arXiv:1412.5184v1 [astro-ph.GA] 16 Dec 2014

# Modeling Physical Processes at Galactic Scales and Above

Nickolay Y. Gnedin

Nickolay Y. Gnedin
Particle Astrophysics Center, Fermi National Accelerator Laboratory, Batavia, IL 60510, USA, e-mail: gnedin@fnal.gov; Kavli Institute for Cosmological Physics and Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637 USA

# Contents

**Modeling Physical Processes at Galactic Scales and Above** . . . . . 1
Nickolay Y. Gnedin
1 In Lieu of Introduction . . . . . 4
2 Physics of the IGM . . . . . 5
2.1 Linear Hydrodynamics in the Expanding Universe . . . . . 5
2.2 Lyman-$\alpha$ Forest . . . . . 7
2.3 Modeling the IGM . . . . . 16
2.4 What Observations Tell Us . . . . . 17
3 From IGM to CGM . . . . . 24
3.1 Large Scale Structure . . . . . 24
3.2 How Gas Gets Onto Galaxies . . . . . 26
3.3 Cool Streams . . . . . 27
3.4 Galactic Halos . . . . . 29
3.5 Diversion: Cooling of rarefied gases . . . . . 31
3.6 Back to Galactic Halos . . . . . 37
4 ISM: Gas In Galaxies . . . . . 41
4.1 Galaxy Formation Lite . . . . . 41
4.2 Galactic Disks . . . . . 43
4.3 Ionized, Atomic, and Molecular Gas in Galaxies . . . . . 47
4.4 Molecular ISM . . . . . 55
5 Star Formation . . . . . 66
5.1 Kennicutt-Schmidt and All, All, All . . . . . 66
5.2 Excursion Set Formalism in Star Formation . . . . . 72
6 Stellar Feedback . . . . . 75
6.1 What Escapes from Stars . . . . . 75
6.2 Unconventional Marriage: Feedback and Star Formation . 79
6.3 Toward The Future . . . . . 82
7 Answers To Brain Teasers . . . . . 84
References . . . . . 85

# 1 In Lieu of Introduction

What should these lectures be? The subject assigned to us is so broad that many books can be written about it. So, in planning these lectures I had several options.

One would be to focus on a narrow subset of topics and to cover them in great detail. Such a subset necessarily would be highly personal and useful to a few readers at best. Another option would be to give a very shallow overview of the whole field, but then it won't be very much different from a highly compressed version of a university course (which anyone can take if they wish so).

So, I decided to be selfish and to prepare these lectures as if I was teaching my own graduate student. Given my research interests, I selected what the student would need to know to be able to discuss science with me and to work on joint research projects. So, the story presented below is both personal and incomplete, but it does cover several subjects that are poorly represented in the existing textbooks (if at all).

Some of topics I focus on below are closely connected, others are disjoint, some are just side detours on specific technical questions. There is an overlapping theme, however. Our goal is to follow the cosmic gas from large scales, low densities, (relatively) simple physics to progressively smaller scales, higher densities, closer relation to galaxies, and more complex and uncertain physics. So, we (you - the reader, and me - the author) are going to follow a "yellow brick road" from the gas well beyond any galaxy confines to the actual sites of star formation and stellar feedback. On the way we will stop at some places for a tour and run without looking back through some others. So, the road will be uneven, but I hope that some readers find it useful.

## 2 Physics of the IGM

Most of the volume of the universe is occupied by gas outside galaxies, the so-called intergalactic medium (IGM). It may seem this gas is located far from galaxies, and should not be relevant to formation of galaxies and stars. Wrong! - IGM is the gas that eventually gets accreted by galaxies and turns into stars. After all, before the first galaxy formed, the whole universe was just IGM.

Hence, as we follow the "yellow brick road" to our goal of modeling star formation in galaxies, we pass through the IGM land first...

### *2.1 Linear Hydrodynamics in the Expanding Universe*

Linear dynamics of the non-relativistic cold dark matter is almost trivial, density fluctuation $\delta_X(t,k)$ with a spatial wavenumber $k$ satisfies a simple ordinary differential equation (ODE),

$$\frac{d^2}{dt^2}\delta_X(t,k) + 2H\frac{d}{dt}\delta_X(t,k) = 4\pi G\bar{\rho}\,\delta_{\rm tot}(t,k), \tag{1}$$

where $a(t)$ is the cosmic scale factor, $H(t) \equiv \dot{a}/a$ is the Hubble parameter and $\bar{\rho}$ is the mean density of the universe. If the universe only contained cold dark matter, then $\delta_{\rm tot} = \delta_X$. A second order ODE has two solutions, one of them is always growing with time,

$$\delta_X(t,k) = D_+(t)\delta_0(k), \tag{2}$$

where $D_+$ is called "the linear growing mode".

In reality, the universe contains gas, which is also subject to pressure forces. Hence, in the linear regime the evolution of the dark matter and gas fluctuations $(\delta_X, \delta_B)$ is described by a system of two coupled equations,

$$\frac{d^2\delta_X}{dt^2} + 2H\frac{d\delta_X}{dt} = 4\pi G\bar{\rho}\left(f_X\delta_X + f_B\delta_B\right), \tag{3}$$

$$\frac{d^2\delta_B}{dt^2} + 2H\frac{d\delta_B}{dt} = 4\pi G\bar{\rho}\left(f_X\delta_X + f_B\delta_B\right) - \frac{c_S^2}{a^2}k^2\delta_B, \tag{4}$$

where $f_X \approx 0.84$ and $f_B \approx 0.16$ are the mass fractions of dark matter and baryons respectively, and $c_S$ is the speed of sound in the gas.

This system of equations is coupled, but if high precision is not required, one can assume $f_B \ll f_X$ and ignore the baryonic contribution in the gravity terms in both equations. In that case the solution for the dark matter fluctuation is still given by equation (2), while the equation for the baryonic fluctuation reduces to

$$\frac{d^2\delta_B}{dt^2} + 2H\frac{d\delta_B}{dt} = 4\pi G\bar{\rho}\,\delta_X - \frac{c_S^2}{a^2}k^2\delta_B. \tag{5}$$

Notice the difference between this equation and an equation for baryonic fluctuations in a static reference frame ($a =$ const, no expansion of the universe) in the absence of dark matter:

$$\frac{d^2\delta_B}{dt^2} = 4\pi G\bar{\rho}\,\delta_B - \frac{c_S^2}{a^2}k^2\delta_B.$$

We know that in the latter case the characteristic scale over which baryonic fluctuations are suppressed by the pressure force is the Jeans scale,

$$k_J \equiv \frac{a}{c_S}\sqrt{4\pi G\bar{\rho}}.$$

Equation (5) cannot be solved analytically in a general case, but the important physics we are after is how baryonic fluctuations deviate from the dark matter ones. Hence, a quantity of interest is the ratio of two fluctuations, which can be expanded in the Taylor series of powers of $k^2$,

$$\frac{\delta_B(t,k)}{\delta_X(t,k)} = r - \frac{k^2}{k_F^2} + O(k^4), \tag{6}$$

where $r =$ const and we will call $k_F(t)$ a *filtering scale*. Because dark matter is expected to be more clustered that baryons (it is not a subject of the pressure force in the linear regime), we can expect that, in a general case $k_F > k_J$ (in the presence of dark matter baryonic fluctuations are less suppressed than in a purely baryonic case).

In the following we will only consider the case of $r = 1$ (baryons trace the dark matter on large scales), since this is an excellent approximation for $z < 10$. However, at higher redshifts this is not the case any more (Naoz & Barkana, 2007), as the different evolution of baryons and dark matter during the recombination epoch is not completely forgotten at these high redshifts (for example, $r \ll 1$ at $z > 1000$).

Substituting equation (6) into (5), it is possible to obtain an expression for $k_F$ in a closed form (Gnedin & Hui, 1998),

$$\frac{1}{k_F^2(t)} = \frac{1}{D_+(t)}\int_0^t dt' a^2(t')\frac{\ddot{D}_+(t') + 2H(t')\dot{D}_+(t')}{k_J^2(t')}\int_{t'}^t \frac{dt''}{a^2(t'')}.$$

While this expression is long and ugly, for reasonable thermal histories of the universe a good rule of thumb at $z \sim 2-4$ is $k_F \approx 2 \times k_J$ (the filtering scale is about half the Jeans scale).

Figure 1 gives an example of scale-dependence of $\delta_B(t,k)/\delta_X(t,k)$ for a representative thermal history of the universe at several redshifts (see Gnedin et al., 2003, for details). Fluctuations on small scales, where the pressure force dominates, are simple sound waves, and the transition to the baryons-trace-the-dark-matter regime is well described by the filtering scale.

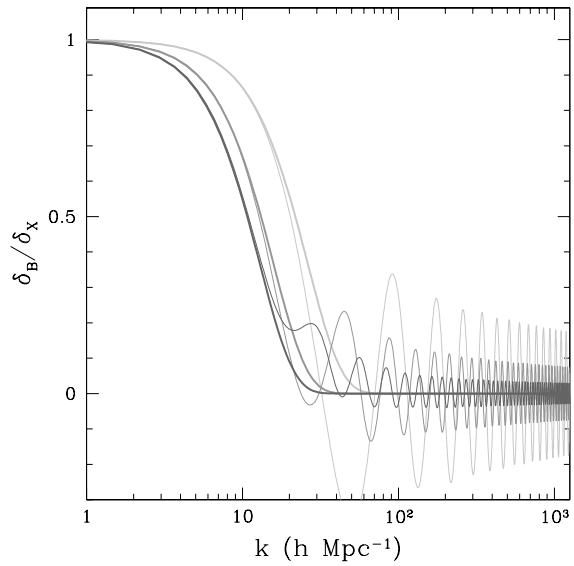



**Fig. 1** Solutions to equations (4) for a representative thermal history of the universe at $z = 4$ (light gray), $z = 1.5$ (medium gray), and $z = 0$ (dark gray); thin lines show the exact solutions, thick lines give the approximation $\delta_B/\delta_X = \exp(-k^2/k_F^2)$ (adopted from Gnedin et al. (2003)).

**Brain teaser #1**: Pressure generates sound waves, and sounds waves in the ideal gas do not dissipate. Why, then, are fluctuations "suppressed" by the pressure force?

## *2.2 Lyman-α Forest*

A well known empirical fact is that the IGM is highly ionized at low and intermediate redshifts, $z < 6$ (we will come back to that fact). To keep the cosmic gas ionized, the universe must be filled with ionizing radiation, the so-called "Cosmic Ionizing Background" (CIB).

Since most of the IGM is hydrogen, let us consider hydrogen first. The ionization balance equation for hydrogen in the expanding universe is simple,

$$\dot{n}_{\rm HI} = -3Hn_{\rm HI} - n_{\rm HI}\Gamma + R(T)n_e n_{\rm HII},$$

where $n_{\rm HI}$, $n_{\rm HII}$, and $n_e$ are number densities of neutral hydrogen, ionized hydrogen, and free electrons respectively, $\Gamma$ is the *photoionization rate* and $R(T)$ is a (temperature-dependent) recombination coefficient.

Often it is more convenient to consider not the actual number density of neutral or ionized hydrogen, but the *neutral fraction* $x \equiv n_{\rm HI}/n_{\rm H}$, because then the Hubble expansion term cancels out,

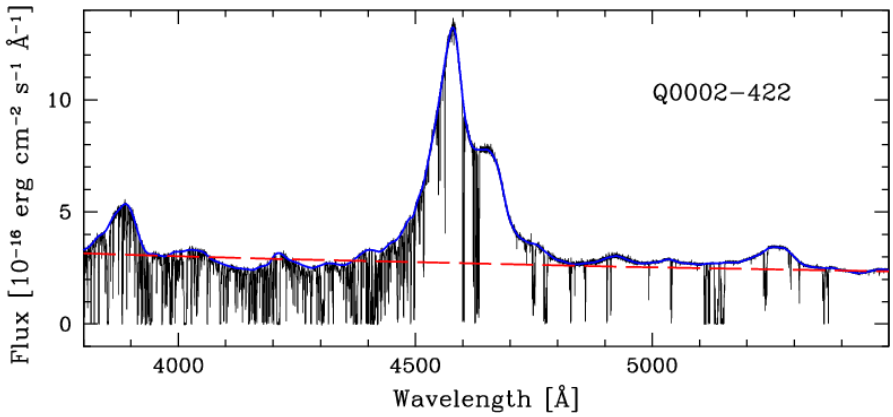



**Fig. 2** Typical $z \sim 3$ quasar spectrum together with the power law continuum fit (dashed red line) and the local continuum fit (blue line; adopted from Dall'Aglio et al. (2008)).

$$\dot{x} = -x\Gamma + R(T)n_e(1-x). \tag{7}$$

In the ionization equilibrium $\dot{x} = 0$, hence

$$x_{\rm eq} = \frac{R(T)}{\Gamma} n_e (1 - x_{\rm eq}),$$

and since the IGM is highly ionized ($x \ll 1$),

$$x_{\rm eq} = \frac{R(T)}{\Gamma} (\bar{n}_{\rm H} + 2\bar{n}_{\rm He})(1+\delta),$$

where we assumed that Helium is fully ionized, $\bar{n}_e = \bar{n}_{\rm H} + 2\bar{n}_{\rm He}$ (denser gas is more neutral).

Let us now consider a light source somewhere in the universe (a quasar, a galaxy, a gamma-ray burst, etc); the light source is at redshift $z_e$ in our reference frame. Let us also imagine that a photon with wavelength $\lambda_e$ is emitted by the source. As it propagates through the universe, the photon is going to be redshifted. At a redshift $z_a < z_e$ (from our reference frame) the photon has a wavelength

$$\lambda_e \frac{1+z_a}{1+z_e}.$$

Hence, for any $1216\text{Å}(1+z_e) < \lambda_e < 1216\text{Å}$ there is such $z_a$ that

$$\lambda_e \frac{1+z_a}{1+z_e} = 1216\text{Å}.$$

When a photon with wavelength of 1216Å (= Lyman-$\alpha$) hits a neutral hydrogen atom, it can get absorbed and excite the atom to $n = 2$ level.

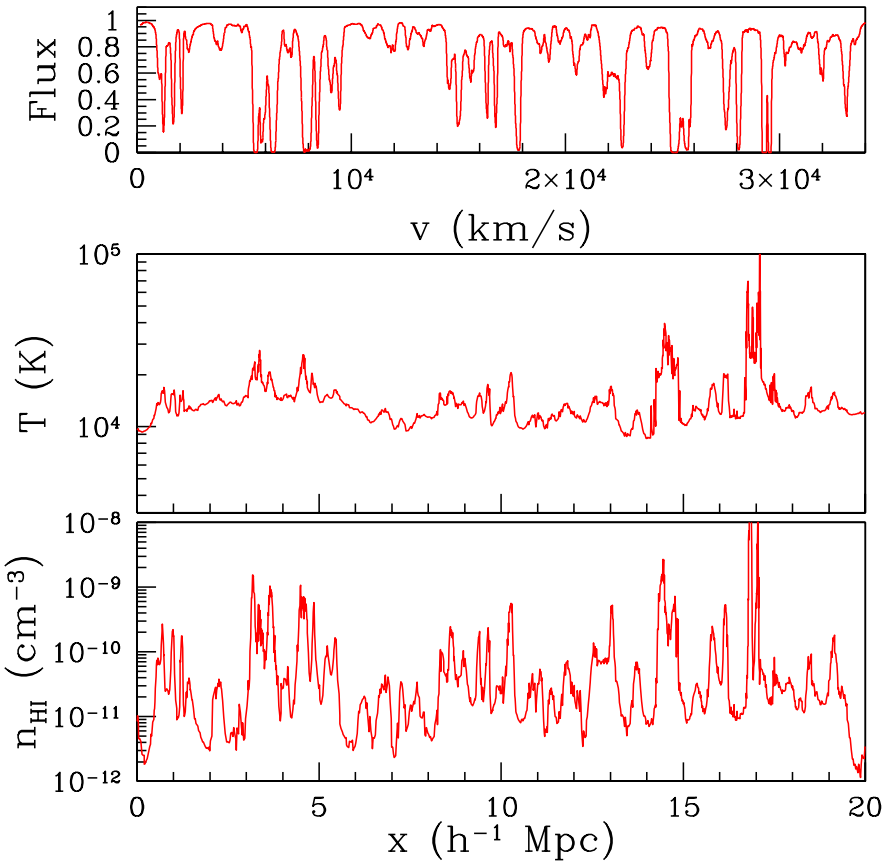



**Fig. 3** Runs of neutral hydrogen density (bottom) and gas temperature (middle) along one line of sight in a numerical simulation of Lyman-$\alpha$ forest at $z \sim 3$. The resultant absorption spectrum is shown in the top panel.

Indeed, this is exactly what happens in the real universe. Figure 2 shows a spectrum of a typical $z \sim 3$ quasar. The broad emission line in the middle is the Lyman-$\alpha$ of the quasar itself, and blue envelope for the observed spectrum is the continuum - i.e. the light that the quasar itself emitted. Black absorption lines come from the gas between us and the quasar, and the numerous forest of them at shorter wavelength is the hydrogen Lyman-$\alpha$ absorption from the neutral gas in the IGM, the so-called *Lyman-$\alpha$ Forest*.

Figure 3 illustrates how fluctuations in the neutral hydrogen density and in the gas temperature combine to produce the Lyman-$\alpha$ forest of absorption features in the spectrum. In order to understand how one goes from the lower two panels to the top one in that figure, we need to refresh the basics of resonant line absorption in the expanding universe.

**Brain teaser #2**: Hydrogen atoms do not sit forever in $n = 2$ state, they decay back into $n = 1$ and a Lyman-$\alpha$ photon is re-created back. How can there be any Lyman-$\alpha$ absorption?

### 2.2.1 Introduction To Resonant Line Absorption

The cross-section for an atom at rest to absorb a photon in the frequency range from $\nu$ to $\nu + \Delta\nu$ to the energy level with the energy $h\nu_0$ is

$$\sigma(\nu) = \frac{\pi e^2}{m_e c \nu_0} f \phi(\nu) \equiv \sigma_0 \phi(\nu),$$

where $f$ is the oscillator strength for the transition and

$$\phi(\nu) = \frac{1}{\pi} \frac{w\nu_0}{(\nu - \nu_0)^2 + w^2} \approx \nu_0 \delta(\nu - \nu_0),$$

where $w$ is the *natural line width* in frequency units. For hydrogen Lyman-$\alpha$ the combination of fundamental constants

$$\sigma_0 = \frac{\pi e^2}{m_e c \nu_0} f = 4.5 \times 10^{-18}\,\mathrm{cm}^2.$$

Atoms, though, are social creatures and rarely live alone. For a cloud of gas of density $n$, size $L$, and temperature $T$ we need to integrate over all atoms to compute the optical depth of the transition at any frequency $\nu$,

$$\tau(\nu) = nL \int \sigma_0 \phi(\nu') \frac{1}{\sqrt{\pi} b} e^{-\frac{(u_\nu - u')^2}{b^2}} du',$$

where $\nu' = \nu_0(1 + u'/c)$ and $u_\nu$ is defined via the expression $\nu = \nu_0(1 + u_\nu/c)$. The quantity

$$b = \left(2\frac{k_B T}{m_H}\right)^{1/2}$$

is called the *Doppler parameter* and the product $nL$ is the *column density*.

In an expanding universe it is not enough just to multiply by the cloud size $L$, since different locations along the line-of-sight are redshifted relative to the observer and project to different locations in the velocity (or frequency) space. Hence, we must integrate along the line-of-sight,

$$\tau(\lambda) = \sigma_0 \int n(x) \frac{c}{\sqrt{\pi} b_x} e^{-\frac{(u_\lambda - u_x)^2}{b_x^2}} \frac{dx}{1 + z_x}, \tag{8}$$

where we switch to from the frequency to the wavelength (as almost all observers tend to live in the wavelength space), and we integrate over the comoving distance $x$ (as almost all theorists tend to live in the comoving space); both $b_x$ and $u_x$ are, in general, functions of position, since the temperature and velocity vary in space. The wavelength is related to the velocity along the line-of-sight through the usual Doppler effect,

$$\lambda = \lambda_0 \left(1 + z_x + \frac{u_\lambda}{c}\right)$$

and $z_x$ is the redshift of location $x$ along the line-of-sight.

The spectrum shown on the top panel of figure 3 is just $\exp(-\tau(\lambda))$ with $\tau(\lambda)$ computed with equation (8) from the two bottom panels of the same figure.

Now we are ready to figure out why the IGM must be highly ionized at $z < 6$. From figure 2 we notice that the forest absorbs about 50% of the quasar flux, so the average optical depth is $\tau \sim 0.5 - 1$. Considering one absorption system stretching for about $\Delta x \sim 100\,\mathrm{kpc}$ and having temperature of $10^4$ K (or $b \sim 10\,\mathrm{km/s}$), an crude estimate for $\tau$ is

$$\begin{aligned}
\tau &\sim \sigma_0 \frac{c}{\sqrt{\pi} b} x_{\mathrm{HI}} n_{\mathrm{H}} a \Delta x \\
&= 4.5 \times 10^{-18}\,\mathrm{cm}^2 \frac{3 \times 10^5\,\mathrm{km/s}}{\sqrt{\pi} 10\,\mathrm{km/s}} x_{\mathrm{HI}} 1.3 \times 10^{-5}\,\mathrm{cm}^{-3} (4a)^{-3} 0.75 \times 10^{23}\,\mathrm{cm}(4a) \\
&= 7 \times 10^4 \frac{x_{\mathrm{HI}}}{(4a)^2}
\end{aligned}$$

at the cosmic scale factor $a$. To get $\tau \sim 1$ the neutral fraction $x_{\mathrm{HI}}$ must be $x_{\mathrm{HI}} \sim 10^{-5}$.

### 2.2.2 Temperature

The final component in modeling the IGM is to know what the temperature of the gas is.

Since the IGM is highly ionized, a process of photo-heating (heating by ionizing radiation) is important. When a high energy photon hits a hydrogen atom, 13.6 eV of its energy goes into ionizing the atom, the rest goes into the energy of the ejected electron. If the electron is not super-energetic (less that $\sim 40$ eV), it thermalizes and adds its energy to the thermal energy of the gas. For more energetic electrons the situation may be more complex, as it can ionize another atom by colliding with it (a so-called *secondary ionization*). That, in turn, produces an energetic electron which may ionize another atom etc. Usually, these secondary ionizations are only important if the gas is substantially (more than a few percent) neutral; for the low redshift IGM with $x_{\mathrm{HI}} \sim 10^{-5}$ secondary ionizations are completely unimportant.

If all the excess energy of an ionizing photon goes into heat, the rate of internal energy increase in the gas due to photo-heating is

$$\frac{3}{2} \frac{d}{dt} (n k_B T) \bigg|_{\mathrm{PH}} = c n_{\mathrm{HI}} \int_{E_0}^{\infty} (E - E_0) \sigma_{\mathrm{HI}}(E) n_E dE,$$

where $E_0 = 13.6\,\mathrm{eV}$ is the hydrogen ionization threshold, $\sigma_{\mathrm{HI}}(E)$ is the hydrogen ionization cross-section, and $n_E$ is the radiation spectrum measured in photons per unit volume per unit energy.

The photoionization rate of hydrogen is

$$\Gamma = c\int_{E_0}^{\infty} \sigma_{\mathrm{HI}}(E) n_E dE,$$

hence

$$\frac{3}{2}\frac{d}{dt}\left(nk_BT\right)\Big|_{\mathrm{PH}} = n_{\mathrm{HI}}\Gamma\langle\Delta E\rangle,$$

where $\langle\Delta E\rangle$ is the average excess energy (over 13.6 eV) of an ionizing photon,

$$\langle\Delta E\rangle \equiv \frac{\int_{E_0}^{\infty}(E-E_0)\sigma_{\mathrm{HI}}(E)n_E dE}{\int_{E_0}^{\infty}\sigma_{\mathrm{HI}}(E)n_E dE}. \tag{9}$$

Let us ignore helium for a moment: $n_e = (1-x)n_{\mathrm{H}}$, $n = n_{\mathrm{H}} + n_e = (2-x)n_{\mathrm{H}}$. Then the thermal balance equation together with the ionization balance equation become

$$\frac{3}{2}\frac{d}{dt}\left((2-x)k_BT\right) = x\Gamma\langle\Delta E\rangle, \tag{10}$$

$$\frac{d}{dt}x = -x\Gamma + R(T)n_{\mathrm{H}}(1-x)^2. \tag{11}$$

Let us start with cold neutral IGM ($x = 1$, $T = 0$, like at very high redshift, before cosmic reionization), and assume that the ionizing radiation pops out of nowhere instantaneously at a cosmic time $t_R$ (a favorite approximation of your CMB friends),

$$\Gamma \propto \theta(t - t_R)$$

($\theta(x)$ is a Heaviside function). In the ionization equilibrium

$$x_{\mathrm{eq}} = \frac{Rn_{\mathrm{H}}}{\Gamma}(1-x_{\mathrm{eq}})^2.$$

Hence, until the ionization equilibrium is established (i.e. while $x \gg x_{\mathrm{eq}}$) $x\Gamma \gg Rn_{\mathrm{H}}(1-x)^2$. In that limit equation (11) becomes simply

$$\frac{d}{dt}x = -x\Gamma$$

and its solution for $t > t_R$ is

$$x(t) = e^{-\Gamma(t-t_R)}.$$

That solution is valid until $x$ becomes small enough ($\sim Rn_{\mathrm{H}}/\Gamma$) for the ionization equilibrium to get established.

Equation (10) can also be solved easily in the same limit,

$$(2-x)k_BT = \frac{2}{3}\langle\Delta E\rangle\left(1-e^{-\Gamma(t-t_R)}\right),$$

and in the limit of small $x$ the gas temperature becomes constant (i.e. gas becomes *isothermal*),

$$T_\infty = \frac{\langle\Delta E\rangle}{3k_B} \tag{12}$$

and independent of density or the photoionization rate. This is an important lesson: **if a region of space is ionized rapidly, its temperature does not depend on the strength of the radiation field**. I.e., you cannot heat up the IGM by cranking up the ionizing source, only by making the source spectrum *harder*.

It is also instructive to plug some numbers into equation (12). For example, for a power-law energy spectrum for ionizing photons, $n_E \propto E^{-\alpha}$, and using the fact that just beyond the ionization edge $\sigma_{\rm HI}(E) \propto E^{-3}$, we find

$$\langle\Delta E\rangle = \frac{\int_{E_0}^{\infty}(E-E_0)\sigma_{\rm HI}(E)n_E dE}{\int_{E_0}^{\infty}\sigma_{\rm HI}(E)n_E dE} = \frac{E_0}{1+\alpha}$$

and

$$T_\infty = \frac{52{,}000\,\mathrm{K}}{1+\alpha} = \left[\begin{array}{l} 26{,}000\,\mathrm{K}\ (\alpha=1) \\ 5{,}000\,\mathrm{K}\ (\alpha=9) \end{array}\right.$$

In other words, the temperature of the photo-ionized gas is about 10,000 K, give-or-take a factor of 2.

Let us now consider what happens next. A region of space was ionized to $x = x_{\rm eq}$ at $t = t_R$ ($a = a_R$), and the temperature of the gas is at this moment constant at $T = T_\infty$. Another important effect is plain adiabatic cooling due to the expansion of the universe, so that the full equation that governs the temperature evolution after ionization equilibrium is established is

$$\frac{dT}{dt} = T\frac{2\dot{n}_{\rm H}}{3n_{\rm H}} + T_\infty x_{\rm eq}\Gamma = T\frac{2\dot{n}_{\rm H}}{3n_{\rm H}} + T_\infty R n_{\rm H}. \tag{13}$$

The recombination coefficient can be well approximated as a power-law function of gas temperature, $R(T) \approx 4.3\times10^{-13}T_4^{-0.7}\,\mathrm{cm^3/\,s}$ ($T_4 \equiv T/10^4\,\mathrm{K}$). It is easy to solve equation (13) for the temperature $T_0$ at the cosmic mean density, $\bar{n}_{\rm H} \propto a^{-3}$,

$$T_0(a) = T_\infty\left(\frac{a_R}{a}\right)^2\left[1+1.34R(T_\infty)\bar{n}_{{\rm H},R}t_R\left(\left(\frac{a}{a_R}\right)^{1.9}-1\right)\right]^{1/1.7}. \tag{14}$$

At late times ($a \gg a_R$) the asymptotic behavior of the temperature at the mean density is

$$T_0(a) \propto \left(\frac{a_R}{a}\right)^{1.5/1.7}.$$

It is less rapid than pure adiabatic expansion $T \propto a^{-2}$ because photo-heating off the residual neutral hydrogen fraction remains non-negligible at all times.

It turns out that for densities other than the mean a power-law ansatz provides a decent approximation for moderate overdensities, $\delta \lesssim 10$,

$$T(\rho) \approx T_0(1+\delta)^{\gamma-1}, \tag{15}$$

where both $T_0$ (as is given above) and $\gamma$ are functions of time (Hui & Gnedin, 1997). Expression for $\gamma(a)$ is rather ugly, but its main important properties are that $\gamma = 1$ right after instantaneous reionization and $\gamma \to 1.62$ at late times (notice, it is 1.62 and *not* $2/3$).

