Dark Matter

Xingjian-Lv

Contents

1 Cosmology Cheat Sheet

1.1 The Homogeneous Universe

The isotropic and homogeneous part of our Universe (whose energy content is close to the critical density) is characterized by the scale factor $a(t)$ (normalized to $a_0 = 1$);

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\mathrm{d}\mathbf{x}^2 \tag{1}$$

Its evolution is described by the Friedmann equation, which tells us how the expansion of the universe (Hubble parameter H) depends on its contents (energy density $\rho = \rho_{\text{mat}} + \rho_{\text{rad}}$, cosmological constant, Λ, and possibly geometry, k):

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} \tag{2}$$

Note Hubble parameter can be viewed as a

1. a distance-velocity relationship, $v = Hd$, cosmologists
2. a timescale (rate), $\tau = H^{-1}$, on which the universe changes by $\mathcal{O}(1)$, particle physicists
3. a distance scale, $c\tau = cH^{-1}$, roughly how big the Universe is

We need to supplement the Friedmann equation with two additional equations to close the equation. The fluid equation, which tells us how density varies due to the expansion of the Universe, and the equation of state, which relates the pressure to the energy density:

$$\dot{\rho} = -3H(\rho + P) \tag{3}$$

$$P = w\rho \tag{4}$$

Remark The Friedmann equation and the fluid equation can be derived from the Einstein field equation based on the Cosmological Principle, i.e., the Universe is homogeneous and isotropic on large scales. The derivation goes like this: the most general metric tensor consistent with the Principle is the RW metric, and the most general "contents" are homogeneous and isotropic perfect fluid, with an energy–momentum tensor of the form

$$T_{\mu\nu} = (P+\rho)u_\mu u_\nu + P g_{\mu\nu} = (P+\rho)\delta^0_\mu \delta^0_\nu + P g_{\mu\nu}$$

Plug these into the field equation, and we get two equations (refereed to as the first and second Friedmann equations). Then with some maneuver, we finally arrive at the Friedmann equation and the fluid equation. There is also a shortcut: we can get the correct form of the Friedmann equation using purely Newtonian gravity. And we can derive the fluid equation (which is the manifestation of local energy-momentum conservation $\nabla^\mu T_{\mu\nu} = 0$) from the first law of thermal dynamics

Now that we have three differential equations for three unknowns, we are in the position to solve for the evolution of our Universe. To know what w is requires some thermal dynamics:

- *Matter* (Non-relativistic gas)

$$\left.\begin{aligned} P &= nkT = \tfrac{\rho_{\text{density}}}{m} kT = \rho_{\text{d}} \tfrac{\langle v^2 \rangle}{3} \\ E &= mc^2\sqrt{1+\gamma^2} \;\Rightarrow\; \rho \simeq \rho_{\text{d}} c^2 \end{aligned}\right\} \Rightarrow w = \tfrac{P}{\rho} \simeq \tfrac{\langle v^2 \rangle}{3c^2} \ll 1$$

  E.g. N_2 at room temperature: $w = 10^{-12}$

- *Radiation* (Relativistic gas)

$$\left.\begin{aligned} &\rho = \alpha T^4 \text{ (Stephan-Boltzmann law)} \\ &P = \tfrac{1}{3}\alpha T^4 \end{aligned}\right\} \Rightarrow w = \tfrac{P}{\rho} = \tfrac{1}{3}$$

  $\therefore$ Gases have $w \subset [0, 1/3]$

- *Dark Energy*

  By definition, DE is a component that accelerates the expansion rate of the

Universe. From the acceleration

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2}(\rho + 3P) = -\frac{4\pi G}{3c^2}\rho(1+3w) \tag{5}$$

We can see that DE must satisfy $w < -\frac{1}{3}$.

E.g., cosmological constant: $w = -1$

Plug the equation of state into the fluid equation, assuming that the fluid equation holds for any individual component (i.e., no interactions between different components), and we have:

$$\rho_i(a) = \rho_{i,0} a^{-3(1+w)} \qquad P_i(a) = w\rho_i(a) = P_{i,0} a^{-3(1+w)} \tag{6}$$

i.e.

$$\rho_i(a) = \begin{cases} \text{radiation} & \rho_{i,0}\, a^{-4} \\ \text{matter} & \rho_{i,0}\, a^{-3} \\ \text{c.c} & \rho_{i,0} \end{cases} \tag{7}$$

All of which have apparent physical meanings.

As for the evolution of the scale factor itself, it is not easy to analytically solve for eqn(2), which can be rewritten as:

$$\left(\frac{\dot{a}}{a}\right)^2 = H^2 = H_0^2 \left[\Omega_M a^{-3} + \Omega_R a^{-4} + \Omega_\Lambda + \Omega_k a^{-2}\right] \tag{8}$$

Where the density parameter Ω_i is the density relative to the critical density $\epsilon_c = 3H^2/8\pi G$

Luckily for us, we can view the Universe as single component-dominated at any time (and the fact that we live in a flat universe) for most purposes. Therefore, the Friedmann equation reads:

$$\dot{a}^2 = \frac{8\pi G \rho_0}{3c^2} a^{-3(1+w)}$$

Solving for which gives:

$$a(t) = \begin{cases} \text{radiation only} & (t/t_0)^{1/3},\ t_0 = 1/2H_0 \\ \text{matter only} & (t/t_0)^{2/3},\ t_0 = 2/3H_0 \\ \text{c.c only} & \exp(H_0(t-t_0)) \end{cases} \tag{9}$$

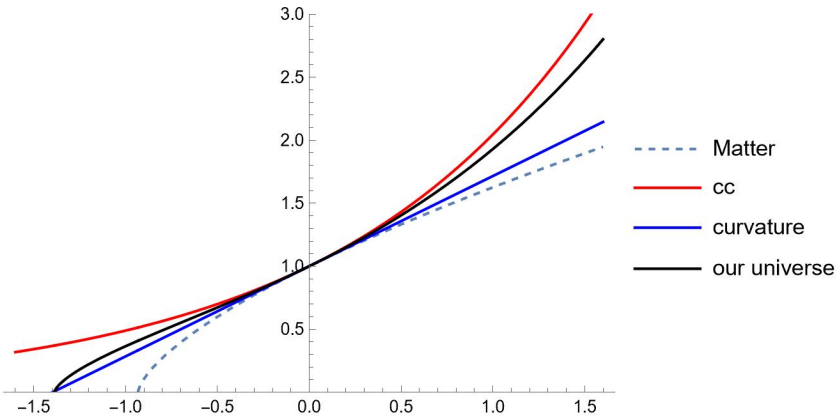


Figure 1: The evolution scale factor for different initial conditions

1.2 Thermal history

The Universe is simple to understand because we believe it started as a hot and dense 'fireball,' which is in an equilibrium state where the abundance of all particle species are determined by a single quantity, the temperature of the Universe.

Fortunately, this simplicity of the primordial Universe did not persist. As the temperature decreased, the *interaction rate* Γ of some particle eventually became smaller than the *expansion rate* H. These particles dropped out of thermal equilibrium and decoupled from the thermal bath. These departures from thermal equilibrium are what make life interesting. Therefore, the key to understanding the thermal history of the Universe (also the production mechanism of dark matter) is the comparison between $H \simeq T^2/M_{\rm Pl}$ (in the early Universe) and $\Gamma = n\sigma v$.

Remark A recurring theme in these sorts of calculations: the equilibrium energy distribution function of photon (the Planck distribution) has a long tail, and there are far more photons than baryons (that is, after matter antimatter annihilation):

$$\eta \equiv \frac{n_{\rm b} - n_{\bar{\rm b}}}{n_\gamma} \approx 5 \times 10^{-9} \tag{10}$$

1.2.1 Some Statistical Mechanics

The basic concept in statistical mechanics is the distribution function, which tells us the probability that a particle chosen randomly has a momentum $\mathbf{p}$. For an equilibrium (maximum entropy), the distribution function is given by either Fermi-Dirac or Bose-Einstein:

$$f(p,T) = \frac{1}{e^{(E(p)-\mu)/T} \pm 1} \tag{11}$$

As we can see, it is characterized by two parameters, the temperature, T, and the chemical potential, μ. The chemical potential may be temperature dependent. Since the temperature changes in an expanding universe, even the equilibrium distribution functions depend implicitly on time.

Chemical potential At early times, the chemical potentials of all particles are much smaller than the temperature and can hence be neglected. Later in the Universe, we need the help of chemical potential to deal with the problem of chemical equilibrium.
Some useful facts:

- Chemical equilibrium is reached when $\sum_i \mu_i = \sum_f \mu_f$
- Photon has *no* chemical potential because photon number is not conserved. This is sometimes written as $\mu_\gamma = 0$.
- In chemical equilibrium, $\mu_X = -\mu_{\bar{X}}$. To see this, consider particle-antiparticle annihilation $X + \bar{X} \leftrightarrow \gamma\gamma$

To relate eqn(11) to macroscopic properties that we really care about, we integrate (sum up):

$$n(T) = \frac{g}{(2\pi)^3} \int \mathrm{d}^3 p f(p,T) \tag{12}$$

$$\rho(T) = \frac{g}{(2\pi)^3} \int \mathrm{d}^3 p f(p,T) E(p) \tag{13}$$

$$P(T) = \frac{g}{(2\pi)^3} \int \mathrm{d}^3 p f(p,T) \frac{p^2}{3E(p)} \tag{14}$$

where $E + \sqrt{m^2 + p^2}$ (ignoring the particles' interaction energy). There is

a $(2\pi)^3$ sitting at the denominator because we are using natural units, $h^3 = (2\pi\hbar)^3 \equiv (2\pi)^3$.

Plug eqn(11) into the above equations and set the chemical potential to 0 gives:

$$\begin{aligned} n &= \frac{g}{2\pi^2}\int_0^\infty \mathrm{d}p \frac{p^2}{\exp\left[\sqrt{p^2+m^2}/T\right]\pm 1} \equiv \frac{g}{2\pi^2}T^3 I_\pm(\frac{m}{T}) \\ \rho &= \frac{g}{2\pi^2}\int_0^\infty \mathrm{d}p \frac{p^2\sqrt{p^2+m^2}}{\exp\left[\sqrt{p^2+m^2}/T\right]\pm 1} \equiv \frac{g}{2\pi^2}T^3 J_\pm(\frac{m}{T}) \end{aligned} \tag{15}$$

Two limits exist. The relativistic limit, $T \gg m$, reads

$$n = \frac{\zeta(3)}{\pi^2} g T^3 \begin{cases} 1 & \text{bosons} \\ \frac{3}{4} & \text{fermions} \end{cases}, \quad \rho = \frac{\pi^2}{30} g T^4 \begin{cases} 1 & \text{bosons} \\ \frac{7}{8} & \text{fermions} \end{cases}. \tag{16}$$

And the non-relativistic limit, $T \ll m$, reads

$$n = g\left(\frac{mT}{2\pi}\right)^{3/2} e^{-m/T}, \quad \rho \approx mn + \frac{3}{2}nTv \approx mn. \tag{17}$$

Comparing the two limits, we can see that n, ρ, and P drop exponentially (Boltzmann suppressed) after T drops below m. This can be interpreted as the annihilation of particles and antiparticles depletes the species, while the reverse process becomes ineffiencient.

The total energy of the early Universe can be written as the sum over all contributions (curvature and DE negligible)

$$\rho = \sum_i \frac{g_i}{2\pi^2} T_i^4 J_\pm\left(x_i\right) = \frac{\pi^2}{30} g_*(T) T^4 \tag{18}$$

where in the second equality, we have defined the *'effective number of DoF'* at temperature T[1] as

$$g_*(T) \equiv \sum_i g_i \left(\frac{T_i}{T}\right)^4 \frac{J_\pm\left(x_i\right)}{J_-(0)} \tag{19}$$

[1]Before about 1s, all species are in equilibrium at the same temperature, $t = T_i$, afterwords the *temperature of the Universe* is typically chosen to be the photon temperature, $T = T_\gamma$

For $T_i \gg m$ (which is automatically away from mass thresholds), it often suffices to include only the relativistic species since $\rho_{\text{rel}} \gg \rho_{\text{NR}}$. So g_* reduces to

$$g_*(T) \equiv \sum_{i=b} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8}\sum_{i=f} g_i \left(\frac{T_i}{T}\right)^4 \tag{20}$$

Entropy

To describe the evolution of the Universe, we would like to track a conserved quantity. Energy is not conserved because the FRW spacetime does not possess a time-like Killing vector. Entropy, on the other hand, is conserved in equilibrium and is often more informative than energy.

Using the first law of thermal dynamics, we can show that the entropy density of the Universe, $s \equiv S/V$, can be written as

$$s = \frac{\rho + P}{T} \tag{21}$$

And the all-important entropy conservation can be written as

$$\frac{d(sa^3)}{dt} = 0 \tag{22}$$

So that entropy is conserved *in equilibrium* (or equivvalently, when the expansion is *adiabatic*) and evolves as $s \propto a^{-3}$.

One can show that, like energy density, the total entropy density can be written as:

$$s = \sum_i \frac{\rho_i + P_i}{T_i} \equiv \frac{2\pi^2}{45} g_{*S}(T) T^3 \tag{23}$$

where the *"effective number of degrees of freedom in entropy"* $g_{*S}(T)$ is defined as[2]

$$g_{*S}(T) \approx \sum_{i=b} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8}\sum_{i=f} g_i \left(\frac{T_i}{T}\right)^3 \tag{24}$$

Useful relation from entropy conservation

►From $\rho_{\text{rad}} = \alpha T^4$ and $\rho_{\text{rad}} \propto a^{-4}$, we have:

$$T \propto g_*^{-1/4} a^{-1} \quad \text{i.e.} \quad aT = \text{const} \tag{25}$$

[2]Like g_*, eqn(23) only holds away from mass thresholds

►Since $s \propto a^{-3}$, the number of particles in a comoving volume is proportional to $N_i \equiv \frac{n_i}{s}$.

►For radiation-dominated era:

$$H = \frac{8\pi G}{3}\rho_{\text{rad}} = \sqrt{\frac{8\pi^3 g_*}{90}}\frac{T^2}{M_{\text{Pl}}} \approx 1.66\sqrt{g_*}\frac{T^2}{M_{\text{Pl}}} \sim \frac{T^2}{M_{\text{Pl}}} \tag{26}$$

Combined with $H = 1/2t$, we have:

$$t = \frac{1}{2}\sqrt{\frac{90}{8\pi^3 g_*}}\frac{M_{\text{Pl}}}{T^2} \approx \frac{0.301}{\sqrt{g_*}\frac{M_{\text{Pl}}}{T^2}} \sim \frac{M_{\text{Pl}}}{T^2} \tag{27}$$

Plug in some numbers. We have:

$$\frac{T}{1\text{MeV}} \simeq 1.5 g_*^{-1/4}\left(\frac{1\text{sec}}{t}\right)^{1/2} \tag{28}$$

It is a helpful rule of thumb that the temperature of the universe 1 second after the Big Bang was about 1 MeV (or 104 K) and evolved as $t^{1/2}$ before that.

1.2.2 First Ten Microseconds

'scale'	Energy	Temperature	Time
Planck ends	10^{19}GeV	10^{32}K	10^{-45}s
GUT ends	10^{16}GeV	10^{29}K	10^{-36}s
Inflation(?)	$10^{15} \sim 10^{9}$GeV	$10^{28} \sim 10^{22}$K	$10^{-36} \sim 10^{-32}$s
EW phase transition	150GeV	10^{15}K	10^{-12}s
QCD phase transition	150MeV	10^{12}K	10^{-5}s

Brief description:

- Quantum Gravity at work
- GUT Higgs acquires none-zero VEV
- Universe expands by $\mathcal{O}(10^{26})$, terminated by the decay of the inflaton field and reheating. It must have been completed before BBN (observationally).
- Electroweak Higgs acquires none-zero VEV. In SM, this is a smooth crossover; in extension to the SM, it can be a first-order phase transition.
- Hadronization of quarks and gluons. In SM, this is a smooth crossover.

1.2.3 Cosmic Neutrino Background

Neutrinos were coupled to the thermal bath only through weak interactions. We therefore expect them to *decouple first* from the thermal plasma. And the various quantities in the interaction rate, Γ, can be approximated by: $v \approx c \equiv 1, \sigma \approx G_{\mathrm{F}}^2 T^2, n \propto T^3$. So

$$\Gamma = n\sigma v \approx G_{\mathrm{F}}^2 T^5 \tag{29}$$

Compare this with the Hubble rate,$H \approx T^2/M_{\mathrm{Pl}}$:

$$\frac{\Gamma}{H} \approx \left(\frac{T}{1\mathrm{MeV}}\right)^3 \tag{30}$$

We conclude that neutrinos decouple around 1 MeV. (A more accurate computation gives a decoupling temperature of 0.8 MeV.) After decoupling, the neutrinos move freely along geodesics and preserve the relativistic Fermi-Dirac distribution (even after becoming non-relativistic later).

Another thing that we need to consider is the 'reheating'[3] of photons by $e^- e^+$ annihilation soon after neutrino decouple (Because the temperature drops below m_e). With the help of entropy conservation ($e^- e^+$ annihilation happens quasi-adiabatically), we can show that

$$T_\nu = \left(\frac{4}{11}\right)^{1/3} T_\gamma \tag{31}$$

1.2.4 Beyond Equilibrium

To understand the world around us, it is crucial to understand deviations from equilibrium. For instance, if a massive particle stays in equilibrium after T drops below m, then its number density is Boltzmann suppressed. For a massive species to survive until the present day, it must drop out of equilibrium before m/T becomes large.

[3]Not the heating of the Universe after inflation.

The main tool to describe the evolution beyond equilibrium is the **Boltzmann equation**:

$$\hat{L}[f] = \hat{C}[f] \tag{32}$$

Refer to Sec 5.1.1 for more details.

1.2.5 Baryogenesis

It could interest us because of the so-called Asymmetric Dark Matter models (in the narrow definition). They are based on the hypothesis that the present-day abundance of dark matter has the same origin as the abundance of ordinary or visible matter: an asymmetry in the number densities of particles and antiparticles. They are largely motivated by the observed similarity in the mass densities of dark and visible matter, with the former observed to be about five times the latter.

1.2.6 Big Bang Nucleosynthesis (TBD)

An excellent probe of baryonic density parameter.
Currently, the most ancient period we can probe.
Lithium problem.

1.2.7 Cosmic Microwave Background (TBD)

The most perfect blackbody ever observed, the best evidence for the hot Big Bang model. First, there is *recombination* happened around $z_{\text{rec}} \approx 1210$, then *photon decoupling* happens about $z_{\text{dec}} \approx 1080$ due to the rapid decrease of ionization rate caused by recombination.

Radiation decoupling does not mean that matter and radiation lose all thermal contact. In fact, the interaction of a few photons with matter keeps the temperatures of matter and radiation equal down to redshifts $z \sim 200$. Only after that does the temperature of baryonic matter begin to decrease faster than that of radiation. There is no trace of this temperature in baryons seen today

because most of them are bound to galaxies where they are ‘reheated’ during gravitational collapse.

Event	time t	redshift z	temperature T
Inflation	10^{-34} s (?)	–	–
Baryogenesis	?	?	?
EW phase transition	20 ps	10^{15}	100 GeV
QCD phase transition	20 μs	10^{12}	150 MeV
Dark matter freeze-out	?	?	?
Neutrino decoupling	1 s	6×10^9	1 MeV
Electron-positron annihilation	6 s	2×10^9	500 keV
Big Bang nucleosynthesis	3 min	4×10^8	100 keV
Matter-radiation equality	60 kyr	3400	0.75 eV
Recombination	260–380 kyr	1100–1400	0.26–0.33 eV
Photon decoupling	380 kyr	1000–1200	0.23–0.28 eV
Reionization	100–400 Myr	11–30	2.6–7.0 meV
Dark energy-matter equality	9 Gyr	0.4	0.33 meV
Present	13.8 Gyr	0	0.24 meV

1.3 The Inhomogeneous Universe

1.3.1 Structure Formation

Derivation There are three competing forces at work when considering structure formation:

- *Gravity* gravitational instabilities cause ”the rich get richer and the poor get poorer,” — Jeans Instability. $t_{\text{dyn}} \sim 1/(4\pi G \bar{\rho}_{\text{density}})^{1/2}$.

- *Pressure* hold against gravity, the reason why Earth does not collapse to a Black Hole. $t_{\text{pre}} \sim R/c_s$, where $c_{\text{s}} = (\mathrm{d}p/\mathrm{d}\epsilon)^{1/2} = \sqrt{w}$ is the speed of sound. Perturbations will grow if $t_{\text{dyn}} < t_{\text{pre}}$ which is the case if R is greater than the Jeans length, $\lambda_J \sim c_{\text{s}} t_{\text{dyn}} \sim c_{\text{s}}/(G\bar{\rho})^{1/2}$.
- H the expansion of the Universe causes overdense regions to become less dense with time

A fully relativistic calculation for growth of density perturbations on sub-horizon scales yields:

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2}\Omega_{\text{m}} H^2 \delta = 0 \tag{33}$$

where δ represents the fluctuation in the density of matter *alone*, $\delta = (\rho_{\text{m}} - \bar{\rho}_{\text{m}})/\bar{\rho}_{\text{m}}$, $\bar{\rho}_{\text{m}}$ is the average matter density.

Now, during radiation dominated era ($\Omega_{\text{m}} \approx 1$ and $a \propto t^{2/3}$): $\delta(t) = B_1 + B_2 \ln t$, while during matter dominated era $\left(\Omega_{\text{m}} \approx 1 \text{ and } a \propto t^{2/3}\right)$: $\delta(t) = D_1 t^{2/3} + D_2 t^{-1}$, where B and D are some constants. Hence, DM perturbations grow as a power law from radiation-matter equality. However, baryons are tightly coupled to photons before decoupling and have $c_s = 1/\sqrt{3}$; hence, baryonic perturbations cannot grow. After decoupling, baryons 'fall into the potential wells' created by DM. As we will see in Sec.2.3.1, these effects lead to evidence for DM matter from amplitudes of temperature anisotropies in CMB.

1.3.2 CMB Anisotropies (TBD)

Several characteristic regions

- The 'Sachs-Wolfe' plateau, low l
  Temperature variations arise from variations from the gravitational potential.
- The acoustic peaks, intermediate l
  from oscillations in photon-baryon fluid due to the competition between gravity and pressure.

- The Silk damping tail, high l
  due to the diffusion of photons during the recombination, fluctuations on small scales are damped.

1.3.3 Polarization and Lensing of CMB photons (TBD)

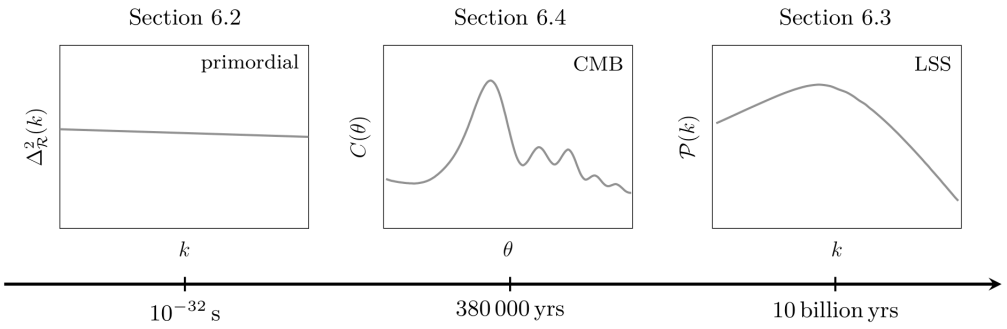


Fig. 6.2 In this chapter, we show how a scale-invariant spectrum of primordial curvature perturbations, $\Delta^2_{\mathcal{R}}(k)$, evolves into the angular power spectrum of the CMB anisotropies, $C(\theta)$, and the matter power spectrum, $\mathcal{P}(k)$. Baumann P212

2 Observational Evidence for DM

ALL the evidence for DM comes from its gravitational effects. That said, the proposal of DM is still the most conservative hypothesis simply for its ability to explain the vast amount of observational evidence. Theories like MOND are simply unable to explain much of the data we have.

2.1 Galaxies

2.1.1 Rotation Curves

First discovered by Rubin et al. in the 1970s. The first definitive evidence for the existence of missing matter.

Note Not all rotation curves are exactly flat.

Derivation Newton told us that:

$$a = \frac{v^2}{R} = \frac{GM(R)}{R^2} \Rightarrow v(R) = \sqrt{\frac{GM(R)}{R}} \tag{34}$$

Therefore we expect:

$$v(R) \propto \begin{cases} R & \text{constant spherical ball} \\ 1/\sqrt{R} & \text{test mass sufficiently far away} \\ \text{Const} & \rho \propto 1/R^2 \end{cases} \tag{35}$$

Rotation curves are flat all the way to the outermost part of typical galaxies, where there is little or no luminous matter, in sharp contrast to the 'Keplerian fall-off' predicted by good old Newtonian gravity[4].

