

Superheavy dark matter and stepped dark radiation for Hubble Tension

Zhijie Xu

zhijie.xu@pnnl.gov

zhijie.xu@pnnl.gov <https://orcid.org/0000-0003-0459-4531>

Article

Keywords: dark matter, dark radiation, Hubble tension

Posted Date: October 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8570078/v3>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Superheavy dark matter and stepped dark radiation for Hubble Tension

Zhijie (Jay) Xu,¹[★]

¹*Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory; Richland, WA 99354, USA*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The persistent Hubble tension motivates early-universe physics that reduces the sound horizon while remaining consistent with CMB and large-scale structure data. We present a framework in which the relic abundance of superheavy nonthermal dark matter (DM) X is set by gravity rather than thermal freeze-out. Several independent considerations point to a common mass scale $m_X \simeq 10^{12}$ GeV: a free-streaming analysis identifies it as the unique mass at which the free-streaming mass equals the particle mass, marking the onset of a Newtonian quantum-gravity regime; the gravity-induced bulk peculiar velocity in the radiation era is then of order km s^{-1} ; and UV/IR mixing together with the DM–DR drag independently select the same scale. These results apply equally to standard dark-matter cosmologies, independently of any overproduction. In the overproduction scenario, the initially overabundant X becomes ultra-cold as the Universe expands; the DM density becomes comparable to the radiation density at $t_X \simeq 10^{-6}$ s. This triggers gravitational collapse to a bound state in which quantum pressure balances gravity, setting a quantum-gravitational scale $r_X \simeq 10^{-13}$ m. Within these states the overlapping wavefunctions convert most of the overabundance into dark radiation (DR) at very low relative velocity—a “cold” freeze-out—leaving a relic $n_\infty = \gamma n_i$ with $\gamma \sim 10^{-9}$, hence $\Delta N_{\text{eff}} \simeq 0.4$ and a raised CMB-inferred H_0 . We realize this in a minimal dark sector built on the stepped WZDR dark-radiation framework: a renormalizable Yukawa generates the DM–DR drag with coupling $\lambda_X \sim 10^{-5}$, while a Planck-suppressed spurion portal induces a tiny coupling $y_K \sim 10^{-14}$. High-scale SUSY breaking $\sqrt{F} \simeq m_X \simeq 10^{12}$ GeV fixes the DM mass, and a no-scale supergravity embedding cancels the tree-level gravity-mediated soft mass, so the scalar acquires a subleading F^2/M_{Pl}^3 contribution naturally at the eV scale. The framework links a macroscopic cold freeze-out to microscopic couplings and suggests targets for CMB and structure surveys and future probes of dark-sector interactions.

Key words: dark matter; dark radiation; Hubble tension; Early universe;

1 INTRODUCTION

The “Hubble tension” is the persistent $\sim 5\sigma$ discrepancy between the H_0 inferred from the CMB within Λ CDM, and the H_0 measured locally with calibrated supernovae (Di Valentino et al. 2021). The tension is robust to changes in both data and methodology: replacing local SH0ES (Riess et al. 2022) with TRGB (Freedman 2021) or Planck with ACT or SPT, and cross-checking with BAO+BBN retains the discrepancy at a comparable significance. Among many proposed resolutions (Karwal & Kamionkowski 2016; Poulin et al. 2019; Lin et al. 2019; Smith et al. 2020), those that modify the sound horizon at recombination r_s by altering the pre-recombination expansion history are especially compelling: they reduce the sound horizon r_s before recombination, thereby raising the CMB-inferred H_0 .

Two early-time directions have been explored. Early Dark Energy (EDE) (Poulin et al. 2019) adds a few-percent energy component that peaks before recombination and then dilutes rapidly, shrinking r_s and lifting H_0 , but is tightly constrained by the CMB’s precision and by large-scale structure data. Dark radiation (DR) adds relativistic degrees of freedom (Pandey et al. 2020). Purely free streaming DR also lowers r_s but imprints phase shifts and damping tail changes that are tightly bound (“drag effect”) (Bashinsky & Seljak 2004; Follin

et al. 2015). These constraints motivate interacting dark radiation scenarios in which strong self-interactions suppress free streaming (Blinov & Marques-Tavares 2020; Buen-Abad et al. 2015; Bagherian et al. 2025). A particularly natural realization involves dark matter charged under a new non-Abelian dark gauge group, whose massless dark gluons contribute to the energy density as self-interacting dark radiation that also interacts with DM (Buen-Abad et al. 2015).

Motivated by the incompatibility between the substantial amounts of interacting radiation required and high- l polarization data, another interesting model is the “step” Wess–Zumino dark radiation (WZDR) (Aloni et al. 2022; Schöneberg et al. 2023). A massive scalar ϕ and a massless fermion ψ remain in equilibrium via a Yukawa coupling; Supersymmetry breaking gives ϕ a small mass m_ϕ . When the hidden-sector temperature drops below m_ϕ , ϕ becomes nonrelativistic and annihilates or decays to ψ , injecting entropy and producing a step-like rise in the DR energy density at a transition redshift z_t controlled by m_ϕ , which distinguishes the low- and high- l multipoles. With m_ϕ in the eV range, the step occurs around $z_t \approx 10^4$, where it lowers r_s without spoiling the tight CMB constraints. Because dark radiation self-interacts, the phase-shift constraints are softened, allowing an efficient lift of H_0 .

Beyond H_0 tension, galaxy and lensing measurements favor less clustering than Λ CDM calibrated to the CMB, the “S8 tension”. Coupling dark matter to the step dark radiation can provide early-time

[★] E-mail: zhijie.xu@pnnl.gov; zhijiexu@hotmail.com

drag that suppresses power on 10–20 Mpc scales without degrading CMB fits (Buen-Abad et al. 2018; Archidiacono et al. 2019). In the WZDR+ extension (Joseph et al. 2023), a Yukawa interaction between X and ϕ mediates X – ψ scattering that is active before equality and shuts off naturally when ϕ becomes Boltzmann suppressed, simultaneously improving H_0 and S8 in joint fits. A closely related model construction employs a “stepped” partially acoustic dark matter scenario: X is a complex scalar charged under a dark $U(1)$. The associated massless gauge boson and a light Dirac fermion ψ of mass m_ψ that is also charged under $U(1)$ constitute the DR fluid (Buen-Abad et al. 2023). As the temperature drops below m_ψ , ψ becomes nonrelativistic and annihilates/decays to dark gauge bosons, generating a step-like increase in DR. In both cases, a step-like enhancement of DR interacting with DM, mediated by t -channel exchange of ϕ or dark gauge bosons, introduces pre-equality drag and acoustic effects that reduce the clustering power at the targeted scales.

All these developments suggest a minimal dark sector: (i) prior to recombination, an extra dark-radiation component is populated by some unknown early Universe process, reducing r_s and lifting H_0 ; (ii) the dark radiation is self-interacting to satisfy CMB phase-shift and damping-tail constraints; (iii) dark radiation history includes a step-like injection that differentially affects low- and high- l CMB multipoles; and (iv) dark radiation interacts with dark matter on 10–20 Mpc scales to mitigate the S8 tension.

A natural next step is to tie the origin, timing, and features of DR to minimal DM microphysics. The benchmark CDM paradigm has emphasized thermal relics that froze out of chemical equilibrium with an electroweak-scale annihilation rate (the “WIMP miracle”). The mass of WIMPs ranges from a few GeV up to the unitarity bound near 100 TeV. However, persistent null results motivate alternatives to the thermal WIMPs. In standard nonthermal scenarios, the initial abundance of DM is determined by the production history rather than a freeze-out. Examples include inflaton decay (Moroi & Yin 2021), purely gravitational production at the end of inflation (Chung et al. 1999), and freeze-in via feeble portal couplings (Hall et al. 2010). Beyond these, we consider a different non-standard scenario, where nonthermal X particles are initially overproduced, with $\rho_{ini} \gg \rho_\infty$ (Fig. 1). The final abundance is then fixed by a subsequent depletion stage without requiring prior equilibrium. Because depletion occurs in a cold, out-of-equilibrium regime (“cold” freeze-out), gravity plays an important role. The interaction rates and kinematics differ qualitatively from the canonical WIMP case, and the usual unitarity bound for hot, equilibrium plasmas does not apply to nonthermal relics. This landscape accommodates a superheavy nonthermal DM senario, which populates the DR and modifies early-time expansion.

The essential dynamics are as follows. Free streaming superheavy X particle becomes nonrelativistic very early (step 1 in Fig. 1) and cools rapidly so that its thermal velocity falls below the gravity-induced velocity sourced by large-scale perturbations (step 2). In this ultra-cold regime, overproduced X particles have a density comparable to radiation, and can gravitationally collapse to form a bound state with de Broglie-scale extent; overlapping wavefunctions enable fast conversion, populating the dark radiation, and depleting the initial overabundance (step 3). This “cold” freeze-out predicts a superheavy mass $m_X \approx 10^{12}$ GeV and an effective cross section $\langle\sigma v\rangle \approx 10^{-21} \text{ m}^3 \text{ s}^{-1}$. Exact solutions produce the full density evolution, implying that only one in a billion (10^9) particles survives the depletion while the rest converted into DR with an energy density of $\Delta N_{eff} \approx 0.4$, sufficient to impact the sound horizon and alleviate the Hubble tension (Riess et al. 2019). Therefore, superheavy nonthermal DM offers a candidate origin for DR: DM microphysics sets the composition, temperature, and interaction of DR. This work develops

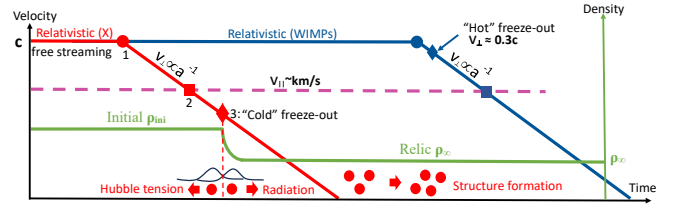


Figure 1. The “cold” freeze-out of nonthermal heavy X particles (red) contrasts with the “hot” freeze-out of thermal WIMPs (blue). WIMPs are thermal relics that become nonrelativistic much later, with a thermal velocity around $0.3c$ at freeze-out. By contrast, nonthermal X particles become nonrelativistic much earlier (1: circle), with thermal velocity redshifting rapidly to far below c . At a subsequent epoch, the thermal velocity $V_\perp$ falls below the gravity-induced velocity $V_\parallel$ sourced by large-scale density perturbations (2: square), enabling the gravitationally bound state to be formed due to the particles’ overabundance, heavy mass, and extreme coldness (3: diamond). Overlapping particle wavefunctions within these bound states permit conversion into extra radiation at extremely low velocities (i.e., “cold” freeze-out). For X particles with an initial overabundance ρ_{ini} , this cold freeze-out sets the relic abundance ρ_∞ ; the density evolution of X particles is shown in green. The additional radiation relative to Λ CDM could alleviate the Hubble tension. Stages 1, 2, and 3 are discussed in sections 2, 4, and 6, respectively.

such a framework, aiming for a simple, testable connection between DM microphysics and the step, self-interacting DR phenomenology that alleviates the Hubble and S8 tension.

