Lecture 4: WIMP & Axion Dark Matter Production

Graduate Course in Astroparticles and Cosmology

Luca Visinelli

Università di Salerno, Spring 2026

Plan of the lecture

Suggested references and resources

Classic papers

- Lee and Weinberg — *Cosmological lower bound on heavy-neutrino masses*
- Srednicki, Watkins and Olive — *Calculations of relic densities in the early Universe*
- Gondolo and Gelmini — *Cosmic abundances of stable particles*
- Griest and Seckel — *Three exceptions in the calculation of relic abundances*

Textbooks and reviews

- Steven Weinberg — *Cosmology*
- Kolb and Turner — *The Early Universe*
- Dodelson and Schmidt — *Modern Cosmology*
- Bertone, Hooper and Silk — *Particle dark matter*

The dark matter problem

A wide range of observations indicates that most of the matter in the Universe is non-luminous and interacts predominantly through gravity. In terms of density parameters defined with respect to the critical density, the present energy budget is approximately

$$\Omega_b \simeq 0.05, \qquad \Omega_{\rm DM} \simeq 0.26, \qquad \Omega_\Lambda \simeq 0.69. \tag{1}$$

Baryons are only a subdominant part of the total matter density. The remaining matter component behaves, at least after recombination, as a pressureless gravitating fluid.

Central question

What particle, or field, accounts for the non-baryonic matter density inferred from cosmology and gravitational dynamics?

Galactic rotation curves

For a test particle in circular orbit inside a spherically symmetric mass distribution, Newtonian gravity gives

$$v^2(r) = \frac{GM(r)}{r}. \tag{2}$$

If the visible matter were concentrated in the stellar disk, one would expect the velocity to decrease at large radius. Observationally, however, spiral galaxies often show

$$v(r) \simeq \text{const.} \tag{3}$$

The enclosed mass must then grow approximately linearly with radius,

$$M(r) \propto r, \tag{4}$$

which is naturally interpreted as evidence for an extended halo of non-luminous matter.

Clusters, lensing, and the CMB

Galaxy clusters provide an independent gravitational laboratory. The velocity dispersion of cluster galaxies, the X-ray temperature of the hot intra-cluster gas, and weak gravitational lensing all point to total masses much larger than those contained in luminous galaxies.

In merging systems such as the Bullet Cluster, the dominant lensing mass is spatially displaced from the X-ray gas. This separation is difficult to explain if all gravitating matter were ordinary collisional baryons.

Cosmological evidence

The acoustic structure of the CMB and the growth of large-scale structure require a dominant non-baryonic component that is already non-relativistic by the time structure begins to form efficiently.

Cold dark matter as an effective description

The observations are summarized by the statement that dark matter is approximately cold, collisionless, non-baryonic, and pressureless on cosmological scales.

At the level of background cosmology, this means

$$p_{\mathrm{DM}} \simeq 0, \qquad \rho_{\mathrm{DM}} \propto a^{-3}. \tag{5}$$

This is a successful macroscopic description, but it does not identify the microscopic nature of the dark component.

Particle-physics problem

One must explain why the relic abundance is of the observed size and why the candidate is stable, neutral, and sufficiently weakly coupled to ordinary matter.

The WIMP hypothesis

A weakly interacting massive particle, denoted by χ, is assumed to be stable on cosmological time scales and to have been in thermal contact with the primordial plasma at early times.

At high temperature, annihilation and production reactions maintain chemical equilibrium,

$$\chi\chi \leftrightarrow f\bar{f}. \tag{6}$$

As the temperature falls below m_χ, the equilibrium number density becomes exponentially small. Eventually the annihilation rate can no longer compete with the expansion rate.

Thermal relic idea

The present abundance is the frozen remnant of an early equilibrium population.

Phase-space Boltzmann equation

The relic abundance is determined by the phase-space distribution $f(t, \mathbf{p})$. For a homogeneous FLRW background, the Boltzmann equation is

$$\frac{\partial f}{\partial t} - H\,\mathbf{p} \cdot \frac{\partial f}{\partial \mathbf{p}} = C[f]. \tag{7}$$

The second term is the redshifting of physical momenta due to cosmic expansion.

The number density is

$$n_\chi(t) = g_\chi \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3} f(t, \mathbf{p}), \tag{8}$$

where g_χ counts internal degrees of freedom.

From phase space to number density

Integrating the Liouville operator over momentum gives

$$g_\chi \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3} \left(\frac{\partial f}{\partial t} - H\mathbf{p} \cdot \frac{\partial f}{\partial \mathbf{p}} \right) = \dot{n}_\chi + 3Hn_\chi. \tag{9}$$

Thus the left-hand side of the integrated Boltzmann equation contains not only the time variation of n_χ, but also the dilution of particles in an expanding volume.

Interpretation

The term $3Hn_\chi$ is purely kinematical. The collision operator contains the microphysics that can create or destroy particles.

Collision term for annihilation

For annihilations $\chi(p_1)\chi(p_2) \leftrightarrow f(p_3)f(p_4)$, the collision operator contains the invariant phase-space measure and the squared matrix element,

$$C[f_1] = \frac{1}{2E_1} \int \prod_{i=2}^{4} \frac{\mathrm{d}^3\mathbf{p}_i}{(2\pi)^3 2E_i} (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p_3 - p_4) \, |\mathcal{M}|^2 \times [f_3 f_4 - f_1 f_2] \, . \tag{10}$$

Here quantum-statistical factors have been omitted, as is appropriate for the Maxwell–Boltzmann approximation used in the standard relic-density calculation.

The number-density equation

Assuming kinetic equilibrium, negligible chemical potentials for the bath particles, and Maxwell–Boltzmann statistics, the integrated collision term takes the simple form

$$\dot{n}_\chi + 3Hn_\chi = -\langle \sigma v_{\rm rel} \rangle \left(n_\chi^2 - n_{\chi,\rm eq}^2 \right) . \tag{11}$$

This single equation contains both directions of the reaction. When $n_\chi > n_{\chi,\rm eq}$, annihilations dominate; when $n_\chi < n_{\chi,\rm eq}$, inverse reactions repopulate the species.

Equilibrium as a fixed point

For sufficiently rapid interactions, the solution is driven toward $n_\chi = n_{\chi,\rm eq}$.

