# Galaxy formation and evolution
# 53863, 5 op, autumn 2020

on Zoom

## Lecture 8: Formation and evolution of gaseous haloes, 30/10/2020

# On this lecture we will discuss

1. Shock heating of infalling gas and hydrostatic equilibrium.
2. Virial equilibrium and the virial temperature.
3. Radiative cooling. The definition of the cooling function.
4. Summary of the main cooling processes and photoionization.
5. The primordial cooling function and the cooling function for non-zero gas metallicity.
6. The role of cooling in galaxy formation. Cold and hot gas flows.
7. The lecture notes correspond to: MBW: pages 366-393 (§8.1-8.4)
   L: pages 477-482 (§16.2)

# 8.1 The cooling and heating of gas in dark matter haloes

- So far during this course we have concentrated on the formation of structure under the influence of gravity alone.
- However, since the galaxies we observe directly are made of baryons and electrons, the role of gas-dynamical and radiative processes must be taken into account in order to understand how the structures we observe form and evolve.
- As demonstrated during the first part of this course, the density perturbation fields of the baryons $\delta_B$ and dark matter $\delta_{DM}$ are expected to be equal in the linear regime.
- The key difference between baryons and dark matter is that baryons can cool by radiating energy away leading to a collapse. As a result the observable baryonic galaxy is an order magnitude smaller than the more diffuse dark matter component.

# Shock heating I

- Let us study a gas cloud of mass $M_{gas}$ falling into a dark matter halo with mass $M_{halo}$ with velocity $v_{in}$. At some point the gas is shocked, either close to the centre, where the flow lines converge or at the accretion shock, which is located at the virial radius.
- Making the assumption that the shock thermalizes all the kinetic energy of the gas flow so that $<v_{gas}>\sim 0$ after it is shocked and assuming that the internal energy of the gas is small compared to the kinetic energy at infall, we get from energy conservation for a mono-atomic gas with $\gamma$=5/3:

$$E_{\mathrm{init,sh}} = \frac{3}{2} N k_B T_{\mathrm{sh}} = \frac{1}{2} M_{\mathrm{gas}} v_{\mathrm{in}}^2$$

- The shock temperature can be solved as: $T_{\mathrm{sh}} = \frac{\mu m_p}{3 k_B} v_{\mathrm{in}}^2$

# Shock heating II

- In addition, if we make the assumption that the gas fall in from a very large distance where the gravitational potential $\Phi$(r)~0, we get the following expression for the initial velocity:

$$v_{\rm in} \simeq v_{\rm esc}(r_{\rm sh}) = \sqrt{2|\Phi(r_{\rm sh})|}$$

- Finally $v_{in}$ can be related to virial velocity $v_{vir}$ by making the common assumption $r_{sh}=r_{vir}$ then:

$$v_{\rm in}^2 = \zeta \frac{GM_{\rm vir}}{r_{\rm vir}} = \zeta v_{\rm vir}^2 \qquad\qquad T_{\rm sh} \simeq \frac{\zeta \mu m_p}{3k_B} v_{\rm vir}^2$$

- The correction factor $\zeta$ is of the order of unity and depends on the details of the dark matter halo profile, i.e. isothermal, NFW, etc.

# Hydrostatic equilibrium I:

- Under the assumption that the gas is non-radiative (no cooling, except for adiabatic expansion), then the shocked gas will settle in hydrostatic equilibrium defined as:

$$\nabla P(r) = -\rho_{\mathrm{gas}} \nabla \Phi(r)$$

- In spherical symmetry we get for the total mass $M(r)=M_{gas}(r)+M_{DM}(r)$ and for an ideal gas, respectively, the following equations:

$$\nabla \Phi(r) = \frac{d\Phi}{dr} = \frac{GM(r)}{r^2} \qquad \nabla P(r) = \frac{dP}{dr} = \frac{k_B}{\mu m_p} \frac{d}{dr}(\rho T)$$

- Combining the two conditions we get for the total mass M(r):

$$M(r) = -\frac{k_B T(r) r}{\mu m_p G} \left[ \frac{d \ln \rho_{\mathrm{gas}}}{d \ln r} + \frac{d \ln T}{d \ln r} \right]$$

# Hydrostatic equilibrium II:

- If the gas is in hydrostatic equilibrium and both T(r) and $\rho_{gas}$(r) are known, one can infer the total mass profile $M_{tot}$(r). This is commonly done for galaxy clusters.
- In this derivation we assumed that all of the pressure is due to thermal gas pressure ($P_{tot}$=$P_{thermal}$). However, it is possible that the total pressure has significant contributions from non-thermal sources, such as non-thermal turbulence, magnetic fields and/or cosmic rays. In this case the equation for hydrostatic equilibrium must be modified:

$$M(r) = -\frac{k_B T(r) r}{\mu m_p G}\left[\frac{d\ln\rho_{\rm gas}}{d\ln r} + \frac{d\ln T}{d\ln r} + \frac{P_{\rm nt}}{P_{\rm th}}\frac{d\ln P_{\rm nt}}{d\ln r}\right]$$

- In addition to hydrostatic equilibrium also other simplifying assumptions must often be made, such as isothermal gas T(r)=const.

