8 Radiative Cooling and Heating

Reading:
Katz et al. 1996, ApJ Supp, 105, 19, section 3
Thoul & Weinberg, 1995, ApJ, 442, 480
Optional reading:
Thoul & Weinberg, 1996, ApJ, 465, 608
Weinberg et al., 1997, ApJ, 477, 8
The latter two address the influence of photoionization on galaxy formation and the interplay between physics and numerical resolution in cosmological simulations.

The thermal energy equation is

$$\frac{\mathrm{D}\epsilon}{\mathrm{D}t} = -\frac{P}{\rho}\vec{\nabla}\cdot\vec{u} - \frac{1}{\rho}\vec{\nabla}\cdot\vec{F} + \frac{1}{\rho}\Psi + \frac{\Gamma - \Lambda}{\rho}.$$

We now turn our attention to the last term, in which

$$\begin{aligned} \Gamma &\equiv \text{volumetric radiative heating rate} \quad [\mathrm{erg\,cm^{-3}\,s^{-1}}] \\ \Lambda &\equiv \text{volumetric radiative cooling rate} \quad [\mathrm{erg\,cm^{-3}\,s^{-1}}]. \end{aligned}$$

Unfortunately, due to a shortage of Greek letters, Γ is also used to represent ionization rates. In this section of the course and notes, therefore, I will denote the radiative heating rate by $\mathcal{H}$, and will reserve Γ for ionization rates.

I will focus on cooling and heating of a primordial composition, H/He plasma because (a) it's simple enough to treat fairly comprehensively, (b) you'll encounter other cases in the ISM class, and (c) it's the case I know the most about.

8.1 Cooling processes in a primordial plasma

The cooling processes that affect a H/He plasma are:

Collisional excitation: free electron impact knocks a bound electron to an excited state; it decays, emitting a photon.

Collisional ionization: free electron impact ionizes a formerly bound electron, taking energy from the free electron.

Recombination: free electron recombines with an ion; the binding energy and the free electron's kinetic energy are radiated away (only the latter counts as a "loss" here — the binding energy was "charged" to the collisional ionization process).

Free-free emission: free electron is accelerated by an ion, emitting a photon. (A.k.a. Bremsstrahlung.)

All of these processes are proportional to a function of temperature (different for each process) times the electron number density times the number density of the relevant ionic species.

At very high densities, formation of H_2 molecules is an additional source of cooling.

8.2 Ionic abundances and ionization equilibrium

In order to calculate cooling rates using, say, the formulas for the above processes given in Table 1 of Katz et al. (1996), one needs to know the density of the various ionic species: $n_{\rm e}$, $n_{\rm H_0}$, $n_{\rm H_+}$, $n_{\rm He_0}$, $n_{\rm He_+}$, $n_{\rm He_{++}}$.

At fixed total gas density, the evolution of these densities is governed by equations like

$$\frac{{\rm d}n_{\rm H_0}}{{\rm d}t} = \alpha_{\rm H_+}(T) n_{\rm H_+} n_{\rm e} - \Gamma_{\rm eH_0}(T) n_{\rm e} n_{\rm H_0} - \Gamma_{\gamma {\rm H_0}} n_{\rm H_0}, \tag{59}$$

where

$$\alpha_{\rm H_+}(T) = \text{hydrogen recombination coefficient } [\,{\rm cm^3\,s^{-1}}] \tag{60}$$

$$\Gamma_{\rm eH_0}(T) = \text{collisional ionization rate } [\,{\rm cm^3\,s^{-1}}] \tag{61}$$

$$\Gamma_{\gamma {\rm H_0}} \equiv \int_{\nu_T}^{\infty} \frac{4\pi J(\nu)}{h\nu} \sigma(\nu) d\nu = \text{photoionization rate } [\,{\rm s^{-1}}], \tag{62}$$

with

$$\begin{aligned}
\nu_T &= \text{ionization threshold frequency (e.g., 13.6 eV}/h \text{ for H}^0) \\
\sigma(\nu) &= \text{ionization cross-section } [\,{\rm cm^{-2}}] \\
J(\nu) &= \text{radiation background intensity } [{\rm erg\,s^{-1}\,cm^{-2}\,sr^{-1}\,Hz^{-1}}].
\end{aligned}$$

Suppose that the medium is $\sim$ 50% ionized ($n_{\rm H_0} \sim n_{\rm H_+} \sim n_{\rm e} \sim n$) and the right hand side equation (59) is far from balance.
The ionic density will evolve on a timescale

$$t \sim \left| n \left(\frac{{\rm d}n}{{\rm d}t} \right)^{-1} \right| \sim \left| n_{\rm e}(\alpha_{\rm H_+}(T) - \Gamma_{\rm eH_0}(T)) - \Gamma_{\gamma {\rm H_0}} \right|^{-1}.$$