Equilibrium number density

The modified Bessel functions that appear in relic-density calculations arise directly from integrating the Boltzmann factor over relativistic phase space. For a particle of mass m_χ in Maxwell–Boltzmann equilibrium,

$$n_\chi^{\rm eq} = g_\chi \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3} f_{\rm eq}(\mathbf{p}), \qquad f_{\rm eq}(\mathbf{p}) = \exp\left[-\sqrt{p^2 + m_\chi^2}/T\right]. \tag{12}$$

changing variables from momentum to energy gives

$$n_\chi^{\rm eq} = \frac{g_\chi}{2\pi^2} \int_{m_\chi}^{\infty} E\sqrt{E^2 - m_\chi^2}\, e^{-E/T}\, \mathrm{d}E. \tag{13}$$

This integral has the exact closed form

$$n_\chi^{\rm eq} = \frac{g_\chi m_\chi^2 T}{2\pi^2} K_2\!\left(\frac{m_\chi}{T}\right) \simeq g_\chi \left(\frac{m_\chi T}{2\pi}\right)^{3/2} e^{-m_\chi/T}. \tag{14}$$

The exponential factor is the essential origin of freeze-out.

Origin of the Bessel function K_1

In the thermal average of the annihilation rate, the relevant variable is instead the total energy of the pair,

$$\sqrt{s} = E_1 + E_2, \tag{15}$$

so that one encounters integrals weighted by $e^{-\sqrt{s}/T}$. After reduction of the phase-space integrals, this leads to

$$\langle \sigma v \rangle \propto \int \mathrm{d}s\, (\cdots)\, K_1\left(\frac{\sqrt{s}}{T}\right). \tag{16}$$

Interpretation

$K_2(m/T)$ describes the thermal weight of a single particle, while $K_1(\sqrt{s}/T)$ encodes the thermal distribution of the center-of-mass energy of an annihilating pair. Both arise unavoidably from relativistic phase-space integrals with the Boltzmann factor.

Thermal average

The thermally averaged annihilation rate is defined by

$$\langle \sigma v_{\rm rel}\rangle = \frac{g_\chi^2}{n_{\chi,\rm eq}^2}\int \frac{d^3\mathbf{p}_1}{(2\pi)^3}\frac{d^3\mathbf{p}_2}{(2\pi)^3} f_{\rm eq}(E_1) f_{\rm eq}(E_2)\,\sigma v_{\rm rel}. \tag{17}$$

The microscopic annihilation cross section is thus averaged over the thermal distribution of the incoming particles.

Physical meaning

The relic abundance is not controlled by a cross section at one fixed energy, but by an average over the energies populated at freeze-out.

Freeze-out condition

At early times the annihilation rate per particle is

$$\Gamma_{\rm ann} = n_{\chi,\rm eq}\langle\sigma v_{\rm rel}\rangle. \tag{18}$$

Chemical equilibrium is maintained as long as

$$\Gamma_{\rm ann} \gg H. \tag{19}$$

Freeze-out occurs when the two rates become comparable,

$$n_{\chi,\rm eq}(T_f)\langle\sigma v_{\rm rel}\rangle \simeq H(T_f). \tag{20}$$

Qualitative consequence

A larger annihilation cross section keeps the species in equilibrium longer and leaves a smaller final abundance.

Yield and entropy conservation

It is useful to divide the number density by the entropy density,

$$Y \equiv \frac{n_\chi}{s}, \qquad s = \frac{2\pi^2}{45} g_{*s} T^3. \tag{21}$$

For adiabatic expansion,

$$\dot{s} + 3Hs = 0, \tag{22}$$

so that the dilution terms cancel in the Boltzmann equation.

One obtains

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = -s\langle\sigma v_{\mathrm{rel}}\rangle \left(Y^2 - Y_{\mathrm{eq}}^2\right). \tag{23}$$

The yield is the natural variable because it measures the abundance per comoving entropy.

Inverse temperature variable

Introducing

$$x \equiv \frac{m_\chi}{T}, \tag{24}$$

and assuming that g_{*s} varies slowly, radiation domination gives

$$H = 1.66\, g_*^{1/2} \frac{T^2}{m_{\rm Pl}}. \tag{25}$$

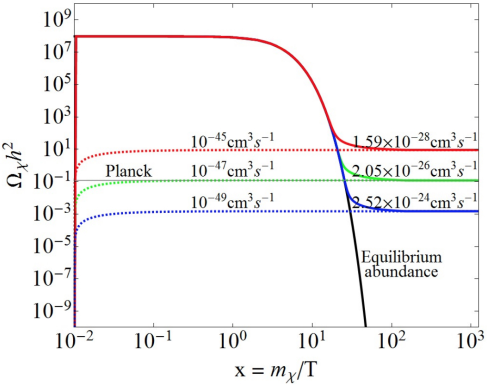
The Boltzmann equation becomes

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = -\frac{s\langle\sigma v_{\rm rel}\rangle}{Hx}\left(Y^2 - Y_{\rm eq}^2\right). \tag{26}$$

Meaning of x

Increasing x corresponds to the cooling of the Universe. Freeze-out occurs at $x_f = m_\chi/T_f \gg 1$.

Freeze-out and freeze-in: relic abundance



DM relic abundance $\Omega_\chi h^2$ vs $x = m_\chi/T$ for different $\langle \sigma v_{\rm rel}\rangle$. For sufficiently large interaction rates, $\langle \sigma v_{\rm rel}\rangle \sim 10^{-26}\,\mathrm{cm}^3\,\mathrm{s}^{-1}$, the particle remains in thermal equilibrium and subsequently freezes out, yielding a relic abundance inversely proportional to the cross section. For much smaller interaction rates, $\langle \sigma v_{\rm rel}\rangle \ll 10^{-26}\,\mathrm{cm}^3\,\mathrm{s}^{-1}$, the particle never thermalizes and is gradually produced through freeze-in. The black curve shows the equilibrium abundance.

Equilibrium yield

Using the exact equilibrium density, the equilibrium yield is

$$Y_{\rm eq} = \frac{45}{4\pi^4}\frac{g_\chi}{g_{*s}}x^2 K_2(x). \tag{27}$$

For $x \gg 1$, this becomes

$$Y_{\rm eq} \simeq \frac{45}{4\sqrt{2}\,\pi^{7/2}}\frac{g_\chi}{g_{*s}}x^{3/2}e^{-x} \simeq 0.145\frac{g_\chi}{g_{*s}}x^{3/2}e^{-x}. \tag{28}$$

The equilibrium abundance decreases exponentially, whereas the expansion rate decreases only as a power of temperature. This mismatch forces departure from equilibrium.