# Virial temperature I

- If the temperature profile of the gas is not known, which is most often the case, we can get a rough estimate of the temperature of the gas by assuming virial equilibrium. The scalar virial theorem states:

$$\frac{1}{2}\frac{d^2 I}{dt^2} = 2K + W + \Sigma$$

- Here I is the moment of inertia ($d^2I/dt^2=0$ in virial equilibrium, $d^2I/dt^2>0$ -> expansion, $d^2I/dt^2<0$ -> contraction), K is the kinetic energy, W is the potential energy and $\Sigma$ is the work done by the surface pressure.
- For a spherical halo of mass $M_{vir}$ and radius $r_{vir}$, with ideal mono-atomic gas:

$$3\frac{M_{\rm gas}}{\mu m_p}k_B T_{\rm vir} - \zeta\frac{G M_{\rm gas} M_{\rm vir}}{r_{\rm vir}} - 4\pi r_{\rm cl}^3 P_{\rm ext} = 0$$

# Virial temperature II

- If we ignore the external pressure $P_{ext}$ we get for the virial temperature exactly the same result as for the shock-heated gas:

$$T_{\rm vir} = \frac{\zeta \mu m_p}{3k_B} v_{\rm vir}^2$$

- As an example we can take the truncated, singular isothermal sphere of gas (no dark matter) for which $\zeta$=3/2. Then the virial theorem implies a virial temperature of:

$$T_{\rm vir} = \frac{\mu m_p}{2k_B} v_{\rm vir}^2 \simeq 3.6 \times 10^5 {\rm K} \left( \frac{{\rm v_{vir}}}{100\ {\rm kms^{-1}}} \right)^2$$

- Here the molecular weight $\mu$=0.59 is appropriate for a fully ionized primordial gas. The virial temperature is a very useful diagnostic, although realistically gas in virialized dark matter haloes is not isothermal and will in general have a radial temperature profile.

# 8.2 Radiative cooling: Cooling function

- Until now we have ignored radiative processes, which can cause both cooling and heating of the gas.
- The unit of the volumetric heating and cooling rates are, respectively:

$$[H] = [C] = \mathrm{erg\ s^{-1}cm^{-3}}$$

- Let us now assume that there is no heating, H=0 and that we operate in the optically thin limit, so that all photons can escape the system.
- It is useful to define the cooling function $\Lambda$(T,Z), which depends on the temperature T and the metallicity Z, but not on the gas number density $n_H$:

$$\Lambda(T) = \frac{C}{n_H^2} \quad [\Lambda] = \mathrm{erg\ s^{-1}cm^3}$$

# Cooling time I

- The cooling time is defined as the time it takes the gas to radiate away its internal energy for a given metallicity and is given by:

$$t_{\rm cool} = \frac{\rho\epsilon}{C} = \frac{\rho\epsilon}{n_H^2\Lambda(T)}$$

- For monatomic ideal gas ($\gamma$=5/3) of primordial composition and fully ionized the cooling time can be written as:

$$t_{\rm cool} = \frac{3nk_BT}{2n_H^2\Lambda(T)} \simeq 3.3\times10^9{\rm yr}\left(\frac{\rm T}{10^6{\rm K}}\right)\left(\frac{\rm n}{10^{-3}{\rm cm}^{-3}}\right)^{-1}\left(\frac{\Lambda({\rm T})}{10^{-23}{\rm ergs}^{-1}{\rm cm}^3}\right)^{-1}$$

- The cooling time $t_{cool} \propto n^{-1} \propto \rho^{-1}$ -> denser gas cools faster.
- The impact of cooling can be estimated by comparing it to the Hubble time ($t_H \propto H(z)^{-1}$) and the dynamical free-fall time ($t_{ff} \propto \rho^{-1/2}$), see Lecture 1 for details.

# Cooling time II

1. $t_{cool} > t_H$: The cooling time is long and cooling is not important. The gas will remain in hydrostatic equilibrium, unless it was recently disturbed.
2. $t_{ff} < t_{cool} < t_H$: The system is in quasi-hydrostatic equilibrium and evolves on the cooling time scale. The gas contracts slowly as it cools, but the system has enough time to maintain and re-establish hydrostatic equilibrium. Note here ff=free-fall.
3. $t_{cool} < t_{ff}$: The cooling is catastrophic. Gas cannot respond fast enough to the loss of pressure. Since the cooling time decreases with increasing density, cooling proceeds faster and faster. Gas falls to the centre on the free-fall time.

- Because $t_{cool} \propto \rho^{-1} \propto (1+z)^{-3}$ and $t_{ff} \propto \rho^{-1/2} \propto (1+z)^{-3/2}$. Cooling is generally more efficient at higher redshifts (for primordial gas).

