



Early kinetic decoupling of dark matter: When the standard way of calculating the thermal relic density fails

Tobias Binder,^{1,*} Torsten Bringmann,^{2,†} Michael Gustafsson,^{1,‡} and Andrzej Hryczuk^{2,§}

¹*Institute for Theoretical Physics, Georg-August University Göttingen, Friedrich-Hund-Platz 1, D-37077 Göttingen, Germany*

²*Department of Physics, University of Oslo, Box 1048, NO-0316 Oslo, Norway*

(Received 31 August 2017; published 15 December 2017)

Calculating the abundance of thermally produced dark matter particles has become a standard procedure, with sophisticated methods guaranteeing a precision that matches the percent-level accuracy in the observational determination of the dark matter density. Here, we point out that one of the main assumptions in the commonly adopted formalism, namely local thermal equilibrium during the freeze-out of annihilating dark matter particles, does not have to be satisfied in general. We present two methods for how to deal with such situations, in which the kinetic decoupling of dark matter happens so early that it interferes with the chemical decoupling process: (i) an approximate treatment in terms of a coupled system of differential equations for the leading momentum moments of the dark matter distribution, and (ii) a full numerical solution of the Boltzmann equation in phase space. For illustration, we apply these methods to the case of scalar singlet dark matter. We explicitly show that even in this simple model the prediction for the dark matter abundance can be affected by up to 1 order of magnitude compared to the traditional treatment.

DOI: 10.1103/PhysRevD.96.115010

I. INTRODUCTION

The leading hypothesis for the cosmological dark matter (DM) [1,2] is a new type of elementary particle [3]. One of the most attractive options to explain the present abundance of these particles consists in the possibility that they have been thermally produced in the early universe. This is particularly interesting for the scenario originally studied by Lee and Weinberg [4], as well as others [5–8], in which nonrelativistic DM particles initially are kept in thermal equilibrium with the heat bath through frequent annihilation and creation processes with standard model (SM) particles. Once the interaction rate starts to fall behind the expansion rate of the universe, the DM number density begins to “freeze-out” and remains covariantly conserved. For weakly interacting massive particles (WIMPs), elementary particles with masses and interaction strengths at the electroweak scale, this scenario automatically leads to a relic abundance in rough agreement with the observed DM density—a fact sometimes referred to as the *WIMP miracle*.

The by now standard treatment [9,10] of calculating the resulting DM abundance in these scenarios implements an efficient and highly accurate method of solving the Boltzmann equation for a given (effective) invariant DM annihilation rate. This approach fully captures, in particular,

the three famous exceptions to the original relic density calculations pointed out in a seminal paper by Griest and Seckel [11], namely coannihilations, threshold effects and resonances. The main assumption entering this formalism is that, during the freeze-out process, DM is still kept in *local* thermal equilibrium with the heat bath by frequent scattering processes with relativistic SM particles. For many WIMP candidates, this is indeed satisfied to a high accuracy and kinetic decoupling typically only happens much later than the chemical decoupling [12].

Here we point out that exceptions to this standard lore do exist, even in very simple scenarios, where kinetic decoupling happens so early that it cannot be neglected during the freeze-out process. We develop both semianalytical and fully numerical methods to solve the Boltzmann equation and to compute the DM relic abundance in these cases. Technically, one of the challenges that had to be overcome for obtaining sufficiently accurate results was to extend the highly nonrelativistic Boltzmann equation, as discussed previously in the literature, to the semirelativistic regime. Numerically, we also succeeded to resolve the evolution of the full phase-space distribution accurately enough to test, for the first time, the underlying assumptions for the standard way of calculating the relic density of WIMPs or other self-annihilating DM candidates (for a recent example where the relic density is instead set by inelastic scattering, rather than self-annihilation, see Ref. [13]). We illustrate our general results by a detailed discussion of the scalar singlet model [14–16], for which we find a DM relic

*tobias.binder@theorie.physik.uni-goettingen.de

†torsten.bringmann@fys.uio.no

‡michael.gustafsson@theorie.physik.uni-goettingen.de

§a.j.hryczuk@fys.uio.no

density that differs by up to an order of magnitude from the standard treatment.

This article is organized as follows. In Sec. II, we start with a general description of the underlying Boltzmann equation that governs the DM phase-space evolution. We then briefly review the standard treatment of solving for the DM number density (Sec. II A), extend this by deriving a *coupled* system of evolution equations for the number density and the velocity dispersion (Sec. II B), and finally introduce our framework for a fully numerical solution (Sec. II C). Section III is devoted to a thorough application of these methods to the scalar singlet model. We comment on our results in Sec. IV, and discuss potential other areas of application, before we conclude in Sec. V. In two Appendices we discuss in detail the evolution of the singlet DM phase-space density for selected parameter points (Appendix A) and comment on the semirelativistic form of the scattering operator in the Boltzmann equation (Appendix B).

II. THERMAL PRODUCTION OF DARK MATTER

Let us denote the DM particle by χ , and its phase-space density by $f_\chi(t, \mathbf{p})$. The evolution of f_χ is governed by the Boltzmann equation which, in an expanding Friedmann-Robertson-Walker universe, is given by [17,18]

$$E(\partial_t - H\mathbf{p} \cdot \nabla_{\mathbf{p}})f_\chi = C[f_\chi]. \quad (1)$$

Here, $H = \dot{a}/a$ is the Hubble parameter, a the scale factor, and the collision term $C[f_\chi]$ contains all interactions between DM and SM particles f . For WIMPs, we are to leading order interested in two-body processes for DM annihilation and elastic scattering, $C = C_{\text{ann}} + C_{\text{el}}$, where

$$\begin{aligned} C_{\text{ann}} = & \frac{1}{2g_\chi} \int \frac{d^3\tilde{p}}{(2\pi)^3 2\tilde{E}} \int \frac{d^3k}{(2\pi)^3 2\omega} \int \frac{d^3\tilde{k}}{(2\pi)^3 2\tilde{\omega}} \\ & \times (2\pi)^4 \delta^{(4)}(\tilde{p} + p - \tilde{k} - k) \\ & \times [|\mathcal{M}|_{\tilde{\chi}\chi \leftrightarrow \tilde{f}f}^2 g(\omega)g(\tilde{\omega}) - |\mathcal{M}|_{\tilde{\chi}\chi \leftrightarrow \tilde{f}f}^2 f_\chi(E)f_\chi(\tilde{E})], \quad (2) \end{aligned}$$

and

$$\begin{aligned} C_{\text{el}} = & \frac{1}{2g_\chi} \int \frac{d^3k}{(2\pi)^3 2\omega} \int \frac{d^3\tilde{k}}{(2\pi)^3 2\tilde{\omega}} \int \frac{d^3\tilde{p}}{(2\pi)^3 2\tilde{E}} \\ & \times (2\pi)^4 \delta^{(4)}(\tilde{p} + \tilde{k} - p - k) |\mathcal{M}|_{\chi f \leftrightarrow \chi f}^2 \\ & \times [(1 \mp g^\pm)(\omega)g^\pm(\tilde{\omega})f_\chi(\tilde{\mathbf{p}}) - (\omega \leftrightarrow \tilde{\omega}, \mathbf{p} \leftrightarrow \tilde{\mathbf{p}})]. \quad (3) \end{aligned}$$

In the above expressions, $|\mathcal{M}|^2$ refers to the respective squared amplitude, summed over *all* spin and other internal degrees of freedom, as well as all SM particles f . We assume the SM particles to be in thermal equilibrium, such that their phase-space distribution is given by $g^\pm(\omega) = 1/[\exp(\omega/T) \pm 1]$. Note that we have neglected Bose

enhancement and Pauli blocking factors for f_χ here, as we assume DM to be nonrelativistic; momentum conservation then implies that, in C_{ann} , we can also neglect these factors for the SM particles.

Assuming CP invariance, and using the fact that in thermal equilibrium annihilation and creation processes should happen with the same frequency, the annihilation term given by Eq. (2) can be further simplified to [9]

$$\begin{aligned} C_{\text{ann}} = & g_\chi E \int \frac{d^3\tilde{p}}{(2\pi)^3} v \sigma_{\tilde{\chi}\chi \rightarrow \tilde{f}f} \\ & \times [f_{\chi,\text{eq}}(E)f_{\chi,\text{eq}}(\tilde{E}) - f_\chi(E)f_\chi(\tilde{E})], \quad (4) \end{aligned}$$

where $v = v_{\text{Møller}} \equiv (E\tilde{E})^{-1}[(p \cdot \tilde{p})^2 - m_\chi^4]^{1/2}$ is the Møller velocity, which in the rest frame of one of the DM particles coincides with the lab velocity $v_{\text{lab}} = [s(s - 4m_\chi^2)]^{1/2}/(s - 2m_\chi^2)$.