Estimate of the freeze-out temperature

Using the non-relativistic expression for $n_{\chi,\mathrm{eq}}$ in the condition $\Gamma_{\mathrm{ann}} \simeq H$, one obtains

$$g_\chi \left(\frac{m_\chi^2}{2\pi x_f}\right)^{3/2} e^{-x_f} \langle \sigma v_{\mathrm{rel}} \rangle \simeq 1.66\, g_*^{1/2} \frac{m_\chi^2}{x_f^2 m_{\mathrm{Pl}}} \,. \tag{29}$$

Solving for x_f gives the implicit equation

$$x_f = \ln \left[\frac{0.038\, g_\chi\, m_{\mathrm{Pl}}\, m_\chi \langle \sigma v_{\mathrm{rel}} \rangle}{g_*^{1/2} x_f^{1/2}}\right] . \tag{30}$$

For weak-scale masses and cross sections, the logarithm is large but slowly varying, giving

$$x_f \simeq 20\text{–}30. \tag{31}$$

Late-time solution

After freeze-out, the equilibrium density is negligible. The yield then satisfies approximately

$$\frac{\mathrm{d}Y}{\mathrm{d}x} \simeq -\frac{s\langle\sigma v_{\rm rel}\rangle}{Hx}Y^2. \tag{32}$$

Integrating from x_f to late times gives

$$\frac{1}{Y_\infty} = \frac{1}{Y(x_f)} + \int_{x_f}^{\infty} \frac{s\langle\sigma v_{\rm rel}\rangle}{Hx}\,\mathrm{d}x. \tag{33}$$

The first term is usually subdominant, so the final abundance is controlled by the accumulated annihilation probability after freeze-out.

Relic density

The present density parameter follows from

$$\Omega_\chi = \frac{m_\chi s_0 Y_\infty}{\rho_{c0}}, \tag{34}$$

where s_0 is the present entropy density. For a smoothly varying annihilation cross section, this leads to the standard estimate

$$\Omega_\chi h^2 \simeq \frac{3 \times 10^{-26}\,\mathrm{cm}^3\,\mathrm{s}^{-1}}{\langle \sigma v_{\mathrm{rel}} \rangle}. \tag{35}$$

The WIMP miracle

An annihilation cross section of the order expected for weak interactions naturally gives a relic density close to the observed dark matter abundance.

Invariant formulation

Gondolo and Gelmini showed that the thermal average can be expressed in terms of Lorentz-invariant quantities and reduced to a one-dimensional integral over

$$s = (p_1 + p_2)^2. \tag{36}$$

The appropriate velocity variable is the Møller velocity,

$$v_{\mathrm{Møl}} = \frac{\sqrt{(p_1 \cdot p_2)^2 - m_\chi^4}}{E_1 E_2}, \tag{37}$$

for which the product $\sigma v_{\mathrm{Møl}}$ is invariant.

Since the phase-space measure and $|\mathcal{M}|^2$ are also invariant, the collision integral may be evaluated in the center-of-mass frame. Introducing the variables

$$s = (p_1 + p_2)^2, \qquad E_+ = E_1 + E_2, \qquad E_- = E_1 - E_2, \tag{38}$$

one can perform the angular integrations and reduce the multidimensional phase-space integral to a single integral over s.

Gondolo–Gelmini thermal average

For Maxwell–Boltzmann statistics, the exact thermal average is

$$\langle \sigma v_{\rm rel}\rangle = \frac{1}{8m_\chi^4 T K_2^2(m_\chi/T)} \int_{4m_\chi^2}^{\infty} \mathrm{d}s\, \sigma(s)(s-4m_\chi^2)\sqrt{s}\, K_1\!\left(\frac{\sqrt{s}}{T}\right). \quad (39)$$

The modified Bessel functions arise from integrating the Boltzmann factor over relativistic phase space:

$$K_2(m_\chi/T) \text{ from } n_\chi^{\rm eq}, \qquad K_1(\sqrt{s}/T) \text{ from pair energy weighting.} \quad (40)$$

Importance

A multidimensional phase-space integral has been reduced to a single integral over an invariant energy variable. This expression remains valid when $\sigma(s)$ varies rapidly, as near resonances or kinematic thresholds.

In such situations, the expansion $\sigma v = a + bv^2 + \cdots$ fails because the thermal distribution samples values of s far from threshold.

Limits of the standard freeze-out approximation

The standard analytic treatment of thermal freeze-out assumes that the annihilation cross section varies smoothly near threshold and is expanded as

$$\sigma v_{\mathrm{rel}} \simeq a + bv^2 + \cdots \,. \tag{41}$$

This approximation is often adequate, but it can fail when the annihilation rate is controlled by dynamics that introduce a strong energy dependence in $\sigma(s)$.

A systematic analysis was given by Griest and Seckel, who identified three important exceptions: resonant annihilation, threshold effects, and coannihilation with nearly degenerate states.

Interpretive remark

The non-relativistic expansion assumes that the annihilation process probes only a narrow range of energies near threshold. Whenever the cross section varies significantly over the thermally accessible range, this approximation ceases to be reliable.

Resonant annihilation

A particularly important situation arises when annihilation proceeds through an s-channel mediator R with

$$2m_\chi \simeq m_R. \tag{42}$$

For non-relativistic particles one has $s \simeq 4m_\chi^2$, so the center-of-mass energy lies close to the mediator pole.
Near the resonance, the cross section takes the Breit–Wigner form

$$\sigma(s) \propto \frac{1}{(s - m_R^2)^2 + m_R^2 \Gamma_R^2}, \tag{43}$$

which varies rapidly with s in a narrow region of width $\sim m_R \Gamma_R$.

During freeze-out the particles have a thermal velocity distribution, so s is not fixed but spread around $4m_\chi^2$. As a result, even when

$$2m_\chi < m_R, \tag{44}$$

the high-energy tail of the distribution can reach the resonance and significantly enhance the thermally averaged rate $\langle \sigma v \rangle$.

Threshold effects

A second failure of the smooth velocity expansion occurs when a new annihilation channel opens near freeze-out. If

$$2m_\chi \simeq 2m_f\,, \tag{45}$$

then the available final-state phase space changes rapidly. Close to threshold, one typically encounters factors of the form

$$\sigma v \propto \sqrt{1 - \frac{m_f^2}{m_\chi^2}}\,. \tag{46}$$

The cross section is then non-analytic at the threshold, and the relic density becomes sensitive to the precise mass spectrum.