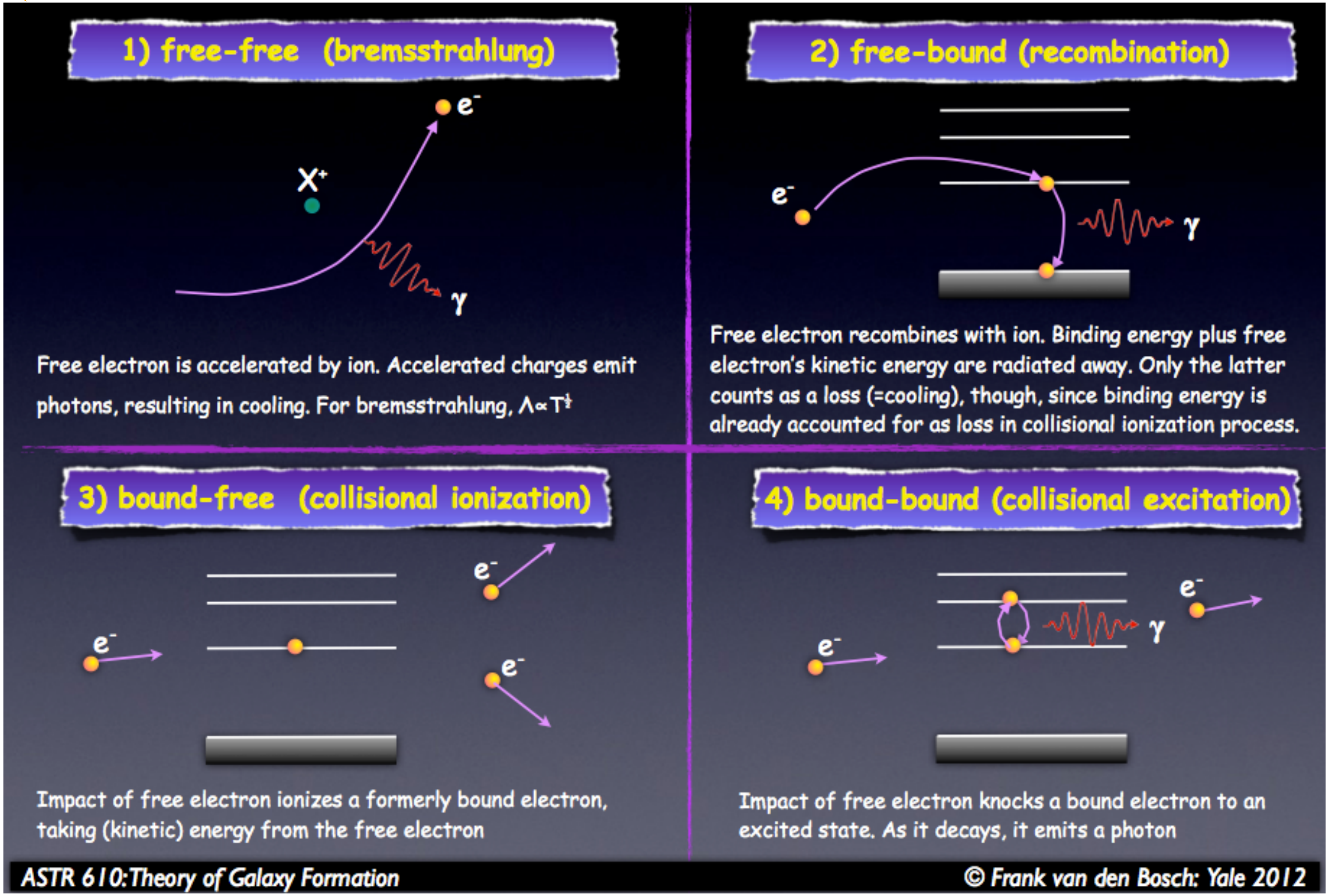
# Cooling processes I

- The primary cooling processes relevant for galaxy formation are two-body radiative processes, in which photons are emitted as a consequence of two-body interactions (for details MBW Appendix B1.3-B1.4): (Due to the low gas densities, processes involving three bodies can be ignored.)

1. Free-free emission (bremsstrahlung) due to the acceleration of electrons as they encounter atomic nuclei: $L_{ff} \propto T^{1/2}$. Important at very high temperatures, $T \geq 10^6$ K. Note here: ff=free-free
2. Free-bound emission (recombination), in which an electron recombines with an ion and emits a photon, hence loosing thermal energy.
3. Bound-free emission (collisional ionization), in which atoms and ions are ionized by collisions with electrons removing from the gas an amount of kinetic energy equal to the ionization threshold.
4. Bound-bound emission (collisional excitation), in which atoms are first excited by collisions with electrons and then decay radiatively to the ground state.

# Cooling processes II

All processes require free electrons and are strongly dependent on temperature.



# Compton cooling and photoionization heating I

- When photons of low energy $E=h\nu$ pass through a thermal gas of non-relativistic electrons with $T_e$, photons and electrons exchange energy due to Compton scattering.
- If the radiation field is a thermal background (CMB) with temperature $T_\gamma << T_e$, the net effect is for electrons to lose energy to the radiation field. $C_{comp} \propto T_\gamma^4 \propto (1+z)^4$, inverse Compton cooling is important at particurlarly high redshifts.
- In addition to cooling, the gas can also be heated due to the photo-ionizing background caused by UV radiation from quasars and star-forming galaxies. The intensity of the UV background peaks at z~2.
- The UV radiation causes photo-ionization that 1) severely suppresses the cooling of low density gas and 2) also directly heats the gas.
- The UV radiation eliminates ionization and line excitation as cooling processes at low densities. At high densities recombination very effective -> cooling is virtually unaffected by the UV background.

# Photoionization heating II

- Photoionization also heats the gas because the photoelectrons carry off residual energy left over from the ionization:

$$H = n_{H_0}\epsilon_{H_0} + n_{He_0}\epsilon_{He_0} + n_{He_+}\epsilon_{He_+}$$

$$\epsilon_i = \int_{\nu_i}^{\infty} \frac{4\pi J(\nu)}{h\nu}\sigma_i(\nu)(h\nu - h\nu_i)d\nu$$

$\epsilon_i$ = residual energy after ionization for a given flux and (frequency-dependent) cross-section.