The paper is organized in two parts. The first is independent of the overproduction scenario and applies equally to standard dark matter. Section 2 analyzes the free-streaming step and identifies a critical particle mass of 10^{12} GeV (Step 1). Section 3 introduces the gravitational bound state, where UV/IR mixing points independently to the same 10^{12} GeV scale. Section 4 then follows the evolution of the thermal and gravity-induced velocities (Step 2). Section 5 discusses the DM-DR drag interaction that is relevant for S8 tension and suggests a similar mass scale of 10^{12} GeV. The second part addresses overproduction. Section 6 develops the “cold” freeze-out of nonthermal relics (Step 3), whose exact solution tracks the density evolution and the conversion into extra dark radiation in the early Universe, and predicts a superheavy mass of 10^{12} GeV, in agreement with the free-streaming analysis. Section 7 connects this framework to the stepped dark-radiation model for the H_0 tension.

2 FREE STREAMING AND CRITICAL PARTICLE MASS

The smallest dark matter structures are set by the free streaming scale: primordial density perturbations below this scale are erased by the thermal motion of the dark matter particles. The corresponding free streaming mass $M_{fs}(m_X)$, i.e., the minimal halo mass, depends on the particle mass m_X . To determine the free streaming mass M_{fs} , we begin with the comoving Jeans length (Schneider et al. 2013):

$$\lambda_J(t) = \frac{\sigma_{DM}(t)}{a(t)} \sqrt{\frac{\pi}{G\bar{\rho}(t)}}, \quad (1)$$

where $\sigma_{DM}(t)$ is the particle velocity dispersion, $a(t)$ is the scale factor, G is Newton’s constant, and the background density scales as $\bar{\rho}(t) \propto a^{-4}$ in the radiation dominated era. Perturbations with wavelengths larger than λ_J collapse gravitationally, whereas smaller-scale perturbations are suppressed. The velocity dispersion evolves with time: particles are relativistic at early times ($\sigma_{DM}(t) = c$ for $t < t_{NR}$) and become nonrelativistic at $t = t_{NR}$ (or $a = a_{NR}$), after

which $\sigma_{DM} \propto a^{-1}$. We consider particles produced relativistically to become nonrelativistic later at time t_{NR} or a_{NR} .

For thermal relics produced in equilibrium with the radiation bath of temperature T , the typical momentum and energy are $p \sim E \sim T \gg m$ at production. Such particles become nonrelativistic when the redshifted temperature falls to $T \sim m$, as well as p and E . For nonthermal relics, however, the production mechanism sets the initial momentum independently of the SM temperature. We therefore parametrize the initial momentum at production by $p_{prod} \equiv \alpha_{prod} T_{ref}$, where T_{ref} is a reference radiation temperature at production scale factor a_{prod} . The parameter α_{prod} is only a kinematic bookkeeping parameter; it does not imply that the X particles were ever in thermal equilibrium with the SM bath. After production, the physical momentum redshifts as $p(a) = p_{prod} a_{prod}/a$. The nonrelativistic transition is determined by $p(a_{NR}) \simeq m$, rather than by the condition $T(a_{NR}) \simeq m$. With radiation temperature scales as $T(a) = T_{ref} a_{prod}/a$ (same as $p(a)$), particles become nonrelativistic at $a_{NR} \simeq \alpha_{prod} T_{\gamma 0}/m$. If one uses $\langle p \rangle \simeq 3T$, this becomes

$$a_{NR} \simeq 3\alpha_{prod} \frac{k_B T_{\gamma 0}}{mc^2}, \quad (2)$$

where $T_{\gamma 0} = 2.7K$ (or 0.235 meV) is the CMB temperature at $z = 0$.

In this work, we assume an initially overproduced nonthermal population and encode the production kinematics through α_{prod} and initial density ρ_{ini} at scale factor a_{prod} . The benchmark case $\alpha_{prod} \sim 1$ corresponds to particles whose initial momentum is of order the reference radiation temperature (same as thermal relics), while larger α_{prod} corresponds to a hotter nonthermal spectrum and shifts the free-streaming mass accordingly. Because these relics are out of thermal equilibrium, once they become nonrelativistic their physical momentum and velocity dispersion redshift as $p \propto a^{-1}$ and $\sigma_{DM} \propto a^{-1}$. During radiation domination, $\bar{\rho} \propto a^{-4}$, and therefore the comoving Jeans length remains approximately constant for $t > t_{NR}$ with $\lambda_J \propto a^0$ (Eq. (1)). By contrast, for thermal relics that remain coupled to the radiation bath after becoming nonrelativistic, $t_{NR} < t < t_{dec}$, the velocity dispersion scales as $\sigma_{DM} \propto a^{-1/2}$, which implies $\lambda_J \propto a^{1/2}$ and hence a larger free-streaming mass. This trend is evident in Fig. 2, where the WIMP case lies above the nonthermal benchmark.

Since Jeans length remains constant for $t > t_{NR}$, we have Jeans length determined by the nonrelativistic time a_{NR} ,

$$\lambda_J = \frac{2ct_{NR}}{a_{NR}} = \frac{c}{H_0} \frac{a_{NR}}{\sqrt{\Omega_{rad}}}. \quad (3)$$

Here, we use the relation in the radiation era ($t \propto a^2$),

$$t = \frac{a^2}{2H_0 \sqrt{\Omega_{rad}}}. \quad (4)$$

In nature unit, we can express the Hubble constant as

$$\frac{H_0}{M_{pl}} \sqrt{\Omega_{rad}} = \frac{T_{\gamma 0}^2}{M_{pl}^2} \sqrt{\frac{4\pi^3}{45} g_*} \approx 3.06 \frac{T_{\gamma 0}^2}{M_{pl}^2}, \quad (5)$$

where Ω_{rad} is the fraction of radiation at $z = 0$, $g_* \approx 3.4$ is the effective number of relativistic degrees of freedom today. Next, the comoving free streaming mass M_{fs} reads (from Eqs. (3)),

$$M_{fs} = \frac{\pi}{6} \lambda_J^3 \Omega_{DM} \bar{\rho}_0 = \frac{\pi}{6} \Omega_{DM} \bar{\rho}_0 \left(\frac{c}{H_0} \frac{a_{NR}}{\sqrt{\Omega_{rad}}} \right)^3, \quad (6)$$

where Ω_{DM} is the mass fraction of dark matter at $z = 0$. The critical density of the Universe is $\bar{\rho}_0 = 3H_0^2/8\pi G$. Inserting a_{NR} into Eq. (6) leads to the final free streaming mass

$$M_{fs} = \frac{1}{16} \left[\frac{\Omega_{DM}}{\Omega_{rad}^{3/2}} \frac{(3\alpha_{prod} k_B T_{\gamma 0})^3}{GH_0 c^3} \right] m_X^{-3} \propto (m_X)^{-3}. \quad (7)$$

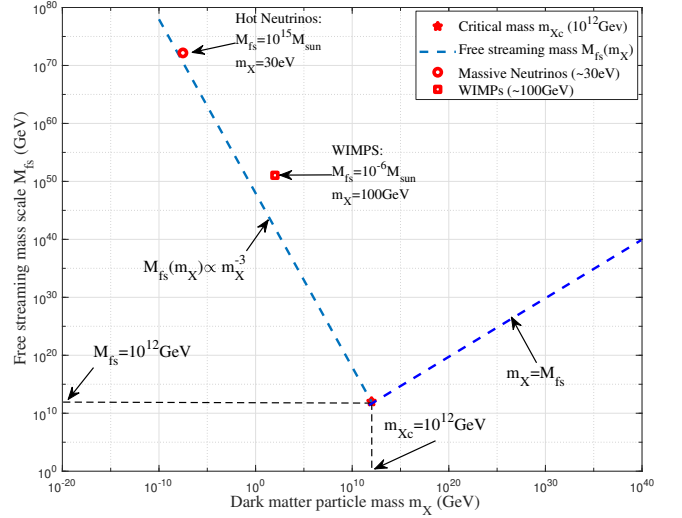


Figure 2. Free streaming mass M_{fs} as a function of particle mass m_X . The computed scaling $M_{fs} \propto m_X^{-3}$ is shown as dashed line (Eq. (7)). The free streaming mass $M_{fs} \approx 10^{15} M_\odot$ for hot neutrinos with $m_X \approx 30$ eV (circle) and $M_{fs} \approx 10^{-6} M_\odot$ for WIMPs with $m_X \approx 100$ GeV (square). The critical mass satisfies $m_{Xc} = M_{fs}(m_{Xc})$ (star) and is $m_{Xc} \approx 10^{12}$ GeV (Eq. (8)). Thus, m_{Xc} represents the minimum structure mass attainable across dark matter candidates and, because smaller structures form earlier, marks the earliest possible structures.

We find the free streaming mass decreases with particle mass as $M_{fs} \propto m_X^{-3}$, as shown in Fig. 2. A critical particle mass m_{Xc} is then identified as the particle mass exactly equal to the free streaming mass, i.e., $M_{fs}(m_{Xc}) = m_{Xc}$, which is the particle mass to form the smallest possible dark matter structure. For $\alpha_{prod} = 1$, taking $M_{fs} = m_X$, we find the critical mass m_{Xc} (red star in Fig. 2):

$$m_{Xc} = \frac{1}{2} \left[\frac{\Omega_{DM}}{\Omega_{rad}^{3/2}} \frac{(3k_B T_{\gamma 0})^3}{GH_0 c^3} \right]^{1/4} \approx \frac{3}{2} \times 10^{-15} \text{ kg} \approx 10^{12} \text{ GeV}, \quad (8)$$

or equivalently with $H_0 \simeq 1.44 \times 10^{-42} \text{ GeV}$,

$$\frac{m_{Xc}}{M_{pl}} = 0.75 \left[\frac{\Omega_{DM}}{\Omega_{rad}^{3/4}} \right]^{1/4} \left(\frac{H_0}{M_{pl}} \right)^{1/8}. \quad (9)$$

The calculation yields a very important mass scale, which we will frequently revisit. The crossing condition $M_{fs}(m_{Xc}) = m_{Xc}$ has a clear physical interpretation: for $m_X < m_{Xc}$, the minimum self-gravitating structure ($M_{fs} \gg m_X$) always contains many particles; for $m_X > m_{Xc}$, $M_{fs} < m_X$ and no meaningful structure can form below the single-particle mass. Thus m_{Xc} is the boundary at which quantum-gravitational single-particle binding becomes relevant, and should be interpreted as a preferred scale delineating this regime. This scale is smaller than the mass of WIMPzilla that is set by the production with $m_X \sim H_I \sim 10^{13}$ GeV (Chung et al. 1999). We offer a structure-formation motivation that leads to a smaller mass scale.

Figure 2 displays the theoretical scaling between particle mass and free streaming mass (dashed curve). For hot, massive neutrinos with $m_X \approx 30$ eV, the free streaming mass is $M_{fs} \approx 10^{15} M_\odot$ (red circle), reflecting their later transition to non-relativistic velocities and the resulting long free streaming length. By contrast, for WIMPs with $m_X \approx 100$ GeV, one finds $M_{fs} \approx 10^{-6} M_\odot$ (the mass of the Earth; red square) (Bertschinger 2006). These examples illustrate the strong inverse dependence of M_{fs} on m_X : larger particle masses yield smaller free streaming masses and permit the formation of smaller

structures. For $m_X < m_{Xc}$, the minimal self-gravitating structure set by free streaming is much larger than the individual particle ($M_{fs} \gg m_X$). For $m_X > m_{Xc}$, the free streaming mass falls below the particle mass, implying that the smallest structure formed is limited by the particle mass scale. Thus, m_{Xc} defines the mass of the smallest dark matter structure attainable across all particle candidates.

