
These informal notes give some background details about the energy balance in the solar photosphere, chromosphere, and corona. These notes do not describe the solar wind or heliosphere; for that, see the other papers included in the web directory where this document is found.

1 Photosphere

Though already probably covered in detail in the lectures on Radiative Transfer, Helioseismology, and Solar Magnetic Fields, there are a few aspects of the photosphere that I’d like to emphasize in order to contrast it with the overlying layers to be discussed.

1. The temperature in the photosphere is essentially determined by a balance between radiative heating (due to matter absorbing photons) and radiative cooling (due to matter emitting photons). In discussions of the condition of “radiative equilibrium” you have seen that $J$ (the mean intensity field) is equal to $S$ (the source function) under this condition. The net rate of heating (i.e., heating minus cooling) is proportional to $J - S$, with each quantity integrated over the frequency-dependent opacity. Thus, the atmosphere comes into equilibrium when the two quantities are equal. The frequency-integrated (gray) versions of these two quantities are given in the Eddington approximation by

$$J(\tau) = \frac{3}{4\pi} \sigma_R T_{\text{eff}}^4 \left( \tau + \frac{2}{3} \right)$$

and by equating the two, one obtains the traditional expression for $T(\tau)$. The photosphere is defined as the height where $\tau = \frac{2}{3}$ and $T = T_{\text{eff}} \approx 5770$ K.

The photosphere essentially “ends” where other heating mechanisms start to dominate over radiative heating.

2. The photosphere is the “solar surface” where most of the magnetic field variability, pulsations, waves, etc., are observed in the greatest detail. Much of our knowledge of, say, the amplitudes of waves and the strength of the magnetic field at higher levels of the atmosphere comes from extrapolation from the well-observed photosphere.

2 Chromosphere

About 500 km above thephotosphere, the temperature reaches a minimum value of about 4000 to 4500 K, and then starts to increase again. Over the next 1500 km or so (i.e., only 0.002 solar radii), $T$ increases slowly back up above $T_{\text{eff}}$ and further to values around 10,000–20,000 K. Somewhere around 2000 km above the photosphere, the temperature shoots up extremely rapidly to $\sim 1$ million K. This “transition region” (TR) is extremely narrow: only a few hundred km. The region between the 500 km temperature-minimum and the 2000 km TR is defined as the chromosphere. Histori-
cally, this “color layer” was observed during total eclipses as a thin red ring around the Sun, or in images of the Sun taken in the narrow wavelength band around the Hα spectral line of hydrogen.

From a physical point of view, the chromosphere is defined as the region where radiative cooling is balanced by some kind of “mechanical” heating—i.e., either kinetic energy of plasma motions or magnetic energy stored in flux tubes—that becomes dissipated (i.e., transferred to heat, like friction).

The community is still divided as to the dominant form of the heating. Understanding the controversy requires a digression into how the solar convection manifests itself at the solar surface.

Even though the photosphere is above the region of true convective instability, the deeper turbulent motions still make their presence known. The primary observational manifestation is the granulation pattern of hot upwelling 1000-km sized cells, surrounded by narrower, cool downflowing lanes. Superimposed on the granulation are other kinds of oscillations that are surface manifestation of interior pulsations. These oscillations are “trapped” acoustic waves that travel up from the convection zone and bounce back downward because the steep gradient in \( T(\tau) \) acts as a “wall.” These motions will be covered in the Helioseismology lectures. Of interest here is a third type of variability: propagating waves that are able to tunnel past the wall and travel up above the photosphere. These waves have higher frequencies (and smaller wavelengths) than the trapped modes, and thus—to them—the wall is not so sharp. These waves are proposed to propagate up into the chromosphere (and some up into the corona; see below) and damp out their energy when they cease to behave as simple linear oscillations.

The controversy involves whether the damping in the chromosphere is dominated by: (1) simple acoustic, or sound waves, that propagate vertically, (2) waves similar in character to acoustic waves, but that are guided or channeled along magnetic flux tubes, or (3) magnetohydrodynamic (MHD) waves that have different propagation characteristics from (1) or (2). I will summarize the traditional idea (1).