- The heating rate decreases with increasing temperature, because the recombination rates, and hence the neutral targets for the photons decline.
- In the presence of photo-ionization what is important is the net heating/cooling rate and now because of the competition between photo-ionization and recombination the cooling function also depends on density, unlike the case without photo-ionization:

$$\Lambda = \frac{C - H}{n_H^2} = \Lambda(T, n_H)$$

# 8.3 Constructing the cooling function

- In order to calculate the cooling function $\Lambda$(T) we need to first determine the densities of the various ionic species. In the case of gas of a primordial composition we have a only hydrogen and helium and the species are $n_e$, $n_{H0}$,$n_{H+}$,$n_{He0}$, $n_{He+}$, $n_{He++}$.
- At fixed total gas densities, the ionic densities are determined by:

$$\frac{dn_{H0}}{dt} = \alpha_{H+}(T) n_{H+} n_e - \Gamma_{eH0}(T) n_e n_{H0} - \Gamma_{\gamma H0} n_{H0}$$

- $\alpha_{H+}$(T) is the hydrogen recombination coefficient [$cm^3\,s^{-1}$].
- $\Gamma_{eH0}$(T) is the collisional ionization rate [$cm^3\,s^{-1}$].
- $\Gamma_{\gamma H0}$ is the photo-ionization rate [$s^{-1}$], which depends on the ionization threshold (13.6 eV for hydrogen) and the ionization cross-section.

# Ionization equilibrium

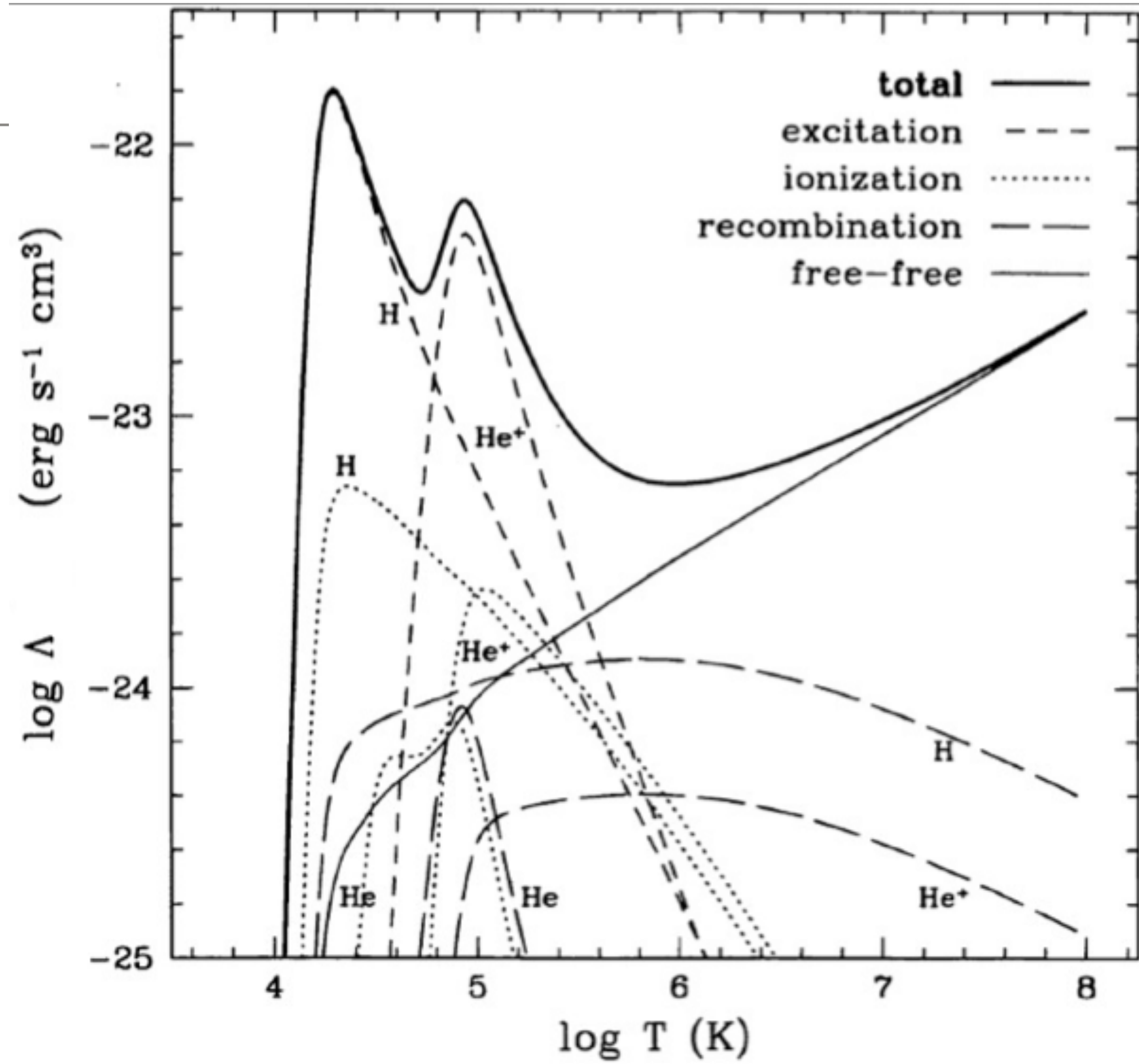
- The typical timescale for photo-ionization for a typical ionizing background is around $t_{phot} \sim (\Gamma_{\gamma H0})^{-1} \sim 3\times10^4$ yrs, which is a much shorter time than the typical dynamical times. Hence, the timescale on which $n_{H0}$ evolves is dominated by photoionization.
- However, in the absence of a photo-ionizing background, $n_{H0}$ evolves on a timescale of: $$\frac{1}{n_e(\alpha_{H+} - \Gamma_{eH0})} \sim 10^6 \mathrm{yr} \left( \frac{n_e}{10^{-5} \mathrm{cm}^{-3}} \right)^{-1}$$
- Both of these timescales are typically short compared to the hydrodynamical timescales. Thus it is usual safe to assume that the system is in ionization equilibrium as the destruction and creation rates are in equilibrium.
- If photo-ionization is ignored one still has the collisional ionization equilibrium (CIE).