2.1.2 Disk Stability

Self-gravitating disks suffer from 'bar instability' unless they have large velocity dispersion, which disagrees with observation. Embedding disks in a massive, more extended, spherical halo is a solution to this. However, there are other ways around this problem.

2.2 Clusters

Galaxy Clusters: the greatest concentrations (or second greatest, after galaxy filaments, according to wiki) of matter in the Universe, consisting of $100 \sim 1000$ galaxies, hot X-ray emitting gas(intracluster medium, ICM; at very high temperatures, up to $T \sim 10^8$K, and is therefore fully ionized) and dark matter, with typical masses ranging from $10^{14} \sim 10^{15}\mathrm{M}_\odot$ and the typical diameter of $1 \sim 5$Mpc, all moving within the same gravitational potential. The spread of velocities for the individual galaxies is about 800–1000 km/s.

Mass fraction: Stars (Galaxies): less than 2%, X-ray emitting intracluster gas: $\sim 10\%$, rest is dark matter.

E.g. $\mathrm{M}_{\mathrm{vir}}^{\mathrm{Coma}} \approx 2 \times 10^{15}\mathrm{M}_\odot$, $\mathrm{M}_{\mathrm{stars}}^{\mathrm{Coma}} \approx 3 \times 10^{13}\mathrm{M}_\odot$, $\mathrm{M}_{\mathrm{gas}}^{\mathrm{Coma}} \approx 2 \times 10^{14}\mathrm{M}_\odot$

[4] galaxies are typically disk-like so that we can treat it like point mass at its center, but the general trend is the same.

2.2.1 Viral Theorem

Proposed by Zwicky in the 1930s, it is considered to be our first hint of dark matter.

Derivation For a system in equilibrium, the virial theorem applies:

$$-2\langle K\rangle = \langle U\rangle \tag{36}$$

where $\langle K\rangle$ and $\langle U\rangle$ are time-averaged kinetic and potential energy, respectively. For the kinetic energy, we can write:

$$K = \frac{1}{2}\sum_i m_i|\dot{\mathbf{x}}|^2 = \frac{1}{2}M\langle v\rangle^2 \tag{37}$$

where $M = \sum_i m_i$ is the total mass and $\langle v\rangle^2 \equiv 1/M\sum_i m_i|\dot{\mathbf{x}}|^2$ is the mean square velocity (weighted by galaxy mass). Similarly, for the potential energy we have:

$$U = -\sum_{i<j} G\frac{M_iM_j}{|\mathbf{x}_i - \mathbf{x}_j|} = -\alpha\frac{GM^2}{r_{\rm h}} \tag{38}$$

In deriving the second equality, we have used that gravity only depends on the distance between two objects. α is a numerical factor of order unity that depends on the density profile of the cluster, and $r_{\rm h}$ is the half-mass radius.

Combine eqn(36-38), we arrive at:

$$M = \frac{\langle v\rangle r_{\rm h}}{\alpha G} \tag{39}$$

Now let us apply eqn(39) to the Coma cluster. First, let us try to estimate $\langle v\rangle$: we can deduce a galaxy's recession velocity relative to us using Hubble's law. Do this for hundreds of galaxies in the Coma, and we can estimate the one-dimensional velocity dispersion of the galaxies, projected along the line of sight to Earth:

$$\sigma_{\rm r} = \langle(v_{\rm r} - \langle v_{\rm r}\rangle)\rangle^{1/2} = 880\text{km s}^{-1} \tag{40}$$

Assuming isotropic velocity dispersion (a reasonable assumption for a relaxed cluster), then the velocity we want is:

$$\langle v^2\rangle = 3\times(880\text{km s}^{-1})^2 \tag{41}$$

The half mass is trickier. All we can do is measure the half-light radius and proceed assuming that the dark matter and baryons are not segregated within the cluster. In this case, $r_{\rm h} \approx 1.5$Mpc.

Finally, let us plug in the numbers, and we find that:

$$\langle \frac{M}{L} \rangle_{\rm Coma} \approx 400 \frac{M_\odot}{L_\odot} \tag{42}$$

Conclusion There are vast amounts of invisible matter in clusters.

2.2.2 Hydrostatic Equilibrium

Apart from the virial theorem, we can also use the temperature and density of the hot intracluster gas.

Derivation Assuming the gas is spherically symmetric and in hydrostatic equilibrium (i.e., the gas is supported by its own pressure against gravitational infall):

$$\frac{1}{\rho} \frac{{\rm d}P}{{\rm d}r} = -\frac{GM(<r)}{r} \tag{43}$$

Notice that $M(<r)$ is the total(DM + baryons) mass inside a sphere of radius r. P and ρ are pressure and density of the gas, respectively. For ideal gas:

$$P = \frac{\rho k T}{\mu m_{\rm p}} \tag{44}$$

Combine the last two equations; we can solve for the total mass of the cluster:

$$M(r) = \frac{kT(r)r}{G\mu m_{\rm p}} \left[-\frac{d\ln\rho}{d\ln r} - \frac{d\ln T}{d\ln r} \right] \tag{45}$$

Thus, it is possible to deduce the cluster mass by tracking $T,\ \rho$ and the composition of the cluster gas from the cluster core to the outskirts. For the Coma cluster, this method gives:

$$M_{\rm hydro}^{\rm Coma} = (1-2) \times 10^{15} {\rm M}_\odot \tag{46}$$

Conclusion Consistent with what we have found using the virial theorem.

2.2.3 Baryon Fraction from SZ Effect

Sunyaev-Zel'dovich Effect: Electrons in the ICM (which is fully ionized after reionization) can transfer energy to CMB photons via *inverse Compton scattering*. This leaves the isotropy of CMB unchanged, but affects its frequency distribution. Detailed calculations show that the change in intensity can be related to the physical properties of the cluster:

$$\frac{\Delta I_\nu^{\mathrm{RJ}}}{I_\nu^{\mathrm{RJ}}} = -2 \int \frac{kT}{m_{\mathrm{e}} c^2} \sigma_{\mathrm{T}} n_{\mathrm{e}} \mathrm{d}l \tag{47}$$

Where 'RJ' stands for 'Raleigh-Jeans' part of the CMB spectrum, and $\sigma_{\mathrm{T}} = 6.65 \times 10^{-25}\ \mathrm{cm}^2$ is the Thompson scattering cross-section. Thus, if we know the cluster's mass from virial or hydrostatic arguments, we can calculate the baryon fraction f_b of the intracluster gas.

Conclusion Typically, $f_b \approx 10 \sim 12\%$, confirms the conclusion that non-baryonic dark matter is the dominant contributor to the gravitational potential of clusters.

2.2.4 Lensing

Gravitational Lensing
Strong lensing: angular separation between source and lens is small; *Microlensing*: a special type of strong lensing when the angular separation is too small to be separated, gives rise to temporarily brightening of source; *Weak lensing*: occurs when the alignment is not close, *cosmic shear*.

We can deduce the mass distribution of clusters from the strong lensing of background galaxies. In general, the cluster masses deduced by modeling their gravitational lensing effect are in good agreement with the masses found by applying the virial theorem to the motions of the galaxies within the cluster or by applying the equation of hydrostatic equilibrium to the X-ray emitting intracluster gas.

A fascinating case is the *Bullet cluster*, which is two merging clusters. Because of the different nature of baryonic matter and dark matter, we observe a spatial offset of the center of the total mass from the center of the baryonic mass peaks, which is really hard to explain in some MOND models.

2.3 CMB Anisotropies

2.3.1 Amplitude of Perturbations

The measured typical amplitude of the fluctuations, $\Delta T/T \approx 10^{-5}$, alone provides evidence for non-baryonic dark matter.

Derivation Remember that baryonic perturbations cannot grow until photon decoupling. Therefore, in a universe without non-baryonic DM, the initial density perturbations have to be larger and produce larger temperature fluctuations in the CMB, $\Delta T/T \approx 10^{-4}$, for observed structures to form.

2.3.2 Characteristic Angular Scale

The baryon and matter densities affect the oscillations in the photon fluid, and hence the heights of the Doppler peaks. Increasing the baryon density increases the amplitude of the odd peaks, while the height of the third peak is sensitive to the cold dark matter density. From the 2018 Planck temperature, polarization data and lensing data:

$$\begin{aligned} \Omega_{\mathrm{b}} h^2 &= 0.02237 \pm 0.00015 \\ \Omega_{\mathrm{CDM}} h^2 &= 0.1200 \pm 0.012 \end{aligned} \tag{48}$$

Conclusion This precise determination of the baryon density parameter is consistent with the independent, and much higher redshift, measurement from nucleosynthesis. There has to be non-baryonic dark matter so that the Universe is what it looks like today.

2.4 Large Scale Structure

Large-scale structure observations are typically not as powerful or clean a probe of cosmological parameters on their own as the CMB anisotropies are (galaxies are biased tracers of the matter distribution, redshift is a combination of expansion and peculiar velocity, etc.). However, different observables have different degeneracies (combinations of parameters that they are insensitive to), so combining data sets can lead to more precise constraints (provided that they are consistent).

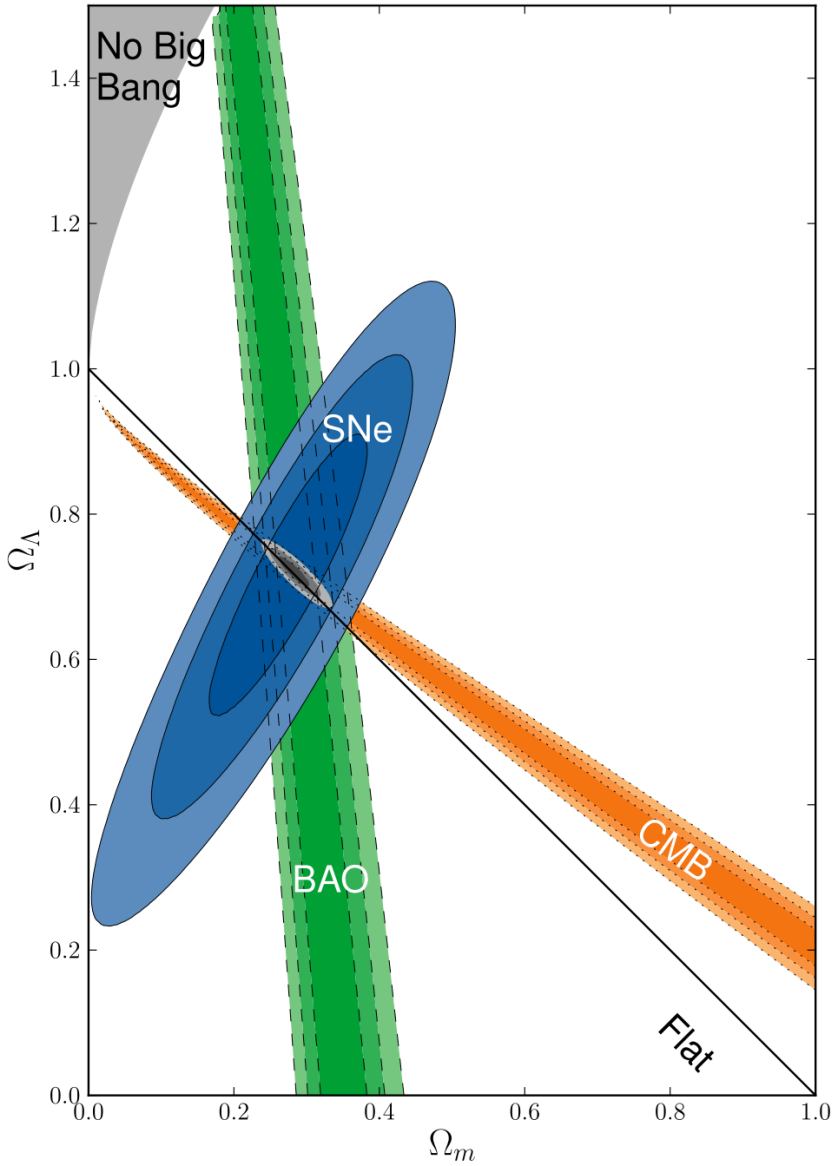


Figure 2: Cosmic concordance

All in all, there is lots of consistent cosmological and astronomical evidence, which tells us that around one-sixth of the total matter in the Universe is unaccounted for, and it is not easy to get away with it by something like MOND.

'dark' Baryons we know that on cosmological scales, baryons just do not work because of constraints from BBN, CMB, and structure formation. And it is hard to imagine that they would pop out somehow *after* CMB.
Notable exceptions: primordial black holes (PBH);
Non-standard baryonic bound states, e.g. Strangelet;
exotica, e.g. Λ di-baryons.

3 DM Density and Velocity Distribution

For indirect searches for DM, we are primarily interested in the spatial distribution of DM (e.g., in galaxies). Occasionally, we are also interested in the velocity distribution of DM (e.g., velocity-dependent thermally averaged cross-section).

3.1 Simulations

There are two types of simulations: N-body simulations (DM-only simulation) and hydrodynamic simulations (take the effects of baryons into account).

Results: Halos in N-body simulations have cuspy NFW/Einasto profiles and contain lots of substructures (with more subhalos at large radii). Hydrodynamic simulations find a) fewer subhalos survive in the inner region; b) density profiles are steepened (but small inner density constant cores can form). c) local velocity distribution is fairly close to Maxwellian.

Challenges: Resolutions of N-body simulations are not high enough (e.g., the effect of tidal stripping on subhalos?). Consequently, we do not have a good handle on the density spectrum of the smallest substructure, which could potentially have a significant impact on the boost factor.

Baryons affect the distribution of DM, especially at the center of galaxies when their density is the highest. There are two competing effects: black hole accretion ('cuspier') and stellar feedback (producing 'cores'). As to which effects take the upper hand, it depends on the size of the halo and how we choose to model baryons.

3.1.1 Small Scale Challenges

The apparent differences between the results of numerical simulations and observations on sub-galactic scales.

- *Cusp-core:*
- *Missing satellites:*
- *'Too-big-to-fall':*

3.2 Density Profile of Halos

The standard halo model (SHM), a.k.a. isothermal sphere[5] is the simplest model of a DM halo, with a density profile $\rho(r) \propto r^{-2}$. It is a solution to a steady-state phase space distribution of collisionless particles that moves under their own gravity.

A scale-invariant DM distribution based on N-body numerical simulation results can be written in a general form as:

$$\rho = \frac{\rho_s}{(r/r_{\rm s})^{\gamma}\left[1+(r/r_{\rm s})^{\alpha}\right]^{(\beta-\gamma)/\alpha}} \tag{49}$$

Where ρ_s and $r_{\rm s}$ are respectively a scale density and a scale radius, whose values are determined by fitting to simulation results. A good choice of parameters is **the NFW profile**, in which $(\alpha, \beta, \gamma) = (1, 3, 1)$:

$$\rho(r) = \frac{\rho_0}{(r/r_{\rm s})\left[1+(r/r_{\rm s})\right]^2} \tag{50}$$

[5]This name arises since the phase-space distribution function has the same form as that of an isothermal self-gravitating sphere of gas.

Where $r_{\rm s}$ is defined as: $(\mathrm{d}\ln\rho/\mathrm{d}\ln r)_{r=r_{\rm s}} = -2$. For $r \ll r_{\rm s}$, $\rho(r) \propto r^{-1}$, while for $r \gg r_{\rm s}$, $\rho(r) \propto r^{-3}$.

A generalization of NFW is the generalized NFW (**gNFW**) profile:

$$\rho(r) = \rho_s(\frac{R_0}{r})^{\gamma}(\frac{R-0+r_s}{r+r_s})^{3-\gamma} \tag{51}$$

where R_0 is the distance between the galactic center and the sun. $\gamma = 1$ corresponds to the standard NFW profile while $\gamma = 0$ is a more conservative cored profile and $\gamma = 1.3$ is a more aggressive cuspy profile.

Another set of parameters that was commonly used in the past is **the Moore profile**, $(\alpha, \beta, \gamma) = (1.5, 3, 1.5)$

$$\rho(r) = \frac{\rho_0}{(r/r_{\rm s})^{1.16}\left[1+(r/r_{\rm s})\right]^{1.84}} \tag{52}$$

Recent high-resolution N-body simulations find density profiles deviate from a pure power law as $r \to 0$. A better fit is the so-called **Einasto profile**:

$$\rho(r) = \rho_{\rm s}\exp\left\{-\frac{2}{\alpha}\left[(r/r_{\rm s})^{\alpha}-1\right]\right\} \tag{53}$$

with shape parameter $\alpha \approx 0.1 - 0.2$. The Einasto profile has a logarithmic slope equal to $2(r/r_{\rm s})^{\alpha}$ i.e. it decreases smoothly as r decreases.

Observations of galactic rotation curves often find a cored profile. Some profiles are proposed in light of this, e.g., **the Burkert profile** and **the truncated Isothermal profile**:

$$\rho_{\rm Bur}(r) = \frac{\rho_s}{(1+r/r_s)\left(1+(r/r_s)^2\right)} \tag{54}$$

$$\rho_{\rm Iso}(r) = \frac{\rho_s}{1+(r/r_s)^2} \tag{55}$$

Technicalities: *virial radius* $r_{\rm vir}$, the radius within which the density is Δ times the background density $\bar{\rho}$ (critical density or matter density), where Δ is the *virial overdensity* (different convention is used, e.g. $\Delta = 200$).The *virial mass* is then the mass within this radius $M_{\rm vir} = 4\pi r_{\rm vir}^3 \Delta \bar{\rho}/3$. A more practical way to parameterize halo size is to use the maximum circular speed: $v_{\rm max} = \sqrt{GM(<r)/r}\Big|_{\rm max}$.
Another quantity that is sometimes useful when describing DM halos is the *concentration*, $c \equiv r_s/r_{\rm vir}$

3.3 Subhalos

Subhalos in DM only simulations have a power-law mass function $dn/dM \propto M/M_\odot)^{-\alpha}$ with $\alpha = 1.90 \pm 0.03$.

4 Dark Matter Models

4.1 Particle Physicists' Classification

- *Axion like particles*
  The pseudo Nambu–Goldstone boson of some spontaneously broken symmetry (which is already weakly broken e.g. by some none perturbative effects) at some yet to probe high energy scale. A good idea to go after them — since they emerge from such a general initial context.
  A prototypical example is the QCD axion.
- *Neutralino*
  The lightest SUSY particles. E.g., in MSSM, there are four neutralinos (These four states are composites of the bino and the neutral wino (which are the neutral electroweak gauginos), and the neutral higgsinos) that are fermions and are electrically neutral, the lightest of which is stable in an R-parity conserved scenario of MSSM.

- *Sneutrion*
  Supersymmetric partner of neutrino. The simplest case is already excluded by direct detection.
- *Neutrino*
  The right-hand partner of neutrinos, $\sim$KeV, is a minimalism extension to SM.
- *LTOP*
  Lightest T-odd particle
- *KKDM*
  Kaluza-Klein Dark Matter, extra dimensions
- *Axiono*
  Supersymmetric partner of the axion.
- *Gravitino*
  Supersymmetric partner of the graviton, non-thermal production (e.g., decay of heavier particles)
- *Asymmetric DM*
  A potential solution to the baryon anti-baryon asymmetry. (the narrow definition of asymm DM)
- *Hidden Sector*
  Dark Force models, an entire dark sector with its own gauge interactions, could be linked to the SM sector through some small coupling parameter ϵ.
- *Primordial Black Holes*
  Not totally died as a DM candidate yet. The parameter space left to check is very narrow.

4.2 Cosmologists' Classification

- *Cold Dark Matter*:
- *Warm Dark Matter*:
- *Fuzzy Dark Matter (Wave DM)*: form a galaxy-sized Bose-Einstein condensate
- *Dark Fluid*: attempt to explain dark matter and dark energy in a single

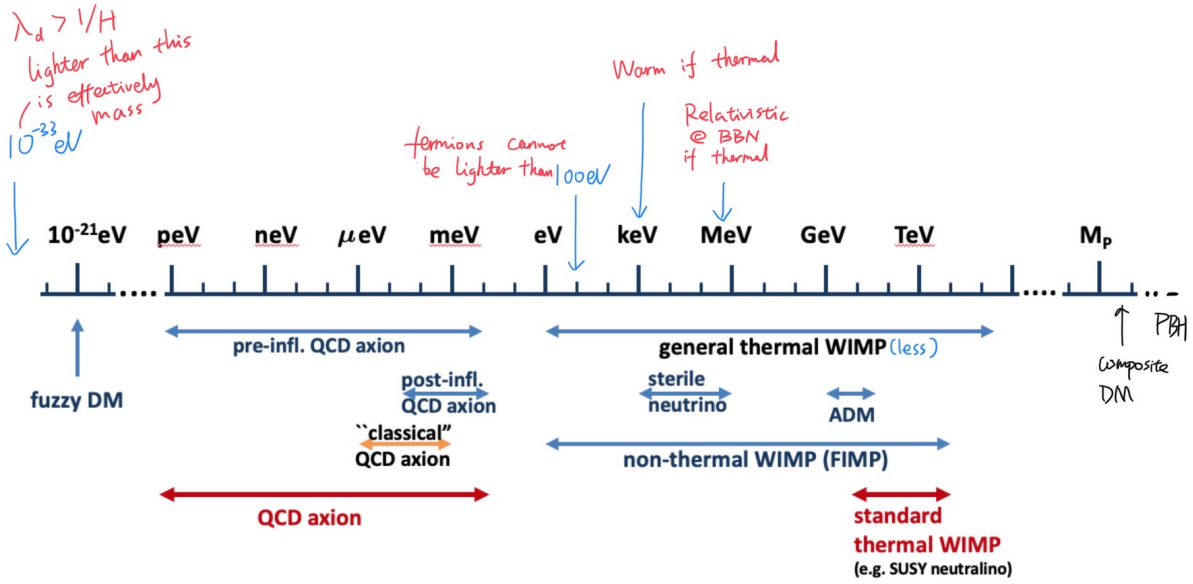
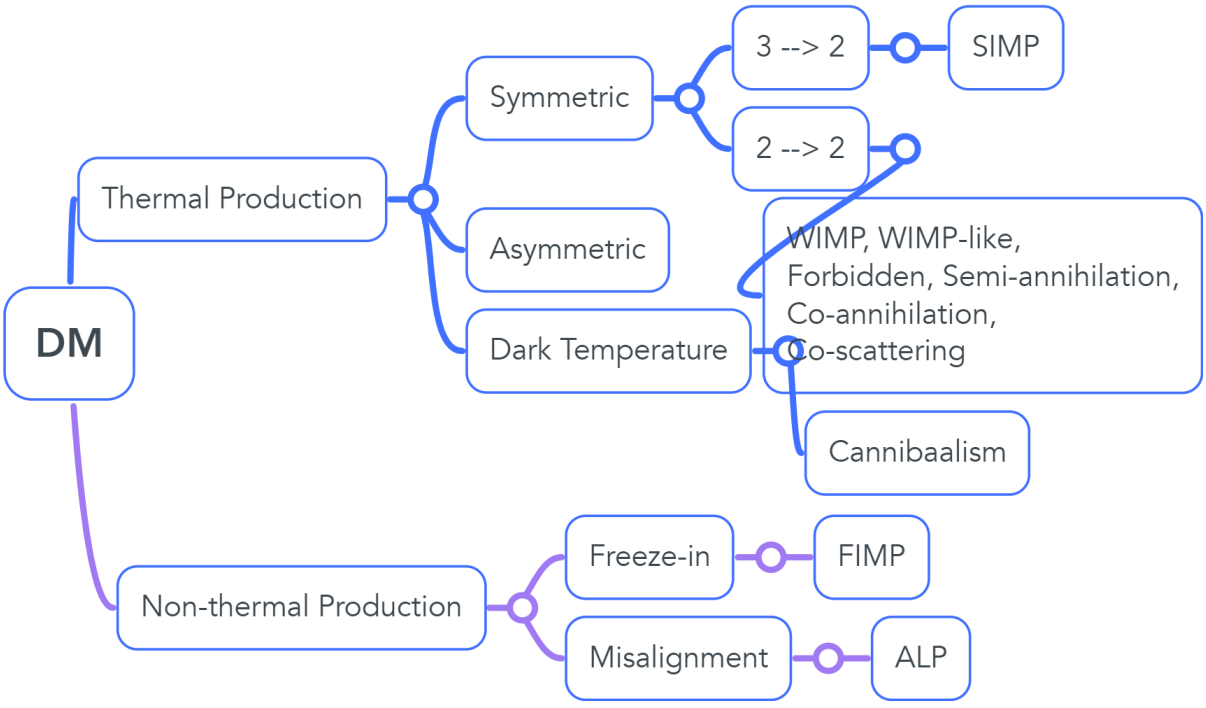


Figure 3: Landscape of DM (incomplete)

framework

- *Self-Interacting Dark Matter*: not collisionless

5 Production Mechanism of DM



5.1 Thermal Relics

The paradigm of *thermal decoupling* is a very successful framework for the origin of species with excellent predictions for recombination and BBN and a detailed description for neutrino decoupling.

A major 'advantage' of the thermal relic is that we do not need to know anything about the *initial* production mechanism of DM since they are considered thermalized (i.e., any initial, high energy information is erased). So the model is highly predictive based on what we know.

