITA satellite has created a map of galaxy clusters that are visible in x-ray range, based on the intergalactic plasma.

9.3 Consequences of magnetohydrodynamics

Magnetic effects are weaker than electric effects, in principle (Box 15.4), but there still are magnetic fields everywhere in space. There are no large scale electric fields because these are screened by counter-ions. Magnetic fields are not screened in the same way because there are no magnetic monopoles.

Magnetic fields in a plasma obey magnetohydrodynamics (MHD). We report a few results without going into the mathematics.

- When an ion travels at right angles to the magnetic field lines, it is forced onto a circular path by the Lorentz force and cannot be transported to beyond this circular path at right angles to the magnetic field. Along the field lines, a spiraling motion does achieve transport. This coupling of the particles to the magnetic field also works in the other direction: If a plasma moves perpendicular to the magnetic field, it can pull the magnetic field along. The magnetic field and plasma are locked together.
- Bodies with differential rotation (fastest in the center) can generate magnetic fields (assuming conductivity). The plasma than acts as a "dynamo" (earth dynamo, solar dynamo, dynamos in neutron stars). Leaving details aside, the field amplification can be made plausible as attempted in Fig. 9.3. There is a magnetic field in a conductive medium and there is a toroidal flow in this medium.[67] Because the magnetic follows the plasma, straight field lines turn into helical lines. Because the field lines become elongated, the energy contained in the magnetic field increases. If there is a mechanism, by which the field lines can diffuse back to the linear structure, the interplay between elongation and back-diffusion results in positive feedback. The magnetic field is enhanced.

[67] There are further requirements for flow fields causing a dynamo. A number of "no-dynamo theorems" formulate these.

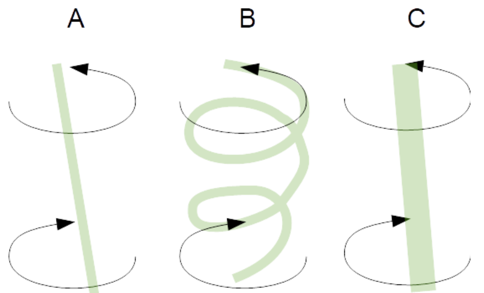


Fig. 9.3
Simplified sketch of how magnetic fields are amplified in a plasma with toroidal flow. A small initial magnetic field is dragged along by the flow and elongated on the way (B). The total energy contained in the magnetic field increases. If the field diffuses back and does not lose all of its energy, the field becomes stronger (C). The effect is caused by an interplay of convective transport of magnetic fields (with the flow) and diffusive transport. The details are complicated.

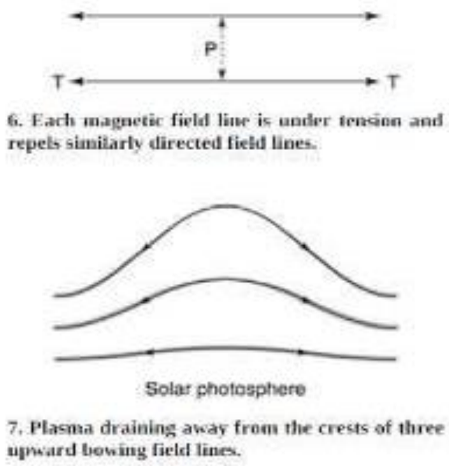


Fig. 9.4
Because the density of matter inside a magnetic flux tube is decreased, a buoyant force lets the tube move upwards to the surface of the sun. The material, which it carries with it, rains back along the field lines, making the tube even lighter and amplifying the buoyancy.
Source: Binney (footnote 1)

– A strong magnetic field generates a pressure (Box 9.1). The plasma is expelled from these regions. This is how magnetic confinement works in a fusion reactor. Conversely, a magnetic field is expelled from the plasma. It then is concentrates into flux tubes, which amounts to a structure formation mechanism. The energy density in the tubes has a contribution from the magnetic pressure. Because the total energy density is about the same as outside, the plasma in the tubes has a lower kinetic energy than the plasma outside. It is a little colder.

– Such tubes exist on the sun. There is an instability that causes these tubes to protrude from the surface of the sun (Fig. 9.4) at the sunspots. The sunspots appear as dark because of the lowered temperature. The magnetic field in the sunspots is noticed, spectroscopically, based on how the Zeeman effect lets certain lines shift in a magnetic field.

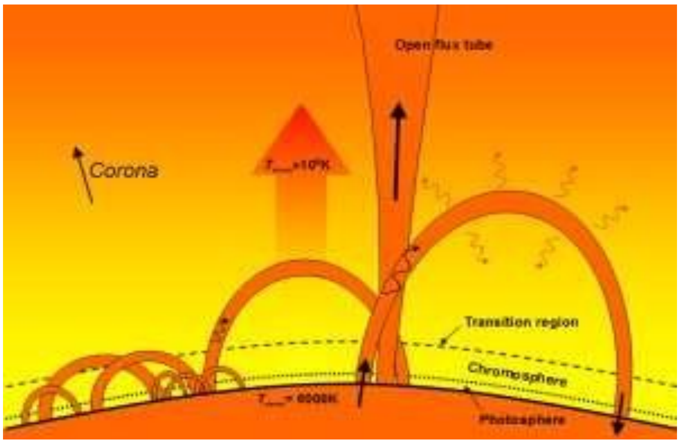


Fig. 9.5
Magnetic flux tubes on the sun
www.wikiwand.com/en/Flux_tube

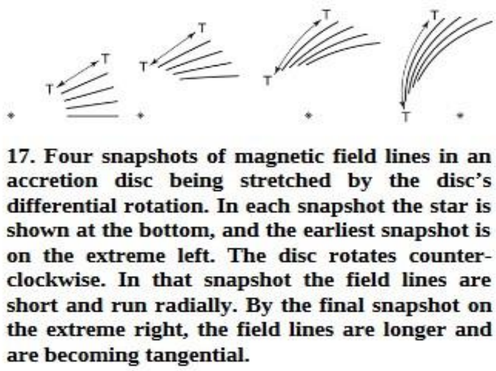


Fig. 9.6
Source: Binney (footnote 1)

- When magnetic field tubes merge ("flux tube reconnection"), large amounts of energy are released. This energy contributes to the heating of the solar corona (Fig. 9.5). Temperatures in the corona reach millions of degrees.
- Fast, charged particles in strong magnetic fields emit synchrotron radiation (section 9.5), which can be recognized examining the spectrum. Synchrotron radiation is often emitted by jets.
- When a plasma becomes denser, the local magnetic field increases. That happens in the accretion disks (Fig. 9.6) and when neutron stars are formed by an implosion.

Remember:

- Long-range magnetic fields permeate the interstellar space and even the intergalactic space.
- In a plasma with large electric conductivity, matter and magnetic fields are locked together.
- An amplification of magnetic fields can occur in a plasma with differential rotation (in a dynamo).
- Magnetic fields may be concentrated into flux tubes.
- The density and temperature of the plasma inside the flux tube is decreased. The flux tubes therefore experience a buoyant force. When flux tubes intersect the surface of the sun, they form sunspots.
- Flux tube reconnection releases large amounts of energy.

9.4 Strong magnetic fields around neutron stars

The increase in the magnetic field during a compaction of a plasma is dramatic in collapse supernovae (section 8.4). Neutron stars are the result of such a collapse (section 8.3). They have magnetic fields of the order of 10^{11} Tesla. Due to the spiraling motion of the charges, neutron stars only eject material at the magnetic poles. They then emit "jets". (Jets also come in other variants.) When neutron stars rotate (they do so rather quickly) and when the magnetic north pole is not on the axis of rotation, the neutron star acts like a lighthouse. The star lights up when the jet crosses our line of sight. These stars are the pulsars.

In special cases, a dynamo effect occurs and the magnetic field increases by one more factor of 1000. Values then are up to 10^{14} Tesla. The superconducting protons play a role in this dynamo.[68] A “magnetar” is formed. Magnetars appear to the observer as "soft gamma repeaters" combined with "radio bursts". The bursts are caused by a reconnection of flux tubes. To date, several dozen magnetars have been identified. The magnetic fields of these extraordinary objects decay over a period of approximately 10 000 years.

The energy density in the magnetic field of magnetars is 10^4 times higher than the energy density of lead (calculated using $E = mc^2$). These magnetic fields are so strong that the vacuum becomes birefringent. Electromagnetic radiation can penetrate a plasma there because the magnetic field prevents the ions from following the alternating electromagnetic field.

Remember:

- When a plasma collapses, the magnetic field collapses with it and then assumes extremely high values (10^{11} Tesla).
- The magnetic field of the neutron stars only allows matter (analogous to the solar wind) to escape at the magnetic poles. If the magnetic pole is not on the axis of rotation, a pulsar is formed.
- The emission at the magnetic poles leads to jets.

---

[68] In superconductivity, there is a coupling between two fermions mediated by the lattice vibrations in such a way that a quasiparticle with zero spin (a boson) is created. This quasiparticle is subject to Bose-Einstein statistics with many consequences. Such couplings exist in different variants. In particular, high-temperature superconductivity is still poorly understood. Coupling protons so strongly that they become superconducting requires high densities (as found in neutron stars).

– In the case of magnetars, the magnetic fields are amplified by a factor of around 1000 as a result of a dynamo.

9.5 Advanced Topic: Synchrotron radiation

The following text is adapted from http://www.jeffstanger.net/Astronomy/emissionprocesses.html

The synchrotron emission mechanism is another example of an important astrophysical process (Figure 3 [here: Fig. 9.7]). Synchrotron radiation is observed in regions where relativistic electrons (those traveling close to the speed of light) spiral around magnetic field lines. This process results in strongly polarized radiation concentrated in the direction of the electrons motion (called "beaming"). Similar to bremsstrahlung, synchrotron has a characteristic shape of its spectra which is a power law spectrum. The shape of the spectrum produced is dependent on the energy distribution of the emitting electrons and is easily distinguishable from thermal black-body radiation.

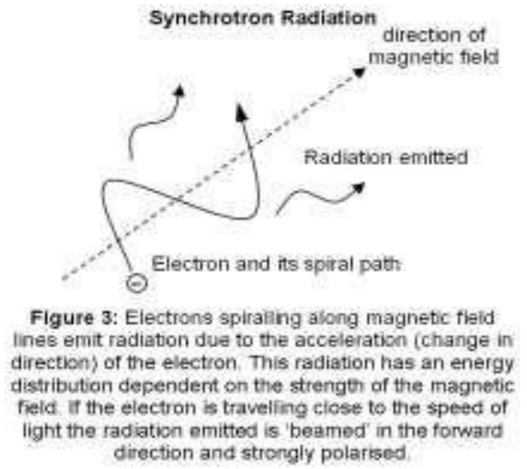


Fig. 9.7
Synchrotron radiation is emitted by electrons and ions on circular paths in a magnetic field.

A knowledge of synchrotron radiation is essential in the study of a large group of astronomical objects called active galactic nuclei (AGNs). A typical spectral profile for an AGN is shown in Figure 4 [Fig. 9.8 left].

AGNs are thought to contain massive black holes and the associated structure in the heart of the galaxy NGC 4261 can be seen in Figure 5. According to the unified theory for AGN the hot and dusty ion torus, shown to the right of Figure 5 [Fig. 9.8 right], emits unpolarized thermal radiation. This emission accounts for the 'UV bump' shown in Figure 4. This emission leads astronomers to believe that the thick ionized disk associated with this torus generates varying magnetic fields of up to 10^4 Gauss (very large) across its surface. This induces a large electric field and accelerates particles to relativistic speeds away from the disk (contributing to bipolar jets). These particles spiral along the magnetic field lines and produce synchrotron radiation which is up to 60% linearly polarized.

This synchrotron radiation is thought to be a contributing source to the power law component of the spectrum in Figure 4 [Fig. 9.8 left]. The spectrum shown in Figure 4 does not exhibit synchrotron emission at low frequencies due to synchrotron self-absorption. This is where the plasma that the synchrotron emitting electrons are part of becomes opaque to their synchrotron emission. This results in the turnover frequency that can be seen in Figure 4 (Fig. 9.8) at infrared (IR) wavelengths.

<u>Remember:</u>

– Fast-moving particles in magnetic fields emit synchrotron radiation.
– Synchrotron radiation has a characteristic spectrum and is polarized.

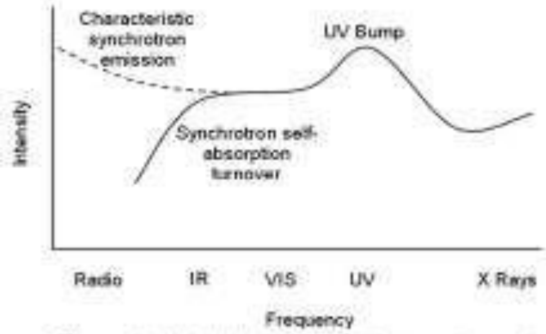


Figure 4: A sketch of the continuum observed for many AGN with intensity vs. frequency.

Figure 5: Active galaxy NGC 4261 at radio and optical wavelengths.

Fig. 9.8
Addressed in the text as "Figure 4" and "Figure 5".
http://www.jeffstanger.net/Astronomy/emissionprocesses.html

10 Neutral gas, molecular clouds

10.1 Gas in the galaxies

The role of gas in galaxies is quite complex. First, the density is so high that the shock waves, that heat the warm and hot intergalactic medium (WHIM), do not penetrate to the inside of the galaxy. The gas mostly is neutral; it is not a plasma. For hydrogen, this can be inferred from the 21-cm line emitted by atomic hydrogen (Section 10.2).

Molecular hydrogen (H_2) does not form in most of the interstellar space. H_2 and other molecules are only found in the molecular clouds. In the molecular clouds, dust catalyzes the reactions forming molecules. The dust also governs the temperature (~ 50 K). Molecular clouds, dust, and star formation come together. Around the young stars in these star formation regions, hydrogen is ionized by the UV-radiation from the young and hot stars (Box 10.1). The HII regions ("II" for the first ionization state) have a characteristic red color. Dust clouds are rarely found in very old galaxies, because the elements heavier than helium have not yet been synthesized in sufficient amounts. Star formation in the first generation of galaxies proceeded differently from how it proceeds today.

The fraction of the mass in gas to the mass in stars can vary between 10^{-2} and 10. Gas is required for star formation. The gas is enriched in the spiral arms of the spiral galaxies. There are shock waves, which in part are caused by central bar of the galaxy interacting gravitationally with the spiral arms. The

Box 10.1
Stars

The larger part of a star's life is rather unspectacular. These stars are referred to as "main sequence stars". The main sequence consists of points in the color-luminosity diagram (Fig. 10.2). In this diagram, a band runs from top left to bottom right. The main sequence is the consequence of a relation between color and luminosity. The variable underlying both parameters is mass. The larger a star is, the hotter is the center, and the faster is nuclear fusion. Heavy stars are bright and blue. They are blue because they are hot (Wien's law). Heavy stars live shorter than low-mass stars because the fuel is used up more quickly. Very heavy stars only live for a few million years. Low-mass stars can live longer than the universe is old now (mathematically: hundreds of billions of years).