# The CIE cooling function

- If there is no photo-ionization ($J(\nu)=0$) then under CIE, the relative abundances of ionic species depend only on temperature $\Lambda=\Lambda(T)$.
- At high T, the gas is fully ionized -> only bremsstrahlung.
- At $T<10^4$ K, all the gas is neutral -> no ions -> no bremsstrahlung at sufficiently low T. The residual free electrons do not have enough energy to excite H to its first excited state (which requires 10.2 eV).
- At T>few $\times 10^4$ K all H is ionized -> H no longer contributes to cooling a local drop in $\Lambda(T)$.
- Helium is responsible for a second peak in $\Lambda(T)$ at around $T\sim10^5$ K.
- If the gas contains metals, many new cooling channels (more cooling transitions) are available mainly between $\sim10^4$ K and $\sim10^7$ K. For a gas with solar metallicity the cooling rate at $T=10^6$ K is boosted by a factor of ~100 with respect to primordial gas.

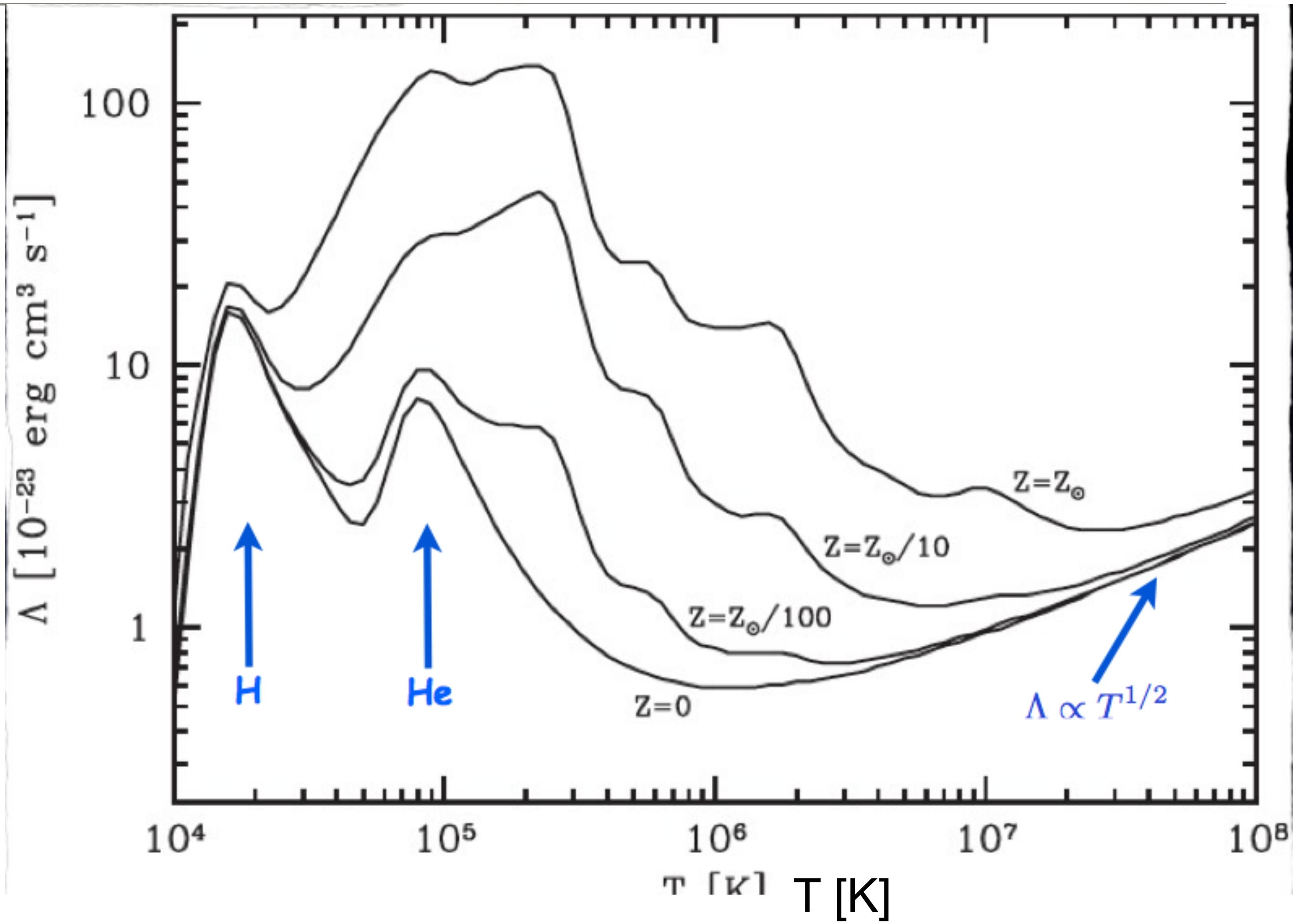
# The primordial cooling function

- The primordial cooling function has two distinct peaks due to collisionally excited hydrogen ($H^0$) and helium ($He^+$).
- The excitation of H and He are the dominant sources of cooling.