For a typical estimate of the radiation background at $z \sim 2$, $J(\nu) \sim$ few$\times 10^{-22}{\rm erg\,s^{-1}\,cm^{-2}\,sr^{-1}\,Hz^{-1}}$,

$$\frac{1}{\Gamma_{\gamma {\rm H_0}}} \sim 10^{12}\,{\rm s} \sim 3 \times 10^4 \ {\rm yr},$$

while for $T \sim 10^5\,{\rm K}$

$$\left| n_{\rm e}(\alpha_{\rm H_+}(T) - \Gamma_{\rm eH_0}(T)) \right|^{-1} \approx \frac{1}{n_{\rm e} \Gamma_{\rm eH_0}(T)} \sim 10^6 \left(\frac{n_{\rm e}}{10^{-5}\,{\rm cm^{-3}}} \right)^{-1} \ {\rm yr}.$$

If photoionization ($\Gamma_{\gamma {\rm H_0}}$) is significant, the timescale is almost always short compared to the dynamical timescale of cosmological systems, and even without photoionization it is usually still shorter than the dynamical timescale.

In many circumstances, therefore, one can safely assume that creation and destruction rates will be balanced for each species, the condition known as ionization equilibrium. In this case, both sides of equations like (59) are equal to zero.
In ionization equilibrium, the abundances of ionized and neutral hydrogen are determined by *algebraic* equations (much easier than differential equations),

$$\begin{aligned}
&\Gamma_{\mathrm{eH_0}} n_{\mathrm{e}} n_{\mathrm{H_0}} + \Gamma_{\gamma \mathrm{H_0}} n_{\mathrm{H_0}} = \alpha_{\mathrm{H_+}} n_{\mathrm{e}} n_{\mathrm{H_+}}, \\
&n_{\mathrm{H_+}} + n_{\mathrm{H_0}} = n_{\mathrm{H}} \\
&n_{\mathrm{e}} = n_{\mathrm{H_+}} + n_{\mathrm{He_+}} + 2n_{\mathrm{He_{++}}}
\end{aligned}$$

together with similar equations

$$\text{destruction rate} = \text{creation rate}$$

for $n_{\mathrm{He_0}}$, $n_{\mathrm{He_+}}$, $n_{\mathrm{He_{++}}}$ (Katz et al. eqs. 25-28).

In a numerical hydrodynamics code, one must decide whether to integrate the ionic abundance equations like (59) or compute abundances assuming ionization equilibrium.
The appropriate choice depends on the physical situation. In some circumstances, departures from ionization equilibrium are physically important. However, one must be sure to integrate the time-dependent equations on a timscale short compared to the abundance evolution timescale, which can be very short compared to other timescales of interest.
When the ionization equilibrium assumption is appropriate, it is much more efficient to use the equilibrium equations.

8.3 The collisional equilibrium cooling curve

The above equilibrium equations can be recast into the form

$$\frac{n_{\mathrm{H_0}}}{n_{\mathrm{H}}} = \frac{\alpha_{\mathrm{H_+}}}{\alpha_{\mathrm{H_+}} + \Gamma_{\mathrm{eH_0}} + \Gamma_{\gamma \mathrm{H_0}}/n_e},$$

with analogous equations for the fractions of neutral helium and singly ionized helium.
If there is no photoionization, $\Gamma_{\gamma \mathrm{H_0}} = 0$, then, since $\Gamma_{\mathrm{eH_0}}$ and $\alpha_{\mathrm{H_+}}$ depend only on temperature (not density), the *relative* abundances of ionic species depend only on temperature. All cooling processes are therefore a function of temperature times the square of the total gas density ρ (or, equivalently, the total hydrogen number density $n_{\mathrm{H}} = X\rho/m_p$, where $X \approx 0.76$ is the hydrogen mass fraction).

Since ionic abundances are determined by equilibrium of collisional processes, this situation is called collisional equilibrium (a.k.a. coronal equilibrium). Cooling rates in collisional equilibrium are fully described by the function $\Lambda(T)/n_{\mathrm{H}}^2$, known as the cooling curve (e.g., Katz et al. Figure 1).

At high temperatures the gas is fully ionized, so the only cooling process is free-free emission, with

$$\frac{\Lambda}{n_{\mathrm{H}}^2} \sim 2.5 \times 10^{-23} \left(\frac{T}{10^8 \ \mathrm{K}}\right)^{1/2}.$$

The dominant processes at low temperatures are collisional line excitation of hydrogen ($T \sim 10^{4.2}$ K) and collisional line excitation of He^+($T \sim 10^5$ K).

Below 10^4 K, collisions are not energetic enough to ionize atoms or even excite them out of the ground state, so the cooling rate drops to zero.