Coannihilation: physical setting

Suppose there are several particles χ_i, with the lightest state $\chi_1 \equiv \chi$ stable, and with masses close to m_χ. Define

$$\Delta_i \equiv \frac{m_i - m_\chi}{m_\chi}. \tag{47}$$

If

$$\Delta_i \lesssim \frac{T_f}{m_\chi} \simeq \frac{1}{x_f}, \tag{48}$$

then the heavier states are still present during freeze-out. Rapid scattering reactions keep the different species in chemical equilibrium with one another.

Consequence

One must track the total abundance $n = \sum_i n_i$, not only the abundance of the lightest particle.

Effective coannihilation cross section

The coupled Boltzmann equations reduce to

$$\dot{n} + 3Hn = -\langle \sigma_{\rm eff} v \rangle \left(n^2 - n_{\rm eq}^2\right), \tag{49}$$

where

$$\langle \sigma_{\rm eff} v \rangle = \sum_{ij} \langle \sigma_{ij} v \rangle \frac{n_i^{\rm eq} n_j^{\rm eq}}{(n^{\rm eq})^2}. \tag{50}$$

For non-relativistic species,

$$n_i^{\rm eq} = g_i \left(\frac{m_i T}{2\pi}\right)^{3/2} e^{-m_i/T}. \tag{51}$$

Thus the heavier states are Boltzmann suppressed but can still dominate the effective rate if their annihilation cross sections are large.

Factoring out the lightest mass and using $x = m_\chi/T$, one obtains

$$\langle\sigma_{\rm eff} v\rangle = \sum_{ij} \langle\sigma_{ij} v\rangle \frac{g_i g_j (1+\Delta_i)^{3/2}(1+\Delta_j)^{3/2} e^{-x(\Delta_i+\Delta_j)}}{g_{\rm eff}^2}, \tag{52}$$

with

$$g_{\rm eff} = \sum_k g_k (1+\Delta_k)^{3/2} e^{-x\Delta_k}. \tag{53}$$

Even a small mass splitting can be important because $x_f \simeq 20$–30. A splitting of only a few percent may noticeably suppress, but not necessarily eliminate, the heavier partner.

Effect on the relic abundance

If a heavier state has stronger gauge interactions than the lightest particle, its contribution to $\langle\sigma_{\rm eff} v\rangle$ may compensate for the Boltzmann suppression.

In that case,

$$\langle\sigma_{\rm eff} v\rangle > \langle\sigma_{\chi\chi} v\rangle, \tag{54}$$

so that the relic abundance is reduced,

$$\Omega_\chi h^2 \propto \frac{1}{\langle\sigma_{\rm eff} v\rangle}. \tag{55}$$

Important point

The final density of the stable particle can be controlled by reactions involving particles that no longer exist today.

The opposite limit

Thermal freeze-out assumes that the dark matter particle was initially in equilibrium with the plasma. This need not be true if its coupling to the Standard Model is extremely small.

In the freeze-in regime, the abundance of χ is always much smaller than its equilibrium value,

$$n_\chi \ll n_{\chi,\mathrm{eq}}. \tag{56}$$

The dark sector is then slowly populated by rare processes in the thermal bath, rather than depleted from an initially thermal abundance.

Freeze-in

The relic abundance grows from an initially negligible value and never reaches thermal equilibrium.

Boltzmann equation for freeze-in

If the dominant process is

$$i + j \to \chi + X, \tag{57}$$

with i and j in thermal equilibrium, backreaction from the dark sector can be neglected. The number-density equation becomes

$$\dot{n}_\chi + 3Hn_\chi \simeq \sum_{i,j} \langle \sigma_{ij\to\chi X} v_{\rm rel} \rangle n_i^{\rm eq} n_j^{\rm eq}. \tag{58}$$

The right-hand side is now a source term. It increases the dark matter abundance instead of driving it toward equilibrium by detailed balance.

Freeze-in yield

In terms of the yield $Y_\chi = n_\chi/s$ and the inverse temperature variable $x = m_\chi/T$, one finds

$$\frac{\mathrm{d}Y_\chi}{\mathrm{d}x} \simeq \frac{s}{Hx}\sum_{i,j}\langle\sigma v_{\rm rel}\rangle Y_i^{\rm eq} Y_j^{\rm eq}. \tag{59}$$

Integrating from reheating to late times gives

$$Y_\chi(\infty) \simeq \int_{x_{\rm RH}}^{\infty} \frac{s}{Hx}\sum_{i,j}\langle\sigma v_{\rm rel}\rangle Y_i^{\rm eq} Y_j^{\rm eq}\,\mathrm{d}x, \tag{60}$$

where $x_{\rm RH} = m_\chi/T_{\rm RH}$.

Relic abundance from freeze-in

The present density is again obtained from the asymptotic yield,

$$\Omega_\chi = \frac{m_\chi s_0}{\rho_{c0}} Y_\chi(\infty). \tag{61}$$

But the dependence on the interaction strength is opposite to the freeze-out case.

For freeze-out,

$$\Omega_\chi h^2 \propto \frac{1}{\langle \sigma v_{\rm rel} \rangle}, \tag{62}$$

whereas for freeze-in,

$$\Omega_\chi h^2 \propto \langle \sigma v_{\rm rel} \rangle. \tag{63}$$

FIMP regime

Particles produced by freeze-in are often called feebly interacting massive particles, because the coupling must be too small to thermalize the dark sector.

Freeze-out versus freeze-in

The two mechanisms differ in their initial conditions and in the role played by the interaction strength. Freeze-out begins in equilibrium and ends when interactions become inefficient. Freeze-in begins with negligible abundance and ends when production from the bath shuts off.

Mechanism	Initial state	Scaling of abundance
Freeze-out	thermal equilibrium	$\Omega_\chi h^2 \propto \langle \sigma v_{\rm rel} \rangle^{-1}$
Freeze-in	negligible abundance	$\Omega_\chi h^2 \propto \langle \sigma v_{\rm rel} \rangle$

Evolution of the abundance

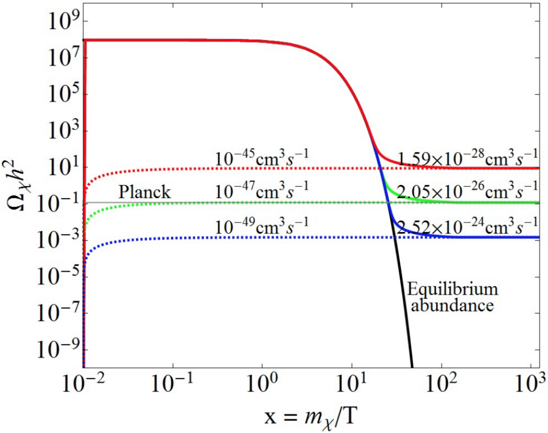


Figure: Qualitative comparison between freeze-out and freeze-in production. In freeze-out the abundance tracks equilibrium before departing from it; in freeze-in the abundance is accumulated slowly from rare bath reactions and remains below equilibrium at all times.