Figure 4 compares the temperature-density relation in the IGM from a real calculation (by following heating and cooling of individual fluid elements, Hui & Gnedin, 1997) with the approximate solution above. One effect that we ignored is photo-heating of helium - heat input from the ionizations of the residual neutral helium will heat the gas a bit more than is given by equation (14), but, overall, our analytical calculation does rather well.

"Hey", a meticulous reader will exclaim, "what about radiative cooling?" After all, gas does cool by emitting radiation. A story of gas cooling, with all its gory details, awaits us in the future, but here let us estimate how important radiative cooling actually is in the Lyman-$\alpha$ forest.

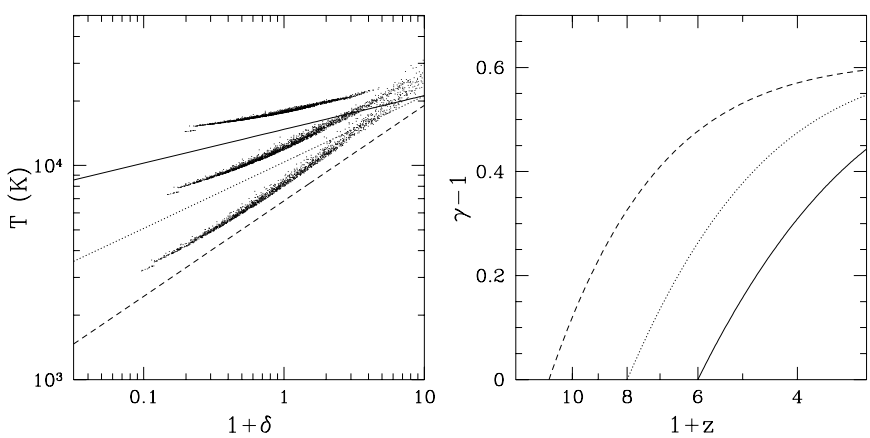



**Fig. 4** (Left) Temperature-density relation for a sudden reionization model at $z = 6$: dots show the the results of a full calculation at $z = 4, 3, 2$ (from top down) while lines are approximation (15) with $T_0$ given by equation (14); the approximate solution slightly underestimates the temperature because it ignores the heat input from helium ionizations. (Right) Time evolution of $\gamma$ for $z_R = 10, 7, 5$ respectively.

In a highly ionized gas the dominant radiative cooling mechanism is recombination cooling,

$$\left.\frac{dU}{dt}\right|_{\rm RC} = -\frac{3}{4}k_B T R(T) n_e n_{\rm HII}.$$

If we compare this term to photoionization heating in ionization equilibrium,

$$\left.\frac{dU}{dt}\right|_{\rm PH} = n_{\rm HI}\Gamma\langle\Delta E\rangle = \langle\Delta E\rangle R(T) n_e n_{\rm H},$$

we see that the radiative cooling is lower by a factor of

$$\frac{3}{4}\frac{k_B T}{\langle\Delta E\rangle} = \frac{T}{4T_\infty}.$$

Hence, radiative cooling makes at most a 25% correction, and well after reionization ($T < T_\infty$) the correction is even smaller.

Is this the complete story? Alas, no, the reality is always more complicated than we are ready to accept and you need to be aware of several caveats when using the temperature-density relation.

- The temperature-density relation is an *approximation*, with 5-10% scatter at low densities and progressively larger scatter as one moves up the density axis, because it misses a major hydrodynamic effects - shocks. Gas motions in the IGM will cause shock waves that will lead to additional gas heating.
- There may exist other heating and cooling mechanisms. For example, Puchwein et al. (2012) argued that heating of the Lyman-$\alpha$ forest by ultra-high energy gamma rays from a population of blazars is important at very low densities. The jury is still out on whether such an effect is important or not, but we should always be aware that we do not known everything.
- Our analysis assumed that the gas is optically thin to ionizing radiation. While this is the case for hydrogen at $z \lesssim 6$, helium is believed to be reionized the second time (from HeII to HeIII) at $z \sim 3$, in which case gas is *not* optically thin to helium ionizing radiation at $z \gtrsim 3$. Non-trivial opacity to ionizing radiation normally leads to increasing photo-heating rates in the gas.

**Brain teaser #3**: The temperature-density relation is sometimes called an "equation of state" (occasionally even without quotes). Do not fall into that trap - it relates the gas temperature and density, but it is *not an equation of state*. Can you explain why?

## 2.3 Modeling the IGM

The most straightforward model of Lyman-$\alpha$ forest is a hydrodynamic simulation with ionization balance. In the 1990-ties several approximate methods have been used, such as a log-normal approximation, Zel'dovich approximation, a pure N-body simulation, Hydro-Particle-Mesh (HPM) approximation. None of these methods is competitive any more and their use can be hardly justified.

The assumption of the ionization equilibrium is very good in the IGM, but it does break down in a few special cases (quasar proximity zones, helium reionization, etc). Hence, the most accurate simulation of the IGM includes (a model of) Cosmic Ionizing Background (CIB), radiative transfer, non-equilibrium ionization, separate fields for each of ionizing species (HI, HII, HeI, HeII, ...). Such a simulation, however, is usually an overkill, except when it is used for a special purpose like modeling non-equilibrium effects in quasar proximity zones.

A standard approach is to include CIB and ionization equilibrium and follow radiative heating (and, optionally, cooling) "on-the-fly" (a-la equation 13). For simulations of the Lyman-$\alpha$ forest alone the temperature-density relation may be assumed, but the computational savings in that case will be modest and it is rarely worth it.

Example of a numerical simulation of the forest is shown in figure 5. The right panel shows the gas density, and looks like a usual image of large-scale structure. The left panel shows the Lyman-$\alpha$ optical depth that would be observed in the corresponding position along the absorption spectrum towards a high redshift quasar. The main thing to take from that figure is that the actual absorption lines we see clearly in the spectra (those with $\tau \gtrsim 0.5$) come from filaments: weaker ones tend to cluster around stronger ones, although a few of the weakest ones do occur in the voids. The higher optical depth systems, those that lead to saturated lines with $\tau \gtrsim 2$ tend to occur at the intersections of filaments, nearer to galaxies.

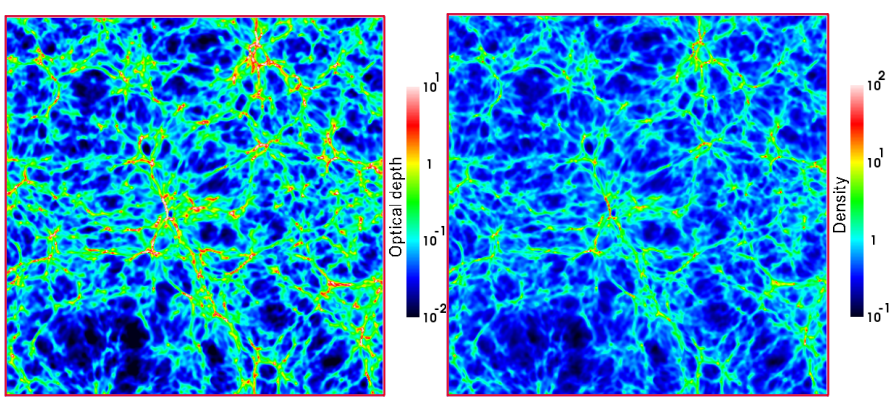



**Fig. 5** Slices of optical depth (left) and gas density (in units of the cosmic mean, right) in a simulation of Lyman-$\alpha$ forest at $z \sim 3$. The box size is $20h^{-1}$ comoving Mpc.

### 2.3.1 Density - Column Density Correlation

What is clear from figure 5 is that the gas density and the optical depth of the corresponding absorption feature are well correlated. Crudely, the relation is

$$\tau \sim (1+\delta)^{1.5},$$

although the slope and normalization of this correlation are redshift dependent.

This correlation is so good, especially on large scales, that it is often used to match directly the gas density into the opacity along the line of sight - such an *ansatz* is called *Fluctuating Gunn-Peterson Approximation*, or FGPA. FGPA is useful for modeling the forest on large scales, but one has always keep in mind that the absorption spectrum is in the velocity space, while the density is sampled in real, physical space, hence FGPA breaks down on sufficiently small scales (roughly less than 1 Mpc).

## *2.4 What Observations Tell Us*

For a long time since the Lyman-$\alpha$ forest was discovered in the 60-ties, it was treated by observers as just another absorption spectrum - as a collection of individual absorption lines, each having a fixed column density $N$ and the Doppler parameter $b$, as if the absorption was coming from discrete clouds. Now we know that this is not a good description - the density, temperature, and neutral fraction fields are continuous, and it is impossible to decompose the realistic spectrum into a set of $(N,b)$ pairs uniquely.

The modern view of the forest is that $\tau(\lambda)$ is a continuous field and should be treated as such. However, there is one application where the $(N,b)$ decomposition is still useful - measuring the temperature-density relation. If we think about a segment of the spectrum that has an "absorption line", the width of the feature is determined by the temperature of the gas plus any velocity gradient across the region that may exist. In some cases that velocity gradient will be very small, so the narrowest features at each column density should be those that are broadened by temperature alone. Hence, looking at the distribution of fitted $b$ parameters at given column density, one can measure $T(N)$ in the forest and, by virtue of the strong correlation between $\rho$ and $\tau$ (and, hence, $N$), translate that measurements into the measurement of $T-\rho$ correlation.

Figure 6 shows an example of such distribution from a single high resolution quasar spectrum (Rudie et al., 2012). The cutoff in the distribution of Doppler parameters for a given $N$ is clearly visible, and the value of the cutoff is well fit by the power-law in $N$, demonstrating the fact that the power-law temperature-density relation is indeed a good approximation.

A compilation of the majority of existing measurements is shown in figure 7 (Ricotti et al., 2000; Schaye et al., 2000; McDonald et al., 2001; Lidz et al., 2010;

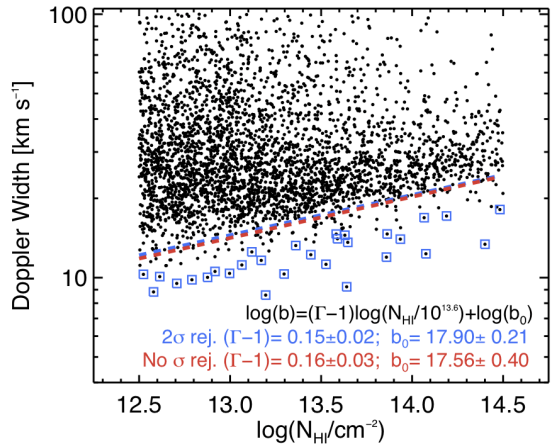



**Fig. 6** Example of the $(N,b)$ distribution for a quasar spectrum at $z = 2.4$. The measurement points labeled by blue squares are contamination from heavy elements. A relatively sharp edge of the distribution of Doppler parameters at a given $N$ is apparent in the figure (adopted from Rudie et al., 2012).

Rudie et al., 2012). The data seem to indicate (albeit rather vaguely) an increase in the temperature and a decrease in $\gamma$ at $z \approx 3$ - a behavior reminiscent of cosmic reionization (equation 12). Indeed, this may correspond to the second reionization of helium (HeII goint into HeIII), thought to occur at redshifts around 3.

An even simpler quantity is the *column density distribution* - a distribution of all $N$ values irrespectively of what their $b$ values are. Altay et al. (2011) show how modern cosmological simulations can match the observed distribution over 10 orders of magnitude in column density (see Figure 8).

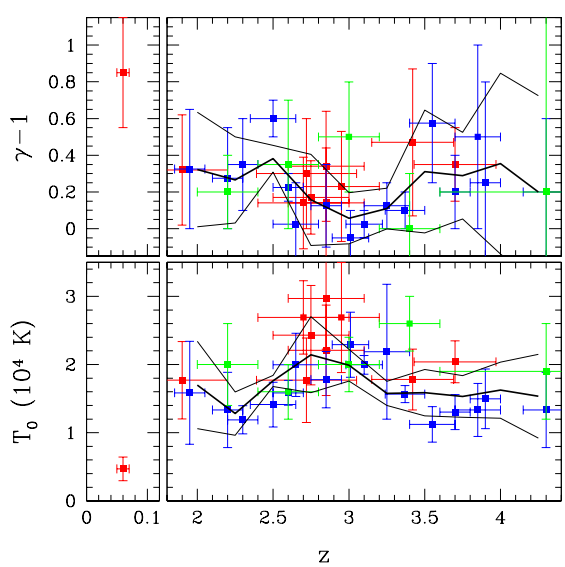



**Fig. 7** Evolution of the temperature-density relation $T \approx T_0(1+\delta)^{\gamma-1}$. Red, blue, and green points show individual measurements from Ricotti et al. (2000); Schaye et al. (2000); Lidz et al. (2010) respectively; thick and thin black lines shows the average values and $1\sigma$ dispersion for all measurements. Possible increase in temperature and dip in $\gamma$ at $z \sim 3$ is attributed to HeII $\rightarrow$ HeIII reionization.

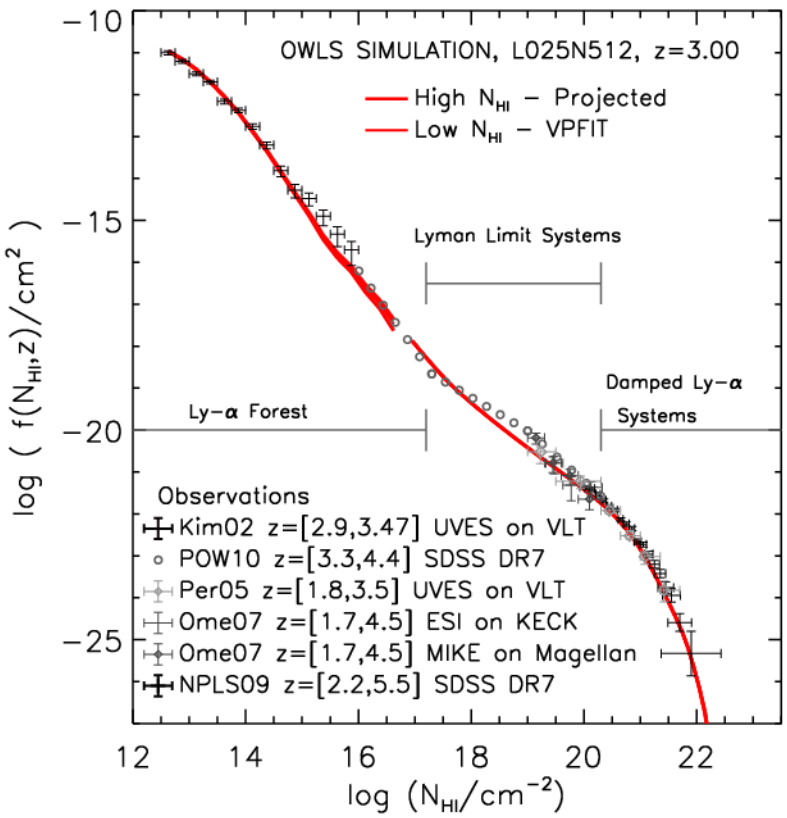



**Fig. 8** Distribution of column densities of Lyman-$\alpha$ absorbing systems (adopted from Altay et al. (2011)).

The column density distribution is a useful observational measurement for other types of hydrogen absorbing systems, such as *Lyman-limit system* ($10^{17}\mathrm{cm}^{-2} < N_{\mathrm{HI}} < 10^{20}\mathrm{cm}^{-2}$) and *Damped Lyman-$\alpha$ systems* ($10^{20}\mathrm{cm}^{-2} < N_{\mathrm{HI}}$), but has not been particularly constraining for the forest.

**Brain teaser #4**: The photo-ionization cross-section for neutral hydrogen at the ionization edge (13.6 eV) is $\sigma_{\mathrm{ion}} = 6.3 \times 10^{-18}\mathrm{cm}^2$. Hence, a column density of $N_{\mathrm{HI}} = 1.7 \times 10^{17}\mathrm{cm}^{-2}$ has an optical depth of $\tau_{\mathrm{ion}} = \sigma_{\mathrm{ion}} N_{\mathrm{HI}} = 1$. Never-the-less, Lyman-$\alpha$ absorbers remain ionized almost all the way to Damped Lyman-$\alpha$ systems, $N_{\mathrm{HI}} \gtrsim 10^{19}\mathrm{cm}^{-2}$ ($\tau_{\mathrm{ion}} \sim 500$). Can you explain why?

### 2.4.1 Lyman-$\alpha$ Power Spectrum

Perhaps the most important use of Lyman-$\alpha$ forest in cosmology is in measuring the evolution of the matter power spectrum. Observations of the forest cover a wide redshift range, from $z \gtrsim 2$ to $z \lesssim 5$; since the observed optical depth is well correlated with the gas density, which, in turn, traces the matter density on large scales (above the filtering scale), the observed spectra of the forest contain hidden information about the clustering of matter and its evolution over the redshift range $2 \lesssim z \lesssim 5$.

Measuring the matter power spectrum is exactly the application for which the Fluctuating Gunn-Peterson Approximation (FGPA) is most suitable. In the theory of large scale structure formation there is a theorem that states that if a locally non-linear field is a function of matter density only ($f = f(\rho)$), then on sufficiently large scales the field $f$ is *linearly biased* with respect to the density field, i.e. for sufficiently small $k$

$$P_f(k) = b_f^2 P(k),$$

where the bias factor $b_f$ is independent of $k$. Hence, one can measure the matter power spectrum $P(k)$ in a few simple steps:

1. measure the 1D power spectrum of the transmitted Lyman-$\alpha$ flux, $P_{1D}(k)$, directly from the observed spectra;
2. convert from a 1D to a 3D flux power spectrum,

$$P_F(k) = -\frac{2\pi}{k}\frac{dP_{1D}}{dk};$$

3. determine the flux bias factor, $b_F$, from numerical simulations,
4. and, finally, compute the matter power spectrum

$$P(k) = \frac{P_F(k)}{b_F^2}. \tag{16}$$

Such a program was first completed by Croft et al. (1998) and later repeated many times with better data. For example, the largest set of observed Lyman-$\alpha$ spectra was obtained as part of the Sloan Digital Sky Survey (SDSS), and is shown in figure 9. A little bit of nuisance is that the flux power spectrum is measured in the velocity space, so the units of $k$ in equation (16) are $(\mathrm{km/s})^{-1}$. That makes it hard to compare with other measurements of matter power spectrum without knowing cosmological parameters. But a good piece of news is that the power spectrum grows (or the plotted quantity, $\Delta(k)^2 = k^3 P(k)/2\pi^2$, decreases) with redshift with the rate prescribed by the standard cosmology, so you have not studied your Introduction To Cosmology in vain...

On a more serious note, this measurement provides extremely powerful constraints on the matter power spectrum at the smallest scales - in fact, the forest probes the smallest scales currently accessible to any observational measurement. Many important cosmological and physical studies use these measurements, from

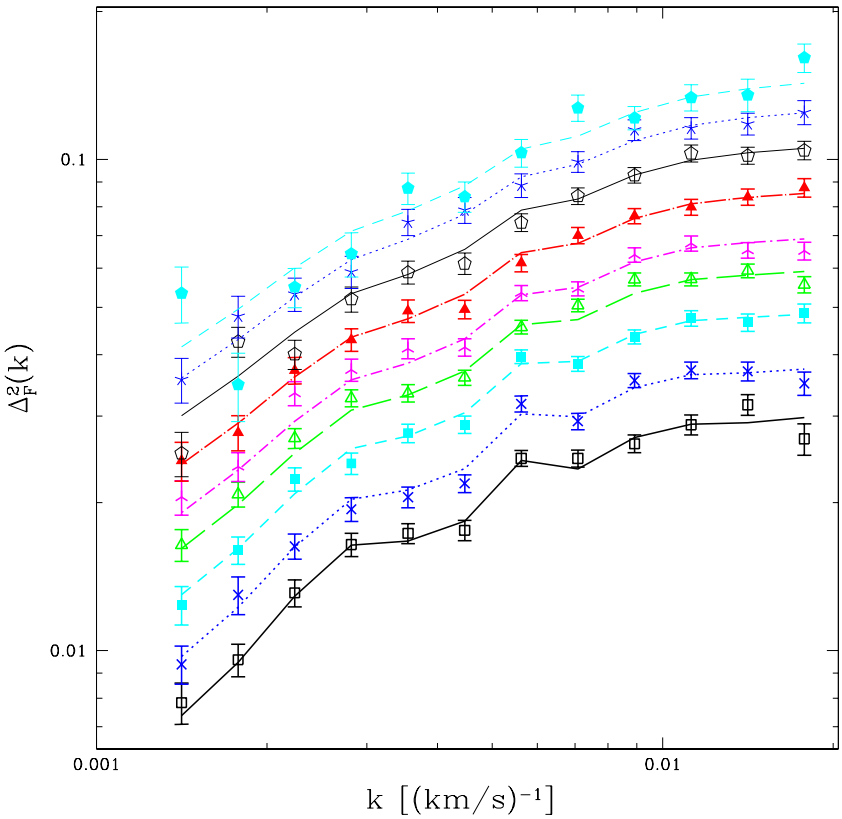



**Fig. 9** Matter power spectra measured from SDSS Lyman-$\alpha$ measurements at a range of redshifts from $z = 2.2$ (bottom) to $z = 4.2$ (top) (adopted from McDonald et al. (2006)).

determining cosmological parameters to constraining neutrino masses (but that is a field I am not going to review in these lectures).

### 2.4.2 Where the Forest Ends

The Lyman-$\alpha$ forest is a small-scale scale counterpart of the large-scale structure - but how small is "small"? In other words, what are the smallest spatial scales on which there is structure in the IGM?

This question is not moot - indeed, the filtering scale tells us where the baryonic fluctuations lag behind the dark matter, but it only applies to linear evolution. The forest is nonlinear, and nonlinear evolution may drive new fluctuations on a variety of scales.

One way to measure structure in any distribution is the, familiar to us already, power spectrum. Using high resolution spectra from 8m-class telescopes one can extend the SDSS measurement to much smaller scales, as is illustrated in figure 10. The decrease in the clustering amplitude is clearly visible at $k > 0.03$ s/km, but is it really the end of the forest? The answer is "unfortunately, no" - unfortunately, because the roll over in the flux power spectra has nothing to do with the actual matter clustering - it is merely an artifact of the thermal broadening of the spectra (the exponential factor in equation 8). Alternatively, one can think of it as the break up of the linear biasing approximation (equation 16).

So, how would one approach the question of studying the smallest scale structure in the forest? One option is offered by spectra of double or gravitationally lensed quasars - if the two quasar images are not too far on the sky, their sightlines probe small spatial scales. Unfortunately, this approach has not been particularly popular among observers - in the only study I am aware of Rauch et al. (2001) demonstrated that, in fact, there is not that much structure in the forest on scales below a kpc. For example, figure 11 shows Lyman-$\alpha$ spectra along two lines of sight to two images of a gravitationally lensed quasar separated by about 0.5 comoving kpc at $z \sim 3$.

Using this measurement, Rauch et al. (2001) placed a strict constraint on the density variation in the forest on small scales,

$$\sqrt{\langle(\Delta \ln \rho)^2\rangle} < 3 \times 10^{-2} \text{ for } \langle \Delta x \rangle = 0.6\,\text{kpc},$$

or, alternatively,

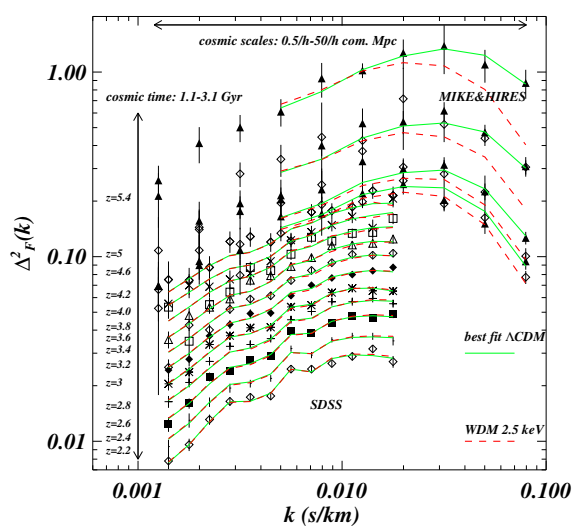



**Fig. 10** Matter power spectra measured from SDSS Lyman-$\alpha$ measurements (as in figure 2) combined with data from high resolution spectra of several quasars (adopted from Viel et al. (2013)).

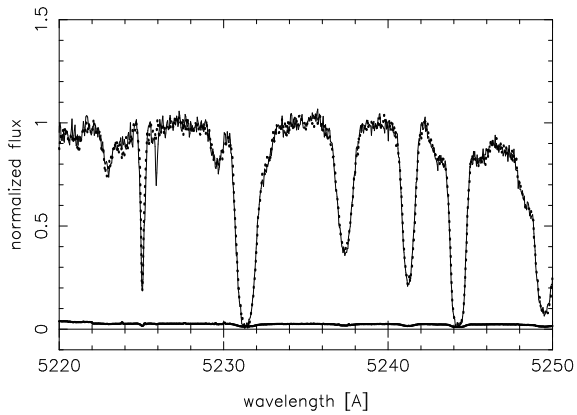



**Fig. 11** Lyman-$\alpha$ spectra along lines of sight to the gravitationally lenses quasar Q1422+231 (images A and C). One spectrum is shown with the solid line, another one with the dotted line beaded with dots. The two spectra are identical to within the observational errors (adopted from Rauch et al. (2001)).

$$\sqrt{\left\langle \left( \frac{\Delta \ln \rho}{\Delta x} \right)^2 \right\rangle} < 0.05\,\mathrm{kpc}^{-1}.$$

A scientifically interesting question is whether the IGM is turbulent on small scales. The Rauch et al. (2001) constraint implies that either the forest is *not* turbulent on these small scales, or that any turbulence that is present is highly sub-sonic (i.e. incompressible). The latter option is possible but is not too likely - density fluctuations in the sub-sonic turbulence scale as Mach number squared, with the flow in the forest becoming transonic at scales $100-200\,\mathrm{kpc}$. If we take a Kolmogorov-like scaling law,

$$\sqrt{\langle (\Delta \ln \rho)^2 \rangle} \approx 1 \left( \frac{\Delta x}{200\,\mathrm{kpc}} \right)^{1/3},$$

then on scale of $0.6\,\mathrm{kpc}$ we find the rms density fluctuation of $\sqrt{\langle (\Delta \ln \rho)^2 \rangle} \approx 0.15$, 5 times higher than the actual observed upper limit. Of course this is not a formal derivation, and factors of several may be lurking here and there, but the estimate serves to demonstrate that the forest is remarkably quiet on scales below a kpc.

**Brain teaser #5**: It is well known in classical hydrodynamics that any flow with Reynolds number in excess of about 1000 becomes turbulent. The viscosity in the IGM is very small, and Reynolds number in the forest is of the order of $10^6$. Hence, the naive expectation is that the IGM must be very turbulent on small scales, but the Rauch et al. (2001) observations suggest it is not. Can you think of an explanation?

## 3 From IGM to CGM

Circumgalactic medium, or CGM, is often understood as the gas within the galactic dark matter halo. I am taking a broader view here, since some of the structures in the universe, like filaments, fall in the border zone between the IGM and CGM, they are not always considered to be part of the Lyman-$\alpha$ forest, but they also are not related to galaxies. They do produce absorption lines in the quasar spectra, but they also stream gas into galactic halos.

### *3.1 Large Scale Structure*

Probably everyone has seen a picture of the large-scale structure of the universe by now (if you have not, check out excellent visualizations of the Millennium simulation at `www.mpa-garching.mpg.de/galform/virgo/millennium`). Since the large-scale structure forms as a result of gravitational clustering from the linear Gaussian fluctuations, it is fully characterized by the linear matter power spectrum. Hence, various scales that we see in the pictures are all related to features in the power spectrum. For example, typical size of voids corresponds to about $1/2$ of the scale at which the power spectrum peaks (which is about $100h^{-1}$ Mpc in comoving units). Hence, in comoving reference frame void sizes do not change - they are as large at $z = 10$ as they are at $z = 0$ (although, these largest voids are, of course, not nearly as empty at $z = 10$ as they are at $z = 0$). Filaments that surround voids are highly non-linear structures and their width is controlled by the the nonlinear scale at each epoch, i.e. the scale at which the amplitude of linear fluctuations reaches unity. Finally, material that makes the largest objects at any time (clusters of galaxies today, galaxies at $z \gg 2$) is assembled from regions roughly the nonlinear scale in size, so masses of these objects are about $4 \times$ (mean density) $\times$ (nonlinear scale)$^3$.

Since our main interest is how gas flows from low to high density regions, the actual motion of matter is of particular importance to us. With time voids become deeper as matter (both dark and gaseous) flows from them onto filaments, and then along the filaments into the galaxies. This pattern of flows is illustrated in figure 12 from a numerical simulation of a local region around the Local Group by Klypin et al. (2003).

As gas flows into a filament from opposite directions, it gets shocked, and the gas temperature is expected to rise above that maintained by photo-heating and adiabatic expansion/contraction - a complication that eventually destroys nice and tight density-temperature relation that exists in the lower density IGM.

The actual structure of the filaments received surprisingly little attention in the literature. In a classical review Shandarin & Zeldovich (1989) showed the profiles of one-dimensional collapse onto a 2D pancake (figure 13). Collapse onto a 1D filament is qualitatively similar, because the physics is the same - gas gets piled up at the center, where the entropy is the lowest, while the dark matter from each side flows through the upcoming stream, creating density caustics on the outside.