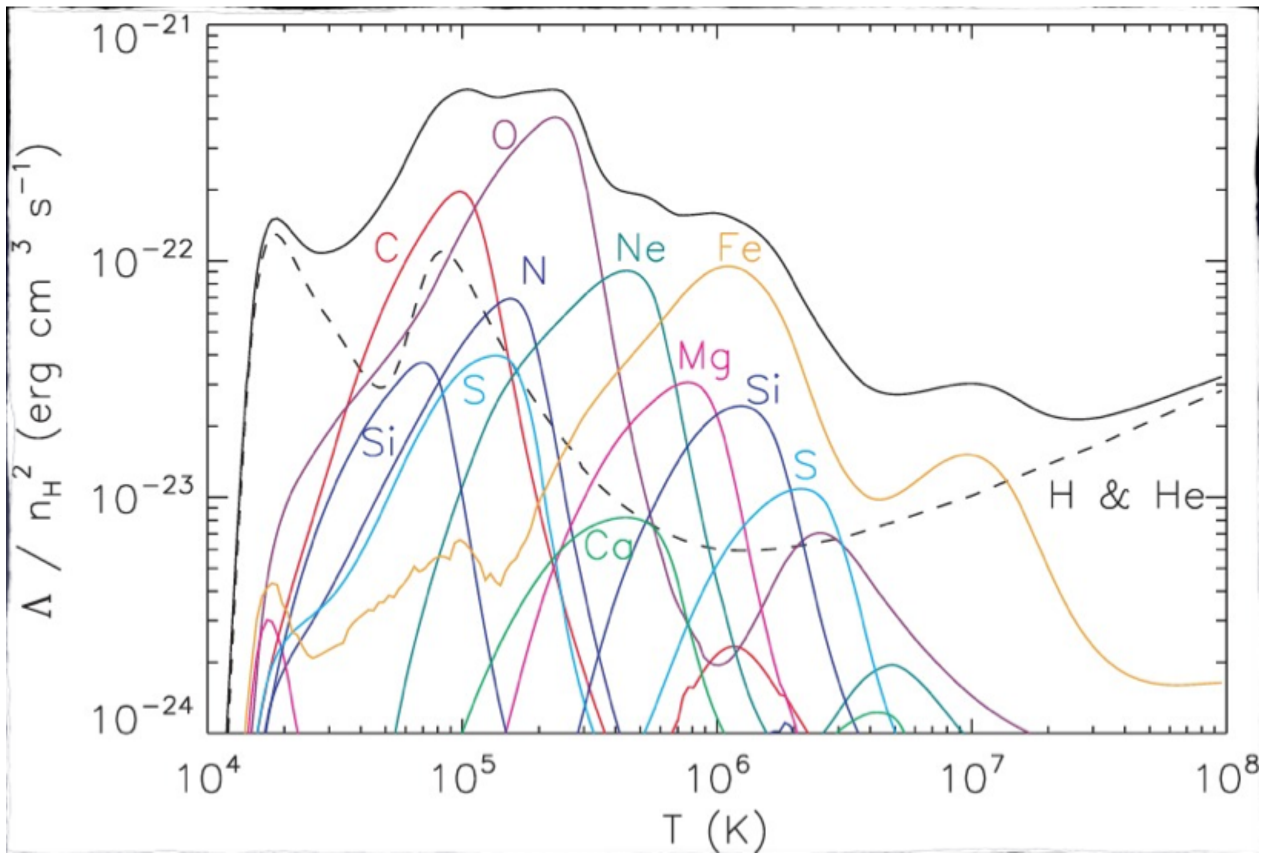
# The cooling function with non-zero metallicity I

- For metal-enriched gas there is strong peak at $T \sim 10^5$ K due to the collisionally excited levels of ions of oxygen, carbon and nitrogen.
- The overall cooling level can be boosted by up to a factor of ~100.



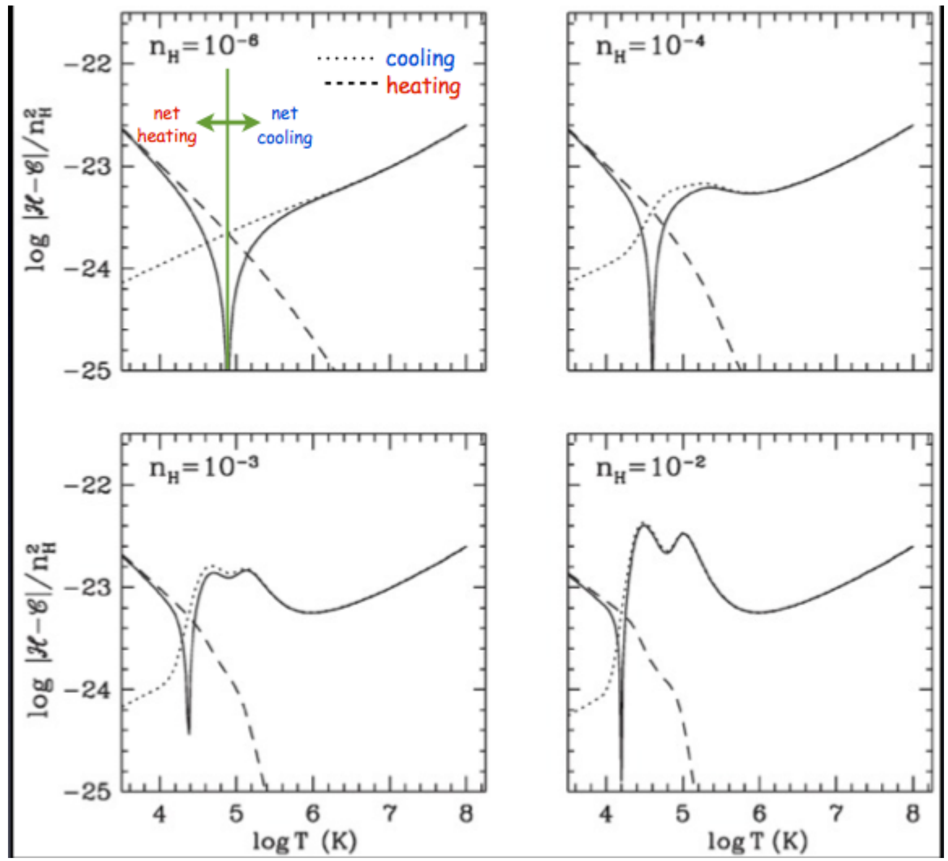
# The cooling function with non-zero metallicity II