3 TWO-BODY GRAVITATIONAL BOUND STATE.

The crossing condition $M_{fs}(m_{Xc}) = m_{Xc}$ marks the boundary at which quantum-gravitational single-particle binding becomes relevant. In this section, we consider the simplest scenario, a gravitationally bound two-particle state (Bohr-like gravitational atom), to provide insights into the compact bound state. Such binding is feasible for sufficiently heavy particles $m_X \geq m_{Xc}$. Within this bound state, we adopt two relations. The first follows from virial equilibrium, connecting kinetic and potential energies ($2KE + PE = 0$)

$$2v_X^2 = \frac{Gm_X}{2r_X}, \quad (10)$$

where r_X denotes a characteristic size (orbital radius) and v_X is the corresponding velocity in the bound configuration. Efficient bound-state conversion requires substantial wavefunction overlap; accordingly, we impose that r_X is of the order of the de Broglie length through the angular-momentum quantization condition

$$m_X v_X \cdot r_X = \hbar. \quad (11)$$

This quantization mirrors the Bohr model, in which the electron angular momenta assume discrete values in units of $\hbar$.

Equations (10) and (11) define a minimal Newtonian (nonrelativistic) quantum-gravity model that couples gravitational binding to quantum-mechanical constraints. Combining these two relations yields closed-form expressions for the relevant physical scales as functions of m_X . The particle size is

$$r_X = \frac{4\hbar^2}{G(m_X)^3} = \frac{4\hbar}{c} \left(\frac{M_{Pl}}{m_X} \right)^2. \quad (12)$$

The resulting length scale $r_X \gg \hbar/(m_X c)$ is a gravity-induced quantum scale, set by the balance between quantum pressure and gravity. The characteristic velocity, formation time, and scale factor read

$$\begin{aligned} v_X &= \frac{G(m_X)^2}{4\hbar} = \frac{c}{4} \left(\frac{m_X}{M_{Pl}} \right)^2, \\ t_X(m_X) &= \frac{\pi}{\sqrt{2}} \frac{r_X}{v_X} = \frac{2^{7/2} \pi \hbar^3}{G^2(m_X)^5} = \frac{2^{7/2} \pi \hbar}{c^2} \frac{(M_{Pl})^4}{m_X^5}, \\ a_X(m_X) &= 14.75 \left(\frac{M_{Pl}^3 T^2}{m_X^5} \right)^{1/2}. \end{aligned} \quad (13)$$

For any time before t_X , the formation of the bound state is not possible because of fast expansion, thus suppressing the bound-state conversion. The velocity v_X is Planck-suppressed, as discussed in Section 7. The gravitational binding energy E_b reads

$$E_b = m_X v_X^2 = \frac{G^2 m_X^5}{16\hbar^2} = \frac{m_X}{16} \left(\frac{m_X}{M_{Pl}} \right)^4 = \frac{(\hbar/r_X)^2}{m_X} = \frac{\pi \hbar}{\sqrt{2} t_X}. \quad (14)$$

Starting with zero energy, these two particles undergo an energy shift akin to the Rydberg energy in the Bohr model. The binding energy and formation time satisfy the "uncertainty principle" $E_b t_X \sim \hbar$.

A single dimensionful quantity—a *specific power* ε (energy per unit mass per unit time, units $\text{m}^2 \text{s}^{-3}$)—appears at both ends of the

scale hierarchy and links them. At the UV end it takes its Planck value; at the m_X end it is fixed by the bound-state structure of Eq. (13):

$$\varepsilon_{pl} = \frac{c^3}{l_p} = \frac{M_{Pl} c^4}{\hbar}, \quad \varepsilon_u = \frac{v_X^3}{r_X} = \frac{1}{2^8} \varepsilon_{pl} \left(\frac{m_X}{M_{Pl}} \right)^9, \quad (15)$$

where l_p is the Planck length. Inverting Eq. (15) gives the mass,

$$\frac{m_X}{M_{Pl}} = 2 \left(\frac{\varepsilon_u}{2 \varepsilon_{pl}} \right)^{1/9}. \quad (16)$$

If ε_u is scale- and time-independent, it may be identified with the large (IR) cosmological scale structure-formation rate $\varepsilon_u \approx u^2 H = u_0^2 H_0 \sim 10^{-7} \text{m}^2 \text{s}^{-3}$ set by the dark-matter velocity dispersion $u_0 \sim 300 \text{km s}^{-1}$ and the Hubble rate, tying the smallest bound state to the largest cosmic scale, suggesting an energy-cascade picture worth further study (Xu 2023a). The resulting ratio $\varepsilon_u/\varepsilon_{pl} \sim (u_0/c)^2 H_0/M_{Pl} \sim 10^{-67}$ is among the smallest dimensionless numbers in cosmology, the square root of the cosmological-constant hierarchy $\rho_\Lambda/\rho_{pl} \sim (H_0/M_{Pl})^2 \sim 10^{-122}$.

Two features make this suggestive rather than a coincidence. First, it is a genuine instance of *UV/IR mixing*: a short-distance mass is controlled by a largest-scale dynamical rate, echoing the short/long-distance entanglement of the Cohen–Kaplan–Nelson bound (Cohen et al. 1999). Second, the $1/9$ power renders the hierarchy *technically natural*: the small ratio $m_X/M_{Pl} \sim 10^{-7}$ is generated robustly rather than tuned, since a large uncertainty in ε_u enters m_X only through its ninth root. The construction does not explain the hierarchy but traces it to $\varepsilon_u/\varepsilon_{pl}$, and ultimately to H_0/M_{Pl} :

$$\frac{m_X}{M_{Pl}} = 2 \left(\frac{\varepsilon_u}{2 \varepsilon_{pl}} \right)^{1/9} = 0.4 \left(\frac{H_0}{M_{Pl}} \right)^{1/9}, \quad (17)$$

i.e. $m_X \sim 10^{12} \text{GeV}$, fixed by UV/IR mixing under the assumption of a scale-independent parameter ε_u , is also in agreement with the critical mass m_{Xc} from free-streaming analysis (Eq. (8)). Being a suggestion of the particle mass, this holds independently of the overproduction scenario and applies equally to standard dark matter cosmologies.

4 GRAVITY-INDUCED DARK MATTER VELOCITY

We now examine the evolution of the dark matter peculiar velocity. In the early Universe, the particle velocity can be decomposed into a longitudinal (gravity-induced) curl-free component $V_{||}$, sourced by large-scale density perturbations, and a transverse component $V_{\perp}$ associated with non-gravitational thermal motions. At early times, the thermal component dominates, but it redshifts rapidly and soon falls below the gravity-induced longitudinal component (Fig. 1). For the transverse part, the comoving velocity $\mathbf{u}_{\perp} = V_{\perp}/a$ satisfies $\dot{\mathbf{u}}_{\perp} + 2H\mathbf{u}_{\perp} = 0$, implying $\mathbf{u}_{\perp} \propto a^{-2}$ and hence $V_{\perp} \propto a^{-1}$. This thermal component sources the vorticity. Because $V_{\perp}$ decays much faster than the longitudinal component $V_{||}$, it is generally an excellent approximation to treat the large-scale dark matter velocity field as irrotational, with negligible vorticity (e.g., (Xu 2023c, 2024a,b)). To compute the gravity-induced velocity $V_{||}$, we consider the comoving longitudinal velocity $\mathbf{u}_{||} = V_{||}/a$, which is coupled to the matter density perturbation δ . The linearized perturbation equations read

$$\begin{aligned} \dot{\delta} + \nabla \cdot \mathbf{u}_{||} &= 0, & (\text{continuity equation}) \\ \dot{\mathbf{u}}_{||} + 2H\mathbf{u}_{||} &= -\frac{1}{a^2} \nabla \delta \Phi, & (\text{momentum conservation}) \\ \nabla^2 \delta \Phi &= 4\pi G \rho_{DM} a^2 \delta, & (\text{Poisson equation}) \end{aligned} \quad (18)$$

where $\delta\Phi$ is the potential perturbation. The gravity-induced component can be expressed as the gradient of potential perturbation $\delta\Phi$, i.e., $\mathbf{u}_{||} \propto -\nabla\delta\Phi$ (Zeldovich approximation). Substituting into Eq. (18), density perturbation δ and comoving velocity $\mathbf{u}_{||}$ read

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G\rho_{DM}\delta, \quad (19)$$

$$\dot{\mathbf{u}}_{||} + 2H\mathbf{u}_{||} = 4\pi G\rho_{DM}\frac{\delta}{\bar{\delta}}\mathbf{u}_{||}.$$

These equations are valid in both radiation and matter eras, where the Hubble parameter satisfies

$$H^2 = \frac{8}{3}\pi G\rho_{DM}\left(1 + \frac{a_{eq}}{a}\right). \quad (20)$$

Neglecting gravitational forcing on large scales, i.e., setting the right-hand side of Eq. (19) to zero, yields the standard logarithmic growth, $\delta \propto \ln t$, together with a rapid decay of the longitudinal velocity $\mathbf{u}_{||} \propto a^{-2}$ during radiation domination. When gravitational terms are retained, Eq. (19) admits the exact solution

$$\delta = 2B/3 + B(a/a_{eq}), \quad (21)$$

with B a constant. Consequently, in the radiation era ($a \ll a_{eq}$) the density contrast is effectively frozen and does not grow in the linear regime on large scales, the Mészáros effect (Mészáros 1974). In the matter era ($a \gg a_{eq}$), perturbations grow linearly $\delta \propto a$ for $a \gg a_{eq}$.

Combining Eqs. (21) and (20), RHS of Eq. (19) for $\mathbf{u}_{||}$ becomes

$$4\pi G\rho_{DM}\frac{\delta}{\bar{\delta}} = \frac{1}{2}\frac{2 + 3a/a_{eq}}{1 + a/a_{eq}}H. \quad (22)$$

Substituting this into Eq. (19) for $\mathbf{u}_{||}$, the equation for $\mathbf{u}_{||}$ reads

$$\frac{d}{dt}(\mathbf{u}_{||}^2) + H\mathbf{u}_{||}^2\frac{2 + a/a_{eq}}{1 + a/a_{eq}} = 0. \quad (23)$$

This yields an exact solution for the gravity-induced velocities,

$$\mathbf{u}_{||}^2 = C(1/a + 1/a_{eq})/a, \quad (24)$$

$$V_{||}^2 = \mathbf{u}_{||}^2 a^2 = C(1 + a/a_{eq}),$$

where C is a constant. Thus, on large scales during radiation domination ($a \ll a_{eq}$), there exists a nonzero, frozen longitudinal velocity with $V_{||}^2 = C$, paralleling the frozen density contrast δ . In a matter-dominant era, the velocity dispersion grows as $V_{||}^2 \propto a$ (Eq. (24)). Taking $V_{||}$ at $z=0$ to be of the order of a few hundred km/s (e.g., 450 km/s from Illustris (Vogelsberger et al. 2014) and 415 km/s from (Sheth & Diaferio 2001)) and $a_{eq} \approx 1/3400$ (Aghanim et al. 2021), we infer $C = 6 \times 10^7 \text{ m}^2/\text{s}^2$, implying a constant gravity-induced velocity $V_{||} \approx 7.7 \text{ km/s}$ in radiation era.