The bright and hot stars are so short-lived that they are only found in the star formation regions. These hot stars ionize the medium around them by UV radiation.

On the surface of stars there is a corona with extremely high temperatures (sun: several million K). The mechanisms of heating are sound and the fusion of magnetic flux tubes (section 9.3). The temperatures are so high because the density becomes lower and lower towards the top. The sound waves travel upwards, but no longer find any material into which they could dissipate their energy. This can be thought of as an analogy to the crack of a whip.

Because the corona is very hot, the ions in the solar wind have a high kinetic energy. Emission lines were found in the spectra from the solar corona that could not be assigned. An element with the name "coronium" was postulated. It turned out that "coronium" consisted of highly ionized atoms, which had not yet been . Such highly ionized atoms do not exist on Earth.

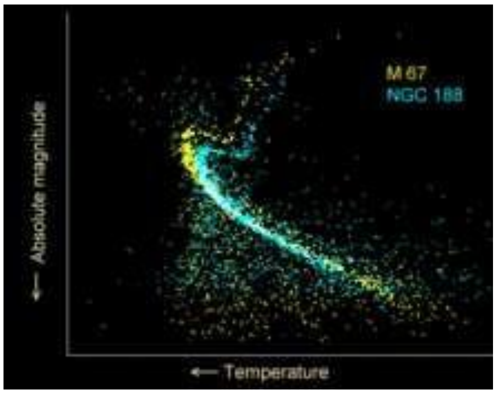


Fig. 10.1
A temperature-luminosity diagram for two open star clusters. NGC 188 (blue) is the older of the two and shows an earlier bend from the main sequence than M67 (yellow). Open star clusters (such as the Pleiades) are clusters of young stars. They are to be distinguished from globular clusters (made of old stars).
studylibde.com/doc/2494993/stars

shock waves trigger star formation. Explosions (nova and supernova) at the end of the lives of short-lived stars trigger further star formation. There are cascades of star formation. Because of star formation, the spiral arms shine brightly.

Remember:

- With the exception of the star forming regions, the gas inside the galaxies is neutral.
- Gas is required for star formation.

10.2 Advanced Topic: The 21-cm line

Neutral atomic hydrogen inside the galaxies is easily seen with radio telescopes making use of the famous 21-cm line. The Doppler shift of the 21-cm line can be used to measure the rotation curves of galaxies (the angular velocity as a function of the distance from the center, Fig. 13.1).

The 21-cm line is worth a digression insofar as the lifetime of the excited state is 11 million years, which can be understood. A reminder: Typical fluorescence lifetimes are a few nanoseconds. A lifetime of one nanosecond is considered long insofar as other electronic processes take place on the femtosecond scale, i.e. 10^6 times faster. Nuclear movements take place on the picosecond scale. They are much faster than fluorescence. The excited state of the 21-cm line is 10^{26} times more long-lived than the excited states of fluorescence.

Radio waves with a wavelength of 21 cm (corresponding to 1.4 GHz) are emitted when the nuclear spin of atomic hydrogen changes from parallel to antiparallel orientation relative to the spin of the electron. This influence of the nuclear spin on the energy levels of hydrogen is part of the "hyperfine structure". The hyperfine structure can be studied with high-resolution laser spectroscopy. Back to the 21-cm line: A first question is why there is no collisional deexcitation over a period of 11 million years. Collisional deexcitation always competes with radiative deexcitation. Collisions are rare in the interstellar medium, but the collision times of a few years are considerably shorter than the lifetime of the spin-parallel H atom. However, collisions do not usually flip the nuclear spin. This is known from spin-polarized ^{3}He. This material can be stored for many hours. It can even be transported by car. All that is needed is cryocontainer with non-magnetic walls.

The long lifetime results from two separate factors, namely the low frequency and a selection rule. The frequency first: We quote Wikipedia for the rate of spontaneous fluorescence:[69]

The rate of spontaneous emission (i.e., the radiative rate) can be described by Fermi's golden rule. The rate of emission depends on two factors: an 'atomic part', which describes the internal structure of the light source and a 'field part', which describes the density of electromagnetic modes of the environment. The atomic part describes the strength of a transition between two states in terms of transition moments. In a homogeneous medium, such as free space, the rate of spontaneous emission in the dipole approximation is given by:

Eq. 10.1

$$\Gamma_{rad}(\omega) = \frac{\omega^3 n |\mu_{12}|^2}{3\pi\varepsilon_0 \hbar c^3} = \frac{4\alpha\omega^3 n |\mu_{12}|^2}{3c^2}, \qquad \frac{1}{\pi\varepsilon_0 \hbar c} = 4\alpha$$

[69] *en.wikipedia*.org/wiki/Spontaneous_emission

where ω is the emission frequency, n is the index of refraction, μ_{12} is the transition dipole moment, ε_0 is the vacuum permittivity, $\hbar$ is the reduced Planck constant, c is the vacuum speed of light, and α is the fine structure constant. [...] The above equation clearly shows that the rate of spontaneous emission in free space increases proportionally to ω^3 [end of quote from Wikipedia].

Important to Eq. 10.1 are the frequency (to the third power) and the square of the (magnetic) transition dipole moment. For this calculation, let the fluorescence wavelength, with which we compare, be 200 nm, corresponding to $1.4 \cdot 10^{15}$ Hz. Based on the pre-factor ω^3, we conclude that the lifetime of the spin-parallel atomic hydrogen is at least $10^{3\times 6} = 10^{18}$ times longer than the lifetime of the excited fluorescent dye. This would result in a lifetime of 10^{18} nanoseconds (i.e. 10^9 sec). However, 11 million years correspond to 2.5 10^{14} seconds: a factor of 10^5 is still missing. This factor results from the fact that the spin flip is a magnetic dipole transition (not an electric dipole transition). Because magnetic effects are smaller than electric effects (section 16.1), magnetic dipole transitions are slower then electric dipole transitions.

Remember

- The Doppler shift of the 21-cm line can be used to map out the rotation curves of galaxies.
- The 21-cm line results from a flip of the nuclear spin in atomic hydrogen.
- The lifetime of the excited state is around 11 million years. The long lifetime results from the ω^3-term in the rate of spontaneous emission and, also, from the transition being a magnetic dipole transition rather than an electric dipole transition.

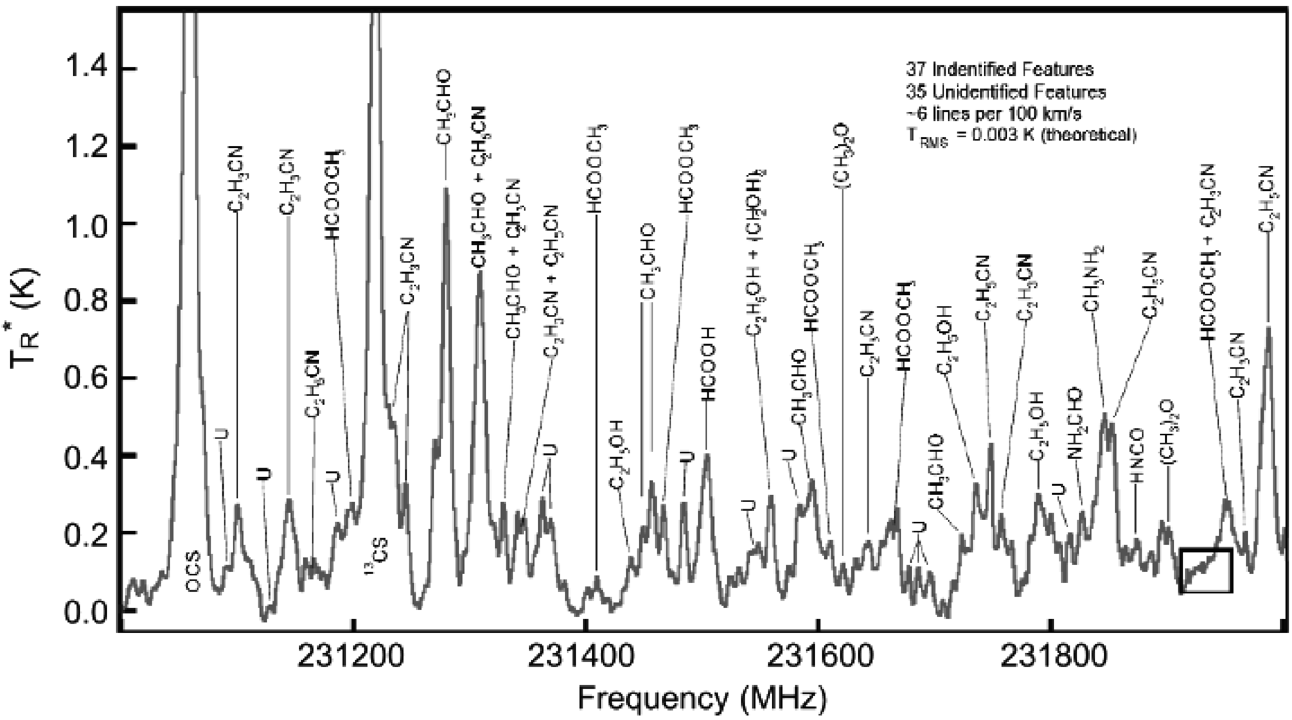


Fig. 10.2
A microwave spectrum from a molecular cloud
This λ = 1.3 mm spectrum of the molecular cloud SgrB2(N) near the Galactic center is completely dominated by molecular lines from known and unknown (U) species (Ziurys et al. 2006, NRAO Newsletter, 109, 11). More than 140 different molecules containing up to 13 atoms have been identified in space.
None of the molecules contain the element silicon. Silicon just about always forms dust grains (section 12.1).
https://www.cv.nrao.edu/~sransom/web/Ch7.html

10.3 Molecules in space

Molecules in space mostly are studied using microwave spectroscopy (rotational spectroscopy, Fig. 10.2, section 11.2).[70] About 200 molecules have been identified so far.[71,72]

Further comments as a list:

- Chemistry in space is less rich than on Earth because UV radiation dissociates molecules.
- Amino acids are found in space but this does not have much to do with life. Amino acids were also found in the experiments by Urey and Miller (which also – as it later turned out – had little to do with prebiotic chemistry[73]).
- These 200 molecules do not include the polyaromatic hydrocarbons (PAHs). These exist in a large variety. PAHs occasionally condense into soot.
- Molecules are found in the molecular clouds. They coexist with dust. The molecular clouds often are star formation regions. Formation of molecules is catalyzed on the surfaces of dust grains.
- Dust brings about a uniform temperature of ~50 K.
- More than half of the known molecules contain carbon. The metastability of the four-bond carbon plays about the same role as on Earth ("carbon chauvinism" Box 15.1).
- Although CO is much less abundant than H_2, it can still serve as a thermometer for molecular clouds (section 11.3).

<u>Remember:</u>

- About 200 molecules have currently been identified in space on the basis of microwave spectra (rotation spectra). Many of them contain carbon.
- Molecules are found in the molecular clouds. They coexist with dust.
- The formation of molecules is heterogeneously catalyzed on dust grains.
- UV-induced photodissociation prevents the formation of many more molecules.
- Molecules are also found in the meteorites.

---

[70] The James Webb telescope recently detected CH_3^- with an IR spectrometer. In this study, H_2 was also observed, evidenced by the quadrupole transition at $\lambda = 2.12$ µm.

[71] If you want to make yourself interesting at a party, let everyone know that there is more alcohol in the Orion nebula than there is whiskey in all of Scotland. Actually, there is more alcohol in the Orion nebula than there is water on Earth.

[72] Polycyclic aromatic hydrocarbons (PAHs) are separate. There is a large diversity among these latter molecules. The different PAHs are difficult to distinguish from each other and they tend to form clusters.

[73] According to current opinion, life originated in hydrothermal vents in the deep sea. Flashes of lightning – as used in the experiments by Urey and Miller – played no role.

11 Peculiarities of spectroscopy in astrophysics

11.1 Forbidden lines are never completely forbidden

Sometimes, lines are seen in the sky that are not known form the laboratory because particles collide so rarely in space that deexcitation by collisions is too slow to compete with radiative deexcitation. Radiative deexcitation also is slow because these lines are "forbidden". "Forbidden" usually means "electric-dipole-forbidden". The electric dipole antenna is the best antenna in terms of efficiency. However, there also is a "magnetic dipole antenna" (a wire forming a coil) and an electric quadrupole antenna. Motions of charges in molecules may turn the molecules into a magnetic dipole antenna or an electric quadrupole antenna.

Fig. 11.1
The green emission from the edge of the Cat's Eye Nebula contains lines unknown from earth
de.wikipedia.org/wiki/Nebulium

The best-known example is "nebulium". In 1864, William Huggins discovered bright lines with wavelengths of 3726, 3729, 4959 and 5007 Å in the Cat's Eye Nebula. These produce a characteristic green color (Fig. 11.1). Huggins suggested a new element, the "nebulium", as the explanation. The periodic table was not yet complete at the time and "helium" (proposed in 1860) indeed was confirmed to be an element later. Nebulium was not confirmed. In 1927, Ira S. Bowen identified the supposed nebulium lines as forbidden lines of ionized oxygen. The two most prominent lines are magnetic dipole transitions.

<u>Remember:</u>

– Because of the low collision rates, forbidden lines, which are no observed on earth because of collisional deexcitation, are seen in the sky.

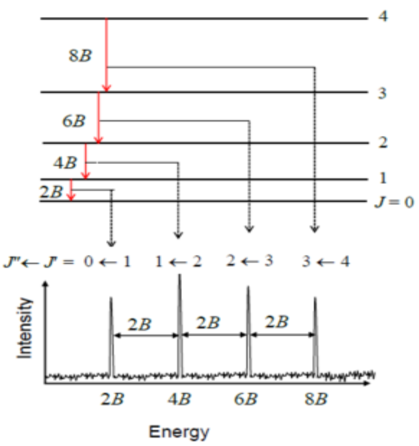


Fig. 11.2
Energy levels of the linear rigid rotator and the corresponding emission spectrum
en.wikipedia.org/wiki/Rotational_spectroscopy/

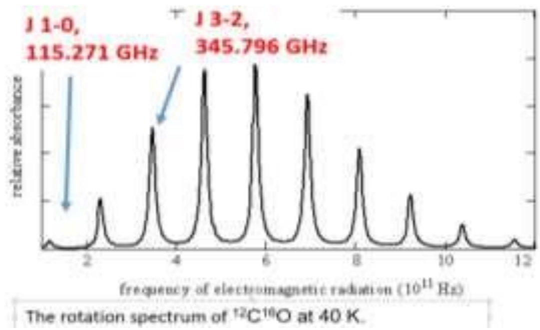


Fig. 11.3
A calculated rotation spectrum of CO
chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Book%3A_Quantum_Sttes_of_Atoms_and_Molecules_(Zielinksi_et_al)/07%

11.2 Recap: Rotational spectroscopy

We need some background on rotational spectroscopy to understand the CO rotation temperature (section 11.3). The energy levels of the rigid linear rotator are

Eq. 11.1

$$E_J = hcBJ\left(J+1\right)$$

The selection rule is $\Delta J = \pm 1$.[74] This leads to a spectrum of evenly spaced lines, where the distance between two lines (in units of wave numbers) is equal to $2B$ with the rotational constant. Fig. 11.2 shows a level diagram and a rotation spectrum. The rotation constant, B, is linked to the moment of inertia, I, by the relation $B = h/(8\pi^2 cI)$. I is equal to μr^2 with $\mu = m_1 m_2/(m_1+m_2)$ the reduced mass and r the distance between the atoms. Cleary, the rotation spectrum gives access to the distance between the two atoms.