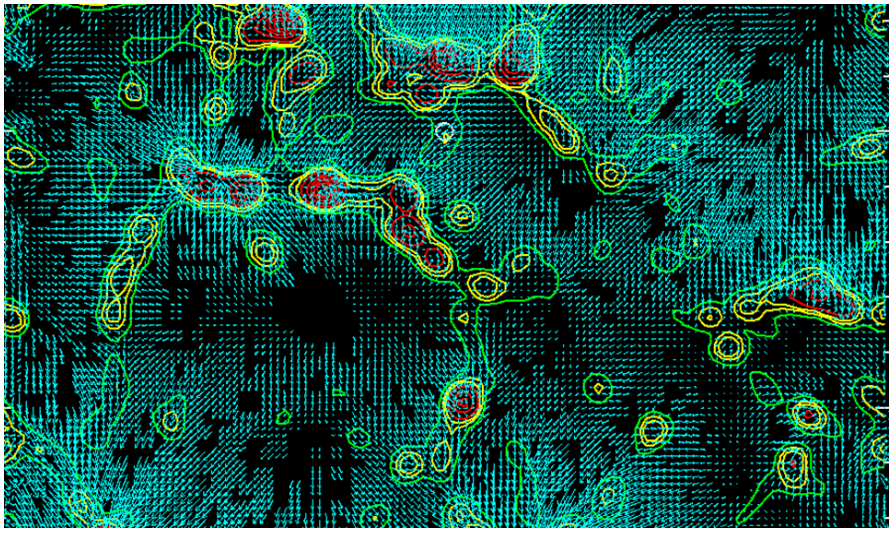


**Fig. 12** Large-scale flows (cyan arrows) on top of the density contours (green, yellow, and red). Flows of matter onto (almost all of the) filaments are clearly visible in this visualization of a numerical simulation (adopted from Klypin et al. (2003)).

What happens next is determined by whether the filaments are self-gravitating - but since they have widths comparable to the nonlinear scale, we know that they, on average, *are* self-gravitating. In a self-gravitating filament the dark matter will stop streaming, turn around, and fall on itself once again and again, increasing the number of intersecting streams as time goes on.

In order to illustrate the large- (and not-so-large-) structure further, I will use a cosmological simulation from Gnedin & Kravtsov (2010). This simulation is not very big, and focuses on the environs of a single, Milky-Way like galaxy, but it will

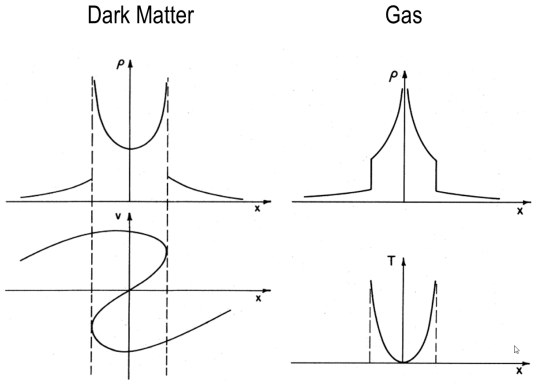



**Fig. 13** Profiles of one-dimensional collapse of dark matter and gas onto a 2D pancake (adopted from Shandarin & Zeldovich (1989)).

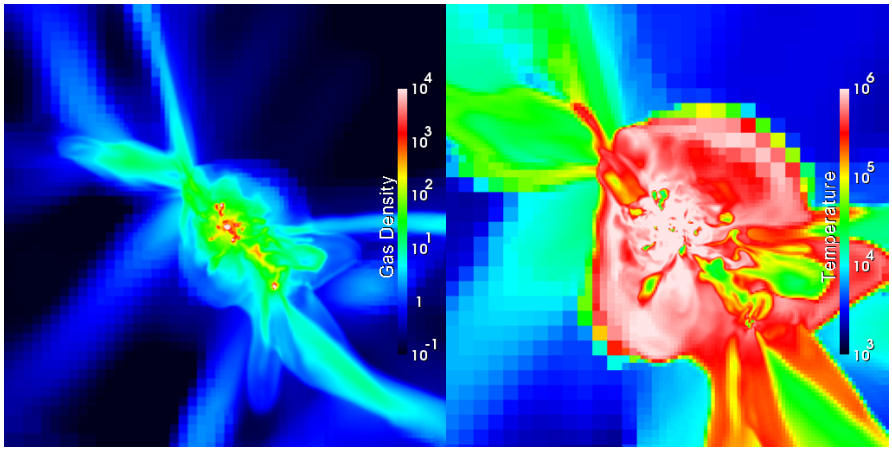



**Fig. 14** Thin slices through cosmological simulation that show the gas density (left) and the gas temperature (right) around a typical galaxy at $z \approx 2$.

suffer for our purpose. Figure 14 shows the gas density and the gas temperature around the main galaxy at $z = 2$.

There are a few features to note. First of all, the gas filaments do appear to be denser and cooler in the middle, similarly to the 1D collapse. Second, in the temperature plot we see really hot (million degrees) gas. Most of that hot gas is concentrated around the galaxy, in the dark matter halo and beyond, but some of it extends way into the filaments - those are the temperature spikes that we see in figure 13.

### *3.2 How Gas Gets Onto Galaxies*

Everyone knows that dense enough regions of the large-scale structure will collapse and *virialize* (i.e. reach, or, at least, approach, the virial equilibrium). The simplest model of such collapse is a *top-hat*,

$$\rho(\mathbf{x}) = \begin{cases} \bar{\rho}(1+\delta_i), & r < r_i \\ \bar{\rho}, & r > r_i \end{cases}$$

where $r_i$ and $\delta_i$ are the initial radius and amplitude of the perturbation. The overdense perturbation collapses, and the evolution of the radius of the perturbation can be solved analytically in the matter-dominated regime ($a \propto t^{2/3}$), albeit parametrically with a parametric variable $\theta$:

$$r = \frac{GM}{\delta_i \dot{r}_i^2}(1 - \cos\theta),$$

$$t = \frac{GM}{\delta_i^{3/2} \dot{r}_i^3}(\theta - \sin\theta).$$

The moment of collapse is defined as $r = 0$, which occurs at the time when $\theta = 2\pi$. A remarkable property of the top-hat solution is that at the moment of collapse $t_f$ the linear density fluctuation

$$\delta_L(t) = \frac{D_+(t)}{D_+(t_i)}\delta_i$$

is just a number, independent of the initial overdensity, size, or the mass of the overdense region,

$$\delta_L(t_f) = \frac{3}{5}\left(\frac{9\pi^2}{4}\right)^{1/3} = 1.69.$$

A perturbation cannot collapse to a point - that would be even less likely than making a pencil stand on a sharp end. A standard assumption is that the collapsing perturbation virializes - i.e. reaches the virial equilibrium - at around the time $t_f$. In that case the average overdensity $\delta_v$ of the final virialized object is $1 + \delta_v = 18\pi^2 \approx 178 \approx 180 \approx 200$.

The virial radius of the dark matter halo in figure 14 is roughly the green roundish region in the density panel (overdensity $\gtrsim 100$), while the million-degree gas extends well beyond it. The virial radius serves as a good approximation of a boundary beyond which any, even imaginable, resemblance of spherical symmetry totally vanishes! As gas falls into potential wells of dark matter halos, it gets shocked and heated to around the virial temperature (also deviations can easily be a factor of 2-3 in each direction). Shocks never stand still (in the reference frame of the gas behind them), so the accretion shock propagates outward. For typical cosmological objects, be it star-forming galaxies at $z \sim 2$ or galaxy clusters at $z = 0$ (or anything in between), it is not uncommon to find the accretion shock extending to 3 virial radii. Since it goes so much beyond the quasi-spherical region, it is highly asymmetric and non-spherical, with some of its protrusions reaching well into voids, up to $\sim 10$ virial radii, while along filaments the accretion shock may not even exist (or do not reach to even a modest fraction of the virial radius).

### *3.3 Cool Streams*

A story of the ”cold streams” is a real-life story of an elephant-in-the-room. In a gesture of non-conformity, I am going to call them ”cool streams”, because in the ISM-speak (which we are going to use for the most of this course) the term ”cold” refers to truly cold gas, below 100 K. Strictly speaking, they should be called ”warm streams”, since $10^4$ K gas is ”warm” in the ISM-speak, but that would confuse too many people...

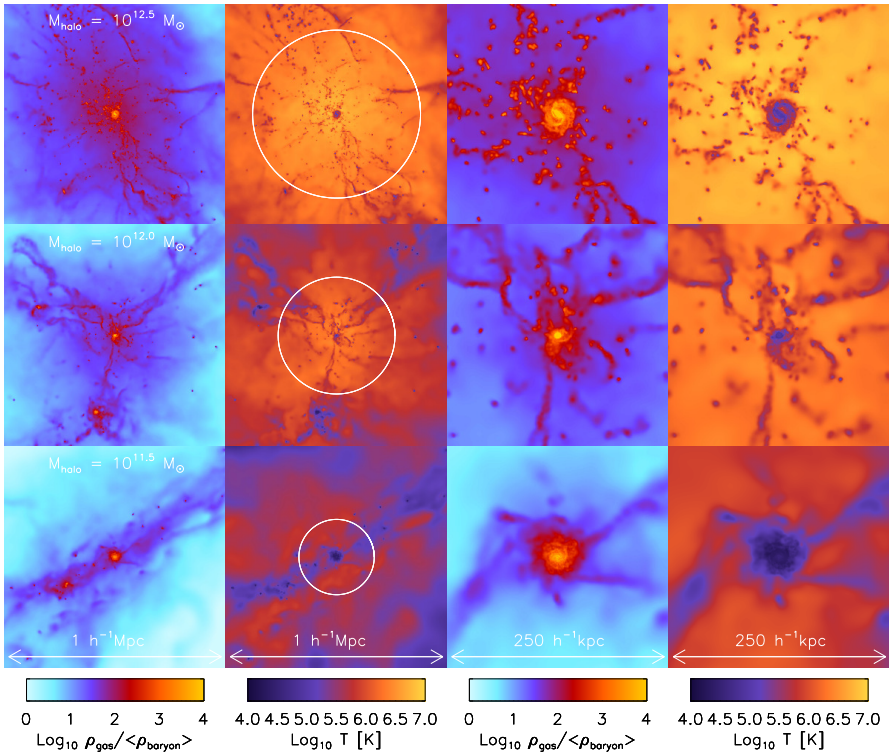



**Fig. 15** Density and temperature images of galaxies of different masses at $z \approx 2$. Cool streams are clearly visible in temperature images as blue blobs and filaments (adopted from van de Voort et al. (2011)).

Every practicing simulator knew about cool streams, but no one paid any attention to them until in 2005 in an influential paper Kereš et al. (2005) showed that at intermediate redshifts - the epoch where galaxies make most of their stars - cool streams deliver significant, or even dominant, fraction of gas onto the galactic disks, where stars actually form. Hence, from the point of view of a galaxy as a gas consumer, cool streams are the primary consumption channel.

Examples of cool steams in cosmological simulations from Overwhelmingly Large Simulation project (OWLS, van de Voort et al., 2011)) are shown in figure 15. As in a weird monster movie, the blue "tentacles" of cool gas try to reach the central galaxy; they break up into individual blobs for a massive one ($M = 10^{12.5}\,\mathrm{M}_\odot$), remain as thin streams for a $M = 10^{12}\,\mathrm{M}_\odot$ one, and completely swamp gas accretion for a Milky-Way type galaxy ($M = 10^{11.5}\,\mathrm{M}_\odot$ at $z = 2$). Images like that can be made from almost any cosmological simulation, and from any modern simulation code, be it an SPH code, an AMR, or a moving mesh code like AREPO[1](Springel, 2010). All simulations agree that the cool flows dominate the gas accretion for halos

[1] For these and other curious abbreviations check out Volker Springel's lectures in this volume.

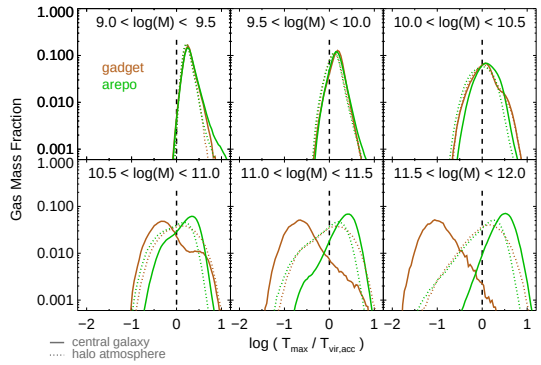



**Fig. 16** Temperature distribution function for galaxies in different mass bins simulated with GADGET and AREPO. The two codes predict significantly different distributions for high mass galaxies (adopted from Nelson et al. (2013)).

above about $M = 10^{11.5}\,\mathrm{M}_\odot$ , with this mass being only weakly (if at all) redshift dependent.

Since most of the gas accretion occurs in low mass galaxies at all times, most of gas that ends up in galactic disks enters the halo as ”cool” - significantly below the virial temperature, but it may still be well above the ”ISM warm” of $10^4$ K - at all cosmic times up to the present epoch. The contribution of cool streams is, however, diminishing with time, so by $z = 0$ they, on average, only deliver about half of the accreting gas onto galactic disks.

A happy concordance is broken, however, when the fate of cool streams inside the halo is explored further. In a recent study, a carefully designed comparison between GADGET (Springel, 2005) and AREPO (Springel, 2010) codes found some disturbing differences (Nelson et al., 2013, shown in figure 16). The two codes have the same gravity and dark matter solvers, but differ in the way gas dynamics is treated (for details, check Volker Springel's lectures in this volume). While in the SPH GADGET simulation the cool streams remain cool inside the halo and reach all the way to the galactic disk, in the mesh-based simulation with AREPO the cools streams heat up as they approach the disk. This discrepancy reflects the well-known dichotomy between SPH and mesh codes - the former do not have enough diffusion (without special fixes), while the latter may have too much numerical diffusion, especially in the poorly resolved regions. Which of the two codes is closer to reality is not yet clear; the progress in this field, though, happens at a relativistic speed, so as you are reading these lectures, the ambiguity may have been already resolved.

## *3.4 Galactic Halos*

Few sane people doubt the existence of dark matter halos. Whether galaxies have gaseous halos is an entirely different matter.

Cosmological simulations generically predict that galaxies like the Milky Way (MW) should be surrounded by hot gaseous halos in quasi-virial equilibrium. For

two decades the actual existence of these hot halos was an even more hotly debated topic. The point of contention was the simple fact that hot gas emits X-rays, hence hot halos must be detectable in X-rays. The cruel reality is that the halo gas is rather tenuous, and for galaxies like the Milky Way it is expected to have temperatures that are very hard to detect observationally.

How much gas one expects to reside in the Milky Way halo actually depends on the halo mass, which has been notoriously difficult to estimate. Proposed values range from $\sim 7 \times 10^{11}\,\mathrm{M}_\odot$ to over $2 \times 10^{12}\,\mathrm{M}_\odot$ (values outside this range are considered extremist and will provoke a French military intervention or an American bombing campaign). For the fiducial value of $10^{12}\,\mathrm{M}_\odot$ the cosmic share of baryons in the MW is $1.6 \times 10^{11}\,\mathrm{M}_\odot$. The stellar mass of the MW is about $6 \times 10^{10}\,\mathrm{M}_\odot$ (although, values up to $8 \times 10^{10}\,\mathrm{M}_\odot$ are sometimes used) and the disk gas mass is $\lesssim 1 \times 10^{10}\,\mathrm{M}_\odot$. Hence, the gaseous halo may contain up to $10^{11}\,\mathrm{M}_\odot$ (it may, of course, be much less if some of the gas is expelled from the Galaxy by stellar feedback and other energetic processes).

The contention about the existence of the hot halo finally has been resolved by *Chandra* - not the brilliant man who resolved so many other contentions, but the remarkably successful space mission named after him. In a ground-breaking observation the Chandra team finally detected the X-ray emission from the hot gas around the Milky Way (Gupta et al., 2012). While measuring the total mass of the halo from Chandra observations is very challenging (try measuring the mass of a giant monster that swallowed you), the limits that the Chandra team has been able to place on the gas mass in the halo are consistent with our estimate of $10^{11}\,\mathrm{M}_\odot$.

X-ray detection of the halo is important, because it is a *direct* evidence for the existence of a massive (from the point of view of the disk) gaseous halo. Historically,

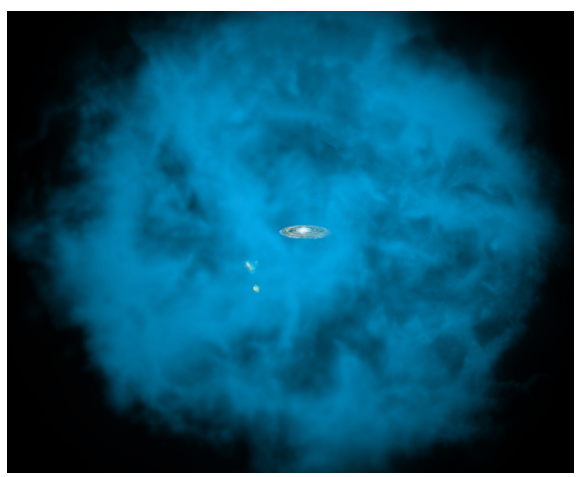


**Fig. 17** Radial density profile at $z = 0$ of the same galaxy shown in figure 14. Blue and red solid lines shows the actual simulated profiles of dark matter and gas, while dashed lines give the best-fit NFW (for dark matter) and rescaled by $0.05 f_b$ NFW (for gas) profiles respectively. The red dotted line is the best-fit beta profile for the gaseous halo. Filled and open symbols are pre-Chandra observational constraints (Weiner & Williams, 1996; Blitz & Robishaw, 2000; Grcevich & Putman, 2009; Quilis & Moore, 2001; Stanimirović et al., 2002; Anderson & Bregman, 2010).

however, a large number of indirect constraints existed that all pointed out towards the same conclusion. In figure 17 I show the $z = 0$ dark matter and gas profiles for the same galaxy we met in figure 14. The hot halo (solid red line) in that simulation is consistent with the existing pre-Chandra observational constraints as well as with the actual Chandra measurement. What is remarkable is that in the simulation all stellar feedback processes were switched off (see Gnedin, 2012, for details about the actual simulation). The galactic disk in the simulation is overly massive and has incorrect density profile, but the halo seems to be ok (at least within the precision of observational constraints). There is, actually a simple reason for it - the main physical process that matters for the gas in the halo is radiative cooling, it is cooling that determines which gas can rain on the disk and which remains in the halo in the hot phase.

Hence, the physics of radiative cooling is our next stop.

## *3.5 Diversion: Cooling of rarefied gases*

Before we proceed further along our yellow brick road, let's step aside for a short while and consider how cosmic gas cools - the process we have already met in the IGM segment of our journey, and which we will be meeting over and over again in the future.

*Radiative cooling* is an "umbrella" name for diverse physical processes through which gas transforms its thermal energy into the radiation that leaves the system. At low enough density, three processes dominate, and all three of them involve a collision of a free electron or an atom/ion with a neutral atom or a partially neutral ion. These three processes are

**line excitation**: a collision excites the neutral atom into a higher energy state, the state decays and the resultant photon leaves the system;

**collisional ionization**: a collision ionizes the neutral atom and the binding energy of the freed electron is charged against the thermal energy account;

**recombination**: an ion captures a free electron and the sum of the kinetic energy of the electron and the binding energy of the neutral atom is emitted as a photon.

All these collisional processes depend on the square of the density, so it is convenient (and customary) to factor our that density dependence explicitly in the cooling rate of the gas,

$$\left.\frac{dU}{dt}\right|_{\rm cool} = -n_b^2 \Lambda(T, ...),$$

where $n_b$ is the number density of baryons (I prefer it to another commonly used parametrization that factors out the hydrogen nucleus number density $n_{\rm H}$, because $n_b$ is directly proportional to the gas mass density for any value of helium abundance or gas metallicity) and $\Lambda$ is commonly called a *Cooling Function*.

In the simplest case of gas in pure collisional equilibrium (no external or internal radiation of any kind - the so-called *collisional ionization equilibrium*, or CIE) the

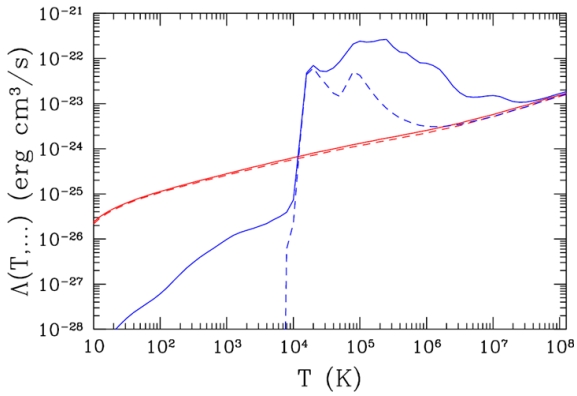



**Fig. 18** Cooling functions for the primordial gas ($Z = 0$, dashed lines) and the gas at solar metallicity (solid lines). Blue lines show the "standard" CIE cooling functions, while red lines show the cooling functions for the fully ionized gas (the only cooling process is Bremsstrahlung).

cooling function is called "standard". If the relative abundance of various chemical elements is fixed and small variations in the helium abundance are neglected, the the cooling function only depends on the gas temperature $T$ and the total metallicity $Z$,

$$\Lambda_{\rm CIE} = \Lambda_{\rm CIE}(T,Z).$$

Examples of this function for $Z = 0$ and $Z = Z_\odot$[2] are plotted in figure 18.

The specific shape of the CIE cooling function, with its "bumps and wiggles", is determined by the interplay between contributions of over a dozen various chemical elements. A good recent review is given by Wiersma et al. (2009), an illustration from which is reproduced here in figure 19. In particular, one has to be aware that many of the atomic cooling rates used to construct the cooling function are know rather poorly, not better than a factor of 2, and that uncertainty propagates into the actual value of the cooling function. In realistic galactic and cosmological simulations this uncertainty is often, however, unimportant: the cooling time-scale is so much shorter than any other physical time-scale in the problem that it does not need to be known very precisely (all gas that can cool will indeed cool rapidly).

Wiersma et al. (2009) paper offers another, much more important lesson, though. As they show, the actual cooling function in the IGM, CGM, and even ISM of galaxies at low and high redshifts may deviate from the "standard" one quite substantially. In other words, the "standard" CIE cooling function is actually highly non-standard and is almost never realized in nature. The reason for that is that low density cosmic gas is always affected by external radiation field.

Figure 20 is a simple illustration of this process. Cooling (and heating) processes in a gaseous halo can be modified in a major way if it straddles too close to a strong source of ionizing radiation, such a bright quasar. Within 1 Mpc from the quasar, the equilibrium temperature in the halo goes all the way up to 200,000 K, twenty times above our "canonical" 10,000 K.

[2] Throughout these lectures I define "solar metallicity" as the metallicity of our galactic neighborhood, $Z_\odot = 0.199$ in absolute units, rather than metallicity of an average-looking single star somewhere in the outskirts of the Galaxy.

So, let us review the cooling function from the very beginning, this time being careful. In a most general case in addition to cooling there is also radiative heating by the radiation field. Hence, the change of the gas internal energy due to radiative processes has two terms with opposite signs,

$$\left.\frac{dU}{dt}\right|_{\text{rad}} = n_b^2\left(\Gamma - \Lambda\right),$$

where $\Lambda$ is our old acquaintance the cooling function and $\Gamma$ is the *heating function*. Both of them depend on a multitude of parameters,

$$[\Gamma, \Lambda] = \mathscr{F}\left(T, n_b, X_{ijl}, J_\nu, \tau_{ijl}\right), \tag{17}$$

where the density dependence reappears because not all processes are two-body, $X_{ijl}$ is the abundance of the chemical element $i = \text{H}, \text{He}, \ldots$ in the ionization state state $j = \text{neutral}, \text{single ionized}, \ldots$ in the quantum state $l$, $J_\nu$ is the spectrum of the incident radiation field that shines on a given (formally infinitesimally small) parcel of gas, and $\tau_{ijl}$ are opacities in each radiative transition (gas may be optically thick to some of its own cooling radiation if our parcel is embedded deep inside a huge cloud). For the sake of brevity in notation, we will use $\mathscr{F}$ to label either $\Gamma$ or $\Lambda$, since both functions always depend on the same set of arguments.

In order to compute the cooling and heating functions in such a detail one needs a highly sophisticated computer code that, in its complexity, rivals modern cosmological simulation codes. Fortunately, such codes exist, and the most famous and widely used of them is *Cloudy*[3]. Conceived by Gary Ferland from the University of Kentucky and contributed to by many people, *Cloudy* is freely available from its website, `nublado.org`, and is well-documented for a fast start-up curve.

There is one problem only with *Cloudy* - it is way too complex to be used in modern simulation codes for computing cooling functions "on the fly". Perhaps in the future, in the era of exa-scale computing, it will be possible to run *Cloudy* as

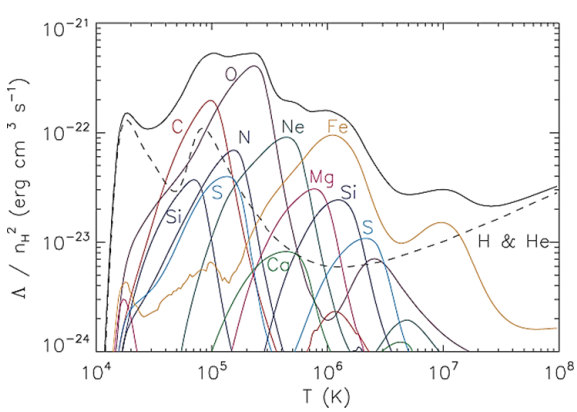



**Fig. 19** Contributions of individual chemical elements to the "standard" CIE cooling function (adopted from Wiersma et al. (2009)).

[3] Notice the convention, *Cloudy* is a name, not an abbreviation.

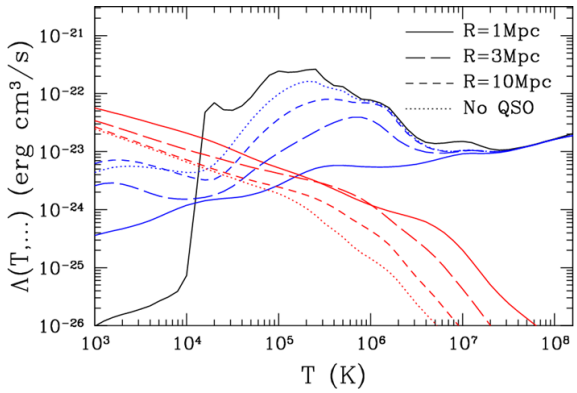



**Fig. 20** Illustration for the role of radiation field in suppressing cooling (blue lines) and enhancing heating (red lines). A gas in the galactic halo (at density 340 over the cosmic mean) is shined upon by the $10^{12}L_{\odot}$ quasar. Sufficiently close, the quasar radiation modifies the cooling and heating functions in a major way.

a "sub-grid" model in real simulations, but for now we need to seek approximate short-cuts.

So, what one can do? Unless densities are very high (hence our focus on *low density gas*), the gas will be optically thin to its own cooling radiation, so the dependence of cooling and heating functions on $\tau_{ijl}$ disappears - for this to be exactly true, we also should exclude all cooling and heating processes due to molecules, since those always require radiative transfer to be followed properly. Thus, if you need to follow molecular cooling/heating as well, you will have to add them "manually" to the cooling and heating functions that we discuss below.

Second, in almost all galactic and cosmological simulations the assumption of the *ionization and excitation equilibrium* is not a bad one. In the ionization equilibrium the distribution of a given chemical element over various ionization states is uniquely determined by density, temperature, and the radiation field. The same is true about various quantum levels in the local thermodynamic equilibrium. If, in addition, we assume that relative abundances of chemical elements are fixed (say, to the solar abundance pattern), then the dependence on $X_{ijl}$ reduces to the simple dependence on the overall gas metallicity $Z$,

$$\mathscr{F} = \mathscr{F}(T, n_b, Z, J_\nu). \tag{18}$$

Very often this latter expression is what actually called a "cooling/heating function". But even the latter form is unusable in modern simulations codes, because it includes an explicit dependence on the radiation spectrum, which is an arbitrary function of frequency (in a strict mathematical sense $\mathscr{F}$ in equation (18) is actually an *operator*, not a function). Hence, we still need to account for that dependence in an approximate manner.

One particular short-cut has been used in many cosmological codes for over a decade. Wiersma et al. (2009) paper again serves as a good reference, although the first known (to me) example of such approach is used by Kravtsov (2003). In the most of the volume of the universe the dominant source of external radiation is the cosmic background that we already met in the previous chapter. The cosmic back-

ground is uniform in space and is a function of the cosmic redshift only, hence in the limit when $J_\nu$ can be approximated by the cosmic background, cooling and heating functions become functions of 4 arguments, temperature, density, gas metallicity, and cosmic redshift, and hence can be easily tabulated and used in simulation codes efficiently via a simple table look-up.