More rigorously, the constant velocity $V_{||}$ is fixed by the linear matter power spectrum $P(k)$. From linear theory, the variance of the large-scale peculiar velocity field in a domain of size R reads (Strauss & Willick 1995)

$$V_{||}^2(z=0) = \frac{H_0^2 [f(\Omega_m)]^2}{2\pi^2} \int_0^\infty P(k, z=0) W^2(kR) dk, \quad (25)$$

where $f(\Omega_m) \approx \Omega_m^{0.55}$ is the linear growth-rate factor, $\Omega_m = \Omega_{DM} + \Omega_b$ is the total matter fraction, and the window function $W(kR)$ approaches 1 on scales larger than individual haloes. Here $V_{||}$ is the center-of-mass bulk velocity of any structure driven by large-scale gravitational perturbations. Because $P(k)$ decays steeply toward small k , the integral is controlled by super-halo scales; consequently, $V_{||}$ is relatively insensitive to the halo mass (Xu 2023b).

To obtain a closed-form estimate, we approximate the integral by its dominant contribution near the equality peak, $\int P dk \approx$

$P(k_{max}, z=0) k_{max}$, where k_{max} is the comoving wavenumber of matter-radiation equality. Using the standard results

$$k_{max} \approx \frac{H_0}{c} \sqrt{\frac{2\Omega_m^2}{\Omega_{rad}}}, \quad P(k_{max}, z=0) \approx \frac{2\pi^2 A_s}{k_{max}^3 a_{eq}^2}, \quad (26)$$

where the factor a_{eq}^{-2} accounts for the suppression of sub-equality modes relative to large-scale modes during radiation domination. Substituting into Eq. (25), we obtain

$$V_{||}(z=0) \approx \frac{f(\Omega_m)}{\Omega_m} \left(\frac{\Omega_{rad} A_s}{2 a_{eq}^2} \right)^{1/2} c. \quad (27)$$

The frozen radiation-era value then follows from Eq. (27):

$$V_{||} = V_{||}(z=0) a_{eq}^{1/2} \approx \frac{f(\Omega_m)}{\Omega_m} \left(\frac{\Omega_{rad} A_s}{2 a_{eq}} \right)^{1/2} c \equiv \lambda_X c. \quad (28)$$

For Λ CDM parameters ($A_s = 2.1 \times 10^{-9}$, $\Omega_m \approx 0.14/h^2$, $\Omega_{rad} \approx 4.2 \times 10^{-5}/h^2$ (Aghanim et al. 2021)), the bulk flow velocity $V_{||} \equiv \lambda_X c$ is a constant of order a few km s^{-1} in the radiation era, consistent with the Illustris streaming velocity (Vogelsberger et al. 2014) and close to the post-recombination baryon sound speed ($\sim 6 \text{ km s}^{-1}$) (Tselikhovich & Hirata 2010). Although the precise value varies with model details, a km s^{-1} -scale $V_{||}$ is robust: during radiation domination it is independent of both the particle mass m_X and time, and is therefore expected to be common across dark-matter candidates.

The velocity fixes a characteristic halo scale. Requiring $V_{||}^2 = Gm_h^*/r_h^*$ with a virialized overdensity $\delta_c = 18\pi^2 \approx 200$,

$$V_{||}^2 = \frac{Gm_h^*}{r_h^*}, \quad m_h^* = \frac{4}{3}\pi r_h^{*3} \rho_{DM}(a_{eq}) \delta_c, \quad (29)$$

gives $m_h^* \sim 100 M_\odot$ and $r_h^* \sim 0.04 \text{ pc}$ at matter-radiation equality, in contrast to today's characteristic mass $m_h^* \sim 10^{12} M_\odot$.

5 DM PARTICLE MASS FROM DM-DR DRAG

In this section, we consider nonrelativistic dark matter X of mass m_X in a dark-radiation bath of temperature T_{DR} ($m_{DR} \equiv T_{DR}/c^2$) and estimate particle mass based on DM-DR drag physics. In the radiation-era constant-drag regime the momentum-transfer rate is velocity-independent (Joseph et al. 2023; Bagherian et al. 2025),

$$\Gamma = \lambda_{\text{eff}} \frac{T_{DR}^2}{m_X}, \quad \lambda_{\text{eff}} \sim \lambda_X^2, \quad R_\Gamma \equiv \frac{\Gamma}{H} \approx \lambda_X^2 \frac{M_{Pl}}{m_X}, \quad (30)$$

where λ_X is a renormalizable Yukawa DM-DR coupling constant in stepped DR framework. The measured drag $R_\Gamma \approx 0.07$ thus fixes one relation between m_X and λ_X . Below we show the same coupling is fixed independently by an intrinsic transport bound, yielding a consistency check on m_X .

(i) **Transport bounds at fixed T_{DR} .** To isolate transport from cosmology we work at fixed T_{DR} and switch off expansion ($H \rightarrow 0$), so Γ alone governs a random walk of step speed V and mean free path $d_X = V/\Gamma$, with diffusivity

$$D_X = \frac{1}{3} V d_X = \frac{V^2}{3\Gamma} = \frac{\hbar}{3m_X} \left(\frac{m_X}{m_{DR}} \right)^2 \frac{(V/c)^2}{\lambda_X^2}. \quad (31)$$

In kinetic equilibrium equipartition sets $V \rightarrow V_{th}$, giving DM thermal speed and diffusivity

$$V_{th} = c \sqrt{\frac{3T_{DR}}{m_X}}, \quad D_{th} = \frac{V_{th}^2}{3\Gamma} = \frac{1}{\lambda_X^2 T_{DR}}. \quad (32)$$

Holding V_{th} fixed in thermal equilibrium, $D_{th} \propto \lambda_X^{-2}$, and the two ends of the coupling range are each bounded:

(a) *Strong-coupling (quantum) floor.* As λ_X grows the step d_X cannot localize below the thermal de Broglie length, so D_X reaches floor at $\hbar/3m_X$ for a maximum coupling λ_{max}

$$d_{min} = \frac{\hbar}{m_X V_{th}} = \left(\frac{1}{3T_{DR} m_X} \right)^{1/2}, \quad (33)$$

$$D_{min} = \frac{\hbar}{3m_X}, \quad \lambda_{max} = \left(\frac{3m_X}{T_{DR}} \right)^{1/2}, \quad \Gamma_{max} = 3T_{DR},$$

which gives the corresponding values for d_X , D_X , and Γ at λ_{max} .

(b) *Weak-coupling (coherence) ceiling.* As $\lambda_X \rightarrow 0$, D_{th} would diverge, which is unphysical: once d_X reaches the largest coherent transport scale, D_X saturates and becomes coupling-independent, $D_{max} \propto 1/\Gamma \propto T_{DR}^2/m_X$ (Eq. (30)). From Eq. (31) this plateau is reached when the step speed is $V \sim \lambda_X c$; hence the critical speed $V_c \sim \lambda_X c$ is not an assumption but the unique speed saturating the bound for the specific DM-DR interaction in Eq. (30). Setting $V_c = V_{th}$ defines the minimum coupling λ_{min}

$$\lambda_X^* c = V_{th} = c \sqrt{\frac{3T_{DR}}{m_X}} \Rightarrow \lambda_X^* \equiv \lambda_{min} = \sqrt{\frac{3T_{DR}}{m_X}}, \quad (34)$$

with ceiling quantities

$$D_{max} = \frac{\hbar}{3m_X} \left(\frac{m_X}{m_{DR}} \right)^2, \quad d_{max} = d_{min} \left(\frac{m_X}{T_{DR}} \right)^2, \quad (35)$$

$$\lambda_{min} = \lambda_{max} \frac{T_{DR}}{m_X}, \quad \Gamma_{min} = \Gamma_{max} \left(\frac{T_{DR}}{m_X} \right)^2.$$

Thus λ_X^* separates two regimes: for $\lambda_X > \lambda_X^*$ the drag is strong, DM thermalizes at V_{th} , and $D_{th} = 1/(\lambda_X^2 T_{DR})$; for $\lambda_X < \lambda_X^*$ the drag is weak and D_X plateaus at the coupling-independent ceiling.

(ii) **Turning on expansion: the sweet spot and the particle mass.** Expansion enters through a single parameter $R_\Gamma = \Gamma/H$. The gravity-induced bulk flow $V_\parallel$ is throttled by drag as $V_d \simeq V_\parallel/(1 + R_\Gamma)$, separating two regimes:

- $R_\Gamma \gg 1$ ($V_\parallel \ll V_c$): trapped and thermalized, $V_d \rightarrow V_{th}$;
- $R_\Gamma \ll 1$ ($V_\parallel \gg V_c$): collisionless, CDM-like, $V_d \sim V_\parallel$.

Since $V_c \sim \lambda_X c$ is the weak/strong DM-DR drag boundary, a marginal sweet spot exists where drag is both cosmologically relevant and yet non-trapping: $V_d \sim V_\parallel \sim V_c \sim \lambda_X c$. Outside it the drag either thermalizes DM ($R_\Gamma \gg 1$) or fully decouples ($R_\Gamma \ll 1$, cosmologically irrelevant for S8 tension). Imposing this window and combining with Eq. (30),

$$m_X \sim \left(\frac{V_\parallel}{c} \right)^2 \frac{M_{Pl}}{R_\Gamma}, \quad V_\parallel \sim c \sqrt{R_\Gamma \frac{m_X}{M_{Pl}}}. \quad (36)$$

For $V_\parallel \sim \text{km s}^{-1}$ and $R_\Gamma \approx 0.07$ this gives $m_X \approx 10^{12}$ GeV, $\lambda_{eff} \approx 10^{-9}$, and a renormalizable Yukawa $\lambda_X \approx 10^{-5}$, tying the large-scale drag to the particle mass and coupling.

(iii) **Coupling and mass from the expansion history.** The critical coupling grows with expansion. Including both Hubble and DM-DR drag, we can write the DM speed V and the equilibrium speed V_{eq} ,

$$\dot{V} = -HV - \Gamma(V - V^*) \Rightarrow V_{eq} = \frac{\Gamma}{\Gamma + H} V^* = \frac{R_\Gamma}{1 + R_\Gamma} \lambda_X^* c. \quad (37)$$

We work in the weak-drag limit near the λ_X^* . Define V^* as a speed

scale such that when DM particles move at speed V^* , they decouple from the drag and only Hubble drag acts. In the weak-drag limit $V^* = \lambda_X^* c$ is the free-streaming onset speed, the weak-coupling ceiling established above. In the absence of expansion ($H = 0$) the equilibrium speed equals the thermal speed, $V_{eq} = V_{th} = V^*$, and λ_X^* reduces to its no-expansion value (Eq. (34)). At fixed equilibrium speed $V_{eq} = V_{th}$, increasing H (smaller R_Γ) demands a larger coupling λ_X^* to offset expansion (and a larger V^*); hence the critical coupling λ_X^* is a function of expansion R_Γ .

We fix the DM-DR coupling to the marginal case—non-trapping yet cosmologically relevant—by setting $\lambda_X = \lambda_X^*$. The relation $V_{th} \approx R_\Gamma \lambda_X c$ for $R_\Gamma \ll 1$ can be obtained from Eq. (37). Combining this relation with Eq. (30), we obtain

$$V_{th} = \lambda_X^3 c \left(\frac{m_X}{M_{Pl}} \right)^{-1} = c \sqrt{\frac{3T_{DR}}{m_X}}. \quad (38)$$

Taking $T_{DR} \sim T_{\gamma 0}/a_X$ comparable to the SM radiation at freeze-out a_X and substituting a_X from Eq. (13), we obtain coupling

$$\lambda_X \simeq \left(\frac{m_X}{M_{Pl}} \right)^{7/12}, \quad R_\Gamma \simeq \left(\frac{m_X}{M_{Pl}} \right)^{1/6}. \quad (39)$$

With $\lambda_X = V_\parallel/c$ (km s⁻¹) or the measured $R_\Gamma = 0.07$, both relations return $m_X \approx 10^{12}$ GeV, consistent with the earlier sections.