The envelope above these evenly spaced lines allows to read the temperature because the strength of an emission line depends on how many molecules are in the respective excited state (in the excited states, plural). The occupation number of these states follows the Boltzmann distribution. We omit a few complications here, but one complication must not be omitted, which is degeneracy. Energy levels are degenerate, if there are different states having that energy. The degeneracy of the rigid linear rotator is $2J+1$.[75] Accounting for degeneracy, the strength of a rotational emission line is proportional to $(2J+1)\exp(-hcBJ(J+1)/(k_B T))$.[76] The quantum number J labels the excited state.

A few side remarks:

- For larger molecules and for non-axisymmetric molecules like water, the rotation spectra are much more complicated than for the rigid linear rotator.
- Rotational spectroscopy has for a long time been viewed as being too difficult to play a role in chemical analytics. That may change; techniques have advanced.

<u>Remember:</u>

- Rotation spectra reveal the molecule's moment of inertia and the temperature of the medium.

11.3 CO rotation temperature

As reported in section 3.2, a temperature can easily be derived form a black-body spectrum. A temperature can also be derived from line spectra. An example is the CO rotation temperature. CO is much less abundant than H_2, but H_2 does not have a permanent dipole and therefore does not emit photons when rotating. (It does, actually, because of its electric quadrupole moment, but quadrupole radiation is much weaker than dipole radiation.) There is enough CO in the molecular clouds to let the temperature be easily determined from CO. CO and H_2 collide frequently, thereby establishing an equilibrium. The CO rotation temperature is the temperature of the entire gas.

In most molecular clouds, the temperature is a few tens of Kelvin. The Boomerang nebula is an exception (Fig. 11.4). In the Boomerang nebula, the temperature in some regions is about 1 K, which is below the temperature of the microwave background. That is surprising. This region has experienced adiabatic cooling (Fig. 11.5).

---

[74] A „selection rule" names all lines, which are not forbidden. For the linear rotator, a minority of transitions is not forbidden (the ones with $\Delta J = \pm 1$.)

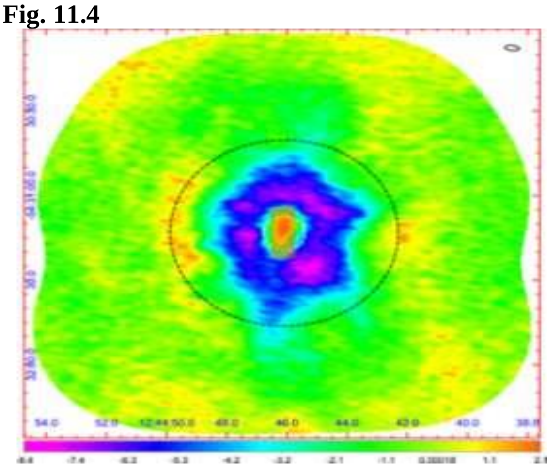
[75] Why is the degeneracy of the p-orbitals in the hydrogen atom equal to 3?

[76] There are more factors on influence on the line strength than the Boltzmann factor and the degeneracy (ignored here).

Remember:

- Because it carries a permanent dipole, CO emits a rotation spectrum (H_2 does not).
- The CO rotation temperature in molecular clouds typically is a few tens of Kelvin.
- The CO rotation temperature in the Boomerang nebula is lower the temperature of the CMB because the boomerang nebula has seen adiabatic expansion.

Fig. 11.4



Map of the Boomerang nebula taken on the $J = 1 \rightarrow 0$ line. An enlarged image of the dual lobe structure in the center is shown on the right.
core.ac.uk/download/pdf/84870536.pdf

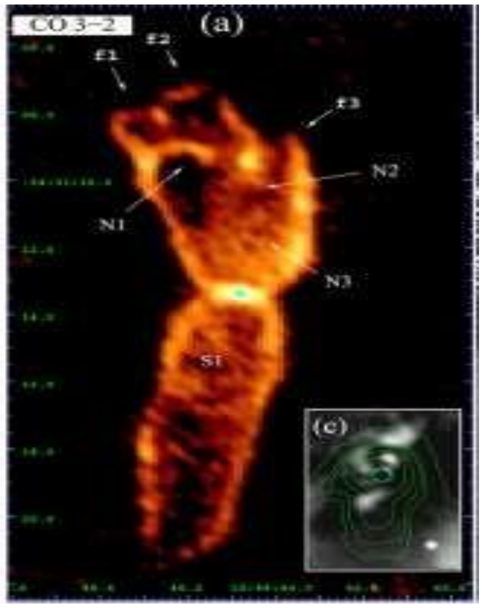


Image of the center of the Boomerang nebula taken by the ALMA observatory on the J 3→2 line
core.ac.uk/download/pdf/84870536.pdf
The rotation temperature in the outer regions of the boomerang nebula is only about 1 K (i.e. lower than the temperature of the CMB).

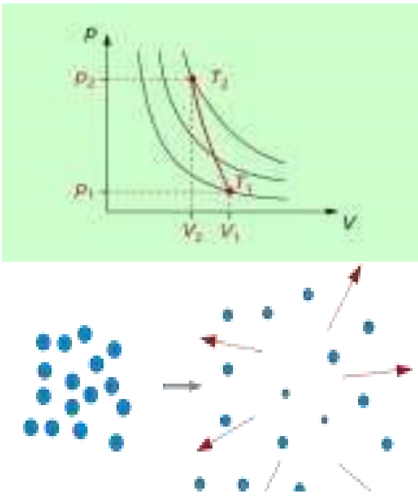


Fig. 11.5
Adiabatic expansion results in cooling of a medium. This phenomenon is observed in the atmosphere; as air rises due to thermal convection, it expands and causes temperatures to decrease at higher altitudes. Refrigerators also utilize adiabatic expansion, typically involving real gases, which leads to the Joule-Thomson effect.
In the context of the Boomerang Nebula, adiabatic expansion can be explained without relying on thermodynamics. High temperatures involve random motion. During expansion, some of this random motion is converted into directed motion, resulting in a reduction of thermal energy.
www.lernhelfer.de/schuelerlexikon/physik-abitur/article/adiabatic-state-changes

11.4 Cosmic masers

Most remarkable about cosmic masers is their existence as such. "Maser" stands for "microwave amplification by stimulated emission of radiation." Stimulated emission requires a "population inversion", that is, requires that the upper level of some transition is more populated than the lower level. In equilibrium, the populations are Boltzmann-distributed and the lower states are more populated than the higher states. Absorption is then more likely than stimulated emission. Population inversion is achieved by “pumping”. In the case of the masers, it is an optical pumping process. The light from the nearby stars excites molecules. There are transitions, where the upper level is moderately long-lived, so that molecules accumulate in this state. The lower level is short-lived, so that is depleted quickly. If that is so, there are more molecules in the upper level and stimulated emission is more likely than absorption.

Are there cosmic **l**asers in the universe?[77] No, these do not exist. Stimulated emission is always in competition with spontaneous emission. If an initial level for a possible stimulated emission has been populated by some kind of pumping process, these molecule stay in the excited state for a limited time. If a stimulating photon passes by during this time: good. Otherwise: just as well. The molecules then spontaneously fall back to the lower state. For reasons that are not quickly explained, the rate of spontaneous emission has a pre-factor of ω^3 with ω the frequency (section 10.2). Inserting values of 10^{15} Hz and 10^9 Hz for the frequencies ν (optical frequencies and microwave frequencies), one sees that microwave-excited states live about 10^{18} times longer than optically excited states. Masing is less demanding than lasing.[78] On earth, both masers and lasers contain mirrors for feedback. A certain photon is repeatedly sent through the active medium. There are no such mirrors in cosmic masers. The size of the active region is so enormous that amplification and coherence occur without feedback.

How is a maser recognized? Which molecules are active? A maser is recognized by a single strong line. The first of these lines (at 1.665 GHz) was discovered in 1965. The surprise was expressed by postulating a new form of interstellar matter with the name “mysterium”. However, it was the OH molecule. The transition is somewhat exotic. One speaks of Λ-doubling. The nuclear spin is involved, just as with the 21-cm line. Since then, masers have also been found at many other wavelengths. Even in the atmosphere of Jupiter there is stimulated coherent emission.

<u>Remember:</u>

- Cosmic masers emit coherent radiation.
- Masing requires a population inversion, that is, a non-equilibrium state of the molecules in the cloud.
- Masing is easier than lasing because the rate of the spontaneous emission is proportional to ω^3.

[77] "l" for "light", i.e. optical wavelengths

[78] The laser was invented in parallel to the maser, put it took longer to get it to work. Pumping is more difficult for the laser. Pumping is particularly difficult for UV lasers (required in semiconductor lithography) and for x-ray lasers. The best x-ray lasers at present are free electron lasers (FELs), e.g. in Hamburg. Extreme-UV lithography (EUV lithography) at λ = 13 nm uses of non-coherent radiation.

12 Condensed matter

The mass fraction of condensed matter is much lower in space than on earth. Among the reasons is the large abundance of hydrogen and helium. The attractive interactions are weak for hydrogen and helium because they are not polar and because they can hardly be polarized due to the small number of electrons.

A choice was made to not caver planets in this course. Liquids only occur on the planets because liquids usually coexist with a vapor phase.[79] Only planets are heavy enough to bind an atmosphere. In free space, droplets evaporate. In a logarithmic pT-diagram, the liquids occupy a relatively small range (Fig. 12.1).

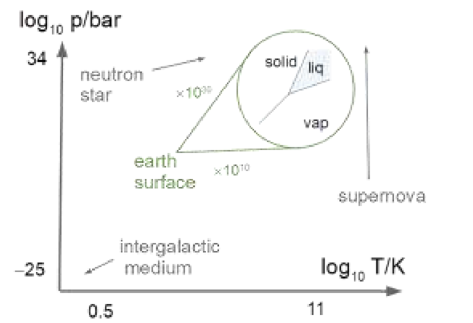


Fig. 12.1
Liquids are not common in space because they are only stable in a narrow range of pressures and temperatures.

The molecules of importance that form liquids are water, ammonia, and some hydrocarbons. Liquid hydrocarbons (including methane) exist on Saturn's moon Titan. Liquid ammonia has not been found in the solar system.[80]

12.1 Cosmic dust

The solid material in the universe mostly is dust. In parts, the dust may be thought of a soot, similar to what comes out exhaust pipes when the combustion of the fuel was incomplete. That statement includes the polycyclic aromatic hydrocarbons (PAHs). If oxidized or aliphatic, the molecules containing carbon mostly stay in the gas phase (Fig. 10.2).

When volatile molecules (H_2O, ammonia, methane) condense, those solids are referred to as “ice”. The outer part of the solar system (and the outer parts of the protoplanetary nebulae) contain ice. There is an “ice line” at some distance from the star. When ice grains cross this line, the evaporate and take part in the dynamics leading the streaming instability (section 6.1). Comets contain much ice. They are "dirty snowballs".

Silicon rarely forms small molecules (which might stay in the gas phase of form liquids in condensed form). In particular, SiH_4 is unstable because the H atoms do not shield the Si atom as well against attack by other atoms or molecules as they do in methane The famous metastability of carbon compounds[81] does not occur with the silicon compounds. The binding energy of silicate crystals (in various morphologies) is so high that silicates immediately form dust. In more technical terms, the Kelvin radius for silicates is only slightly larger than the SiO_2 molecule.[82] There is virtually no nucleation barrier.

The silicates in the cosmic dust differ from the silicates in the crust of the earth in two regards. First, they contain iron an nickel. On earth, large amounts of iron and nickel have sedimented to the earth’s core. Also, “pyroxenes” occur frequently. Pyroxenes contain linear chains of tetrahedra. They usually are metastable. On earth, silicates mostly crosslink into two-dimensional[83] of three-dimension

[79] Exceptions are the ionic liquids and the silicone oils
[80] Ammonia is a more polar liquid than the hydrocarbons. The chemistry in liquid ammonia would be more interesting than the chemistry in the lakes on Titan, which contain methane.
[81] That leads to the carbon chauvinism, Box 15.1
[82] The Kelvin radius is the critical size of a cluster in a vapor phase, beyond which the cluster grows, following the Kelvin equation. The Kelvin radius of water is a few nanometers (depending on the supersaturation). In supersaturated water vapor (and even in undersaturated vapor for the special case of water), there are numerous clusters of water, which do not turn into fog, but rather are in a dynamic equilibrium of formation and decay
[83] Two-dimensional structure make out of alumino-silicates form the nano-clays, more generally, the clays.

networks, which are more stable and less mobile than the pyroxenes. Pyroxenes played some role in the prebiotic chemistry, which occurred in the hydrothermal vents at the bottom of sea.

Dust grains have sizes of than a micron. In the protoplanetary disks, physical collisions followed by sticking increase the size, but not to beyond a few millimeters (section 6.1).

Remember:

- The condensed matter in the universe mainly is dust.
- Dust mainly consists of silicates, but also of carbonaceous molecules and ice.

Box 12.1
The discovery of C_{60}

There is experimental evidence for a high silicate content in the dust in the form of a broad emission at 10 μm. Such broad absorption can also be seen in model dusts. The constituents (Si, C, Al, other elements) are first evaporated and then condensed on a cold plate ($T \approx 70°K$) in the presence of argon. The argon forms a solid matrix which traps small particles that have formed on the way to the plate. One takes spectra of these films and hopes for a match with the emission spectra from the dust clouds. Such an agreement is found for silicate particles.

There is a historical remark note to this. At some point, four bands were seen in spectra taken on these “dusts” with no correspondence either on earth or in the sky. The guy scratched his head and took a mental note. At some point he met someone who was familiar with the IR spectra of C_{60}, which at this time were based on calculations with no counterpart in experiment. The IR bands from the soot matched these predictions. All doubts dispersed when ^{13}C NMR showed one single line. Since then, C_{60} exists as a material. Before that it had been a hypothesis and – occasionally – a peak in mass spectroscopy.