5.1.1 Classical WIMPs

In Sec.1.1 we calculated the evolution of the Universe in the absence of interactions between different species. Now we are ready to go beyond the zeroth order approximation and add some interactions. Our starting point is the Boltzmann equation, which lies at the root of non-equilibrium thermal dynamics:

$$\hat{L}[f] = \hat{C}[f] \tag{56}$$

where $f = f(\overrightarrow{p}, \overrightarrow{x}, t)$[6] is the phase space density[7], $\hat{L}$ is the *Liouville operator* describing the change in time of f, and $\hat{C}$ is the *collision operator* describing collisions (interactions). Since we are interested in particle number density, defined by

$$n(t) = \sum_{\text{spin}} \int \frac{\mathrm{d}^3 p}{(2\pi)^3} f(E,t) \tag{57}$$

Thus we take eqn(56) and integrate (for definiteness, let us consider a process of the type $1+2 \to f$, where we are interested in the number density of particle 1, and where we assume that final states are in thermal equilibrium.)

$$\text{LHS} \Rightarrow g_1 \int \hat{L}[f_1] \frac{\mathrm{d}^3 p}{(2\pi)^3} = \frac{\mathrm{d}n_1}{\mathrm{d}t} + 3H \cdot n_1$$

$$\text{RHS} \Rightarrow g_1 \int \hat{C}\left[f_1\right] \frac{\mathrm{d}^3 p}{(2\pi)^3} = -\left\langle \sigma \cdot v_{\text{Møl}} \right\rangle \left(n_1 n_2 - n_1^{\text{eq}} n_2^{\text{eq}}\right)$$

[6] In a homogenous and isotropic universe, $f = f(|\overrightarrow{p}|, t)$, or equivalently, $f = f(E,t)$

[7] Actually f is *one-particle distribution function*

where $\sigma = \sum_f \sigma_{12\to f}$ indicates the invariant, polarized total cross-section, and where the 'Møller velocity' is defined as:

$$v_{\text{Møl}} \equiv \frac{\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}}{E_1 E_2} \tag{58}$$

Note When carrying out the integration, we used the principle of detailed balance: $f_3^{eq} f_4^{eq} = f_1^{eq} f_2^{eq}$ (assuming only one final state $f = 34$) and unitarity.
In the rest frame of 1 (or 2; what we can think of as the "lab frame"), $v_{\text{Møl}} \to v_{\text{rel}} = |\overrightarrow{v}_1 - \overrightarrow{v}_2|$

Lastly, the thermal average is defined by:

$$\langle \sigma v_{\text{Møl}} \rangle = \frac{\int \sigma v_{\text{Møl}} \mathrm{d}n_1^{\text{eq}} \mathrm{d}n_2^{\text{eq}}}{\int \mathrm{d}n_1^{\text{eq}} \mathrm{d}n_2^{\text{eq}}} \tag{59}$$

Equate two sides; we arrive at the familiar expression:

$$\dot{n}_1 + 3Hn_1 = -\langle \sigma v_{\text{Møl}} \rangle (n_1 n_2 - n_1^{\text{eq}} n_2^{\text{eq}}) \tag{60}$$

If particle 1 and 2 are identical (i.e. Majorana particle), then the density of the species $n = n_1 = n_2$ satisfies

$$\dot{n} + 3Hn = -\langle \sigma v_{\text{Møl}} \rangle \left(n^2 - n_{\text{eq}}^2\right) \tag{61}$$

On the other hand, if particle 2 is the antiparticle of particle 1, $n = n_1 + n_2$. And if the species has negligible chemical potential, $n_1 = n_2$. Then we would have an extra $1/2$ on the RHS, contrary to naive expectation. In the following, we will not write out the $1/2$ explicitly.

Now we want to take the thermal average. Let us only consider particle species whose equilibrium distribution is Maxwell Boltzmann distribution[a], $f \propto \exp(-E/T)$.

$$\langle \sigma v_{\text{Møl}} \rangle = \frac{\int \sigma v_{\text{Møl}} \mathrm{d}n_1^{\text{eq}} \mathrm{d}n_2^{\text{eq}}}{\int \mathrm{d}n_1^{\text{eq}} \mathrm{d}n_2^{\text{eq}}} \tag{62}$$

$$= \frac{\int \sigma v_{\text{Møl}} \mathrm{e}^{-E_1/T} \mathrm{e}^{-E_2/T} \, \mathrm{d}^3 p_1 \, \mathrm{d}^3 p_2}{\int \mathrm{e}^{-E_1/T} \mathrm{e}^{-E_2/T} \, \mathrm{d}^3 p_1 \, \mathrm{d}^3 p_2} \tag{63}$$

$$= \frac{1}{8m^4 T K_2^2(m/T)} \int_{4m^2}^{\infty} \sigma \left(s - 4m^2\right) \sqrt{s} K_1(\sqrt{s}/T) \mathrm{d}s \tag{64}$$

[a] the result is applicable to other statistics provided $T \lesssim 3m$

Now, all we have to do is solve eqn(61). For simpleness, we will only consider the limiting case

- When the interaction rate is large, $\Gamma \equiv n \langle \sigma v_{\text{Møl}} \rangle \gg H$, the RHS of eqn(61) dominates, so the natural state of the system is to be in equilibrium.[8], $n \sim n_{\text{eq}}$
- When the interaction rate drops below Hubble rate, $\Gamma \ll H$, the RHS of eqn(61) gets suppressed and the comoving density of the particles approaches a constant relic density, $N_1 \equiv na^3 \to$ const.

The transition point (period) between the two limits is referred to as *"freezeout"*. A precise calculation of the eventual thermal relic density requires numerically solving the evolution equation, eqn(61). But we can get a pretty decent estimation by a simple order of magnitude approximation:
At zeroth order of approximation, we can simply take the time of freezeout to be when

$$\Gamma \sim H \Rightarrow n\langle \sigma v \rangle \sim t^2/M_{\text{Pl}} \qquad \text{assuimg in radiation dominated era} \tag{65}$$

$$\Rightarrow n_{\text{f.o.}} \sim \frac{T_{\text{f.o.}}^2}{M_P \langle \sigma v \rangle} \tag{66}$$

[8] See Baumann P105 for explanation

As always, there are two limiting cases:

$$n_{\text{rel}} \sim T^3 \quad \text{for } m \ll T \tag{67}$$

$$n_{\text{non-rel}} \sim (mT)^{3/2} \exp\left(-\frac{m}{T}\right) \quad \text{for } m \gg T \tag{68}$$

Let us first take the example of neutrino freezeout. We estimate the relevant cross-section in the Fermi four-fermion contact interaction approximation, and we take $E \sim T_\nu$, so that $\sigma \sim G_{\text{F}}^2 T_\nu^2$. Assuming neutrino was still a relativistic species when it decoupled, we have:

$$n\langle\sigma v\rangle = H \quad \rightarrow \quad T_\nu^3 G_F^2 T_\nu^2 = T_\nu^2 / M_P \tag{69}$$

$$\rightarrow T_\nu = \left(G_F^2 M_P\right)^{-1/3} \simeq \left(10^{-10} \times 10^{18}\right)^{-1/3} \text{GeV} \sim 1\text{MeV} \tag{70}$$

Cheerfully, we verify that T_ν indeed satisfies $T_\nu \gg m_\nu$, a fact we have used in assuming for the form of $n(T)$.

Now let us move on to the relic density. First, let us introduce the notation $Y \equiv n/s$, where n is the number density of a given species, and s is the entropy density of the Universe. We introduce such a funny-looking entity because the total entropy of the Universe is conserved[9], $s \cdot a^3 = \text{const}$, therefore, $Y \sim na^3$ is thus a "comoving number density". Assuming no entropy production from decoupling to present day, we have $Y_{\text{today}} = Y_{\text{freeze-out}}$. Therefore,

$$Y_{\text{f.o.}} = Y_0 = \frac{n_0}{s_0} = \frac{\rho_{\nu,0}}{m_\nu s_0} \tag{71}$$

where the third equality holds because $T_{\nu,0} = 1.9\text{K}(1.7 \times 10^{-4}\text{eV}) \ll m_\nu$. So the energy density of neutrino today is

$$\rho_{\nu,0} = m_\nu Y_{\text{f.o.}} s_0 \tag{72}$$

For the SM neutrino, a more careful calculation yields its critical density to be:

$$\Omega_\nu h^2 = \frac{\rho_\nu}{\rho_{\text{crit}}} h^2 \simeq \frac{m_\nu}{91.5\text{eV}} \tag{73}$$

[9] in the absence of violating events such as phase transition.

While the normalization depends on the relevant cross-section, it is a general fact that *a hot relic's relic abundance scales linearly with the relic's mass*. For a weakly interacting dark matter particle, requiring that the thermal dark matter density be less or equal than the observed matter density leads to the so-called *Coswik-McClelland limit*, $m_{\text{hot}} \ll 10eV$, on the mass of a hot dark matter relic. In reality, structure formation provides much better constraints on HDM.

Now let us turn to the more relevant case of cold relics:

$$n\langle\sigma v\rangle = H \rightarrow (m_\chi T)^{3/2} \exp\left(-m_\chi/T\right) \langle\sigma v\rangle = T_\nu^2/M_P \tag{74}$$

$$\rightarrow \frac{m_\chi^3}{x^{3/2}} e^{-x} = \frac{m_\chi^2}{x^2 M_P \langle\sigma v\rangle} \tag{75}$$

$$\rightarrow \sqrt{x} \cdot e^{-x} = \frac{1}{m_\chi M_P \langle\sigma v\rangle} \tag{76}$$

where we have introduced the notation $x \equiv m_\chi/T$.

WIMP Miracle Plug in weak scale numbers into the above equation, i.e. $m_\chi \sim m_{\text{weak}} \sim 100\text{GeV}$,$\langle\sigma v\rangle \sim g_{\text{weak}}^4/m_{\text{weak}}^2$, so we have $\sqrt{x} \cdot e^{-x} \sim 10^{-14}$. Solve numerically for x; we find that $x_{\text{f.o.}} \simeq 20 \rightarrow 25$[10]. Now,

$$\Omega_{\chi,0} = \frac{m_\chi \cdot n_\chi\left(T = T_0\right)}{\rho_c} = \frac{m_\chi T_0^3}{\rho_c} \frac{n_0}{T_0^3} \tag{77}$$

Again, using conservation of entropy, eqn(25), $aT \sim$ const,

$$\frac{n_0}{T_0^3} \simeq \frac{n_{\text{f.o.}}}{T_{\text{f.o.}}^3} \tag{78}$$

We have

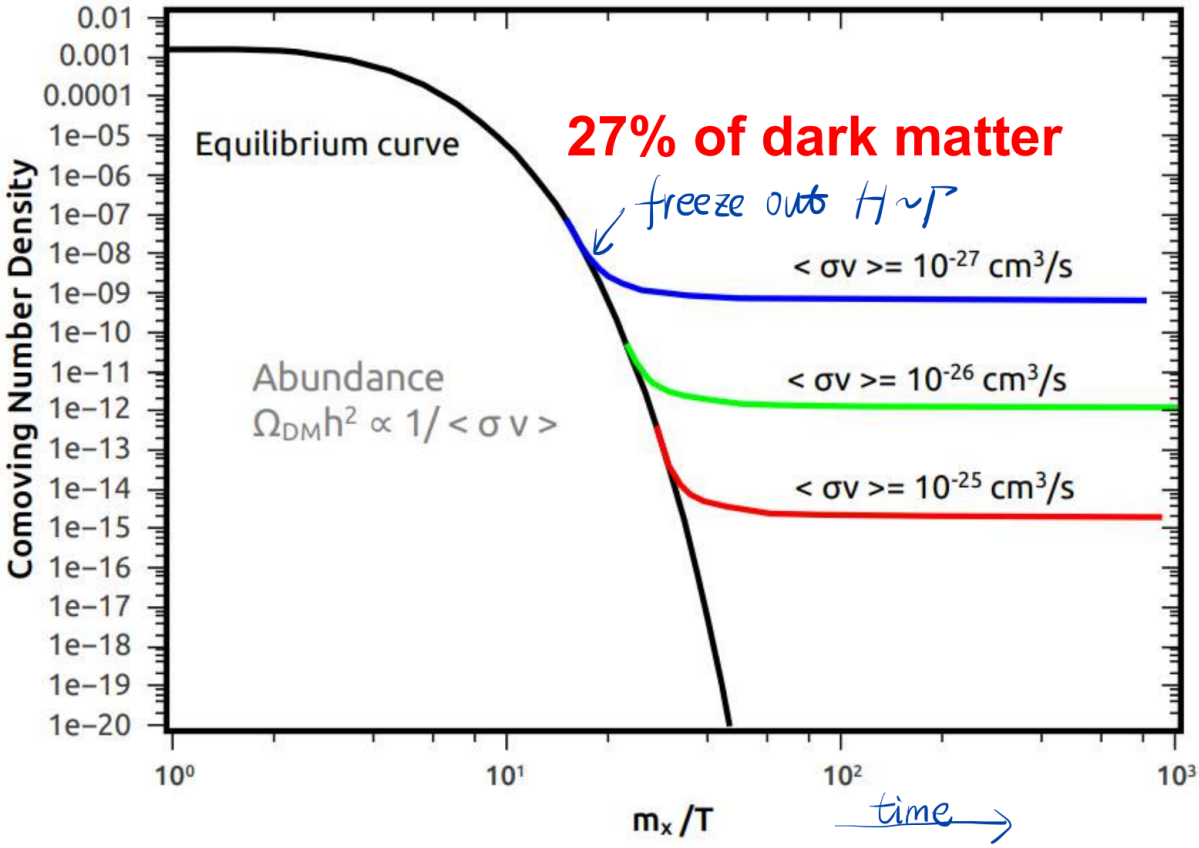
$$\Omega_\chi = \frac{m_\chi T_0^3}{\rho_c} \frac{n_{\text{f.o.}}}{T_{\text{f.o.}}^3} = \frac{T_0^3}{\rho_c} x_{\text{f.o.}} \left(\frac{n_{\text{f.o.}}}{T_{\text{f.o.}}^2}\right) = \left(\frac{T_0^3}{\rho_c M_P}\right) \frac{x_{\text{f.o.}}}{\langle\sigma v\rangle} \tag{79}$$

Now comes the miracle: plug in the measured quantity, e.g., $T_0 = 2.75\text{K} \sim 10^{-4}eV$, we find that (converting back to SI units):

$$\left(\frac{\Omega_\chi}{0.2}\right) \simeq \frac{x_{\text{f.o.}}}{20} \left(\frac{3 \times 10^{-26}\text{cm}^3\text{s}^{-1}}{\langle\sigma v\rangle}\right) \tag{80}$$

But $\langle\sigma v\rangle \sim 3 \times 10^{-26}\text{cm}^3\text{s}^{-1}$ is roughly what we would have found for a weak scale interaction, $\sigma \sim 10^{-8}\text{GeV}^{-2}$, and $v \sim c/3$ (for $x \sim 20$).

[10] Notice that the naive expectation would be a species freeze out shortly after T drops below m, when n becomes exponentially (Boltzmann) suppressed, i.e., However, M_{Pl} is large, and the Universe expands slowly



Schematic representation of freezeout mechanism. Source: 1703.07364

Note Much of the above derivation is in fact not limited to weak scale. Eqn(80) is derived without referring to weak scale. The only place where we have borrowed numbers from weak interactions is $x_{\text{f.o.}} \simeq 20 \to 25$, but notice that $x_{\text{f.o.}}$ only has a logarithmic dependence on the DM mass and annihilation cross-section. Therefore, at the lowest order, we can take $x_{\text{f.o.}}$ to be around 25 for most thermal relics.

Having seen the miracle, let us now turn to the problem of the possible mass range of WIMPs. The upper limit comes from unitarity, roughly:

$$\langle \sigma v \rangle \lesssim \frac{4\pi}{m_\chi^2} \tag{81}$$

which implies

$$\frac{\Omega_\chi}{0.2} \gtrsim 10^{-8}\text{GeV}^{-2} \cdot \frac{m_\chi^2}{4\pi} \tag{82}$$

Therefore, demanding $\Omega_\chi \lesssim 0.2$ implies

$$\left(\frac{m_\chi}{120\text{TeV}}\right)^2 \lesssim 1 \tag{83}$$

or $m_\chi \lesssim 120$TeV, this is the *unitarity bound*[11].

The lower limit, like the Coswik-McClelland limit for HDM, comes from the observed energy density of DM. Consider a particle interacting via the weak interaction, $\langle\sigma v\rangle \sim G_\text{F}^2 m_\chi^2$, in which case

$$\Omega_\chi h^2 \sim 0.1 \frac{10^{-8}}{\text{GeV}^{-2}} \cdot \frac{1}{G_F^2 m_\chi^2} \sim 0.1 \left(\frac{10\text{GeV}}{m_\chi}\right)^2 \tag{84}$$

This implies that $m_\chi \gtrsim 10$GeV for WIMPs, a limit known as the *Lee-Weinberg limit*.

Note Because $x \sim 20$, freeze out is at $T \sim 5$GeV and $t \sim$ ns, not at $T \sim 100$GeV and $t \sim$ ps. Also, what we have been calculating is called *chemical freeze out* (no particle number changing through $\chi\chi \leftrightarrow ff$), which is to be distinct from *kinetic freeze out* (no energy exchange through $f\chi \leftrightarrow f\chi$

To wrap up, the relation between Ω_χ and annihilation strength is wonderfully simple: $\Omega_\chi \propto 1/\langle\sigma v\rangle \sim m_\chi^2/g_\chi^2$. Keeping track of the constants, we find that for $m_\chi \sim 100$GeV, $g_\chi \sim 0.6 \rightarrow \Omega_\chi \sim 0.1$. And the possible mass range for WIMPs is $10\text{GeV} \sim 120\text{TeV}$.

'The Discrete WIMP Miracle' A natural solution to the gauge hierarchy problem is introducing some new particle at the weak scale. However, this would have a direct consequence on precise EW measurements. To avoid such constraints, a straightforward solution is to introduce a discrete parity, so all interactions require pairs of new particles. This also makes the lightest new particle stable, which makes for an excellent DM candidate[12].

[11] The physical picture is clear for this one: as the mass gets larger, the number density of DM gets smaller for a given energy density, so the annihilation cross-section has also to get larger in order not to freeze out too early. At some point, the cross-section gets so large that it violates unitarity.

[12] All heavy elementary particles that we konw of, $m >$ GeV, are unstable on cosmological times, which makes the longevity of DM unusual.

5.1.2 WIMPless Dark Matter

In the above derivation we have used the weak scale as a benchmark scale, $\Omega_\chi \propto 1/\langle\sigma v\rangle \sim m_\chi^2/g_\chi^2$, $g_\chi \sim 1$. But we could also let g_χ vary freely; this can happen if dark matter particles belong to their own hidden sector, with their own gauge coupling strength. By doing this, we opened up a whole new set of dark sector signals in particle physics and cosmology, all with the same WIMP miracle pedigree.

Because the coupling strength is now arbitrary, the lower mass bound is weaker than classic WIMP. Consider a cross-section of the form $\sigma \sim g^4/m_\chi^2$. The condition $\sigma \sim 10^{-8}$GeV then translate to

$$g^2 \sim \frac{m_\chi}{10\text{TeV}} \tag{85}$$

independent of which scale m_χ is at. For it to be a cold relic, we require that $x \gg 1$, which translates to

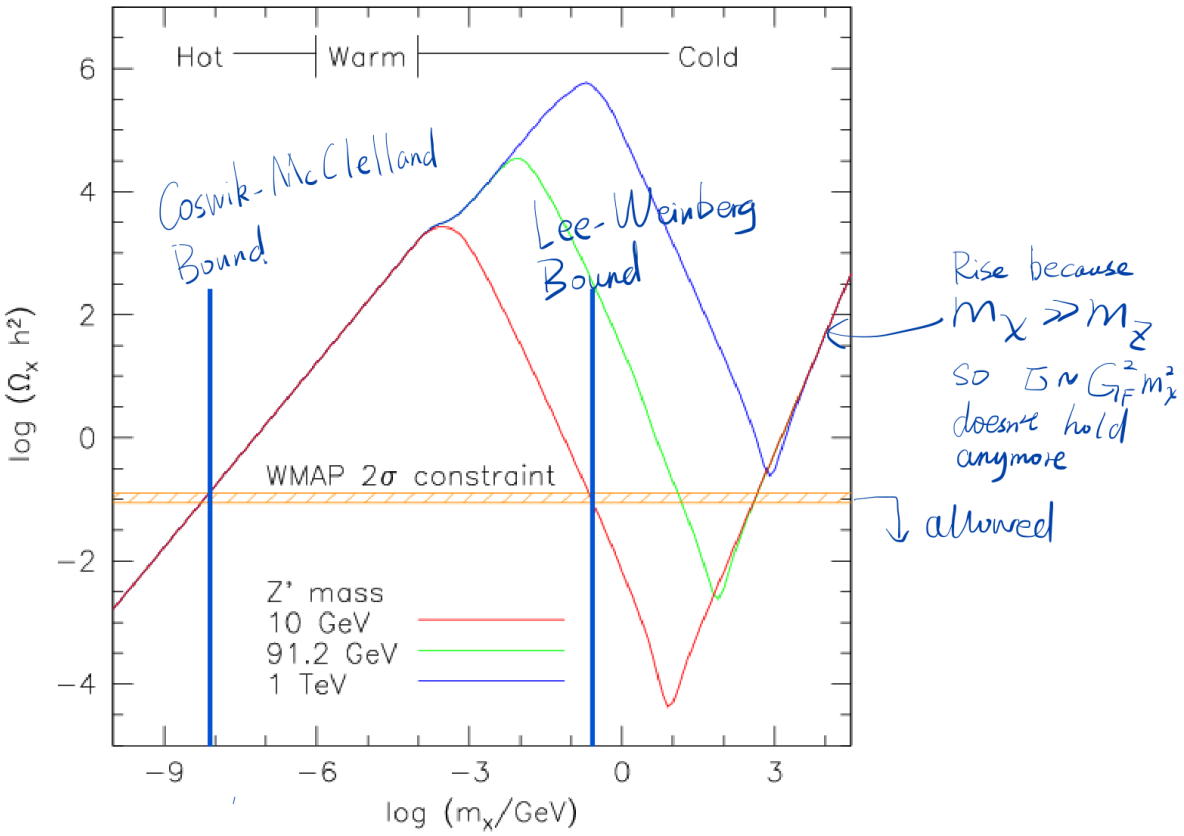
$$m_\chi \cdot M_P \cdot \langle\sigma v\rangle \gg 1 \tag{86}$$

via eqn(76). Combine the above two conditions; we find that $m_\chi \gg 0.1$eV.

the Final Example Below is the thermal relic density of a weakly interacting particle as a function of the particle's mass. the cross-section is assumed to be of the form

$$\sigma \sim \frac{m_\chi^2}{\left(s - m_{Z'}^2\right)^2 + m_{Z'}^4} \tag{87}$$

The mass of the mediator Z' is taken to be 10GeV, 91.2GeV (the Z mass), and 1TeV. The asymptotic hot and cold relic behaviors are clearly visible and match the predictions we made above: a hot relic density scales linearly with mass; a cold relic with $m_\chi \gg m_{Z'}$ has $\Omega \sim 1/\langle\sigma v\rangle \sim m_\chi^2$, and a cold relic in the regime where $m_\chi \ll m_{Z'}$ has $\Omega \sim m_{Z'}^4/m_\chi^2$.



5.1.3 Caveats to the Standard Story

Velocity Dependence

Looking at the form of eqn(64), we can see that it is the convolution of the annihilation cross-section and the temperature. Therefore, the thermally averaged cross-section could have vastly different values at different periods. Typical examples are:

- Low-velocity suppression:

  p-wave suppression Can happen when the s-wave is strongly suppressed (e.g., Majorana fermions annihilating to SM scalars, s-wave contribution involves CP-violation)

  Freeze out happens through multi-body annihilation, involving three or more DM particles in the initial state.

  Channels that are (exponentially) kinematically suppressed at late times.

  Thresholds: the cross-section in eqn(64) suddenly increases as, e.g. $s >$

$4m_t^2$, where m_t is the particle in which pair our annihilating particle can go /etc.

- Low-velocity enhancement

  "Sommerfeld enhancement" can happen when some long-range force carriers are present.

  Resonances: annihilation happens through some intermediate particle $m_a \simeq 2m_\chi$, depending on whether $m_a \lesssim 2m_\chi$ or $m_a \gtrsim 2m_\chi$, could increase/suppress annihilation at late times.

Co-annihilation At early times, the DM may interact with, annihilate against, or be partially comprised of another species that no longer exists in the late Universe. Examples including

- *Asymmetric DM* where DM and anti-DM are distinct particles with different abundances (This is analogous to how the relic abundance of ordinary matter is determined.). Then the $\chi\bar{\chi}$ annihilation cross-section could be in principle much larger than the natural value.