Unfortunately, most of the volume in the universe contains only a modest fraction of the mass, and even smaller fraction of action. The radiation field in the ISM (and, in at least part of the CGM) of galaxies is dominated by local radiation sources (for example, the UV radiation field in the solar neighborhood is 500 times higher than the cosmic background; at the center of the galaxy that ratio jumps to 5,000). Since stars form in the ISM, any galactic or cosmological simulation that attempts to model star formation cannot use cooling and heating functions which only account for the cosmic background.

How one can attempt to construct a more accurate short-cut? After all, the effect of external radiation is in ionizing some of the chemical elements and/or exciting particular levels, and ionization and excitation rates are all integrals over the radiation spectrum with some cross-sections, which are broad and relatively slowly varying functions. Let's imagine the following thought experiment: we take a given spectrum and increase the radiation intensity in a narrow frequency bin between some $\nu_0$ and $\nu_0 + \Delta\nu$. If the increase is large, the cooling and heating functions will be affected. Now shift the frequency bin to $\nu_0 - \Delta\nu$ to $\nu_0$. Most of ionization and excitation rates will be barely affected (unless we choose $\nu_0$ very carefully to correspond exactly to the ionization/excitation threshold of an important cooling channel), since cross-sections of most physical processes will not change significantly between the two narrow bins. Hence, in order to compute the cooling and heating functions accurately, we do not need to know the radiation spectrum in excessive detail (say, in hundreds of frequency bins), but it may be sufficient to describe it by several "broadband filters".

There can be infinitely many choices for these filters. In a specific implementation of this idea, Nick Hollon and I decided to use photoionization rates of several chemical elements as "broadband filters". After all, the ionization balance is controlled by photoionization rates, so it makes sense from the atomic physics perspective. We have explored over 20 various chemical elements and their ionization states, and the best approximation that we have been able to come up depends on just 4 ionization rates (Gnedin & Hollon, 2012).

Specifically, we adopt the approximation in which the metallicity dependence of the cooling and heating functions is expanded into the Taylor series in gas metallicity,

$$\mathscr{F}(T, n_b, Z, J_\nu) = \sum_{i=0}^{n} \left(\frac{Z}{Z_\odot}\right)^i \mathscr{F}_i(T, n_b, J_\nu), \tag{19}$$

with $n = 2$ providing a highly accurate approximation for $Z < 5Z_\odot$. Each of the expansion coefficients is approximated as

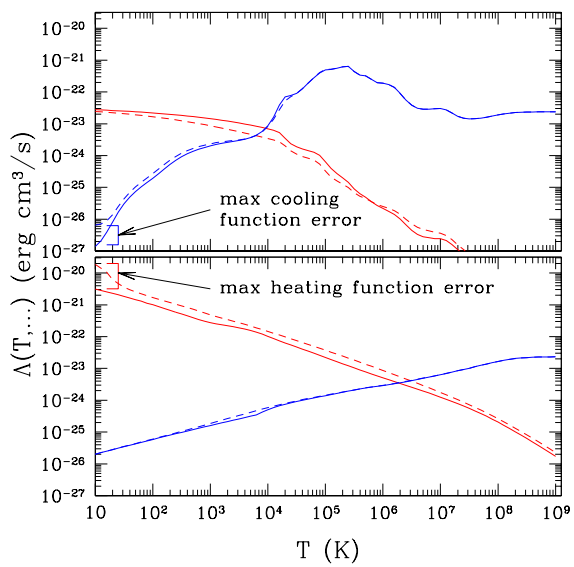



**Fig. 21** Cooling (blue lines) and heating (red lines) functions for our test models that maximize the error in the cooling function (top panel) and the heating function (bottom panel). Approximate functions from equations (19-21) are shown as dashed lines, while exact calculations from Cloudy are shown with solid lines.

$$\mathscr{F}_i(T, n_b, J_\nu) \approx \mathscr{F}_i(T, \{r_j\}, n_b), \tag{20}$$

with several parameters $r_j$ encapsulating the full dependence of the cooling and heating functions on the external radiation field.

A parameter set that we found to work well is defined as follows:

$$\begin{aligned} r_1 &= \frac{P_{\rm LW}}{n_b}, \\ r_2 &= \left(\frac{P_{\rm HI}}{P_{\rm LW}}\right)^{0.353} \left(\frac{P_{\rm HeI}}{P_{\rm LW}}\right)^{0.923} \left(\frac{P_{\rm CVI}}{P_{\rm LW}}\right)^{0.263}, \\ r_3 &= \left(\frac{P_{\rm HI}}{P_{\rm LW}}\right)^{-0.103} \left(\frac{P_{\rm HeI}}{P_{\rm LW}}\right)^{-0.375} \left(\frac{P_{\rm CVI}}{P_{\rm LW}}\right)^{0.976}, \end{aligned} \tag{21}$$

where $P_{\rm LW}$ is the rate of photo-destruction of molecular hydrogen (molecules are excluded from the cooling and heating functions, since they cannot be treated without radiative transfer, so we use $P_{\rm LW}$ just as a convenient "broadband filter" for the radiation below the hydrogen ionization threshold) and $P_{\rm HI}$, $P_{\rm HeI}$ and $P_{\rm CVI}$ are photoionization rates of HI (ionization edge of 1 Ry), HeI (ionization edge of 1.8 Ry), and CVI (ionization edge of 36 Ry). These rates sample a large range of photon energies, and serve as a good set of more-or-less independent "broadband filters".[4]

[4] They are not fully independent, of course - a photon ionizing CVI can also ionize neutral hydrogen, but it is convenient to use photoionization rates rather that some other, arbitrary filter shapes, since the same rates can be useful in the simulation code for other purposes - for example, for computing the ionization balance of hydrogen, helium, or other chemical elements.

The main problem with approximation (19-21) is that it occasionally results in ”catastrophic errors” - for example, if you choose the radiation field, gas temperature, density, and metallicity at random, in about 1 case out of the million the approximate cooling or heating function will deviate from the actual Cloudy calculation by a factor of several (that is a consequence of not being able to fully represent all possible variations in the radiation field by just 3 coefficients $r_j$, $j = 1,2,3$). Figure 21 demonstrates the worst-case catastrophic error of the approximation.

The good news is that these catastrophic errors occur for either highly implausible or completely irrelevant values of parameters - for example, the large error in the heating function at $T \sim 10\,\mathrm{K}$ in the bottom panel of figure 21 is not very important because the heating function there is much larger than the cooling function, and the equilibrium temperature (blue and red lines cross) is $T_{\mathrm{eq}} \approx 2 \times 10^6$ K. If the gas at 10 K finds itself suddenly in such conditions, it will be heated to above million Kelvins rapidly, quickly leaving the parameter space where the approximation is inaccurate.

Similarly, the large error in the cooling function at $T \sim 10\,\mathrm{K}$ in the top panel of figure 21 is irrelevant, because the heating function in those conditions is more than 3 orders of magnitude larger than the cooling function, hence it is not important to know the cooling function at all.

Undoubtedly, a better approximation for the cooling and heating functions is possible, but in the absence of such, equations (19-21) provide a practical way to fully account for the effect of the radiation field in modern cosmological and galactic simulations.

### *3.6 Back to Galactic Halos*

Armed with the understanding of the cooling and heating functions, we can now return to the fate of gas in galactic halos. As gaseous halos are expected to become denser at the center, the cooling time will decrease towards the center. Hence, there must exist a *cooling radius* $R_C$ at which the cooling time is equal to the age of the halo. Gas inside $R_C$ is able to cool efficiently and condense towards the halo center, while the gas outside $R_C$ cools too slowly and will remain in the (quasi-) hydrostatic equilibrium.

A detailed analysis of the cooling process is well presented in Maller & Bullock (2004), although they were not the first group who considered that process. In figure 22, adopted from that paper, the final profile of the hot gas is shown with the dashed red line. The density profile is cored - all the gas above some threshold density is able to cool, and the core density is set by the requirement that the cooling time in the remnant of the core gas is longer than the age of the halo.

The gas that is able to cool will stream towards the center and will settle into a galactic disk. It can do that, however, in two distinct ways: it can either develop a cooling flow and smoothly flow in a quasi-spherical way all the way to the center, or it can experience thermal instability, split into individual dense clouds, which then

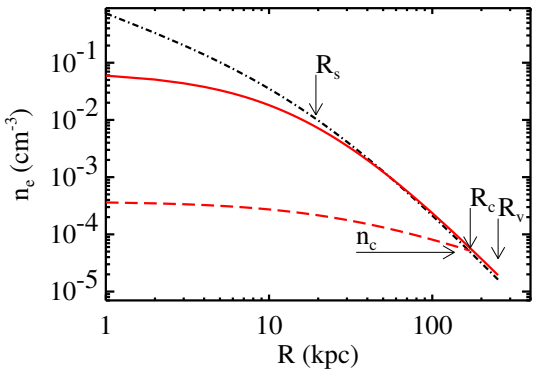



**Fig. 22** Density profiles of the hot phase of halo gas in Maller & Bullock (2004) model in the absence of cooling (solid red line) and with cooling properly accounted for (dashed red line) for a Milky Way like galaxy at $z = 0$. The dot-dashed black line shows the NFW profile (adopted from Maller & Bullock (2004)).

fall onto the disk along parabolic orbits like rain drops fall on the ground. Which of these two ways dominates is still a completely open question, with the observational evidence being sparse and inconclusive.

Clouds of neutral hydrogen (hence dense and cool) are indeed detected in the halo of the Milky Way, they are commonly known as "high velocity clouds" (HVC), since they are detected in radio observations as neutral hydrogen at velocities significantly offset from the gas in the galactic disk (clouds that are not offset in the velocity would not be distinguishable from the disk itself).

For example, the recent GALPHA-HI survey by Arecibo telescope uncovered a large number of new clouds (Saul et al., 2012, as shown in figure 23). Unfortunately, from the radio observations alone it is very hard to determine the distances to those

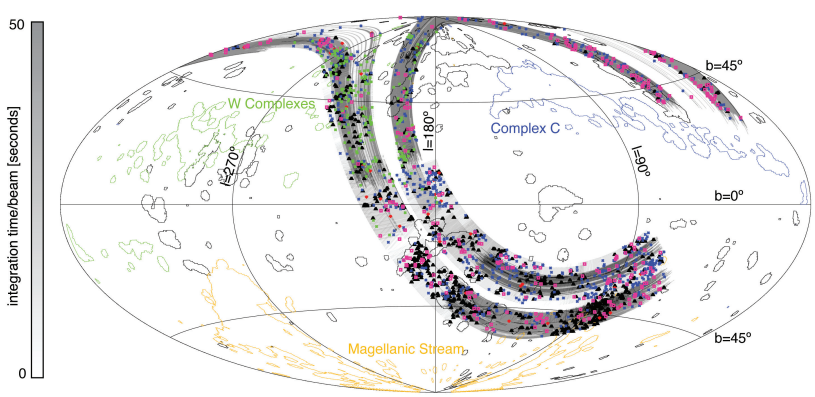



**Fig. 23** Neutral hydrogen clouds in the halo of the Milky Way discovered by the GALPHA-HI survey. Various colors mark the cloud type, with HVC plotted in black (adopted from Saul et al. (2012)).

clouds. Perhaps, they are not located in the halo but form the so-called ”galactic fountain”, with the gas being thrown up by stellar feedback.

One way to resolve the ambiguity is to search for high velocity clouds in external galaxies. Alas, even in our neighbor Andromeda galaxies none have been found. Andromeda is sufficiently far away for sufficiently small clouds to remain undetected, so the jury is still out on whether HVCs are indeed the halo gas raining onto the galactic disks or the disk gas pushed (temporarily) into the halo.

One part of the problem is that the 21 cm line that is used to detect neural hydrogen in radio observations is one of the weakest lines in this universe. Neutral hydrogen also has one of the strongest lines - Lyman-$\alpha$. However, it is not easy to excite $n = 2$ level in the hydrogen atom, hence Lyman-$\alpha$ is usually seen in absorption.

That is where other chemical elements come to rescue. Even while we are primarily after hydrogen, a trace amount of heavy elements may produce enough absorption in some of their, more easily excitable and observable lines. One such element is Magnesium, the ionization threshold of its singly ionized state is just 15 eV, very close to the hydrogen ionization threshold of 13.6 eV. Because of that, MgII has been used as a proxy for neutral hydrogen in absorption studies of galaxies for several decades. Figure 24 shows a plot from a recent compilation of observational constraints in several ions by Chen (2012). A general feature of all observations is that MgII drops precipitously further away that about 100 kpc from a galaxy (with a mild dependence on the galaxy luminosity). It is highly tempting to associate this drop with the cooling radius for the halo, and MgII with the cool clouds formed by thermal instability, but in the absence of additional evidence such a proposition will remain no more than a plausible conjecture.

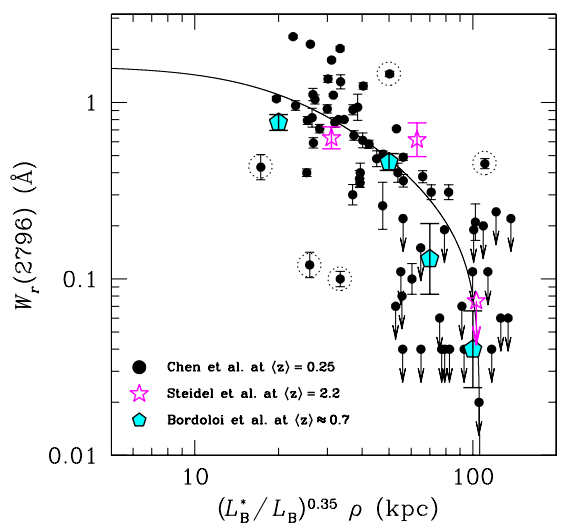



**Fig. 24** Strength of MgII absorption as a function of distance from he host galaxy. There is a sharp drop in absorption for distances in excess of about 100 kpc, probably indicating the cooling radius for halos (adopted from Chen (2012)).

One way or the other the gas from the halo (and beyond) ends up in the galactic disk, making up the Interstellar Medium (ISM) of galaxies. This is where our yellow brick road leads us next.

# 4 ISM: Gas In Galaxies

The field of Interstellar Medium takes easily a quarter of all of Astronomy. Any attempt to review it at any reasonable level will result in me still writing these lectures on my deathbed. Hence, our journey through the ISM realm will be brief and highly focused - we will be mainly concerned with "gas in galaxies", i.e. gas as a medium (forget about chemistry, except for one very specific topic), and gas as a galaxy component (i.e. not small-scales behavior of gas, but rather the role of gas as a citizen of a galaxy). Even with these restrictions, the journey that lays ahead is extremely biased towards my own research interests and topics I find fascinating.

## *4.1 Galaxy Formation Lite*

Galaxies are rather complex creatures; understanding galaxy formation and evolution is the current frontier of extragalactic astronomy and cosmology. Never-theless, the basic sketch of how galaxies form and evolve has been developed - it is captured by the Mo et al. (1998) model (hereafter MMW98).

The cornerstone assumption of MMW98 model is that the cool ($\sim 10^4$ K) gas is delivered to the bottom of the potential well of a dark matter halo - either by radiative cooling in the halo or by inflow along cool flows. The specific way by which gas is delivered is unimportant; what matters is that the angular momentum is conserved, and hence the cool gas settles into a rotationally-supported disk.

It is convenient to parametrize the mass of the disk $M_d$ as a fraction $m_d$ of the halo mass $M_h$,

$$M_d = m_d M_h,$$

and the disk angular momentum $J_d$ as a fraction $j_d$ of the halo angular momentum $J_h$,

$$J_d = j_d J_h.$$

For an exponential disk with constant circular velocity $V_c$ and the surface density profile

$$\Sigma(R) = \Sigma_0 \exp(-R/R_d),$$

$M_d = 2\pi\Sigma_0 R_d^2$ and $J_d = 4\pi\Sigma_0 R_d^3 V_c$. From these two equations the disk density profile (parameters $\Sigma_0$ and $R_d$) can be expressed as functions of $m_d$, $j_d$, and $V_c$.

The distribution of angular momenta for dark matter halos is usually quantified by the *spin parameter*

$$\lambda = \frac{J_h |E_h|^{1/2}}{G M_h^{5/2}},$$

where $E_h$ is the binding energy of the halo (which depends on the actual adopted density profile). In hierarchically clustered universe spins of dark matter halos are induced by tidal torques from the surrounding material (Heavens & Peacock, 1988).

The distribution of spin parameters of halos of various masses turns out to be surprisingly independent of anything else (halo mass, shape of the matter power spectrum, cosmological parameters, redshift, etc) and is approximately lognormal,

$$p(\lambda)d\lambda = \frac{1}{\sqrt{2\pi}\sigma_\lambda} \exp\left(-\frac{\ln^2(\lambda/\bar{\lambda})}{2\sigma_\lambda^2}\right)\frac{d\lambda}{\lambda},$$

with $\bar{\lambda} \approx 0.05$ and $\sigma_\lambda \approx 0.5$ - that result remains unchanged from the first N-body simulations (Barnes & Efstathiou, 1987) to the present day (Trowland et al., 2013).

The final step in the MMW98 model is the connection between the disk circular velocity $V_c$ and the virial velocity of the halo,

$$V_{\rm vir} = \left(\frac{GM_h}{r_{\rm vir}}\right)^{1/2}.$$

In the original MMW98 model the coefficient of proportionality between $V_c$ and $V_{\rm vir}$ was assumed to be 1, but it does not have to be. For example, for the NFW profile

$$\left(\frac{V_c(r)}{V_{\rm vir}}\right)^2 = \frac{1}{x}\frac{\ln(1+cx) - cx/(1+cx)}{\ln(1+x) - c/(1+c)}$$

where $x \equiv r/r_{\rm vir}$ and $c$ is the concentration of the halo. In this case, however, $V_c$ is a function of radius and is not constant, so which one should we use? One solution is to consider the "maximal" disk, i.e. take the largest value of $V_c$ for any radius, commonly referred to as $V_{\rm max}$, as the disk circular velocity. That value is mildly dependent on the halo concentration $c$,

$$\begin{aligned} V_{\rm max} &= 1.0\,V_{\rm vir} \text{ for } c = 3, \\ V_{\rm max} &= 1.2\,V_{\rm vir} \text{ for } c = 10, \\ V_{\rm max} &= 1.6\,V_{\rm vir} \text{ for } c = 30. \end{aligned}$$

The MMW98 model is controlled by two main parameters, $m_d$ and $j_d$. In principle, they can be arbitrary. However, recently an interesting property of real galaxies has been noticed by Kravtsov (2013): disk sizes (both for stellar disks and gaseous disks) are linearly proportional to the virial radii, with the scatter in the relation entirely consistent with the distribution of $\lambda$ parameters for halos of a given mass. In other words, parameters $m_d$ and $r_d$ must be such that for stellar disks $R_d \approx 0.01 R_{\rm vir}$ and for gaseous disks it is about a factor of 2.5 larger.

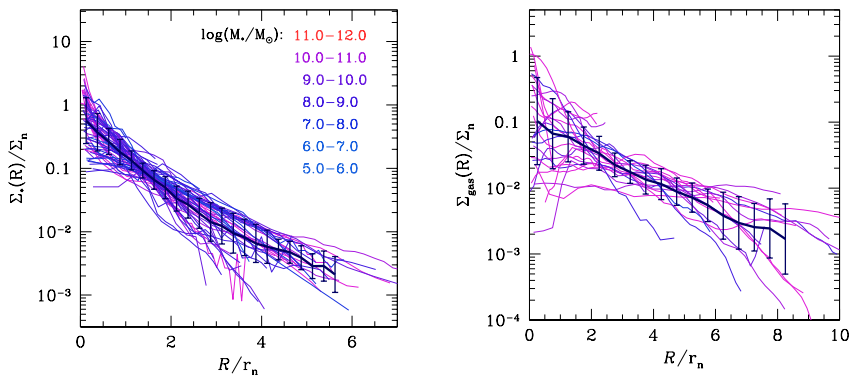



**Fig. 25** Normalized surface density profiles of stars and neutral gas for late-type galaxies (adopted from Kravtsov, 2013)).

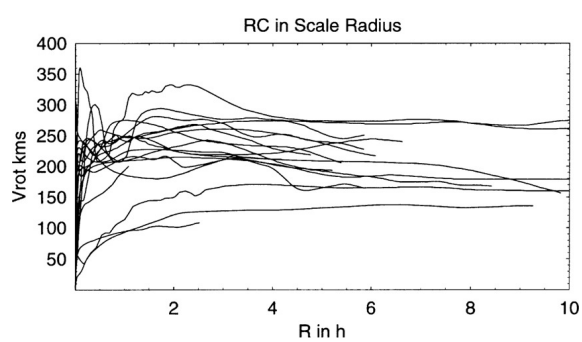
## *4.2 Galactic Disks*

We now descend into the actual galactic disks. The common lore is that disks are exponential, rotationally supported, and have flat rotation curves. While all these statements are kind of true, they are very far from being exact.

Disks come with a variety of density profiles and a variety of rotation curves. For example, figure 25 shows surface density profiles for stars and gas for several samples of disk galaxies (Kravtsov, 2013). On average profiles are indeed exponential, but deviations of individual galaxies from the mean can easily reach a factor of several.

Similarly, rotation curves of individual galaxies (figure 26) show large deviations from the canonical flat shape - some rotation curves are rising, some are falling, some remain truly flat all the way to the outer edge of the disk.

**Fig. 26** Rotation curves of several spiral galaxies from Sofue et al. (1999).



Disk dynamics in general is a very complex affair. A large number of various disturbances and waves can propagate over the disks - in addition to spiral arms, there exit bending modes, bars, warps, etc. All these perturbations cause orbits of stars and gas to deviate from spherical symmetry. For example, spiral arms are shock waves, gas changes its velocity abruptly by a large factor (up to several times its sound speed) as it crosses the shock, and hence the gas in front of and behind the spiral arm shock cannot remain on the same circular orbit - one of the sides has to deviate substantially. For example, in the classical example of the grand design spiral, M51, the deviations of the gas rotational velocity from the circular velocity reach 20 km/s *almost everywhere in the disk* (Hitschfeld et al., 2009).

Such deviations, in fact, may be responsible, at least partially, for the notorious cusp-core controversy. Some of the "observed" cusps may, in fact, be just an erroneous consequence of the incorrect assumption that the rotational velocity is equal to the circular velocity for gas (Valenzuela et al., 2007).

### 4.2.1 Disk Stability

How one would investigate such waves and features? Nonlinear treatment would require numerical simulations, but some widely known (and not so widely known) results can be obtained analytically for the linear stability of disk systems. A standard approach to studying linear stability of any system is to impose small fluctuations on the system and derive their dispersion relation. For an infinitely thin disk one can represent the radially perturbed (i.e. a perturbation remains azimuthally symmetric) surface density $\Sigma(t,R)$ as

$$\Sigma(t,R) = \bar{\Sigma}(R) + \Delta\Sigma(t,R),$$

where the perturbation $\Delta\Sigma(t,R)$ is assumed to be a collection of linear waves, each wave characterized by the frequency $\omega$ and the wavevector $\mathbf{k} = (k_R, k_\phi)$. Let's first focus on purely radial perturbations, $k_\phi = 0$. In that case the dispersion relation for the gaseous disk becomes (Binney & Tremaine, 1987)

$$\omega^2 = \kappa^2 - 2\pi G\bar{\Sigma}|k_R| + c_s^2 k_R^2, \tag{22}$$

where $\kappa^2 \equiv R(d\Omega^2/dR) + 4\Omega^2$ is the so-called *epicyclic frequency* and $\Omega(R)$ is the disk angular velocity, $V_c(R) = R\Omega$.

The disk is stable when the right hand side is always positive, which is achieved if and only if

$$Q \equiv \frac{c_s\kappa}{\pi G\bar{\Sigma}} > 1. \tag{23}$$

This condition is universally known as *Toomre stability criterion*, although for gaseous disks it has been obtained earlier by Safronov (1960), while Alan Toomre derived a similar relation for stellar disks (Toomre, 1964), a much more difficult exercise.

When $Q < 1$, some of the radial modes in the disk become unstable,

$$\frac{\kappa}{Qc_s}\left(1-\sqrt{1-Q^2}\right) < k_{\text{unstable}} < \frac{\kappa}{Qc_s}\left(1+\sqrt{1-Q^2}\right).$$

An interesting property of this relation is that only a limited range of wavenumbers become unstable, the disk remains stable at very large ($k \to 0$) and very small ($k \to \infty$) scales.

### 4.2.2 Beyond Toomre

Toomre stability criterion is often used in galactic and extragalactic studies. However, it is, unfortunately, often forgotten that it is incomplete. No disk is infinitely thin, and no perturbation is perfectly radial.

A case of arbitrary, not necessarily radial, perturbations was considered by Polyachenko & Polyachenko (1997), who found that the critical value for the $Q$ parameter is actually larger than 1. This is not surprising - at $Q = 1$ radial perturbations go unstable; however, for the disk to become unstable it is only enough for *some* waves to become unstable, and these first unstable waves do not have to be radial. Thus, some of the non-radial (i.e. non-axially-symmetric) perturbations may become unstable when all radial perturbations remain stable with $Q > 1$.

The critical value of the $Q$ parameter turns out to depend on the disk density profile,

$$Q^2_{\text{crit}} = \frac{3\alpha^2-3}{2\alpha^2-3} > 1,$$

where

$$\alpha^2 = \frac{2\Omega(R)}{R|d\Omega/dR|}.$$

For example, for a flat rotation curve ($\Omega \propto R^{-1}$) $\alpha^2 = 2$ and

$$Q_{\text{crit}} = \sqrt{3}.$$

This is the reason why most actively star-forming (and, thus, instability-developing) disk galaxies have $Q$ parameters above unity but not significantly greater than 2 (Leroy et al., 2008).

Another generalization of the Toomre stability criterion is obtained when the finite thickness of a disk is taken into account. In that case the dispersion relation has been introduced by Begelman & Shlosman (2009), although in a highly convoluted form it has been derived earlier by Safronov (1960),

$$\omega^2 = \kappa^2 - 2\pi\frac{G\bar{\Sigma}|k_R|}{1+|k_R|h} + c_s^2 k_R^2, \tag{24}$$

where $h$ is the disk *scale height*, $\bar{\Sigma}(z) \propto \exp(-z/h)$. For a non-exponential vertical profile the dispersion relation becomes more complex and is not presentable analytically in a closed form.

Relation (24) is remarkable in that in the limit of very small scales, well below the disk scale height, $kh \gg 1$ (in which case the disk cannot be considered as a flattened system any more), it reduces to

$$\omega^2 = -4\pi G\bar{\rho} + c_s^2 k_R^2,$$

(with $\bar{\Sigma} = 2\bar{\rho}h$), which is nothing else as a usual Jeans stability dispersion relation, familiar to any astrophysicist since kindergarten.

### 4.2.3 Modeling Disks

Modeling disks numerically is a subject of itself, and cannot be covered in these lectures. However, a word of caution is in order here. Let's imagine one is trying to model a galactic disk (or, for that matter, a disk around a supermassive black hole, or any other self-gravitating disk). A natural setup is to start with an axially-symmetric disk and let the instabilities develop.

So, you prepared your symmetric disk as the initial condition for your powerful numerical code that includes all kind of important physical processes (cooling, star formation, feedback, etc). To be specific, let's say you set the gas temperature to $10^4$ K in the disk with the circular velocity of 200 km/s.

You press the magic button, simulation starts, and in an instant your disk cools off to the lowest temperatures your cooling module allows (indeed, cooling times in astrophysical environments are often very short), the $Q$ parameters plunges to very small values, and your disk fragments into tiny clumps of size comparable to the wavelength of fastest growing instability mode $\lambda_{\rm fast}$,

$$R \sim \lambda_{\rm fast} = 2\pi Q \frac{c_s}{\kappa}.$$

Such a state, however - cold homogeneous disk - is *unphysical*, there is no plausible physical process that can create such a system: after all, you started with an artificial initial condition; try running it backward in time, the disk is still cooling, so shortly before your initial moment it should have been blazingly hot, at $10^7 - 10^8$ K, and how would you propose to keep $10^8$ K plasma in a disk with 200 km/s circular velocity?

Ok, that does not work. Let's now start with an initially stable disk ($Q \gg 1$) and let it become unstable gradually (either by artificially introducing cooling gradually, or disabling cooling below $10^4$ K, or, even better, gradually adding mass to the disk). As $Q$ decreases gradually, at some moment it will reach a critical value $Q_{\rm crit} > 1$. At that moment some non-radial perturbations become unstable and start growing, turning into non-linear waves; any non-linear wave in the gas steepens to a shock; any shock in a differentially-rotating disk becomes an oblique spiral wave; oblique shocks are known to generate an energy cascade a-la turbulence (although it may not be turbulence in the exact meaning of that word). Turbulence will provide extra

support to the gas, replacing the sound speed $c_s$ in equation 22 with $\sqrt{c_s^2+\sigma_t^2}$ and will limit the fragmentation scales to $R \sim 2\pi\sqrt{c_s^2+\sigma_t^2}/\kappa$.

In other words, in the latter scenario the $Q$ parameter never had a chance to become much lower than the critical value, but must linger at around it, maintaining the disk in the just-unstable-enough state to generate enough turbulence. Hence the conclusion that the author arrived at himself after much suffering and erring: if your disk simulation has $Q \ll 1$, you are doing something wrong...