12.2 Meteorites

All meteorites hitting the earth originate from the solar system. Many are fragments of asteroids. Some originate from Moon or Mars. Others originate from comets. Of course, meteorites have seen a transformation when clumping together from dust, but the general opinion still is that they talk to us about the nature of solids in space.

Worth a mention are the carbonaceous chondrites. These contain more carbon than other meteorites. Presumably they formed in the outer solar system, where the temperature is low enough to let molecules condense. "Ice" is part of these bodies as long as they are cold. Ice does not reach the ground of the earth, though. In astronomy, "ice" refers to all solids which only are solids at low temperature. Among the molecules forming ice are ammonia, methane, water, CO, CO_2 and of course the larger hydrocarbons.

When meteorites hit Earth, an isotope analysis can be carried out. The isotope analysis on the lunar rocks brought back by the Apollo missions suggest that the Moon was ejected from the Earth by a violent collision with a third body early in the Earth's history. Meteorites from Mars can also be recognized as such by their isotope stamp.[84] It is estimated that there are between 95 and 148 parent bodies for the meteorites hitting Earth. One (out of a few) pathways leading to the isotope stamps is discussed in section 12.4.

Thousands of different molecules have been found in meteorites. The carbonaceous chondrites, i.e. meteorites with a high carbon content, stand out. "Cosmochemistry" often refers to a detailed isotope

[84] For evidence that meteorites with this isotopic stamp are from Mars, see *en.wikipedia*.org/wiki/Martian_meteorite

analysis (Section 12.4). There are "fingerprints". The amino acids in the Murchison meteorite are preferentially left-handed (for whatever reason[85]).

12.3 Recap: The partition sum and its relation to free energy

Among the questions addressed time and again in physical chemistry is the relation between the microscopic properties of molecules, on the one hand, and the macroscopic properties of the materials, on the other. The "partition sum" plays a pivotal role in this connection. All equilibrium properties (specific heat, compressibility, polarizability, …) depend on the thermodynamic potentials and their derivatives in one way or another. Among the thermodynamic potentials are the internal energy, *U*, and the free energy, *A*. Systems in a thermal equilibrium with an environ minimize their free energy, *A*. *A* is a state function with no dependence on history, once equilibrium was reached. *A* depends on volume and temperature.

A can be related to the microscopic properties in a rather simple way. The equation is

Eq. 12.1

$$A = -k_B T \ln Q$$

Q is the partition sum:

Eq. 12.2

$$Q = \sum_{microstates,\, i} \exp\left(\frac{-\varepsilon_i}{k_B T}\right)$$

One may think of the partition sum as the number of thermally accessible states. The ε_i in Eq. 12.2 are the energy eigenvalues from quantum mechanics. The partition sum is where quantum mechanics and materials properties meet.

Arguably, these two equation look a bit simpler than they actually are. Macroscopic bodies have very many microstates. Determining the energy levels of all these states can be a formidable task. In the next section we will not actually compute these levels. We will simply argue that the vibrational energy levels decrease when the material is enriched with heavy isotopes and that the partition function increases, in consequence. For now, remember Eq. 12.1 and Eq. 12.2.

Remember:

– The partition sum (the number of thermally accessible states) links the energy levels from quantum mechanics to the equilibrium properties of the macroscopic media.

12.4 Enrichment of heavy isotopes in condensed phases

Because most processes in chemistry and physical chemistry weakly depend on the weights of the nuclei, isotopic fingerprints can be time capsules.[86] Isotope fractionation – once achieved – remains unchanged over millions of years, if not billions.

---

[85] Crystallization will typically involve molecules with one handedness, only. When such crystals grow, the molecules with the wrong handedness are not incorporated and evaporate more easily than the molecules incorporated in the crystals. If a solid body contains few crystals, that mechanism may lead to an asymmetry in the abundance of both types. Preferred handedness in the Murchison meteorite does not prove that life originated outside the earth.

[86] There are two cases, where the isotopic composition noticeable affects chemical reactivity. The first one is the hydrogen/deuterium pair. Deuterium is twice as heavy as hydrogen. Isotope effects are strong for that reason. Heavy water is poisonous to humans. The other exception is ozone. The different isotopomers behave much differently in chemical reactions – for reasons that cannot be explained quickly.

Mars meteorites can be recognized by their isotope ratios. The $^{18}O/^{16}O$ ratio at a certain depth of a drill core from antarctic ice correlates with the climate at the time of formation. The origin of ores – potentially sourced from the Congo, where such materials are subject to an embargo – can be traced using the isotope stamp. Isotope stamps can also be obtained from small sample quantities, using mass spectroscopy. The phase equilibria between condensed phases (liquid or solid) and vapor are particularly important for isotope fractionation in the bodies of the solar system. (There are other factors, including cosmic radiation and radioactive decays.)

In the following, the focus is on why heavy isotopes are enriched in condensed phases. The pair $H_2{}^{18}O/H_2{}^{16}O$ is an example.[87] When there is a transport equilibrium between liquid water and water vapor, the liquid will be enriched with $H_2{}^{18}O$. Because the enrichment increases with decreasing temperature, snow in Canada contains more $H_2{}^{18}O$ than rain in Florida.

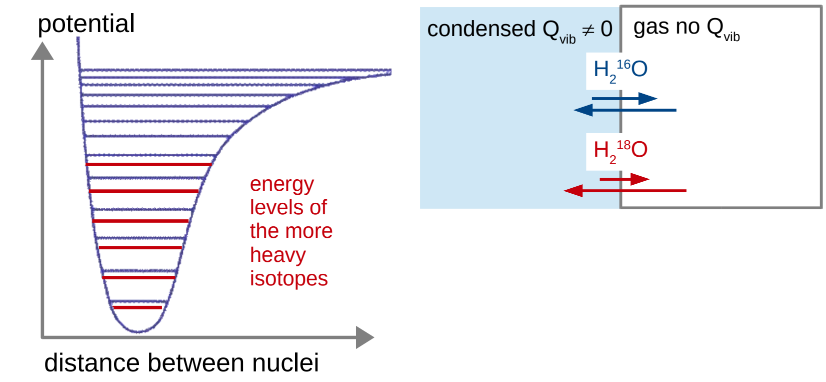


Fig. 12.2
Left: Heavier isotopes possess lower vibrational energy levels, resulting in more thermally accessible states. This increases the partition function and decreases free energy.
Center: There is no vibrational partition function in gases; the phenomenon described on the left occurs only in liquids or solids. Heavy isotopes have lower free energy in the condensed phase than light isotopes, leading to their enrichment.

A large part of the thermal energy in condensed media is contained in vibrations. We consider local vibrations, only. We assume the interaction potential to be parabolic, as in the lower part of the potential shown on the left-hand side in Fig. 12.2, The energy levels are given as

Eq. 12.3

$$\varepsilon_v = \hbar\omega\left(v + \frac{1}{2}\right)$$

v is the vibrational quantum number, ω is the frequency. The frequency is related to the spring constant, κ, and the reduced mass, μ, by

Eq. 12.4

$$\omega = \sqrt{\frac{\kappa}{\mu}}$$

This same relation holds for the classical harmonic resonator.

Increased mass slows down the vibration, thereby decreasing the energy levels (left in Fig. 12.2). The number of accessible levels increases. Because $A = -k_B T \ln Q$, an enrichment in $H_2{}^{18}O$ lets the free

[87] Do not confuse $H_2{}^{18}O$ with heavy water. Heavy water is D_2O with D for deuterium.

energy of the liquid phase decrease. Nature likes low free energy and therefore enriches the condensed phases with the heavy isotopes. The above text glossed over a few complications.[88]

Remember:

- Isotope stamps are highly conserved.
- Heavy isotopes are enriched in condensed matter because the vibrational partition sum increases with increasing mass of the nuclei.

[88] M.M. White: Isotope Geochemistry, Chapter 8. *www.geo.cornell.edu/geology/classes/Geo656/656notes13/IsotopeGeochemistryChapter8.pdf*

13 Dark matter

The existence of dark matter is inferred from
– the velocities of the galaxies in the Coma cluster being incompatible with the virial theorem[89]
– the velocities of stars in globular clusters being incompatible with the virial theorem.
– the rotation curves of galaxies (Fig. 13.1)
– how dark matter bends light, forming gravitational lenses
– the expansion rate of the universe.

Interestingly, accelerated motion of stars is not part of this list. Velocities of stars (in addition to their positions) have been determined for many stars by the Gaia satellite. Because acceleration is the 2nd time derivative of position, its determination requires highly precise positions. At this point, the distribution of dark matter cannot be inferred from the accelerated motion of stars.

The "cold dark matter" in the Λ–CDM model obeys thermodynamics. Otherwise it could not be cold. Cold dark matter is trapped by gravitation. Hot dark matter escapes gravity.

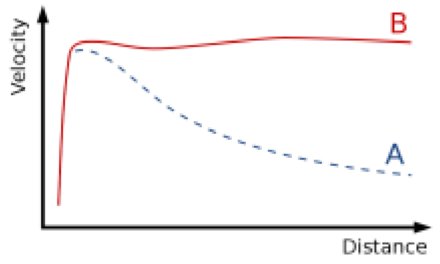


Fig. 13.1
Rotation curve of a typical spiral galaxy:
A: Predicted from centripetal force exerted by visible matter:. **B**: Measured from the Doppler shift of the 21-cm line (Section 10.2).
The difference is attributed to dark matter.
en.wikipedia.org/wiki/Spiralgalaxy

At present, we only know what dark matter is not made of. It does not consist of neutrinos, because these would be hot. It does not consist of a large number of compact, cold objects that cannot be seen by optical telescopes due to their low brightness. These "MACHOs" (massive compact halo objects) would be detected indirectly as gravitational micro-lenses. When compact objects enter the line of sight to a star, they produce a brief, transient brightening of the star (Fig. 13.3). This effect exists, but it is not very common. If there were a large number of compact objects (baryonic or not), these would be seen in the form of micro-lensing.

Dark matter probably does not consist of "weakly interacting massive particles" (WIMPS), i.e. particles that are detected based on the weak interaction. (Neutrinos are detected based on the weak interaction). There have been extensive searches for WIMPS, but researchers so far have come back empty-handed.

At present, dark matter is only evidenced by gravitation. There are maps of the distribution of dark matter. These show that the dark matter in spiral galaxies is less concentrated to the disk than the baryonic matter (stars, gas, dust).

Because gravitational lensing is most efficient on large scales, it is easiest observed in clusters of galaxies. Whether or not one sees gravitational lensing depends on the ratio of the Schwarzschild radius to the geometric radius.[90] (If the Schwarzschild radius were larger than the geometric radius, the object would be a black hole.) "Weak gravitational lensing" turns a circular galaxy located behind the lens into a galaxy with a slightly elliptical appearance. For any individual galaxy, a slight distortion due to gravitational lensing is indistinguishable from true ellipticity. However, the images of very many galaxies can be statistically analyzed (Fig. 13.2). This is one of the research goals of the "Euclid" space telescope.

---

[89] That was how Fritz Zwicky first noticed dark matter in 1933.
[90] The ratio $R_{\text{Schwarzschild}}/R_{\text{geometric}}$ estimates how large the effects of general relativity are for a certain problem. The fact that this ratio is larger for galaxy clusters than for galaxies is not trivial.

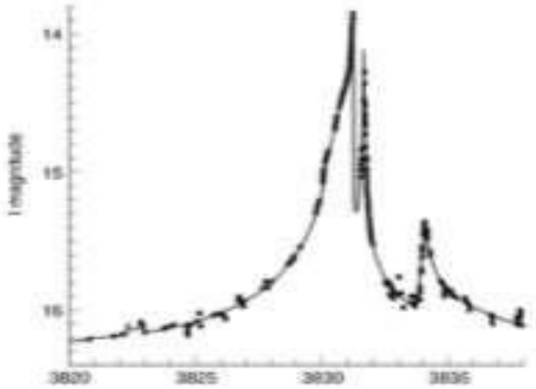
Because the nature of dark matter is so elusive, people have claimed that its numerous effects actually are the consequence of an incomplete understanding of gravity on our side (meaning: no dark matter).

Remember:

– The universe contains more dark matter than baryonic matter.
– Dark matter is only noticed from its gravitation
– Dark matter is subject to thermodynamics. The dark matter in this universe is "cold" in the sense that it is trapped in minima of the gravitational potential.

32. Weak lensing in the galaxy cluster Abell 2218. The images of galaxies that lie behind the cluster are stretched by the cluster's gravitational field into arcs perpendicular to the field direction.

Fig. 13.2
If the galaxies in a certain region are elliptical with a certain relative orientation to a center, one infers "weak gravitational lensing" by dark matter.
Source: Binney (footnote 1)



33. The micro-lensing event OGLE-2006-BLG-109. The brightness of the star measured at 12 observatories is plotted against time in units of a day. The gravitational fields of two planets generate extremely rapid brightness fluctuations. These data yield the mass of the star as 0.51 M? and the planets' masses as 231 M_e and 86 M_e similar to Jupiter and Saturn.

Fig. 13.3
If a star transiently becomes slightly brighter, a compact, invisible body may have crossed the line of sight, causing "microlensing".
Source: Binney (footnote 1)

14 Neutrinos

14.1 Neutrino oscillations prove that neutrinos have a mass

There are about as many neutrinos in the universe as protons (give and take a factor of 100), but they are almost invisible because of their weak interaction with other particles. In the famous Homestake experiment from 1970s, researchers counted the neutrinos that reach Earth from the sun. This was accomplished using a tank filled with tetrachloroethene, where chlorine reacts with neutrinos to produce ^{37}Ar, which has a half-life of 35 days as it decays back into chlorine. To detect the decay, the tank was periodically flushed with helium as a carrier gas. The argon was collected in a cold trap and underwent purification before being subjected to a decay analysis, lasting for 250 to 400 days. Given the low count rate of approximately one event per week, meticulous precautions were necessary to reduce background.

The Homestake team was confident that they had sufficiently calibrated their detectors. They also believed to have understood the production rate of neutrinos in the sun. Having calibrated the detector, knowing distance to the sun, and knowing the production rate, they expected a count rate about three times as high as what they saw, which is the "solar neutrino problem." Not everyone shared their confidence at the time, but thirty years later, it was confirmed that two out of three neutrinos transform into different types while traveling to Earth.[91] The process is called "neutrino oscillations". The Homestake experiment does not see these other neutrinos. The "neutrino oscillations" prove that neutrinos possess a rest mass.

For neutrinos, there is a difference between "reaction states" and "energy eigenstates." Such a difference is not typically for molecules, but it does exist for ammonia (Fig. 14.1). The chemical reactions that produce ammonia result in a tetrahedral configuration, where the nitrogen atom can occupy positions either to the left of the triangle formed by the three hydrogen atoms ("L") or to the right ("R"). However, these positions do not represent energy eigenstates because the nitrogen atom can tunnel between L and R. The tunneling is also referred to as "pyramidal inversion". The energy eigenstates correspond to the symmetric ("symm") and antisymmetric ("anti") superpositions of L and R. The "symm" and the "anti" states are also called "mass states," reflecting the relationship $E = mc^2$.