The scattering term, on the other hand, is in general considerably more difficult to manage. Analytic expressions have, however, been obtained in the highly non-relativistic limit of the DM particles, and assuming that the momentum transfer in the scattering process is much smaller than the DM mass [12,18–22]:

$$C_{\text{el}} \simeq \frac{m_\chi}{2} \gamma(T) \left[T m_\chi \partial_p^2 + \left(p + 2T \frac{m_\chi}{p} \right) \partial_p + 3 \right] f_\chi, \quad (5)$$

where the momentum exchange rate is given by

$$\gamma(T) = \frac{1}{48\pi^3 g_\chi m_\chi^3} \int d\omega g^\pm \partial_\omega (k^4 \langle |\mathcal{M}|^2 \rangle_t), \quad (6)$$

with

$$\langle |\mathcal{M}|^2 \rangle_t \equiv \frac{1}{8k^4} \int_{-4k_{\text{cm}}^2}^0 dt (-t) |\mathcal{M}|^2 = 16\pi m_\chi^2 \sigma_T, \quad (7)$$

and $k_{\text{cm}}^2 = (s - (m_\chi - m_f)^2)(s - (m_\chi + m_f)^2)/(4s)$ evaluated at $s = m_\chi^2 + 2\omega m_\chi + m_f^2$. Here, $\sigma_T = \int d\Omega (1 - \cos \theta) d\sigma/d\Omega$ is the standard *transfer cross section* for elastic scattering. In Appendix B, we discuss how the scattering term is expected to change in the semirelativistic case, i.e. when the assumption of highly nonrelativistic DM is slightly relaxed. For reference, we will in the following use

$$C_{\text{el}} \simeq \frac{E}{2} \gamma(T) \left[T E \partial_p^2 + \left(p + 2T \frac{E}{p} + T \frac{p}{E} \right) \partial_p + 3 \right] f_\chi \quad (8)$$

when explicitly addressing this regime.

A. The standard treatment

In order to calculate the DM relic abundance, we can integrate the Boltzmann Eq. (1) over \mathbf{p} . This results in

$$\frac{dn_\chi}{dt} + 3Hn_\chi = g_\chi \int \frac{d^3p}{(2\pi)^3 E} C_{\text{ann}}[f_\chi], \quad (9)$$

which has to be solved for the DM number density

$$n_\chi = g_\chi \int \frac{d^3p}{(2\pi)^3} f_\chi(\mathbf{p}) \quad (10)$$

(note that C_{el} vanishes once it is integrated over). In order to evaluate the rhs of this equation, the usual assumption [9] is that during chemical freeze-out one can make the following ansatz for the DM distribution:

$$f_\chi = A(T) f_{\chi,\text{eq}} = \frac{n_\chi}{n_{\chi,\text{eq}}} f_{\chi,\text{eq}}, \quad (11)$$

where $A(T) = 1$ in full equilibrium, i.e. before chemical freeze-out. This is motivated by the fact DM-SM scattering typically proceeds at a much faster rate than DM-DM annihilation, because the number density of relativistic SM particles is not Boltzmann suppressed like that of the nonrelativistic DM particles. In that case, DM particles are kept in *local* thermal equilibrium even when the annihilation rate starts to fall behind the Hubble expansion and chemical equilibrium can no longer be maintained.

Approximating furthermore $f_{\chi,\text{eq}}(E) \approx \exp(-E/T)$, i.e. neglecting the impact of quantum statistics for nonrelativistic particles, five of the six integrals in Eq. (9) can be performed analytically. This by now standard treatment, as established by Gondolo and Gelmini [9], results in the often-quoted expression

$$\frac{dn_\chi}{dt} + 3Hn_\chi = \langle \sigma v \rangle (n_{\chi,\text{eq}}^2 - n_\chi^2), \quad (12)$$

where $n_{\chi,\text{eq}} = g_\chi m_\chi^2 T K_2(m_\chi/T)/(2\pi^2)$ and

$$\begin{aligned} \langle \sigma v \rangle &\equiv \frac{g_\chi^2}{n_{\chi,\text{eq}}^2} \int \frac{d^3p}{(2\pi)^3} \frac{d^3\tilde{p}}{(2\pi)^3} \sigma v_{\tilde{\chi}\chi \rightarrow \tilde{f}f} f_{\chi,\text{eq}}(\mathbf{p}) f_{\chi,\text{eq}}(\tilde{\mathbf{p}}) \\ &= \int_1^\infty d\tilde{s} \sigma_{\tilde{\chi}\chi \rightarrow \tilde{f}f} v_{\text{lab}} \frac{2m_\chi \sqrt{\tilde{s}-1} (2\tilde{s}-1) K_1(\frac{2\sqrt{\tilde{s}} m_\chi}{T})}{TK_2^2(m_\chi/T)}. \end{aligned} \quad (14)$$

Here, K_i are the modified Bessel functions of order i , and we have introduced $\tilde{s} \equiv s/(4m_\chi^2)$. While there are various ways to state the final result for $\langle \sigma v \rangle$, the form given above stresses that physically one should indeed think of this quantity as a thermal average of σv_{lab} rather than any other combination of cross section and velocity (in the sense that we strictly have $\langle \sigma v \rangle = \sigma v_{\text{lab}}$ for $\sigma v_{\text{lab}} = \text{const}$; for e.g., $\sigma v_{\text{CMS}} = \text{const}$, on the other hand, with $v_{\text{CMS}} = 2(1 - 4m_\chi^2/s)$, we instead have $\langle \sigma v \rangle \rightarrow \sigma v_{\text{CMS}}$ only in the limit $T \rightarrow 0$).

By introducing dimensionless variables

$$x \equiv m_\chi/T, \quad (15)$$

$$Y \equiv n_\chi/s, \quad (16)$$

and assuming entropy conservation, finally, the above Boltzmann equation for the number density, Eq. (12), can be brought into an alternative form that is particularly suitable for numerical integration:

$$\frac{Y'}{Y} = \frac{sY}{x\tilde{H}} \langle \sigma v \rangle \left[\frac{Y_{\text{eq}}^2}{Y^2} - 1 \right]. \quad (17)$$

Here, $s = (2\pi^2/45)g_{\text{eff}}^s T^3$ denotes the entropy density, $' \equiv d/dx$ and $\tilde{H} \equiv H/[1 + \tilde{g}(x)]$, where

$$\tilde{g} \equiv \frac{1}{3} \frac{T}{g_{\text{eff}}^s} \frac{dg_{\text{eff}}^s}{dT}. \quad (18)$$

The value of Y today, $Y_0 \equiv Y(x \rightarrow \infty)$, can then be related to the observed DM abundance by [9]

$$\Omega_\chi h^2 = 2.755 \times 10^{10} \left(\frac{m_\chi}{100 \text{ GeV}} \right) \left(\frac{T_{\text{CMB}}}{2.726 \text{ K}} \right)^3 Y_0. \quad (19)$$

We note that Eq. (17) is the basis for the implementation of relic density calculations in all major numerical codes [23–29].

B. Coupled Boltzmann equations

The main assumption that enters the standard treatment reviewed above is contained in Eq. (11), i.e. the requirement that during chemical freeze-out, or in fact during any period when the comoving DM density changes, local thermal equilibrium with the heat bath is maintained. If that assumption is not justified, one has in principle to solve the full Boltzmann equation in phase space, Eq. (1), numerically (see next subsection). As first pointed out in Ref. [30], however, it sometimes suffices to take into account the second moment of Eq. (1), instead of only the zeroth moment as in the previous subsection. This leads to a relatively simple *coupled* system of differential equations that generalizes Eq. (17).

The starting point is to define, in analogy to Y for the zeroth moment of f_χ , a dimensionless version of the *second* moment of f_χ :

$$y \equiv \frac{m_\chi}{3s^{2/3}} \left\langle \frac{\mathbf{p}^2}{E} \right\rangle = \frac{m_\chi}{3s^{2/3}} \frac{g_\chi}{n_\chi} \int \frac{d^3p}{(2\pi)^3} \frac{\mathbf{p}^2}{E} f_\chi(\mathbf{p}). \quad (20)$$

For a thermal distribution, the DM particles thus have a *temperature*

$$T_\chi = y s^{2/3} / m_\chi. \quad (21)$$

We note that for nonthermal distributions we could still view this last equation as an alternative *definition* of the DM “temperature,” or velocity dispersion, in terms of the second moment of f_χ as introduced above. This allows, e.g., a convenient characterization of kinetic decoupling as the time when T_χ no longer equals T but instead starts to approach the asymptotic scaling of $T_\chi = T_{\text{kd}}(a/a_{\text{eq}})^{-2}$ for highly nonrelativistic DM [12,18].