  In the simplest models, the indirect detection signal at late times is tiny, as the abundance of anti-DM is exponentially suppressed. However, if the anti-DM population can be regenerated at some later time, there can potentially be huge indirect-detection signals due to the large annihilation cross-section.
- *Co-annihilation* χ can annihilate with another particle species χ'. The relative abundance of such a state at the temperature of freezeout can roughly be estimated as follows:

$$\begin{aligned} \frac{n_{\chi'}}{n_\chi} &\sim \frac{e^{-m_{\chi'}/T_{\rm F}}}{e^{-m_\chi/T_{\rm F}}} \\ &= e^{-\Delta m_\chi/T_{\rm F}} \\ &\sim e^{-20\Delta} \end{aligned} \tag{88}$$

  where $\Delta \equiv (m_{X'} - m_X)/m_X$ is the mass splitting. For large Δ, χ' will play little roles at freeze out. For smaller splitting ($\Delta \lesssim 0.1$), things are more interesting. To calculate the impact of co-annihilations on the thermal

relic abundance, we introduce the following effective cross-section (with self-explanatory notations):

$$\sigma_{\rm eff}(T) \equiv \sum_{i,j} \sigma_{i,j} \frac{g_i g_j}{g_{\rm eff}^2(T)} \left(1+\Delta_i\right)^{3/2} \left(1+\Delta_j\right)^{3/2} e^{-m_X(\Delta_i+\Delta_j)/T} \tag{89}$$

where $g_{\rm eff}(T) \equiv \sum_i g_i \left(1+\Delta_i\right)^{3/2} e^{-m_X \Delta_i/T}$. As a simple example, consider two nearly degenerate states ($\Delta \sim 0.1$) with equal internal DoF ($g_\chi = g_{\chi'}$). In this case, $\sigma_{\rm eff} \simeq 0.5\sigma_{\chi\chi} + 0.5\sigma_{\chi'\chi'} + \sigma_{\chi\chi'}$. If $\sigma_{\chi\chi'} \gtrsim \sigma_{\chi\chi}$, co-annihilation plays a major role in the deletion of χ abundance. In the opposite case ($\sigma_{\chi\chi'} \lesssim \sigma_{\chi\chi}, \sigma_{\chi'\chi'}$), then χ and χ' each freeze out and contribute to the final dark matter abundance independently.

Entropy Production Suppose at some point after a WIMP has frozen out and is thus decoupled from the Universe's thermal bath, the entropy density changes from $s \to \gamma \cdot s, \gamma > 1$ from, e.g., decaying relics (such as relic gravitinos, moduli. . .) or from a first-order phase transition. Then

$$Y_{\rm today} \to \frac{Y_{\rm today}}{\gamma} \quad \text{and} \quad \Omega_\chi \to \frac{\Omega_\chi}{\gamma} \tag{90}$$

For a sufficiently large γ, the relic abundance of almost any over-abundant WIMP can be "diluted" enough to match the observed dark matter density, i.e., the mass of WIMP can be very light. The additional requirement is, of course, that the entropy injection happens at a temperature smaller than the WIMP freezeout, which sets a (weak) constraint on the WIMP mass: to maintain the successful predictions of light elemental abundances, entropy injection cannot happen too close to the era of BBN.

Alternate Expansion History So far, we have dealt with exceptions to the left-hand side of the Boltzmann equation (or of its shorthand version $\Gamma = n\langle\sigma v\rangle = H$). We can also change the RHS, i.e., fiddle around with the H, the expansion history of the Universe. E.g., *late-time inflation* that occurs after freezeout.

5.2 Freeze-In

The freeze-in scenarios concern Feebly Interacting Dark Matter Particles (FIMPs): these particles have such feeble couplings to the standard particles that they are initially decoupled from the primordial thermal plasma. Freeze-in is a mechanism whereby rare interactions within the SM thermal bath slowly build up an abundance of DM. (In the usual freeze-in story, it is thus assumed that dark sector particles are not produced at an appreciable level through the decay of the inflaton during reheating.)

Unlike freezeout, freeze-in scenarios need a description of the production mechanisms. Therefore, freeze-in scenarios are predictive only when they derive from a high-energy theory that can predict their original density.

5.3 Decay Products

Suppose a particle species ψ, with $m_\psi > m_\chi$ is produced in the early Universe with an abundance Ω_ψ, and that ψ decays to χ, which is the stable dark matter particle, at a temperature when χ is out of equilibrium. The relic density that χ will then inherit (up to contribution from the decay of other particle species and thermal production etc.) is simply

$$\Omega_\chi \simeq \Omega_\psi \frac{m_\chi}{m_\psi} \tag{91}$$

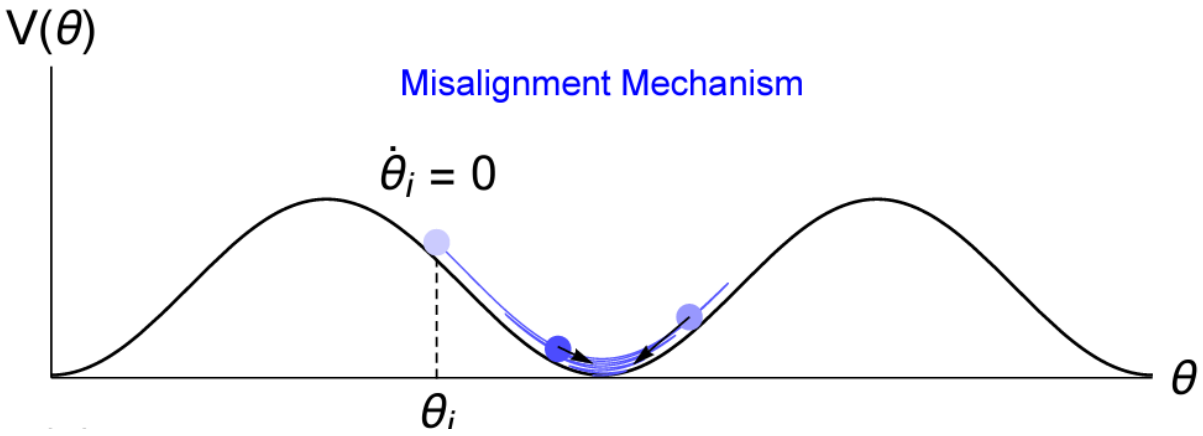
where the $\simeq$ sign indicates that additional effects (such as some entropy production in the decay process) can enter.

5.4 Misalignment

Misalignment is a common mechanism for the production of wave dark matter. In the early Universe, if the initial value of the field, θ_{initial}[13], is away from the

[13]The symbol θ is chosen because we have in our mind some sort of angular field which would be massless if the bottom of the Mexican hat is flat, i.e., the spontaneously broken symmetry is exact.

minima, the axion field starts to oscillate at a temperature T^* when $m_\theta \sim 3H$, where H is the Hubble expansion rate. These oscillations, illustrated in the figure below, can account for the observed abundance of dark matter, with the right choice (tuning?) of the initial value of θ.



Derivation The equation governing the evolution of the axion field in the Universe is

$$\ddot{\theta} + 3H\theta + \frac{\partial V}{\partial \theta} = 0 \tag{92}$$

Assume the initial value of ϕ to be $\mathcal{O}(F)$ (a natural choice for the initial value, too close to the bottom or the top of the potential will be considered fine-tuned). In the early Universe, the Hubble friction (second term on the LHS) dominates, so θ is "freezes" to be its initial value. Until $H \sim m$, θ then rolls down to the bottom and oscillates, leading to the production of DM. Besides, the Hubble friction causes the oscillation amplitude to become smaller, and we can show that the energy density of θ decreases as $\rho \propto 1/a^3$. Therefore, even though θ is very light, it can fill the role of a particle dark matter.

Next, we give an estimation of the abundance of axions. For simplicity, let us assume the potential V takes the form of a quadratic function, $V \sim m^2\theta^2$. Then the total energy stored before rolling down is $\sim m^2F^2$. Denote the scale factor when θ starts to roll down as a_*, then the energy density today is

$$\rho_0 \sim m^2F^2 \left(\frac{a_*}{a_0}\right)^4 \tag{93}$$

For an estimation of a_*, we take $H(a_*) \sim m$, using $H \sim T_*^2 = T_0(a_0/a_*)^2$, we have

$$m^{1/2} \sim \frac{a_0}{a_*} \tag{94}$$

Finally, combine the above two equations, we arrive at

$$\rho_0 \sim m^{1/2} F^2 \tag{95}$$

We can see that ρ_0 depends more strongly on F than on m.

Note The production of QCD axion is slightly more complicated. The Mexican hat does not tilt until the QCD phase transition (i.e., its potential is temperature dependent).

6 DMID

FILL IN THE BLANKS

Dark Matter annihilates/decays in *a place* to *some particles*, which are detected by *an experiment* (at a different location).

$$N_{\text{signal}} = \phi_\chi \cdot A_{\text{eff}} \cdot T_{\text{exp}} \tag{96}$$

where ϕ_χ indicates the relevant dark matter-induced event rate, cm^2s^{-1}, A_{eff} is an effective area and T_{exp} represents the relevant "exposure time". Our goal is to have some signal events with high enough signal-to-noise ratio, $N_{\text{signal}} > (\#\sigma)\sqrt{N_{\text{bkg}}}$

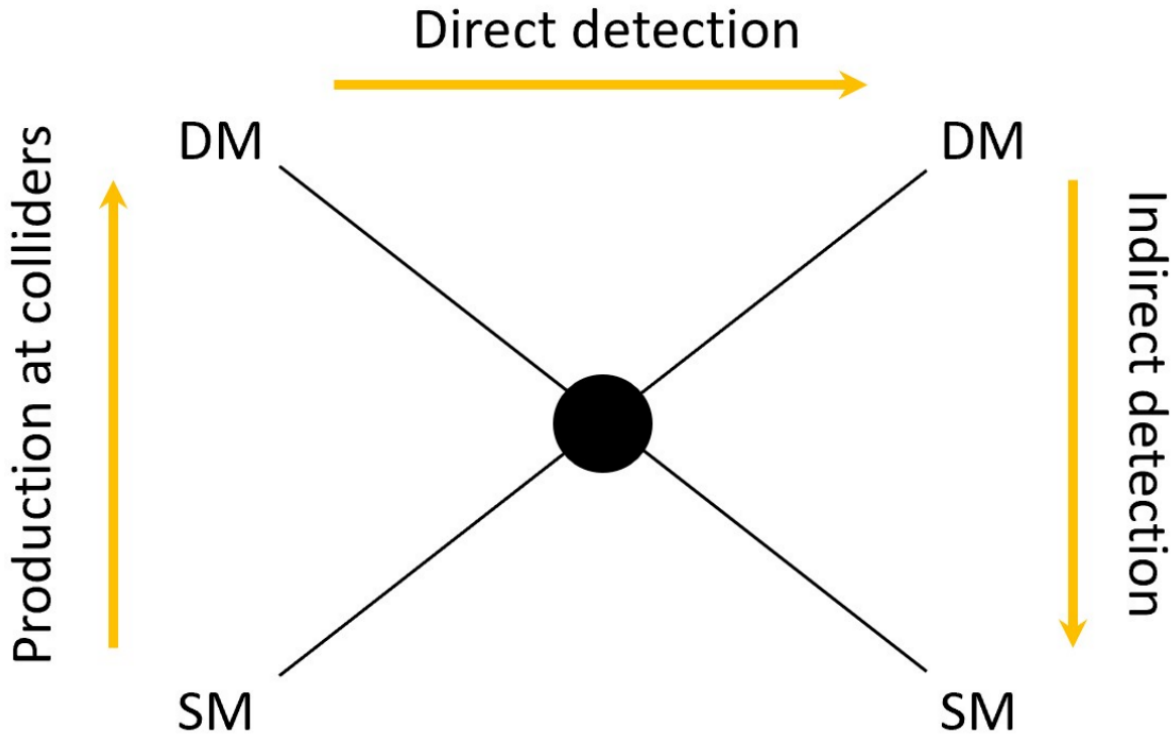


Figure 4: Schematic representation of three different types of DM particle searches and their interplay

Note Generally, we cannot view this as a Feynman diagram that we can rotate to go from one to another. We can easily cook up a scenario where one signal is large while the others are vanishing small.

6.1 Very Indirect Detection

This category includes effects induced by dark matter on astrophysical objects or cosmological observations.

6.1.1 Effects on Astrophysical Objects

A very diverse field, lots of amazing ideas, including but not limited to:

- Solar Physics (dark matter can affect the Sun's core temperature, the sound speed inside the Sun...)

- Neutron Star Capture (accretion), possibly leading to the formation of black holes (notably, e.g., in the context of asymmetric dark matter)
- Supernova and Star cooling (see the excellent book by Georg Rafelt)
- Protostars (e.g., WIMP-fueled population-III stars)
- Planets warming

6.1.2 Effects on Cosmological Observations

Lots of work here, spanning effects on BBN, CMB, reionization, structure formation, and many more.

Structural Information Deviation from 'vanilla' collisionless CDM can affect the DM distribution; we will take two examples:

- *Warm Dark Matter*: simulations show that WDM halos have cuspy profiled like CDM, but there are fewer light halos. Therefore, probing the linear power spectrum/halo mass function could help us distinguish between the two. Methods used to measure the power spectrum include Lyman-alpha forest, Milk Way satellites, gaps in stellar streams (A DM sub-halo passing by a stellar stream gives the stars a kick and perturbs their orbits), strong lensing (looking for flux ratio anomalies), gravitational imaging. The current limits are roughly $m_\chi \gtrsim 5\text{keV}$.
- *Self-Interacting Dark Matter*: simulations show that SIDM halos have constant density cores, but the sub-halo mass function is similar to CDM. Therefore, detailed measurements of density profile (cusp or core, ellipticity) can help a lot. SIDM can also be constrained by observations on merging clusters, including the bullet cluster. The current limits are about $\sigma/m \lesssim 1\text{cm}^2\text{g}^{-1}$.

Energy Injection Even though DM particles do not annihilate/decay at a significant rate after freezeout, the energy they pump into the SM bath can still be appreciable. Let us first consider the case of DM annihilation. We will take a thermal relic DM with s-wave annihilation as a benchmark. Then the fraction

of DM that annihilates in a Hubble time is thus given approximately by:

$$f_{\rm ann} \approx \frac{n^2 \langle \sigma v_{\rm rel} \rangle}{n} H^{-1} \propto a^{-3} \langle \sigma v_{\rm rel} \rangle H^{-1} \tag{97}$$

where we have used the fact that n scales as $n \propto a^{-3}$ after freezeout. Here we are assuming the DM is its own antiparticle, but the scaling relations we derive here do not change if the DM has an antiparticle with equal abundance. For different eras of the Universe, we have

$$\text{Annihilation frac.} \begin{cases} \text{radiation} \quad H \propto a^{-2} \Rightarrow f_{\rm ann} \propto a^{-1} \\ \text{matter} \quad H \propto a^{-3/2} \Rightarrow f_{\rm ann} \propto a^{-3/2} \\ \text{c.c} \quad H = \text{const} \quad \Rightarrow f_{\rm ann} \propto a^{-3} \end{cases} \tag{98}$$

Now it is time for some order of magnitude magic. At freezeout, by definition $n \langle \sigma v_{\rm rel} \rangle \sim H$, i.e. a $\mathcal{O}(1)$ fraction of DM particles annihilates per Hubble time, and so we have $f_{\rm ann} \sim 1$. Freeze out is expected to happen before BBN not to mass up anything, which implies $T_{\rm f.o.} \sim \mathcal{O}(1\text{MeV})$ at the lowest (it may be much higher); matter-radiation equality occurs $\sim$ 1eV, Thus we expect $f_{\rm ann}$ to drop to $< 10^{-6}$ by matter-radiation equality. Today the temperature of the Universe is a few 10^{-4}eV, and most of the remaining expansion (about a factor of 3000 in a) has occurred during matter domination, meaning $f_{\rm ann}$ has decreased by a further factor of about 6×10^{-6}. Thus, we expect the fraction of DM that annihilates per Hubble time at the present day to be $\mathcal{O}(10^{-11})$ *at the very most*, in regions where the DM density takes its cosmological average value. To get a $\mathcal{O}(1)$ change to the local DM abundance from annihilation, we would need a density *11 orders of magnitude higher* than the cosmological density. For comparison, the DM density in the neighborhood of the Earth is roughly a factor of 4×10^5 higher than the cosmological density, so even at this upper bound (which we will see is strongly excluded), we would expect only about one in 10^6 DM particles to annihilate per Hubble time.

Conclusion Only a negligible amount of DM particles have been annihilated since freezeout, and it does nothing to the number density of DM. However, as we will now see, even such a small fraction of annihilating DM can have very marked effects on the history of the cosmos.

If we are interested in the effects on ordinary matter from DM annihilation, it is often helpful to examine the energy liberated by DM annihilation per baryon per Hubble time. At late times there is roughly 5 GeV of energy stored in DM for every baryon (since the mass of a proton/neutron is roughly 1 GeV), and this ratio is fixed, so we can just multiply $f_{\rm ann}$ by 5 GeV to obtain the energy injection per baryon in a Hubble time,

$$\epsilon \sim 5\text{GeV}\, f_{\rm ann} \left(\sim 5\text{GeV} \frac{T_0}{T_{\rm f.o.}} \right) \tag{99}$$

where the last approximate equality holds during the radiation epoch (ignoring changes in the number of relativistic degrees of freedom, assuming temperature scales as 1/a). Now let us see what this amount of energy can do to the evolution of the cosmos. Let us take a 100 GeV thermal relic DM, which freezeout at a temperature around 5 GeV as a benchmark ($x_{\rm f.o.} \sim 20$).

Present Day: Down to redshift 200 ($T \sim 0.1$eV), the CMB temperature is equal to the baryon temperature (i.e., in kinetic equilibrium). Plug in the number, we find that $\epsilon \sim 0.1$eV, so this means DM annihilation is expected to inject enough energy in every Hubble time to change the kinetic energy of all baryons in the Universe by $\mathcal{O}(1)$ amount throughout the radiation-dominated epoch, for 100 GeV thermal relic DM.[14] Lighter DM, which freezes out later, will inject even more energy.

BBN occurs around 1 MeV. Plug in the numbers, we find that 100 GeV thermal relic DM will inject roughly 1 MeV of energy for every baryon in the Universe per Hubble time during the BBN epoch. This amount of energy injection has the potential to affect subdominant nuclear abundances during nucleosynthesis.[15]

Recombination when the CMB radiation is released, the temperature of the photon bath is around 0.2 eV, and the Universe has recently become matter-

[14] In practice, this effect does not literally change the temperature of the baryons by this factor, as the CMB is tightly coupled to the baryons and can act as a heat sink.

[15] Note that this is distinct from the BBN constraint based on the number of relativistic degrees of freedom, which constrains light DM with $m_\chi \leq 1$ MeV

dominated. Our quick estimate above suggests that the annihilation of 100 GeV thermal relic DM should release roughly 0.2 eV per baryon in a Hubble time during the recombination epoch. Since the ionization energy of hydrogen is 13.6 eV, this means such annihilation has the power to ionize roughly 1-2% of the hydrogen in the Universe!

Recombination is characterized by a sharp drop in the ambient ionization level and a corresponding increase in the amount of neutral hydrogen; an increase in the post-recombination ionization level by 1-2% would be very visible, as the extra free electrons would provide a screen to the photons of the CMB. A careful calculation suggests that $f_{\rm ann,CMB} \leq 10^{-11}$. The net effect is that thermal relic DM with a velocity independent σv can be ruled out below mass scales of $\mathcal{O}(10-30)$ GeV[16], depending on the annihilation channel (e.g., if all energy goes into SM neutrinos, then we will not see any signal at all).

Temperature of baryons can also set some limits. However, these limits are usually weaker than the corresponding CMB limits due to the high baseline temperature[17] (Before $z \sim 200$, baryons and CMB photons are coupled together, after $z \sim 2-6$, the temperature of the Universe is expected to increase rapidly due to photoheating from stars).

Given the existing CMB limits, we can see that our previous upper limit of $\sim 10^{-11}$ of DM annihilating today, corresponding to $T_{\rm f.o.} \sim 1$MeV for a standard thermal relic, was far too high; we would have observed an enormous signal in the CMB if that were the case! Taking $T_{\rm f.o.}$ to be 5 GeV instead of 1 MeV, as appropriate for 100 GeV DM, we would need to lower our estimate of the upper bound by a factor of 5000; thus, outside bound structures, in the present day, one would expect only a few in $\sim 10^{-15}$ DM particles to annihilate in a Hubble time. In the neighborhood of the Earth, we might expect a few 10^{-10} DM particles to annihilate per Hubble time.

Conclusion Clearly, DM annihilation is not expected to deplete the DM content

[16] A reminder: $T_{\rm f.o.} = m_\chi/x \sim 0.05 m_\chi$ and $\epsilon \sim 5{\rm GeV} T_0/T_{\rm f.o.}$, so the lower the mass, the lower $T_{\rm f.o.}$, and the higher ϵ, the higher the ionization power.

[17] See Sec 1.2.7

of our Galactic halo anytime soon!

For decaying DM, the calculation is very similar. The fraction of DM decaying in a Hubble time is $f_{\rm dec} = \Gamma H^{-1}$. This quantity scales with redshift just as H^{-1}, so

$$\text{Decay frac.} \begin{cases} \text{radiation} \quad H \propto a^{-2} \Rightarrow f_{\rm dec} \propto a^2 \\ \text{matter} \quad H \propto a^{-3/2} \Rightarrow f_{\rm dec} \propto a^{3/2} \\ \text{c.c} \quad H = \text{const} \Rightarrow f_{\rm dec} \propto \text{const} \end{cases} \tag{100}$$

In contrast to the annihilation scenario, decaying DM signals tend to be dominated by low redshifts / late times. Therefore, looking for late-time signals for decaying DM models will be advantageous.

Consider the estimated limits on heating from Lyman-alpha and (in the future) 21 cm observations: we estimated these would have sensitivity to $f_{\rm dec} \sim 2 \times 10^{-10}$ (Lyman-alpha) and in future $f_{\rm dec} \sim 2 \times 10^{-13}$ (21 cm) during epochs where $H^{-1} \sim 10^{15-16}$s. This suggests Lyman-alpha observations could have sensitivity to $\tau_{\rm dec} \sim \mathcal{O}(10^{25-26})$s, comparable to the CMB bounds.

6.2 Indirect Detection

Now it is time to go beyond simple total energy transfer and talk about *direct observation* of the particles produced by annihilation/decay.

We first need to know the spectra of annihilation/decay products to calculate the observed spectrum. The procedure usually goes as follows

$$\begin{aligned} \text{DM} &\to \text{Rate of SM produced} \to \text{Rate of stable SM produced} \\ &\to \text{Spectra of} e^{\pm}\ \gamma\ \nu\bar{\nu}\ p\bar{p},\ \text{heavier nuclei} \\ &\to \text{What we see at Earth \& how to tell them apart from bkg.} \end{aligned}$$

The calculation of SM final states can be done using public codes such as `Pythia` and `Herwig` or by checking tabulated results from `PPPC4DMID`. We

will only consider two-body final states[18].

Very roughly speaking, there are four broad categories of SM final states:

1. *Hadronic/photon-rich continuum*: $\chi\bar{\chi} \rightarrow (\tau$, gauge bosons, $q\bar{q}) \xrightarrow{decay} \pi^{\pm}\pi^0$, $\pi^0 \rightarrow 2\gamma$ (branching ratio 99%), producing a broad spectrum of photons; $\pi^{\pm} \rightarrow \nu\bar{\nu}\ \mu\bar{\mu}$, then $\mu\bar{\mu} \rightarrow e^{\pm}\ \nu\bar{\nu}$,implying copious production of $e^{\pm}\ \nu\bar{\nu}$ with broad energy spectra. Given sufficient CoM energy, heavier hadronic states can also be produced, including nuclei and anti-nuclei.
2. *Leptonic/photon-poor*:$\chi\bar{\chi} \xrightarrow{mostly} e^{\pm}\ \mu^{\pm}$, γ are produced directly only as part of 3-body final states, by final state radiation or internal bremsstrahlung; photon production rate is suppressed, and the photon spectrum is typically quite hard, peaked toward the DM mass. Copious charged leptons are produced, along with $\nu\bar{\nu}$ in the case of $\mu\bar{\mu}$ final state; these spectra also tend to be harder/more peaked than those produced by hadronic final states.
3. *Photon lines* (dream scenario):$\chi\bar{\chi} \xrightarrow{directly} \gamma\gamma$ or γX, a clear detection of a gamma-ray spectral line would be very difficult to explain with conventional astrophysics.
4. *Neutrinos only* (nightmare scenario): $\chi\bar{\chi} \xrightarrow{solely} \nu\bar{\nu}$, typically quite difficult to observe. At sufficiently high DM masses, production of $\nu\bar{\nu}$ implies that W/Z can be radiated from the interaction point, and the decays of these W/Z give a contribution similar to that from hadronic states as discussed above. In some circumstances, these radiative corrections are more detectable than the primary signal.

Besides the prompt gamma rays discussed above, there are also 'secondary' gamma rays that arise from Inverse Compton Scattering (ICS) of $e^{\pm}$ produced by DM on the ambient light or synchrotron radiation from $e^{\pm}$ propagating in a magnetic field.