## *4.3 Ionized, Atomic, and Molecular Gas in Galaxies*

Everyone knows that ISM consists of several gas phases. The ionized gas comes in two flavors, as hot ($\sim 10^6$ K) coronal gas and warm/cool ($\sim 10^4$ K) ionized gas (known under many names: warm ionized medium (WIM), diffuse ionized gas (DIG), Reynolds Layer); atomic gas exists as warm/cool ($\sim 10^4$ K) and cold ($\sim 10^2$ K) neutral media (WNM and CNM respectively); finally, molecular gas is almost always cold ($< 10^2$ K).

### 4.3.1 Ionized Gas

A story of coronal gas is misty and messy - it is not even clear how much of it there is in the Milky Way ISM, or what fraction of it comes from stellar feedback processes and what fraction is merely halo gas intermixed into the ISM due to various disk instabilities. Warm ionized medium is understood better because it is primarily located at the outer edges of the disk.

What causes WIM? We can get a hint on its origin from its temperature - gas at $10^4$ K is likely to be photo-ionized. If we recall that only gods have the power to switch off the Cosmic Ionizing Background, the ionizing source is there too - plus whatever ionizing radiation escapes from star-forming regions inside the Milky Way disk.

An example of how the relative distribution of neutral and ionized gas may look like in the Milky Way galaxy (or other similar galaxies) is shown in figure 27. The WIM contribution stays more-or-less constant at about $0.5\,\mathrm{M}_\odot/\,\mathrm{pc}^2$ (column density $N_\mathrm{H} = 6\times10^{19}\,\mathrm{cm}^{-2}$) in the outer disk, but increases to several $\mathrm{M}_\odot/\,\mathrm{pc}^2$ inside the solar radius because of the increased radiation field and a contribution of coronal gas. Broadly, such behavior is consistent with actual observations of the ionized gas in the Milky Way and other galaxies. For example, in the Milky Way the contribution of ionized gas at the solar radius is about $1\,\mathrm{M}_\odot/\,\mathrm{pc}^2$.

The outer parts of the disk are consistent with being ionized by Cosmic Ionizing Background, and the transition from neutral to ionized gas is often very sharp. However, consistency does not imply causality. There could be other ionizing sources, such as stellar radiation escaping from star-forming regions or cosmic rays. Since stars do not form in the ionized gas (as far as we can tell), we leave the WIM-land on

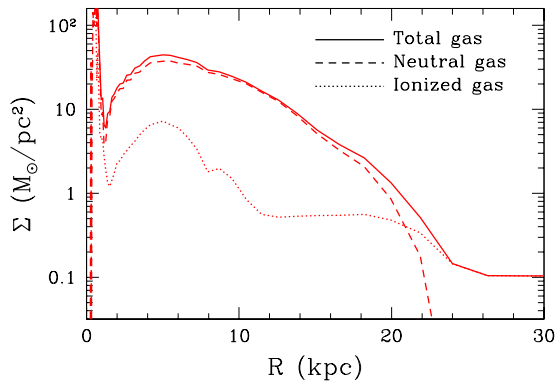



**Fig. 27** Surface density profiles for the total, ionized, and neutral (atomic and molecular) gas for a model Milky-Way-like galaxy (from the simulation described in Gnedin (2012)).

our way to denser and colder domains; interested readers should check an excellent recent review by Haffner et al. (2009).

### 4.3.2 From Atomic To Molecular Gas

Stars (at least most of them) form from molecular gas. Few astronomers would question this conjecture. While a minority of all stars may form in the atomic gas (at least Pop III stars certainly form in gas that is 99% atomic), on this journey we are chasing the bulk of star formation. Hence, the transition from atomic to molecular gas is a necessary condition for (the bulk of) star formation.

Chemistry of molecular hydrogen is not particularly complex; $H_2$ forms through two physically distinct channels: in numerous reactions in the gaseous phase, from rare ions H− and $H_2{}^+$ (the best reference for these processes is Glover & Abel (2008)), and on the surface of cosmic dust, which serves as a catalyst. The gas processes are slow exactly because H− and $H_2{}^+$ are rare; fraction of molecular hydrogen forming in the gas phase saturates at $10^{-3} - 10^{-2}$ and only jumps to close to 1 when 3-body reactions become sufficiently efficient (which only happens at densities above about $10^{12}\,\mathrm{cm}^{-3}$). This channel of $H_2$ formation does not require any metals and can proceed in the primordial gas (indeed, this is how Pop III stars form).

Formation of $H_2$ on dust grains is not fully understood. It is usually assumed that atomic hydrogen accumulates on grains where two atoms can find each other much more easily (young couples tend to live in cities). The formation rate $R_D$, defined as

$$\left.\frac{dn_{\mathrm{H}_2}}{dt}\right|_{\mathrm{dust}} = R_D n_{\mathrm{H}} n_{\mathrm{HI}},$$

has been modeled (somewhat inconclusively) theoretically and measured observationally by Wolfire et al. (2008):

$$R_D = D_{\rm MW} R_0,$$

with $R_0 \approx 3.5 \times 10^{-17}$ cm$^3$/ s, where from now on I will use a convenient parameters $D_{\rm MW}$ that measures the abundance of dust relative to the solar neighborhood; i.e. $D_{\rm MW} = 1$ implies the same abundance of dust per unit mass of gas as in the Milky Way ISM around us.

It is not, however, enough to know the formation rate to determine the abundance of molecular hydrogen - like predator and prey, ultraviolet radiation plays with $H_2$ the game of life and death. Particularly deadly for molecular hydrogen is radiation in the so-called Lyman and Werner bands, at energies between 11.3 and 13.6 eV (actually, the bands extends further, but hydrogen ionizing radiation is often well shielded by neutral atomic ISM). In addition, molecular hydrogen is destroyed by collisions with atoms and other molecules when gas temperatures raise above about 5,000 K. Hence, in order to predict the abundance of molecular hydrogen in specific conditions, we need to know the Interstellar Radiation Field (ISRF).

ISRF is not measured directly, but rather modeled based on the observations of various line ratios in the ISM. Two canonical references to such modes are Draine (1978) and Mathis et al. (1983), which are perfectly consistent with each other. In the solar neighborhood $J_0 \approx 10^6$phot/cm$^2$/s/eV/rad, but in the Galaxy the radiation field changes with the distance from the center. At the center it is up to 10 times higher than around the Sun.

Just like masses and luminosities are convenient to measure in solar units, in galactic studies it is convenient to measure the radiation field and other quantities (like dust abundance) in the Milky Way units. Hence, hereafter we will also use $U_{\rm MW} \equiv J_{\rm LW}/J_0$ (where $J_{\rm LW}$ is the average radiation field in the Lyman and Werner bands). By definition, $U_{\rm MW} = 1$ in the solar neighborhood, but in high redshift galaxies it can be large, $U_{\rm MW} = 30 - 300$ at $z \sim 2$ (Chen et al., 2009).

Even the Milky Way radiation field is extremely strong from the molecular hydrogen point of view - if it could shine on typical molecular clouds unimpeded, the molecular fraction would only be $10^{-6} - 10^{-5}$. The only reason molecular clouds exist in the universe is because all that radiation is *shielded*.

There are two distinct shielding processes: dust shielding and molecular self-shielding. Dust absorbs radiation over a very large range of wavelengths, from infrared to X-rays. Dust opacity is a smooth function of wavelengths, and in the first approximation it can be considered constant over a narrow Lyman and Werner bands (for detailed plots of dust opacity see Weingartner & Draine, 2001). In different galaxies the dust opacity is different, but in the three galaxies it was studied best - Milky Way and two Magellanic clouds - it is roughly proportional to the dust-to-gas ratio,

$$\sigma_{\rm LW} = D_{\rm MW} \sigma_0$$

with $\sigma_0 = 1.7 \times 10^{-21}$cm$^2$ for the Milky Way ($D_{\rm MW} = 1$),$\sigma_0 = 1.6 \times 10^{-21}$cm$^2$ for the LMC ($D_{\rm MW} \approx 0.5$), and $\sigma_0 = 2.2 \times 10^{-21}$cm$^2$ for the SMC ($D_{\rm MW} \approx 0.2$). Thus, it is possible to simply take $\sigma_0$ as a universal constant,

$$\sigma_0 \approx 2 \times 10^{-21}{\rm cm}^2.$$

Accounting for continuum shielding over a narrow band is easy; the molecular hydrogen photo-destruction rate $\Gamma$ is then simply

$$\Gamma = c\sum_j \int_{\nu_1}^{\nu_2} \sigma_j(\nu)\ \underbrace{e^{-\sigma_d(\nu)N_{\rm H}} n_\nu}_{\text{radiation field}}\ d\nu \approx e^{-\bar{\tau}_d}\Gamma_{\rm LW},$$

where $N_{\rm H}$ is the total hydrogen column density, $\bar{\tau}_d \equiv \bar{\sigma}_d N_{\rm H}$ is the average dust opacity in the Lyman and Werner bands, $\Gamma_{\rm LW}$ is the so-called "free space" photo-destruction rate (i.e. photo-destruction rate in the absence of any shielding), and the sum is taken over all $H_2$ lines in the Lyman and Werner bands. It is convenient to define a *shielding factor* $S_D$ that parametrizes the suppression of the free space field by dust shielding, $\Gamma = S_D\Gamma_{\rm LW}$, with

$$S_D(D_{\rm MW}, N_{\rm H}) = e^{-D_{\rm MW}\sigma_0 N_{\rm H}}.$$

Self-shielding of molecular hydrogen is much more complicated. Lyman and Werner bands consists of numerous lines of various strengths (figure 28). Absorbing a photon in one of those lines may or may not lead to the destruction of the hydrogen molecule, and the probability of dissociation varies significantly for different lines.

Hence, the shielded photo-destruction rate can be represented as a sum over individual lines, each with its own cross section $\sigma_j(\nu)$,

$$\Gamma = c\sum_j \int_{\nu_1}^{\nu_2} \sigma_j(\nu)\ \underbrace{e^{-\sigma_j(\nu)N_{\rm H_2}} n_\nu}_{\text{radiation field}}\ d\nu \approx \sum_j e^{-\bar{\tau}_j}\Gamma_{{\rm LW},j} = S_{\rm H_2}(N_{\rm H_2})\Gamma_{\rm LW}. \qquad (25)$$

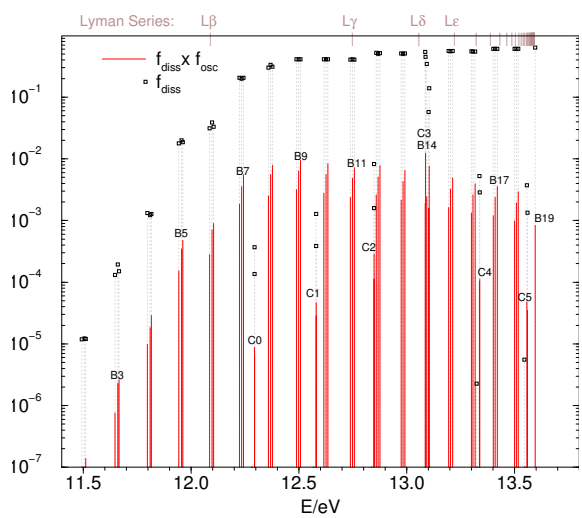



**Fig. 28** Molecular lines in the Lyman and Werner bands. A hydrogen molecule has a non-zero probability to be photo-dissociated $f_{\rm diss}$ when it is excited into any of these states (adopted from Haiman et al. (2000)).

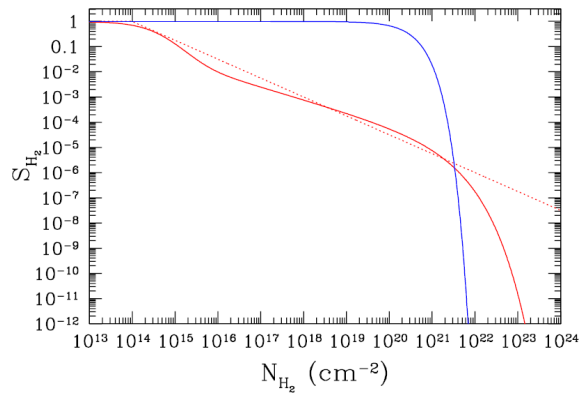



**Fig. 29** Draine & Bertoldi (1996) molecular self-shielding factor as a function of $H_2$ column density (solid red line). For comparison, exponentially falling off shielding factor (dust shielding with Milky Way dust and fully molecular gas, $N_{H_2} = 2N_{H_2}$)is shown as a blue line. Red dotted line is a power-law approximation for the self-shielding factor, $S_{H_2} \propto N_{H_2}^{-0.75}$ that has been also used in the past.

The self-shielding factor $S_{H_2}(N_{H_2})$ is much harder to compute, but its general behavior may be guessed. As individual Lyman and Werner bands lines become optically thick, some of the terms in the sum in equation (25) become small, but weaker lines will remain optically thin and un-shielded for much higher column densities than the stronger lines, thus allowing the destructing radiation to sneak deeper into a molecular cloud. Hence, as the column density of molecular hydrogen increases, the self-shielding factor will fall at a rate, which is much slower than the exponential decline of an individual line.

The self-shielding of molecular hydrogen has been modeled extensively; a specific approximation for the self-shielding factor that is most commonly used is due to Draine & Bertoldi (1996),

$$S_{H_2} = \frac{0.965}{(1+x/b_5)^{\alpha}} + \frac{0.035}{\sqrt{1+x}} \exp\left(-\frac{\sqrt{1+x}}{1180}\right), \tag{26}$$

where $x \equiv N_{H_2}/5 \times 10^{14}\,\mathrm{cm}^{-2}$, $b_5 \equiv b/\,\mathrm{km/s}$, and in the original approximation $\alpha = 2$. Wolcott-Green et al. (2011) suggested that at higher temperatures a better fit is $\alpha = 1.1$, but the first term in equation (26) is not important anyway.

Figure 29 shows the Draine & Bertoldi (1996) approximation as a function of the molecular column density. A gradual decline of the self-shielding factor ($S_{H_2}$ going approximately as $N_{H_2}^{-0.75}$) is apparent for almost 8 orders of magnitude. However, at very high column densities, $N_{H_2} > 10^{22}\,\mathrm{cm}^{-2}$, the fall-off becomes steeper, with the last factor in equation (26) dominating. What could cause such a steep decline?

Our deduction above that the weaker lines remain optically thin and serve as avenues for the radiation to sneak into a molecular cloud remain correct for as long as each absorption line can be treated as independent. However, just like in human society neighbors sooner or later will put a stop on a weak person misbehaving, so in the society of Lyman and Werner bands stronger lines begin to interfere in the affairs of weaker one at sufficiently high column densities. Since each excited state in an atom or molecule lives for a finite time, lines have non-trivial *natural width* (see section 2.2.1). In the high column density limit the natural width dominates,

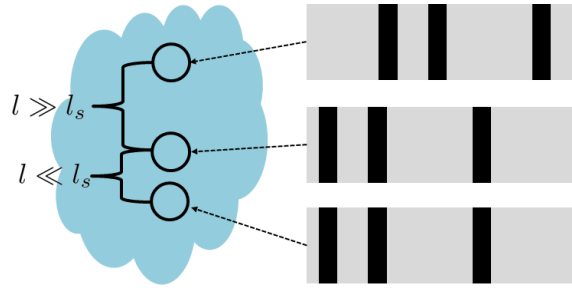



**Fig. 30** A cartoon illustrating the role of ISM turbulence in suppressing self-shielding of molecular hydrogen on scales above the sonic length.

and the equivalent width of a line (the area of the spectrum the line takes out) grows as $N_{\mathrm{H}_2}^{1/2}$. As the strongest lines begin to overlap, the nature of self-shielding changes - instead of individual lines absorbing UV radiation each by itself, the absorption cross-section now becomes a continuous function of frequency, with cross-sections of individual lines all blending together into a single, continuum-like absorption. Hence, self-shielding becomes much stronger, and that is manifested in the drop-off in the Draine & Bertoldi (1996) formula at $N_{\mathrm{H}_2} > 10^{22}\,\mathrm{cm}^{-2}$.

Finally, we need to figure out what $N_{\mathrm{H}_2}$ actually is. Let's imagine that we have a line-of-sight through a molecular cloud with the total hydrogen column density $N_{\mathrm{H}}$. The first inclination is to simply use $N_{\mathrm{H}_2} = 0.5 N_{\mathrm{H}}$ (let's assume the cloud is fully molecular), but that is actually *wrong*!

Equation (26) is suitable for the idealized case of a slab of gas with no internal motions. Real molecular clouds are, however, supersonically turbulent on scales above the *sonic length*, $l_s \lesssim 1$ pc. In other words, if you take two parcels of gas inside a molecular cloud separated by a distance $l$, the rms velocity dispersion between them satisfies what is known as Larson's law,

$$\delta v(l) \approx c_s \left(\frac{l}{l_s}\right)^{0.5}$$

with $c_s$ being the gas sound speed (in fact, the definition of the sonic length is that $\delta v(l_s) = c_s$). For $l \gg l_s$, the velocity difference between them would be much larger than the width of each Lyman and Werner bands line $b \sim c_s$. Hence, these two fluid elements would shield each other only if they happen accidentally to fall at the same line-of-sight velocity, which would occur with the probability $b/\delta v$.

This is illustrated in a cartoon fashion in figure 30. Hence, a fluid element inside a molecular cloud sees a column density of about $N_{\mathrm{H}_2} \sim \langle n_{\mathrm{H}_2}\rangle_s L_{\mathrm{MC}} b/\delta v \approx N_{\mathrm{H}_2} \sim \langle n_{\mathrm{H}_2}\rangle_s (l_s L_{\mathrm{MC}})^{1/2}$, where $\langle n_{\mathrm{H}_2}\rangle_s$ is the average molecular hydrogen density on a sonic scale at the location of interest and $L_{\mathrm{MC}}$ the width of the whole molecular cloud. This is valid, however, only until individual lines do not overlap. With line overlap relative velocity shifts between different fluid elements become unimportant (lines overlap anyway). In other words, at sufficiently large column densities line radiative transfer in the Lyman and Werner bands effectively behaves as con-

tinuum radiative transfer, and the effective length over which the column density is accumulated approaches $L_{\rm MC}$.

In equation (26) the line overlap is described by the last exponential factor. To account for the supersonic turbulence inside the molecular cloud, equation (26) can be modified as

$$S_{\rm H_2} = \frac{0.965}{(1+x_1/b_5)^2} + \frac{0.035}{\sqrt{1+x_1}} \exp\left(-\frac{\sqrt{1+x_2}}{1180}\right), \tag{27}$$

where $x_1 \equiv \langle n_{\rm H_2}\rangle_s (l_s L_{\rm MC})^{1/2}/5\times 10^{14}\,{\rm cm}^{-2}$ is proportional to the $H_2$ column density over the sonic length, while $x_2 \equiv \langle n_{\rm H_2}\rangle_{\rm MC} L_{\rm MC}/5\times 10^{14}\,{\rm cm}^{-2}$ accounts for the column density of the whole molecular cloud. Obviously, $x_2 \gg x_1$.

Armed with understanding of dust and self-shielding, we can consider some interesting limiting cases. In the kinetic equilibrium the rates of photo-destruction and molecular hydrogen formation balance, hence

$$\Gamma_{\rm LW} S_{\rm H_2} e^{-\sigma_{\rm LW} N_{\rm H}} n_{\rm H_2} = R_D n_{\rm H} n_{\rm HI}.$$

The free-space radiation field is parametrized by the introduced above $U_{\rm MW}$ parameter, $U_{\rm MW} \equiv \Gamma_{\rm LW}/\Gamma_0$. Hence,

$$\frac{f_{\rm H_2}}{(1-f_{\rm H_2})} = \frac{D_{\rm MW}}{U_{\rm MW}} \frac{R_0}{S_{\rm H_2}\Gamma_0} e^{D_{\rm MW}\sigma_0 N_{\rm H}} n_{\rm H}. \tag{28}$$

As we already know, in low metallicity environments self-shielding is expected to dominate over dust shielding,

$$\frac{f_{\rm H_2}}{(1-f_{\rm H_2})} = \frac{D_{\rm MW}}{U_{\rm MW}} \frac{R_0}{S_{\rm H_2}\Gamma_0} n_{\rm H}.$$

Let's say we are interested in densities at which the gas becomes 50% molecular ($f_{\rm H_2} = 0.5$). In that case

$$S_{\rm H_2} \propto \frac{D_{\rm MW}}{U_{\rm MW}},$$

and for high enough column density, when

$$S_{\rm H_2} \sim e^{-{\rm const}\times N_{\rm H_2}^{1/2}},$$

we find

$$N_{1/2} \equiv N_{\rm H}(f_{\rm H_2} = 1/2) \propto \ln^2\left(\frac{U_{\rm MW}}{D_{\rm MW}} \times {\rm const}\right),$$

i.e. the column density of the atomic-to-molecular transition depends only weakly on the dust abundance or the interstellar radiation field.

In the opposite extreme, in high radiation fields the dust shielding dominates,

$$\frac{f_{\mathrm{H}_2}}{(1-f_{\mathrm{H}_2})} = \frac{D_{\mathrm{MW}}}{U_{\mathrm{MW}}}\frac{R_0}{\Gamma_0}e^{D_{\mathrm{MW}}\sigma_0 N_{\mathrm{H}}}n_{\mathrm{H}},$$

hence

$$N_{1/2} \propto \frac{1}{D_{\mathrm{MW}}}\ln\left(\frac{U_{\mathrm{MW}}}{D_{\mathrm{MW}}}\times \mathrm{const}\right).$$

As could have been easily guessed, higher dust abundance pushes the atomic-to-molecular transition towards lower (column) densities.

How should we go now from shielding factors for individual parcels of gas to the factors that should be used in actual numerical simulations? Modern cosmological or galactic scale simulation may not resolve molecular clouds at all or may resolve them down to parsec scales. Hence, in the most general case we can imagine whole space being tessellated into regions (say, simulation cells) of size $L$ some of which include pieces of molecular clouds. Each such piece has a distribution of column density inside it, $\phi_j(N_{\mathrm{H}_2})$, where $j$ refers to a given piece. Hence, the average shielding factor is

$$\langle S_{\mathrm{H}_2}\rangle_j = \int S_{\mathrm{H}_2}(N_{\mathrm{H}_2})\phi_j(N_{\mathrm{H}_2})dN_{\mathrm{H}_2} = S_{\mathrm{H}_2}(N_{\mathrm{eff},j})\int \phi_j(N_{\mathrm{H}_2})dN_{\mathrm{H}_2} = S_{\mathrm{H}_2}(N_{\mathrm{eff},j})$$

since $\int \phi_j(N_{\mathrm{H}_2})dN_{\mathrm{H}_2} = 1$ by definition. If the distribution $\phi_j(N_{\mathrm{H}_2})$ was known, one can also compute $N_{\mathrm{eff},j}$, but at present there are no models that attempt to determine $\phi_j$. Hence, we need to come up with an ansatz for $N_{\mathrm{eff},j}$. For example, in the absence of a better alternative, we can simply take equation (27) with the sonic length $l_s$ being fixed to some small value ($0.1-1\,\mathrm{pc}$) and a model for the size of molecular cloud $L_{\mathrm{MC}}$.

Perhaps the simplest such model is a "Sobolev-like" approximation that Andrey Kravtsov and I introduced a few years ago (Gnedin & Kravtsov, 2011),

$$L_{\mathrm{MC}} \equiv \frac{\rho}{2|\nabla\rho|}.$$

With such an approximation the complete set of equations is obtained. The dependence of the characteristic density of the transition on the environmental parameters on the particular spatial scale $L = 65\,\mathrm{pc}$ (read "resolution of your simulation") is shown in figure 31 - the two limiting regimes are easily noticeable in the figure.

In figure 32 I show how such a model fares in matching the observed surface densities of atomic and molecular gas on larger scales, where they are actually measured. The main achievement of models like this one is that they capture the observed saturation of the atomic surface density at about $10\,\mathrm{M}_\odot/\,\mathrm{pc}^2$ (for $D_{\mathrm{MW}} = U_{\mathrm{MW}} = 1$ case; the saturation level does depend on the environment, just like $n_{1/2}$). A detailed description of the latest edition of Gnedin & Kravtsov (2011) model is presented in Gnedin, Kravtsov, & Draine (2013, in preparation).

An alternative model for the atomic-to-molecular transition is due to Krumholz et al. (2009) - that model is simpler to implement, but does not account for line over-

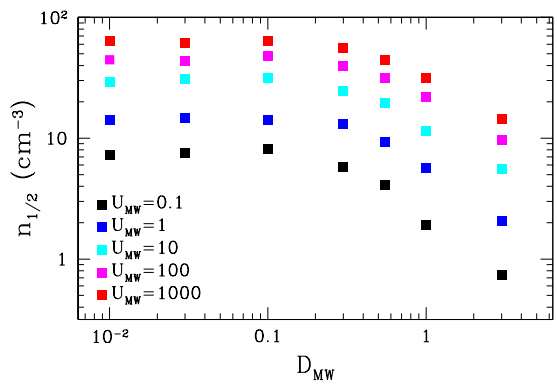



**Fig. 31** Average total hydrogen number density of atomic-to-molecular gas transition (defined as $f_{H_2} = 1/2$) as a function of the dust-to-gas ratio $D_{MW}$ and the interstellar radiation field $U_{MW}$ on scales $L = 65$ pc.

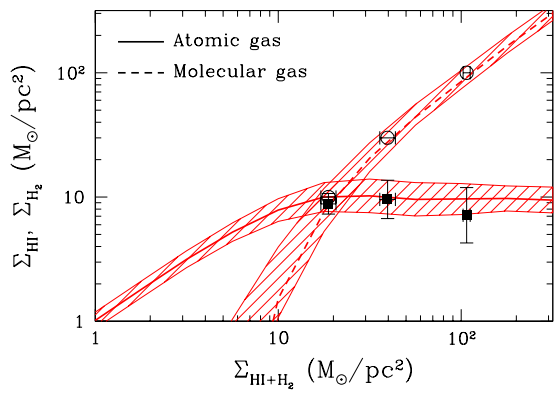



**Fig. 32** Average atomic and molecular gas surface densities as functions of the total (neutral) hydrogen gas surface density averaged over 500 pc scale for the $(D_{MW} = 1, U_{MW} = 1)$ simulation case (red lines/bands for mean/rms). Filled squares and open circles with error bars mark the observed average and rms atomic and molecular hydrogen surface densities from Wong & Blitz (2002).

lap, and, hence, breaks down for metallicities (or, rather, dust-to-gas ratios) below about 20% of the Milky Way value.

## *4.4 Molecular ISM*

Ok, we arrived into the molecular ISM. Now what? Why do we even care about the molecular gas? After all, many experts in star formation will tell you that molecules are *not* required for star formation. Now we know this is not quite true - line overlap makes $H_2$ self-shielding important at low dust abundances, and, hence, in that regime molecules *are required* for star formation.

A second answer to that question is offered by Krumholz et al. (2011) (and by the nature herself, but that story is still well ahead). Shielding in molecular gas actually performs two functions at once - it protects hydrogen molecules from photodestruction by Lyman and Werner bands photons, but it also allows gas to cool to the state that is properly called *cold* (100 K and below) - without shielding, UV and op-

tical photons can eject energetic electrons from dust grains by photoelectric effect (the one Einstein got the Nobel prize for); these energetic electrons thermalize in the gas, effectively transferring the energy of radiation into the gas thermal energy. With shielding, this process becomes much less efficient and the gas can cool to low temperatures - and, hence, fragment into small clumps from which stars can form.

Thus, even in the regime when molecular self-shielding is not important, molecular gas plays a role of a ”paint” someone poured into the ISM - the ”painted” (i.e. molecular) gas is cold and can form stars, while gas without ”paint” is too hot for star formation to take place there. Think of this as a lucky ”coincidence”, if you like.

### 4.4.1 Thermodynamics of $H_2$

Before we move further down the yellow brick road towards star forming regions, let us pause for a short while and refresh what we know about the hydrogen molecule. After all, it is the simplest molecule one can imagine, containing just two atoms (hence *diatomic*), and its thermodynamics can be solved (almost) exactly.