Integrating Eq. (1) over $g_\chi \int d^3 p / (2\pi)^3 / E$ and $g_\chi \int d^3 p / (2\pi)^3 \mathbf{p}^2 / E^2$, respectively, we find

$$\frac{Y'}{Y} = \frac{m_\chi}{x\tilde{H}} C_0, \quad (22)$$

$$\frac{y'}{y} = \frac{m_\chi}{x\tilde{H}} C_2 - \frac{Y'}{Y} + \frac{H}{x\tilde{H}} \frac{\langle p^4 / E^3 \rangle}{3T_\chi}, \quad (23)$$

where

$$\langle p^4 / E^3 \rangle \equiv n_\chi^{-1} g_\chi \int \frac{d^3 p}{(2\pi)^3} \frac{\mathbf{p}^4}{E^3} f_\chi(\mathbf{p}) \quad (24)$$

and we introduced the moments of the collision term as

$$m_\chi n_\chi C_0 \equiv g_\chi \int \frac{d^3 p}{(2\pi)^3 E} C[f_\chi], \quad (25)$$

$$m_\chi n_\chi \left\langle \frac{\mathbf{p}^2}{E} \right\rangle C_2 \equiv g_\chi \int \frac{d^3 p}{(2\pi)^3 E} \frac{\mathbf{p}^2}{E} C[f_\chi]. \quad (26)$$

Plugging in $C = C_{\text{ann}} + C_{\text{el}}$ as provided in Eqs. (4) and (5), finally, we arrive at a coupled set of equations that constitutes one of our main results¹:

$$\frac{Y'}{Y} = \frac{sY}{x\tilde{H}} \left[\frac{Y_{\text{eq}}^2}{Y^2} \langle \sigma v \rangle - \langle \sigma v \rangle_{\text{neq}} \right], \quad (27)$$

¹This extends the results presented in [30]. Compared to that reference, we have kept terms proportional to Y_{eq} (see also [31]) and adopted a fully relativistic temperature definition in Eqs. (20) and (21). The latter indeed turns out to be important outside the highly nonrelativistic regime and is the origin of the last term in Eq. (28), as well as the corrected form of $\langle \sigma v \rangle_2$ —which now (unlike in its original form) can be seen as a proper thermal average in the sense that a constant σv_{lab} leads to $\langle \sigma v \rangle_2 = \sigma v_{\text{lab}}$ for all values of T (i.e. not only for $T \rightarrow 0$).

We note that both $\langle p^4 / E^3 \rangle$ and the integral over ϵ_+ can be expressed in terms of a series of Bessel functions when expanding E in the denominator around $E = m$. Since this series does not converge very fast for the relatively small values of x that we will be interested in here, however, we do not display these series.

$$\begin{aligned} \frac{y'}{y} = & \frac{\gamma(T)}{x\tilde{H}} \left[\frac{y_{\text{eq}}}{y} - 1 \right] + \frac{sY}{x\tilde{H}} [\langle \sigma v \rangle_{\text{neq}} - \langle \sigma v \rangle_{2,\text{neq}}] \\ & + \frac{sY}{x\tilde{H}} \frac{Y_{\text{eq}}^2}{Y^2} \left[\frac{y_{\text{eq}}}{y} \langle \sigma v \rangle_2 - \langle \sigma v \rangle \right] + \frac{H}{x\tilde{H}} \frac{\langle p^4 / E^3 \rangle_{\text{neq}}}{3T_\chi}. \end{aligned} \quad (28)$$

Here, in addition to $\langle \sigma v \rangle$ in Eq. (13), we also introduced another, temperature-weighted thermal average:

$$\langle \sigma v \rangle_2 \equiv \frac{g_\chi^2}{T n_{\chi,\text{eq}}^2} \int \frac{d^3 p d^3 \tilde{p}}{(2\pi)^6} \frac{p^2}{3E} \sigma v_{\tilde{\chi}\chi \rightarrow \tilde{f}f} f_{\chi,\text{eq}}(\mathbf{p}) f_{\chi,\text{eq}}(\tilde{\mathbf{p}}) \quad (29)$$

$$\begin{aligned} = & \int_1^\infty d\tilde{s} \sigma_{\tilde{\chi}\chi \rightarrow \tilde{f}f} v_{\text{lab}} \frac{4\tilde{s}(2\tilde{s}-1)x^3}{3K_2^2(x)} \\ & \times \int_1^\infty d\epsilon_+ e^{-2\sqrt{\tilde{s}}x\epsilon_+} \left[\epsilon_+ \sqrt{(\tilde{s}-1)(\epsilon_+^2-1)} \right. \\ & \left. + \frac{1}{2\sqrt{\tilde{s}}} \log \left(\frac{\sqrt{\tilde{s}}\epsilon_+ - \sqrt{(\tilde{s}-1)(\epsilon_+^2-1)}}{\sqrt{\tilde{s}}\epsilon_+ + \sqrt{(\tilde{s}-1)(\epsilon_+^2-1)}} \right) \right], \end{aligned} \quad (30)$$

where we have used $\epsilon_+ \equiv (E + \tilde{E}) / \sqrt{s}$. The “out-of-equilibrium average” $\langle \sigma v \rangle_{2,\text{neq}}$ is defined as in Eq. (29), but for arbitrary $n_\chi, f_\chi(\mathbf{p})$ —and hence also $1/T \rightarrow 1/T_\chi$ in the normalization; the last equality, Eq. (30), thus does not hold in this case. Correspondingly, $\langle \sigma v \rangle_{\text{neq}}$ is defined in analogy to Eq. (13), but equals in general not the expression given in Eq. (14).

Two comments about this central result are in order. The first comment, more important from a practical point of view, is that the set of Eqs. (22) and (23) includes higher moments of f_χ , and hence does not close with respect to the variables Y and y . Concretely, we need additional input to determine the quantities $\langle \sigma v \rangle_{\text{neq}}$, $\langle \sigma v \rangle_{2,\text{neq}}$ and $\langle p^4 / E^3 \rangle_{\text{neq}}$ in Eqs. (27) and (28) in terms of only y and Y . We will make the following ansatz for these quantities:

$$\langle \sigma v \rangle_{\text{neq}} = \langle \sigma v \rangle|_{T=ys^{2/3}/m_\chi}, \quad (31)$$

$$\langle \sigma v \rangle_{2,\text{neq}} = \langle \sigma v \rangle_2|_{T=ys^{2/3}/m_\chi}, \quad (32)$$

$$\langle p^4 / E^3 \rangle_{\text{neq}} = \left[\frac{g_\chi}{2\pi^2 n_{\chi,\text{eq}}(T)} \int dp \frac{p^6}{E^3} e^{-\frac{E}{T}} \right]_{T=ys^{2/3}/m_\chi}. \quad (33)$$

These expressions would, in particular, result from a DM phase-space distribution of the form

$$f_\chi = \frac{n_\chi(T)}{n_{\chi,\text{eq}}(T_\chi)} \exp \left(-\frac{E}{T_\chi} \right) \Big|_{T_\chi=ys^{2/3}/m_\chi}, \quad (34)$$

which describes a situation in which the DM particles follow a Maxwellian velocity distribution with a temperature different from that of the heat bath (as expected, e.g., if the DM particles exhibit significant self-scattering

[30,32–34]). We emphasize, however, that from the point of view of solving the coupled set of Eqs. (27) and (28), there is no need to make such a relatively strong assumption about $f_\chi(\mathbf{p})$: any form of f_χ that leads to (very) similar results for the quantities given in Eqs. (31)–(33) will also lead to (very) similar results for $Y(x)$ and $y(x)$. In other words, we expect our coupled system of Boltzmann equations to agree with the full numerical solution discussed in the next section—concerning the evolution of Y and y —if and only if the ansatz in Eqs. (31)–(33) coincides with the corresponding averages numerically determined from the “true” phase-space distribution. As we will see later, this is indeed very often the case.

The second comment concerns the first term on the rhs of Eq. (28), which is proportional to the second moment of the elastic scattering term given in Eq. (5). As that latter expression is valid only to lowest order in $p^2/E^2 \sim p^2/m_\chi^2 \sim 1/x$, we had for consistency also to neglect any higher-order corrections in these quantities to the elastic scattering part of C_2 when deriving our final result. As discussed in Appendix B, in fact, there is no simple way of determining the next-to-leading order corrections to C_{el} . If we use our default semirelativistic scattering term given in Eq. (8), however, including the resulting corrections from subleading orders corresponds to replacing in Eq. (28),

$$\begin{aligned} T_\chi \left[\frac{y_{\text{eq}}}{y} - 1 \right] &= T - T_\chi \\ &\rightarrow T - T_\chi + \frac{1}{6} \left\langle \frac{p^4}{E^3} \right\rangle - \frac{5}{6} T \left\langle \frac{p^2}{E^2} \right\rangle + \frac{1}{3} T \left\langle \frac{p^4}{E^4} \right\rangle. \end{aligned} \quad (35)$$

By construction, see Appendix B, this operator must still be an attractor to the equilibrium solution, and hence be proportional to (some power of) $T - T_\chi$; for the ansatz of Eq. (34), e.g., this can easily be verified directly. In practice, this replacement has very little impact on the evolution of Y and y , even at times as early as $x \sim 10$. We can think of the resulting small differences as a measure of the intrinsic uncertainty associated to our treatment of the scattering term.