More exotic examples include suppressed two-body final state (one needs

[18]One can, of course, cook up some model in which the two-body final state is suppressed, and one needs to take three or more body final states into consideration. But such situations are model dependent, and we will not bother with them in this note.

to take three or more body final states into consideration) and DM first annihilates/decays into some dark sector particle, then these subsequently decay back to SM.

6.2.1 Charged Cosmic Rays

6.2.2 Neutral Particles

γ and ν travel in (roughly) straight lines, so in addition to the energy spectrum, we also need to give a prediction of the spatial distribution on the sky of the signal.

In general, we only have a two-dimensional view of the sky, so what we observe will be the number of γ/ν[19] arriving at our *detector* from within a particular *solid angle* on the sky, within a particular *time interval*. With obvious notations, the signal arising from a volume dV located at (r, θ, ϕ), where the Earth is at $r = 0$. Supposing energy spectrum does not change on its way to Earth (i.e., neglecting redshift, absorption etc.), we have:

$$\frac{dN_\gamma}{dEdtdV} = \left(\frac{dN_\gamma}{dE}\right)_0 \frac{A}{4\pi r^2} \times \begin{cases} \frac{1}{2}\left\langle \sigma v_{\text{rel}} \right\rangle n(\vec{r})^2 & \text{annihilation} \\ \frac{n(\vec{r})}{\tau} & \text{decay} \end{cases} \tag{101}$$

where $(dN_\gamma/dE)_0$ denotes energy spectrum at production of each annihilation/decay. Use $dV = r^2 dr d\Omega$ and integrating along the line of sight (l.o.s), we find

$$\frac{dN_\gamma}{dEdtd\Omega} = \frac{A}{4\pi}\left(\frac{dN_\gamma}{dE}\right)_0 \times \begin{cases} \frac{\langle \sigma v_{rel} \rangle}{2m_{\text{DM}}^2} \int_0^\infty \rho(\vec{r})^2 dr & \text{annihilation} \\ \frac{1}{m_{\text{DM}}\tau} \int_0^\infty dr \rho(\vec{r}) & \text{decay} \end{cases} \tag{102}$$

notice how the two r^2 nicely cancel each other. If we are talking about a localized source, it is often more useful to integrate over $d\Omega$ subtended by the object and obtain the full signal from that source

$$\frac{dN_\gamma}{dEdt} = \frac{A}{4\pi}\left(\frac{dN_\gamma}{dE}\right)_0 \times \begin{cases} \frac{\langle \sigma v_{\text{rel}} \rangle}{2m_{\text{DM}}^2} \int_{\text{los}} \int_{\Delta\Omega} dr d\Omega \rho(\vec{r})^2 & \text{annihilation} \\ \frac{1}{m_{\text{DM}}\tau} \int_{\text{los}} \int_{\Delta\Omega} dr d\Omega \rho(\vec{r}) & \text{decay} \end{cases} \tag{103}$$

[19] Below we will treat γ and ν synonymously.

Now it is time for a useful notation that separates the particle physics part of the puzzle from the astrophysical part. Introducing ***J-factor***[20]:

$$J_{\text{ann}} = \iint dr d\Omega \rho(\vec{r})^2 \qquad J_{\text{dec}} = \iint dr d\Omega \rho(\vec{r}) \tag{104}$$

so that the expected signal can be neatly written as

$$\frac{1}{A}\frac{dN_\gamma}{dEdt} = \frac{1}{4\pi}\frac{\langle \sigma v_{\text{rel}} \rangle}{m_{\text{DM}}^2}\left(\frac{dN_\gamma}{dE}\right)_0 J_{\text{ann}} \quad \text{annihilation} \tag{105}$$

$$\frac{1}{A}\frac{dN_\gamma}{dEdt} = \frac{1}{8\pi}\frac{1}{m_{\text{DM}}\tau}\left(\frac{dN_\gamma}{dE}\right)_0 J_{\text{dec}} \quad \text{decay} \tag{106}$$

Since different sources differ only in their J-factor, it is helpful if we built some initiation for it. Consider the simple example of DM annihilating in a spherical dwarf galaxy of radius Δr, uniform density ρ, and located at a distance $d \gg \Delta r$, then the j-factor is given by

$$J_{\text{ann}} = \rho_\chi^2 \int_{\text{los}} \int_{\Delta\Omega} dr d\Omega \simeq \rho_\chi^2 \cdot \frac{\pi \Delta r}{d^2} \cdot \int_d^{d+\Delta r} dr \simeq \frac{\Delta r^3 \rho_\chi^2}{d^2} \tag{107}$$

From this simple estimation, we see that the most promising targets of gamma-ray searches for DM are those that

1. Have a high density of DM ($J \propto \rho_\chi^2$)
2. Nearby ($J \propto d^{-2}$)
3. Large volume ($J \propto V$)
4. Low or well-understood astrophysical BKG

We can draw a signal-noise diagram for gamma-ray search using the above conditions.

[20] Notice that different authors use different conventions, e.g., whether to include the $4pi$ or put some normalization factors to make J dimensionless.

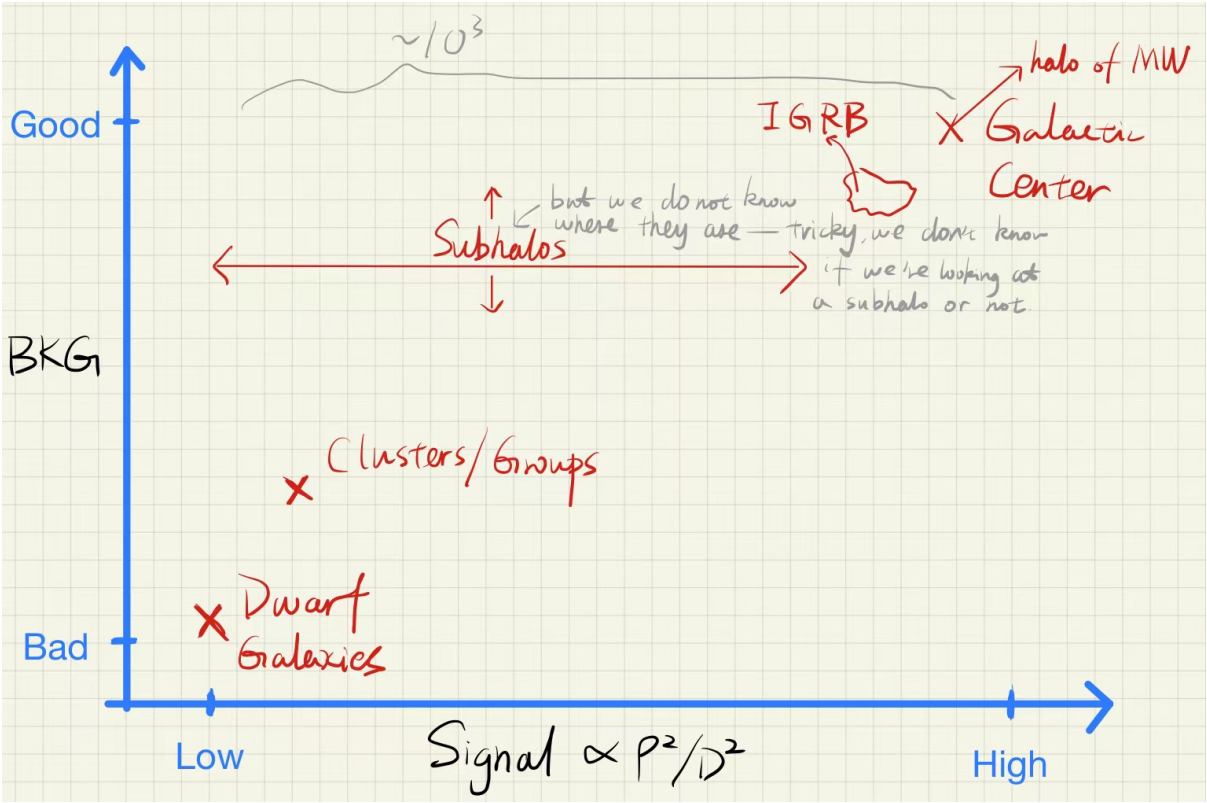


Figure 5: Schematic diagram of gamma-ray searches for DM

For the actual numbers, if we assume an NFW profile, then the dwarf satellite galaxies of the Milky Way have J-factors in the neighborhood $J_{\rm ann} \approx 10^{17-20}{\rm GeV}^2/{\rm cm}^5$; the region within 1 degree of the Milky Way's center has $J_{\rm ann} \approx 10^{22}{\rm GeV}^2/{\rm cm}^5$. Notice that the profile chosen can have significant impact in some cases.

Boost factor The J-factors listed above include only the contribution from the smooth NFW density profile; in reality, the presence of small-scale substructure could potentially greatly increase $J_{\rm ann}$, for $\langle\rho^2\rangle > \langle\rho\rangle^2$. The effects of subhalos grow with the size of the halo and are more prominent for clusters than for galaxies, in the former case, could potentially give rise to an $\mathcal{O}(10^3)$ enhancement to $J_{\rm ann}$, though with large uncertainty. $J_{\rm dec}$, on the other hand, is not affected by substructures because it only depends linearly on $\langle\rho\rangle$, i.e. only sensitive to the total amount of DM energy density, not to its spatial distribution.

A The Galactic and Extragalactic Environments

These environments are extremely *rarified*. Typical benchmark number densities of ordinary matter are

- $\lesssim 1\text{cm}^{-3}$ in the Galaxy,
- $\lesssim 10^{-9}\text{cm}^{-3}$ for CR in the Galaxy,
- $\lesssim 10^{-6}\text{cm}^{-3}$ on average, in a cosmological setting

They also have very *long* time scales and much *bigger* distance scales, compared to terrestrial scales. Referred to Append. F for actual numbers.

The interstellar medium (ISM) of our Galaxy, as well as external ones, is magnetized. Typical benchmark values are $1 - 10$ μG for galactic fields[21]. As for extragalactic medium, there are indications that its magnetic fields exceeds 10^{-19} G at least, and not exceeding ~ 22 nG in the truly extragalactic medium.

As to extragalactic radiation (photon) fields, the important ones are the CMB (with energy density $u_{\text{CMB}} \sim 0.3$ eV cm^{-3} today) and the EBL (extragalactic background light), mostly due to starlight and dust reflection. Inside galaxies, $u_{\text{CMB}} \ll u_{\text{star light}}$, outside, $u_{\text{CMB}} \gg u_{\text{EBL}}$.

B CR Propagation

B.1 Random Motion

Mean free path and *interaction rate*

$$\ell = \frac{1}{\sigma n}, \quad \Gamma = \sigma \beta n = \frac{\beta}{\ell} \tag{108}$$

where β is the particle's velocity.

The average distance moved after N collisions is

$$\langle X \rangle_N = \left\langle \sum \vec{r}_i \right\rangle = 0 \tag{109}$$

[21] As a comparison, the Earth's magnetic field is of the order 1 G

This does not mean that the particle stays still, for that the variance of distance moved after N collisions is not zero:

$$\left\langle X^2 \right\rangle_N = \left\langle \left(\sum_i \vec{r}_i \right) \cdot \left(\sum_j \vec{r}_j \right) \right\rangle = \sum_i \left\langle \vec{r}_i^2 \right\rangle + \ell^2 \sum_{i \neq j} \left\langle \cos \theta_{ij} \right\rangle$$
$$= N \left\langle \vec{r}_1^2 \right\rangle + 0 = \ell^2 N \quad (110)$$

The above equation is the discrete version of a diffusive propagation, with N proportional to the time elapsed via the constant Γ. We can thus guess the continuum limit:

$$\left\langle X^2 \right\rangle (t) = \ell^2 \Gamma t = \ell \beta t, \quad \text{or} \quad X^2 \propto Dt, \text{ where } D \propto l\beta \quad (111)$$

Here we see the prototypical feature of diffusive motion: the variance grows *linearly* with time[22], $X \propto \sqrt{t}$ and that the proportionality factor is called *the diffusion coefficient*.

B.2 Magnetohydrodynamics

Alfvén wave magnetic disturbances parallel to the pre-existing magnetic field in a medium

B.3 Diffusion-loss Equation

Basic Concepts:
motion of charged particle in a constant magnetic field (full relativistic treatment):

$$\frac{\mathrm{d}(m\gamma \mathbf{v})}{\mathrm{d}t} = q\mathbf{v} \times \mathbf{B}_0 \Rightarrow m\gamma \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q\mathbf{v} \times \mathbf{B}_0 \quad (112)$$

since γ is a constant (Lorentz force does no work). Similarly, v, $p = \gamma m v$, *the pitch angle* $\mu \equiv p_z/p$, $p_z = p\mu$, and $p_\perp = \sqrt{1-\mu^2}p$ are also constants. Using some high school physics, we have:

$$\frac{\mathrm{d}v_\perp}{\mathrm{d}t} = \frac{q}{m\gamma} v_\perp B_0 \text{ and } \frac{\mathrm{d}v_\perp}{\mathrm{d}t} = \frac{v_\perp^2}{r} \Rightarrow r = \frac{m\gamma v_\perp}{qB_0} \quad (113)$$

[22] The $\propto$ sign is used because we have been omitting numerical constants depending on the space dimensions.

With respects to the non-relativistic *Larmor radius* or *gyroradius* r_g, *angular gyrofrequency* or *cyclotron frequency* ω_g, and *gyrofrequency* ν_g defined as

$$r_g \equiv \frac{v_\perp}{\omega_g}, \quad \omega_g \equiv \frac{qB_0}{m}, \quad \nu_g = \frac{\omega_g}{2\pi} \equiv \frac{qB_0}{2\pi m} = 2.8\ \mathrm{Hz} Z \left(\frac{m_e}{m}\right)\left(\frac{B_0}{\mu\mathrm{G}}\right) \quad (114)$$

the relativistic generalizations [23] are thus

$$r_L = \gamma r_g = \sqrt{1-\mu^2}\frac{\mathcal{R}}{B_0} \simeq 10^{-6}\sqrt{1-\mu^2}\frac{\mathcal{R}}{\mathrm{GV}}\frac{\mu\mathrm{G}}{B_0}\mathrm{pc} \quad (115)$$

where we introduced *the rigidity* $R \equiv p/q$ (measured typically in GV), and

$$\Omega = \frac{\omega_g}{\gamma} = \frac{qB_0}{E} \simeq 10^{-2} Z \frac{B_0}{\mu G}\frac{\mathrm{GeV}}{E}\mathrm{rad/s} \quad (116)$$

Note that the timescales or equivalently spatial scales of this movement are *very small* for Galactic astrophysics standards.

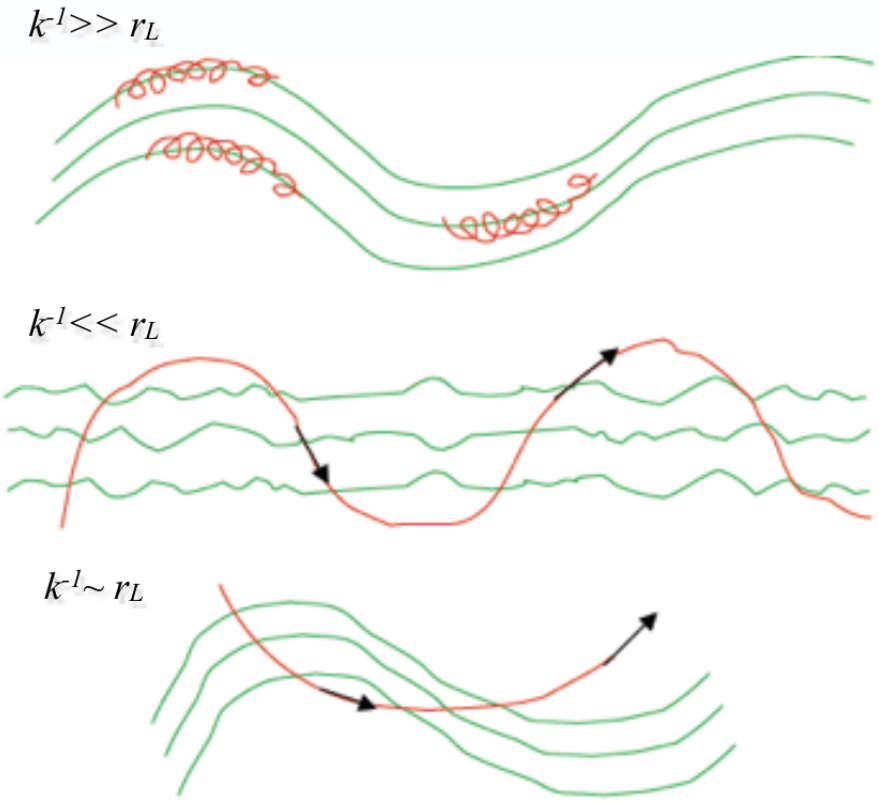
Heuristic derivation for collisionless diffusion:

If we add an ensemble of small-scale, stochastic perturbations to the B-field, orthogonal to its regular value, i.e. $|\delta\mathbf{B}| \ll |\mathbf{B}_0|$ and $\delta\mathbf{B} \perp \mathbf{B}_0$, we will find *diffusion in pitch angle*, μ. For simplicity, let us add $\delta\mathbf{B} = |\delta\mathbf{B}|\{\cos(-kz+\psi), \sin(-kz+\psi), 0\}$, and the result is *pitch angle scattering*

$$\frac{\mathrm{d}\left\langle\Delta\mu^2\right\rangle}{\mathrm{d}t} \simeq \pi C^2 \delta(w) = \pi\left(1-\mu^2\right)\Omega\frac{|\delta\mathbf{B}|^2}{B_0^2}k_{\mathrm{res}}\delta\left(k-k_{\mathrm{res}}\right), \quad k_{\mathrm{res}} \equiv \frac{\Omega}{v\mu} \quad (117)$$

where the average is over the ensemble over phase ψ (assuming all the $\delta\mathbf{B}$s have the same wavelength). Schematically, the result can be viewed as the following:

[23] The conclusion that the gyrofrequency does not depend on velocity only applies to non-relativistic cases.



i.e. a CR "surfs" along field lines whose fluctuations have very long wavelengths, "ignores" fluctuations of the field at too small scales compared with its *Larmor radius*, but undergoes a significant deflection with respect to its unperturbed trajectory if the perturbation matches r_L.

From the analogy with Eqn(111), we introduce the diffusion coefficient of the pitch angle, $D_{\mu\mu}$, according to

$$D_{\mu\mu}(k) \equiv \frac{1}{2}\left\langle \frac{\mathrm{d}\Delta\mu^2}{\mathrm{d}t} \right\rangle_{\psi} \equiv \left(1-\mu^2\right)\nu_{\theta\theta} = \frac{1}{4}\left(1-\mu^2\right)\Omega\frac{1}{B_0^2}\int dx e^{ikx}\delta B^2(x) \tag{118}$$

Which can be written in the more intuitive form:

$$\nu_{\theta\theta}\left(k_{\mathrm{res}}\right) \sim \Omega\left(\frac{\delta B}{B_0}\right)^2\left(k_{\mathrm{res}}\right) \tag{119}$$

i.e. the typical frequency over which the pitch angle μ of CRs with wavenumber k_{res} changes by order one is $\nu_{\theta\theta}$ [24].

[24] We ignore the numerical factors because this simplified approach cannot account for quantitative subtleties of the CR propagation problem.

Conclusion The CR movement is essentially a collisionless (rather than a collisional) diffusive process.Namely, CR do not scatter on other particles, but on inhomogeneities (or more in general, "waves") of the magnetic field.

the CR transport equation (collisionless):
Starting from the Boltzmann equation (essentially conservation of particle number) for the phase space distribution function f (or one-particle distribution function if we are being pedantic)

$$\hat{L}[f] = \hat{C}[f] \tag{120}$$

Since we only care about the average behavior, let us split f into an ensemble-averaged part and a fluctuating part: $f = \langle f \rangle + \delta f$. Averaging fluctuations over $\delta\mathbf{B}$ and/or $\delta\mathbf{E}$ as well, we get for **magnetostatic perturbations**:

$$\partial_t \langle f \rangle + \mathbf{v} \cdot \nabla_{\mathbf{x}} \langle f \rangle - (\mathbf{\Omega} \times \mathbf{p}) \cdot \nabla_{\mathbf{p}} \langle f \rangle = \langle (\delta\mathbf{\Omega} \times \mathbf{p}) \cdot \nabla_{\mathbf{p}} \delta f \rangle \tag{121}$$

where we introduced the gyrofrequency vector of the *ensemble averaged field* $\mathbf{\Omega} \equiv q\langle \mathbf{B} \rangle / E$, as well as the one associated to fluctuations $\delta\mathbf{\Omega} \equiv q\delta\mathbf{B}/E$. We have argued that the effects of the "collisional term"[25] is to relax f to an isotropic distribution function, in the frame of the scattering centers, over a timescale $\nu_{\theta\theta}$. Entering *BKG Ansatz*:

$$\langle i\delta\mathbf{\Omega} \cdot \mathbf{L}\delta f \rangle \simeq -\nu_{\theta\theta} \left(\langle f \rangle - n \right) \tag{122}$$

where the *isotropic mean* is defined[26] as $n \equiv (4\pi)^{-1} \int d\Omega_{\mathbf{p}} \langle f \rangle$. Taylor expand (multipole expansion) $\langle f \rangle = (n + 3\hat{\mathbf{p}} \cdot \mathbf{w})$, with the *flux*[27] given by $\mathbf{w} \equiv (4\pi)^{-1} \int d\Omega_{\mathbf{p}} \hat{\mathbf{p}} \langle f \rangle$ and $\hat{\mathbf{p}} \equiv \mathbf{p}/p$. Plug this into eqn(121) and neglect $\partial_t \mathbf{w}$,

[25] Note that this collisional term is conceptually different from RHS of eqn(120). The former is just a piece from the LHS of eqn(120) pulled to the RHS, acting as an "effective" collisional term (acting on the ensemble-averaged $\langle f \rangle$). While the latter is a "true" collisional term arise from particle-particle interactions. Therefore, eqn(121) is actually not a Boltzmann equation.

[26] $(4\pi)^{-1}$ is introduced because $\int_{\Omega_{\mathbf{p}}} d\Omega_{\mathbf{p}} = 4\pi$

[27] w is related to the usual current by $\mathbf{j} = v\mathbf{w} = p\mathbf{w}/\mathrm{E}$

which is much smaller than $\nu_{\theta\theta}^{-1}$, and we are left with a closed equation (*Fick's law*) for the *angular average of the ensemble average*, n

$$\partial_t n = \partial_i \left(D_{ij} \partial_j n \right) \tag{123}$$

where the *spatial* diffusion tensor is

$$D_{ij} = \frac{1}{3\nu_{\theta\theta}} \left(\frac{p}{E} \right)^2 \frac{\nu_{\theta\theta}^2 \delta_{ij} + \nu_{\theta\theta} \Omega_k \epsilon_{ijk} + \Omega_i \Omega_j}{\nu_{\theta\theta}^2 + \Omega^2} \tag{124}$$

with eigenvalues $1/3\nu_{\theta\theta}$ and $1/3\left(\nu_{\theta\theta} \pm i\Omega\right)$ corresponding to diffusion parallel and perpendicular to the magnetic field, respectively.

Conclusion It is indeed a diffusive process, with the spatial diffusion coefficient inversely proportional to the pitch angle scattering diffusion coefficient. Also, since in the limit of weak turbulence one has $\nu_{\theta\theta} \ll |\Omega|$, this result confirms that the predominant diffusion is parallel to the background field.

The above treatment is severely limited in terms of practical usage for it assumes magnetic fields are static in the Lab (Galactic) frame. In reality, the plasma is moving with respect to this frame, and the ISM, as any other magnetized plasma, can support a number of collective excitation modes. The most famous example is the *Alfvén wave*, propagating along the magnetic field direction with a characteristic speed of $v_A = B_0/\sqrt{4\pi\rho}$, where ρ is the mass density of the medium.