What has been established

The WIMP relic-density calculation reduces a microscopic annihilation process to a cosmological prediction. The essential equation is

$$\dot{n}_\chi + 3Hn_\chi = -\langle \sigma v_{\rm rel} \rangle \left(n_\chi^2 - n_{\chi,{\rm eq}}^2 \right) , \tag{64}$$

whose solution shows how an initially thermal population freezes out once $\Gamma_{\rm ann} \simeq H$.

The simple estimate

$$\Omega_\chi h^2 \simeq \frac{3 \times 10^{-26}\,{\rm cm}^3\,{\rm s}^{-1}}{\langle \sigma v_{\rm rel} \rangle} \tag{65}$$

explains why weak-scale interactions are cosmologically interesting.

The strong CP problem

Quantum chromodynamics allows for a CP-violating term in the Lagrangian,

$$\mathcal{L}_\theta = \frac{\alpha_s}{8\pi}\,\theta\,G^a_{\mu\nu}\tilde{G}^{a\,\mu\nu}. \tag{66}$$

This term induces an electric dipole moment for the neutron,

$$d_n \sim 10^{-16}\,\theta\,e\,\mathrm{cm}, \tag{67}$$

while experiments require

$$|\theta| \lesssim 10^{-10}. \tag{68}$$

Puzzle

The parameter θ is dimensionless and a priori of order unity. Its extreme smallness is not protected by any symmetry of the QCD Lagrangian.

Peccei–Quinn mechanism

The Peccei–Quinn mechanism promotes the CP-violating parameter to a dynamical field by introducing a global $U(1)_{\mathrm{PQ}}$ symmetry, spontaneously broken at a scale f_φ.

The associated pseudo-Nambu–Goldstone boson is the axion field $\varphi(x)$, which replaces

$$\bar{\theta} \;\to\; \bar{\theta} + \frac{\varphi(x)}{f_\varphi}. \tag{69}$$

Non-perturbative QCD effects generate a periodic potential,

$$V(\varphi) = m_\varphi^2 f_\varphi^2 \left[1 - \cos\left(\frac{\varphi}{f_\varphi}\right)\right]. \tag{70}$$

Key idea

The axion dynamically relaxes to the minimum of the potential, driving the effective CP-violating parameter to zero.

Axion mass and couplings

Expanding the potential around its minimum yields

$$m_\varphi^2 = \frac{\chi_{\rm QCD}}{f_\varphi^2}, \tag{71}$$

with lattice results giving

$$m_\varphi \simeq 5.7\,\mu\text{eV}\left(\frac{10^{12}\,\text{GeV}}{f_\varphi}\right). \tag{72}$$

Axions couple to photons through

$$\mathcal{L}_{\varphi\gamma\gamma} = -\frac{1}{4}g_{\varphi\gamma\gamma}\varphi F_{\mu\nu}\tilde{F}^{\mu\nu}. \tag{73}$$

Scaling

All axion interactions are suppressed by f_φ^{-1}, so larger f_φ implies lighter and more weakly coupled particles.

Vacuum misalignment mechanism

In an expanding Universe, the homogeneous axion field obeys

$$\ddot{\varphi} + 3H\dot{\varphi} + m_\varphi^2(T)\varphi = 0. \tag{74}$$

At early times $H \gg m_\varphi$, and the field is frozen at

$$\varphi_i = \theta_i f_\varphi. \tag{75}$$

Oscillations begin when

$$3H(T_{\rm osc}) \simeq m_\varphi(T_{\rm osc}). \tag{76}$$

Physical picture

The axion behaves as an overdamped field at early times and as a harmonic oscillator once the Hubble friction becomes subdominant.

Axion condensate as dark matter

After the onset of oscillations, the axion field evolves approximately as

$$\varphi(t) \simeq \varphi_i\, a^{-3/2} \cos(m_\varphi t)\,, \tag{77}$$

with energy density

$$\rho_\varphi(t) \simeq \frac{1}{2} m_\varphi^2 \varphi_{\rm osc}^2 \left(\frac{a_{\rm osc}}{a(t)}\right)^3 . \tag{78}$$

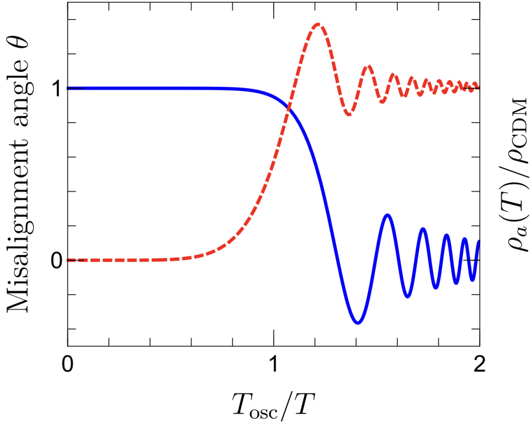
Averaged over oscillations, the pressure vanishes and the energy density redshifts as

$$\rho_\varphi \propto a^{-3}. \tag{79}$$

Interpretation

A coherently oscillating scalar field behaves as a pressureless fluid, providing a non-thermal realization of cold dark matter.

Onset of axion oscillations



The evolution of the axion field is controlled by the competition between Hubble friction and the axion mass. At early times, when $H \gg m_\varphi$, the field is overdamped and remains frozen at its initial value θ_i (blue curve). As the Universe expands, the Hubble rate decreases and oscillations begin when $3H(T_{\rm osc}) \simeq m_\varphi(T_{\rm osc})$.

Axion relic abundance

Evolving the energy density from the onset of oscillations to today yields

$$\Omega_\varphi h^2 \approx 0.12\,\theta_i^2\,F(\theta_i)\left(\frac{f_\varphi}{5\times 10^{11}\,\mathrm{GeV}}\right)^{1.19}. \tag{80}$$

The nontrivial exponent arises from the temperature dependence of the axion mass near the QCD transition.