Reactions states differing from mass states also exit for neutrinos (and also for quarks). Similar to how the nitrogen atom tunnels between L and R, the neutrinos oscillate between the different reaction states (electron neutrino, muon neutrino and tau neutrino). Similar to how "symm" and "anti" have different energy in the ammonia molecule, the energy eigenstates have different energy eigenvalues. Because of $E = mc^2$, they differ in mass, which means that they must have mass.

A recent study from Karlsruhe[92] aims to determine the mass of neutrinos and reports that, as of 2025, the mass is less than 0.45 eV. The most stringent upper limit on neutrino mass comes from astrophysical observations. The universe is contains about 300 neutrinos per cm^3. If the sum of the masses of all three types of neutrinos were more than 0.12 eV, their presence would delay the expansion of the universe, contradicting experimentally observed rates of expansion.

Neutrinos are classified as "hot dark matter," distinct from "cold dark matter". While cold dark matter plays an essential role galaxy formation, neutrinos, due to their minimal mass, pass through the precures of the galaxies undisturbed.

---

[91] Nucleosynthesis in the sun produces electron neutrinos. The other types of neutrinos are the muon neutrinos and the tau neutrinos.

[92] https://de.wikipedia.org/wiki/KATRIN

Remember:

– There is difference between reaction states and energy eigenstates for the ammonia molecule, for the neutrinos, and for the quarks.
– Neutrino oscillations proves that neutrinos have a mass.

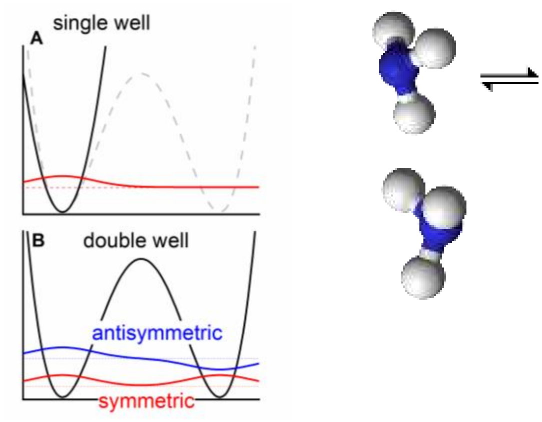


Fig. 14.1
A: The Gaussian solves the time-independent Schrödinger equation for the parabolic potential. **B**: For two neighboring potential wells, the solution can be approximated by the sum and the difference of the solutions to the single parabolic potential. The "pyramidal inversion" is illustrated on the right.
commons.wikimedia.org/wiki/File:Nitrogeninversion-3D-balls.png

14.2 Neutrino mixing may explain the imbalance between matter and antimatter in the universe

For the ammonia molecule, the transformation between {L,R} and {symm, anti} amounts to a matrix. For ammonia, symmetry enforces a simple form, which is

Eq. 14.1

$$\begin{pmatrix} \Psi_{symm} \\ \Psi_{anti} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{pmatrix} L \\ R \end{pmatrix}$$

For ammonia, the matrix is not only simple, it is also purely real. What if it were complex? A complex matrix would create relative phases, which would not revert sign under time reversal. The term iωt, on the other hand, does revert sign. The consequences are of the most astonishing type.

For illustration, let "anti" have the phase φ relative to "symm". anti then lags behind symm in time (or runs ahead in case the phase is negative). A factor exp(–iφ) occurs in addition to the factor exp(–iωt), accounting for different frequencies of the anti- and the symm-state. The wave functions read as

Eq. 14.2

$$\Psi_{symm} = \frac{1}{\sqrt{2}}(L+R)\exp(\mathrm{i}\omega t)$$
$$\Psi_{anti} = \frac{1}{\sqrt{2}}(L-R)\exp(\mathrm{i}\varphi + \mathrm{i}\omega t)$$

If time were to flow backwards, the delayed state would turn into the state running ahead and vice versa because φ stays unchanged.

The matrix of neutrino mixing (the “PNMS matrix”) *is* complex. The matrix for quark mixing also is complex. The complex matrix violates the time-reversal invariance at the level of microscopic dynamical laws. It may be regarded as an "arrow of time", but it has nothing to do with the increase in entropy, which also is an arrow of time (section 5.12).

The time-reversal asymmetry is of moderate importance as far as time is concerned. There are other, more interesting arrows of time. More importantly, time-reversal asymmetry implies a violation of the symmetry between matter and antimatter. All gauge theories are "CPT-invariant".[93] "C", "P", and "T" stand for a reversal of charge, parity (handedness), and time. If a process violates T-invariance, it must also violate CP-invariance. CP transforms matter into antimatter and vice versa. Time-reversal asymmetry creates a difference between matter and antimatter. It might explain why the universe contains more matter than antimatter.

That argument can be made quantitative. A CP violation also exists for quarks, but this one is too weak to explain the observed imbalance between matter and antimatter. For neutrinos, CP violation is known, but the value of the phase is not yet known well enough to relate it to the matter/antimatter excess. That may change in about 10 years. Large, funded neutrino experiments will then start operation and – hopefully – answer one of most fundamental open questions of current physics.

Remember:

- The matrix, which connects the reaction states of the neutrinos to the mass states, is complex.
- Why bother? Because that violates time reversal symmetry.
- Neutrinos are an "arrow of time" but not a very important one, compared to the second law of thermodynamics. Time reversal asymmetry, however, implies an asymmetry between matter and antimatter. The predominance of matter in the universe may be rooted in the non-trivial phase of neutrino mixing.

[93] The "gauge theories" are the formalism underlying the standard model of physics.

15 The anthropic principle in cosmology

15.1 Fine tuning of the fundamental constants

The question of variable fundamental "constants" comes up in the context of the fine-tuning problem. It can be argued that life would not exist if the fine structure constant (governing the strength of electromagnetism, α) and the coupling constant of the strong interaction (g) had values slightly different from what we know. This fortunate circumstance may be viewed as a consequence of the anthropic principle. The anthropic principle applied to cosmology postulates that the universe is much larger than what we see.[94] We see to a horizon, which is 46 billion light years away.[95] Only in this small bubble do α and g have the favorable values (favorable for us, the complex living entities). Elsewhere they are different. We live in a nice corner of this large universe. For the same reason, we live on Earth instead of Mars. That is not surprising. We live here *because* it is nice here.

Box 15.1:
We live where it's nice

The proponents of the fine-tuning problem claim that life is only possible in the narrow bar on the left in Fig. 2.2. If the nuclear forces were stronger (top left), there would be no hydrogen atoms, because two hydrogen nuclei would fuse to form a diproton. Without hydrogen, there would be no small molecules, because hydrogen atoms terminate the growth of (e.g.) carbon chains. Fluorine, chlorine, bromine and iodine are too rare to replace hydrogen in this respect. Without small molecules, there would be no liquids with low viscosity.

If the nuclear force were much weaker, the stars would not shine because nuclear fusion would not be efficient enough (bottom left).

If the Coulomb repulsion between the protons in the nucleus were greater, carbon would decay radioactively just like uranium (bottom right). No carbon, no life ("carbon chauvinism").

If both α and g were large, the atoms would be so stable that they would no longer react chemically (top right).

This diagram has been criticized because it ignores alternative mechanisms. Other fundamental constants would not only close existing entries to the path of evolution, but also open new ones.

There are similar diagrams in which one of the axes is the strength of gravity.

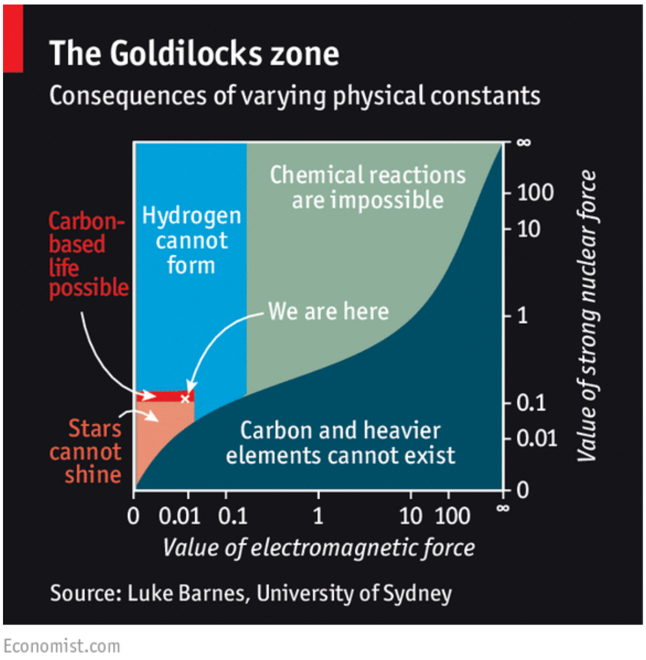


Fig. 15.1
A map of the problems that would arise if the fine structure constant α and the coupling constant of the strong interaction g had different values than they have.
https://www.niepelt.ch/wp-content/uploads/2015/08/20150815_BBC928_0.png

That fundamental constants might vary in time or space can be motivated with the dielectric constant, ε. ε is materials parameter, depending on temperature ($\varepsilon \approx 78$ for water at room temperature). If there were intelligent creatures living in water at some fixed temperature (not knowing anything other than water), these would think that ε was a fundamental constant. Similar creatures might live in ice and

[94] Sometimes this larger universe is called the "multiverse", but the multiverse also has other meanings.
[95] 46 billion light years is the diameter of the visible universe *today*. The universe is only 13.7 billion years old. Today's diameter is greater than 13.7 billion light years because of the cosmic expansion.

not be aware of the fact that ice can crystallize in different ways, and that, further, ε depends on the crystal modification. These creatures would be surprised to hear that ε takes values other than the one they know in other crystal phases.

What we perceive as the vacuum is not empty, at all. It is vibrant with vacuum fluctuations.[96] Some of those fluctuations we know and understand. However, the smallest spatial scales visible to us are at about 10^{-16} m. The Planck length, below which space and time lose their meaning, is at about 10^{-35} m (Fig. 15.2, see also Box 15.2). There are some 19 decades not visible to us. What are fundamental constants to us, might depend on funny things going on in these 19 decades. They might depend on temperature (similar to ε in liquid water) and they might vary in space in case the funny things have cooled into different states at different places in the universe.

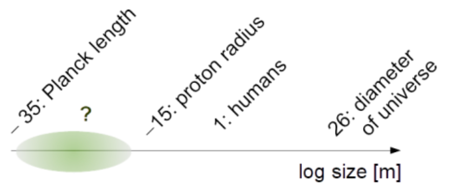


Fig. 15.2
The concept of space loses its meaning below the Planck length and above the diameter of the universe. There are some 20 decades below the proton radius, which we know nothing about. Possibly, the "fundamental constants" are materials parameters of whatever lives and thrives in that size range.

In the following, we first introduce a mechanism that can bring about a very large universe (cosmic inflation, sec-

Box 15.2
The Planck length

The Planck scale of 10^{-35} m, on which the effects of quantum gravity become noticeable, arises if one postulates an elementary particle that forms a black hole on its own. According to everything we know, electrons are point particles. Why are they not black holes? A black hole is an object whose *entire* mass is located within the Schwarzschild radius associated with its mass, given as

$$r_S = \frac{2Gm}{c^2}$$

G is the gravitational constant. To make that relation plausible, multiply it by mc^2 and divide by r_S:

$$mc^2 = \frac{2Gm^2}{r_S}$$

For black holes, the energy contained in gravity is about as large as the energy contained in the rest mass.

In the context of quantum mechanics, even a point particle is not a black hole as long as the extension of the wave package, Δx, is greater than the Schwarzschild radius. If the mass increases, the Schwarzschild radius increases and Δx as estimated from the uncertainty relation decreases because of $\Delta x \approx \hbar/(2\Delta p)$ with Δp proportional to the mass. When Δx and r_S are equal, the Planck mass has been reached.

Write the size of the wave package as $\Delta x = \hbar/(2\Delta p)$, assume the relativistic energy-momentum relation ($E^2 = (mc^2)^2 + p^2c^2$), and neglect the rest mass. For relativistic particles, $E \approx cp$ and $\Delta p \approx \Delta E/c$. Assume $\Delta E \approx E$ and replace the mass m in the equation for the Schwarzschild radius with E/c^2:

$$r_S = \frac{2GE}{c^4} \approx \frac{G\Delta E}{c^4} \approx \frac{G\, c\Delta p}{c^4} \approx \frac{cG\hbar}{c^4\Delta x} = \frac{G\hbar}{c^3\Delta x}$$

The factor 2 has disappeared behind the ≈. Requiring $l_P \approx r_S \approx \Delta x$ leads to

$$l_P \approx \sqrt{\frac{G\hbar}{c^3}}$$

Inserting values leads to $l_P \approx 10^{-35}$ m. Such a particle would have a mass of around 20 µg (the "Planck mass"), corresponding to 10^{19} proton masses.

[96] Some people argue that the standard deviation of the temperature in Fig. 3.2 (18µK, which is a fraction of about 10^{-5} of the temperature), should count as a fundamental constant. The number is really important, agreed. Still: Those 18 µK are the result of a vacuum fluctuation in the very early universe.
Martin Rees: "Just Six Numbers: The Deep Forces That Shape The Universe", Basic Books 2001.

tion 15.2). Second, it is proposed that the fundamental "constants" depend on temperature,[97] where "high temperatures" are temperatures beyond 10^{15} K. Third, it is proposed that the cooling that took place after the Big Bang has led to different values for the low-temperature values of the fundamental constants in the different regions of the large universe.

Remember:

- The world around us would be much different if the strength of the electromagnetic interaction and of the strong interaction were slightly different from what they are.
- If the nuclear force was stronger, there would be diprotons. There would be no protons and no hydrogen atoms. There would be few small molecules.
- If the electric repulsion between protons was stronger, carbon would undergo α-decay. That would prevent life because of the carbon chauvinism.
- If the nuclear force was weaker, the stars would not shine.
- The fundamental constants might not be fundamental. They might result from things happening at small scales, which we do not understand. In that regard, they might be similar to materials constants, such as the dielectric permittivity of water, ε.
- Nothing can be smaller than 10^{-35} m, which is the Planck length. l_p is 20 decades below the proton radius. Experimental studies on such mall objects are out of reach. Possibly, there is a rich set of phenomena below our bottom of spatial scales, which governs what we perceive as fundamental constants.

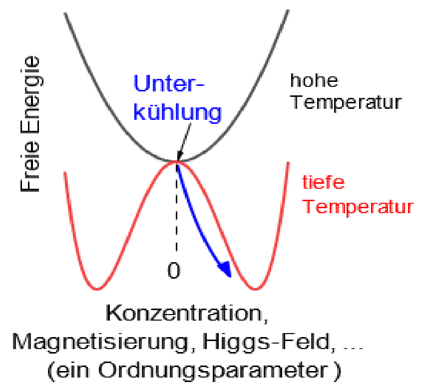


Fig. 15.3
Spontaneous symmetry breaking upon cooling a system to below its critical point.