To take the movement of the scattering center into account, we give the magnetic field some velocity $\mathbf{V} \equiv \mathbf{u} + \delta\mathbf{v}$, which gives rise to an associated electric field[28], $\mathbf{E} = \mathbf{B} \times \mathbf{V}$. Now we write down an effective Boltzmann Equation for $\langle f \rangle$ (to avoid cluttering, below this will be referred to simply as f)

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{x}} f + \mathbf{F} \cdot \nabla_{\mathbf{p}} f = \left(\frac{\partial f}{\partial t} \right)_{\text{collision}} \tag{125}$$

[28] As an additional benefit, now we can accelerate particles, which is impossible to do with magnetic field alone.

where $\mathbf{F} = -\boldsymbol{\Omega} \times (\mathbf{p} - E\mathbf{u})$. Proceed as before, we expand f to the first two terms[29]

$$f(\mathbf{x}, \mathbf{p}, t) \simeq n(\mathbf{x}, p, t) + 3\mathbf{w}(\mathbf{x}, p, t) \cdot \hat{\mathbf{p}} \tag{126}$$

where n and $\mathbf{w}$ are defined the same as before. The collisional term is a bit more involved than before. As we have argued, "collisions" drives f towards isotropy:

$$\left(\frac{\partial f}{\partial t}\right)_c \simeq -\nu_{\theta\theta}\,(f - n) = -3\nu_{\theta\theta} w_i(p) \frac{p_i}{p} \tag{127}$$

The only complication is that now it happens in a moving frame. So we boost. And after some further algebraic maneuver and set $\partial_t \mathbf{w} = 0$, we finally arrive at the (collisionless) **transport equation**[30]

$$\frac{\partial n}{\partial t} - \frac{\partial}{\partial x_i} D_{ij} \frac{\partial n}{\partial x_j} + u_i \frac{\partial n}{\partial x_i} - \frac{1}{3} \frac{\partial u_i}{\partial x_i} \left(p \frac{\partial n}{\partial p} \right) = \frac{1}{p^2} \frac{\partial}{\partial p} \left(p^2 D_{pp} \frac{\partial n}{\partial p} \right) \tag{128}$$

where the spatial diffusion tensor is the same as eqn(124), and momentum space coefficient is defined as

$$D_{pp} \equiv \frac{\nu_{\theta\theta} E^2 \left\langle \delta v^2 \right\rangle}{3} \tag{129}$$

Compare with eqn(123), we see three new terms (from left to right):

- *convection/advection* It accounts for spatial transport due to large scale movements like Galactic winds. For typical values of $u = \mathcal{O}(10)$ km s^{-1}, it is relevant at $\mathcal{O}(1)$ GeV, but its relevance decreases at higher energies.
- *adiabatic energy losses/gains* a particularly crucial way to accelerate particles (e.g. in SNR)
- *reacceleration* If we adopt for the spatial diffusion coefficient its parallel value, we arrive at the estimate $D_{pp} D_{zz} = p^2 \langle \delta v^2 \rangle / 9$, where $\langle \delta v^2 \rangle$ is typically set to the square of the Alfvén velocity v_A.

If we use eqn(128) for actual CR propagation, then the diffusion is expected to be strongly anisotropic as long as $\delta B \ll B_0$. However, observed CRs are highly isotropic. This apparent contradiction is resolved once we realized that B_0 is actually also randomized on galactic scales. So we need to take an ensemble

[29] Note that n and $\mathbf{w}$ are functions of only $|\mathbf{p}|$.

[30] This applies for article acceleration as well, very general result.

average over B_0 as well. Practically, the diffusive propagation is often modeled as an *effective* isotropic diffusion coefficient, $D_ij \to D\delta_ij$. A good fit to actual data yields is obtained for $D = 0.3(\mathcal{R}/10\mathrm{GV})^{0.5}\mathrm{kpc}^2/\mathrm{Myr}$

What we have neglect:

- We require that the scattering centers are moving non-relativistically.
- Neglect higher orders (starting from two particle interaction) in the *BBGKY* hierarchy.
- Treat CRs as test particles. The CR themselves contribute to generate the fields in which they propagate. This coupling has been ignored. Including these effects leads to *nonlinear propagation* (See Sec. B.5).

Collisions

We have stressed that the propagation of CRs is essentially a collisionless process. Why so we consider them now? Here are some reasons:

- CRs travel in the Galaxy for a really long time. Even the subleading collisions are important in shaping its spectrum.
- Collisions also generate secondary signals (photons, neutrinos), which is important (at least) from a diagnostic point of view.

Whatever the underlying nature of an interaction, it belongs to one of two categories: continuous energy losses (e.g., Bremsstrahlung) or catastrophic losses (e.g., spallations).

Catastrophic losses

These are treated as "source and sink terms" at the RHS of the propagation equation, since the species "changes nature". Decay is the easiest to deal with, we simply add $-f/\left(\gamma(p)\tau_0\right)$ to the RHS of eqn(128), where the γ factor is added to account for relativistic time dilation. For a species subject to collisional interactions that make it disappear, we replace $/\left(\gamma(p)\tau_0\right) \to \Gamma$, with the interaction rate defined in eqn(108). Finally, for the secondaries that these interactions produced, we add a source term to the RHS. We can symbolically write this "secondary" source term for a species α as $\sum_\beta \Gamma_{\beta\to\alpha} n_\beta$. The true

expression is more complicated and requires a convolution over energies (and the differential cross-sections).

Continuous energy losses

Under this category fall ionization and Coulomb losses (however only important at energies below ∼GeV). In addition, electrons and positrons interact with the ISM emitting bremsstrahlung (again, only important at ∼ few GeV energies), but also synchrotron radiation on the galactic magnetic fields and IC on interstellar radiation fields, which are instead very important at tens of GeV or above.

To account for their effects, we write a continuity equation in E space

$$\frac{\partial \mathcal{N}}{\partial t} + \frac{\partial}{\partial E}(\dot{E}\mathcal{N}) = Q(E) \tag{130}$$

Written in terms of phase space density $n \propto \mathcal{N}(E)/p^2$, it can be shown that the term analogous to $\partial \dot{E}\mathcal{N}/\partial E$ writes

$$-\frac{1}{p^2}\frac{\partial}{\partial p}\left[p^2 \left(\frac{dp}{dt}\right)_{\text{loss}} n\right] \tag{131}$$

Put everything together, we can write the full **diffusion-loss equation**[31] as

$$\frac{\partial n_\alpha}{\partial t} - \frac{\partial}{\partial x_i} D_{ij} \frac{\partial n_\alpha}{\partial x_j} + u_i \frac{\partial n_\alpha}{\partial x_i} - \frac{1}{3}\frac{\partial u_i}{\partial x_i}\left(p\frac{\partial n_\alpha}{\partial p}\right) + \frac{1}{p^2}\frac{\partial}{\partial p}\left[p^2 \left(\frac{dp}{dt}\right)_\ell n_\alpha\right]$$
$$= Q - \Gamma n + \sum_\beta n_\beta \Gamma_{\beta\to\alpha} + \frac{1}{p^2}\frac{\partial}{\partial p}\left(p^2 D_{pp}\frac{\partial n_\alpha}{\partial p}\right) \tag{132}$$

[31] Without knowing anything about astrophysics, we could suspect that acceleration and propagation happen on similar time scales, and thus should be treated simultaneously. But empirical evidence suggests that these two can be factorized to a good approximation. That is why we can factor out a source term Q on the RHS of eqn(132).

Note Number of particles dN in d^3r in $(p, p+dp)$ (independent of the direction of **p**) is:

$$dN = 4\pi p^2 n(\boldsymbol{r}, p, t) d^3r dp \tag{133}$$

People often define the differential number density $\psi(\boldsymbol{r}, p, t)$ as the number of particles in d^3r about r with momentum in the interval $(p, p+dp)$ with

$$dN = \psi(\boldsymbol{r}, p, t) d^3r dp \tag{134}$$

So that

$$\psi(\boldsymbol{r}, p, t) = 4\pi p^2 n(\boldsymbol{r}, p, t) \tag{135}$$

with units of "particles per volume interval per unit momentum".

Written in terms of ψ, the master equation reads

$$\begin{aligned}\frac{\partial \psi(\vec{r}, p, t)}{\partial t} =& q(\vec{r}, p, t) + \vec{\nabla} \cdot \left(D_{xx} \vec{\nabla} \psi - \vec{V} \psi \right) \\ &+ \frac{\partial}{\partial p} p^2 D_{pp} \frac{\partial}{\partial p} \frac{1}{p^2} \psi - \frac{\partial}{\partial p} \left[p\psi - \frac{p}{3} (\vec{\nabla} \cdot \vec{V}) \psi \right] - \frac{1}{\tau_f} \psi - \frac{1}{\tau_r} \psi\end{aligned} \tag{136}$$

Note Detectors do not measure ψ, but rather *differential intensity*, $\mathcal{J}_p$. If particle velocity has magnitude v, the differential intensity is

$$\mathcal{J}_p(\boldsymbol{r}, p, t) = v\psi(\boldsymbol{r}, p, t) \tag{137}$$

i.e. "number of particles / unit momentum / unit surface area / unit time" that goes from ALL directions (i.e. 4π steradian) through the detector.

Another commonly used normalization is

$$\mathcal{J}_p(\boldsymbol{r}, p, t) = \frac{1}{4\pi} v\psi(\boldsymbol{r}, p, t) \tag{138}$$

i.e. "number of particles / unit momentum / unit surface area / steradian / unit time"

In practice we measure particles with respect to *kinetic energy* intervals and not momentum intervals. $E^2 = p^2 + m^2 \Rightarrow dT = dE = (p/E)dp = vdp$.

Since we specify the same particles in two ways, we have that $\mathcal{J}_T dT = \mathcal{J}_p dp$ it follows that $\mathcal{J}_T = \mathcal{J}_p dp/dT = \mathcal{J}_p/v$, so that

$$\mathcal{J}_T = \frac{\psi}{4\pi} = p^2 n \tag{139}$$

IMPORTANT: $\mathcal{J}_T$ measured by experimentalists has simple relationship to quantities ψ and n used by theoreticians!

Lastly, to make comparison with observed data, we need to take the effect of *solar modulation*. The most widely used model is the so called *force-field approximation*

$$\mathcal{N}_{\rm TOA}(E) = \frac{E^2 - m^2}{(E + |Z|e\Phi)^2 - m^2} \mathcal{N}_{\rm LIS}(E + |Z|e\Phi) \tag{140}$$

where TOA stands for "top of the atmosphere" and LIS stands for "local interstellar".

B.4 Analytic Solutions

Notice that all terms in the diffusion-loss equation are of the form n divided by some timescale, e.g., $\tau_{\rm loss} \equiv -E/(dE/dt)$. By comparing different timescales in different scenarios, we can decide which effect(s) is the dominant one and arrive at some approximations.

E-loss dominated propagation
If continuous energy loss timescales are the shortest ones, the steady state equation approximates to

$$-\frac{1}{p^2}\frac{\partial}{\partial p}\left[p^2 \left(\frac{dp}{dt}\right)_\ell n_\alpha\right] = Q \Longrightarrow n(p) \propto -\frac{1}{p^2 (dp/dt)_\ell} \int^p \mathrm{d}p' Q\left(p'\right) p'^2 \tag{141}$$

which, for $Q \propto p^{-s}$ and $(dp/dt)_l \propto -p^l$, leads to

$$n(p) \propto p^{-s-\ell+1} \tag{142}$$

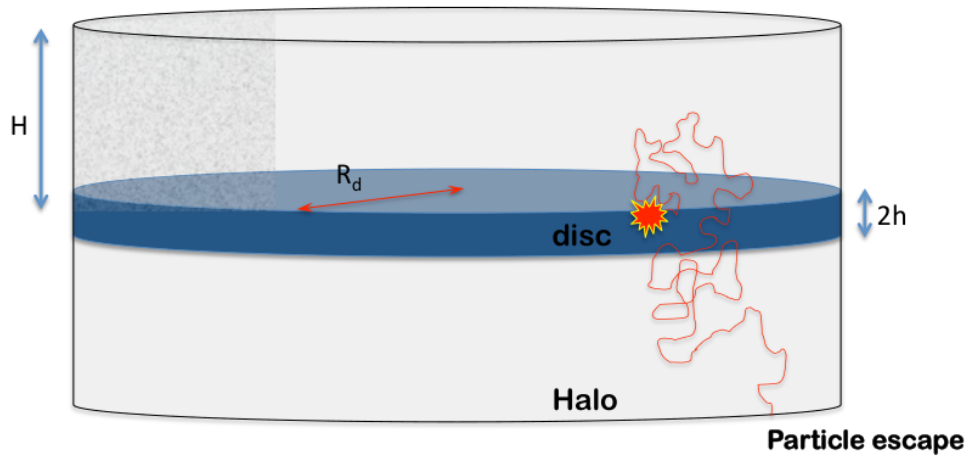
Namely, the resulting spectrum is $l - 1$ softer than the injected one. It turns out that for CR leptons, the above situation is close to truth, with ionization

and Coulomb Energy losses ($l \simeq 0$) dominating at low-energies (spectrum is harder than the source), eventually overcome by bremsstrahlung energy losses ($l \simeq 1$, spectrum matching the source one) and Compton-Synchrotron energy losses ($l \simeq 2$, steeper spectrum).

Diffusion dominated propagation

To a pretty good approximation, we can treat the diffusion as a 1-D problem, in which the Galaxy is considered as a (radially) "infinite" gas thin disk of uniform surface density, sandwiched in a thicker diffusive halo, only the vertical coordinate is relevant:

$$\frac{\partial n}{\partial t} - \frac{\partial}{\partial z}\left(D\frac{\partial n}{\partial z}\right) + u\frac{\partial n}{\partial z} - \frac{1}{3}\frac{\mathrm{d}u}{\mathrm{d}z}p\frac{\partial n}{\partial p} = Q - \Gamma n \tag{143}$$



"Leaky box"

Let us assume that $u = 0$, and sources as well as catastrophic losses are confined to an infinitesimal disk. The *steady state* transport equation simplifies into

$$-\frac{\partial}{\partial z}\left(D\frac{\partial n}{\partial z}\right) = 2q_0(p)h\delta(z) - 2h\Gamma_\sigma n\delta(z) \tag{144}$$

where $\Gamma_\sigma \equiv \sigma v n_{\rm ISM}$. Notice that there are different ways to normalize. For instance, we can also write it as $2h\ (\sigma v n_{\rm ISM})\ n\delta = (\mu\sigma v/m_{\rm CR})\ n\delta$, where μ is a grammage parameter of the disc. $q_0 = N(E)\mathcal{R}/2\pi D_d^2$, where $N(E)$ is the

energy spectrum of the source, $\mathcal{R}$ the rate of supernova explosion[32].

Solve the above equation with respect to boundary condition $n(z \geq H) = 0$, and we find

$$n(z,p) = \frac{Q(H-|z|)}{D(p) + n_{\rm ISM}\sigma v h H} \equiv n_0(p)(1-|z|/H) \tag{145}$$

where we denote quantities on the disk (i.e. $z = 0$) with a subscript 0 and

$$n_0 = q_0(p)\tau_{\rm eff}, \quad \text{where } \tau_{\rm eff}^{-1} \equiv \tau_d^{-1} + \Gamma_\sigma(p) \tag{146}$$

and

$$\tau_d(p) \equiv \frac{Hh}{D(p)} \approx 10^7{\rm yr}\frac{H}{3{\rm kpc}}\frac{h}{100{\rm pc}}\frac{10^{28}\ {\rm cm^2\ s^{-1}}}{D}, \tag{147}$$

$$\tau_\sigma(p) \approx 10^7{\rm yr}\left(\frac{1\ {\rm cm^{-3}}}{n_{\rm ISM}}\right)\left(\frac{100{\rm mb}}{\sigma}\right) \tag{148}$$

Conclusion: the resulting spectrum is the same anywhere in the diffusion zone, with linearly decreasing magnitude towards the boundary. We also note that, as long as $D(p)$ is a growing function of p, the diffusive timescale τ_d dominates over collisional losses at sufficiently high energies (spallation cross-section does not depend on energy that much). This remains true in more general models.

Note τ_d has no physical meaning. It is just a convenient device we made up to make comparisons. This is simple to see: any timescale related to diffusion cannot involve h. The reason h appears in τ_d is due to the way we chose to normalize q_0, which is itself unphysical. A better way to define things is in terms of Q_0. For sake of simplicity, let us ignore spallation, then the solution is of the form $\propto (H/D)Q_0$, which can be rewritten in the more informative form $(H^2/D)Q_0/H$. Now we can associate (physical) diffusion time to H^2/D, and energy injection per unit volume *in the diffusive zone* to Q_0/H.

Secondary over primary

Secondary production just replaces the source term q by the spallation term

[32]Note that q_0, which is defined in $\cdots$ per unit volume, does not have any physical meaning because we assume that sources are confined to an infinitesimal disk. Then the only meaningful quantity is $Q_0 \equiv hq_0$, defined in ... per unit surface area.

$n_p(0)n_{\rm ISM}\sigma v$:

$$n_s(z,p) = \frac{n_p(0)n_{\rm ISM}\sigma_p vh}{D(p) + n_{\rm ISM}\sigma_s vhH}(H - |z|) \tag{149}$$

where σ_p and σ_s are the primary and secondary destruction cross-section. So the secondary/primary ratio in the disk is:

$$n_s(0)/n_p(0) = \frac{n_{\rm ISM}\sigma_p vhH}{D(p) + n_{\rm ISM}\sigma_s vhH} = \frac{\sigma_p v(n_{\rm ISM}h)}{D(p)/H + \sigma_s v(n_{\rm ISM}h)} \tag{150}$$

Notice that only the combination $(hn_{\rm ISM})$ has a will defined physical meaning, which is the surface density of particles in the galactic disk. And H and D are strongly degenerated.

Unstable particles

The master equation in this case is:

$$-\frac{\partial}{\partial z}\left(D\frac{\partial n}{\partial z}\right) = 2q_0(p)h\delta(z) - \frac{n}{\tau_r} \tag{151}$$

Outside the source region $q(z) = 0$:

$$D\frac{d^2n}{dz^2} - \frac{n}{\tau_r} = 0 \tag{152}$$

The general solution is of the form $n = Ae^{-kz} + Be^{+kz}$ where $k = 1/\sqrt{(D_{xx}\tau_r)}$. Plug in the boundary condition, we have:

$$n = Ae^{-kz}(1 - e^{2k(z-H)}) \tag{153}$$

To determine the constant A, we integrate over the disk as usual:

$$\int q(z)dz = \frac{1}{2}D_{xx}\left[\frac{dn}{dz}\right]_-^+ = A\left[\left(-ke^{-kz} - ke^{kz}\right)\right]_{z=0} = -2AkD_{xx} \tag{154}$$

Therefore the solution is:

$$n_p = \frac{Q}{2kD}e^{-kz}(1 - e^{2k(z-H)}) \tag{155}$$

We are interested in e.g. $10Be/9Be$ i.e. unstable/stable secondaries at z = 0.

$$\frac{n_{us}}{n_{ss}} = \frac{\sigma_{us}}{\sigma_{ss}}\frac{(1 - e^{-2H/\sqrt{D\tau_r}})}{2H/\sqrt{D\tau_r}} \tag{156}$$

For $\tau_r \to 0$, i.e. rapid decay, the exponential $\to 0$ and

$$\frac{n_{us}}{n_{ss}} = \frac{\sqrt{D\tau_r}}{2H} \tag{157}$$

This can be interpreted as the diffusion distance $\sqrt{D\tau_r}$ before decay compared with the halo height H, giving the ratio of decaying to non-decaying particles. This shows how the combination of secondary/primary which constrain H/D and unstable/stable secondaries which constrain $\frac{\sqrt{D\tau_r}}{H}$ together allow both H and D to be determined.

Convection Only

The master equation:

$$\nabla \cdot (\vec{V} n) = 2q_0(p)h\delta(z) \tag{158}$$

B.5 Non-linear Aspects

in this section we explore the action of cosmic rays on the environment in which the transport and/or acceleration take place. When cosmic rays propagate outside the acceleration region, such action is mainly in two forms:

1) they generate *hydromagnetic waves*, through streaming instabilities, leading to a dependence of the scattering properties of the medium on the spectrum and spatial distribution of the energetic particles, and

2) they exert a dynamical action on the plasma, which may cause the launching of *cosmic ray driven Galactic winds*.

Self-generatrd Waves

The general idea is as follows: as soon as there are energetic particles streaming at a speed that is larger than the local Alfvén speed, Alfvén waves of appropriate wavelength become unstable and grow in amplitude. This phenomenon is referred to as *streaming instability*.

There are two kinds of streaming instability: the non-resonant instability and the resonate one. For particle transport in the Galaxy, the resonate branch is the relevant one[33]. The growth of resonate instability is induced by fast

[33]The non-resonant instability is important where the CR density and/or velocity are especially high, and hence

streaming particles at wavelengths that are resonant with their Larmor radius $r_{\rm L}$, with perturbation wavenumber that are resonant with $k = 1/r_{\rm L}(p)$. For a back of an envelope estimation, we can write the growth rate of the non-resonant instability as

$$\Gamma_{\rm CR}^{\rm RES}(k) = \frac{\pi^2 e}{B_0}\frac{v_{\rm A}}{c}\left[\frac{D(p)}{H}4\pi p^3 n(p)\right]_{p=p_{\rm res}} \tag{159}$$

where $p_{\rm res}(k) = eB_0/ck$, alone the direction of the background magnetic field B_0.

The magnetic perturbations will effectively grow only if the rate $\Gamma_{\rm CR}^{\rm RES}$ is faster than the damping processes at work. The main damping mechanisms traditionally considered for these waves are non-linear damping (NLD) and ion-neutral damping (IND). There could be places where this condition is satisfied, e.g., the vicinity of the source and high altitude in the Galaxy.

So what does this self-generated wave do to the classic picture of CR propagation? Here we give two possibilities.

1) In all treatments of CR propagation problems, we assume that the scattering is due to irregularities in the galactic magnetic fields. But this ignores the problem of *where do these turbulences come from*. On one hand, one could assume that turbulence is injected, for instance by supernova explosions and eventually cascades towards the small scales relevant for particle scattering. On the other hand, it is difficult to imagine that such turbulence would exist at high altitude above the Galactic disc. CR self-generated waves give a natural origin to these turbulences. And one can show that the growth of waves induced by CR streaming is the dominant effect at lower energies. As a result, the diffusion coefficient will change its energy dependence around some threshold energy.

2) Propagation in the vicinity of sources. Near the source, both the CR density and gradient are large enough that one can expect non-linear effects to dominate transport. Recent studies show that *accelerated particles constitute a primary source of turbulence in the vicinity of their accelerators*. In some

within CR sources or in their vicinity.

case the turbulences generated by CRs are so powerful that it results in longer confinement time and higher grammage accumulated near the source. And the latter can actually be *nonnegligible* with respect to the overall grammage. This could have interesting implications on the interpretation of secondary-to-primary ratios (e.g., B/C) in terms of grammage traversed in the Galaxy.

CR Induced Galactic Wind

The launching of a CR induced wind results from the interplay between CRs and gravity. A wind is launched only if the gas can be accelerated to supersonic speeds [34] and its properties connected smoothly with the boundary conditions at infinity.

Whether this outflowing Galactic wind is "external" or self-generated, is still undermined. But at least in the near-disk region, which is most perturbed by the direct action of repeated supernova explosions and where the Alfvén waves are severely damped due to ion-neutral damping, one could speculate that this region may host a local, non self-generated, turbulence directly associated to hydrodynamical turbulence.

One advantage of considering advection is that it acts as a physical "boundary" to CR propagation, whereas in the standard picture this is provided rather ad hocly by imposing a free escape boundary at height H. The way this is done is that we treat the distance $s_*(p)$ at which advection (which we assume increase linearly with distance, $v_A \sim \eta z$) dominates upon diffusion as an effective boundary. Then the role of H is played by the physical quantity $s_*(p)$ which however is an output of the problem (not imposed by hand) and depends on the particles' momentum.

C Collisions

[34] So that the material is not simply lifted up and fall down in what is known as Galactic fountains.

C.1 Electromagnetic Interactions

Electrons
- Excitation & Ionization
- Coulomb Scattering (+ Multiple Scattering)
- Bremsstrahlung
- Synchrotron Radiation & Inverse Compton
- Cherenkov radiation, Transition Radiation

Ionization Losses

The main source of energy loss mechanism for charged particles at low energy, $\beta\gamma < 1000$. E.g., it is an effective mechanism for heating of the interstellar gas and important for particle detection. The energy loss rate (including quantum and relativistic effects) is given by the famous *Bethe-Bloch* equation[35]

$$-\frac{dE}{dx} = Kz^2\frac{Z}{A}\frac{1}{\beta^2}\left[\ln\frac{2m_ec^2\beta^2\gamma^2}{I} - \beta^2 - \frac{\delta}{2}\right] \tag{160}$$

where

$x \equiv \int \rho dr$:mass per unit area (surface density) measured in g/cm^2
$K = 4\pi N_{\rm A} r_e^2 m_e c^2 = 0.307$MeV cm^2 mol^{-1}
z : charge of incident heavy particle
$Z\&A$: charge and atomic mass of the medium
I : mean excitation energy (main obstacle in applying Bethe-Bloch)
β^2, the second term in the parentheses is due to quantum correction
$\delta/2$, the third term in the parentheses is due to density correction

dE/dx is called the *mass stopping power*, with unit MeV$/[$g cm$^{-2}]$ while $\rho dE/dx$ is called the *linear stopping power*, with unit MeV/cm.