Key result

For $\theta_i = \mathcal{O}(1)$, the observed dark matter abundance is reproduced for

$$f_\varphi \sim 10^{11}\text{–}10^{12}\,\mathrm{GeV}. \tag{81}$$

Axions produced in this way never reach thermal equilibrium and constitute a purely non-thermal dark matter component.

Two cosmological histories

The cosmology of the axion depends crucially on when the PQ symmetry is broken relative to inflation.

If PQ symmetry is broken before inflation and not restored afterwards, our observable Universe inherits a single homogeneous value of θ_i, up to inflationary quantum fluctuations.

If PQ symmetry is broken after inflation, different Hubble patches choose uncorrelated values of the axion field. Topological defects form, and the relic abundance receives additional contributions.

Important distinction

Pre-inflationary axions are constrained mainly by isocurvature perturbations; post-inflationary axions are accompanied by strings, domain walls, and small-scale density fluctuations.

Global axion strings

In the post-inflationary scenario, the spontaneous breaking of the global $U(1)_{\rm PQ}$ symmetry produces global cosmic strings. Around such a string, the axion field winds by 2π.

The string tension is approximately

$$\mu \sim \pi f_\varphi^2 \ln\left(\frac{L}{\delta}\right) \sim \pi f_\varphi^2 \ln\left(\frac{f_\varphi}{H}\right), \tag{82}$$

where $\delta \sim f_\varphi^{-1}$ is the core size and L is an infrared cutoff of order the Hubble radius.

Interpretive remark

The logarithm reflects the fact that a global string stores energy in the long-range axion field, not only in its microscopic core.

Scaling string network

After formation, the string network evolves toward a scaling regime in which the number of strings per Hubble volume remains roughly constant.

Maintaining this scaling solution requires the network to lose energy. For axion strings, the dominant loss mechanism is radiation into axions, with momenta typically of order the Hubble scale.

$$k_{\mathrm{emitted}} \sim H. \tag{83}$$

The emitted axions are highly non-thermal and contribute to the cold dark matter abundance after they become non-relativistic.

Open issue

The spectrum of axions emitted by the string network is one of the main uncertainties in the post-inflationary relic abundance.

Domain walls and discrete vacua

When non-perturbative QCD effects turn on, the continuous Peccei–Quinn symmetry is explicitly broken to a discrete subgroup. The axion potential develops multiple degenerate minima,

$$\theta = \frac{2\pi k}{N_{\rm DW}}, \qquad k = 0, \dots, N_{\rm DW} - 1. \tag{84}$$

In the post-inflationary scenario, different Hubble patches settle into different vacua. As a result, **domain walls** form at the boundaries between regions with different values of θ.

Domain wall number

The integer $N_{\rm DW}$ counts the number of distinct vacua.

String–wall network

Domain walls do not form in isolation. They attach to the network of global axion strings produced when the PQ symmetry is spontaneously broken.

Each string is connected to $N_{\rm DW}$ domain walls, forming a coupled string–wall system that evolves dynamically.

Physical picture

Strings correspond to locations where the axion field winds by 2π, while domain walls interpolate between adjacent minima of the potential. The combined network stores a significant fraction of the energy density of the Universe.

The domain wall problem

The cosmological fate of the string–wall network depends crucially on the domain wall number $N_{\rm DW}$.

Case $N_{\rm DW} = 1$:
Each string is attached to a single domain wall. The wall tension pulls on the string, making the network unstable. The system collapses efficiently and decays into a population of cold axions.

Case $N_{\rm DW} > 1$:
Each string is attached to multiple domain walls. The forces can balance, leading to a stable network. Its energy density redshifts too slowly and would eventually dominate the Universe.

Cosmological constraint

A viable QCD axion model must either have $N_{\rm DW} = 1$, or include a small explicit PQ-breaking term that lifts the degeneracy and causes the walls to decay.

Total abundance in the post-inflationary scenario

In the post-inflationary scenario, the axion relic abundance receives several contributions,

$$\Omega_\varphi = \Omega_{\rm misalignment} + \Omega_{\rm strings} + \Omega_{\rm walls}. \tag{85}$$

A commonly quoted estimate for the total abundance is

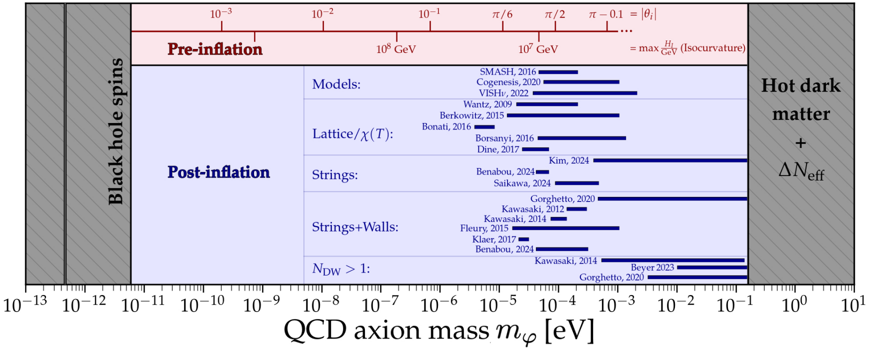
$$\Omega_\varphi h^2 \simeq 0.12 \left(\frac{f_\varphi}{\mathcal{O}(10^{11})\,\mathrm{GeV}}\right)^{1.19}, \tag{86}$$

although the normalization remains uncertain at the level of a factor of a few.

Origin of the uncertainty

Simulations must extrapolate over the enormous hierarchy between the Hubble scale and the microscopic string core scale.

QCD axion parameter space



The QCD axion mass is inversely related to the PQ scale, so different cosmological histories select different mass ranges. In the pre-inflationary scenario, the initial misalignment angle is homogeneous across the observable Universe and CMB isocurvature bounds constrain the inflationary scale. In the post-inflationary scenario, strings and domain walls contribute to the relic abundance, leading to predictions in the region around $m_\varphi \sim 10^{-5}$–$10^{-3}\,\mathrm{eV}$, with residual uncertainty from simulations.

Initial fluctuations after inflation

If the PQ symmetry is broken after inflation, the axion field takes uncorrelated values in different Hubble patches,

$$\varphi_i = \theta_i f_\varphi, \tag{87}$$

with $\theta_i = \mathcal{O}(1)$ varying randomly in space.

When the field begins to oscillate, the axion energy density is proportional to θ_i^2. Spatial variations of the initial angle translate into order-one density fluctuations on the horizon scale.