6 COLD FREEZE-OUT FOR NONTHERMAL RELICS

The free-streaming analysis, the gravitational bound state, the gravity-induced bulk flow, and the DM-DR drag interaction developed above are all independent of the overproduction scenario and apply equally to standard dark-matter cosmologies, all pointing to a common mass scale $m_X \sim 10^{12}$ GeV.

In this section, we focus on the overproduction scenario and develop the cold freeze-out for nonthermal relics X , which leads to the depletion of an initial overabundance ρ_{ini} to a final relic abundance ρ_∞ . Our goal is to predict the particle mass m_X , the depletion ratio $\gamma = \rho_\infty/\rho_{ini}$, and the production parameter α_{prod} . The physical picture is as follows. Near the cold freeze-out epoch t_X , the X population is nonthermal, ultracold, and initially overabundant. Its density can be comparable to the radiation density at t_X , so the X component may undergo collective self-gravitating infall at freeze out. Collapse proceeds until the gravitational bound state where quantum pressure balances gravity. In this high-compact configuration, overlapping wavefunctions covert dark matter into dark radiation (and/or Standard Model radiation),

$$X + X \rightarrow \text{Dark Radiation} + \text{SM Radiation}. \quad (40)$$

The freeze-out is therefore modeled as a collective gravitational collapse followed by a bound-state conversion cycle. From this description, dominant equations can be established for cold freeze-out:

(i) **Overabundance and energy conversion budget.** If the depletion of the overproduced X population at t_X proceeds efficiently, the released rest-mass energy must be partitioned between dark radiation and the Standard Model bath. Let f_{DR} be the fraction of the converted X energy that becomes dark radiation, with $(1 - f_{DR})$ reheating the SM bath. Normalizing all densities at t_X to the observed (post-injection) SM radiation density $\rho_{rad}^{SM} \equiv 1$ (Table 1), and writing $x_{DM} \equiv \rho_X^-/\rho_{rad}^{SM} = \rho_X^-$ for the pre-conversion dark matter abundance, the SM injection gives the pre-conversion SM radiation abundance,

$$\rho_{SM}^- = 1 - (1 - f_{DR}) x_{DM} (1 - \gamma), \quad (41)$$

Table 1. Component energy densities before and after the cold freeze-out.

DM (ρ_X)	Dark rad. (ρ_{DR})	SM rad. (ρ_{rad}^{SM})
x_{DM}	0	$1 - x_{DM}(1 - \gamma)(1 - f_{DR})$
γx_{DM}	$x_{DM}(1 - \gamma)f_{DR}$	1

i.e. the SM radiation *before* t_X is reduced by the amount later created in conversion. Without losing generality, taking $x_{DM} \equiv \rho_X^- = y_{DM} \rho_{SM}^-$ leads to (mean-field gravitational instability requires X to dominate the pre-injection bath, i.e., $y_{DM} \geq 1$)

$$x_{DM}(f_{DR}) = \frac{1}{\frac{1}{y_{DM}} + (1 - \gamma)(1 - f_{DR})}. \quad (42)$$

The dark-radiation contribution follows,

$$\Delta N_{\text{eff}} = \kappa(g_*) f_{DR} (1 - \gamma) x_{DM} = \frac{\kappa(g_*) f_{DR}}{\frac{1}{y_{DM}(1 - \gamma)} + 1 - f_{DR}}, \quad (43)$$

where κ reflects g_* , the effective number of relativistic degrees of freedom, and is slightly dependent on the freeze out time t_X , with $\kappa \simeq 3.44$ at freeze-out temperature $T_X \sim 1 \text{ GeV}$ and $\kappa \simeq 7.44$ for late freeze-out after BBN.

Inverting Eq. (43), the branching is fixed by the measured ΔN_{eff} ,

$$f_{DR} = \frac{\Delta N_{\text{eff}}}{\kappa + \Delta N_{\text{eff}}} \left(1 + \frac{1}{y_{DM}(1 - \gamma)} \right). \quad (44)$$

In the overproduction scenario $y_{DM} \geq 1$, and a cosmologically relevant dark-radiation component ($\gamma \ll 1$), freeze-out before BBN ($\kappa \simeq 3.44$) is constrained by $\Delta N_{\text{eff}} \leq 0.4$, giving $f_{DR} \leq 0.21$ and $x_{DM} \leq 0.56$. For standard DM scenario without overproduction, the relations reduce to: $\gamma = 1$, $x_{DM} = y_{DM}$, and $\Delta N_{\text{eff}} = 0$. The relic dark matter density should satisfy (using Eq. (13) for cold freeze-out scale factor a_X)

$$\frac{\Omega_{DM} a_X^{-3}}{\Omega_{rad} a_X^{-4}} = \gamma x_{DM} \Rightarrow \left(\frac{m_X}{M_{pl}} \right)^{\frac{5}{2}} = \frac{14.75}{\gamma x_{DM}} \cdot \frac{\Omega_{DM} T_{\gamma 0}}{\Omega_{rad} M_{pl}}. \quad (45)$$

This establishes the first relation between m_X , γ , and x_{DM} .

Next, we require that the virialized configuration of the collapsed overdensity, with number density $n_\infty \delta_c / \gamma$, is precisely the bound state: the interparticle spacing equals the bound-state size r_X ,

$$\left(\frac{n_\infty \delta_c}{\gamma} \right)^{-\frac{1}{3}} a_X = 2 r_X \Rightarrow \left(\frac{m_X}{M_{pl}} \right)^{\frac{5}{2}} = 0.18 \frac{\delta_c}{\gamma} \frac{\Omega_{DM} T_{\gamma 0}}{\Omega_{rad} M_{pl}}, \quad (46)$$

where the present dark-matter density is

$$\rho_{DM} = \frac{3}{8\pi} H_0^2 M_{pl}^2 \Omega_{DM} = 3.06^2 \frac{3}{8\pi} \frac{\Omega_{DM}}{\Omega_{rad}} T_{\gamma 0}^4, \quad (47)$$

with relic number density $n_\infty = \rho_{DM} / m_X$ and virialization contrast $\delta_c \sim 200$. This is equivalent to

$$\Gamma(a_X) = \frac{\delta_c n_\infty a_X^{-3}}{\gamma} \cdot (\pi r_X^2 v_X) \sim \frac{1}{t_X} \sim H_X, \quad (48)$$

i.e., the reaction rate of virialized state is comparable with Hubble expansion. Comparing Eq. (46) with the relic condition Eq. (45) gives $x_{DM} = 82 / \delta_c \approx 0.5$. Thus at freeze-out an overproduced dark-matter component carrying about half the SM radiation density must collapse into the bound state and convert, provided the gravitational collapse dynamics permit it that we will discuss.

(ii) **Jeans length and coldness condition.** At freeze-out t_X the dark matter can dominate over radiation, so the mean homogeneous background becomes gravitationally unstable. To completely convert into

DR, collapse must begin at the smallest single-particle scale, where Poisson (shot) noise drives the density contrast to $\delta \sim 1$. For this collapse to proceed all the way to the gravitationally bound state, the dark matter must be cold: its thermal velocity must have redshifted down to the order of the bound-state velocity v_X , i.e. $V_\perp \sim v_X$.

What sets whether collapse can reach that bound state is the competition between the inward gravitational pull and the internal thermal pressure that resists it. The boundary between the two is the Jeans length λ_J , the smallest scale on which perturbations are unstable and grow; below λ_J , pressure halts collapse. Continuous collapse down to the bound state is therefore possible only if the Jeans length reaches the bound-state scale, $\lambda_J \sim 2 r_X$, which fixes the coldness condition (using Eqs. (2) for a_{NR} , (12) for r_X , and (13) for a_X):

$$\frac{c a_{NR} a_X}{H_0 \sqrt{\Omega_{rad}}} = \frac{8h}{c} \frac{M_{pl}^2}{m_X^3} \Rightarrow \left(\frac{m_X}{M_{pl}} \right)^{1/2} \approx 1.8 \alpha_{prod}. \quad (49)$$

This connects the production parameter α_{prod} to particle mass m_X . The coldness requirement $V_\perp \sim v_X$ is also automatically met, which permits the gravitational collapse to the bound state (Eq. (13)),

$$V_\perp = \frac{a_{NR}}{a_X} c = v_X \Rightarrow \left(\frac{m_X}{M_{pl}} \right)^{1/2} \approx \alpha_{prod}. \quad (50)$$

Equivalently, in terms of the free-streaming mass: for an overproduced population the initial comoving density is $n_i = n_\infty / \gamma$, so requiring the free-streaming mass to equal the particle mass, $M_{fs} = \delta_c \Omega_{DM} \bar{\rho}_0 (\pi \lambda_J^3 / 6) / \gamma = m_X$, gives the critical particle mass (using Eqs. (7) and (5))

$$\frac{m_{Xc}}{M_{pl}} = 0.75 \left[\frac{\alpha_{prod}^3 \delta_c}{\gamma} \cdot \frac{\Omega_{DM}}{\Omega_{rad}^{3/4}} \right]^{\frac{1}{4}} \left(\frac{H_0}{M_{pl}} \right)^{\frac{1}{8}}, \quad (51)$$

where $\delta_c \sim 200$. For bookkeeping scenario $\gamma = \alpha_{prod} = \delta_c = 1$, this reduces to Eq. (9).

(iii) **Kinematic condition for the relic ratio γ .** At t_X the overproduced dark matter dominates the radiation density ($y_{DM} \geq 1$), so gravity dominates expansion. The thermal velocity $V_\perp$ supports no structure above the Jeans length $\lambda_J \sim 2 r_X$ (Eq. (50)), which lies below the particle spacing r_{DM} . Because $\lambda_J < r_{DM}$, thermal pressure does not prevent perturbations on the interparticle and larger scales from becoming gravitationally unstable. Collapse is seeded at the single-particle scale, where Poisson fluctuations give $\delta \sim O(1)$. Self-gravity drives collapse to $\lambda_J \sim 2 r_X$ in one free-fall time (Eq. (49)), where conversion to radiation occurs. In the idealized case with pre-existing bulk flow $V_\parallel = 0$, every particle is swept into range and depletion is complete with $\gamma \rightarrow 0$. The process is thus entirely gravity-driven, in contrast to the canonical WIMP case.