- At $T<10^4$ K most of the electrons have recombined and the cooling level drops by several orders of magnitude.
- At these low temperatures cooling proceeds either by metals or by $H_2$ molecules in the primordial case.



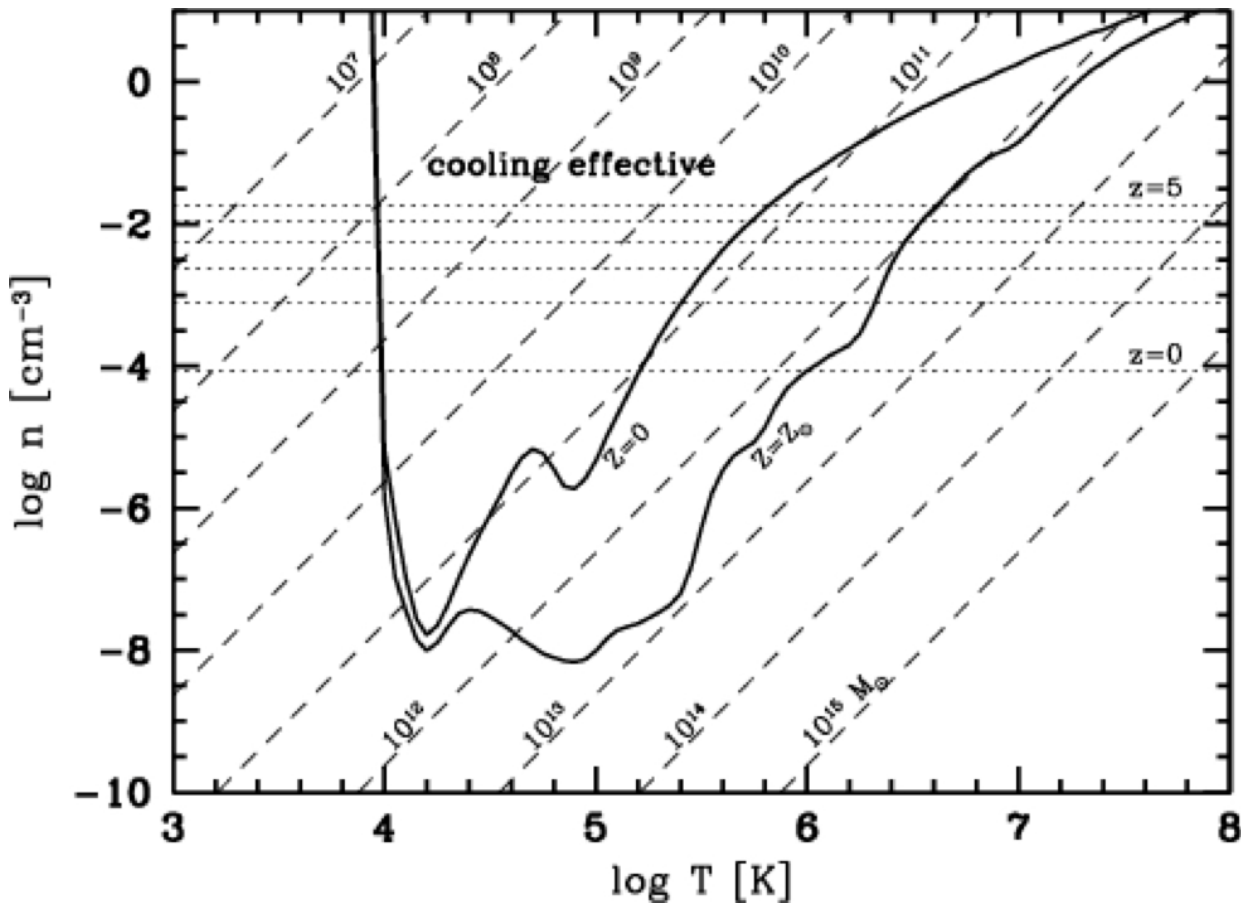
# The cooling function with a photo-ionizing background

- The net heating/cooling rates for gas of a primordial composition with a UV radiation background of typical intensity.
- The resulting curve is shown for four different densities. Heating dominates at low T and photo-ionization suppresses the H-He cooling peaks for low density gas.



# 8.4 Cooling and galaxy formation I

- The solid lines show $t_{cool}=t_{ff}$ for two metallicities. Cooling is efficient above the lines.
- The tilted lines are lines of constant gas mass.
- The horizontal dotted lines show the gas densities expected for virialized halos ($\delta$=200) at different redshifts.