C. The full phase-space density evolution

We now turn to solve the Boltzmann Eq. (1) at the full phase-space density level. This is numerically more challenging, but allows to assess the validity of the assumptions in previous sections and to track deviations (as we will see can occur) from the standard Maxwell Boltzmann velocity distribution. To achieve this, we start by reexpressing Eq. (1) in the two dimensionless coordinates,

$$x(t, p) \equiv m_\chi/T \quad \text{and} \quad q(t, p) \equiv p/T,$$

where the monotonic temperature $T(t)$ replaces as before the time parameter t via our $x(T)$, and q is now the

“momentum” coordinate that depends on both t and p . In these variables, we can rewrite the Liouville operator on the lhs of Eq. (1) as

$$(\partial_t - H\mathbf{p} \cdot \nabla_{\mathbf{p}}) = \partial_t - Hp\partial_p = \tilde{H}(x\partial_x - \tilde{g}q\partial_q). \quad (36)$$

Here, we used the fact that the system is isotropic and assumed, as in the previous sections, that entropy is conserved. With the collision terms given in Eqs. (4) and (8), the Boltzmann equation for f_χ now becomes

$$\begin{aligned} \partial_x f_\chi(x, q) &= \frac{m_\chi^3}{\tilde{H}x^4} \frac{g_{\tilde{\chi}}}{2\pi^2} \int d\tilde{q} \tilde{q}^2 \frac{1}{2} \int d\cos\theta v_{\text{M}\phi} \sigma_{\tilde{\chi}\chi \rightarrow \tilde{f}f} \\ &\times [f_{\chi,\text{eq}}(q)f_{\chi,\text{eq}}(\tilde{q}) - f_\chi(q)f_\chi(\tilde{q})] \\ &+ \frac{\gamma(x)}{2\tilde{H}x} \left[x_q \partial_q^2 + \left(q + \frac{2x_q}{q} + \frac{q}{x_q} \right) \partial_q + 3 \right] f_\chi \\ &+ \tilde{g} \frac{q}{x} \partial_q f_\chi, \end{aligned} \quad (37)$$

where $x_q \equiv \sqrt{x^2 + q^2}$ and θ is the angle between \mathbf{q} and $\tilde{\mathbf{q}}$.

The benefits of this rewriting are twofold. First, the interpretation of the Boltzmann equation becomes very transparent, in the sense that this “comoving” phase-space density $f_\chi(x, q)$ clearly stays unaltered for $\tilde{g}(x) = 0$ and vanishing annihilation and scattering rates (being proportional to $\sigma_{\tilde{\chi}\chi \rightarrow \tilde{f}f}$ and γ , respectively). The new coordinates thus absorb how momentum and DM density change exclusively due to the Hubble expansion. (For nonvanishing \tilde{g} , these quantities continue to scale in the same way with the scale factor a , but taking into account that $a \propto g_{\text{eff}}^{-1/3} T^{-1}$.) Second, the use of a comoving momentum $q \equiv p/T$ significantly helps numerical calculations that extend over a large range in $x = m_\chi/T$. In fact, $f_\chi(x, q)$ is expected to stay unchanged in shape both in the early semirelativistic and kinetically coupled regime, where $f_\chi \sim e^{-p/T_\chi} = e^{-q}$ given that $T_\chi = T$, as well as in the late nonrelativistic kinetically decoupled regime, where $f_\chi \sim e^{-p^2/(2mT_\chi)} \propto e^{-q^2/(2m)}$ given that $T_\chi \propto T^2$ in this case—at least as long as $\tilde{g} = 0$ and the DM phase-space distribution remains close to Maxwellian as in Eq. (34).

Let us stress that here, unlike for our discussion in the previous subsection, it is indeed mandatory to use the semirelativistic form of Eq. (8) for the scattering operator when discussing the evolution of the phase-space density, in the sense that it must drive the distribution function $f_\chi(q)$ towards the fully relativistic form $\propto e^{-E/T}$ [and not as Eq. (5) to the nonrelativistic approximation $\propto e^{-\frac{q^2}{2m}/T}$]. The importance of this can be seen by comparing the second and the third line of Eq. (37). The factor in the second line will always drive DM annihilation to occur unless an equilibrium distribution f_{eq} is reached. The term in the third line determines towards which equilibrium shape the

scattering operator will drive the DM distribution $f_\chi(q)$. If the scattering attractor distribution would not match the $f_{\text{eq}}(q)$ of the second line, then scattering could artificially drive annihilation to occur. For more discussions of the semirelativistic aspects of the scattering term, see Appendix B.

We then use a technique that discretizes the unbounded momentum variable q into a finite number of q_i with $i \in \{1, 2, \dots, N\}$. This allows to rewrite our *integro partial differential equation* into a set of N coupled ordinary differential equations (ODEs):

$$\begin{aligned} \frac{d}{dx} f_i = & \frac{m_\chi^3}{\tilde{H} x^4} \frac{g_{\tilde{\chi}}}{2\pi^2} \sum_{j=1}^{N-1} \frac{\Delta \tilde{q}_j}{2} [\tilde{q}_j^2 \langle v_{M\phi} \sigma_{\tilde{\chi}\chi \rightarrow \tilde{f}f} \rangle_{i,j}^\theta (f_i^{\text{eq}} f_j^{\text{eq}} - f_i f_j) \\ & + \tilde{q}_{j+1}^2 \langle v_{M\phi} \sigma_{\tilde{\chi}\chi \rightarrow \tilde{f}f} \rangle_{i,j+1}^\theta (f_i^{\text{eq}} f_{j+1}^{\text{eq}} - f_i f_{j+1})] \\ & + \frac{\gamma(x)}{2\tilde{H}x} \left[x_{q,i} \partial_q^2 f_i + \left(q_i + \frac{2x_{q,i}}{q_i} + \frac{q_i}{x_{q,i}} \right) \partial_q f_i + 3f_i \right] \\ & + \tilde{g} \frac{q_i}{x} \partial_q f_i, \end{aligned} \quad (38)$$

where $f_i \equiv f_\chi(x, q_i)$, and the derivatives $\partial_q f_i$ and $\partial_q^2 f_i$ are determined numerically by finite differentials using several neighboring points to f_i . $\langle v_{M\phi} \sigma_{\tilde{\chi}\chi \rightarrow \tilde{f}f} \rangle_{i,j}^\theta$ is the velocity-weighted cross section averaged over θ (which is evaluated analytically or numerically) as a function of q_i and \tilde{q}_j , and $\Delta \tilde{q}_j \equiv \tilde{q}_{j+1} - \tilde{q}_j$. Finally, the DM number density in Eq. (10) is determined by trapezoidal integration.

Numerous numerical tests have been performed to ensure stability of our solutions to the ODEs of Eq. (38) and that imposed conditions on the now emerged boundary points (at q_1 and q_N) are physically sound. It turns out that very small step sizes over a large range in q are required for solving these stiff ODEs. We typically used the range $q_1 = 10^{-6}$ to $q_N = 50$ with about thousand steps in between, and set the two last terms of Eq. (38) to zero at q_N while using forward derivatives to evaluate them at q_1 . By the use of the ODE15s code in MATLAB, and by analytically deriving internally required Jacobians, we are able to efficiently calculate the full phase-space evolution for the freeze-out after optimizing numerical settings. On the time scale of a few minutes we can derive the relic abundance for a given DM model. The code is general enough to be adapted to any standard single WIMP setup.

III. SCALAR SINGLET DARK MATTER

The simplest example of a renormalizable model providing a WIMP DM candidate is the scalar singlet model [14–16], originally proposed as DM made of “scalar phantoms” by Silveira and Zee [14]. In this model, the only addition to the standard model is a real gauge-singlet scalar field S which is stabilized by a \mathbb{Z}_2 symmetry and never obtains a nonvanishing vacuum expectation value. The

simplicity of the model has in itself triggered considerable interest [35–44], with a further boost of attention after the discovery of the Higgs boson [45–58]. Recently, the GAMBIT [59] collaboration presented the so far most comprehensive study of this model by performing a global fit taking into account experimental constraints from both direct, indirect and accelerator searches for DM [60].

Interestingly, the resulting parameter region with the highest profile likelihood in this global fit is the one where the scalar singlet mass m_S is about half that of the SM Higgs mass, m_h , and where the DM abundance today is set by the resonant annihilation of two DM particles through an almost on-shell Higgs boson. As we will see, it is exactly in this parameter region that the standard way of calculating the relic density, as implemented in all previous studies of this model, fails because kinetic decoupling happens so early that it essentially coincides with chemical decoupling. Instead, the formalism introduced in the previous section provides a reliable calculation of the relic abundance of scalar singlet DM.