Consider a plane-parallel 1D atmosphere, which has a roughly constant temperature \( T \). In an isothermal atmosphere, the pressure \( P \) and density \( \rho \) are linked by an equation of state

\[
P = \rho k_B T / m
\]

where \( k_B \) is Boltzmann’s constant and \( m \) is the mean mass of a gas particle (typically close to \( m_{\text{H}} \)).

The equation of hydrostatic equilibrium,

\[
\frac{\partial P}{\partial z} = -\rho g
\]

can then be integrated to find

\[
\rho(z) = \rho_0 \exp(-z/H) ,
\]

where \( H = k_B T/(mg) \) is a constant density scale height.

Acoustic waves that are launched at the photosphere \((z = 0)\) carry equal amounts of energy in kinetic motions (upward and downward oscillations) and density fluctuations (compressions and rarefactions), such that the energy densities are equal:

\[
\frac{1}{2} \rho (\delta v_j)^2 = \frac{1}{2} \frac{(\delta \rho)^2}{\rho} \rho_s^2
\]
where the sound speed $c_s = \sqrt{\gamma P/\rho}$ and $\gamma = 5/3$. For an isothermal atmosphere in the absence of damping, the energy density of an acoustic wave is conserved. Thus, with $\rho(\delta v)\|^2$ being constant and $\rho$ decreasing exponentially, the velocity amplitude $\delta v$ wants to increase exponentially with height. But this cannot happen indefinitely. When $\delta v$ starts to exceed the sound speed $c_s$, the wave becomes nonlinear and it begins to “break” (i.e., the faster crests overtake the slower troughs). A sinusoidal wave train steepens into a sawtooth-shaped train of shocks.

The full analysis of how shocks heat an atmosphere is beyond the scope of this lecture. Shock waves involve an irreversible gain in entropy from one side to the other, and thus they dissipate as they propagate. The heating rate (in units of power per unit volume) is

$$Q_{\text{shock}} = \frac{\rho T \Delta S}{2\pi/\omega}$$

(7)

where $\Delta S$ is the entropy gain at one shock, and the denominator takes into account that there is a periodic train with frequency $\omega$. In other words, the shocks’ entropy is “dumped in” once per period, and the heating rate is an average. The gain in internal energy at a shock is the product $T \Delta S$, and this is a function of the wave/shock amplitude $\delta v$. In a plane-parallel atmosphere, though, shock dissipations tends to lead to a simple kind of saturation: rather than having $\delta v$ increase exponentially, it reaches a limiting constant value as a result of shock damping. Thus, for our purposes, we can forget about the height dependence of the entropy term and just consider $Q_{\text{shock}} \approx C \rho$, with a constant value of $C$.

The acoustic/shock heating is balanced by radiative cooling. In the optically thick photosphere, this was proportional to the emission of photons in radiative equilibrium (i.e., the source function $S$). In the optically thin upper atmosphere, though, the densities are low enough that the emission due to lines and continuum processes can be “counted up” much more simply without worrying about the detailed geometry or opaqueness of the radiation field. In other words, each emitted photon is allowed to escape. In the upper atmosphere, atomic excitations and ionizations are dominated by collisions between the atoms/ions and free electrons. These excitations and ionizations are followed by de-excitations and recombinations that emit photons, thus removing energy from the system and cooling the gas. The collision rate is proportional to the product of the densities of the “projectiles” (electrons) and the “targets” (atoms/ions), and we can write the cooling rate as

$$Q_{\text{rad}} = n_e n_H \Lambda(T)$$

(8)

where we use number densities $n = \rho/m$ rather than mass densities, and $n_e$ and $n_H$ are the number densities of free electrons and hydrogen. The temperature-dependent function $\Lambda(T)$ essentially contains the product of three tabulated quantities: (1) the abundances of each element, which are normalized to hydrogen, (2) the relative stages of ionization for each element, which are mainly a function of temperature, and (3) the rates of continuum and line cooling for each ionization stage of each element. The figure below (from Owocki 2003) illustrates this function.