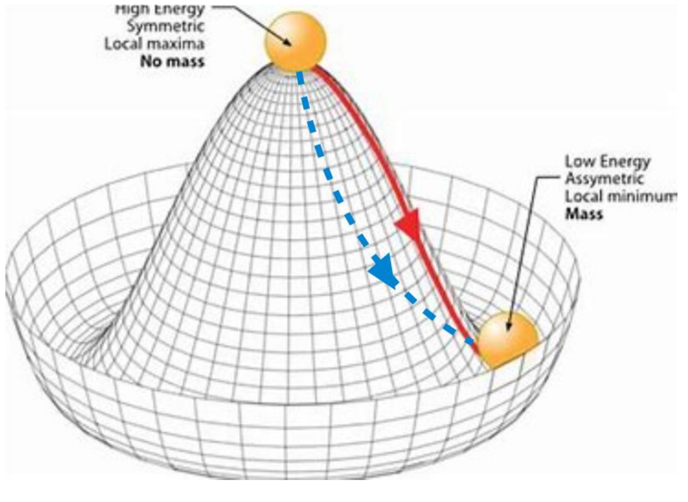


Fig. 15.4
The energy contained in the Higgs field does not have its minimum at zero field. This does not matter at high temperatures. Averaged over time, the Higgs field then still is zero. At low temperatures (< 10^{15} K), the time-average of the Higgs field shifts to the groove, which spontaneously breaks a symmetry. The blue dashed arrow indicates the possibility of supercooling.
Adapted from *physics.stackexchange.com/questions/486003/how-can-the-higgs-field-sense-that-the-temperature-is-high-enough-to-restore-t*

15.2 Cosmic inflation

The reason for formulating the inflation hypothesis was not the fine-tuning problem, but rather were two peculiarities of the Big Bang model. First, the microwave background has almost the same temperature everywhere, although the different regions at the time were unable to communicate with each other due to the finite speed of light. The regions are mutually outside the respective "causal horizons". Heat

[97] Such "gliding coupling constants" are also known from the standard model of particle physics.

did not flow between these regions. This is the "horizon problem". Second, the mass density of the universe is astonishingly close to the critical mass that marks the transition between a continuing expansion and a transition into the "big bounce". People searching a reason for this coincidence call it the "flatness problem".

Allan Guth explained these (and a few other) peculiarities in 1981 by claiming that the universe visible to us is part of a much larger universe. This would require the universe to have expanded much faster in an early phase than what is predicted by the conventional Big Bang model. The expansion was dubbed "cosmic inflation". Since cosmic inflation, there are many such bubbles and the bubbles are all locally flat. The local flatness entails a mass density close to the critical mass density.

The driver for inflation has an analogy in ferromagnetism, sketched in simplified form in Fig. 15.3. The importance of entropy increases at high temperatures. The entropy takes is maximum value when the elementary magnets are distributed randomly and when the macroscopic magnetization vanishes, in consequence. At temperatures below the critical temperature (here: below the "Curie temperature"), the influence of enthalpy increases and the magnets align themselves parallel to each other. In between, there is a 2nd-order phase transition, which causes a spontaneous symmetry breaking because the magnets (together) might point into all directions.

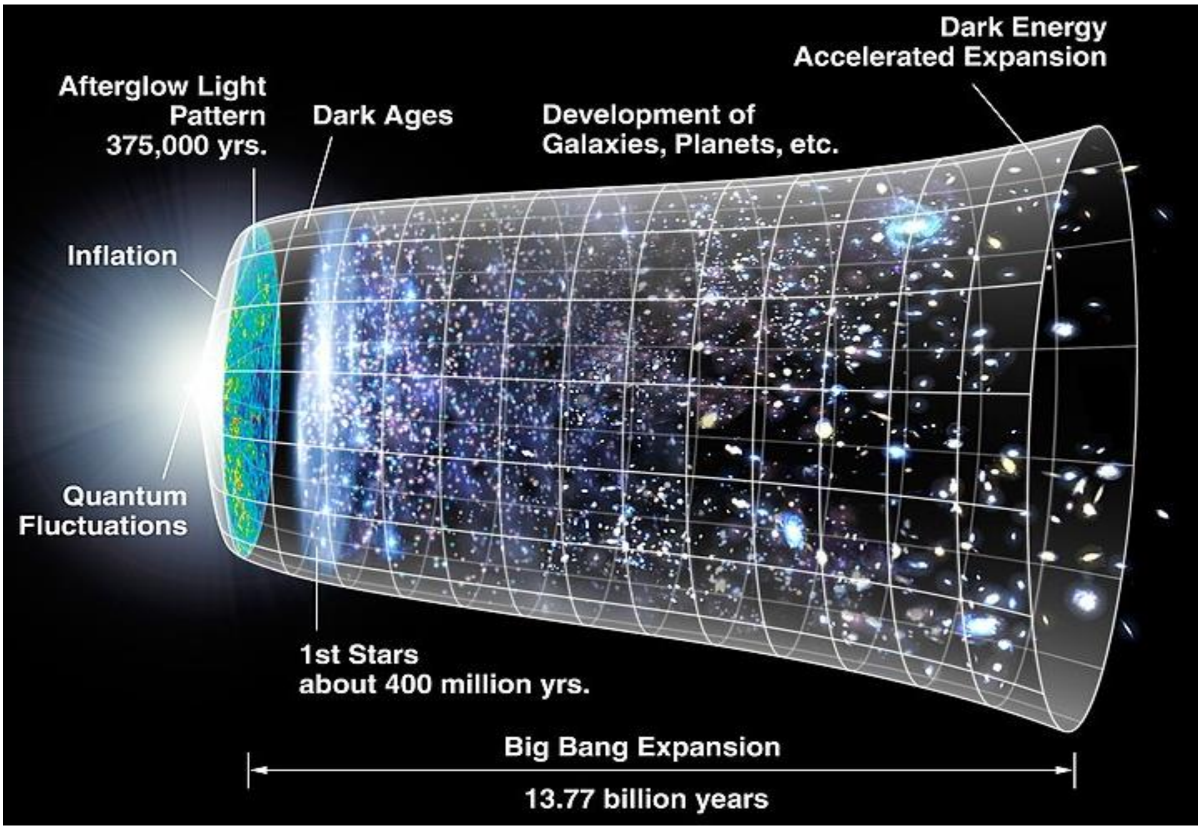


Fig. 15.5
The current cosmological model including inflation and the current accelerating expansion
en.wikipedia.org/wiki/Expansion_of_the_universe

There is no such 2nd-order phase transition for electromagnetic fields in vacuum. The vacuum does contain quantum fluctuations of E, but E fluctuates around zero at all temperatures. $E = 0$ is the ground state. The field-free vacuum also is the ground state for the fields of the weak and the strong interaction.

The respective particles (the vector bosons W^+, W^-, and Z_0 as well as the gluons) are created spontaneously for short times, but the wave functions still fluctuate around zero.

Only the Higgs particle is different (Fig. 15.4). The energy landscape above the Higgs field has its lowest energy at a finite value of the Higgs field. More detailed work following up onto the first proposal by Guth has shown that inflation cannot actually be driven by the Higgs field. It is believed that there is some other, currently unknown field (the "inflaton field"), which drove inflation.

Guth postulated two things. First, the inflaton field fluctuated around zero in the early universe. It became nonzero upon cooling. Second, the cooling was so incredibly fast that the inflaton field could not follow. This "supercooling"[98] is indicated by the dashed arrow in Fig. 15.4. Supercooling is associated with an energy in the same way that a supercooled melt contains a latent heat which is released when the crystallization is nucleated (as in heat cushions[99]). The analogy of the latent heat is the "dark energy" of the vacuum, which created a negative pressure and rapidly expanded the universe.

As a side remark, Fig. 15.5 shows a well-known graph summarizing the current model of cosmology, which is the Λ-CDM model. Λ stands for the dark energy, CDM stands for cold dark matter.

<u>Remember:</u>

– We may be living in a small bubble of a much larger universe, where the "fundamental" constants have by chance taken the values, which allow for life.
– The large universe came about by inflation.

15.3 A time dependence of the fine structure constant?

The fine structure constant, α, can be measured with good precision and one can search for a temporal variation of α ($\dot{\alpha}$, pronounced "alpha-dot"). The name "fine structure constant" suggests that it would pertain to a detail from atomic spectroscopy. However, the fine structure constant may well be called the most important constant of nature. It is more important than the speed of light, c, Planck's quantum of action, h, and the vacuum permeability, ε_0, because c, h and ε_0 are not fundamental constants in the narrow sense. Their values depend on the choice of the units meter, second, and joule. In order to avoid this dependence of the fundamental constants on units, so-called "natural units" can be defined. This is done by writing $\hbar = c = 4\pi\varepsilon_0 = 1$. The units joule, meter and coulomb disappear. For instance, the light-second is the unit of length. Only the second remains.[100] The fine structure constant – being dimensionless – does not depend on the definition of the units. α is defined as

Eq. 15.1

$$\alpha = \frac{e^2}{4\pi\varepsilon_0 hc}$$

e is the elementary charge. The numerical value of the fine structure constant is $\alpha = 7.297352566\times10^{-3}$. People also remember the reciprocal value, which is $\alpha^{-1} \approx 137$. In natural units, the fine structure con-

[98] This is a special form of supercooling insofar, as the supercooled state is unstable rather than metastable. Supercooled liquids and glasses are metastable. In unstable supercooled media, any local symmetry breaking immediately grows, growth only being limited by kinetics. In metastable media the transition requires nucleation.

[99] *en.wikipedia*.org/wiki/Latent *heat storage*

[100] In the context of thermodynamics, $k_B = 1$ is also sometimes written. In this case, the temperature has the unit of energy. (The energy for this case is not reduced to the second by choosing $\hbar = 1$).

stant is the square of the elementary charge. The charge of the electron pervades all of physics. Implications of the fine structure constant for the physics of atoms and molecules are discussed in section 16.1. Since the fine structure in the atomic spectra is caused by magnetic interactions (leading to spin-orbit coupling, Box 15.3), the fine structure constant can be determined spectroscopically from the absorption spectra of atoms. Atomic spectroscopy can be carried out both in the laboratory and in astronomy.

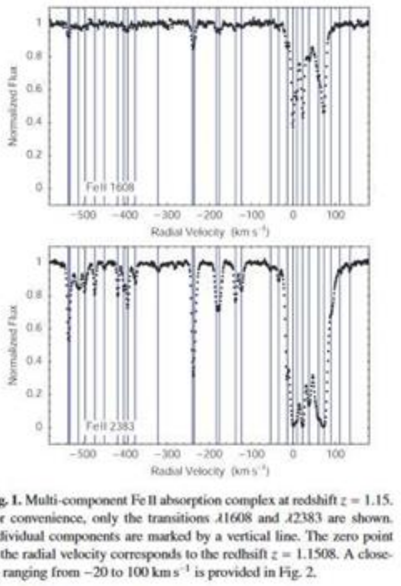


Fig. 15.6
Spectra of quasar HE 0515-4414.
The splitting results from the Doppler shifts in different regions of the source. The analysis focuses on the difference between the wavelengths of the two underlying iron lines (upper and lower graph).
Quast, R.; Reimers, D.; Levshakov, S. A., Probing the variability of the fine-structure constant with the VLT/UVES. *Astronomy & Astrophysics* **2004,** 415, (2), L7-L11.

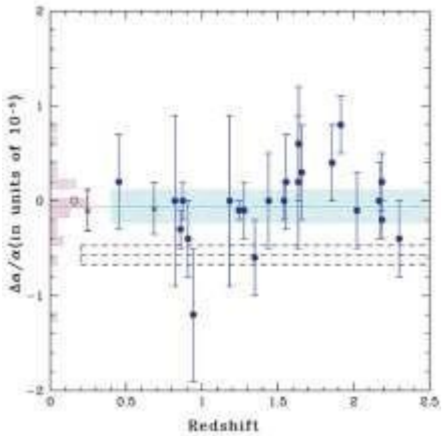


Fig. 15.7
The UVES measurement values of the relative change of the fine structure constant alpha from the sample of absorption systems, plotted as a function of the redshift and the corresponding look-back time. The open circle is the measurement from the Oklo natural reactor (near Gabon, West Africa). The horizontal long dashed lines show the limits of the previous claim of a variable fine structure constant. Clearly, the UVES data are inconsistent with this range.
www.eso.org/sci/publications/messenger/archive/no.116-jun04/messenger-no116-25-28.pdf
"UVES" stands for Ultraviolet and Visual Echelle Spectrograph.
α takes an influence on the distribution of isotopes generated in a fission reactor (such as the Oklo natural reactor)

This raises the question of whether α has the same value everywhere and has had that value in all past. α might change on cosmic timescales and this is what the authors of Ref. 101 have claimed. The dashed horizontal lines in Fig. 15.7 are their limits. Spectroscopy on sources in the sky will always be less accurate than spectroscopy in the laboratory because the Doppler shift turns all lines into bands. This is also the case in Fig. 15.6. Two different spectral ranges are shown. The source is the quasar[102] HE 0515–4414. The patterns are similar because the Doppler shift has a similar effect on both lines. These are two iron lines in absorption. They are absorption lines because the iron atoms are contained in

[101] J. K. Webb; M. T. Murphy; V. V. Flambaum; V. A. Dzuba; J. D. Barrow; C. W. Churchill; J. X. Prochaska; Wolfe, A. M., Further Evidence for Cosmological Evolution of the Fine Structure Constant. *Phys. Rev. Lett.* **2001,** 87, 091301

[102] "Quasar" stands for "quasi-stellar object". A quasar is a type of active galactic nucleus (AGN), essentially the intensely luminous core of a distant galaxy. These cores are powered by supermassive black holes at the center of the galaxy, devouring vast amounts of surrounding gas and dust. Quasars are called "quasi-stellar", because they are not spatially resolved by telescopes. Quasars have a large distance to the earth because they only existed in the early universe.

the outer, colder regions of source. Light with a wavelength matching the absorption line, which penetrates from the interior of the source, is partly absorbed in the colder outer layer.[103] (The same is true for the Fraunhofer lines in Fig. 2.1.) Heavy elements have strong spin-orbit coupling (Box 15.3). □□might also be determined from the famous splitting of the sodium D-line, but the measurement is easier for iron.

The authors of Ref. 101 remained alone with their results. Subsequent investigations have shown that $\dot{\alpha}/\alpha$ is compatible with zero even on cosmic time scales. Fig. 15.7 gives an overview. The 3σ–confidence interval for $\dot{\alpha}/\alpha$ from the astrophysical measurements stretches from $-2.5\cdot10^{-16}$ year^{-1} to $+1.2\cdot10^{-16}$ year^{-1}.