A finite bulk flow $V_\parallel \neq 0$ spoils this completeness and fixes the relic. The flow is curl-free ($\nabla \times \vec{V}_\parallel = 0$), converging on overdensities and diverging in voids, and is coherent from scale $V_\parallel / H$ down to particle spacing $\sim r_{DM}$. Its effect is indirect: the streaming kinetic energy adds to the energy density, raising the expansion rate. From $H^2 = 8\pi G \rho / 3$,

$$\frac{\delta \rho}{\rho} \sim \frac{V_\parallel^2}{c^2} \Rightarrow \frac{\delta H}{H} \sim \frac{V_\parallel^2}{c^2}, \quad (52)$$

with $V_\parallel^2$ the isotropically-averaged dispersion relative to the cosmological frame. Before t_X this effect is negligible ($V_\parallel^2 \ll \rho_{rad}$); at freeze-out, where dark matter dominates, the kinetic contribution shifts the Hubble rate by δH . At fixed comoving density n_i for small

perturbation δH , the balance $\Gamma \simeq H$ then reads (similar to Eq. (48))

$$\begin{aligned} \delta_c n_i (\pi r_X^2 v_X) &\sim (H_X + \delta H) (a_X - \delta a)^3 \\ &\sim H_X a_X^3 \left(1 + \frac{\delta H}{H_X} - 3 \frac{\delta a}{a_X}\right) \Rightarrow \frac{\delta H}{H_X} = 3 \frac{\delta a}{a_X}, \end{aligned} \quad (53)$$

so the faster expansion moves perfect freeze-out time from a_X to $a_X - \delta a$ for a complete conversion, leaving an extra increment of expansion δa : during the collapse–conversion window the enhanced expansion dilutes the dark matter faster than infall, and the fraction δV corresponding to this extra dilution is left behind as the relic. Fixing γ as the unconverted volume fraction,

$$\gamma = \frac{\delta V}{V} = 3 \frac{\delta a}{a_X} = \frac{\delta H}{H_X} \sim \frac{V_{\parallel}^2}{c^2}, \quad (V_{\parallel} \ll c). \quad (54)$$

The low-velocity limits are clear: pre-existing $V_{\parallel} \rightarrow 0$ leaves expansion unperturbed and depletion complete ($\gamma \rightarrow 0$); For the present model $V_{\parallel} \sim \text{km s}^{-1}$ gives $\gamma \sim (V_{\parallel}/c)^2 \sim 10^{-9}$, deep in the low-velocity regime, so depletion is efficient.

With the energy budget (Eq. (44)), the relic density (Eq. (45)), the coldness condition (Eq. (49)), and the kinematic condition for depletion (Eq. (54)), we now determine the particle mass m_X , the depletion ratio γ , and the production parameter α_{prod} .

(i) **Particle mass and depletion ratio.** The kinematic condition (Eq. (54)) leads to γ for km/s scale $V_{\parallel}$,

$$\gamma = \frac{V_{\parallel}^2}{c^2} \sim 10^{-9} - 10^{-8}. \quad (55)$$

Combining with the relic density (Eq. (45)), particle mass m_X reads

$$\frac{m_X}{M_{pl}} = \left(\frac{14.75}{x_{DM}} \cdot \frac{\Omega_{DM} T_{\gamma 0}}{\Omega_{rad} M_{pl}} \right)^{\frac{2}{3}} \left(\frac{V_{\parallel}}{c} \right)^{-\frac{4}{3}}. \quad (56)$$

For overproduction scenario with $x_{DM} \sim \mathcal{O}(1)$ and a km/s scale bulk flow velocity $V_{\parallel}$, this also yields $m_X \sim 10^{12}$ GeV, again, in agreement with suggestions in previous sections. The remaining freeze-out quantities follow directly: $t_X \simeq 5.2 \times 10^{-7}$ s ($a_X \simeq 1.5 \times 10^{-13}$ or $H_X \simeq 10^6/\text{s}$), bound-state size $r_X \simeq 1.2 \times 10^{-13}$ m, thermal velocity $v_X \simeq 5 \times 10^{-7}$ m s⁻¹, and binding energy $E_b \sim 10^{-9}$ eV (Eqs. (12) and (13)). Equation (55) then places the depletion ratio in the range $\gamma \sim 10^{-9} - 10^{-8}$: cold freeze-out near t_X converts the bulk of the population into dark/SM radiation, cutting the number density from n_i to $n_{\infty} = \gamma n_i$. Only this minute surviving fraction remains as relic dark matter to seed structure formation.

(ii) **Particle production parameter.** For $m_X = 10^{12}$ GeV, the coldness condition (49) requires $\alpha_{prod} \sim 10^{-4}$: the particles are produced with momentum well below the ambient radiation temperature, i.e. nonthermally, which suggests production mechanism.

For gravitational production, the produced momentum tracks the inflationary Hubble rate, $p_{prod} \sim H_I$, so

$$\left(\frac{m_X}{M_{pl}} \right)^{1/2} \approx 1.8 \alpha_{prod} = 1.8 \frac{p_{prod}}{T_{ref}} = \left(\frac{H_I}{M_{pl}} \right)^{1/2}, \quad (57)$$

using the production-era temperature $T_{ref} \approx 1.8 \sqrt{H_I M_{pl}}$. Therefore, the coldness condition requirement combined with gravitational production enforces $m_X \sim H_I$, i.e., the heaviest mass that avoids exponential production suppression. For $m_X = 10^{12}$ GeV this demands high-scale inflation, $H_I \sim 10^{12}$ GeV, but still consistent with the Planck limit on the tensor-to-scalar ratio, which caps the inflationary

scale at $H_I < 10^{14}$ GeV (Kolb & Long 2024). Energy conservation then bounds the reheating temperature (Kolb et al. 1998),

$$T_{RH} \leq 0.24 \sqrt{H_I M_{pl}} \approx 10^{15} \text{ GeV}. \quad (58)$$

A second, independent constraint comes from the relic abundance. For a gravitationally produced, conformally coupled massive scalar ($\xi = 1/6$) with $H_{RH} < m_X$ (late reheating), with spin-1/2 fermions following a similar trend (Kolb & Long 2024), the relic density is (Eq. (A4) in Appendix A)

$$\frac{\Omega_{DM} h^2}{0.12} = 0.15 \gamma \left(\frac{m_X}{10^{12} \text{ GeV}} \right)^2 \frac{T_{RH}}{10^6 \text{ GeV}}. \quad (59)$$

Matching the observed abundance with $m_X = 10^{12}$ GeV and $\gamma \sim 10^{-9}$ to 10^{-8} gives $T_{RH} \sim 10^{15}$ GeV.

Cold freeze-out combined with gravitational production therefore requires high-scale inflation, $H_I \sim m_X \sim 10^{12}$ GeV, and efficient reheating, $T_{RH} \sim 10^{15}$ GeV, a value of H_I that sources a primordial tensor-to-scalar ratio

$$r = \frac{P_t}{P_s} = \frac{16}{\pi A_s} \left(\frac{H_I}{M_{pl}} \right)^2 \approx 10^{-5}. \quad (60)$$

With the scalar amplitude fixed at $A_s \simeq 2.1 \times 10^{-9}$, this Hubble scale corresponds to a slow-roll parameter $\epsilon = r/16 \simeq 10^{-6}$ and a tensor-to-scalar ratio $r = 16\epsilon \sim 10^{-5}$, well within the current bound $r < 0.032$, and is potentially testable by future CMB observations.

(iii) **The complete cold freeze-out picture.** We can now assemble the quantitative picture. Superheavy X particles of $m_X \sim 10^{12}$ GeV are overproduced nonthermally, with momentum $\sim 10^{-4}$ of the radiation temperature at production, and become nonrelativistic extremely early ($a_{NR} \sim 10^{-28}$). By freeze-out ($t_X \sim 10^{-6}$ s, $T \sim \text{GeV}$), they become extremely cold with thermal velocity redshifted to $v_X \sim 10^{-7}$ m s⁻¹, and Jeans length $\lambda_J \sim 10^{-13}$ m. At this epoch ΛCDM gives radiation density 10^{21} kg m⁻³ and relic dark matter density 10^{12} kg m⁻³. With $\gamma \sim 10^{-9} - 10^{-8}$, the initial pre-freeze-out overabundance is comparable to or dominant over radiation at t_X . The mean spacing $n_i^{-1/3} \sim 10^{-12}$ m is then only several times the Jeans length and bound-state size $r_X \sim 10^{-13}$ m, so the mean background is unstable and gravitational collapse starts from particle scale down to the Jeans scale. Collapse proceeds to the Jeans/bound-state scale, where overlapping wavefunctions trigger near-instantaneous conversion of the rest energy (plus the $V_{\parallel}$ kinetic energy) into dark and/or SM radiation. A large-scale, curl-free bulk flow $V_{\parallel} \sim \text{km s}^{-1}$ gravitates, slightly enhancing the expansion rate and setting the relic ratio γ . Freeze-out leaves a relic with spacing increased to $n_{\infty}^{-1/3} \sim 10^{-9}$ m. The SM radiation thereafter follows the ΛCDM , while the additional dark radiation alters the expansion history, helping relieve the Hubble tension when combined with the step-radiation framework.

7 HUBBLE TENSION FROM COLD FREEZE-OUT

We propose a concrete realization alleviating the Hubble tension in which, at cold freeze-out t_X , superheavy X particles were converted into dark radiation via a bound-state conversion. A small relic population survives, with a relic density $n_{\infty} = \gamma n_i$ after freeze-out, where $\gamma \approx 10^{-9}$ to 10^{-8} . We assume the converted energy $\rho_{ini}(1-\gamma)f_{DR}$ is deposited into the dark radiation and thermalizes rapidly, producing a dark sector temperature T_{DR} . This extra radiation modifies the early Universe expansion history away from ΛCDM . The fractional excess radiation $\delta_r = x_{DM} f_{DR}$ and the increase in the effective number of

relativistic species ΔN_{eff} reads (Eq. (45))

$$\Delta N_{eff} = \kappa(g_*) f_{DR} x_{DM} = \kappa(g_*) \frac{f_{DR}}{2 - f_{DR}} \leq 0.4. \quad (61)$$

For $\kappa(g_*) \approx 3.44$ at freeze-out $t_X = 10^{-6}$ s, $f_{DR} \leq 0.21$ and $x \leq 0.56$, this corresponds to a maximum $f_{DR} x_{DM} \approx 10\%$ increase in the radiation density relative to Λ CDM at freeze-out, or $\Delta N_{eff} = 0.4$.

The amount of extra radiation is constrained by big bang nucleosynthesis (BBN), the cosmic microwave background (CMB), and large-scale structure. Recent analyses indicate that the Hubble tension can be mitigated for $0.2 \leq \Delta N_{eff} \leq 0.5$ (Seto & Toda 2021). Joint constraints from the CMB, BBN, and the observed D/H abundance favor at the 2σ level, with $\Delta N_{eff} \leq 0.41$ to 0.44 (Nollett & Holder 2011; Aloni et al. 2023; Yeh et al. 2022). With our predicted $\Delta N_{eff} \approx 0.4$, the bound-state conversion of superheavy X particles provides a candidate mechanism to populate dark radiation and potentially alleviate the Hubble tension. The dark radiation is produced at $t_X \approx 10^{-6}$ s (temperature $T_X \approx 1$ GeV), before BBN begins at $T \approx 1$ MeV. What enters BBN is $N_{UV} \approx 0.4$, the pre-step dark radiation. The WZDR step at $z_t \approx 10^4$ (after BBN) then raises this to $N_{IR} = 1.29 N_{UV} \approx 0.5$, which is the relevant quantity for the CMB. Thus the BBN constraint applies to N_{UV} (our predicted value ≈ 0.4 , at the 2σ boundary of (Nollett & Holder 2011) and (Yeh et al. 2022)), while the CMB constraint applies to $N_{IR} \approx 0.5$. Future work should include a dedicated CMB+BAO+LSS likelihood analysis to confirm or constrain this scenario.