# Cooling and galaxy formation II

- The plot assumes that the gas mass is related to the virial mass by the universal baryon fraction, $M_{gas}$~0.15 $M_{vir}$.
- Haloes with $M_{vir}<10^7\ M_\odot$ cannot cool except by inefficient molecular cooling.
- Haloes with $M_{vir}>10^{12}\ M_\odot$ (Z=0) and $M_{vir}>10^{13}\ M_\odot$ ($Z=Z_\odot$) cannot cool efficiently either -> a possible explanation for exponential cut-off in the luminosity function in early galaxy formation papers.
- However, this explanation is not valid, since massive galaxies form hierarchically -> smaller galaxies can cool, need another source for the exponential cut-off, AGNs? Also the curves are calculated only for $\delta$=200, the inner parts of the galaxies are considerably more overdense and can thus cool more efficiently.

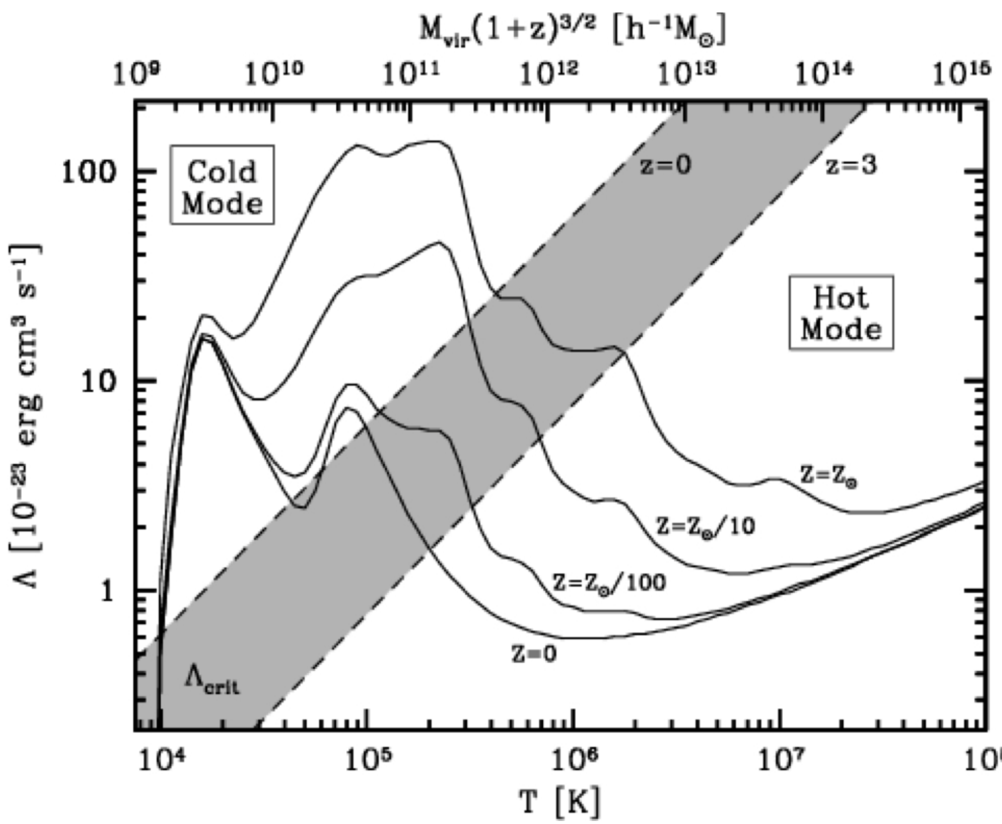
# Overcooling

- Already early investigations in the late 1970s studied the accumulation and condensation of gas in small dark matter haloes that form hierarchically. This resulted in a prediction that most of the gas should have cooled and formed stars, vastly overpredicting the number of observed faint galaxies. This is the ***overcooling problem***.
- Thus, the conclusion is that some extra process, such as supernova and/or supermassive black hole feedback must heat the gas.
- In semi-analytical models of galaxy formation it is common to define a cooling radius, which defined as the radius at which the cooling time equals the free-fall time. Using the cooling radius, the cooling rate can be defined as:

$$t_{\rm cool}(r) = \frac{3n(r)k_B T(r)}{2n_H^2(r)\Lambda(T)} \quad \dot{M}_{\rm cool}(r) = 4\pi\rho_{\rm gas}(r_{\rm cool})r_{\rm cool}^2 \frac{dr_{\rm cool}}{dt}$$

# Hot and cold gas flows

- When $r_{cool} >> r_{vir}$ we are in the regime of catastrophic cooling. If there is no hot gas, there is no accretion shock and the gas can flow in cold -> cold mode accretion.
- When $r_{cool} << r_{vir}$ only the inner gas can cool. In this case, the halo will have a hot atmosphere and an accretion shock. Newly accreted gas is shock-heated and then slowly cools -> hot mode accretion.




The band $\Lambda_{crit}$ indicates the boundary between hot and cold mode accretion.

# What have we learned?

1. Gas that flows into massive dark matter haloes will shock at the virial radius and heat to the ~virial temperature. In the absence of radiative cooling the gas will settle in hydrostatic equilibrium.
2. The cooling time is defined as the timescale it takes the gas to radiate away its thermal energy and it scales inversely with density. Higher density gas cools faster.
3. The primary cooling processes in galaxy formation are two-body processes, which scale with the density squared, whereas the heating processes scale only linearly with density.
4. Including metals will increase the efficiency of cooling by a factor of ~100. Lower mass haloes can accrete material in cold gas flows, whereas more massive haloes accrete gas in the hot mode.