The timescale on which gas can radiate away its thermal energy is

$$t_{\rm cool} \sim \left| \epsilon \left(\frac{\mathrm{D}\epsilon}{\mathrm{D}t} \right)^{-1} \right| \sim \left| \frac{\epsilon \rho}{\Lambda} \right| \propto \epsilon \rho^{-1}, \tag{63}$$

since $\Lambda \propto \rho^2$.
If the gas is confined by gravity or by ambient pressure, loss of thermal energy is usually accompanied by an increase in density (to restore the pressure gradient or equilibrium with the ambient medium).
The increase in density usually decreases the cooling timescale, so cooling tends to be a runaway process.
For a primordial composition cooling curve, gas that *starts* to cool (i.e., has $t_{\rm cool}$ less than the age of the system) usually cools fairly rapidly to 10^4 K, at which point it can cool no further (at least by atomic processes).

8.4 Cooling and galaxy formation

In an expanding universe, the gravitational collapse of a homogeneous spherical perturbation produces a virialized object whose average density is roughly 200 times the background critical density at the time of collapse.
For a given virial mass, one can use this density to compute a characteristic radius R, velocity dispersion $\sigma^2 \sim GM/R$, and virial temperature $T \sim GMm_p/(kR)$, given a collapse redshift (which determines the cosmic background density).
A plausible assumption is that any gas that participates in this collapse is heated to this virial temperature by shocks, and one can then calculate a cooling time from the temperature and density.

Some of the early analytic papers on galaxy formation (Binney 1977, ApJ, 215, 483; Silk 1977, ApJ, 211, 638; Rees & Ostriker 1977, MNRAS, 179, 541) argued that the characteristic mass scale of galaxies was determined by the requirement that this cooling time be shorter than the dynamical time (or, perhaps, than the age of the universe).

More massive objects have higher virial temperatures and tend to collapse later (at lower density), so they tend to have longer cooling times, and the gas in them would therefore be unable to cool and form stars. The cooling argument therefore implies an upper limit on galaxy masses, at a scale that is not too far from what is observed.

In my view, this argument is fatally flawed by the assumption that all of the collapsed gas is at the same density. More realistically, the gas will have a density profile, and the dense gas near the center will be able to cool even if the more diffuse gas at larger radii cannot.

White & Rees (1978, MNRAS, 183, 341) introduced a more realistic picture, in which the amount of gas that cools in a collapsed dark matter halo is determined by the *cooling radius*, the radius at which the cooling time is equal to the time that the object is around before merging with another object of comparable mass (which reheats the gas and "resets the clock.") — or the age of the universe if that is longer.

In this case, the fraction of the mass that cools decreases as the mass of the halo increases, but the total cooled mass still increases. This model therefore does not yield a sharp upper cutoff in galaxy masses.

The cooling radius approach, combined with a more firmly based cosmological model and a more sophisticated framework for calculating halo merger rates, is the basis of "semi-analytic" methods for modeling galaxy formation. Gas cooling must be supplemented with a model for how the gas density determines the star formation rate and, crucially, how the star formation influences the surrounding gas ("feedback").

Apart from the uncertainties associated with star formation and feedback, the key uncertainty in this approach is whether the density profiles and temperatures inferred from the spherical collapse picture are realistic at the desired level of accuracy. In particular, there are suggestions from 3-d simulations of the hierarchical build-up of galaxies that much of the gas that ends up in the cold component is never heated to the virial temperature of a typical galaxy halo. It may yet turn out that this suggestion is a flaw in the numerical simulations themselves, but if it is correct then it calls into question one of the key assumptions of the semi-analytic methods.

8.5 Photoionization

At redshift $z < 6$ (and perhaps higher, but that is still a matter of debate), the universe is pervaded by an ultraviolet radiation background, produced by quasars and (perhaps) star-forming galaxies. (There is no question that high mass stars produce some UV photons energetic enough to ionize hydrogen and even helium. The unknown factor is what fraction of these photons escape from the ISM to become part of the intergalactic UV background.)

At $z \sim 2-3$, when the quasar population is at its peak, the hydrogen photoionization rate is estimated to be

$$\Gamma_{\gamma \mathrm{H}_0} \sim 10^{-12}\ \mathrm{s}^{-1},$$

which corresponds to a background intensity at the Lyman limit frequency ν_T of $J(\nu) \sim$ few $\times$ $10^{-22} \mathrm{erg\,s^{-1}\,cm^{-2}\,sr^{-1}\,Hz^{-1}}$ (see eq. 62).
The helium photoionization rates $\Gamma_{\gamma \mathrm{He}_0}$ and $\Gamma_{\gamma \mathrm{He}_+}$ are more uncertain, probably lower by a factor of a few ($\Gamma_{\gamma \mathrm{He}_0}$) and by a factor of 10-100 ($\Gamma_{\gamma \mathrm{He}_+}$).
For gas near the cosmic mean density, the photoionization rate is much higher than the recombination rate, so the hydrogen is nearly all ionized and the helium is nearly all doubly ionized.