Our minimal model setup incorporates the stepped dark radiation (WZDR) framework (Aloni et al. 2023). WZDR contains a strongly self-interacting dark fluid described by a single chiral multiplet $S = (\phi, \psi)$ with Yukawa interaction $\lambda\phi\psi\psi$ and $\lambda^2(\phi^*\phi)^2$. Supersymmetry breaking gives the scalar a tiny mass $m_\phi \sim \text{eV}$ while the fermion ψ remains effectively massless. When the dark sector temperature T_{DR} drops below m_ϕ at a transition redshift z_t , ϕ becomes nonrelativistic and transfers its entropy to ψ . The dark radiation energy density, therefore, exhibits a “step” between BBN and CMB. In the extended WZDR+ scenario, the same scalar also couples to cold dark matter X through a Yukawa interaction $\lambda_X\phi XX$, producing drag between X and ψ . For $T_{DR} \geq m_\phi$ the t-channel scattering rate scales as $\Gamma = \lambda_{eff} T_{DR}^2 / m_X$, where $\lambda_{eff} \sim \lambda_X^2$, so the ratio $R_\Gamma = \Gamma/H$ is constant during radiation domination. Here, $H \propto T_{DR}^2$ with $T_{DR} \propto a^{-1}$ comparable to the SM temperature; after the step, the interaction shuts off rapidly as $(T_{DR}/m_\phi)^4$. WZDR+ is specified by three parameters: (i) the amount of dark radiation $N_{IR} = 1.29 N_{UV}$, where N_{UV} is the amount produced by some unknown process before the step z_t , while N_{IR} is the abundance after z_t ; (ii) a transition redshift $z_t \approx 10^4$ favored by current data; and (iii) a preferred drag strength $R_\Gamma \approx 0.07$ (Joseph et al. 2023; Bagherian et al. 2025).

WZDR+ assumes a total N_{eff} at BBN close to the Λ CDM value and posits that dark radiation is generated between BBN and the CMB, without specifying the underlying mechanism. It further prefers a low-scale SUSY, $M_{SUSY} \sim 100$ TeV, so that gravity mediation naturally yields $m_\phi \sim (M_{SUSY})^2 / M_{Pl} \sim \text{eV}$. Motivated by these limitations, we consider: (i) populating DR through the bound-state conversion of an initially overproduced superheavy dark matter component, such that $N_{UV} \leq 0.4$ (Eq. (61)); (ii) a Planck-suppressed “portal” that, after SUSY breaking, becomes an effective renormalizable coupling, enabling the conversion of dark matter into dark radiation; and (iii) extend to a no-scale supergravity with a post-stabilization high SUSY breaking scale $\sqrt{F} \equiv M_{SUSY} \approx 10^{12}$ GeV, comparable to the dark matter mass m_X . This “no-scale” feature also generates an eV-scale scalar mass.

Our model contains the WZDR fluid as the superfield $S = (\phi, \psi)$

and a dark-matter superfield $X = (\tilde{\chi}, \chi)$, with $\tilde{\chi}$ the scalar and χ the Weyl fermion. The relic dark matter is χ ; the scalar $\tilde{\chi}$ does not survive to late times, and “fermionic dark matter” always denotes χ . We use the non-reduced Planck mass $M_{Pl} = \sqrt{\hbar c/G} \approx 10^{19}$ GeV throughout. Three discrete symmetries organize the superpotential. (i) A discrete R-symmetry Z_4^R , which may also address the μ problem (Lee et al. 2011), assigns $(R_S, R_X, R_Z) = (2, 0, 2)$, requiring every superpotential term to carry R-charge $2 \pmod{4}$; with $R_X = 0$ the operators X, X^2, X^3 are forbidden, so no bare dark-matter mass or self-coupling arises. (ii) An approximate Z_3 in the S sector with $(q_S, q_X, q_Z) = (1, 1, 0) \pmod{3}$ and a Z_3 -neutral spurion: this forbids the linear S tadpole and S^2 that Z_4^R leaves allowed, keeps S^3, SX^2 , and $Z^2 SX^2$ invariant; we treat it as a UV boundary matching condition: ZX^2 is present, generating $m_X = \lambda_Z \sqrt{F}$. (iii) An exact dark parity $Z_2^X : X \rightarrow -X, S \rightarrow S, Z \rightarrow Z$ to stabilize the dark matter. The minimal consistent superpotential is

$$W \supset \frac{\lambda_Z}{2} ZX^2 + \frac{\lambda_S}{3} S^3 + \frac{\lambda_X}{2} SX^2 + \frac{\lambda_K}{2} \frac{Z^2}{M_{Pl}^2} SX^2, \quad (62)$$

where the spurion Z acquires a VEV $\langle z \rangle \approx \sqrt{F} \approx 10^{12}$ GeV that breaks SUSY and, through ZX^2 . The ZX^2 term fixes the dark-matter mass $m_X = \lambda_Z \sqrt{F} \approx 10^{12}$ GeV, so $\sqrt{F} = m_X$ for $\lambda_Z \sim 1$. The renormalizable SX^2 gives the Yukawa $\lambda_X \phi \chi \chi$ responsible for the DM–DR drag. Because the momentum-transfer rate and the Hubble rate share the same temperature dependence, $\Gamma = \lambda_{eff} T_{DR}^2 / m_X$ and $H \propto T_{DR}^2$, the drag ratio $R_\Gamma \equiv \Gamma/H$ is time- and temperature-independent in the radiation era (Appendix 5) and is fixed by cosmology at $R_\Gamma \approx 0.07$. With $R_\Gamma \approx \lambda_{eff} M_{Pl} / m_X$ this gives $\lambda_{eff} \approx 10^{-9}$, and since $\lambda_{eff} \sim \lambda_X^2$, the coupling $\lambda_X \approx 10^{-5}$.

The Planck-suppressed portal $(Z^2/M_{Pl}^2)SX^2$, the minimal dimension-5 supergravity operator, Z_3 -invariant and needing no extra spurion, induces after SUSY breaking

$$y_K = \lambda_K \frac{F}{M_{Pl}^2} \approx \left(\frac{m_X}{M_{Pl}} \right)^2 \approx 10^{-14}, \quad (63)$$

tiny and since $\sqrt{F} = m_X$, equal to $(m_X/M_{Pl})^2$. The binding of the transient $\chi\chi$ bound state is gravitational. The bound-state orbital time $t_X \sim y_K^{-2} m_X^{-1}$ and hence the cold-freeze-out epoch of Eq. (13). The internal velocity likewise follows as $v_X \sim y_K c \sim 10^{-7} \text{ m s}^{-1}$. Once a bound state forms, its decay to DR through $\lambda_X SX^2$

$$\tau_{ann} = \frac{8\pi\hbar}{\lambda_X^2 m_X c^2} \approx 10^{-25} \text{ s} \quad (64)$$

is prompt on the orbital timescale. Free X particles are therefore cosmologically stable: single-particle decay is forbidden by Z_2^X , and the allowed $\chi\chi \rightarrow \text{DR}$ channel shuts off once the population is too dilute to form the required bound state. The hierarchy $V_{||}/v_X \sim \lambda_X/y_K \sim 10^{10}$ simply reflects the difference between a renormalizable Yukawa and a Planck-suppressed, SUSY-breaking portal, whose smallness is fixed by $F/M_{Pl}^2 = (m_X/M_{Pl})^2$.

Finally, the light mass of ϕ arises in a no-scale supergravity embedding. Ordinary supergravity with $\sqrt{F} = 10^{12}$ GeV would give soft masses of order $m_{3/2} \sim F/M_{Pl} \sim 10^2$ TeV, far too heavy; no-scale supergravity cancels this leading ($n = 1$) contribution and suppresses the anomaly-mediated piece, leaving the same portal coupling y_K acting on the gravitino scale,

$$m_\phi \sim y_K m_{3/2} = \frac{F}{M_{Pl}^2} \cdot \frac{F}{M_{Pl}} = \frac{F^2}{M_{Pl}^3}. \quad (65)$$

Equivalently, a leading operator with n spurion insertions gives $m \sim F^n / M_{Pl}^{2n-1}$; no-scale forbids $n = 1$, so $n = 2$ survives. The same y_K

thus also fixes the step-generating scalar mass. Numerically Eq. (65) gives $m_\phi \sim \text{eV}$ with the non-reduced M_{Pl} used here.

8 CONCLUSIONS

We propose a nonthermal, fermionic, superheavy dark-matter component X and analyze its free-streaming scale, Jeans scale, gravitational bound-state, gravity-induced bulk flow, and DM–DR drag for S_8 tension — all independent of the overproduction scenario. Together these point to a benchmark mass $m_X \sim 10^{12} \text{ GeV}$.

We then focus on the overproduction scenario, in which X is initially overproduced relative to its final relic abundance. Unlike a thermal WIMP, the abundance is fixed not by hot equilibrium freeze-out but by a later cold, out-of-equilibrium depletion ("cold freeze-out"), in which most of the initial population is converted into dark and/or SM radiation ("big depletion"). This epoch occurs at $t_X \sim 10^{-6} \text{ s}$, once the particles are ultra-cold and gravity, rather than thermal equilibrium, controls their evolution. The picture is that the X population, with density comparable to radiation at t_X , undergoes gravitational infall into compact bound states; within each state quantum pressure balances gravity at $r_X \sim 10^{-13} \text{ m}$, where the dark matter converts into radiation.

Cold freeze-out proceeds in stages: (i) X becomes nonrelativistic early and cools rapidly; (ii) the overabundant, ultra-cold population undergoes gravitational infall and virialization; (iii) compact bound states form, with enhanced wavefunction overlap enabling conversion into dark/SM radiation; (iv) the process shuts off at a relic level set by the gravity-induced bulk flow $V_\parallel$. The surviving fraction is $\gamma \equiv n_\infty/n_i \sim 10^{-9}$ —about one X per 10^9 initially produced—and the converted energy feeds a dark-radiation bath with $\Delta N_{\text{eff}} \sim 0.4$, in the range that raises the CMB-inferred H_0 .

The particle-physics embedding ties these scales to high-scale supersymmetry breaking. The dark-matter superfield $X = (\tilde{\chi}, \chi)$ has the fermion χ as the late-time relic. A high SUSY-breaking scale $\sqrt{F} \sim 10^{12} \text{ GeV}$ sets m_X , while the Planck-suppressed portal $Z^2 SX^2/M_{Pl}^2$ induces a tiny coupling $y_K \sim m_X^2/M_{Pl}^2 \sim 10^{-14}$, and the renormalizable $\lambda_X SX^2$ gives a Yukawa coupling $\lambda_X \sim 10^{-5}$ providing the DM–DR drag. The drag is active before the WZDR step and shuts off once the scalar mediator becomes Boltzmann-suppressed; by damping clustering on 10–20 Mpc scales it may ease both the S_8 and Hubble tensions. The framework thus unifies superheavy nonthermal dark matter, early dark-radiation production, the stepped WZDR fluid, and DM–DR drag in a single dark-sector construction.

Three assumptions serve as inputs: (1) *Nonthermal overproduction*. X is produced nonthermally with an abundance exceeding the final relic density; the mechanism (gravitational production, inflaton decay, freeze-in) is left unspecified. All other results follow from this input together with the measured Ω_{DM} . (2) *Two-body Bohr quantization*. The bound state is the minimal Newtonian quantum-gravity configuration: virial equilibrium (Eq. (10)) with angular-momentum quantization $m_X v_X r_X = \hbar$ (Eq. (11)), fixing r_X , v_X , t_X in terms of m_X alone. The ground-state ($n = 1$) choice is a strong assumption; excited or many-body configurations would shift the prefactors. (3) *Relic control by $V_\parallel$* . The gravity-induced, curl-free bulk velocity $V_\parallel \sim \lambda_X c$ gravitates, enhances the expansion rates slightly, and sets where it shuts off, fixing γ .

AUTHOR CONTRIBUTION

X. Z developed the theory, performed the computations, and prepared the manuscript.

DATA AVAILABILITY

All data generated or analyzed during this study are included in this published article.

FUNDING

This research was supported by Laboratory Directed Research and Development at Pacific Northwest National Laboratory (PNNL). PNNL is a multiprogram national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under contract no. DE-AC05-76RL01830.