On large scales, the resulting power spectrum is approximately white noise,

$$P_{\delta_a}(k) \simeq \text{const.} \qquad (k \lesssim k_{\rm osc}), \tag{88}$$

where

$$k_{\rm osc} = a_{\rm osc} H_{\rm osc} \tag{89}$$

corresponds to the horizon scale at the onset of oscillations.

Axion miniclusters

Large axion density fluctuations can become nonlinear around horizon entry and collapse gravitationally into compact bound objects known as axion miniclusters.

The characteristic mass is set by the axion mass contained in the Hubble volume at the onset of oscillations,

$$M_{\rm mc} \sim \frac{4\pi}{3}\rho_a H^{-3}\Big|_{T_{\rm osc}} \sim 10^{-12}\text{–}10^{-10}\, M_\odot. \quad (90)$$

The corresponding comoving scale is

$$k_{\rm mc} \sim a_{\rm osc} H_{\rm osc} \sim 10^6\text{–}10^7\,{\rm Mpc}^{-1}. \quad (91)$$

Physical scale

Axion miniclusters probe structure formation on scales enormously smaller than those relevant for galaxies.

Small-scale power spectrum

The post-inflationary axion scenario predicts a distinctive matter power spectrum. On scales larger than the horizon at the onset of oscillations, fluctuations inherit the white-noise spectrum of the uncorrelated misalignment field.

On smaller scales, correlations are absent and the spectrum is suppressed. The result is a feature peaked around

$$k \sim k_{\rm mc}. \tag{92}$$

The corresponding physical length scale is approximately

$$\lambda_{\rm mc} \sim k_{\rm mc}^{-1} \sim 10^{-7}\text{–}10^{-6}\,{\rm pc}. \tag{93}$$

Contrast with standard CDM

Instead of purely hierarchical formation, dense axion structures may form very early on extremely small scales.

Axion stars

Miniclusters may undergo further gravitational evolution, relaxation, and collapse. In their densest regions, axions can form self-gravitating Bose condensates known as axion stars.

These objects are stabilized by gradient pressure on the dilute branch, while self-interactions become important for denser configurations. Their detailed properties depend on the axion mass and self-coupling.

Possible signatures

Miniclusters and axion stars can enhance local direct-detection rates, produce microlensing events, and affect pulsar timing or radio signals in dense environments.

Inflationary fluctuations of the axion

If PQ symmetry is broken during inflation and not restored afterwards, the axion is light during inflation and acquires quantum fluctuations on super-horizon scales.

Here, H_I is the Hubble scale during inflation, evaluated when CMB modes exit the horizon. For $m_\varphi \ll H_I$, the field fluctuation is

$$\delta\varphi \simeq \frac{H_I}{2\pi}. \tag{94}$$

In terms of the misalignment angle,

$$\delta\theta \simeq \frac{H_I}{2\pi f_\varphi}. \tag{95}$$

Important distinction

These fluctuations perturb the axion density without necessarily perturbing the radiation density in the same way.

From angle fluctuations to density fluctuations

For small initial angle, the axion energy density produced by misalignment scales as

$$\rho_\varphi \propto \theta_i^2. \tag{96}$$

A fluctuation in the initial angle produces

$$\frac{\delta n_\varphi}{n_\varphi} \simeq 2\frac{\delta\theta}{\theta_i}. \tag{97}$$

These perturbations are isocurvature perturbations because the axion number density fluctuates relative to the radiation bath.

$$S_a \equiv \frac{\delta n_\varphi}{n_\varphi} - \frac{3}{4}\frac{\delta\rho_\gamma}{\rho_\gamma}. \tag{98}$$

Interpretation

Isocurvature modes initially redistribute energy among components rather than perturbing the total energy density.

Isocurvature power

The dimensionless power spectrum of axion isocurvature perturbations is approximately

$$\mathcal{P}_S(k) \simeq \left(\frac{H_I}{\pi f_\varphi \theta_i}\right)^2 \left(\frac{\Omega_\varphi}{\Omega_{\rm DM}}\right)^2 . \tag{99}$$

The factor $\Omega_\varphi/\Omega_{\rm DM}$ appears because the observable isocurvature signal is diluted if axions are only a subcomponent of dark matter.

Cosmological analyses often parameterize the isocurvature fraction as

$$\beta_{\rm iso} = \frac{\mathcal{P}_S}{\mathcal{P}_\mathcal{R} + \mathcal{P}_S}. \tag{100}$$

CMB signature

Isocurvature perturbations shift the acoustic pattern relative to adiabatic initial conditions and are tightly constrained.

Constraint on the inflationary scale

CMB measurements require the isocurvature fraction to be small,

$$\beta_{\rm iso} \lesssim 0.04 \tag{101}$$

at the usual CMB pivot scale.

If axions make up a significant fraction of the dark matter and PQ symmetry is broken during inflation, this implies an upper bound on the inflationary Hubble rate. A representative estimate is

$$H_I \lesssim 10^7\,{\rm GeV}\left(\frac{f_\varphi}{10^{11}\,{\rm GeV}}\right)^{0.4}. \tag{102}$$

Conceptual point

Axion dark matter links a low-energy strong CP solution to the energy scale of inflation.

How isocurvature constraints can be avoided

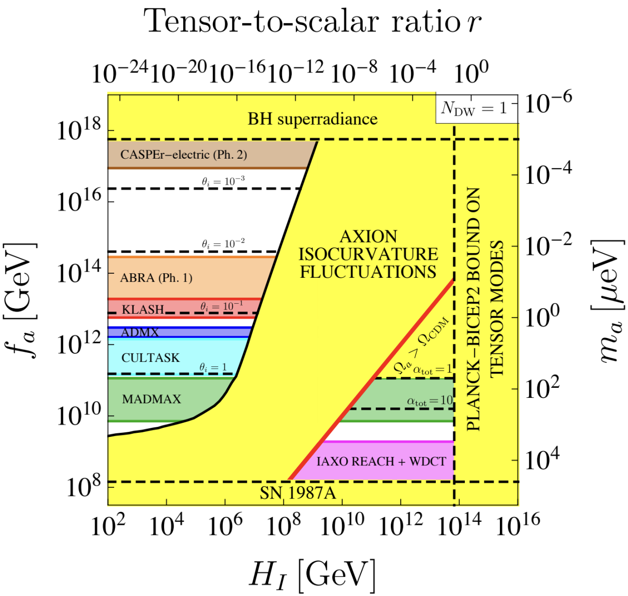
The strength of the isocurvature bound depends on cosmological history. If PQ symmetry is restored after inflation, the large-scale inflationary axion fluctuation is erased.