A. Model setup

The model symmetries, along with the requirement of renormalizability, uniquely determine the form of the Lagrangian to be

$$\mathcal{L}_{\text{SZ}} = \mathcal{L}_{\text{SM}} + \frac{1}{2} \partial_\mu S \partial^\mu S - \frac{1}{2} \mu_S^2 S^2 - \frac{1}{2} \lambda_S S^2 H^\dagger H - \frac{1}{4!} \lambda_{SS} S^4, \quad (39)$$

where H is the Standard Model Higgs doublet. The S boson mass receives contributions from both the bare mass term, μ_S , and from electroweak symmetry breaking, leading to $m_S = \sqrt{\mu_S^2 + \frac{1}{2} \lambda_S v_0^2}$, where $v_0 = 246.2$ GeV is the Higgs vacuum expectation value. We adopt the Higgs mass and the total width from decay to SM particles to be $m_h = 125.09$ GeV [61] and $\Gamma_{h, \text{SM}} = 4.042$ MeV [62]. For the moment, we neglect the quartic self-coupling λ_{SS} , but will later comment on its potential (minor) impact on relic density calculations.

The annihilation cross section of DM pairs to SM particles, apart from hh final states, is given by [49]

$$\sigma v_{\text{CMS}} = \frac{2\lambda_S^2 v_0^2}{\sqrt{s}} |D_h(s)|^2 \Gamma_{h \rightarrow \text{SM}}(\sqrt{s}), \quad (40)$$

where $\Gamma_{h \rightarrow \text{SM}}(\sqrt{s})$ is the partial decay width of a Standard Model Higgs boson of mass \sqrt{s} , and

$$|D_h(s)|^2 = \frac{1}{(s - m_h^2)^2 + m_h^2 \Gamma_h^2}. \quad (41)$$

The total Higgs width Γ_h in the above propagator, but not elsewhere, includes not only all SM channels but also the

$h \rightarrow SS$ channel if it is open. For $\Gamma_{h \rightarrow SM}(\sqrt{s})$, as in [49], we use tabulated values for $\sqrt{s} < 300$ GeV from [62] and analytic expressions at higher \sqrt{s} . Note however that the latter high \sqrt{s} region has no impact on the relic density in the studied scalar singlet mass range. Likewise, the channel $SS \rightarrow hh$ lies outside our kinematic region of interest.

For the elastic scattering processes, we take into account DM scattering with all SM fermions. Being mediated only by a Higgs in the t -channel, the corresponding squared amplitude takes a particularly simple form,

$$|\mathcal{M}_{Sf \rightarrow Sf}|^2 = \frac{N_f \lambda_S^2 m_f^2}{2} \frac{4m_f^2 - t}{(t - m_h^2)^2}, \quad (42)$$

where m_f is the mass of the SM fermion and the color factor is $N_f = 3$ for quarks and $N_f = 1$ for leptons. Averaging over the transferred momentum, as in Eq. (7), we thus find

$$\langle |\mathcal{M}|^2 \rangle_t = \sum_f \frac{N_f \lambda_S^2 m_f^2}{8k^4} \left[\frac{2k_{\text{cm}}^2 - 2m_f^2 + m_h^2}{1 + m_h^2/(4k_{\text{cm}}^2)} - (m_h^2 - 2m_f^2) \log(1 + 4k_{\text{cm}}^2/m_h^2) \right]. \quad (43)$$

Note that the sum here runs over all relevant fermions and antifermions *separately*.

The hierarchical Yukawa structure of the Higgs couplings leads to the scattering rate being dominated by the interactions with the heaviest fermions that for a given temperature are still sufficiently abundant in the plasma. In the range of DM masses m_S that we are interested in, freeze-out happens around $T \sim \mathcal{O}(1 \text{ GeV})$, which is not far from the temperature of the QCD phase transition. Consequently, the details of this transition and the SM plasma can have a significant impact on the scattering rate; a study which goes beyond the scope of this work. Therefore, we follow the literature and adopt two extreme scenarios that can be thought of as bracketing the actual size of the scattering term:

- (A) All quarks are free and present in the plasma down to temperatures of $T_c = 154 \text{ MeV}$ (largest scattering scenario, as adopted in [20]).
- (B) Only light quarks (u, d, s) contribute to the scattering, and only for temperatures above $4T_c \sim 600 \text{ MeV}$, below which hadronization effects start to become sizable [63] (smallest scattering scenario, as adopted in [12]).

Finally, we adopt the recent results from Drees *et al.* [64] for the effective number of relativistic degrees of freedom $g_{\text{eff}}(T)$ that enter the calculation of the Hubble rate during radiation domination, $H = \sqrt{4\pi^3 g_{\text{eff}}/45} T^2/m_{\text{Pl}}$, as well as the entropy degrees of freedom entering for example in the calculation of $\tilde{g}(T)$ as defined in Eq. (18).

B. Relic density of scalar singlet dark matter

Let us first compute the relic density following the standard treatment adopted in the literature. To this end, we numerically solve Eq. (17) for a given set of parameters (m_S, λ_S) and determine the resulting asymptotic value of Y_0 . The blue dashed line in Fig. 1 shows the contour in this plane that results in Y_0 corresponding to a relic density of $\Omega h^2 = 0.1188$, cf. Eq. (19). We restrict our discussion to values of m_S in the kinematic range where $\langle \sigma v \rangle$ is enhanced due to the Higgs propagator given in Eq. (41), and the coupling λ_S that results in the correct relic density is hence correspondingly decreased. This curve agrees with the corresponding result obtained in Ref. [49].

For comparison, we show in the same figure the required value of λ_S that results when instead solving the coupled system of Boltzmann equations (27) and (28), or when numerically solving the full Boltzmann equation as described in Sec. II C. Here, the solid (dashed) line shows the situation for the B (A) scenario for scatterings on quarks. Outside the resonance region, the coupled Boltzmann equations lead to identical results compared to the standard approach, indicating that kinetic decoupling indeed happens much later than chemical decoupling and that the assumption of local thermal equilibrium during chemical freeze-out thus is satisfied. For DM masses inside the resonance region, on the other hand, we can see that the

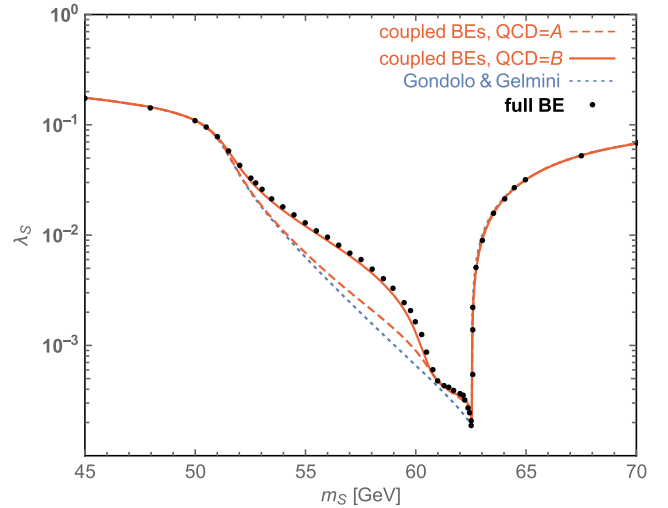


FIG. 1. The required value of the singlet-Higgs coupling λ_S , as a function of the scalar singlet mass m_S , in order to obtain a relic density of $\Omega h^2 = 0.1188$. The blue dashed line shows the standard result as established by Gondolo and Gelmini [9], based on the assumption of local thermal equilibrium during freeze-out. For comparison, we also plot the result of solving instead the coupled system of Boltzmann equations (27) and (28) for the maximal (“B”) and minimal (“A”) quark scattering scenarios defined in the main text (red solid and dashed lines, respectively). Finally, we show the result of fully solving the Boltzmann equation numerically, for the maximal quark scattering scenario and with no DM self-interactions included (“full BE”).