Box 15.3:
Spin-orbit coupling

In quantum physics, only the total angular momentum of an electron is conserved. The orbital angular momentum and the spin are not conserved individually. If, for example, the spin has the value +1/2 at the beginning of an experiment, angular momentum is exchanged back and forth between spin and orbit over the course of time. The energy eigenstate has neither a defined orbital angular momentum nor a defined spin, but only a defined total angular momentum. This is another example where the "reaction states" and the "energy eigenstates" are different. The reaction state has a well-defined spin, because the process of formation dictates that.

The process of spin-orbit coupling can be illustrated using an example from astronomy. The moon has undergone such a coupling. It used to rotate around its own axis independently of its orbit around the Earth. Today, however, it performs a bound rotation. Today, we always see the same side of the moon. In the case of the moon, the spin-orbit coupling was an irreversible process. It was brought about by the tidal forces. Although the moon has no oceans, the rotation led to a small deformation, analogous to ebb and flow on Earth. On Earth, ebb and flow also lead to spin-orbit coupling. Over time, these effects lengthen the days and shorten the years. On the time scale of a few years, other, stronger effects are superimposed on this effect (e.g. redistribution of mass between the Earth's core and mantle, which changes the moment of inertia, similar to the star quakes discussed in section 8.3). 400 million years ago, the day lasted only 21.5 hours. The duration of a day can be inferred from the fact that the year had 405 days, as evidenced by the growth rings on trees. Because the total angular momentum is constant, a longer year implies shorter days.

There are no tidal forces for electrons. This role is taken by magnetism. Place the center of the coordinate system at the position of the electron. From the electron's point of view, the nucleus cycling around it amounts to an electric current that generates a magnetic field. Whether the spin is parallel or antiparallel to this field makes a small difference to the electron's energy. This splitting of energy levels is referred to as the "fine structure" (section 16.1).

$\dot{\alpha}$ can also be determined in the laboratory.[104] The frequencies of two transitions of the ytterbium atom are determined with high precision and these measurement are repeated after some time. The fine structure constant can be computed from the ratio of the two frequencies. The result was $\dot{\alpha}/\alpha = (-0.7\pm2.1)\cdot10^{-17}$ year^{-1}.[105] Since the uncertainty is larger than the mean, the result is compatible

---

[103] Depending on the wavelength, you look into the source to different depths. At the wavelength of a strong line, the frequent absorptions and re-emissions mean that only the outer (colder and darker) areas of the photosphere are visible. Next to the line you see the hotter and brighter interior of the source.

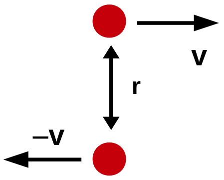
[104] Godun, R. M.; Nisbet-Jones, P. B. R.; Jones, J. M.; King, S. A.; Johnson, L. A. M.; Margolis, H. S.; Szymaniec, K.; Lea, S. N.; Bongs, K.; Gill, P., Frequency Ratio of Two Optical Clock Transitions in Yb-171(+) and Constraints on the Time Variation of Fundamental Constants. *Physical Review Letters* **2014,** 113, (21).

[105] In the foreseeable future, optical clocks shall be anchored to an electromagnetic transition occurring in an atomic nucleus (replacing the transitions in the electron shell, which are currently used). This will be the "thorium 229 clock". Because transitions in the nucleus are less disturbed by the environment, the thorium 229 clock may well be a factor of 100 more precise than the current clocks.
The problem here is that high-precision spectroscopy is more accurate than the theory underlying it. This is unlikely to change because the radius of the proton enters the predictions of the energy levels. The proton radius results from the strong interaction and cannot be predicted with the required accuracy from first principles. So why the thorium 229 clock? First answer: Let's see, who knows. Second answer: The accuracy in the determination of $\dot{\alpha}/\alpha$ can be increased with the thorium 229 clock. Determining $\dot{\alpha}/\alpha$ only requires repeated measurements, no theory.

with zero. The bounds from the laboratory experiments are closer to zero than the bounds from astronomy.

Remember:

- Arguably, the fine structure constant is the most important fundamental constant.
- As of 2025, the fine structure constant does not vary with time within the limits of detection.



Box 15.4
Magnetic forces are weaker than electric forces. Magnetic effects increase in strength in heavy atoms.

In chemistry, magnetic interactions are weaker than electric interactions because the electrons move at a speed much below the speed of light. Consider two charges at distance r, moving into opposite directions with velocities +v and –v. The magnitude of the electric force is $|F_{\mathrm{el}}| = -e^2/(4\pi\varepsilon_0 r^2)$. The magnetic force is the Lorentz force ($e(\mathbf{v} \times \mathbf{B})$). We only care about magnitudes: $|F_{\mathrm{mag}}| = |e\mathrm{v}B|$. Following the law of Biot and Savart, the magnetic field caused by a charge in motion is $|B| = \mu_0 \mathrm{v}e(4\pi r^2)$. The ratio of two forces is

$$\frac{|F_{mag}|}{|F_{el}|} = \mu_0\varepsilon_0 \mathrm{v}^2 = \frac{\mathrm{v}^2}{c^2}$$

The relation $c^2 = (\varepsilon_0\mu_0)^{-1}$ was used in the second step. As long as the electrons are non-relativistic (that is, move slower than the speed of light, c) magnetic forces are weak.

In heavy atoms, electrons move fast because the centrifugal force must match the centripetal force and the because the attraction towards the nucleus increases with increasing charge of the nucleus. Magnetic effects therefore are most easily studied using heavy atoms like iron.

Spin-orbit coupling is among the magnetic effects because the spinning electron carries a magnetic moment. The fine structure in spectroscopy causes a small splitting between lines because it is based on magnetism. The ratio v^2/c^2 is equal to α^2 with the α fine structure constant.

15.4 A time dependence of Λ?

OK…. no drift of α, for now. It happens that another fundamental constant was found to depend on time in 2024,[106] namely the density of dark energy as quantified by the cosmological constant, Λ. The standard model of cosmology, the Λ-CDM model, assumes constant Λ, currently valued at 1.4657×10^{-52} m^{-2}. Λ governs the outward curvature on the right-hand side in Fig. 15.5. Working hard, people have measured the relation between the redshift of galaxies (that is, the velocity, by which galaxies recede from us) and their distance with unprecedented accuracy. These results suggest that Λ decreases with time. The statistical evidence is not yet good enough to declare a "discovery", but people talk about it anyway.

Remember:

- Precise measurements of the redshift of galaxies as a function distance suggest that the density of the dark energy as quantified by the cosmological constant, Λ, decreases with time. Λ might not be a fundamental constant in the narrow sense.

[106] https://arxiv.org/abs/2404.06444

16 Appendices

16.1 The role of the fine structure constant

We base the discussion of α on the Bohr model. Values will always pertain to the ground state of the hydrogen atom (the "H1s state", $n = 1$, $Z = 1$). In the Bohr model, angular momentum is quantized in units of $\hbar$:

Eq. 16.1

$$\mu v r = n\hbar$$

μ is the reduced mass, v is the velocity, r is the radius of the circular orbit, and n labels the shell ($n = 1$ in the following). It follows that

Eq. 16.2

$$v = \frac{n\hbar}{\mu r}$$

Apply Newton's second axiom (force = mass × acceleration). For circular motion, the acceleration a is given as

Eq. 16.3

$$a = \frac{v^2}{r}$$

Inserting the electrostatic attraction for the force results in

Eq. 16.4

$$\frac{e^2}{4\pi\varepsilon_0 r^2} = \frac{\mu v^2}{r}$$

e is the elementary charge and ε_0 is the dielectric permittivity of the vacuum. All relations discussed below follow from Eq. 16.1 and Eq. 16.4. .

To calculate the velocity of the H1s electron, insert Eq. 16.2 for *one* of the two v's in Eq. 16.4:

Eq. 16.5

$$\frac{e^2}{4\pi\varepsilon_0 r^2} = \frac{\mu v^2}{r} = \frac{\mu}{r} v \frac{\hbar}{\mu r}$$

Solving for the velocity leads to

Eq. 16.6

$$v = \frac{e^2}{4\pi\varepsilon_0 \hbar}$$

We divide by the speed of light:

Eq. 16.7

$$\frac{v}{c} = \frac{e^2}{4\pi\varepsilon_0 \hbar c}$$

The right side in Eq. 16.7 is the fine structure constant α:

Eq. 16.8

$$\alpha = \frac{e^2}{4\pi\varepsilon_0 \hbar c}$$

The circulation time, *T*, is

Eq. 16.9

$$T = \frac{2\pi}{\omega} = 2\pi\frac{r}{v} = 2\pi\frac{a_0}{c\alpha}$$

a_0 is the Bohr radius (0.5 Å). Its value also follows from the Bohr model, but is not derived here. The wavelength of the photons emitted by the atoms is

Eq. 16.10

$$Tc = \lambda = \frac{2\pi}{\alpha}a_0$$

For *T* in the femtosecond range, the corresponding wavelengths are in the UV-Vis range. ($c\cdot$(1 fs) ≈ 300 nm.) Atoms and molecules are smaller than the wavelength of the light because they move much slower than light.

Because electrons move slower than light in the atom, magnetic forces are weaker then electric forces (Box 15.4).

16.2 Proof of Eq. 5.14 (the virial theorem applied to the time-averaged energy and potential)

The argument starts from

Eq. 16.11

$$2E_{kin} = 2\sum_i \frac{1}{2} m_i \vec{v}_i \cdot \vec{v}_i = \sum_i \vec{p}_i \cdot \vec{v}_i = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_i \vec{p}_i \cdot \vec{r}_i\right) - \sum_i \vec{r}_i \cdot \frac{\mathrm{d}}{\mathrm{d}t}\vec{p}_i = \frac{\mathrm{d}G}{\mathrm{d}t} - \sum_i\sum_j \vec{r}_i \cdot \vec{F}_{ij}$$

The last term on the right-hand side can be transformed as

Eq. 16.12

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \vec{r}_i \cdot \vec{F}_{ij} = \sum_{j>i}\left(\vec{r}_i \cdot \vec{F}_{ij} + \vec{r}_j \cdot \vec{F}_{ji}\right) = \sum_{j>i}\left(\vec{r}_i - \vec{r}_j\right)\cdot \vec{F}_{ij} = \sum_{j>i} \vec{r}_{ij} \cdot \vec{F}_{ij}$$

$\vec{F}_{ij}$ is the force, which particle *j* exerts on particle *i*. It was used that $\vec{F}_{ij} = -\vec{F}_{ji}$. It follows that

Eq. 16.13

$$2E_{kin} = \frac{\mathrm{d}G}{\mathrm{d}t} - \sum_{j>i} \vec{r}_{ij} \cdot \vec{F}_{ij}$$

G is the sum of the scalar products of the momenta and the positions of all particles:

Eq. 16.14

$$G = \sum_{i=1}^{N} \vec{p}_i \cdot \vec{r}_i$$

The average of d*G/*d*t* over long time periods is zero*:*

Eq. 16.15

$$\left\langle \frac{\mathrm{d}G}{\mathrm{d}t} \right\rangle_{Zeit} = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau \left(\frac{\mathrm{d}G}{\mathrm{d}t} \right) \mathrm{d}t = \lim_{\tau\to\infty} \frac{1}{\tau} \left(G(t) - G(0) \right) = 0$$

It was assumed that the velocities and positions of the particles are bounded. That is realized for periodic orbits, and it is realized for self-gravitating systems with not evaporation. Let the force be the gradient of a potential, *V*: $\vec{F}_{ij} = \mathrm{d}\vec{p}_i/\mathrm{d}t = -\mathrm{d}V/\mathrm{d}\vec{r}_{ij}$. If *V* is proportional to $|\vec{r}_{ij}|^{-1}$, it follows from Eq. 16.13 that

Eq. 16.16

$$2\langle E_{kin,i} \rangle = \left\langle \sum_{j>i} \vec{r}_{ij} \cdot \frac{\mathrm{d}V}{\mathrm{d}\vec{r}_{ij}} \right\rangle = -\langle V_i \rangle$$

The angle brackets denote an average over time (not over the ensemble).

Wikipedia calls this relation (same as Eq. 5.14) “the virial theorem”.

16.3 Thermodynamic variant of the virial theorem, equipartition theorem

Eq. 16.16 is concerned with a time average. It does in involve thermodynamics. There is a related version of the virial theorem, which does involve temperature. Consider a dynamic system consisting of *N* particles with *N* positions, q_i, and *N* momenta, p_i. These 6*N* coordinates span the "phase space". The infinitesimal phase space volume is called dΓ. There is an energy function (the "Hamilton function"), $H(\{q_i,p_i\})$. According to Boltzmann, the probability of finding the system at the coordinates $\{q_i,p_i\}$ is proportional to by $\exp(-H(\{q_i,p_i\}))$.

The basis of the results from this paragraph is the relation:

Eq. 16.17

$$\left\langle q_i \frac{\mathrm{d}H}{\mathrm{d}q_i} \right\rangle = k_B T$$

$<...>_{th}$ denotes the Boltzmann-weighted average. An analogous relation applies to the momenta, p_i.

Ref. 107 calls Eq. 16.17 “the virial theorem". Wikipedia calls Eq. 5.16 the virial theorem. Eq. 5.16 does not involve temperature, while Eq. 16.17 does.

For the proof of Eq. 16.17 we first express the Boltzmann-weighted average as:

Eq. 16.18

$$\left\langle q_i \frac{\mathrm{d}H}{\mathrm{d}q_i} \right\rangle = \frac{\int q_i \frac{\mathrm{d}H}{\mathrm{d}q_i} \exp\left(-\frac{H}{k_B T} \right) \mathrm{d}\Gamma}{\int \exp\left(-\frac{H}{k_B T} \right) \mathrm{d}\Gamma} = \frac{1}{Q} \int q_i \frac{\mathrm{d}H}{\mathrm{d}q_i} \exp\left(-\frac{H}{k_B T} \right) \mathrm{d}\Gamma$$

In the second step, the denominator was called *Q* (the "partition integral"). We use the relation

[107] Schwabl, F., *Statistical Physics*. Springer **2006**, Chapter 2.6.4.1

Eq. 16.19

$$\frac{\mathrm{d}}{\mathrm{d}q_i}\exp\left(-\frac{H}{k_BT}\right)=\exp\left(-\frac{H}{k_BT}\right)\left(-\frac{1}{k_BT}\right)\frac{\mathrm{d}H}{\mathrm{d}q_i}$$

It follows that

Eq. 16.20

$$\frac{\mathrm{d}H}{\mathrm{d}q_i}\exp\left(-\frac{H}{k_BT}\right)=-k_BT\frac{\mathrm{d}}{\mathrm{d}q_i}\exp\left(-\frac{H}{k_BT}\right)$$

Insert Eq. 16.20 into Eq. 16.18 results in

Eq. 16.21

$$\left\langle q_i\frac{\mathrm{d}H}{\mathrm{d}q_i}\right\rangle=\frac{1}{Q}\int q_i\frac{\mathrm{d}}{\mathrm{d}q_i}\exp\left(-\frac{H}{k_BT}\right)\left(-k_BT\right)\mathrm{d}\Gamma$$

We assume that the function exp(–$H(\{q_i,p_i\})$) approaches zero at infinity. This is the case for parabolic potentials. There are no walls. Walls are discussed around Eq. 16.31. If exp(–$H(\{q_i,p_i\})$) is zero at infinity, Eq. 16.21 can partially integrated:

Eq. 16.22

$$\left\langle q_i\frac{\mathrm{d}H}{\mathrm{d}q_i}\right\rangle=\left[\frac{1}{Q}\int x_i\exp\left(-\frac{H}{k_BT}\right)\left(-k_BT\right)\mathrm{d}\Gamma\right]_{-\infty}^{\infty}-\frac{-k_BT}{Q}\int\frac{\mathrm{d}q_i}{\mathrm{d}q_i}\exp\left(-\frac{H}{k_BT}\right)\mathrm{d}\Gamma$$

The square brackets refer to the evaluation on the boundaries of the integration volume (i.e. at infinity). This term vanishes according to the assumption. Using $\mathrm{d}q_i/\mathrm{d}q_i = 1$ leads to

Eq. 16.23

$$\left\langle q_i\frac{\mathrm{d}H}{\mathrm{d}q_i}\right\rangle=k_BT\frac{1}{Q}\int\exp\left(-\frac{H}{k_BT}\right)\mathrm{d}\Gamma=k_BT$$

In the last step it was used that the integral on the right is the partition integral. This proves Eq. 16.17.