If you recall some college thermodynamics, you may remember that the diatomic gas has a polytropic index of $7/5$ (or, equivalently, specific heat $c_V = 5/2$). If you have forgotten that, it should not be hard to re-derive that result! After all, the partition function for the diatomic molecule is simply

$$Z = e^{-E_n/(k_B T)} Z_{\rm rot} Z_{\rm vib},$$

where the vibrational part is

$$Z_{\rm vib} = \sum_{v=0}^{\infty} e^{-\hbar\omega(v+1/2)/(k_B T)}.$$

The rotational part is a bit tricky, but really just a bit - since $H_2$ is a symmetric molecule and two protons are indistinguishable, two nuclear states (with the spins aligned, total nuclear spin is 1 and the spins anti-aligned, total nuclear spin is 0) behave almost like two different molecules (transitions between the two states are possible, but highly suppressed and only occur at high enough densities). The state with the nuclear spin of 1 is called an *ortho*-hydrogen molecule, and only allows odd values for the total angular momentum $J$, while the state with the 0 nuclear spin is a *para*-hydrogen molecule and only has even values of the angular momentum. Ortho-$H_2$ has a higher statistical weight than the para-state, hence

$$Z_{\rm rot} = \frac{3}{4} Z_{\rm ortho} + \frac{1}{4} Z_{\rm para},$$

where

$$Z_{\rm ortho} = \sum_{J=1,3,\ldots} (2J+1) e^{-\hbar^2 J(J+1)/(2Ik_B T)},$$

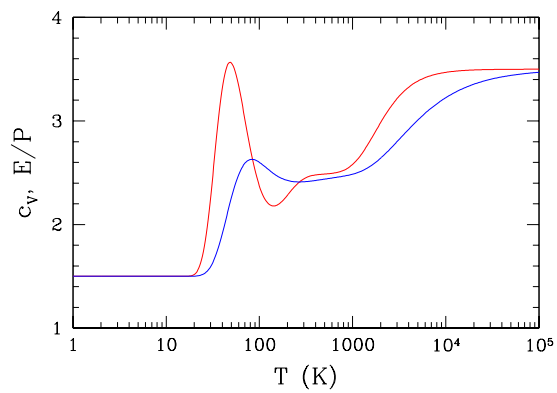



**Fig. 33** Specific heat $c_V$ (red) and the internal energy over pressure (blue) for molecular hydrogen gas as a function of temperature. Notice that $H_2$ never behaves as classic diatomic gas ($c_V = E/P = 5/2$).

$$Z_{\text{para}} = \sum_{J=0,2,\ldots} (2J+1)e^{-\hbar^2 J(J+1)/(2Ik_BT)}. \tag{29}$$

The partition function $Z$ is a magic wand of thermodynamics, all other quantities are derived from it: free energy

$$F = -k_BT \ln\left[\frac{V}{N!}\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{3/2} Z\right],$$

internal energy

$$E = F - T\left.\frac{\partial F}{\partial T}\right|_V,$$

specific heat

$$c_V = \frac{1}{k_BN}\left.\frac{\partial E}{\partial T}\right|_V,$$

etc.

For example, figure 33 shows $c_v$ and $E/(k_bT)$ for $H_2$ gas with 3 : 1 ratio of ortho-to-para molecules. If that plot does not surprise you, then you are a true expert in quantum thermodynamics - **molecular hydrogen actually never behaves as classic diatomic gas with $c_V = 5/2$ (or, equivalently, $\gamma = 7/5$).** More than that, it does not even behave as *polytropic* gas with $E = P/(\gamma - 1)$ except for very low temperatures ($T < 20\,\text{K}$) where it behaves as *monoatomic* gas with $c_V = 3/2$! If you did not know that, you can be excused - some of highly distinguished astrophysicists made that error too...

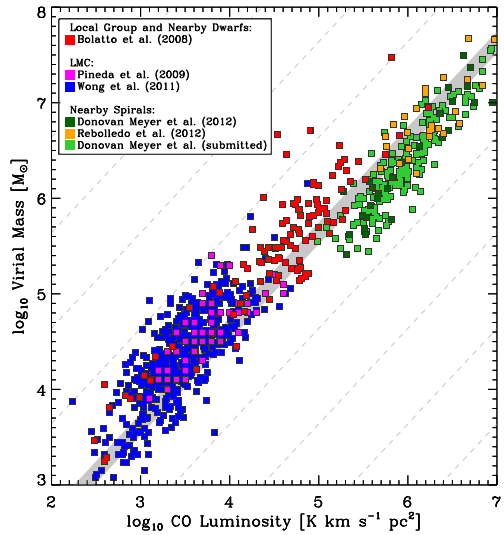



**Fig. 34** CO luminosity versus the virial mass for extragalactic molecular clouds. The gray band shows the average values in the Milky Way (adopted from Bolatto et al. (2013)).

### 4.4.2 Cosmic Pandora Box: The X-Factor

We are now approaching the most confused, abused, and misused subject in the studies of molecular ISM - CO emission and the $X_{\mathrm{CO}}$ factor.

Molecular hydrogen is a great example of a classic catch-22 - $H_2$ has to be shielded from the outside to exist, hence the outside (i.e. us observing it) cannot actually see its emission in the Lyman and Werner bands. And to add insult to injury, the same dust obscures background sources, making absorption spectroscopy extremely difficult. Historically, by far the most common method to observe molecular gas was via its CO emission.

Rotational transitions of the CO molecule are equally spaced in frequency, $\nu_J = \hbar J/(2\pi I)$ (the molecule is asymmetric, so we do not need to worry about ortho/para mess). For the most common $^{12}C^{16}O$ isotope the first ($1 \to 0$) transition is located at $\nu_{1\to 0} = 115\,\mathrm{GHz}$ (or $\lambda_{1\to 0} = 0.26\,\mathrm{cm}$). This is a major convenience, since CO emission lines are easy to identify (just look for a uniform fence in the millimeter wavelengths). On the other hand, there is no a priori reason why CO should be a good tracer of $H_2$: CO needs higher dust shielding to form and it gets saturated at too high column densities. Hence, CO emission comes from a narrow range of column densities, both cloud outskirts and cloud centers emit little.

Never-the-less, whenever a mass of molecular gas can be estimated by other means (usually the virial theorem), observations show a good correlation between the CO luminosity and the gas mass, albeit with substantial scatter from one cloud to another (figure 34).

In galactic studies the relevant conversion factor between the molecular gas and CO luminosity is the infamous *X-factor*,

$$X_{\rm CO} \equiv \frac{N_{\rm H_2}}{W_{\rm CO}},$$

where $W_{\rm CO}$ is the equivalent width of a CO emission line (which will be different for different transitions),

$$W_{\rm CO} = \int T_A(v) dv$$

with $T_A$ being the antenna temperature of the radio emission. The canonical Milky Way value for the X-factor is $X_{\rm CO} = 2 \times 10^{20}\,{\rm cm}^{-2}\,{\rm K}^{-1}\,({\rm km/s})^{-1}$ (enjoy the elegance of units!). The reason for this particular combination is that a measurement of the equivalent width in your telescope beam can be directly converted into the column density of molecular hydrogen along the line-of-sight.

In extra-galactic studies most of the time a galaxy is not spatially resolved (at least until the full ALMA comes online), so a single observation measures the total CO luminosity $L_{\rm CO}$ of a galaxy, and a convenient quantity is

$$\alpha_{\rm CO} \equiv \frac{1.36 M_{\rm H_2}}{L_{\rm CO}}$$

(the factor of 1.36 is a contribution of Helium, and it really should be $1/(1-Y)$, since $Y$ does depend slightly on the metallicity). The Milky Way value is $\alpha_{\rm CO} = 4.3\,{\rm M}_\odot/\,{\rm pc}^2/\,{\rm K}/\,({\rm km/s})$. Notice, that the connection between $X_{\rm CO}$ and $\alpha_{\rm CO}$ is somewhat non-trivial; $\alpha_{\rm CO}$ can be re-written as

$$\alpha_{\rm CO} \propto \frac{\int N_{\rm H_2} dA}{\int W_{\rm CO} dA} = \frac{\langle N_{\rm H_2} \rangle}{\langle W_{\rm CO} \rangle},$$

where $A$ is the area on the sky. Hence, $\alpha_{\rm CO}$ is *not* directly proportional to the average $X_{\rm CO}$ for a galaxy. Rather, it is proportional to the ratio of average $N_{\rm H_2}$ to the average $W_{\rm CO}$. Alternatively, we can re-interpret the averaging procedure for $X_{\rm CO}$ in a highly non-trivial way,

$$\bar{X}_{\rm CO}^{-1} \equiv \frac{\langle W_{\rm CO} \rangle}{\langle N_{\rm H_2} \rangle} = \frac{\langle (W_{\rm CO}/N_{\rm H_2}) N_{\rm H_2} \rangle}{\langle N_{\rm H_2} \rangle} = \left\langle \frac{1}{X_{\rm CO}} \right\rangle_{N_{\rm H_2}}.$$

I.e., $X_{\rm CO}$ should be averaged harmonically and with the $H_2$ column density weighing.

So, how should we approach modeling CO emission in modern cosmological or galactic-scale simulations? Scales on which CO emission originates are not yet resolvable in modern simulations, hence, it needs to be followed with a sub-grid model. However, since CO emission is not important dynamically, it can be modeled in post-processing, after the simulation had been completed. There exist many approaches to constructing a sub-grid model, and the best (at least in principle) sub-grid model is a someone's else simulation!

The field of modeling internal structure of molecular clouds with sufficient physics is rather new, with only a few attempts made so far, but it certainly developing rapidly. One example of how CO emission can be modeled is the work

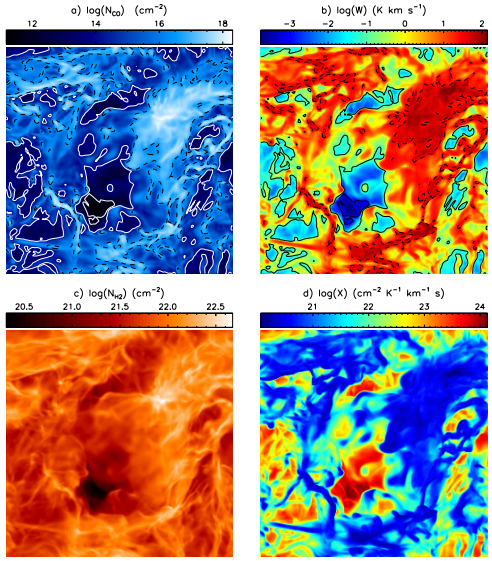



**Fig. 35** Images of $H_2$ and CO column densities, CO equivalent width $W_{CO}$, the $X_{CO}$ factor from Shetty et al. (2011a) simulations.

that was led by Robert Feldmann in two series of paper in 2012 (Feldmann et al., 2012a,b). This is just an illustration, one can follow a similar path with newer, better small-scale simulations for an undoubtedly better result.

One of the very first attempt to model CO emission directly in GMC-scale simulations was done by Simon Glover and collaborators (Glover & Mac Low, 2011; Shetty et al., 2011a,b). Images of $H_2$ and CO column densities, CO equivalent width $W_{CO}$, the $X_{CO}$ factor from these simulations are shown in figure 35. As can be expected in a turbulent ISM, there are large variations in the $X_{CO}$ factor on very small, sub-pc scales. Never-the-less, when averaged over the whole simulated region, $X_{CO}$ dependence on the properties of the molecular cloud exhibits remarkable regularity - Glover & Mac Low (2011) found that the main parameter that controls the $X_{CO}$ factor is (surprise!) the dust opacity (sometimes parametrized as the visual extinction $A_V \sim \tau_D$).

Figure 36 shows the mass-weighted molecular and CO fractions from Glover & Mac Low (2011) simulations. Using these tabulated values, Feldmann et al. (2012a,b) developed a sub-grid model that can be used in cosmological and galactic-scale simulations for computing the $X_{CO}$ factor in each simulation cell. Realistic simulated galaxies have complex ISM, with gas densities, metallicities, dust abundances, and interstellar radiation field varying from place to place. Hence, one can and *should* expect the $X_{CO}$ factor to vary significantly inside a given galaxy and from galaxy to galaxy.

As the result, the Feldmann et al. (2012a,b) model predicts a range of values for $X_{CO}$ even for a given metallicity and $U_{MW}$, not a single number, as is shown in figure 37. Overall, the predictions of the model are within the existing observational

measurements, although observations are still too imprecise to provide a serious constraint on the theoretical models.

Using the Feldmann et al. (2012a,b) model, we can explore why observers are often extremely stubborn in using a constant value for $X_{\rm CO}$ (or, alternatively, for $\alpha_{\rm CO}$). In figure 38 I show the dependence of the $X_{\rm CO}$ factor on the molecular hydrogen column density on small (GMC) scales and on large (galactic) scales. Averaging over large scales performs a miracle - almost all the complicated variations in the $X_{\rm CO}$ factor with various environmental parameters disappear (except for the mild residual dependence on the metallicity) and $\bar{X}_{\rm CO} \propto \alpha_{\rm CO}$ becomes a surprisingly robust conversion factor from the observed CO luminosity to the total mass of the molecular gas in a distant galaxy (this is indeed nothing short of a miracle).

Before we depart from the domain of sub-grid modeling of the $X_{\rm CO}$ factor, a word of caution is in order. Such modeling is, obviously, not unique. In addition, the existing observational constraints that can be used to calibrate such modeling are still in their infant stage. Hence, any sub-grid model for the $X_{\rm CO}$ factor will remain highly imprecise for some time. For example, an alternative model was proposed by Narayanan et al. (2011) in which $X_{\rm CO}$ is a *decreasing* function of $H_2$ column density - the dependence that has the *opposite* sign to the left panel of figure 38. That does seem somewhat inconsistent with the data from Heiderman et al. (2010), but the measurements are not yet fully constraining. In any event it is clear that if two different models predict opposite signs, there is a large amount of work laying ahead...

### 4.4.3 Cosmic Pandora Box, level 2: The X-Factor in ULIRGS

Cosmic Pandora boxes are like Russian Matrioshka dolls, inside one there is always another one...

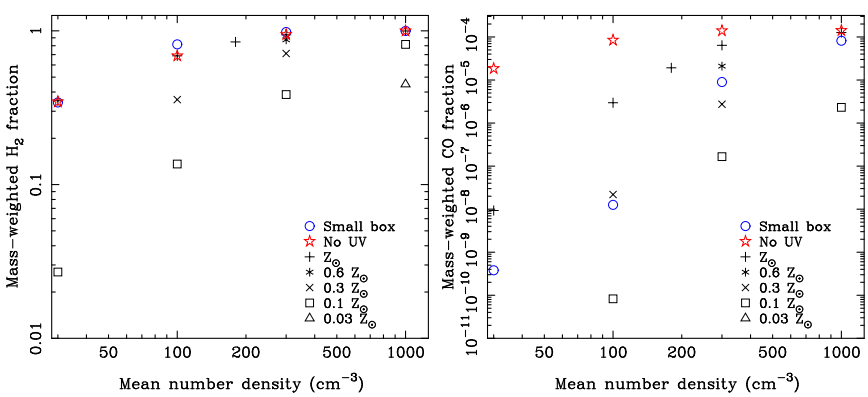



**Fig. 36** Mass-weighted fractions of $H_2$ and CO as a function of average gas density in Glover & Mac Low (2011) simulations.

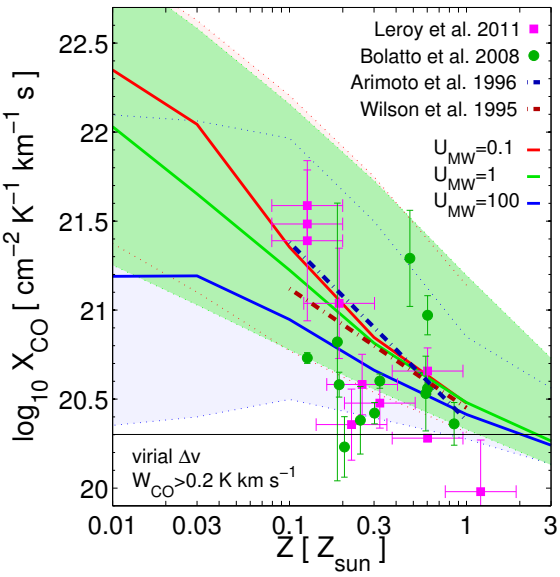



**Fig. 37** Dependence of the $X_{\rm CO}$ factor on the environmental parameters: gas metallicity $Z$ and the interstellar radiation field $U_{\rm MW}$, on $\sim 50$ pc scales. Colored bands show the variation over different locations in a single simulated galaxy (adopted from Feldmann et al. (2012a)).

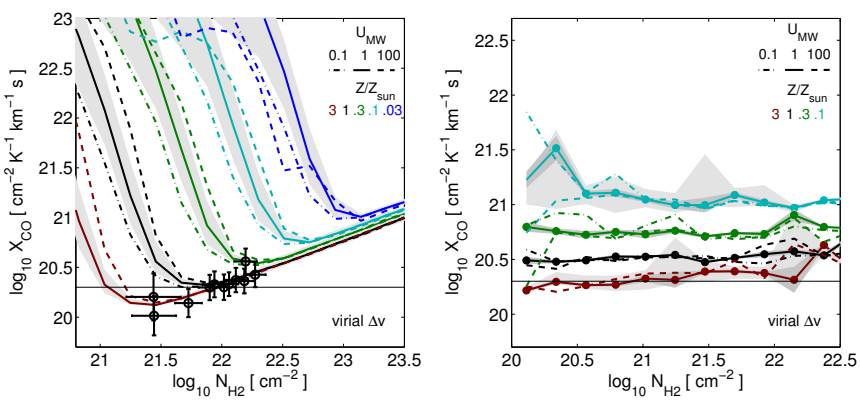



**Fig. 38** $X_{\rm CO}$ factor as a function of $H_2$ column density for a variety of values for the gas metallicity and the interstellar radiation field $U_{\rm MW}$ on small, 50 pc scale (left) and large, $\sim 1$ kpc scale (right). The data points on the left panel are from Heiderman et al. (2010). The role of large-scale averaging in making $X_{\rm CO}$ approximately constant is apparent (adopted from Feldmann et al. (2012a)).

The remarkable property of the $X_{\rm CO}$ factor to average out on large scales has been used extensively in many extragalactic studies. From an observer's point of view, it is very convenient to be able to determine the molecular gas mass of a distant galaxy by a simple multiplication. There is, however, a complication. For an optically thick emission, like CO, the equivalent width of the line $W_{\rm CO} = T_B \int \beta(v) dv$, where $T_B$ is the brightness temperature of the emitting gas and $\beta(v)$ is the escape probability from a parcel of gas with velocity $v$. The second factor can be thought of as an effective line width $\Delta v_{\rm eff}$, so that

$$X_{\rm CO} = \frac{N_{\rm H_2}}{T_B \Delta v_{\rm eff}}.$$

Variations in $N_{\rm H_2}$ and $\Delta v_{\rm eff}$ do average out, so it is $T_B$ that we are concerned with now. In LTE brightness temperature is equal to the gas temperature. In normal galaxies molecular gas is very cold, $T_B \sim 10\,{\rm K}$, but dust is usually warmer than the gas, $T_{\rm dust} = 40 - 60\,{\rm K}$. Hence, if dust and gas couple thermally, $T_B$ can increase systematically in at least some molecular clouds, causing a systematic decrease in the $X_{\rm CO}$ factor that will not average out.

For dust and gas to couple, densities must be really high, significantly higher than is achieved in normal molecular clouds, so in normal galaxies coupling occurs only in a tiny fraction of the most dense molecular gas. The situation is different in Ultra-Luminous IR Galaxies (ULIRG), which are major merger of large galaxies. In mergers substantial fraction of the total gas in both galaxies gets channeled towards the center, where it gets extremely dense, piling up to many thousands of solar masses per square parsec (versus a few tens for galaxies like the Milky Way). At such high densities (and column densities) dusts starts coupling to (and, hence, heating) the gas.

In a classical study Solomon et al. (1997) explored that effect in several nearest ULIRG, and concluded that the $X_{\rm CO}$ factor (or, rather, $\alpha_{\rm CO}$, since we are talking about external galaxies) could be as low as $\alpha_{\rm CO,min} = 0.8\,{\rm M_\odot}/\,{\rm pc}^2/\,{\rm K}/\,({\rm km/s})$. That value, however, was only a strict lower limit, as their results depended on several assumptions that all added a factor of 2 factors on top of $\alpha_{\rm CO,min}$. Alas, in an ironic mis-interpretation of the Solomon et al. (1997) paper many observers took that lower limit as the actual value, and for almost 2 decades it was quite common to hear a fairy tale of "two modes of star formation", each with its own value of $\alpha_{\rm CO}$ (0.8 and 4.3).

Obviously, such a "bimodality" makes no physical sense - a miracle of nature may make $\alpha_{\rm CO}$ a universal constant, but if it is not, then there must be either a distribution of $\alpha_{\rm CO}$ for different galaxies or a systematic trend of the average $\alpha_{\rm CO}$ value with some of galaxy properties, like the mean surface density or IR luminosity.

Fortunately, the dust settled (or, more precisely, was observed directly), thanks to Herschel (again, not a somewhat eccentric, clever, and compassionate man but a space telescope). Measurements of dust emission over several bands between 100 and 1000 microns, when taken together with optical and sub-mm observations from the ground, allow to fit detailed models of dust spectral energy distribution and, hence, derive dust temperature and mass, in a substantial sample of ULIRG over a wide redshift range, all the way to $z \sim 3$ (c.f. Magdis et al., 2012).

These observations can then be combined with measurements of gas metallicities and CO luminosities in the same galaxies in two different ways.

1. If one assumes dust-to-gas ratio as a function of metallicity $M_{\rm gas}/M_{\rm dust}(Z)$ (for example, by calibrating from the measurements of nearby galaxies), then from $M_{\rm dust}$ one gets $M_{\rm gas}$, and under the assumption that all gas is molecular, $M_{\rm gas}$ and $L_{\rm CO}$ give $\alpha_{\rm CO}$.

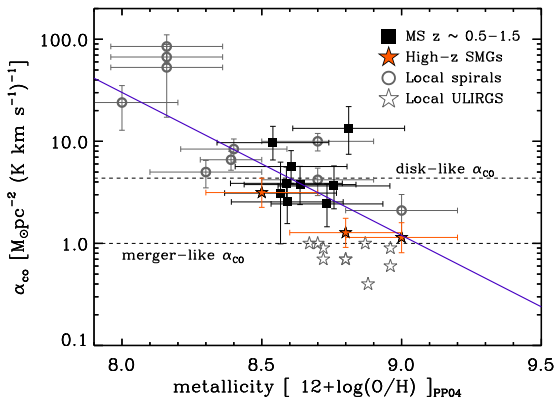



**Fig. 39** $\alpha_{CO}$ as a function of gas metallicity for several samples of local and high redshift galaxies (adopted from Magdis et al. (2012)).

2. Alternatively, if one adopts a value for $\alpha_{CO}$, the dust-to-gas ratio can be derived in the reverse order of steps.

The measurements of $\alpha_{CO}$ vs $Z$ for Magdis et al. (2012) sample and other available samples are shown in figure 39. The data are inconclusive - a trend with metallicity, a wide distribution, even bimodality cannot yet be excluded, but the main conclusion is clear - the $X_{CO}$ *is not universal*.

### 4.4.4 Cosmic Pandora Box, level 3: Which Transition Dominates?

If, by now, you are totally disenchanted with the $X_{CO}$ factor, here is an insult to your injury - in figure 40 I show a distribution of CO emission over the rotational transitions $J \rightarrow (J-1)$ for several galaxies. Even in our own Milky Way CO emits most of its energy in the $2 \rightarrow 1$ transition, in more active/merging galaxies the peak of the emission is shifted to even high transitions (i.e. higher gas temperatures). Hence, the $X_{CO}$ factor is different for different $J \rightarrow (J-1)$ transitions, so to compare apples to apples, we need to convert different observed transitions to one baseline one (say, $1 \rightarrow 0$). These new conversions factors will also depend on the galactic environment, dust temperature, perhaps redshift, etc. A hierarchy of nested Pandora boxes never ends...

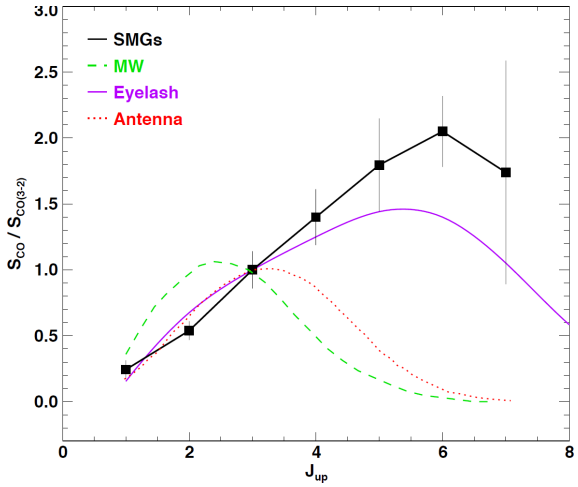



**Fig. 40** Distribution of CO emission over the rotational transition $J$ for several galaxies (colors) and the average distribution for high-redshift sub-millimeter galaxies (SMG). For most galaxies the $1 \rightarrow 0$ transition is not the dominant one (adopted from Bothwell et al. (2013)).

# 5 Star Formation

If the field of ISM is large, what one can say about star formation - it is at least another quarter of all Astronomy research. So, we must thread very carefully, or we will be lost forever in the jungle of clouds, disks, and outflows. We will attempt to stay on largest scales, and will look only on the most generic relations between gas and stars. We are not even going to paint the broad picture, we will just look at the frame...

## *5.1 Kennicutt-Schmidt and All, All, All*

For us, looking down on star formation from galactic scales and above, the story of star formation begins in March 1959, with the classical paper by Maarten Schmidt (Schmidt, 1959), who noticed that the *surface* density (and let us be precise here, we still have very little observational clues on what the *volumetric* density of star formation is doing) of star formation correlates with the surface density of gas approximately as a power-law,

$$\Sigma_{\rm SFR} \propto \Sigma_{\rm gas}^{n},$$

with $n = 1-2$.

This relationship was firmed up later by Rob Kennicutt (Kennicutt, 1989, 1998), resulting in what is nowadays commonly referred to as the Kennicutt-Schmidt (KS) relation[5],

$$\Sigma_{\rm SFR} = (2.5 \pm 0.7) \times 10^{-4} \frac{\rm M_\odot}{\rm kpc^2\,yr} \left( \frac{\Sigma_{\rm gas}}{1\,{\rm M_\odot}/\,{\rm pc}^2} \right)^{1.4\pm0.15}. \tag{30}$$

In this form the KS relation survived for 10 years. But THINGS does matter (orthography is correct), The Nearby HI Galaxies Survey was an important step in shaping our modern understanding and interpretation of the KS relation, in large part because in addition to HI, the THINGS team assembled a large amount of other data on their target galaxies, from CO emission to UV and H-$\alpha$ measurements of star formation rates.

The THINGS survey unambiguously proved what everyone knew in their hearts: stars form from molecular gas[6]. In figure 41 there is a clear strong correlation between $\Sigma_{\rm SFR}$ and the surface density of the molecular gas, but there is almost no correlation with the atomic gas. Hence, we have not wasted our time discussing the atomic-to-molecular transition, it is one of the bottlenecks that controls star formation in galaxies.

---

[5] God save you from calling it a ”law” in the presence of a devout physicist!

[6] At least, the vast majority of them - by itself, the THINGS result does not exclude a possibility of a small fraction of stars forming in the atomic gas.

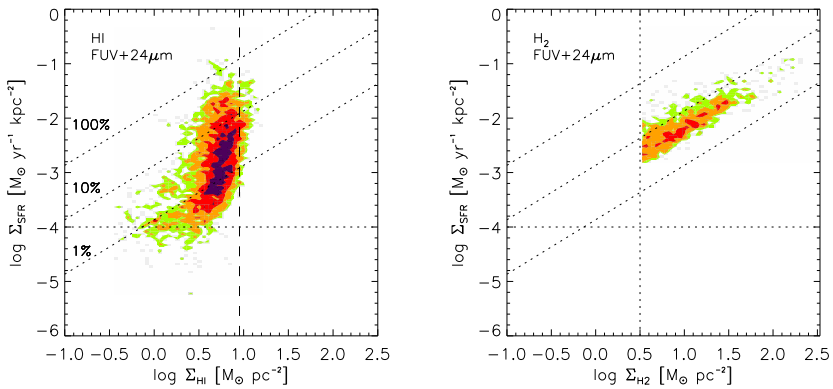



**Fig. 41** Star formation surface density as a function of HI (left) and $H_2$ surface densities from the THINGS survey (adopted from Bigiel et al. (2008)).

Historically, it was common to represent the KS relation as the relation between the star formation surface density and the surface density of the "total" gas, which actually meant the sum of atomic and molecular (i.e. *neutral*) gas. The left panel of figure 42 shows this "classical" form of KS relation from the THINGS data, together with the original measurements from Kennicutt (1998) (although the latter are tricky to interpret, since they use a different value of $\alpha_{CO}$ to convert CO emission to the molecular gas surface density).

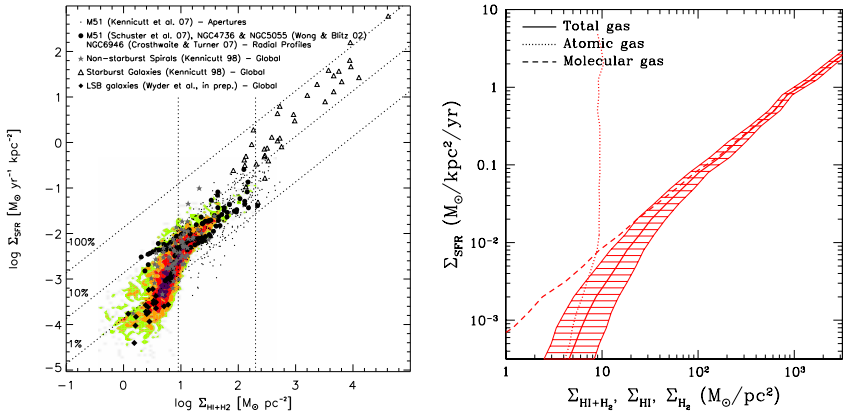



**Fig. 42** *Left*: "classical" Kennicutt-Schmidt relation from THINGS (adopted from Bigiel et al. (2008)). *Right*: separate average KS relations for the atomic (dotted), molecular (dashed), and neutral (solid) gas from the already familiar to us cosmological simulation of the Milky Way like galaxy. The solid line is the sum of the dotted and dashed along the horizontal direction.