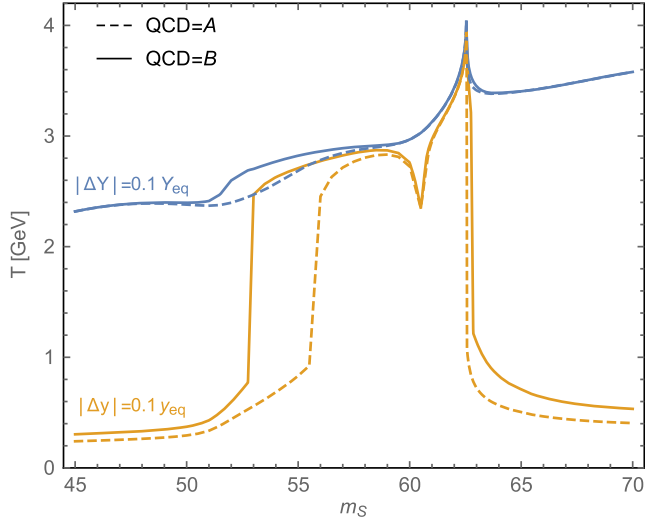


FIG. 2. Temperatures at which DM number density and velocity dispersion (“temperature”) start to deviate from their equilibrium values, defined for the purpose of this figure as $|Y - Y_{\text{eq}}| = 0.1 Y_{\text{eq}}$ and $|y - y_{\text{eq}}| = 0.1 y_{\text{eq}}$, respectively. These curves are based on solving the coupled system of Boltzmann equations (27) and (28), for the same parameter combinations as in Fig. 1 (resulting thus in the correct relic density).

two methods can give significantly different results, implying that this assumption must be violated. For the same reason, a smaller scattering rate (as in scenario B) leads to an even larger deviation from the standard scenario than the maximal scattering rate adopted in scenario A.

This interpretation is explicitly confirmed in Fig. 2, where we plot the temperatures at which the DM number density and temperature start to deviate from the equilibrium values: in the parameter range that we focus on here, kinetic decoupling happens indeed very close to chemical decoupling. The reason for this very early kinetic decoupling is straightforward to understand as the result of a strongly suppressed momentum transfer rate $\gamma(T)$, compared to the annihilation rate, due to two independent effects: (i) the small coupling λ_5 needed to satisfy the relic density requirement, without a corresponding resonant enhancement of $\gamma(T)$, and (ii) the scattering rate being proportional to the Yukawa coupling squared, which favors scattering with Boltzmann-suppressed heavy fermions. We note that the latter point also explains the relatively large difference between the two extreme quark scattering scenarios used here for illustration (in scenario B, the largest Yukawa couplings do not contribute to the scattering).

In order to emphasize the importance of our improved treatment of the decoupling history, we plot in Fig. 3 also the *ratio* of the resulting relic density to that of the standard approach (for parameter values satisfying the relic density constraint for the latter, i.e. corresponding to the blue dashed curve in Fig. 1). Let us stress that, compared to the observational uncertainty in this quantity of about 1%, these corrections are by no means small even in the

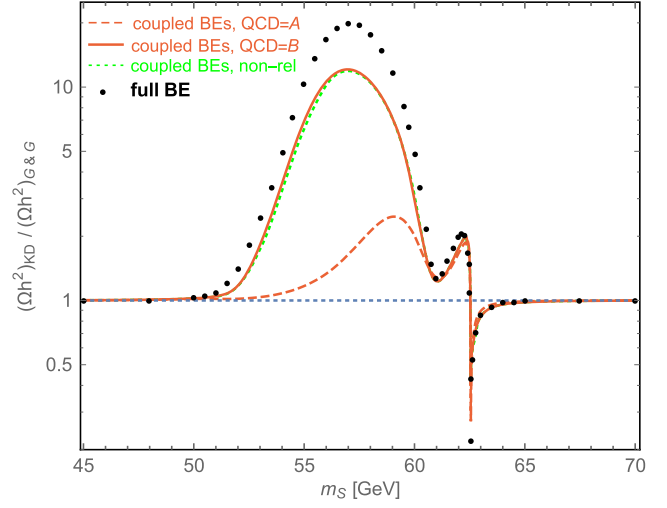


FIG. 3. The impact of the improved treatment of the kinetic decoupling on the relic density for parameter points that would satisfy the relic density constraint in the standard approach (dotted line in Fig. 1), both for the minimal (solid) and maximal (dashed) scenario for scattering with quarks. The numerical result (full BE) implements minimal quark scattering; note that this does not take into account the effect of DM self-interactions (while the other curves are consistent with assuming a maximal self-scattering rate). The green dashed curve shows the impact of implementing the elastic scattering term in the highly non-relativistic limit, cf. Eq. (5).

minimal scattering scenario A. In the same figure, we also compare our result for the coupled system of Boltzmann equations (27) and (28) to the full numerical solution of the Boltzmann equation in phase space, as described in Sec. II C (black dots). Before getting back to these results, let us briefly comment on the green dashed line in Fig. 3, which implements the highly nonrelativistic scattering term C_{el} of Eq. (5), and hence *not* the replacement (35) in Eq. (28) which we otherwise adopt as our default. Clearly, the impact of this choice is very limited for this approach. We note that the quantitative importance of the relativistic correction term proportional to $\langle p^4/E^3 \rangle$ in Eq. (28) lies in the same ballpark, affecting the relic density by at most $\sim 10\%$ in the region very close to the resonance (and below the percent level elsewhere).

In Appendix A we discuss in depth the time evolution of both the coupled Boltzmann equations and the full phase-space density in the resonance region. Let us here just mention that the characteristic features of the curves displayed in Figs. 2 and 3 can indeed all more or less directly be understood in terms of the highly enhanced annihilation rate in a relatively narrow kinematic region around the resonance, $\sqrt{s} \sim m_h \pm \Gamma_h$. As the full numerical solution reveals, furthermore, the shape of $f_\chi(p)$ can in some cases be quite different from the Maxwell-Boltzmann form (34) that is consistent with the coupled system of Boltzmann equations (27) and (28). Whether this has a

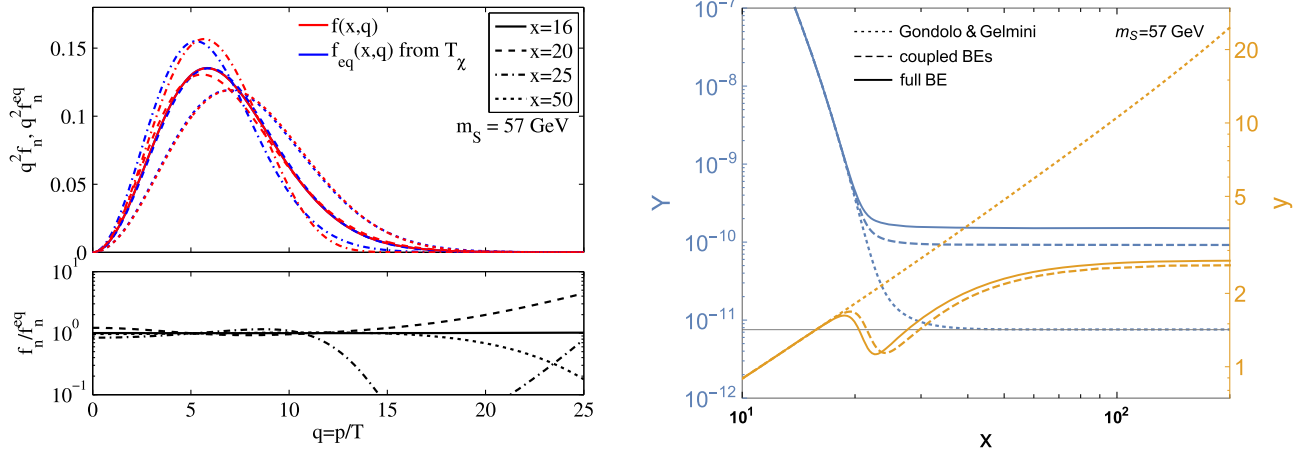


FIG. 4. Phase-space distributions and their evolution for a scalar singlet DM particle with $m_S = 57$ GeV. Left panel: Unit normalized phase-space distributions $f_n(q)$ from our full numerical solution of the Boltzmann equation (red lines) and thermal equilibrium distributions $f_n^{\text{eq}}(q)$ (blue lines) at four different temperatures $x = m_S/T = 16$ (solid), 20 (dashed), 25 (dot dashed) and 50 (dotted). The equilibrium distributions f_n^{eq} are Maxwell-Boltzmann distributions evaluated at the temperatures T_χ , as defined in Eqs. (20) and (21). The bottom part shows the fractional deviation from the respective thermal distribution $f_n(q)/f_n^{\text{eq}}(q)$. Right panel: The evolution of Y (blue) and y (yellow), assuming a Higgs-scalar coupling that leads to the correct relic density in the standard approach (dotted line in Fig. 1). We show these curves for the standard case (dotted lines), the approach using coupled Boltzmann equations (dashed) and the full numerical result (solid). The thin gray line indicates the asymptotic value of Y corresponding to the observed relic density.

noticeable impact on the resulting relic density (like for $m_S \sim 57$ GeV) or not (like for $m_S \sim m_h/2$) again mostly depends on whether or not the shape is affected for those momenta that can combine to $\sqrt{s} \sim m_h$ during chemical freeze-out.