The chromospheric temperature is determined by the balance between acoustic/shock heating and radiative cooling. Very approximately, this demands

$$C \rho \approx \rho^2 \Lambda$$

(9)

(where the conversion between number density and mass density has been absorbed into $C$), and thus $\Lambda \propto \rho^{-1}$. The radiative cooling function needs to increase exponentially with height to keep
the chromosphere in equilibrium:

\[ \Lambda(T[z]) \propto e^{z/H} \quad (10) \]

As illustrated in the figure, for any finite amount of heating this required increase will lead to a steadily higher temperature, from below 10^4 K up to about 10^5 K. Upon reaching the local maximum in the cooling curve at 10^5 K, though, a radiative balance can no longer be maintained without a drastic jump to a much higher temperature. This “thermal instability” happens at a critical density of \( \rho \approx C/\Lambda_{\text{max}} \), which defines the beginning of the TR between the chromosphere and the corona.

3 Corona

If radiation were the only cooling mechanism, the “new” equilibrium temperature (above the TR) would greatly exceed 10^7 K (off the scale on the right of the figure!). This does not happen, though, because another cooling mechanism kicks in: heat conduction. We will see below how conduction tends to diffuse out the energy deposited in the corona to prevent the huge swings in temperature that radiation alone would have dictated.

First, though, we need to consider the heating of the coronal plasma. This is a hugely active field of research, and the “coronal heating problem” has been with us since the 1930s. It is very likely that different mechanisms apply to the bright, compact “active regions” seen mainly in X-rays and to the dimmer open-field regions that feed the solar wind. This discussion will focus on the latter, and also on waves in a similar vein as in the previous section. We should note, though, that the acoustic waves/shocks that were so efficient at heating the chromosphere are essentially fully damped away by the time \( T \) starts to exceed 10^6 K. Thus, we want to focus on other kinds of waves that can remain propagating up into the corona. This is where the strong magnetic field of the Sun comes into play.

As illustrated in the lecture presentation, the photospheric granulation (upwelling cells surrounded by downdrafting lanes) contains a third, tiny-scale component: thin flux tubes in the
dark lanes that contain extremely strong magnetic fields ($B \approx 1500$ G, in comparison to Sun-averaged values of about 10 G). These flux tubes extend up vertically into the outer atmosphere and into interplanetary space. They are shaken more-or-less transversely by the convective granulation motions, and these perpendicular perturbations in both velocity ($\delta v_\perp$) and the magnetic field ($\delta B_\perp$) propagate up the flux tubes as so-called Alfvén waves. (The closest real-world analogy is a transverse wave on a “plucked” wire; the wire’s tension is the restoring force in a similar way as “magnetic tension” is the restoring force for Alfvén waves.) Because Alfvén waves do not contain velocity fluctuations in the vertical direction, they cannot steepen in the same way as acoustic waves. Thus they don’t form shocks, and they survive up into the corona to be damped over much larger length scales.

What is the heating rate associated with Alfvén wave damping? This is a current topic of active research, and there are no less than a dozen different ideas floating around. My own favorite ideas revolve around a turbulent cascade from large eddies, to small eddies, and further to much smaller scales where the damping happens by collisionless plasma-physics processes such as Landau damping and ion-cyclotron resonance. Note that for Alfvén waves the concept of an “eddy” is different than that in fluid hydrodynamics: the magnetic field constrains the free “swirling” motions that are often seen in fluid turbulence.

The rate of volumetric energy dissipation (in units of power transferred to heat per unit volume) due to turbulence can be pinned down rather remarkably by dimensional analysis. This was in fact how Kolmogorov did it in 1941. If all one knows about a turbulent fluid is its density $\rho$, the length-scale $\ell$ of the largest eddies (i.e., the scale on which one “stirs up” the fluid), and the velocity amplitude $v$ of the eddy motions on that scale, there is only one combination of these quantities that yields a volumetric energy dissipation rate:

$$Q \sim \frac{\rho v^3}{\ell}$$

and, to within an order-unity correction factor, this is the energy transfer rate that is seen in many different forms of astronomical, geophysical, and laboratory turbulence!