In order to illustrate how that particular shape appears, the right panel shows the KS relation from a numerical simulation of the Milky-Way-like galaxy that we already met several times in two previous chapters. In the simulation the star formation rate is postulated to be linearly proportional to the molecular gas surface density,

$$\Sigma_{\rm SFR} = \frac{1.36\Sigma_{\rm H_2}}{\tau_{\rm SF}}, \tag{31}$$

where the factor 1.36 is, again, to account for Helium, and $\tau_{\rm SF}$ is the gas *depletion time*, assumed to be constant $\tau_{\rm SF} = 1.5\,$Gyr in the simulation (we will come back to that number shortly).

The "classical" KS relation then forms from the separate atomic $\Sigma_{\rm HI}$ and molecular $\Sigma_{\rm H_2}$ surface densities as

$$\begin{aligned} \Sigma_{\rm SFR} &= \frac{1.36}{\tau_{\rm SF}}\Sigma_{\rm H_2} + 0\times\Sigma_{\rm HI}, \\ \Sigma_{\rm HI+H_2} &= \Sigma_{\rm H_2} + \Sigma_{\rm HI}. \end{aligned}$$

The steepening of the KS relation at low surface densities is simply due to gas becoming predominantly atomic, and is fully explained by the physics of the atomic-to-molecular transition. Indeed, observations support this interpretation (but we won't dive into that question here, it is too wide and deep for us to linger in it, as we are rushing along our yellow brick road).

### 5.1.1 How we should think about star formation

If star formation correlates well with molecular gas, it is useful to think about equation (31) as our primary ansatz, and consider how $\tau_{\rm SF}$ may depend on other properties (for example, density). That thinking, however, is *totally wrong*!

A simple fact that we often forget is that density is *not even defined* without a particular scale. After all, $\rho = M/V$, and if there is no $V$, there is no $\rho$. Hence, both theoretically and observationally, we need to explicitly consider the range of spatial scales that is relevant for our problem.

Let us take some spatial scale $L$. One can imagine the whole universe divided into boxes of size $L$, like in a super-huge numerical simulation, or the universe observed with a given telescope resolution. If we average gas densities on scale $L$, they become meaningfully defined. Thus, equation (31) should really be replaced with

$$\langle\dot{\rho}_*\rangle_L = \frac{\langle\rho_{\rm mol}\rangle_L}{\tau_{\rm SF}}, \tag{32}$$

where $\rho_{\rm mol} = 1.36\rho_{\rm H_2}$ is the density of the molecular gas, and averaging is done over the spatial scale $L$. In that case depletion time becomes the function of other gas properties on scale $L$,

$$\tau_{\rm SF} = \tau_{\rm SF}(L, \langle\rho_{\rm mol}\rangle_L, ...).$$

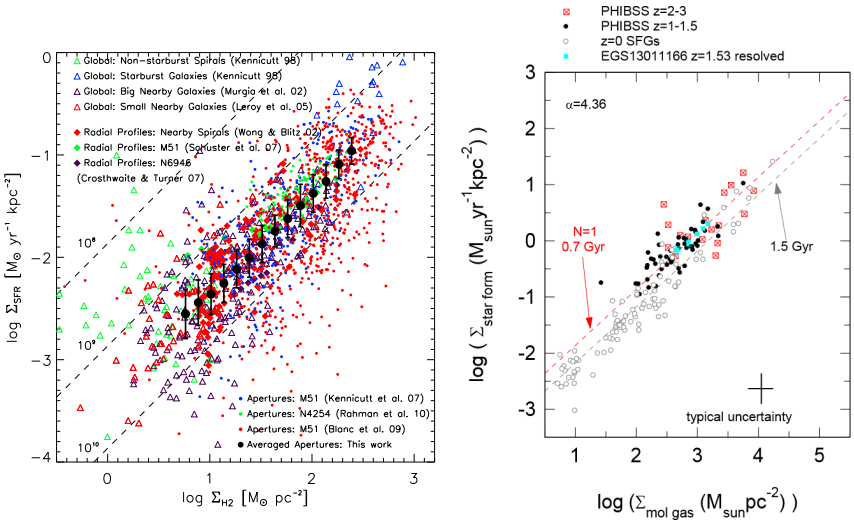



**Fig. 43** Star formation rate surface density versus the surface density of the molecular gas for local galaxies (left, Bigiel et al., 2011) and high redshift normal star forming galaxies (right, Tacconi et al., 2013).

In other words, we need to explicitly think of star formation relation as (at least) a two-dimensional relation on the plane $(L, \langle\rho_{\rm mol}\rangle_L)$, or, perhaps, even a higher-dimensional relation if chemistry, magnetic fields, properties of ISM turbulence, etc are also important.

Armed with this understanding, we can now reinterpret the existing observational constraints on various scales on a uniform basis.

In figure 43 I show the molecular KS relation for normal star forming galaxies (i.e. not ULIRGs, with their complicated CO $\rightarrow$ $H_2$ conversion) in the local universe and at high redshift. These observations sample star formation on large scales (from many hundreds of parsecs to several kilo-parsecs). They all are consistent with roughly linear KS relation,

$$\tau_{\rm SF} \approx {\rm const}(L \gg 100\,{\rm pc}),$$

although with substantial *intrinsic* (i.e. exceeding the formal observational error) scatter, and the actual value for $\tau_{\rm SF}$ noticeably different at $z = 0$ ($\tau_{\rm SF} \approx 2\,{\rm Gyr}$) and high redshift ($\tau_{\rm SF} \approx 0.7\,{\rm Gyr}$).

At the present moment (Sep 2013) it seems difficult to make any further inference from these measurements. For example, is $2 = 0.7$? In fact, they might, since local and high redshift measurements probe very different, non-overlapping ranges of gas surface density: few local observations reach $100\,{\rm M}_\odot/\,{\rm pc}^2$, while all high redshift measurement are way above that limit. The Feldmann et al. (2012a) model for the $X_{\rm CO}$ factor predicts that $X_{\rm CO}/\alpha_{\rm CO}$ factor increases gradually with the gas surface density; Feldmann et al. (2012b) present, as an example, a cosmological simulation

with constant in time and space $\tau_{\rm SF} = 1.5\,{\rm Gyr}$, which is consistent with both the low and high redshift measurements. We do not yet know how accurate the Feldmann et al. (2012a) model is, but, at the very least, there exists a plausible counterexample to any potential claim that low- and high-redshift KS relation are inconsistent with each other.

The same uncertainty applies to the scatter around the mean KS relation. We do know that the $X_{\rm CO}$ factor varies across single galaxies and between different galaxies, so some fraction of the scatter should be due to scatter in $X_{\rm CO}$. In addition, there is scatter from the time dimension that we have so far ignored: CO emission from the molecular gas is essentially instantaneous, but observational estimates of star formation are not. For example, Schruba et al. (2010) show that the depletion time is systematically higher around peaks of CO emission (molecular clouds where star formation is just starting) than around peaks of H-$\alpha$ emission (star forming regions where star formation is well under way).

This difference is purely due to the fact that we do not measure the instantaneous rate of star formation, but use observational indicators that return a time-averaged star formation rate over some characteristic time-scale ($\sim 20\,{\rm Myr}$ for UV light, $\sim 5\,{\rm Myr}$ for H-$\alpha$). Hence, if we point our telescope on a freshly formed molecular cloud, we will see a lower star formation rate than the actual instantaneous one - if the cloud has been forming stars for only $1\,{\rm Myr}$ and we use H-$\alpha$, we will measure a $5\,{\rm Myr}/1\,{\rm Myr} = 5$ times lower star formation rate than the true one. Now, if we point it at a mature star forming region, we will measure a higher time-averaged star formation rate than the instantaneous one, because $5\,{\rm Myr}$ ago the region contained more molecular gas (and, hence, higher instantaneous star formation rate) than it has right now.

The combined scatter due to variations in the $X_{\rm CO}$ factor and the finite time-averaging is easily quantifiable, though, and appears to be less than the actual observed scatter in the KS relation (Feldmann et al., 2012b). This should not be particularly surprising, though - it is hard to imagine that the nature is so kind to us as to make each region of space with the same surface density of molecular gas to have *exactly* the same star formation rate, surely there must be random or systematic variation from place to place that affect star formation rate, and that will appear as the true intrinsic scatter in equation (32).

There exist several other constraints we can place on $\tau_{\rm SF}(L, \langle\rho_{\rm mol}\rangle_L, ...)$. Lada et al. (2010) found that on the scale of individual star-forming cores ($\sim 1\,{\rm pc}$) the depletion time is also constant (i.e. independent of density) and is about $20\,{\rm Myr}$, but only if the density is above $\rho_{\rm min} = 700\,{\rm M}_\odot/\,{\rm pc}^3$. A threshold must exist in that case, since any small-scale relation must be consistent with the large-scale one. Namely, if on $1\,{\rm pc}$ scale we have

$$\langle\dot{\rho}_*\rangle_1 = \begin{cases} \dfrac{\langle\rho_{\rm mol}\rangle_1}{20\,{\rm Myr}}, & \rho_{\rm mol} > \rho_{\rm min}, \\ 0, & \rho_{\rm mol} < \rho_{\rm min}, \end{cases}$$

and on $500\,{\rm pc}$ scale we have a usual molecular KS relation,

$$\langle \dot{\rho}_* \rangle_{500} = \frac{\langle \rho_{\rm mol} \rangle_{500}}{2\,{\rm Gyr}},$$

then these two relations can be mutually consistent if and only if exactly 1% of the molecular gas sits above the small-scale density threshold $\rho_{\rm min}$ - after all,

$$\langle \dot{\rho}_* \rangle_{500} = \left\langle \langle \dot{\rho}_* \rangle_1 \right\rangle_{500}.$$

Another commonly used ansatz for the star formation rate is constant efficiency per free-fall time,

$$\tau_{\rm SF}(L, \langle \rho_{\rm mol} \rangle_L, ...) = \frac{\tau_{\rm ff}(\langle \rho_{\rm mol} \rangle_L)}{\varepsilon_{\rm SF}} = \varepsilon_{\rm SF}^{-1} \sqrt{\frac{3\pi}{32G\langle \rho_{\rm mol} \rangle_L}},$$

or, in a more familiar form,

$$\langle \dot{\rho}_* \rangle_L = \varepsilon_{\rm SF} \frac{\langle \rho_{\rm mol} \rangle_L}{\tau_{\rm ff}} = \varepsilon_{\rm SF} \frac{\langle \rho_{\rm mol} \rangle_L^{3/2}}{\sqrt{3\pi/(32G)}}. \tag{33}$$

The origin of that formula disappears in the depths of time; it is often used without any attention to the scale under consideration. In an influential paper, Krumholz & Tan (2007) argued that many observational constraints are consistent with that ansatz[7] with $\varepsilon_{\rm SF} \approx 1 - 2\%$ for a wide array of molecular densities, from average molecular clouds to molecular cores.

However, observational constraints used by Krumholz & Tan (2007) sample not only various densities, but also *various spatial scales*; namely, they all fall along a particular track $L^2 \times \langle \rho_{\rm mol} \rangle_L \approx 10^4\,{\rm cm}^{-3}\,{\rm pc}^2$ in the two-dimensional plane $(L, \langle \rho_{\rm mol} \rangle_L)$. In other words, observational constraints that support the "constant efficiency per free-fall time" are equally well support the "constant efficiency per unit scale",

$$\langle \dot{\rho}_* \rangle_L = \varepsilon_{\rm SF} \frac{\langle \rho_{\rm mol} \rangle_L}{\tau_{\rm ff}} \approx \varepsilon_{\rm SF} \frac{\langle \rho_{\rm mol} \rangle_L}{\tau_{\rm ff}(10^4\,{\rm cm}^{-3})(L/1\,{\rm pc})}.$$

The two alternatives cannot be distinguished at present without additional observational constraints.

In fact, I am going to make a bold claim (and challenge anyone to refute it) that *all of the existing observational constraints are consistent with the linear (in density) star formation ansatz* in which the depletion time is function of scale only,

$$\langle \dot{\rho}_* \rangle_L = \begin{cases} \dfrac{\langle \rho_{\rm mol} \rangle_L}{\tau_{\rm SF}(L)}, & \rho_{\rm mol} > \rho_{\rm min}(L), \\ 0, & \rho_{\rm mol} < \rho_{\rm min}(L), \end{cases} \tag{34}$$

[7] One should never forget that the "constant efficiency per free-fall time" model is no more than an ansatz; molecular clouds are turbulent and the free-fall time has no physical relevance on scales above the sonic length.

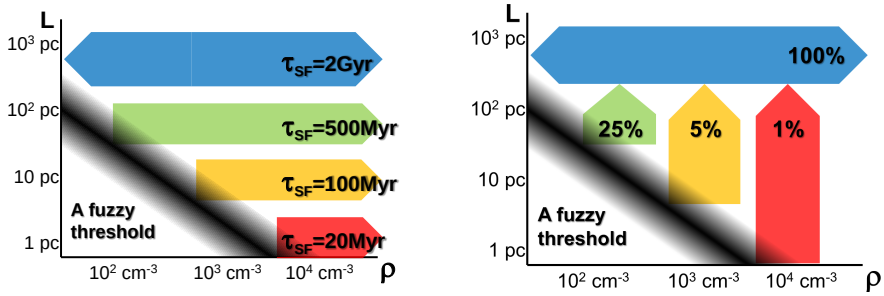



**Fig. 44** Cartoon version of contours of constant depletion time (shown as different colored bands) on the $(L, \langle\rho_{\rm mol}\rangle_L)$ plane. The left panel shows the linear star formation relation (34), while the right panel shows the constant-efficiency-per-free-fall-time star formation relation (33). In the latter case the depletion function must transition to a constant on the largest scale somehow to be consistent with large-scale KS relation.

with

$$\tau_{\rm SF}(L) \sim 2\,{\rm Gyr} \times \min\left(1, \frac{L}{L_0}\right),$$

$$\rho_{\rm min}(L) \sim \rho_0 \times \min\left(1, \frac{L_0^2}{L^2}\right),$$

and $L_0$ is in the range of a few hundred parsecs (for example, the scale height of the gaseous disk). In the Milky Way galaxy $\rho_0$ is such that the Lada et al. (2010) result is matched ($\rho_{\rm min}(L) \approx 700\,{\rm M}_\odot/\,{\rm pc}^3$ at $L \sim 1\,{\rm pc}$), but in other galaxies it may be different (for example, being proportional to the density of the atomic-to-molecular transition).

Figure 44 shows two alternatives (linear star formation relation 34 and constant-efficiency-per-free-fall-time star formation relation 33) in a cartoon fashion. At present, either one is a sensible model, as well as any other, intermediate or more complex model, that still matches the observational constraints.

## *5.2 Excursion Set Formalism in Star Formation*

The idea of using Excursion Set formalism in star formation is based on a well established fact: in isothermal supersonic turbulence the density PDF is lognormal, in a direct analogy with the Gaussian distribution of the linear overdensity $\delta$. Such an approach was first attempted by Padoan & Nordlund (2002), picked up later by Hennebelle & Chabrier (2008) and developed much further by Phil Hopkins in a recent series of papers (Hopkins, 2012a,b, 2013).

### 5.2.1 Refresher: Excursion Set Formalism

Excursion Set formalism (sometimes also called "Press-Schechter formalism") deals with a *Gaussian random field* $\delta(\mathbf{x})$ (and $\delta$ can be anything, for supersonic turbulence it will be $\ln(\rho/\rho_0)$). For a Gaussian random field different wavenumbers of the Fourier transform

$$\delta_{\mathbf{k}} \equiv \int d^3x \delta(\mathbf{x}) e^{i\mathbf{kx}}$$

of the field are uncorrelated,

$$\langle \delta_{\mathbf{k}_1} \delta^*_{\mathbf{k}_2} \rangle = P(k_1) \delta^3_D(\mathbf{k}_1 - \mathbf{k}_2).$$

One can reverse the Fourier transform,

$$\delta(\mathbf{x}) = \int d^3k \sqrt{P(k)} \lambda_{\mathbf{k}} e^{-i\mathbf{kx}}, \tag{35}$$

with uncorrelated, normally distributed random numbers $\lambda_{\mathbf{k}}$ satisfying the relation $\langle \lambda_{\mathbf{k}_1} \lambda^*_{\mathbf{k}_2} \rangle = \delta^3_D(\mathbf{k}_1 - \mathbf{k}_2)$. Equation (35) should be considered in a generalized sense (similar to Dirac delta-function), because for some $P(k)$ the integral may actually diverge. In that case $\delta(\mathbf{x})$ should be considered as a limit of the smoothed density field,

$$\delta(\mathbf{x}) \equiv \lim_{R \to 0} \delta(\mathbf{x}; R) = \int d^3k \sqrt{P(k)} \lambda_{\mathbf{k}} W(kR) e^{-i\mathbf{kx}},$$

where $W(kR)$ is a low-pass filter ($W(0) = 1$, $W(\infty) = 0$). An example of a Gaussian random field at 3 different resolutions is shown in figure 45.

Excursion Set formalism considers $\delta(\mathbf{x}; R)$ as a function of the filter scale $R$ and compares it with some barrier function $b(R)$. Obviously, $\delta(\mathbf{x}, R = \infty) = 0$. As $R$ decreases, $\delta(\mathbf{x}; R)$ starts deviating from zero. For some value of $R$ it may cross the barrier for the first time. The fraction of all $\delta(\mathbf{x}; R)$ that cross the barrier at $R$ is called the first crossing distribution. For example, in the canonical Press-Schechter

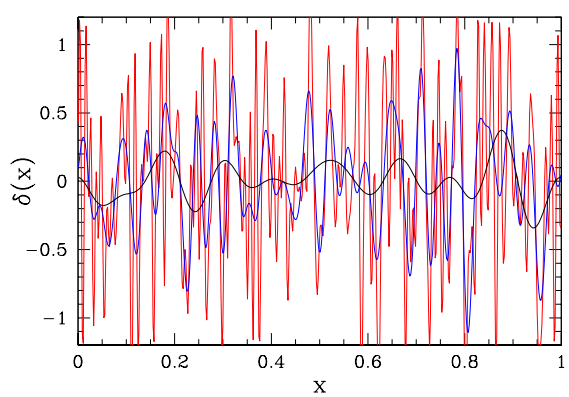



**Fig. 45** Example of the Gaussian random field at three different values for the smoothing scale $R$.

formalism the barrier is constant, $b = \delta_L(t_f) = 1.69$. Then the first crossing distribution becomes (half) the mass function of dark matter halos with $M_h = 4\pi\bar{\rho}_m R^3/3$.

In modeling star formation Excursion Set formalism can be used for several goals:

- First crossing distribution gives the mass function of largest bound objects - molecular clouds.
- Last crossing distribution gives the mass function of smallest bound objects - molecular cores/stars.
- It is useful for other purposes too: distribution of holes in the ISM, clustering of stars, etc.

One only needs to define a barrier - but we have already considered it! After all, gas collapses when it becomes gravitationally unstable, hence the barrier is simply the stability criterion for the disk with finite thickness, our equation (24) - with a minor correction of adding turbulent velocity dispersion $\sigma_t^2$ to the gas sound speed $c_S^2$, since turbulence also provides support against gravitational collapse.

Excursion Set formalism makes predictions that are computable analytically and match a large variety of observations unexpectedly well (see figure 46). The final verdict on this novel approach is still pending, with opinions ranging from "it should never work" to "it solves all the problems". So, if you are bold enough, make your bet...

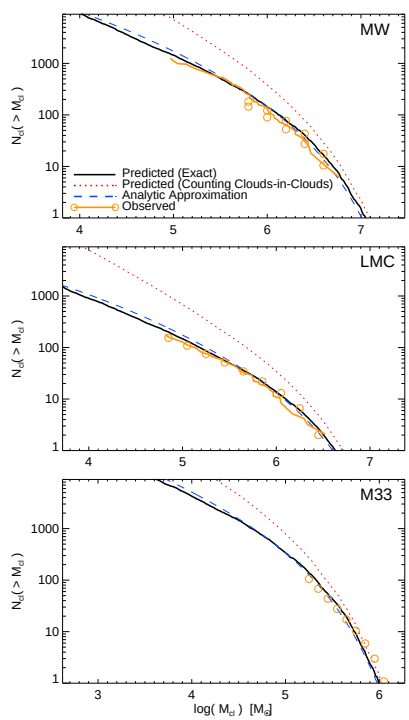
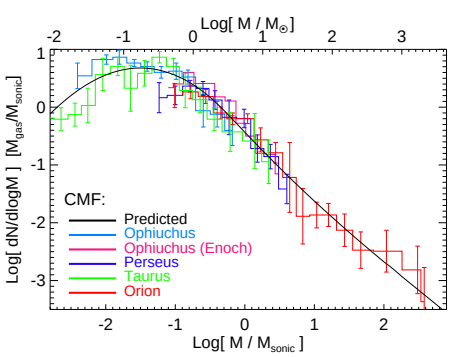



**Fig. 46** Several predictions of the Excursion Set formalism as a theory of star formation: GMC mass functions for the Milky Way, LMC, and M31 (right, adopted from Hopkins (2012a)) and clump mass function (bottom, adopted from Hopkins (2012b)).



# 6 Stellar Feedback

Stars affect their environments by their *feedback* - anyone reading these lines knows that well, without the stellar feedback we would not even exist (as there would not exist planets made out of heavy elements).

## *6.1 What Escapes from Stars*

However, stellar feedback is not just supernovae (sometimes that's the impression one gets by reading simulation papers). Stars affect their environments in several ways: *supernovae* (both type II and type Ia), *stellar winds* from massive stars, *mass loss* from AGB and planetary nebulae, and, of course, *radiation*. Each of these modes inject into surrounding gas *energy*, *momentum*, *mass*, *metals*, *dust*, and *cosmic rays*. The various pathways the inputs and outpust are connected are shown in figure 47. If you think the feedback is complicated, then you are right!

It is easy to get lost in this maze of feedback pathways. But one important fact should light our way (literally) - by far the largest energy output of stars is light! Just as an illustration, I show in figure 48 the energy production rate as a function of time for a normal stellar population at solar metallicity. The bolometric luminosity of stars dwarfs all other feedback channels at all times. And we know that at least half of that energy is re-radiated in the infrared by dust, hence a substantial fraction of stellar light is indeed absorbed by the surrounding gas. We should, therefore, consider that feedback channel very seriously.

### 6.1.1 Radiation Pressure

Massive (hence, young) stars spend a substantial fraction of their lives heavily embedded into the surrounding gas and dust; for heavily obscured stars most of their

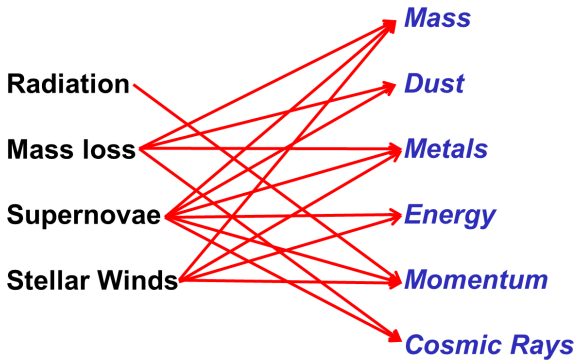



**Fig. 47** Various pathways for stellar feedback.

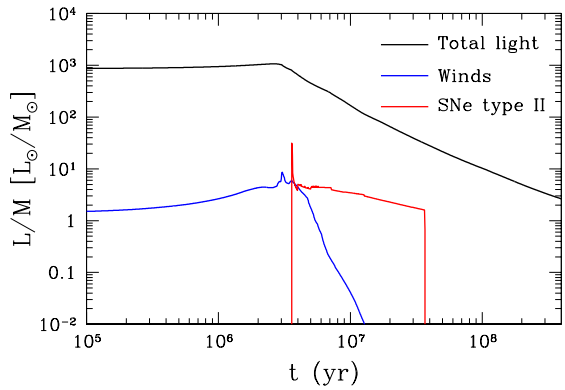



**Fig. 48** Energy injection rate (= luminosity) per unit mass for the total radiation, winds, and supernovae as a function of time for a normal stellar population at solar metallicity.

light is absorbed. Since photons have momentum, absorbing all light from a star/star cluster of luminosity $L$ injects momentum into the surrounding gas,

$$\dot{p}_1 = \frac{L}{c}.$$

The energy, however, is conserved - the absorbed bolometric luminosity of the star must be re-emitted by dust in the infrared. If there is enough dust around a young massive star (and, at least in the Milky Way, there is), the dust will be optically thick to its own IR radiation. That radiation will do work on the surrounding gas, i.e. will inject extra momentum, so that the total momentum injection rate is

$$\dot{p}_{\rm tot} = (1 - f_{\rm esc} + \tau_{\rm IR})\,\frac{L}{c}, \tag{36}$$

where, in order to be completely general, I included the fraction $f_{\rm esc}$ of stellar radiation (of all frequencies) that escapes the star forming region. The new factor $\tau_{\rm IR}$ is easy to derive for a homogeneous medium (Gayley et al., 1995). Since energy is conserved, the radiation flux at each radius $R$ from the star is still

$$F_R = \frac{L}{4\pi R^2}.$$

Hence, the momentum (in the radial direction) imparted on the gas between $R$ and $R + dR$ is simply

$$d\dot{p}_{\rm IR} = 4\pi R^2 \frac{F_R}{c}\kappa dR = \frac{L}{c} d\tau,$$

and, hence,

$$\dot{p}_{\rm IR} = \tau_{\rm IR}\frac{L}{c}.$$

In the infrared dust opacity is

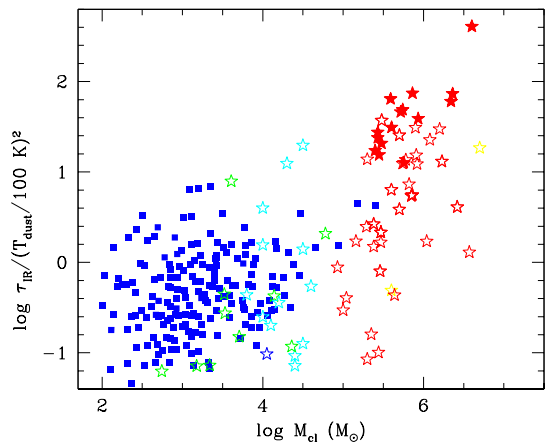



**Fig. 49** Observational estimates of $\tau_{IR}$ for GMC clumps (blue squares) and young stellar clusters (stars). Adopted from Agertz et al. (2013a).

$$\kappa_{IR} \approx 3\frac{\mathrm{cm}^2}{\mathrm{g}}\left(\frac{T_d}{100\mathrm{K}}\right)^2$$

(Semenov et al., 2003). Observational estimates of $\tau_{IR}$ at $T_d = 100\,\mathrm{K}$ are shown in figure 49; radiation pressure is particularly important for large stellar clusters.

### 6.1.2 All Feedback, All the Time

Stars, of course, do not have a freedom to selectively fire only some of their feedback channels, they all work all the time. In figure 50 the time evolution of the momentum and energy injection is shown for several dominant feedback modes. Supernovae form the last episode in the feedback fireworks - by the time they start in earnest (after about 10 Myr after the onset of star formation) all other feedback channels have already finished.

In fact, with all likelihood, supernovae are not important to actually destroying molecular clouds (and, hence, controlling the efficiency of star formation, Fall et al., 2010). They may be important for heating the overall ISM, for stabilizing the disk, for driving galactic winds, but star formation they control not.

The relative role of different feedback channels can be understood even better with a simple numerical exercise - a single computational cell ”simulation”. Figure 51 shows the fate of a such a cell when various feedback channels are switched on and off.

The first important lesson is that purely thermal feedback - injecting all of the supernova energy as thermal energy into the parent cell (or particle in case of SPH) of even a large stellar cluster does not do anything, the cooling times are always so short that the thermal energy is quickly radiated away. This is not a new result, it has been known since the dawn of numerical galaxy formation, and re-discovered

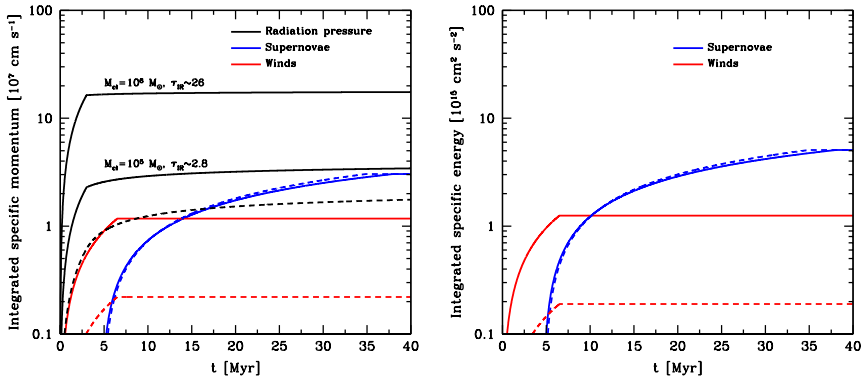



**Fig. 50** Time evolution of the cumulative injected specific momentum (left) and specific energy (right) for various feedback channels at solar metallicity (solid lines) and at $Z = 0.01Z_{\odot}$ (dashed lines). Notice that supernovae kick in after all other feedback channels already fired. Adopted from Agertz et al. (2013a).

independently by many research groups; but it does pose a dilemma for cosmological and even galactic-scale simulations - the only direct way of implementing stellar feedback does not work, and one has to use a *sub-grid model*, i.e. a specific recipe about how to implement the feedback in a numerical code.