For illustration, we pick a DM mass of $m_S = 57$ GeV and show in Fig. 4 the full phase-space distribution for a few selected values of x (left panel) as well as the relevant evolution of Y and y (right panel). For models with DM masses in this range, the relatively large difference between full solution and coupled equations (as visible in Fig. 3) can mostly be understood in terms of the *dip* in the *ratio* of DM phase-space distributions at intermediate values of $q = p/T$ that starts to develop for $x \gtrsim 20$. Concretely, the fact that the actual distribution for those momenta is slightly suppressed compared to a distribution fully characterized only by its second moment, as in Eq. (34), causes the DM particles to annihilate less efficiently, $\langle \sigma v \rangle_{\text{neq}} < \langle \sigma v \rangle$, because this is the momentum range probed by the resonance for these x values. This in turn leads to the DM particles falling out of chemical equilibrium earlier, and hence a larger asymptotic value of Y . The reason for this momentum suppression to develop in the first place is also to be found in the particularly efficient annihilation close to the resonance, which leads to a depletion of DM particles with corresponding momenta because the scattering rate is no longer sufficiently large to redistribute the phase-space distribution to a thermal shape. We note that the bulk part of this effect is actually well captured by the coupled Boltzmann system, cf. the dashed vs solid lines in the right panel of Fig. 4. For further details, we refer again to Appendix A.

IV. DISCUSSION

From the above discussion, we have learned that very early kinetic decoupling is not just a theoretical possibility. It can appear in simple WIMP models, like the scalar singlet case, and affect the DM relic density in a significant way. We note that the size of the latter effect is, as expected, directly related to the size of the momentum exchange rate and hence to just *how* early kinetic decoupling happens compared to chemical decoupling. Let us stress that, from a general point of view, this is a much more important message connected to our choice of considering two scattering scenarios than the question of which of those scenarios is more realistic for the specific model we have studied here.

We have also seen that the coupled system of Boltzmann equations (27) and (28) provides a qualitatively very good description for the resulting DM abundance, see in particular Fig. 1, even though for high-precision results it seems mandatory to actually solve the full Boltzmann equation in phase space. As discussed in Appendix A, differences can arise when the true phase-space distribution is not of the Maxwellian form assumed in Eq. (34)—though the two methods *can* actually still give almost identical results for the relic abundance even when the two distribution differ vastly. The question of under which conditions the coupled system of equations provides an accurate description of the relic density is thus a somewhat subtle one, and requires a careful discussion of the velocity dependence of the annihilation term in the Boltzmann equation.

An exception to this general complication is a DM *self-interaction* rate large enough to force the DM distribution into the form given by Eq. (34) [30,32–34] and hence

render the coupled system of Boltzmann equations (27) and (28) *exactly* correct (up to, as discussed, corrections due to quantum statistics). Sizable self-scattering rates can for example arise due to corresponding contact interactions, like the quartic coupling λ_{SS} in the scalar singlet case, or by adding light mediators that couple to the DM particle (which was indeed the first time such a coupled system of Boltzmann equations was considered [30], albeit in a different context). For the case of resonant annihilation, furthermore, the same resonance also mediates an enhanced self-interaction. For future work, it would hence be worthwhile to extend our numerical framework to even include those DM self-interaction processes. For the scalar singlet case, in particular, we expect that adding the process $SS \rightarrow h^* \rightarrow SS$ would bring all numerical results for the full Boltzmann equation—e.g., those shown in Fig. 3—even closer to those resulting from the coupled system of Boltzmann equations.

Let us finally stress that both the coupled Boltzmann equations and the numerical setup that we have described here are very general, and can be used to consistently study early kinetic decoupling for a much larger range of models than the scalar singlet case. Obvious applications are other scenarios where resonant annihilation and/or annihilation to heavy final states is important in setting the relic abundance, see also Ref. [31]. Further examples where the ratio of the scattering rate to the annihilation rate can be smaller than usual, hence potentially leading to early kinetic decoupling, include Sommerfeld-enhanced annihilation [30,34,65,66] (if the light mediators are not abundant enough to take part in the scattering process) and annihilation to DM bound states [67,68]. Quite in general, our methods provide a powerful means to check whether the DM particles are indeed in local thermal equilibrium with the heat bath around the time when their abundance freezes out—which is the usual assumption, though rarely explicitly tested, not only in WIMP-like scenarios but also when so-called semiannihilations [69] are important in setting the relic density, when computing the relic abundance for modified expansion histories [70,71], or in scenarios that go beyond simple $2 \rightarrow 2$ annihilation processes [72–74].

V. CONCLUSIONS

The standard way of calculating the thermal relic density of self-annihilating DM particles rests on the assumption of local thermal equilibrium during freeze-out, and that hence kinetic decoupling occurs much later than chemical decoupling. Here, we demonstrated for the first time that departure from kinetic equilibrium can instead happen much earlier, even *simultaneously* with the departure from chemical equilibrium.

By introducing a coupled system of equations for the DM number density and its “temperature”, or rather velocity dispersion, we improved the standard way of calculating the relic density in such cases. For an even

higher accuracy in predicting the DM abundance, we also found a way of solving the full Boltzmann equation numerically. The latter approach has the additional advantage of obtaining the full phase-space distribution, rather than only the number density, which in particular allows to test in detail the assumption of a Maxwellian velocity distribution adopted in the standard approach. A numerical solver for the coupled system of Boltzmann equations, s. (27) and (28), will be available in an upcoming version of DARKSUSY [28] and our implemented solver for the full Boltzmann equation at the phase-space level, s. (38), will be released separately.²

Applied to the simplest renormalizable WIMP model—the scalar singlet, extensively discussed in the literature—we somewhat surprisingly found that the relic abundance predicted in the standard approach can differ by up to an order of magnitude from the correct treatment presented in this paper. This is rather remarkable not only in view of the simplicity of this model, but also because the affected region in parameter space happens to coincide with the best-fit region resulting from most recent global scans. We thus expect our results to have a noticeable phenomenological impact, and that our treatment will prove useful also when applied to other examples of relic density calculations in cases where the standard assumption of local thermal equilibrium during freeze-out is not exactly satisfied.

ACKNOWLEDGMENTS

We thank Mateusz Duch, Joakim Edsjö, Bohdan Grzadkowski, Andreas Hohenegger and Ayuki Kamada for very useful conversations during the preparation of this work. A. H. is supported by the University of Oslo through the Strategic Dark Matter Initiative (SDI). M. G. and T. Binder have received funding from the European Union’s Horizon 2020 research and innovation program under Grants Agreements No. 690575 and No. 674896. T. Binder gratefully acknowledges financial support from the German Science Foundation (DFG RTG 1493).

Note added.—Recently, we became aware of a dedicated study on resonant DM annihilation [31], which also found that DM can kinetically decouple much earlier than usual in this case.

APPENDIX A: PHASE-SPACE DENSITY EVOLUTION OF THE SCALAR SINGLET

In Sec. III B, we investigated the impact of our improved treatment of the Boltzmann equation on the expected DM relic abundance in the scalar singlet model. Here, we supplement this by discussing in some more detail the evolution of the DM phase-space density. The main focus

²Please contact any of the authors if you need these numerical routines prior to their public release.