Photoionization has two important effects on gas cooling.

(1) It eliminates line excitation and ionization as cooling processes at low densities, by eliminating H^0, He^0, and He^+. Recombination cooling increases, but the net effect is to severely reduce cooling rates at $T \sim 10^4 - 10^5$ K for low density gas. At high densities (where the recombination rate becomes comparable to or larger than the photoionization rate), the cooling rates move back towards their collisional equilibrium values.

(2) It *heats* the gas because photoelectrons carry off residual energy. The heating rate is

$$\mathcal{H} = n_{\mathrm{H_0}}\epsilon_{\mathrm{H_0}} + n_{\mathrm{He_0}}\epsilon_{\mathrm{He_0}} + n_{\mathrm{He_+}}\epsilon_{\mathrm{He_+}}, \tag{64}$$

where, for example

$$\epsilon_{\mathrm{H_0}} = \int_{\nu_T}^{\infty} \frac{4\pi J(\nu)}{h\nu}\sigma(\nu)(h\nu - h\nu_T)d\nu \quad [\mathrm{erg\ s^{-1}}].$$

The heating rate decreases with increasing temperature because the recombination rates (and hence the fraction of neutral "targets" for the photons) decline.

At low densities, $\mathcal{H} \propto n_H^2$ because, e.g.,

$$n_{\mathrm{H_0}} \propto n_{\mathrm{H}} \times \frac{\text{recombination rate per atom}}{\text{photoionization rate per atom}} \propto n_{\mathrm{H}}^2.$$

The net rate of radiative energy change per unit volume is $\mathcal{H} - \Lambda$. The leading dependence on density is $\mathcal{H} - \Lambda \propto n_{\mathrm{H}}^2$, but there is a slow change with $n_{\mathrm{H_0}}$ because of the competition between photoionization and recombination rates. Thus there is a density dependent "cooling curve"

$$\frac{\mathcal{H} - \Lambda}{n_{\mathrm{H}}^2}(T),$$

(e.g., Katz et al. 1996, Figure 2).

8.6 The equilibrium temperature

In the presence of photoionization, primordial (H/He) gas gets heated at low temperatures and cooled at high temperatures. There is thus a value of T at which

$$\mathcal{H}(T_{\mathrm{eq}}) - \Lambda(T_{\mathrm{eq}}) = 0.$$

If other processes (e.g., shock heating, compression, expansion) can be ignored, the gas will eventually settle to this equilibrium temperature. It is then said to be in thermal equilibrium. (Note that this has nothing to do with local thermodynamic equilibrium.)

For photoionized H/He, T_{eq} is always in the range $10^4 - 10^5$ K. It is higher for a harder (bluer) ionizing background spectrum because the residual photoelectron energy is higher if the typical ionizing photons are more energetic.
T_{eq} decreases with increasing density because at higher densities ionization and line cooling processes become more important.

The timescale for achieving thermal equilibrium is the cooling time, equation (63), but now with Λ replaced by $\mathcal{H} - \Lambda$.
At low densities or high temperatures, t_{cool} can be very long, so thermal equilibrium is a less robust assumption than ionization equilibrium. Systems often change on a timescale much shorter than that required to reach thermal equilibrium.

8.7 Some general lessons

There are *many* other radiative heating and cooling processes that are important in astrophysical situations, e.g.

metal-line cooling (the CII fine-structure line is especially important)
Compton cooling (fast electrons lose energy by upscattering photons; at $z \gtrsim 7$, Compton cooling off microwave background photons can be the dominant cooling process)
synchrotron cooling
molecular line cooling
molecular formation cooling
Compton heating by X-rays or γ-rays
Cosmic ray heating

However, the H/He plasma illustrates a number of features that appear in many situations.
(1) For a general treatment, one must integrate time-dependent equations to compute the abundances of each species, then use these abundances to compute cooling rates.
(2) If ionization and recombination timescales are short, one can compute abundances using ionization equilibrium.
(3) If ionization equilibrium applies, cooling/heating rates are still usually strongly dependent on temperature. In simple cases they are $\propto \rho^2$, though in general the density is more complex.
(4) If there are both heating and cooling processes, then there is usually an equilibrium temperature where $\mathcal{H} - \Lambda = 0$. The gas will relax to this equilibrium temperature if it has sufficient time, though at low densities it often does not.

In some cases, the dependence of equilibrium temperature on density leads to *thermal instability*. A small amplitude perturbation can cause the gas to spontaneously separate into two phases that are in pressure equilibrium at different values of $T_{\rm eq}$, a cool dense phase and a warm diffuse phase with equal pressure. Thermal instability is probably important in determining the structure of the ISM. (See discussion in Shu, chapter 8.)