**Fig. 51** Density and temperature in a computational cell evolving in response to various feedback channels: only energy from supernovae (red), energy and momentum from supernovae (magenta), only momentum feedback from all channels (blue), and all feedback channels acting together (black). Adopted from Agertz et al. (2013a).

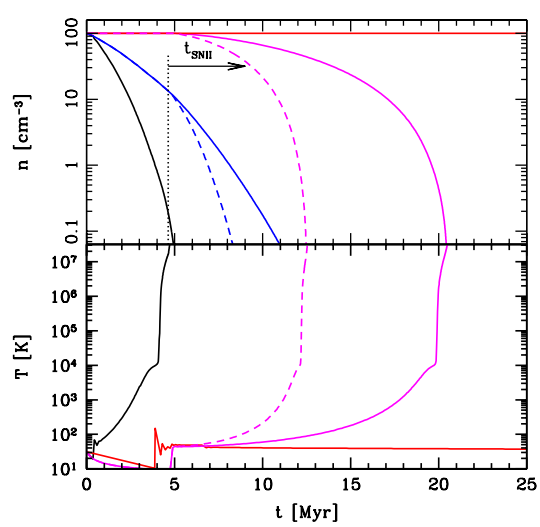


Figure 51 may offer a clue how such a sub-grid model may be implemented: other feedback channels produce a large effect on the dynamics of a single cell, and, hence, may have a significant effect on the dynamics of larger scales as well. There is just one problem with that approach - actual stellar feedback like equation (36) is operating on scales of molecular cores and their very vicinities, on sub-parsec scales. Whenever we use, say, the radiation pressure formula in cosmological (or even galactic-scale) simulations, we are injecting the momentum on scales of many tens, even hundreds of parsecs, well beyond the range of scales where it is actually operating. Hence, using equation (36) in a cosmological code is also a *sub-grid model*, an ansatz that is a priori as good or bad as any other sub-grid model. Not surprising, then, that the radiation pressure is gradually falling out of fashion.

Before we trash all sub-grid models or pick one of them and place it on the throne, it is worth taking a step back and re-thinking what we are actually trying to achieve.

## *6.2 Unconventional Marriage: Feedback and Star Formation*

If you did not skip the previous chapter, my dear reader, then you know that star formation is *inefficient* - the depletion time $\tau_{\rm SF}$ is of the order of a Gyr (give or take a factor of 2-3), while molecular clouds are short-lived ($10-20$ Myr). During their lifetimes molecular clouds convert only a small fraction, mere percents, of their gas into stars. A natural conclusion from that fact is, since star formation is inefficient, then so must be the feedback. And that conclusion is utterly wrong!

Do you recall a simulation of a Milky Way like galaxy that I used to illustrate the properties of the gaseous halo (figure 17)? Have you wondered why I never showed you the circular velocity profile for that galaxy? There is a good reason I have not - I am ashamed to! While the gaseous halo for that galaxy may look ok, the disk is totally wrong, it has an extremely dense spike at the center, with the circular velocity peaking at 450 km/s, more than twice the rotation velocity of the Milky Way. The reason for such a huge discrepancy is the absence of any feedback process in the simulation.

If we learned anything after 20+ years of modeling galaxy formation, then it is that the central spikes in circular velocity (caused by unrealistic central mass concentrations) can only be destroyed by strong feedback. No other physics can do the trick - in fact, as simulations grew more sophisticated, included more physics, and reached higher resolution, the central mass concentration problem became worse. It is a real, physical problem, not a numerical one - the high-redshift progenitors of normal galaxies are too dense, and these early dense gaseous concentrations survive all the subsequent adventures of galaxy evolution; if not blown out, they will become large stellar bulges.

Indeed, that was commonly occurring in simulations until only a few years ago - for example, check out beautiful pictures of center-heavy galaxies in Stinson et al. (2010). At the same time, as observers figured out the difference between the real

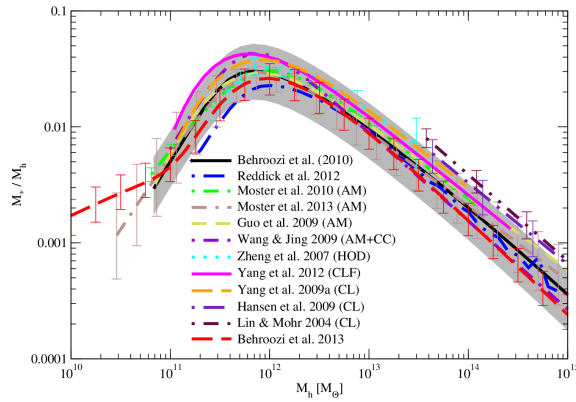



**Fig. 52** Abundance matching in action: stellar vs total mass for dark matter halos from several independent groups; they all agree that star formation is inefficient. Adopted from Behroozi et al. (2013).

bulges and pseudo-bulges (central features formed by secular evolution from barred disks), they realized that a significant fraction, perhaps as much as 50%, of galaxies are actually *bulgeless*, pure disks.

All these examples illustrate one crucially important conclusion about star formation and feedback - while start formation is *inefficient*, the feedback is actually *strong*. These two facts may be deeply connected, but we are not going to dive into the connection between star formation and feedback, for our purpose what is important is this apparent dichotomy in behavior.

A good illustration of that dichotomy comes from the so-called "abundance matching" exercise - a match between the *observationally derived*[8] stellar mass and the theoretically known mass function of dark matter halos. Such a match results in a one-to-one correspondence between the stellar mass and the halo mass for individual halos (or, in a more complex implementation of the abundance matching idea, a distribution of stellar masses for a given halo mass). Figure 52 shows a comparison (and uncertainty) of the stellar mass - halo mass relation for several independent applications of that approach from Behroozi et al. (2013).

The most important result of the abundance matching exercise is that stellar masses of low mass halos are very small, roughly $M_* \propto M_h^{2.5}$ for $M_h \ll 10^{12}\,\mathrm{M}_\odot$. To get such a behavior, it is not enough to make star formation inefficient - that would still result in stellar mass being proportional to halo mass, the inefficiency would only make the coefficient of proportionality small - but it also requires the feedback to be progressively more efficient in lower mass galaxies to sculpt the inferred $M_* \propto M_h^{2.5}$ relation.

Attempts to model the feedback as a sub-grid model are as old as the galaxy formation simulations themselves. It is not too instructive to review all of them, as until 2010 none of the sub-grid models were particularly successful. The important thing to remember about any sub-grid model is that it would only work over a finite range of spatial scales. If the model is good, that range would be sufficiently large

[8] Never forget that stellar masses are *not* observed, they are always derived from observations of luminosity functions, with all the inherent in spectral synthesis uncertainties and biases.

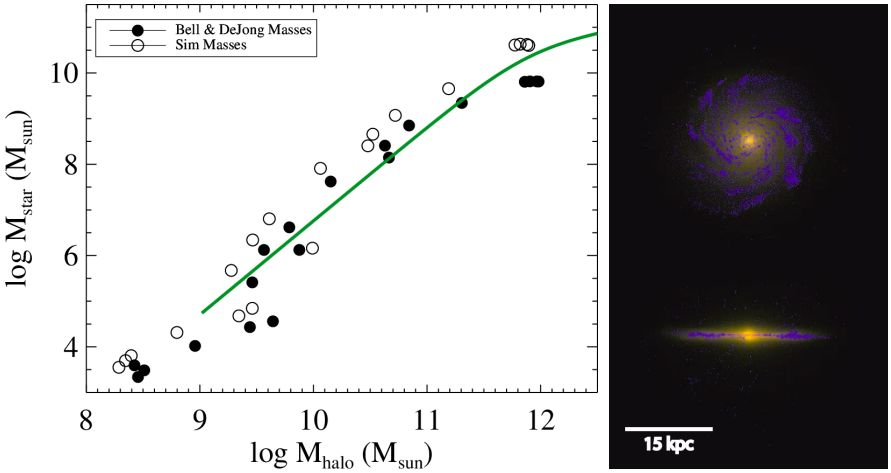



**Fig. 53** *Left:* stellar versus total mass from abundance matching (green line) and modern galaxy formation simulations (open and filled circles - adopted from Munshi et al. (2013)). *Right:* face- and edge-on projections of Eris simulation of the Milky Way galaxy (adopted from Guedes et al. (2011)).

(say, a decade in spatial scale); if the model is bad, the range may be zero. Even if the model is good, but its range of validity does not match the resolution of simulations, then it would not work well.

Indeed, that is what have happened with one simple sub-grid feedback model. In 1997 in his PhD thesis, Jeroen Gerritsen proposed a simple way to make feedback strong - simply to disable cooling in star forming regions for several tens of Myr (we now call this method "delayed cooling"). The model did not work too well with the spatial resolution simulations were able to reach in 1997. However, miracles do happen - as the resolution improved, the delayed cooling model appeared to work better and better, until, finally, in 2010 it was declared to be a panacea for galaxy formation (Governato et al., 2010)!

Figure 53 gives two examples of how well modern simulations with delayed cooling feedback reproduce observations, but similarly impressive examples for various observational constraints are abound.

### 6.2.1 Why Delayed Cooling Works

While it is easy to declare the delayed cooling a success, it is much harder to understand what it actually means. As such, it is just a numerical trick, without any serious physical justification. The fact that it works may be a pure coincidence; alternatively, it can be a manifestation of a real physical process that operates on sub-parsec scales, but its consequences on $\sim 100$ pc scales appear as if cooling was switched off. In fact, it is easy to come up with several real physical processes that will all manifest themselves as delayed cooling on large scales:

- radiation pressure from massive stars (we now know it is important) provides support for gas that "does not cool", i.e. if treated as an effective additional pressure, that pressure would not be affected by the cooling processes in the gas, but will diminish after about 10 Myr;
- coronal gas - the hot, million-degree gas produced in supernova explosions may accumulate in regions of low density in a supersonically turbulent IGM; cooling times in such gas will depend on its density, but generally will be of the order of several to several tens of Myr;
- as stellar feedback continue to stir supersonic turbulence in molecular clouds on small scales, the energy of the kinetic motions will accumulate to the point at which the dissipation rate will approximately equal the production rate; while the dissipation time-scale is likely to be short, the supersonic turbulence (i.e. highly super-thermal additional pressure in the gas) will be maintained for the duration of stellar feedback, several tens of Myr;
- cosmic rays produced in supernova explosions are observationally known to provide significant additional support in the magnetized molecular clouds; cosmic rays diffuse out of GMC on time-scales of tens of Myr.

I am sure that list can be easily extended, but it already serves our purposed well - numerous real small-scale physical processes may hide themselves under the large-scale mask of "delayed cooling", and one, several, all of them, or different combinations of them in different environments may be the actual feedback process(es) that is/are responsible for making the real galaxies as they are...

### *6.3 Toward The Future*

So, where do we go from there? If only we could figure out which of the actual feedback channels hides behind the mask of delayed cooling, the galaxy formation (of normal disk galaxies - the AGN feedback is entirely different story) will be essentially solved (well, hopefully you do not take me as being too optimistic).

These feedback processes are more-or-less understood, as, hopefully, I persuaded you in the beginning of this chapter. Actually modeling them in the cosmological and galactic-scale simulations is not trivial, but big strides in that directions have been already made. We may still argue occasionally how to do it better, or what the most appropriate value for, say, $\tau_{\rm IR}$ should be, but the importance of the modeling feedback correctly is not a subject of debate any more.

The good piece of news is that even if various feedback channels are tuned to match the basic observational constraints like the stellar mass vs total halo mass, Kennicutt-Schmidt relation, rotational velocity curves, etc, simulated galaxies in runs with different feedback channels still look amazingly different (figure 54), and there lies the key to the eventual success.

Hence, the plan for the future is to identify the best observational probes that will help us understand which of all of the potential feedback channels are important in

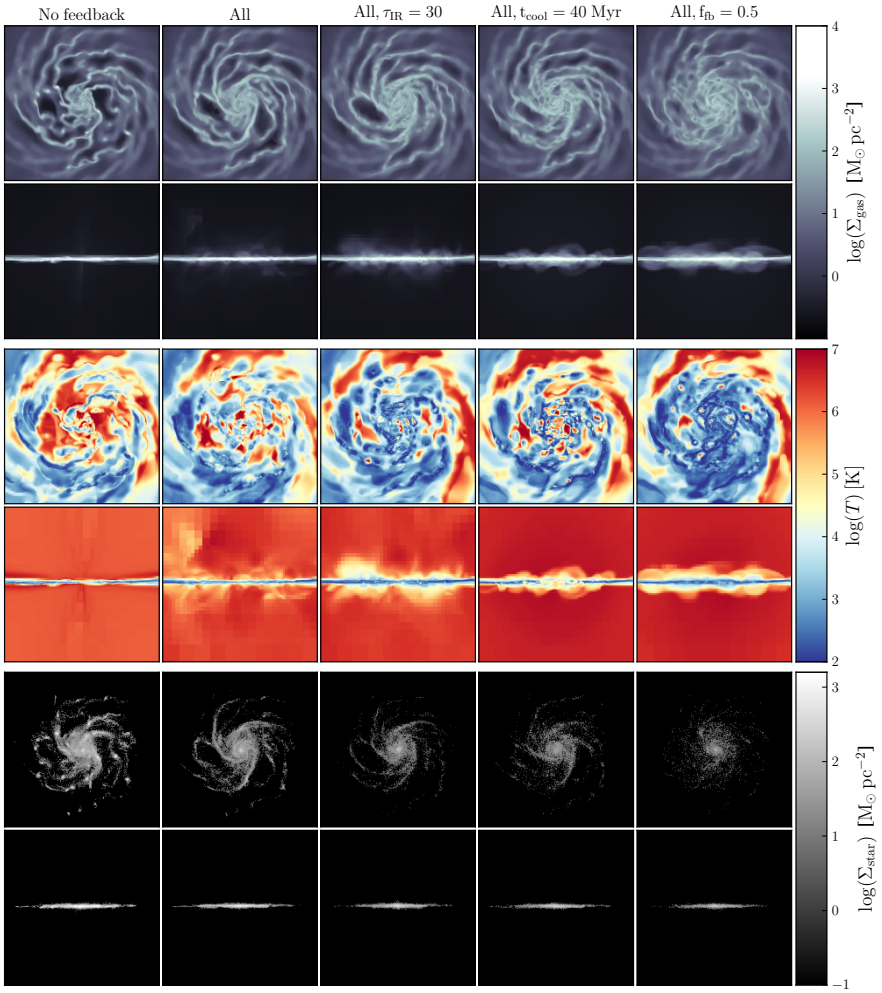



**Fig. 54** Face-on and edge-on maps of simulated galactic disks with different small-scale feedback models. Separate rows show gas surface density (top), mass weighted average gas temperature (middle), and stellar surface density (bottom). Columns from left to right are no feedback, all feedback channels from figure 47, all feedback with extra radiation pressure, delayed cooling feedback, feedback model with extra energy variable (adopted from Agertz et al. (2013b)).

which environments and on what spatial scales. This is left as an exercise for the reader...

## 7 Answers To Brain Teasers

1. Sound waves indeed do not dissipate. However, they also do not grow with time, since they are stable perturbations, while large-scale, unstable perturbations in both dark matter and gas grow. Hence, relative to the large-scale perturbations, the small-scale sound waves become smaller and smaller, i.e. they appear to be "suppressed".
2. The proper term for "Lyman-$\alpha$ absorption" is *resonant scattering*. A Lyman-$\alpha$ photon is re-emitted by the atom, but in the meantime that atom experienced a large number of collisions with other atoms and ions, so its momentum is now unrelated to the momentum it had at the moment of absorption. Hence, the re-emitted Lyman-$\alpha$ photon will be send out into a random direction in the frame of the atom, and will not reach our telescope. For us, that photon is lost, hence we, sometimes, call it absorption.
3. The term "equation of state" relates the perturbations in the gas pressure (or temperature) to those of the density. If we impose (adiabatically) a perturbation $\delta\rho$ to the gas density, the instantaneous response to the pressure will be identical to the ideal gas, $\delta P = c_S^2 \delta\rho$. Only with time adiabatic expansion and photoheating will bring that perturbation back to the temperature-density relation.
4. A typical ionizing photon is not sitting at the Lyman edge, it has the energy of $E_0 + \langle \Delta E \rangle$ (see equation 9), which is about $40 - 50\,\mathrm{eV}$ for the cosmic background. The ionizing cross-section falls off with energy as $E^{-3}$, hence the typical cross-section is $\sim (1-2) \times 10^{-19}\,\mathrm{cm}^2$ instead of $6.3 \times 10^{-18}\,\mathrm{cm}^2$. In addition, the typical ionization level in the forest is $10^{-5}$, which requires $\tau = \ln(10^5) \approx 10$ to neutralize. Hence, hydrogen absorbers only become fully neutral at column densities of $N_H \sim (0.5-1) \times 10^{20}\,\mathrm{cm}^{-2}$.
5. This one is really tricky. In fact, I do not know the full answer to it. One possible reason why Lyman-$\alpha$ forest is not turbulent was suggested to me by Andrea Ferrara: for turbulence to develop, the gas needs to have vorticity, but in the linear regime vorticity in cosmic gas decays, so there should be no vorticity at $\delta \approx 0$ in the forest. Non-linear evolution will generate some vorticity, but since most of the forest is not extremely non-linear, it is plausible that the vorticity generated in the forest may not be enough to create a full turbulent cascade.

## References


Agertz, O., Kravtsov, A. V., Leitner, S. N., & Gnedin, N. Y. 2013a, ApJ, 770, 25

—. 2013b, ApJ, 770, 25

Altay, G., Theuns, T., Schaye, J., Crighton, N. H. M., & Dalla Vecchia, C. 2011, ApJL, 737, L37

Anderson, M. E. & Bregman, J. N. 2010, ApJ, 714, 320

Barnes, J. & Efstathiou, G. 1987, ApJ, 319, 575

Begelman, M. C. & Shlosman, I. 2009, ApJL, 702, L5

Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57

Bigiel, F., Leroy, A., Walter, F., Brinks, E., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2846

Bigiel, F., Leroy, A. K., Walter, F., Brinks, E., de Blok, W. J. G., Kramer, C., Rix, H. W., Schruba, A., Schuster, K., Usero, A., & Wiesemeyer, H. W. 2011, ApJL, 730, L13+

Binney, J. & Tremaine, S. 1987, Galactic dynamics

Blitz, L. & Robishaw, T. 2000, ApJ, 541, 675

Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ArXiv e-prints

Bothwell, M. S., Smail, I., Chapman, S. C., Genzel, R., Ivison, R. J., Tacconi, L. J., Alaghband-Zadeh, S., Bertoldi, F., Blain, A. W., Casey, C. M., Cox, P., Greve, T. R., Lutz, D., Neri, R., Omont, A., & Swinbank, A. M. 2013, MNRAS, 429, 3047

Chen, H.-W. 2012, MNRAS, 427, 1238

Chen, H.-W., Perley, D. A., Pollack, L. K., Prochaska, J. X., Bloom, J. S., Dessauges-Zavadsky, M., Pettini, M., Lopez, S., Dall'aglio, A., & Becker, G. D. 2009, ApJ, 691, 152

Croft, R. A. C., Weinberg, D. H., Katz, N., & Hernquist, L. 1998, ApJ, 495, 44

Dall'Aglio, A., Wisotzki, L., & Worseck, G. 2008, A&Ap, 491, 465

Draine, B. T. 1978, ApJS, 36, 595

Draine, B. T. & Bertoldi, F. 1996, ApJ, 468, 269

Fall, S. M., Krumholz, M. R., & Matzner, C. D. 2010, ApJL, 710, L142

Feldmann, R., Gnedin, N. Y., & Kravtsov, A. V. 2012a, ApJ, 747, 124

—. 2012b, ApJ, 758, 127

Gayley, K. G., Owocki, S. P., & Cranmer, S. R. 1995, ApJ, 442, 296

Glover, S. C. O. & Abel, T. 2008, MNRAS, 388, 1627

Glover, S. C. O. & Mac Low, M.-M. 2011, MNRAS, 412, 337

Gnedin, N. Y. 2012, ApJ, 754, 113

Gnedin, N. Y., Baker, E. J., Bethell, T. J., Drosback, M. M., Harford, A. G., Hicks, A. K., Jensen, A. G., Keeney, B. A., Kelso, C. M., Neyrinck, M. C., Pollack, S. E., & van Vliet, T. P. 2003, ApJ, 583, 525

Gnedin, N. Y. & Hollon, N. 2012, ApJS, 202, 13

Gnedin, N. Y. & Hui, L. 1998, MNRAS, 296, 44

Gnedin, N. Y. & Kravtsov, A. V. 2010, ApJ, 714, 287

—. 2011, ApJ, 728, 88

Governato, F., Brook, C., Mayer, L., Brooks, A., Rhee, G., Wadsley, J., Jonsson, P., Willman, B., Stinson, G., Quinn, T., & Madau, P. 2010, Nature, 463, 203

Grcevich, J. & Putman, M. E. 2009, ApJ, 696, 385

Guedes, J., Callegari, S., Madau, P., & Mayer, L. 2011, ApJ, 742, 76

Gupta, A., Mathur, S., Krongold, Y., Nicastro, F., & Galeazzi, M. 2012, ApJL, 756, L8

Haffner, L. M., Dettmar, R.-J., Beckman, J. E., Wood, K., Slavin, J. D., Giammanco, C., Madsen, G. J., Zurita, A., & Reynolds, R. J. 2009, Reviews of Modern Physics, 81, 969

Haiman, Z., Abel, T., & Rees, M. J. 2000, ApJ, 534, 11

Heavens, A. & Peacock, J. 1988, MNRAS, 232, 339

Heiderman, A., Evans, II, N. J., Allen, L. E., Huard, T., & Heyer, M. 2010, ApJ, 723, 1019

Hennebelle, P. & Chabrier, G. 2008, ApJ, 684, 395

Hitschfeld, M., Kramer, C., Schuster, K. F., Garcia-Burillo, S., & Stutzki, J. 2009, A&Ap, 495, 795

Hopkins, P. F. 2012a, MNRAS, 423, 2016
—. 2012b, MNRAS, 423, 2037
—. 2013, MNRAS, 430, 1653
Hui, L. & Gnedin, N. Y. 1997, MNRAS, 292, 27
Kennicutt, Jr., R. C. 1989, ApJ, 344, 685
—. 1998, ApJ, 498, 541
Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2
Klypin, A., Hoffman, Y., Kravtsov, A. V., & Gottlöber, S. 2003, ApJ, 596, 19
Kravtsov, A. V. 2003, ApJL, 590, L1
—. 2013, ApJL, 764, L31
Krumholz, M. R., Leroy, A. K., & McKee, C. F. 2011, ApJ, 731, 25
Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 693, 216
Krumholz, M. R. & Tan, J. C. 2007, ApJ, 654, 304
Lada, C. J., Lombardi, M., & Alves, J. F. 2010, ApJ, 724, 687
Leroy, A. K., Walter, F., Brinks, E., Bigiel, F., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2782
Lidz, A., Faucher-Giguère, C.-A., Dall'Aglio, A., McQuinn, M., Fechner, C., Zaldarriaga, M., Hernquist, L., & Dutta, S. 2010, ApJ, 718, 199
Magdis, G. E., Daddi, E., Béthermin, M., Sargent, M., Elbaz, D., Pannella, M., Dickinson, M., Dannerbauer, H., da Cunha, E., Walter, F., Rigopoulou, D., Charmandaris, V., Hwang, H. S., & Kartaltepe, J. 2012, ApJ, 760, 6
Maller, A. H. & Bullock, J. S. 2004, MNRAS, 355, 694
Mathis, J. S., Mezger, P. G., & Panagia, N. 1983, A&Ap, 128, 212
McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., & Cen, R. 2001, ApJ, 562, 52
McDonald, P., Seljak, U., Burles, S., Schlegel, D. J., Weinberg, D. H., Cen, R., Shih, D., Schaye, J., Schneider, D. P., Bahcall, N. A., Briggs, J. W., Brinkmann, J., Brunner, R. J., Fukugita, M., Gunn, J. E., Ivezić, Ž., Kent, S., Lupton, R. H., & Vanden Berk, D. E. 2006, ApJS, 163, 80
Mo, H. J., Mao, S., & White, S. D. M. 1998, MNRAS, 295, 319
Munshi, F., Governato, F., Brooks, A. M., Christensen, C., Shen, S., Loebman, S., Moster, B., Quinn, T., & Wadsley, J. 2013, ApJ, 766, 56
Naoz, S. & Barkana, R. 2007, MNRAS, 377, 667
Narayanan, D., Krumholz, M., Ostriker, E. C., & Hernquist, L. 2011, MNRAS, 418, 664
Nelson, D., Vogelsberger, M., Genel, S., Sijacki, D., Kereš, D., Springel, V., & Hernquist, L. 2013, MNRAS, 429, 3353
Padoan, P. & Nordlund, Å. 2002, ApJ, 576, 870
Polyachenko, V. L. & Polyachenko, E. V. 1997, Soviet Journal of Experimental and Theoretical Physics, 85, 417
Puchwein, E., Pfrommer, C., Springel, V., Broderick, A. E., & Chang, P. 2012, MNRAS, 423, 149
Quilis, V. & Moore, B. 2001, ApJL, 555, L95
Rauch, M., Sargent, W. L. W., Barlow, T. A., & Carswell, R. F. 2001, ApJ, 562, 76
Ricotti, M., Gnedin, N. Y., & Shull, J. M. 2000, ApJ, 534, 41
Rudie, G. C., Steidel, C. C., & Pettini, M. 2012, ApJL, 757, L30
Safronov, V. S. 1960, Annales d'Astrophysique, 23, 979
Saul, D. R., Peek, J. E. G., Grcevich, J., Putman, M. E., Douglas, K. A., Korpela, E. J., Stanimirović, S., Heiles, C., Gibson, S. J., Lee, M., Begum, A., Brown, A. R. H., Burkhart, B., Hamden, E. T., Pingel, N. M., & Tonnesen, S. 2012, ApJ, 758, 44
Schaye, J., Theuns, T., Rauch, M., Efstathiou, G., & Sargent, W. L. W. 2000, MNRAS, 318, 817
Schmidt, M. 1959, ApJ, 129, 243
Schruba, A., Leroy, A. K., Walter, F., Sandstrom, K., & Rosolowsky, E. 2010, ApJ, 722, 1699
Semenov, D., Henning, T., Helling, C., Ilgner, M., & Sedlmayr, E. 2003, A&Ap, 410, 611
Shandarin, S. F. & Zeldovich, Y. B. 1989, Reviews of Modern Physics, 61, 185
Shetty, R., Glover, S. C., Dullemond, C. P., & Klessen, R. S. 2011a, MNRAS, 412, 1686

Shetty, R., Glover, S. C., Dullemond, C. P., Ostriker, E. C., Harris, A. I., & Klessen, R. S. 2011b, MNRAS, 415, 3253
Sofue, Y., Tutui, Y., Honma, M., Tomita, A., Takamiya, T., Koda, J., & Takeda, Y. 1999, ApJ, 523, 136
Solomon, P. M., Downes, D., Radford, S. J. E., & Barrett, J. W. 1997, ApJ, 478, 144
Stanimirović, S., Dickey, J. M., Krčo, M., & Brooks, A. M. 2002, ApJ, 576, 773
Stinson, G. S., Bailin, J., Couchman, H., Wadsley, J., Shen, S., Nickerson, S., Brook, C., & Quinn, T. 2010, MNRAS, 408, 812
Tacconi, L. J., Neri, R., Genzel, R., Combes, F., Bolatto, A., Cooper, M. C., Wuyts, S., Bournaud, F., Burkert, A., Comerford, J., Cox, P., Davis, M., Förster Schreiber, N. M., García-Burillo, S., Gracia-Carpio, J., Lutz, D., Naab, T., Newman, S., Omont, A., Saintonge, A., Shapiro Griffin, K., Shapley, A., Sternberg, A., & Weiner, B. 2013, ApJ, 768, 74
Toomre, A. 1964, ApJ, 139, 1217
Trowland, H. E., Lewis, G. F., & Bland-Hawthorn, J. 2013, ApJ, 762, 72
Valenzuela, O., Rhee, G., Klypin, A., Governato, F., Stinson, G., Quinn, T., & Wadsley, J. 2007, ApJ, 657, 773
van de Voort, F., Schaye, J., Booth, C. M., Haas, M. R., & Dalla Vecchia, C. 2011, MNRAS, 414, 2458
Viel, M., Becker, G. D., Bolton, J. S., & Haehnelt, M. G. 2013, ArXiv e-prints
Weiner, B. J. & Williams, T. B. 1996, AJ, 111, 1156
Weingartner, J. C. & Draine, B. T. 2001, ApJ, 548, 296
Wiersma, R. P. C., Schaye, J., & Smith, B. D. 2009, MNRAS, 393, 99
Wolcott-Green, J., Haiman, Z., & Bryan, G. L. 2011, MNRAS, 418, 838
Wong, T. & Blitz, L. 2002, ApJ, 569, 157
Springel, V., 2010, MNRAS, 401, 791
Springel, V., 2005, MNRAS, 364, 1105