Current models of turbulent heating in the solar corona involve replacing $v$ with the velocity amplitude of Alfvén waves and $\ell$ with the cross-sectional diameter of the flux tubes along which the waves propagate. There are also other correction factors that multiply $Q$ above to take account of the ‘efficiency’ of the cascade process.

Putting aside the exact mechanism(s), let us just presume that a given amount of heat is being deposited into the corona. At a height in the mid-corona of $r = 1.5 R_\odot$, let us estimate $\rho \approx 10^{-15}$ g cm$^{-3}$, $\delta v_\perp \approx 10$ km/s, and $\ell \approx 10^9$ cm. This gives a heating rate $Q_h$ of $10^{-6}$ erg s$^{-1}$ cm$^{-3}$. Note that this value is appropriate for the first one or two solar radii of the corona, but it drops off steeply at larger heights.

Thus, how is the coronal temperature determined? There are two regions where the energy balance is reasonably straightforward:

1. In the “low corona” (i.e., the first few tenths of a solar radius above the TR), the heat flux deposited in the corona is conducted downwards. The classical conduction flux due to ion collisions is a vector $F_c$, and its magnitude can be written

$$F_c = -KT^{5/2} \frac{dT}{dr}$$

(12)
(where $K \approx 6 \times 10^{-7}$ erg s$^{-1}$ cm$^{-1}$ K$^{-7/2}$). The local heat deposition rate due to conduction is $|\nabla \cdot F_c|$. In the low corona, which is still close to plane-parallel in its geometry, we can equate the roughly constant heating rate $Q_h$ to $|dF_c/dr|$ and integrate twice to obtain the temperature,

$$T(r) = \left[ T_0^{7/2} + \frac{7Q_h}{4K}(r - R_\odot)^2 \right]^{2/7}.$$

As $(r - R_\odot) \to 0$, the slope of $T(r)$ approaches infinity, which agrees qualitatively with the known steepness of the thin TR layer. This function prescribes an ever-increasing coronal temperature, but since in actuality $Q_h$ “cuts off” above a few solar radii, this relation is valid only in the low corona. In reality, $T(r)$ reaches a maximum value in the middle corona $(r \approx 1.5$ to $3 R_\odot)$ then declines at larger distances. For a height of $r = 1.5 R_\odot$, the numbers given above yield a realistic temperature of about $1.4 \times 10^6$ K. This is substantially larger than the chromospheric temperature ($T_\odot \approx 10^4$ K) but it is much lower than the values in excess of $10^7$–$10^8$ K that pure radiative cooling would have demanded.

2. In the outer part of the corona (above the heights where the energy flux $Q_h$ is deposited) conduction “acts alone” to draw out the heat to larger distances than it would have propagated by itself. At these large radii, the spherical geometry of the corona also needs to be considered when writing the divergence of $F_c$. Thus, for a conduction-dominated energy equilibrium in a spherically symmetric corona, the divergence of the conduction flux approaches zero:

$$\nabla \cdot F_c = \frac{1}{r^2} \frac{d}{dr} \left( r^2 K T^{5/2} \frac{dT}{dr} \right) = 0.$$

Upon double integration, this gives a temperature that declines slowly outward from its coronal maximum, as $T \propto r^{-2/7}$.

### 4 Solar Wind

The hot corona provides enough gas pressure to counteract the Sun’s gravity and accelerate a steady flow of gas into interplanetary space. The attached article “New views of the solar wind with the Lambert W function” gives an overview of how the solar wind equations are derived and solved.
5 References

Books:

Recommended online lecture notes:

Other solar web pages:
http://nineplanets.org/sol.html
http://sohowww.nascom.nasa.gov/explore/
http://solar-center.stanford.edu/
http://www.windows.ucar.edu/
http://www.cfa.harvard.edu/~scranmer/ (your humble lecturer’s page)