Multiple Scattering

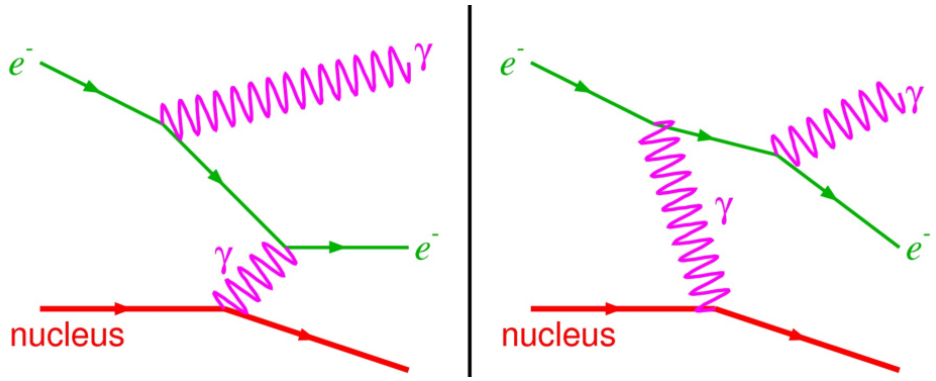
Interactions with the nuclei in the medium. Cannot transfer energy (because $m_{\rm N} \gg m_e$). Only result in direction changes.

[35] Bethe-Bloch formula does not apply to light particles like electron, but the general features of energy loss of electron remain the same. Another condition for Bethe-Bloch to hold is that the velocity of the particle cannot be too slow, $0.1 \lesssim \beta\gamma \lesssim 1000$.

Conclusion: Not of interest to CR propagation. However, it is important for detectors and extensive air showers.

Bremsstrahlung

The previously described process is an (quasi)elastic scattering between a charged energetic particle and electrons in the medium. But the energetic particle can also radiate a photon under the braking action of the nuclear electric fields.



Dominates at high energy. The energy loss rate of an electron through bremsstrahlung is nearly proportional to its energy:

$$-\left(\frac{dE}{dx}\right)_{\text{BS}} = \frac{E}{X_0}, \quad \text{where} \quad \frac{1}{X_0} = \frac{4e^4}{137c^4}\frac{N}{A}\frac{1}{m_{\text{inc}}^2}\ln\frac{183}{Z^{1/3}} \tag{161}$$

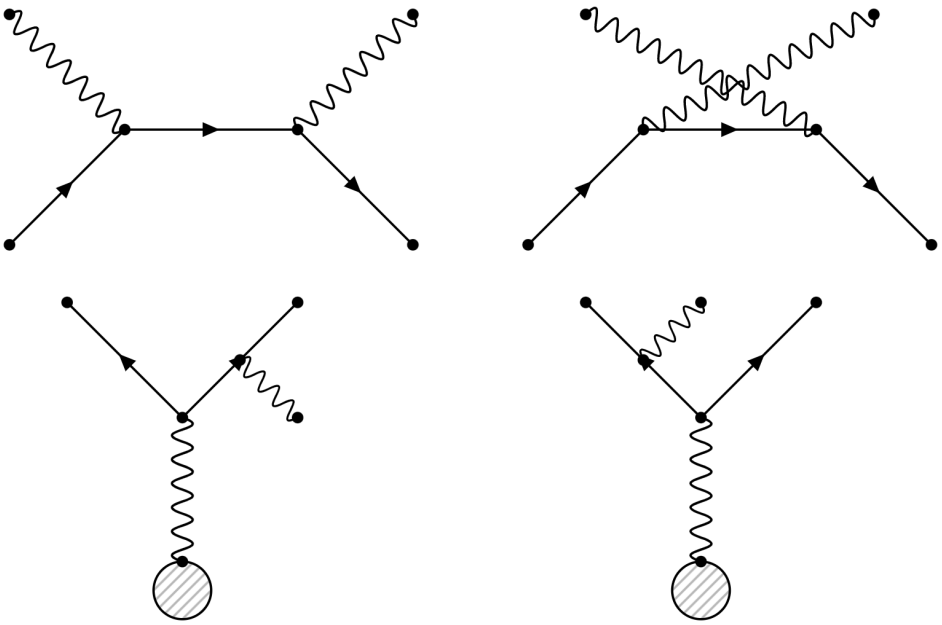
The *radiation length* X_0 can be viewed as the length required to emit one photon. Written in terms of energy loss per unit time

$$-\left(\frac{dE}{dt}\right)_{\text{BS}} = \alpha Z^2 \sigma_T n E, \tag{162}$$

where $\sigma_T = \frac{8\pi}{3}\frac{q^4}{m^2}$ is the Thomson cross-section.

IC and Synchrotron Radiation

IC and synchrotron radiation (a.k.a. magnetobresstrahlung) are basically the same processes, the only difference being the in the former e collides with a real photon while in the latter it collides with a virtual photon (pulled from the magnetic field).



At low energy (roughly speaking, when $E_\gamma \ll m_e$), classic E&M applies. And we can show that the power radiated is (written "quantum mechanically")

$$\langle P \rangle = \sigma_T u \tag{163}$$

where $u \equiv \langle |\mathbf{E}|^2/8\pi + \mathbf{B}|^2/8\pi \rangle = E_0^2/8\pi$ is the energy density in the e.m. field. Apply this formula to synchrotron radiation, where the electric field felt by the incident particle is provided by a Lorentz transformation, $\mathbf{E}' = -\gamma \mathbf{v} \times \mathbf{B}$. Hence, the synchrotron power reads

$$P_s = \frac{2q^4\gamma^2}{3m^2} v^2 B^2 \sin^2\theta \tag{164}$$

Eqn(163), which is derived in non-relativistic case, is actually also valid in more general since, loosely speaking, it's a dE/dt and energy and time transform the same way. After doing the appropriate Lorentz transformation, and taking the energy lost by the electrons per unit time being the difference of the scattered power minus incoming power $\sigma_T u$ (impinging photons had some energy, too!), we have

$$-\frac{dE}{dt} = \sigma_T u \left[\gamma^2 \left(1 + \frac{\beta^2}{3}\right) - 1\right] = \frac{4}{3}\gamma^2\beta^2 u \sigma_T \simeq \frac{4}{3}\gamma^2 u \sigma_T \tag{165}$$

If u is interpreted as the energy density of both photon fields and magnetic field, this formula describes both synchrotron and IC losses (in the Thomson regime)!

A full QED calculation gives the *Klein-Nishina* formula

$$\frac{d\sigma}{d\Omega} = \frac{3}{16\pi}\sigma_T \left(\frac{\epsilon_f}{\epsilon_i}\right)^2 \left(\frac{\epsilon_i}{\epsilon_f} + \frac{\epsilon_f}{\epsilon_i} - \sin^2\theta\right) \tag{166}$$

so that

$$\begin{aligned} \sigma(x) &\simeq \sigma_T(1 - 2x + \ldots) \quad \text{for} \quad x \ll 1 \text{ (Thomson)} \\ \sigma(x) &\simeq \frac{3}{8}\sigma_T \frac{1}{x}\left(\ln 2x + \frac{1}{2}\right) \quad \text{for} \quad x \gg 1 \quad (\text{ extreme KN}) \end{aligned} \tag{167}$$

where $x \equiv \epsilon_i/m_e$.

Beaming effect from relativistic motion

From the Lorentz transformation law for velocity, we can show that for a boost in the x direction:

$$\tan\theta \equiv \frac{u_y}{u_x} = \frac{u'_y}{\gamma\left(u'_x + \beta\right)} = \frac{u'\sin\theta'}{\gamma\left(\beta + u'\cos\theta'\right)} \tag{168}$$

where the source is in the primed frame and the Lab (observer on Earth) in the unprimed frame. This means that a photon ($u = c$) emitted at $\theta' = 0$ travels at $\theta = 0$, while a photon emitted at $\theta' = \pi/2$ travels at $\tan\theta \simeq \theta \simeq 1/\gamma$ as illustrated in Fig. 6 (left).

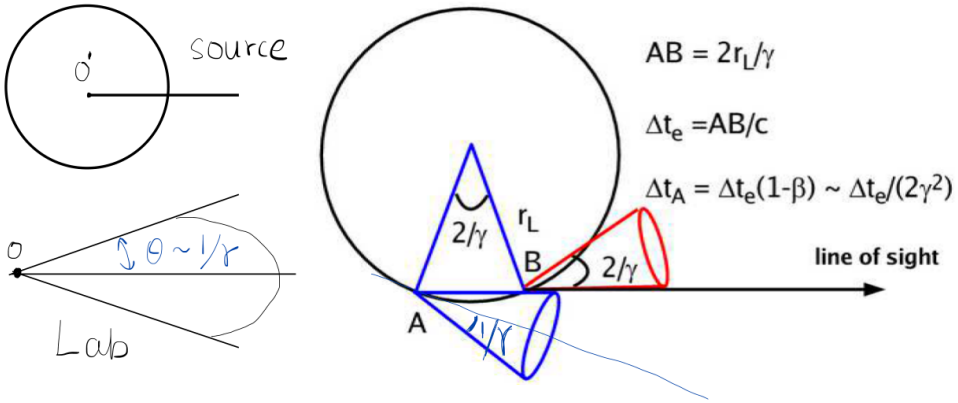


Figure 6: Beaming effect

This relativistic beaming alters the characteristic frequency of the photon emitted, because now only in a certain period of time can Earth's observer receive the signal, as illustrated in Fig. 6 (right). A careful calculation (using

the full power of Fourier transform) shows that the power emitted per unit frequency goes like $\sim \nu^{1/3} e^{-\nu/\nu_c}$, with the characteristic frequency of the order

$$\nu_s \simeq \frac{\gamma^3}{r_L} = \gamma^2 \frac{\omega_g}{2\pi} \Rightarrow E_s \simeq 500 \mu\text{eV} \frac{B}{\mu G} \left(\frac{E_e}{\text{GeV}} \right)^2 \tag{169}$$

which is calculated from Fig. 6 (right) using high school physics.

Synchrotron-Self-Compton (SSC)
The synchrotron photons produced by an energetic electron populations also constitute the main target for further upscatter via inverse Compton.

Photons
- Photoelectric Effect
- Compton Effect
- Pair Production
- Rayleigh Scattering, Photo-nuclear Reactions

C.2 Hadronic Interactions

For a particle of mass $m = Am_p$ and charge Ze, the Thomson cross-section σ_T scales as $\sigma_{\text{e.m.}}(m, Z) = Z^4 A^{-2} (m_e/m_p)^2 \sigma_T \simeq 3 \times 10^{-7} (Z^4/A^2) \sigma_T$, which immediately shows how much more inefficient the direct scattering photon-nucleus is compared to the photon-electron.

Therefore, we only need to consider strong interactions for nucleons and nuclei. We give a brief classification of hadronic interactions:

- *Spallation* nucleon-nucleon reactions: strong but short-range; not plant-like, rather extended objects of radius $R \simeq 1.2A^{1/3}$ fm. Together, these properties give a good rule of thumb $\sigma_{\text{tot}} \sim \pi R^2 \sim 45 A^{2/3} \text{mb}$, which is typically *energy independent*.

  Another thing to keep in mind is the *"superposition model"*, in which most of the target nucleus remains essentially unaffected, and the nucleon typically strikes only one or a few nucleons. This is confirmed by the fact that the energy per nucleon is approximately conserved in such processes.

These are the most relevant catastrophic interactions for nuclei propagating in a "matter-rich" environment (as the Galactic disk, as opposed to the extragalactic space).

- *Inelastic pp collisions* $p + p \to p + p(n) + \pi^0(\pi^+)$, the main energy-loss phenomenon affecting protons of Galactic Cosmic rays, with threshold energy $E_p \simeq 1.2$ GeV.
- For particles propagating over cosmological distances
  - *Adiabatic energy losses* i.e. "cosmological stretching" due to the expansion of the Universe:
    $-(1/E)dE/dt = H(z) = H_0\sqrt{\Omega_\Lambda + \Omega_m(1+z)^3}$
  - *Bethe-Heitler process* $p + \gamma_{\rm CMB} \to p + e^+ + e^-$, with threshold energy $E_p \simeq 2 \times 10^{18}$ eV and rather low inelasticity ($2m_e/m_p \simeq 10^{-3}$ at threshold).
  - *Nuclear photondisintegration* $A+\gamma \to (A-1)+N+\gamma$ in the EBL and CMB. For typical nuclear binding energies of $\mathcal{O}(10)$ MeV, the threshold lies at $E > 10^{19}$ eV. It typically damps the propagation of light nuclei before heavier one, with the Fe case whose flux is only affected by this process closer to 10^{20} eV.
  - *GZK cut off* $p\gamma$ onto CMB photons.

D Statistics

As usual, let us begin by throwing some words around the ART of statistics, that wonderful gray area that is not quite a branch of mathematics — and just as surely not quite a branch of science[36].

The progress of **Science** is impossible without the knowledge of probability and statistics. To see this let's consider how **Scientific Knowledge** evolves.

- We start with a problem (e.g., a new observation, or something unexplained in an existing theory). Probability and statistics are already required at this

[36] Quoted from numerical recipes: the art of scientific computing, third edition, page 720.

stage to make sure we have detected something at all.

- We then conjecture a set of theories, explaining the problem. These theories need to be able to give testable predictions. The predictions can be of a statistical nature, i.e., the mean value of a property of a population of objects or events. We then need probability and statistics to help us decide which theories pass the test and which do not.
- Statistics can never **prove** things, only **disprove** them! At best, you can substantiate a hypothesis by ruling out, statistically, a whole long list of competing hypotheses, every one that has ever been proposed. After a while your adversaries and competitors will give up trying to think of alternative hypotheses, or else they will grow old and die, and then your hypothesis will become accepted.

Sounds crazy, but that's how science works[37]!

Different areas of statistics are often overlapped, so it is not easy to give a clear-cut division of this filed, at least pedagogically. With that in mind, we make a somewhat arbitrary classification:

- **Detection of signals**: How do we know if we've found a new object or detected a new signal? *e.g., detection of astronomical objects, detection of fluctuations in temperature or polarization in the CMB, detection of galaxy clustering.*
- **Detection of correlations**: How do we detect any correlation between two quantities? And when should we believe it's real? *e.g., Hubble diagram.*
- **Tests of hypotheses**: How do we rule out a false theory? Or pass another? *e.g., isotropy of the Universe, the existence and nature of dark matter and dark energy in the Universe.*
- **Model-fitting to data and Estimation of parameters**: How should we compare our models with data? And if the model does fit the data, how do we extract estimates of the model parameters that best fit the data? *e.g., what are the parameters in the ΛCDM model.*
- **As a theoretical tool**: How do we model or predict the properties of

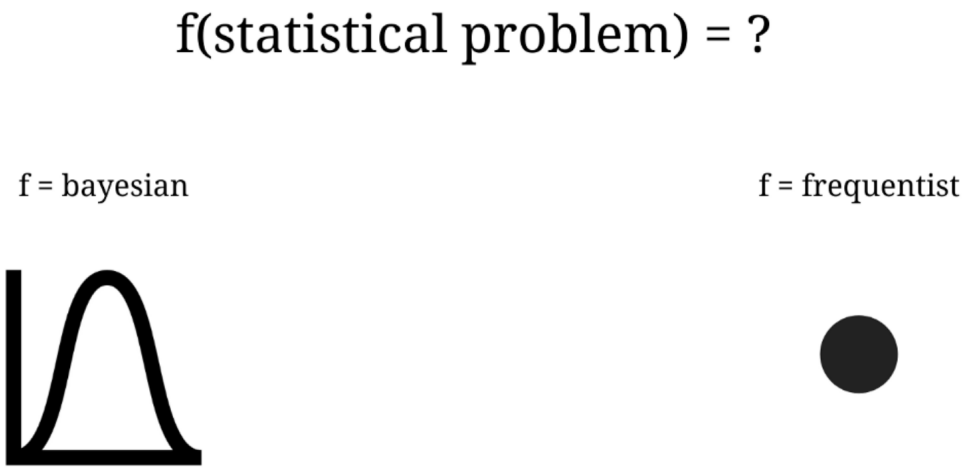
[37] "Science advances one funeral at a time." —Max Planck (attributed)

populations of objects, or study complex systems? *e.g., Statistical properties of galaxy populations, or stars, simulations of stellar clusters or large-scale structure.*

D.1 What is a Probability

There are two main schools of statistics, the **frequentist**, a.k.a. **classical** approach (think P values and confidence intervals) and the **Bayesian** approach, governed by Bayes' theorem.

The fundamental difference is their interpretation of uncertainty and probability: frequentists assign probabilities to *data*, not to hypotheses, whereas Bayesians assign probabilities to *hypotheses*, not to data. Furthermore, Bayesian models incorporate *prior knowledge* into the analysis, updating hypotheses probabilities as more data become available. Interpretation of results is more intuitive with a Bayesian approach compared with the frequentist approach, which can often be misinterpreted.



Frequentist Statistics

Frequentist statistics is all about probability *in the long run*; the data set collected and analyzed is one of *many hypothetical data sets* addressing the same question, and uncertainty is due to *sampling error alone*. We can have a more concrete feeling of it by looking at its definition of probability:

"the number of times the event occurs over the total number of trials, in the limit

of an infinite series of equiprobable repetitions."

The paradigmatic approach to frequentist hypothesis testing is the *tail text* or *p-value test*, which goes like follows:

- Begins by assuming a null hypothesis *to be true* before data are collected.
- Collect data, then apply some formula to the data to compute "a statistic".
- See where the value of that statistic falls in a probability distribution that is computed based on the "null hypothesis" and ask, "How surprising is my result?" (i.e. compute the p-value).
- If it falls in a very unlikely spot, way out on a tail of the distribution, conclude that the null hypothesis is false for your data set.
- If it falls in a reasonable spot, we *cannot* conclude the null hypothesis as proved, because the inference began by assuming the null hypothesis.

The key here is that probabilistic statements (i.e. p-values) can *only be made about the data*, not about hypotheses or parameters. Frequentists simply do not believe in some universe of hypotheses. A given hypotheses/parameter is either right or wrong, and there is no probability assigned to them.

Another common mistake regarding frequentist approach is the interpretation of confidence intervals. A 95% confidence interval tells us that if we were to repeat the experiment over and over (remember, frequentist statistics are long run), 95% of the computed confidence intervals would contain the true mean. This is different from saying there is 95% chance the true mean lies within the interval, because frequentist statistics cannot assign probabilities to parameters—the true mean either lies within the interval or it does not.

As for actual estimating parameters and giving the corresponding 68% confidence interval (C.L.), what a frequentist should do according to the discipline is (a) and (b), but people usually just uses decreases in the likelihood

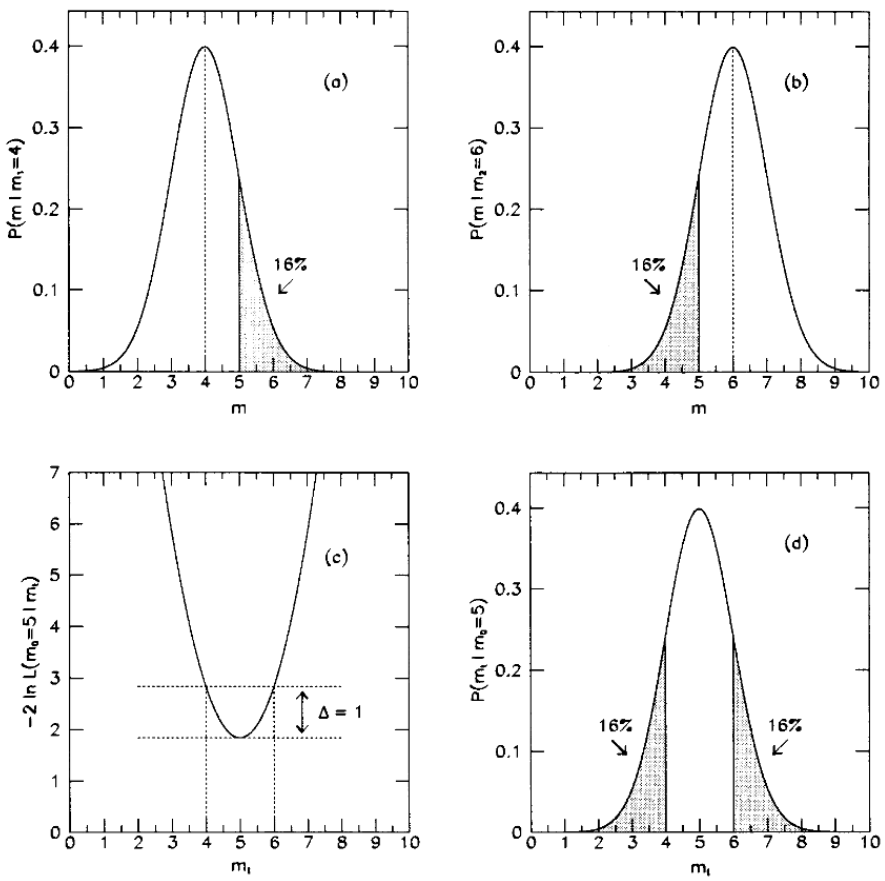


FIG. 1. Methods for 68% C.L. confidence interval construction in case of Gaussian pdf with known $\sigma_m = 1.0$, after single measurement yields $m_0 = 5.0$. (a) and (b) Classical construction of left and right interval endpoints; (c) increase of $-2\ln\mathcal{L}$ by one unit (d) Bayesian with uniform prior.

Bayesian Statistics

In contrast to the frequentist approach, parameters and hypotheses are seen as probability distributions and the data as fixed.This idea is perhaps more intuitive (and often more constant) because generally the data we collect are the only data set we have.

For a Bayesian, probability distributions summarize the current state of knowledge about a parameter or hypothesis[38] and can be updated as more data becomes available using Bayes' theorem.

$$p(\theta \mid \text{Data}) = \frac{p(\text{Data} \mid \theta) \cdot p(\theta)}{p(\text{Data})} \tag{170}$$

Where $p(\theta \mid \text{Data})$ is the posterior, $p(\text{Data} \mid \theta)$ is the likelihood (a function

[38] Bayesian definition of probability: *"probability is a measure of the **degree of belief** about a proposition"*.

of θ), $p(\theta)$ is the prior, and $p(\text{Data})$ is the Bayesian evidence (a.k.a. marginal likelihood, a normalizing constant). Sometimes formulating a prior is not that easy or clear. In such cases, an ~~*"uninformative" prior* that provides no additional information~~[39] could be used, or multiple priors could be tested to determine the sensitivity of the results to particular priors.

From an epistemological perspective, the posterior probability contains *everything* there is to know about an uncertain proposition (such as a scientific hypothesis, or parameter values), given prior knowledge and a mathematical model describing the observations available *at a particular time*. And one of the big advantages of Bayesian inference, is that the posterior distribution can also be used to directly calculate the probability of different hypotheses.

For example, we can extract the mean and 95% credible interval from the posterior. The 95% credible interval is different from a frequentist 95% confidence interval; it is the parameter range that has a 95% probability of including the true parameter value. The frequentist confidence interval is often misinterpreted in this way.

Marginalized probability (very important for dealing with *nuisance parameters*[40]) is also trivial to compute in Bayesian statistics: just integrate out the things that you do not care about:

$$p(\theta \mid \text{Data}) \propto \int \mathcal{L}(\theta, \theta_{\text{nui}}) \cdot p(\theta, \theta_{\text{nui}}) \mathrm{d}\theta_{\text{nui}} \tag{171}$$

Note The difference between Bayesian and frequentist is more than a philosophical one. Aside from different perspective on probability and procedures taken when dealing with a problem, they often give quite different confidence intervals[41]:

[39] Every choice of prior, no matter how innocent, introduces bias. To see this, just ask yourself if you are going to use a "flat" prior, in what measure is the prior "flat", linear? Logarithmic? or something else?

[40] Parameters that have an influence on the data but are of no interest for us. Frequentist methods offer no simple way of dealing with nuisance parameters (the very name derives from the difficulty of accounting for them in classical statistics).

[41] Credible intervals, in a Bayesian language, Though in the literature confidence/credible intervals are often mixed together.

- We cannot say which approach is more conservative (i.e. larger intervals). It depends on the problem and the Bayesian prior chosen.
- Both methods rely critically on knowing $P(m_{\text{obs}} \mid m_{\text{true}})$ correctly[42].
- Operationally, the line of demarcation between frequentists' and Bayesian methods essentially lies in *whether or not $\mathcal{L}$ is integrated*. (It is illegitimate to integrate over the likelihood as if it were a probability density for a frequentist).
- Because frequentists cannot integrate over anything, they heavily rely on mathematical theorem to tell them what are the asymptotic distributions of the estimators. These theorems often depend one way or another on the Central Limit Theorem[43]. So classical method for extracting intervals $\mathcal{L}$ often yields only *approximate* confidence intervals. But nobody seems to care about this caveat, they just use whatever number of samples they happen to be working with[44].

Perhaps the best way to summarize the debate among statisticians is a quote from Louis Lyons:

"Bayesians address the question everyone is interested in by using assumptions no one believes, while frequentists use impeccable logic to deal with an issue of no interest to anyone".

D.2 Averages and Error Bars

In this section we include so-called *descriptive statistics* (another rather arbitrary division) that characterize a data set in general terms: its mean, variance, covariance and so on. A general setup in this category is the following. We are handed a set of values $x_i, y_i, \cdots, i = 1 \cdots N$, where N is the number of measurements. The first thing we will want to do is to estimate various average values, and determine error bars on those estimates.