The constraint is also weakened if axions constitute only a fraction of the dark matter, or if the effective axion decay constant during inflation differs from its value today.

Interpretive remark

Isocurvature is not an unavoidable prediction of all axion models; it is a sharp prediction of the pre-inflationary PQ-breaking scenario.

Axion Parameter Space



Thermal production

In addition to non-thermal production by misalignment and defects, axions can be generated through interactions with the primordial plasma.

If the interaction rate exceeds the Hubble rate, axions thermalize and later decouple as a relativistic relic. If the interaction rate remains below the Hubble rate, axions are produced out of equilibrium through freeze-in.

$$\Gamma_\varphi \approx H\,, \tag{103}$$

separates the thermal and non-thermal regimes.

Physical distinction

Thermal axions remember the temperature of the plasma; freeze-in axions remember the rare processes that produced them.

Production channels

The dominant production processes depend on the axion model and on temperature. In hadronic, KSVZ-type models, high-temperature production involves quarks and gluons,

$$g + g \leftrightarrow g + \varphi, \qquad q + g \leftrightarrow q + \varphi. \tag{104}$$

Below the QCD phase transition, hadronic reactions involving pions become important,

$$\pi + \pi \leftrightarrow \pi + \varphi. \tag{105}$$

In DFSZ-type models, additional couplings to charged leptons can also contribute to production and thermalization.

Model dependence

The cosmological abundance of thermal axions depends not only on f_φ, but also on the axion couplings to quarks, gluons, pions, and leptons.

Decoupling temperature

Parametrically, at high temperature the axion interaction rate scales as

$$\Gamma_\varphi \sim \alpha_s^3 \frac{T^3}{f_\varphi^2}, \tag{106}$$

up to model-dependent coefficients. During radiation domination,

$$H = 1.66\, g_*^{1/2} \frac{T^2}{m_{\rm Pl}}. \tag{107}$$

Equating these rates gives an estimate of the decoupling temperature,

$$T_{\rm dec} \sim \frac{f_\varphi^2}{1.66\, g_*^{1/2} \alpha_s^3 m_{\rm Pl}}. \tag{108}$$

Scaling

Larger f_φ means weaker interactions, earlier decoupling, and a colder thermal axion population today.

Thermal axions as dark radiation

If axions decouple while relativistic, they contribute to the radiation density as an additional light species,

$$\rho_\varphi = \frac{\pi^2}{30} g_\varphi T_\varphi^4, \tag{109}$$

with $g_\varphi = 1$ for a pseudoscalar.
After decoupling, entropy release into the photon bath makes the axion temperature lower than the photon temperature,

$$\frac{T_\varphi}{T_\gamma} = \left(\frac{g_{*s}(T_\gamma)}{g_{*s}(T_{\rm dec})}\right)^{1/3} . \tag{110}$$

Interpretation

The earlier a species decouples, the less it is heated by later annihilations of Standard Model particles.

Contribution to N_{eff}

The radiation density is conventionally written as

$$\rho_r = \rho_\gamma \left[1 + \frac{7}{8} \left(\frac{4}{11} \right)^{4/3} N_{\text{eff}} \right], \tag{111}$$

with the Standard Model prediction $N_{\text{eff}}^{\text{SM}} \simeq 3.045$.
A relativistic thermal axion contributes

$$\Delta N_{\text{eff}} = \frac{4}{7} \left(\frac{T_\varphi}{T_\nu} \right)^4 . \tag{112}$$

For decoupling above the QCD phase transition, one typically finds

$$\Delta N_{\text{eff}} \sim 0.02\text{–}0.05, \tag{113}$$

while later decoupling can give larger values.

Thermal axions as hot dark matter

Once thermal axions become non-relativistic, they contribute as a hot dark matter component. Their abundance is analogous to that of massive neutrinos, with a temperature rescaling,

$$\Omega_\varphi h^2 \simeq \frac{m_\varphi}{94\,\mathrm{eV}} \left(\frac{T_\varphi}{T_\nu}\right)^3 . \tag{114}$$

Because hot relics free stream, they suppress the growth of structure on small scales. Large-scale structure and CMB observations require approximately

$$m_\varphi \lesssim \mathcal{O}(0.1)\,\mathrm{eV}. \tag{115}$$

Physical point

Thermal axions do not behave like the cold condensate from misalignment; if sufficiently light, they are radiation, and if massive enough, they are hot dark matter.

Freeze-in production of axions

If the axion interaction rate is always below the Hubble rate,

$$\Gamma_\varphi < H\,, \tag{116}$$

thermal equilibrium is never reached. Axions are then produced gradually through rare scatterings in the thermal bath, for example

$$\pi + \pi \to \pi + \varphi, \qquad q + g \to q + \varphi\,. \tag{117}$$

The freeze-in yield is obtained by integrating the production rate over the thermal history,

$$Y_\varphi \equiv \frac{n_\varphi}{s} \simeq \int_{T_0}^{T_R} \frac{\gamma_\varphi(T)}{s(T)H(T)}\,\frac{\mathrm{d}T}{T}, \tag{118}$$

where γ_φ is the reaction density producing axions.
For UV-dominated production, one often has parametrically

$$Y_\varphi \sim \frac{m_{\rm Pl}\,T_R}{f_\varphi^2}, \tag{119}$$

up to dimensionless coupling, phase-space, and g_* factors.

Several axion populations

The cosmological axion abundance can contain several components with different origins and phase-space distributions.

For large f_φ, thermal production is usually subdominant and the dominant component is cold, non-thermal axion dark matter from misalignment and possibly topological defects.

For smaller f_φ, thermal axions may contribute appreciably as dark radiation or hot dark matter. Freeze-in can add a further non-thermal population whose spectrum depends on the production channel.

Synthesis

The phrase "axion dark matter" can refer to a coherent condensate, axions radiated by defects, a thermal relic, or a freeze-in population. The observational consequences depend on which component dominates.

Summary: two paradigms for dark matter

- WIMPs are thermal relics whose abundance is set by freeze-out, $\Omega_\chi h^2 \propto \langle \sigma v \rangle^{-1}$.
- Freeze-in gives the opposite scaling, $\Omega_\chi h^2 \propto \langle \sigma v \rangle$, and never reaches equilibrium.
- Axions are non-thermal dark matter from coherent field oscillations, with additional contributions from defects depending on PQ history.
- Pre-inflationary axions are constrained by isocurvature; post-inflationary axions predict strings, walls, and miniclusters.