Consider a harmonic oscillator with the potential

Eq. 16.24

$$V=\frac{1}{2}\kappa x^2$$

κ is the spring constant. We assume that the oscillator is coupled to a heat bath. In consequence, the probabilities for certain values of x are Boltzmann-distributed. Eq. 16.17 can be applied and one may replace H by V, which results in

Eq. 16.25

$$\left\langle x\frac{\mathrm{d}H}{\mathrm{d}x}\right\rangle=\left\langle\kappa x^2\right\rangle=2\left\langle V\right\rangle=k_BT$$

The mean potential energy of the oscillator in equilibrium therefore is 1/2 k_BT.

The time-averaged kinetic energy of a particle in one dimension is

Eq. 16.26

$$E_{kin,1D} = \frac{1}{2m} p_x^2$$

We replace H with E_{kin} in Eq. 16.17 and arrive at

Eq. 16.27

$$\left\langle p_x \frac{\mathrm{d}E_{kin,1D}}{\mathrm{d}p_x} \right\rangle = \left\langle \frac{1}{m} p_x^2 \right\rangle = 2\left\langle E_{kin,1D} \right\rangle = k_B T$$

and

Eq. 16.28

$$\left\langle E_{kin,1D} \right\rangle = \frac{1}{2} k_B T$$

In three dimensions the factor of 1/2 turns into 3/2:

Eq. 16.29

$$\left\langle E_{kin} \right\rangle = \frac{3}{2} k_B T$$

These arguments can be repeated for all quadratic contributions to the energy function. In particular, they can be repeated for the rotational energy, given as 1/2 $I\Omega^2$ with I the moment of inertia and Ω the angular velocity. The mean energy attached to such quadratic contributions is 1/2 k_BT. (It is 3/2 k_BT for the kinetic energy because of the three dimensions of space.)

If the Hamiltonian only contains quadratic contributions, the number of these contributions is called f and the contributions are referred to as "degrees of freedom". The CO_2 molecule has 13 such degrees of freedom (f_{trans} = 3, f_{rot} = 2, f_{vib} = 8 for the kinetic and the potential energy contained the 4 modes of vibration). There are only 2 degrees of freedom for rotation because rotation around the molecular axis is an electronic excitation that is frozen at room temperature. (The vibrational degrees of freedom are partially frozen, as well.)

If quantization is ignored, the mean thermal energy of a system with f degrees of freedom is given by f/2 k_BT. That proves some relations from chemical thermodynamics, namely

Eq. 16.30

$$\bar{U} = \frac{f}{2} RT, \quad c_V = \frac{\mathrm{d}\bar{U}}{\mathrm{d}T} = \frac{f}{2} R$$

$\bar{U}$ is the internal energy per mol and c_V is specific heat capacity at constant volume. Again: A degree of freedom is a quadratic contribution to the energy function. If the potential is not a parabolic potential, the equipartition theorem does not apply. For example, the equipartition theorem does not apply to the potential energy of charges in the diffuse double layer above charged surfaces.

The potential may be approximately harmonic. An example is the 6-12 potential close to its minimum. In the condensed phases, the interparticle distances match the minimum of the potential. The potential can be Taylor-expanded around the minimum, resulting in a parabola. The equipartition theorem applies, which leads to the rule of Dulong and Petit. In a real gas, the interparticle distance are not at the minimum of the 6-12-potential. For real gas, the specific heat is mostly contained in the kinetic energy. There is a small contribution from the interaction[108] and this contribution does not obey the equipartition theorem.

Box 16.1

Does the ideal gas law apply in the center of the sun?

Atkins' textbook reports that the ideal gas law holds in the center of the sun, although the density is ρ = 150 g/cm^3, which is greater than the density of condensed matter on earth. Atkins argues that the atoms are completely ionized and that the collision cross section was governed by the size of the nucleus rather than the size of the corresponding atom. The kinetic gas theory assumes that particles move with constant velocity on straight lines and that they rarely collide. The distance between two collisions is the mean free path, λ. λ is the key parameter in the kinetic theory of gases.

This argument is problematic insofar as the particles primarily interact electrostatically. The nuclear forces do play a certain role (after all there is nuclear fusion), but they do not govern the pressure-volume relation. The collision rate is not govern by the size of the nucleus. In a plasma, electrons almost never fly on straight lines because the Coulomb interaction ($\propto 1/r$) prevents that due to its long range. (Applying the standard equation for the calculation of the collision cross section to the $1/r$-potential leads to a collision cross section, which is infinite). The plasma does not obey kinetic gas theory.

Why does the ideal gas law still hold? This has to do with a variant of the virial theorem, reported in Eq. 16.31. Eq. 16.31 only assumes thermodynamic equilibrium. It has nothing to do with kinetic gas theory. Eq. 16.31 contains the "inner virial" on the right-hand side. The inner virial gives rise to the difference between ideal and real gases. If the mean energy in the interactions is smaller than the mean kinetic energy, then the correction to the ideal gas law according to Eq. 16.31 is small. This is the case for real gases and this is the case for plasma in the center of the sun. The ideal gas law applies because the kinetic energy dominates.

On earth, a gas usually is confined by walls. The forces exerted by the walls must be taken into account. These forces generate to the pressure exerted by the gas onto the wall. A somewhat lengthy calculation results in[107]

Eq. 16.31

$$pV = \frac{2}{3}\langle E_{kin}\rangle - \frac{1}{3}\left\langle \sum_{m>n} |x_{mn}| \frac{\mathrm{d}V(|x_{mn}|)}{\mathrm{d}|x_{mn}|}\right\rangle$$

For gases, the first term on the right-hand side is Nk_BT. The second term on the right-hand side is the "inner virial". (m must be larger than n to avoid interactions between particles being counted twice.) For non-interacting systems the inner virial is zero. Eq. 16.31 then reproduces the ideal gas law. For real gases, the inner virial is nonzero. Evaluating the virial in detail is anything but trivial. As part of the "virial expansion", the virial is Taylor-expanded in the number density, N/V, which leads to the "virial coefficients". The second virial coefficient is of particular importance.

Remember:

– For systems in thermodynamic equilibrium, the relations $\langle q_i \, \mathrm{d}H/\mathrm{d}q_j\rangle_{th} = k_B T\, \delta_{ij}$ and $\langle p_i \, \mathrm{d}H/\mathrm{d}p_j\rangle_{th} = k_B T \delta_{ij}$ hold. q_i and p_i are the positions and the momenta.

---

[108] The specific heat capacity of the noble gases at constant pressure, $c_p = c_V + R$ is slightly larger than 5/2 R. The heavier the atom, the larger the polarizability, the larger the van-der-Waals attraction, and the stronger the deviation of the specific heat capacity from 5/2 R.

- The equipartition theorem follows from the relation above for all quadratic contributions to the energy function.
- Eq. 16.31 explains how the inner virial affects the volume-pressure relation of real gases.

16.4 Proof of Eq. 12.1 (partition sum and free energy)

Start from the definition of the Helmholtz free energy ($A = U - TS$). U is the sum of all energy levels, weighted by occupancy. The entropy, S, in this case is not $k_B \ln\Omega_s$ because the ensemble is not closed. It is in a thermal equilibrium with an environment. The different microstates are not equally probable, but rather are Boltzmann-distributed. The definition of the entropy is adapted suitably by replacing Ω_s with $1/p_i$, where p_i is the probability of microstate i. If all microstates are equally probable, $p_i = 1/\Omega_s$. Otherwise, p_i is replaced by its average:

Eq. 16.32

$$S = k_B \ln \frac{1}{\langle p_i \rangle} = -k_B \ln\langle p_i \rangle = -k_B \sum p_i \ln p_i$$

Using the definitions of U and S, the free energy turns into

Eq. 16.33

$$A = U - TS = \langle \varepsilon_i \rangle - T\left(-k_B \langle \ln p_i \rangle\right) = \sum p_i \left(\varepsilon_i + k_B T \ln p_i\right)$$

The probabilities are given by the Boltzmann distribution:

Eq. 16.34

$$p_i = \frac{\exp\left(-\frac{\varepsilon_i}{k_B T}\right)}{\sum_i \exp\left(-\frac{\varepsilon_i}{k_B T}\right)} = \frac{\exp\left(-\frac{\varepsilon_i}{k_B T}\right)}{Q} \quad \text{with} \quad Q = \sum_i \exp\left(-\frac{\varepsilon_i}{k_B T}\right)$$

The algebra proceeds as

Eq. 16.35

$$A = \sum p_i \left(\varepsilon_i + k_B T \ln\left(\frac{\exp\left(-\frac{\varepsilon_i}{k_B T}\right)}{Q}\right)\right) = \sum p_i \left(\varepsilon_i + k_B T\left(-\frac{\varepsilon_i}{k_B T} - \ln Q\right)\right)$$
$$= k_B T\left(-\ln Q\right)\sum p_i$$
$$= -k_B T \ln Q$$

16.5 Topics from physical chemistry touched upon in these notes

The list also contains brief mentions of some topics, which are not explained in any detail

- Black-body radiation, the Rayleigh-Jeans law, and its roots in the equipartition theorem (section 3.2)
- Deterministic chaos (section 4.1)
- Debye Hückel theory as an example for a mean-field theory (section 4.2).
- Hartree-Fock method as an example for a mean-field theory (section 4.2).
- Fermi resonances, combination tones (Box 4.1)
- Correlations in quantum chemistry, density functional theory, DFT (section 4.4)
- The definition of entropy as the log of the number of microstates (section 5.1)
- The Boltzmann factor (Box 5.2)
- Temperature, thermal equilibrium, and the stability thereof (section 5.2)
- The Debye law predicting the specific heat capacity of dielectrics (below Eq. 5.28)
- Arrows of time (section 5.12)
- Coherent and incoherent detectors, the pair correlation function (section 6.4)
- NIR spectroscopy (Box 6.1)
- The uncertainty relation (section 8.1)
- Electrons in metals are in a degenerate state, forming bands (section 8.1)
- The exchange interaction (section 8.1)
- The adiabatic index (section 8.2).
- Gradual transition from the neutral gas to the plasma (footnote 63)
- The partition sum and its relation to the free energy (section 12.3)
- Lifetime of excited states, fluorescence lifetime (section 10.2)
- Selection rules, magnetic transitions (section 11.1)
- Energy levels of the rigid linear rotator, rotation spectra (section 11.2)
- Adiabatic expansion and adiabatic cooling (Fig. 11.5)
- Stimulated emission (section 11.4)
- The Kelvin radius (footnote 82)
- Reaction states and energy eigenstates of the ammonia molecule (section 14.1).
- Spontaneous symmetry breaking (Fig. 15.3)
- Magnetic effects in spectroscopy, fine structure in spectroscopy (Box 15.4)
- Spin-orbit coupling (Box 15.3)
- The fine structure constant (section 16.1)
- The equipartition theorem (section 16.3)
- The virial expansion (text below Eq. 16.31)
- The kinetic gas theory (Box 16.1)

16.6 Proof of footnote 54

We start from the pressure at constant entropy:

$$-p = \left(\frac{\mathrm{d}U}{\mathrm{d}V}\right)_S = \left(\frac{\mathrm{d}U}{\mathrm{d}V}\right)_T + \left(\frac{\mathrm{d}U}{\mathrm{d}T}\right)_V \left(\frac{\mathrm{d}T}{\mathrm{d}V}\right)_S = 0 + \frac{3}{2} N k_B \left(\frac{\mathrm{d}T}{\mathrm{d}V}\right)_S$$

The equipartion theorem was used in step 3. Following Sackur and Tetrode, the entropy of the ideal gas follows the relation

$$\frac{S}{k_B N} = \ln\left(\frac{V}{N\Lambda^3}\right) + \frac{5}{2}$$

We exploit constant entropy:

$$\mathrm{d}S = 0$$

$$\mathrm{d}\ln V - 3\,\mathrm{d}\ln\Lambda = 0$$

$$\mathrm{d}\ln V + \frac{3}{2}\mathrm{d}\ln T = 0$$

$$\frac{\mathrm{d}V}{V} + \frac{3}{2}\frac{\mathrm{d}T}{T} = 0$$

$$\left(\frac{\mathrm{d}T}{\mathrm{d}V}\right)_S = -\frac{2}{3}\frac{T}{V}$$

Next, we need to express the temperature as a function of entropy and volume:

$$\exp\left(\frac{S}{k_B N} - \frac{5}{2}\right) = \frac{V\left(2\pi m k_B T\right)^{3/2}}{N h^3}$$

$$T = \left(\frac{N h^3 \exp\left(\frac{S}{k_B N} - \frac{5}{2}\right)}{V}\right)^{2/3} \frac{1}{2\pi m k_B} = \left(\frac{N}{V}\right)^{2/3} \frac{h^2}{2\pi m k_B} \exp\left(\frac{2}{3}\frac{S}{k_B N} - \frac{5}{3}\right)$$

This result can be inserted into the first equation

$$p = 0 + \frac{3}{2} N k_B \left(\frac{\mathrm{d}T}{\mathrm{d}V}\right)_S = -\frac{3}{2} N k_B \left(\frac{\mathrm{d}T}{\mathrm{d}V}\right)_S = \frac{h^2}{2\pi m} \exp\left(\frac{2}{3}\frac{S}{k_B N} - \frac{5}{3}\right)\left(\frac{N}{V}\right)^{5/3}$$

That proves footnote 54.