[42] The difference between P and $\mathcal{L}$ is in the viewpoint: the former is a function of the observed data m_{obs}, while the latter is a function of m_{true}

[43] any likelihood function is asymptotically Gaussian near its maximum

[44] The Central Limit Theorem is an often very good approximation for sample numbers like 20 or 30.

Using only information from the data, we can show that the (unbiased) mean is given by

$$\mu = \bar{x} \pm \sigma_{\bar{x}} \tag{172}$$

where

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \sigma_{\bar{x}} = \frac{1}{\sqrt{N}} s \text{ and } s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{173}$$

The rationale for $N-1$ is that the N terms in the last equation are not all independent since $\bar{x}$, which depends on all the x_i, is subtracted. So there are only $N-1$ independent variables (the number of degree of freedom) remain.

s^2 does not vary systematically with N, at large N (where we neglect the factor of -1 compared with N) and so we see that

the error bar in the mean goes like $1/\sqrt{N}$ for large N.

This is discouraging, but is a fact of life when dealing with random noise.

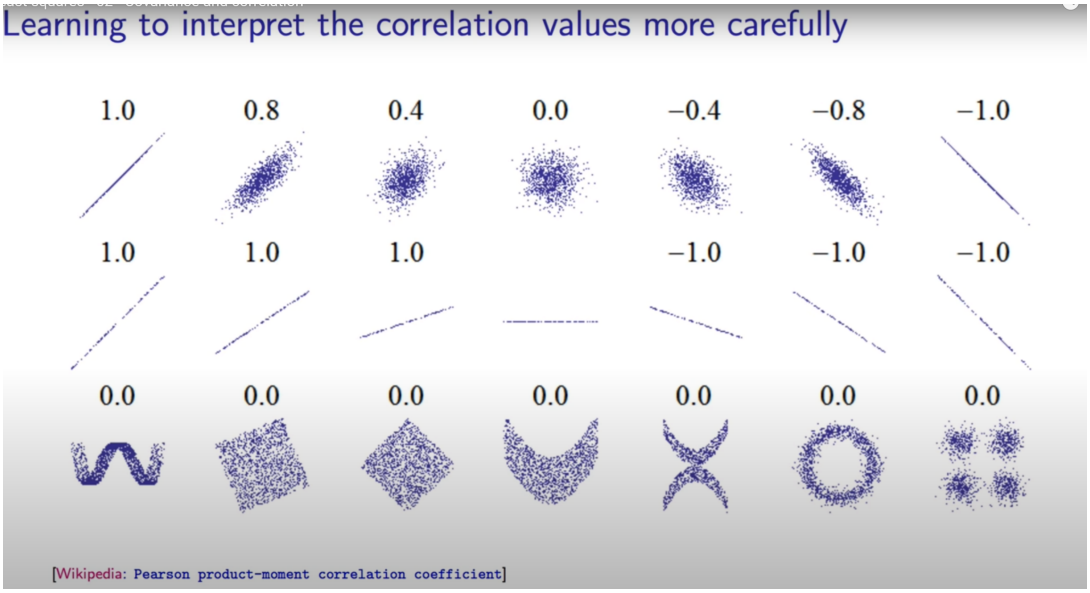
For eqn(173) to be truly useful, we need to know the probability that the true answer μ lies more than $\sigma_{\bar{x}}$ away from our estimate $\bar{x}$. Fortunately, for large N , the central limit theorem tells us (for distributions where the first two moments are finite) that the distribution of x is a Gaussian. Even if N is not very large, so there are some deviations from the Gaussian form, the above numbers are often a reasonable guide.

Note Be aware of the **outliers**!

For multiple parameters, we can give the *covariance matrix* of the parameter set $\vec{\theta}$ as well:

$$\mathcal{C}ov_{ij} = \frac{1}{N}\sum_{k=1}^{N}\left(\theta_i^k - \bar{\theta}_i\right)\left(\theta_j^k - \bar{\theta}_j\right) \tag{174}$$

We can also define the corresponding *correlation matrix* (nothing but normalized covariance matrix) $r_{ij} = \mathcal{C}ov_{ij}/\sqrt{\mathcal{C}ov_{ii}\mathcal{C}ov_{jj}}$. Note that a vanishing correlation matrix does not necessarily mean that two parameters are uncorrelated:



So covariance matrix only fully ~~captures the relations between parameters when the parameters follow multivariable Gaussian~~.

D.2.1 Error propagation

If the errors fellow a multivariable Gaussian distribution, then we can propagate the errors easily with the help of the *covariance matrix* of the original errors. If it is a linear model, then the propagated errors are still Gaussian. If not, then there is not much we can say about it.

However, it is now more common to use either *"jackknife"* or *"bootstrap"* procedures. At the price of some additional computation, which is no difficulty when done on a modern computer, these methods *automate* the calculation of the error bar.

D.3 Fitting Data to a Model

A fitting procedure should contain three basic parts:

1. give the values of the fit parameters, but also
2. provide error[45] estimates on those parameters, and

[45] a better name is *uncertainty*

3. provide a measure of how good the fit is.

If the result of part 3 is that the fit is very poor, the results of parts 1 and 2 are probably *meaningless*. *Chi-by-eye* is NOT OK!

D.3.1 Classical approach

Classical approach to parameter estimation uses the related methods **Maximum Likelihood** and **Least Square**. Although there are some mathematical theorems (go by the name of *Gauss-Markov theorem*) that justify these two methods, we will apply Bayesian idea because then the logic is clearer. Start with the Bayes theorem:

$$p(\theta \mid \text{Data}) = \frac{p(\text{Data} \mid \theta) \cdot p(\theta)}{p(\text{Data})}$$

Let us see how method of maximum likelihood and method of least squares naturally fellow.

Suppose that we are fitting N data points $(x_i, y_i)\,, i = 0, \cdots, N-1$, to a model that has M adjustable parameters $a_j, j = 0, \cdots, M-1$. The model predicts a functional relationship between the measured independent and dependent variables,

$$y(x) = y\left(x \mid a_0 \cdots a_{M-1}\right) \tag{175}$$

To be more quantitative, suppose that each data point y_i has a measurement error that is *independently* random and distributed as a *Gaussian distribution* (with mean 0 and variance σ_i) around the "true" model $y(x)$. Then the likelihood reads

$$P(\text{data} \mid \text{model}) \propto \prod_{i=0}^{N-1} \left\{ \exp\left[-\frac{1}{2} \left(\frac{y_i - y\left(x_i\right)}{\sigma_i} \right)^2 \right] \Delta y \right\} \tag{176}$$

As Bayesians, we proceed by invoking Bayes' theorem with a *"uninformative" prior*. The most probable model, then, is the one that maximizes eqn(176) or, equivalently, minimizes the negative of its logarithm,

$$\left[\sum_{i=0}^{N-1} \frac{\left[y_i - y\left(x_i\right)\right]^2}{2\sigma_i^2} \right] - N \log \Delta y \tag{177}$$

Since N, σ_i and Δy are all constants, minimizing this equation is equivalent to minimizing

$$\chi^2 \equiv \sum_{i=0}^{N-1} \left(\frac{y_i - y\left(x_i \mid a_0 \ldots a_{M-1}\right)}{\sigma_i} \right)^2 \tag{178}$$

called the "chi-square".

A generalization of the ordinary least squares (OLS) estimator is the generalized least squares (GLS) estimator. It is used to deal with situations in which the OLS estimator is not BLUE (best linear unbiased estimator) because one of the main assumptions of the Gauss-Markov theorem, namely that of homoskedasticity and absence of serial correlation, is violated. In such situations, provided that the other assumptions of the Gauss-Markov theorem are satisfied, the GLS estimator is BLUE:

$$\chi^2 \equiv \sum_{i,j} (y_i - y(x_i)) \mathcal{C}ov_{ij}^{-1} (y_j - y(x_j)) \tag{179}$$

where $\mathcal{C}ov$ is the covariance matrix between different measurements

Conclusion The least-squares fitting gives an answer that is both (i) the most probable parameter set in the Bayesian sense, assuming a flat prior, and (ii) the maximum likelihood estimate of the fitted parameters, in both cases if the measurement errors are independent and normally distributed. Notice that we made no assumption about the linearity or nonlinearity of the model $y\left(x \mid a_0 \ldots a_{M-1}\right)$in its parameters $a_0 \cdots a_{M-1}$.

So Maximum Likelihood/Least Square turn parameter estimation into a problem of minimization/maximization. And there are many methods we can achieve this, for instance,
Linear models: Singular Value Decomposition (SVD), recommended in all circumstance; Solution by use of the normal equation (e.g., Gauss-Jordan elimination, QR decomposition)
Nonlinear models: Take a step down the gradient; Levenberg-Marquardt Method.

Goodness of fit

For models that are linear in the a's, it turns out that the probability distribution

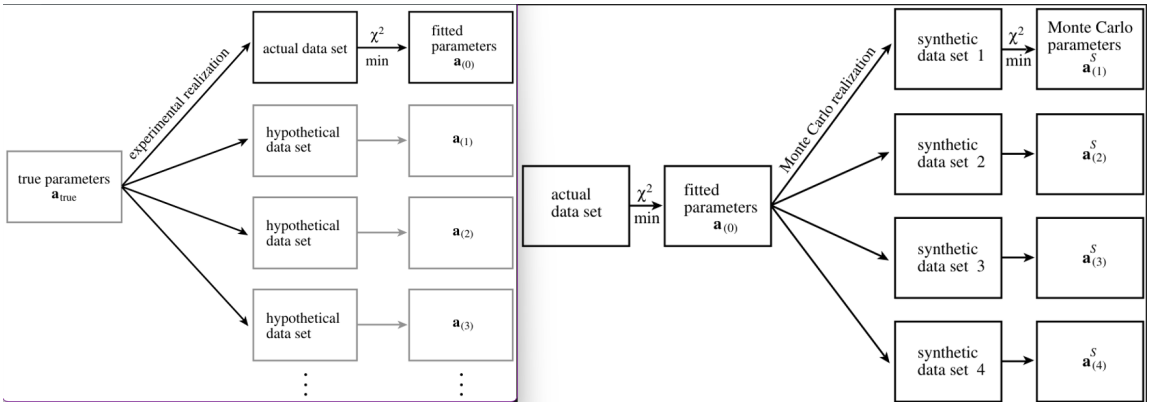
for different values of χ^2 at its minimum can nevertheless be derived analytically, and is the chi-square distribution for $\nu \equiv N - M$ degrees of freedom[46]. We can then compute Q, which is a measure for the goodness-of-fit of the model.

If Q is a very small probability for some particular data set, then the apparent discrepancies are unlikely to be chance fluctuations. At the opposite extreme, if Q is too large, too near to 1, then the experimenter probably overestimated his or her measurement errors.

A rule of thumb is that a "typical" value of χ^2 for a "moderately" good fit should be in the range $\nu^2 \pm \sqrt{2\nu}$.

Confidence Limits

Rather than present all details of the probability distribution of errors in parameter estimation, it is common practice to summarize the distribution in the form of confidence limits[47]. The full probability distribution is a function defined on the M dimensional space of parameters $\vec{\mathbf{a}}$. A confidence region (or confidence interval)is just a region of that M-dimensional space (hopefully a small region) that contains a certain (hopefully large) percentage of the total probability distribution. The experimenter get to pick both the confidence level and the shape of the confidence region.



[46]Because for each parameter, we need one constraint of the type $\partial\chi^2/\partial\theta_p = 0$, which decrease our freedom by one. From this we can also see that different parameters need to be independent of for $\nu \equiv N - M$ to apply. For nonlinear models, the question of DoF is tricky since that different parameters may be correlated. Nevertheless, we can use $N - M$ when the number of samples is large since then ν is dominated by N.

[47]For a frequentist, this is the only choice since there is no such thing as the probability distribution of errors in parameter estimation

Note The commonly used numbers for confidence intervals, 68.3%, 95.4%, and 99.73%, and the use of ellipsoids, are purely matters of convention nowadays. Historically, they had something to do with a normal distribution.

When the parameters are obtained by a chi-square minimization, then there is a natural choice for the shape of confidence intervals, whose use is almost universal. The difference[48] in chi-square, $\Delta\chi \equiv \chi^2 - \chi^2_{\text{min}}$ (often written as $2\Delta \ln \mathcal{L}$), which satisfies a χ^2 distribution of ν Dof, defines an M-dimensional confidence region around the best fit $\mathbf{a}_0$.

D.3.2 Bayesian Approach

As promised, Bayesian approach to parameter estimation is very straight forward and intuitive. And the "errors" can only come from insufficient MC sampling size, not from approximation of the Central Limit.

Just calculate the likelihood (as we have said before, the likelihood is the same for frequentists and Bayesian), then choose your prior based on previous knowledge. Now you know what the posterior looks like (If you do not care about model comparison, then the Bayesian evidence is of no interest to you), all that is left is to *draw samples* from the posterior and *replace* the actual posterior with the sample.

So the only problem remains is how to draw samples accurately and efficiently. Fortunately this is not a major hurdle thanks to the recent increase of cheap computational power.In particular, numerical inference often employs a technique called *Markov Chain Monte Carlo*, which allows mapping out numerically the posterior distribution even in the most complicated situations.

MCMC

Definition: Markov chain Monte Carlo (MCMC) methods comprise a class of algorithms for sampling from a probability distribution. By constructing a Markov chain that has the desired distribution as its equilibrium distribu-

[48] Another equivalent choice is the *ratio* between χ^2's, which satisfies the F-distribution.

tion, one can obtain a sample of the desired distribution by recording states from the chain. Various algorithms exist for constructing chains, including the Metropolis–Hastings algorithm.

The huge advantage of MCMC is that it "automatically" puts its sample points preferentially where $\pi(\mathbf{a}) \equiv P(\text{Data} \mid \mathbf{a})P(\mathbf{a})$ (i.e. the unnormalized posterior) is large (in fact, in direct proportion). In a high-dimensional space, or where $\pi(\mathbf{a})$ is expensive to compute, this can be advantageous by many orders of magnitude.

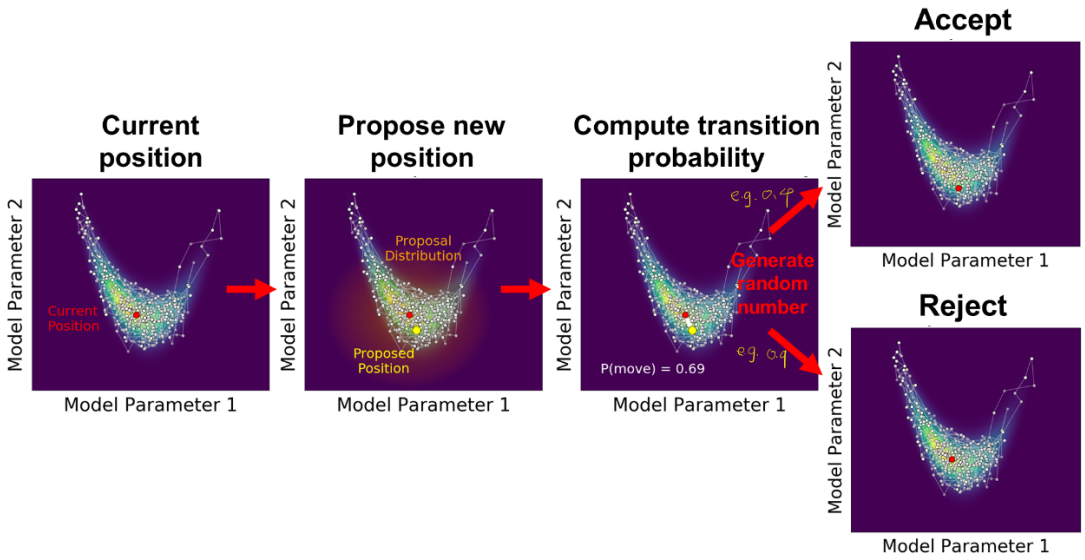


Figure 9: A schematic illustration of the Metropolis-Hastings algorithm. At a given iteration i, we have generated a chain of samples $\{\Theta_1 \to \cdots \to \Theta_i\}$ (white) up to the current position Θ_i (red) whose behavior follows the underlying posterior $\mathcal{P}(\Theta)$ (viridis color map). We then propose a new position Θ'_{i+1} (yellow) from the proposal distribution (orange shaded region). We then compute the transition probability $T(\Theta'_{i+1}|\Theta_i)$ (white) based on the posterior $\mathcal{Q}(\Theta)$ and proposal $\mathcal{Q}(\Theta'|\Theta)$ densities. We then generate a random number u_{i+1} uniformly from 0 to 1. If $u_{i+1} \leq T(\Theta'_{i+1}|\Theta_i)$, we accept the move and make our next position in the chain $\Theta_{i+1} = \Theta'_{i+1}$. If we reject the move, then $\Theta_{i+1} = \Theta_i$. See

Once a Markov chain has been constructed, obtaining Monte Carlo estimates of expectations for any function of the parameters becomes a trivial task:

$$\langle f(\theta) \rangle \approx \frac{1}{M} \sum_{t=0}^{M-1} f\left(\theta^{(t)}\right) \tag{180}$$

As for marginal posterior for a particular (or two/three, it really does not matter) parameter $p\left(\theta_1 \mid d\right) = \int p(\theta \mid d) \mathrm{d}\theta_2 \ldots \mathrm{d}\theta_n$, we simply divide the range of θ_1

in a series of bins and *count the number of samples falling within each bin*.

Credible intervals are also easy to obtain: just find a region in the whole parameter space that contains 68% (or 95%, whatever you like) of all the sample points. An added benefit is that not only can we give the error bars on the estimated parameters, we can actually plot the whole probability distribution of any estimated parameters. This is particular helpful when the marginalized posterior is multimodal.

There are no good fully Bayesian methods for assessing goodness-of-fit, that is, for comparing the probability of a best-fit model to that of a *nonspecific* alternative hypothesis like "the model is wrong." The problem is that Bayesian model comparison maintains that it is pointless to reject a theory unless an alternative explanation is available that fits the observed facts better. Sensible Bayesians usually fall back to p-value (i.e. tail probability) like chi-square probability Q when they really need to know if a model is wrong.

Note If we fit some function to N data points with Gaussian errors, we have to expect that 31.7% of all data points deviate from this fit by more than one sigma. In other words, If you are presented fitted data where the fit goes through all 1σ-errorbars, you should definitely be skeptical.

D.4 Model Comparison

D.4.1 Classical approach – Likelihood-Ratio Test

The likelihood-ratio test assesses the goodness of fit of two competing statistical models based on the ratio of their maximum likelihoods, specifically one found by maximization over the entire parameter space and another found after imposing some constraint (i.e. the competing models are *nested*).

By Wilks' theorem, the ln likelihoods ratio

$$\lambda_{\mathrm{LR}} = -2\ln\left[\frac{\max_{\theta\in\Theta_0}\mathcal{L}(\theta)}{\max_{\theta\in\Theta}\mathcal{L}(\theta)}\right] \sim \chi^2(r) \tag{181}$$

where the null hypothesis is often stated by saying that the parameter Θ is in

a specified subset Θ_0 of Θ. The alternative hypothesis is thus that Θ is in the complement of Θ_0, Θ_0/Θ. And the DoF r of the resulting χ^2 distribution is given by the number of nested parameters.

One should also pay attention to the so-called **look elsewhere effect**, which happens in an hypothesis test when a nuisance parameter is present only under the alternative. For illustration, when searching for a new resonance somewhere in a possible mass range, the significance of observing a local excess of events must take into account the probability of observing such an excess anywhere in the range.

A straightforward way of quantifying the look elsewhere effect may be simply running many Monte-Carlo simulations of background only experiments, and finding for each one the largest fluctuation that resembles a signal. Approximations based on large sample limits also exist.

D.4.2 Bayesian Factor

The evaluation of a model's performance in the light of the data is based on the *Bayesian evidence* (a.k.a. marginal likelihood or model likelihood):

$$p(\text{Data} \mid \mathcal{M}) \equiv \int_{\Omega_{\mathcal{M}}} p(\text{Data} \mid \theta, \mathcal{M}) p(\theta \mid \mathcal{M}) \mathrm{d}\theta \tag{182}$$

From the evidence, the *model posterior probability* given the data is obtained by using Bayes' Theorem:

$$p(\mathcal{M} \mid \text{Data}) = \frac{p(\mathcal{M}) p(\text{Data} \mid \mathcal{M})}{\text{Normalization}} \tag{183}$$

where $p(\mathcal{M})$ is the model prior (often set to be unity). When comparing two models, $\mathcal{M}_0$ versus $\mathcal{M}_1$, one is interested in the ratio of the posterior probabilities, or posterior odds, given by

$$\frac{p\left(\mathcal{M}_0 \mid \text{Data}\right)}{p\left(\mathcal{M}_1 \mid \text{Data}\right)} = K_{01} \frac{p\left(\mathcal{M}_0\right)}{p\left(\mathcal{M}_1\right)} \tag{184}$$

and the all important **Bayes factor** K_{01} (denoted B_{01} or Z_{01} in some literatures) is the ratio of the models' evidences:

$$K_{01} \equiv \frac{p\left(d \mid \mathcal{M}_0\right)}{p\left(d \mid \mathcal{M}_1\right)} \quad \text{(Bayes factor)} \tag{185}$$

After K is computed, we use an empirical scale for evaluating the strength of evidence (so–called "Jeffreys' scale"). The probability column refers to the posterior probability of the favored model, assuming non–committal priors on the two competing models, i.e. $p(\mathcal{M}_0) = p(\mathcal{M}_1) = 1/2$ and that the two models exhaust the model space, $p(\mathcal{M}_0|\text{Data}) + p(\mathcal{M}_1|\text{Data}) = 1$.

$\lvert \ln K_{01} \rvert$	Odds	Probability	Strength of evidence
< 1.0	$\lesssim 3:1$	< 0.750	Inconclusive
1.0	$\sim 3:1$	0.750	Weak evidence
2.5	$\sim 12:1$	0.923	Moderate evidence
5.0	$\sim 150:1$	0.993	Strong evidence

A major advantage of Bayesian model selection is that it automatically implements *Occam's Razor*: a simpler theory with compact parameter space will have a larger evidence than a more complicated one, unless the latter is significantly better at explaining the data. **Caveats**:

- Hidden prior information.
- No well–defined prior measure.
- Lindley's paradox
- Using the data twice (related to look elsewhere effect).

E Astronomy

Some jargon

- It is customary in astronomy to use the Roman numeral I for neutral atoms, II for singly-ionised and III for doubly-ionised, etc. E.g., HI is neutral hydrogen, HII is H^+ OIII is O^{++}.

Coordinate Systems

F Useful Quantities

Gravitational Constant

$$G_N = \frac{1}{M_{\rm Pl}^2} \simeq (\frac{1}{10^{19}\text{GeV}})^2$$

Planck Mass

$$M_p = \sqrt{\frac{c\hbar}{G}} = 1.22 \times 10^{19} GeV \tag{186}$$

Reduced Planck Mass

$$M_p = \sqrt{\frac{c\hbar}{8\pi G}} = 2.4 \times 10^{18} GeV \tag{187}$$

Critical Density

$$\rho_{\rm crit} \simeq 3 \times 10^{10} \frac{M_\odot}{\text{Mpc}^3} \simeq 10^{-6} \frac{\text{GeV}}{\text{cm}^3} \tag{188}$$

Natural Units $c = \hbar = k_{\rm B} \equiv 1$

- $c = 3 \times 10^8 \text{m} \cdot \text{s}^{-1}$, $\hbar = 6.582 \times 10^{-22}$ MeV·s, $\hbar c = 197.3$MeV·fm
- 1 eV $\approx 1.1604 \times 10^4$ K
- 1 J = 10^7 erg = 6.25×10^{18} eV = 1.11×10^{-17} kg
- 1 s = 3×10^8 m = 1.5×10^{15} eV^{-1}
- e $\simeq \sqrt{\alpha} \simeq 0.085$
- 1 Gauss = 1.95×10^{-20} GeV^2, 1.9×10^{-18} C =1, 1 V = 12 eV

Distances

- Distance from Earth to the Sun $\equiv$ 1AU=1.5×10^{13}cm
- Distance at which 1AU subtends 1 sec of arc = 3.1×10^{18}cm$\equiv$ 1pc $\simeq$ 3.3ly
- At low redshift, $d \simeq 40(z/0.01)$ Mpc
- distance from the Sun to the nearest star (Proxima Centauri) = 1.3pc
- distance from the Sun to the GC = 8.5 kpc
- size of Local Group = 1Mpc

- distance to the Virgo cluster = 16 Mpc
- Hubble Radius = 4450 Mpc

Masses

- $1M_\odot = 2 \times 10^{33}$ g
- Stellar mass of Milk Way $\simeq 1 \times 10^{11} M_\odot$
- Dynamical mass of Milk Way $\simeq 1 \times 10^{12} M_\odot$ (Dark Matter !)
- Mass of the Virgo cluster $\approx 10^{15} M_\odot$

Densities

- Neutron Stars $\sim 10^{15}$g/cm^3
- the Sun ~ 1g/cm^3 (i.e. water)
- the Galaxy $\lesssim 1$ particle cm^{-3}
- in a cosmological setting, $\lesssim 10^{-6}$ particle cm^{-3}, on average

Clusters