REFERENCES

- Aghanim N., et al., 2021, *Astronomy & Astrophysics*, 652
- Aloni D., Berlin A., Joseph M., Schmaltz M., Weiner N., 2022, *Phys. Rev. D*, 105, 123516
- Aloni D., Joseph M., Schmaltz M., Weiner N., 2023, *Phys. Rev. Lett.*, 131, 221001
- Archidiacono M., Hooper D. C., Murgia R., Bohr S., Lesgourgues J., Viel M., 2019, *Journal of Cosmology and Astroparticle Physics*, 2019, 055
- Bagherian H., Joseph M., Schmaltz M., Sivarajan E. N., 2025, *Phys. Rev. D*, 111, 043513
- Bashinsky S., Seljak U. c. v., 2004, *Phys. Rev. D*, 69, 083002
- Bertschinger E., 2006, *Phys. Rev. D*, 74, 063509
- Blinov N., Marques-Tavares G., 2020, *Journal of Cosmology and Astroparticle Physics*, 2020, 029
- Buen-Abad M. A., Marques-Tavares G., Schmaltz M., 2015, *Phys. Rev. D*, 92, 023531
- Buen-Abad M. A., Schmaltz M., Lesgourgues J., Brinckmann T., 2018, *Journal of Cosmology and Astroparticle Physics*, 2018, 008
- Buen-Abad M. A., Chacko Z., Kilic C., Marques-Tavares G., Youn T., 2023, *Journal of Cosmology and Astroparticle Physics*, 2023, 005
- Chung D. J. H., Kolb E. W., Riotto A., 1999, *Physical Review D*, 59
- Cohen A. G., Kaplan D. B., Nelson A. E., 1999, *Phys. Rev. Lett.*, 82, 4971
- Di Valentino E., et al., 2021, *Classical and Quantum Gravity*, 38, 153001
- Follin B., Knox L., Millea M., Pan Z., 2015, *Phys. Rev. Lett.*, 115, 091301
- Freedman W. L., 2021, *The Astrophysical Journal*, 919, 16
- Hall L. J., Jedamzik K., March-Russell J., West S. M., 2010, *Journal of High Energy Physics*, 2010, 80
- Joseph M., Aloni D., Schmaltz M., Sivarajan E. N., Weiner N., 2023, *Phys. Rev. D*, 108, 023520
- Karwal T., Kamionkowski M., 2016, *Phys. Rev. D*, 94, 103523
- Kolb E. W., Long A. J., 2024, *Rev. Mod. Phys.*, 96, 045005
- Kolb E. W., Chung D. J. H., Riotto A., 1998, WIMPZILLAS! ([arXiv:hep-ph/9810361](https://arxiv.org/abs/hep-ph/9810361)), <https://arxiv.org/abs/hep-ph/9810361>
- Lee H. M., Raby S., Ratz M., Ross G. G., Schieren R., Schmidt-Hoberg K., Vaudrevange P. K., 2011, *Nuclear Physics B*, 850, 1
- Lin M.-X., Benevento G., Hu W., Raveri M., 2019, *Phys. Rev. D*, 100, 063542
- Moroï T., Yin W., 2021, *JHEP*, 03, 301
- Mészáros P., 1974, *A&A*, 37, 225
- Nollett K. M., Holder G. P., 2011, *arXiv*
- Pandey K. L., Karwal T., Das S., 2020, *Journal of Cosmology and Astroparticle Physics*, 2020, 026
- Poulin V., Smith T. L., Karwal T., Kamionkowski M., 2019, *Phys. Rev. Lett.*, 122, 221301

- Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, *The Astrophysical Journal*, 876, 85
- Riess A. G., et al., 2022, *The Astrophysical Journal Letters*, 934, L7
- Schneider A., Smith R. E., Reed D., 2013, *Monthly Notices of the Royal Astronomical Society*, 433, 1573
- Schöneberg N., Abellán G. F., Simon T., Bartlett A., Patel Y., Smith T. L., 2023, *Phys. Rev. D*, 108, 123513
- Seto O., Toda Y., 2021, *Phys. Rev. D*, 103, 123501
- Sheth R. K., Diaferio A., 2001, *Monthly Notices of the Royal Astronomical Society*, 322, 901
- Smith T. L., Poulin V., Amin M. A., 2020, *Phys. Rev. D*, 101, 063523
- Strauss M. A., Willick J. A., 1995, *Physics Reports*, 261, 271
- Tselikhovich D., Hirata C., 2010, *Phys. Rev. D*, 82, 083520
- Vogelsberger M., et al., 2014, *MNRAS*, 444, 1518
- Xu Z., 2023a, *Scientific Reports*, 13, 4165
- Xu Z., 2023b, *Scientific Reports*, 13, 16531
- Xu Z., 2023c, *Physics of Fluids*, 35, 077105
- Xu Z., 2024a, *Physics of Fluids*, 36, 075146
- Xu Z., 2024b, *Physics of Fluids*, 36, 117158
- Yeh T.-H., Shelton J., Olive K. A., Fields B. D., 2022, *Journal of Cosmology and Astroparticle Physics*, 2022, 046

APPENDIX A: GRAVITATIONAL PRODUCTION OF X

This section focus on the gravitational production. We first relate the relic density to the particle density produced during inflation. Following (Chung et al. 1999), we assume the X and inflaton energy densities redshift at the same rate from the end of inflation ($t_I \sim H_I^{-1}$) to the end of reheating (T_{RH} , $t_{RH} \sim \Gamma_\phi^{-1}$), when the inflaton energy converts to radiation. The relic density traces the density at H_I :

$$\frac{\Omega_{DM} h^2}{\Omega_{rad} h^2} = \gamma \frac{T_{RH}}{T_{\gamma 0}} \frac{\rho_X(T_{RH})}{\rho_{rad}(T_{RH})} \approx \gamma \frac{T_{RH}}{T_{\gamma 0}} \frac{\rho_X(H_I)}{\rho_{univ}(H_I)}. \quad (A1)$$

The depletion parameter γ encodes the gap between the produced and relic densities, $n_\infty = \gamma n_i$. Most studies take $\gamma = 1$; here the freeze-out picture instead gives $\gamma \sim 10^{-9}$ – 10^{-8} , which reshapes the production constraints.

The densities at H_I follow from Einstein's equation and the produced number density reads

$$\rho_{univ}(H_I) = \frac{3H_I^2}{8\pi G} = \frac{3}{8\pi} H_I^2 M_{pl}^2, \quad \rho_X(H_I) = m_X n_X, \quad (A2)$$

where n_X depends on the particle's coupling to gravity. For a conformally coupled ($\xi = 1/6$) massive scalar with late reheating ($H_{RH} < m_X$), the particle number density reads

$$n_X = \alpha_2 m_X H_I^2, \quad \alpha_2 \approx 0.0018, \quad (A3)$$

with $\xi = 0$ (minimal) and $\xi = 1/6$ (conformal) bracketing the coupling; spin-1/2 fermions follow a similar trend (Kolb & Long 2024). Substituting into Eq. (A1) relates the relic density to m_X , and T_{RH} :

$$\left(\frac{\Omega_{DM} h^2}{0.12} \right) = \frac{8}{3} \pi \alpha_2 \gamma \left(\frac{m_X}{10^{12} \text{ GeV}} \right)^2 \frac{T_{RH}}{10^5 \text{ GeV}}. \quad (A4)$$

The cold freeze-out picture combined with this relation suggest high-scale inflation and efficient reheating (Eq. (59)).

APPENDIX B: COINCIDENCES AND ALTERNATIVES

Several quantities here share the same gravitational order; read together they are *suggestive* of a single gravitational-strength coupling.

(i) **Suggestive coincidences.** With the single scale $\sqrt{F} = m_X$, every coupling is a power of the gravitational ratio $\mu \equiv m_X/M_{pl}$:

$$\lambda_X = R_\Gamma^{1/2} \mu^{1/2} \sim \mu^{7/12} \sim 10^{-4} \quad (\text{DM-DR drag, Eq. (39)}), \quad (B1)$$

$$R_\Gamma \sim \mu^{1/6} \sim 10^{-1} \quad (\text{DM-DR drag, Eq. (39)}), \quad (B2)$$

$$\alpha_{\text{prod}} \sim \mu^{1/2} \sim 10^{-4} \quad (\text{production, Eq. (49)}), \quad (B3)$$

$$y_K \sim \mu^2 \sim 10^{-14} \quad (\text{portal, Eq. (64)}), \quad (B4)$$

$$\gamma \sim \lambda_X^2 \sim R_\Gamma \mu \sim \mu^{7/6} \sim 10^{-8} \quad (\text{depletion, Eq. (55)}), \quad (B5)$$

with $R_\Gamma = \lambda_X^2 M_{pl}/m_X \sim \mathcal{O}(0.1)$. The coincidence $\lambda_X \sim \alpha_{\text{prod}}$ permits a purely gravitational DM-DR interaction, whereby the physics producing the dark matter also couples it to the dark radiation. Operator dimension fixes $y_K \propto \mu^2$ (dimension-5 portal), while the Yukawa power is fixed by the drag condition, so $y_K \propto \lambda_X^4/R_\Gamma^2$ —so the couplings are not independent, and with $V_\parallel = \lambda_X c$ the depletion fraction of Sec. 6 follows $\gamma \propto \lambda_X^2 \sim (V_\parallel/c)^2$.

(ii) **A possible CP origin, for future study.** The same scaling may have a deeper origin in a matter-antimatter asymmetry of the dark sector, though whether it does depends on the field content. With a single dark-matter flavor, any CP-violating phase can be rotated away; the resulting asymmetry is then portal-suppressed, of order y_K/λ_X , and disappears as the portal is switched off. A second dark flavor changes this. Promoting X to a pair of states turns the couplings into matrices, which carry a genuine, $\mathcal{O}(1)$ CP phase that survives even when the portal vanishes. The decay of the heavier state through an interfering quantum loop then leaves an asymmetry set by the unsuppressed coupling, $\epsilon_X \sim \lambda_X^2 \sim (V_\parallel/c)^2$, up to an order-unity factor—the dark-sector counterpart of multi-flavor leptogenesis, in which the small relic reflects quantum interference rather than a kinematic condition. Reading the relic as this asymmetry would require a long-lived dark charge and efficient depletion of the symmetric part. We do not develop this here; we flag the $\epsilon_X \propto \lambda_X^2$ branch as a natural two-flavor extension of the minimal model and a promising direction for future work.

(iii) **An alternative without dark-matter depletion.** The cold freeze-out mechanism provides dark radiation for stepped-DR. However, it is therefore worth noting that the phenomenological stepped-DR successes on the resolution of the H_0 and S_8 tensions can be retained in an alternative in which the dark matter is not overproduced and depleted and dark radiation has alternative production mechanism. While the 10^{12} GeV mass scale remains a valid scale from free streaming calculation, UV/IR mixing relation, and DM-DR drag physics (Eqs. (8), (17), and (39)). In this alternative, the dark matter χ is a 10^{12} GeV fermion produced nonthermally with no overabundance ($\gamma = 1$). For gravitational production this reproduces $\Omega_{DM} h^2 = 0.12$ directly, with a modest reheating temperature $T_{RH} \sim 10^7$ GeV, well within all cosmological bounds (Eq. (A4)). No subsequent DR conversion is required. This alternative preserves the fermionic nature of the dark matter, the 10^{12} GeV mass, and the joint resolution of the H_0 and S_8 tensions, while discarding the overabundance-plus-depletion structure. The non-depleting alternative provides a conservative fallback.