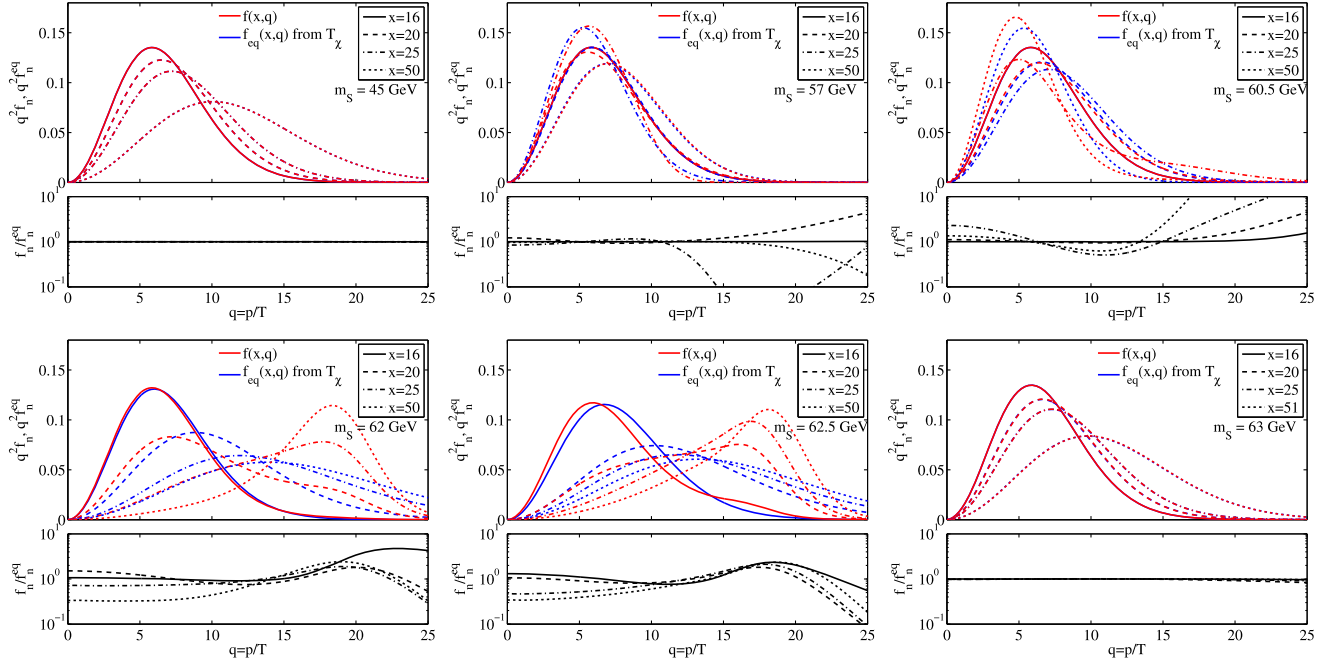


FIG. 5. Same as the left panel of Fig. 4 in the main text, but now for comparison for various DM masses $m_S = 45, 57, 60.5, 62, 62.5, 63$ GeV. Note that for cases where the equilibrium distributions appear to be missing in the top panels, it is just because it agrees very well with the actual phase-space distribution (as also visible in the fractional deviation plotted in the bottom panels).

of this discussion, however, will be a more thorough qualitative understanding of the specific features seen in Figs. 2 and 3, and the underlying interplay of chemical and early kinetic decoupling. Specifically, we can distinguish three mass regimes:

- (1) A regime with $53 \text{ GeV} \lesssim m_S \lesssim 60.5 \text{ GeV}$, which we will refer to as *subresonant* because f_χ starts to deviate from its equilibrium value, $f_{\chi, \text{eq}} = \exp(-E/T)$, at a temperature where the typical DM momenta are too small to hit the resonance, i.e. $\sqrt{s} \lesssim m_h - \Gamma_h$. As a result, we have $\langle \sigma v \rangle_{(\text{neq})} < \langle \sigma v \rangle_{2, (\text{neq})}$ ³ during the whole freeze-out process in this regime—this is because $p^2 f_\chi(p)$ peaks at a higher value of p than $f_\chi(p)$, which brings its bulk distribution closer to (or even on) the cross-section resonance.
- (2) A regime with $60.5 \text{ GeV} \lesssim m_S \lesssim 62.5 \text{ GeV}$ that we will refer to as *resonant*. Here, we have $\langle \sigma v \rangle > \langle \sigma v \rangle_2$ around the time when the DM particles start to leave thermal equilibrium, because the larger mass combines with the relevant momenta to $s \sim m_h^2$. At slightly later times, on the other hand, still relevant in changing the DM abundance, the DM momenta have redshifted so much that we are

back to a situation where typically $\sqrt{s} \lesssim m_h - \Gamma_h$ and hence $\langle \sigma v \rangle < \langle \sigma v \rangle_2$.

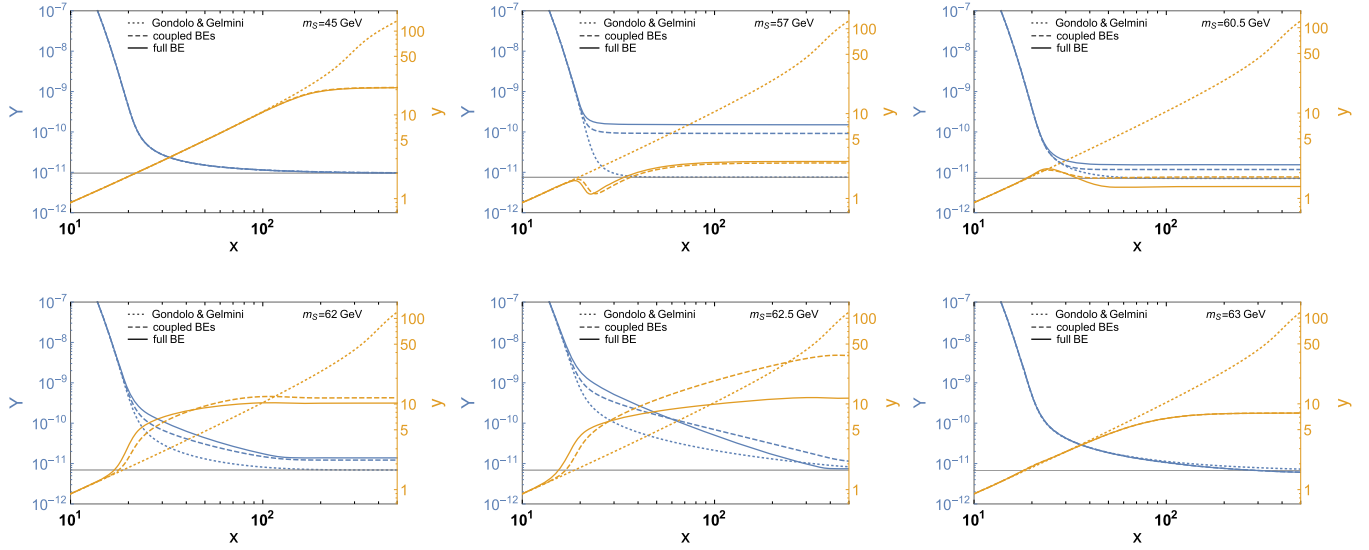
- (3) Finally, there is a *superresonant regime* with $62.5 \text{ GeV} \lesssim m_S \lesssim 65 \text{ GeV}$, where decoupling occurs at such high temperatures that we have $\langle \sigma v \rangle > \langle \sigma v \rangle_2$ during the whole time it takes for $Y(x)$ to reach its asymptotic value (determining the relic density).

To help our discussion, let us look at a selection of benchmark points with scalar singlet masses $m_S = 45, 57, 60.5, 62, 62.5, 63$ GeV and coupling constants $\lambda_S(m_S)$ that result in the correct relic density in the standard approach (dotted line in Fig. 1). In Fig. 5, we show the DM distribution function for these benchmark points that we find with our full numerical approach, for selected values of x , and in Fig. 6 the full evolution of $Y(x)$ and $y(x)$ for the different approaches. These figures thus extend the information in Fig. 4 by covering a range of DM masses.

The first thing to note, as exemplified by the benchmark points with $m_S = 45$ GeV and $m_S = 63$ GeV, is that for masses sufficiently far away from the resonance we find a phase-space distribution which remains almost exactly Maxwellian in shape. For these points, we therefore find as expected a very good agreement for the evolution of $Y(x)$ and $y(x)$ when comparing the numerical solution and the coupled Boltzmann approach, as well as with Y in the standard Gondolo and Gelmini setup (which assumes $T = T_\chi$). We note that this provides an important consistency check for both methods.

An example for a model in the *subresonant* region is the case with $m_S = 57$ GeV, which we discussed in the main

³For the sake of better readability, we will suppress the subscript “neq” for the remainder of this section. We note that, once chemical decoupling has started, the contribution of thermal averages without this subscript is suppressed by a factor of Y_{eq}/Y in Eqs. (27) and (28).

FIG. 6. Evolution of $Y(x)$ and $y(x)$, for the same DM masses as shown in Fig. 5.

text. Here, the resonant annihilation depletes $f_\chi(q)$ for momenta *just above the peak* of the distribution, leading to a relative decrease with respect to a thermal distribution at these momenta, and hence a decrease in the DM velocity dispersion (also known as “temperature”). This effect is visible in Fig. 5 starting with a slight suppression at $q \sim 8$ for the curve with $x = 20$ (note that the relative enhancement at larger values of q is not relevant for our discussion given that f_χ is already highly suppressed here), and results in the decrease in the evolution of y seen in Fig. 6. The latter can also directly be understood from inspection of Eq. (28): in the subresonant regime we have $\langle\sigma v\rangle < \langle\sigma v\rangle_2$, which drives y to smaller values after decoupling (with a strength proportional to Y —which explains why the scattering term $\propto \gamma(T)$ can increase y again, slightly, once the DM abundance has decreased sufficiently). A second effect of this depletion in $f_\chi(q)$ is that $\langle\sigma v\rangle_{\text{neq}}$ decreases, which in turn leads to an earlier chemical decoupling and hence an increased relic density. The difference between the numerical and the coupled Boltzmann approach can in this case thus exclusively be understood as resulting from the slight offset in the $y(x)$ curves during the freeze-out [which in turn results from the fact that the scattering term is not strong enough to maintain an exact Maxwellian shape of $f_\chi(q)$ when the velocity dispersion decreases as explained above].

As we increase the DM mass, we leave the subresonant regime and enter the *resonant regime*, with the transition point marked by the benchmark model with $m_S = 60.5 \text{ GeV}$. We note that this transition is also clearly visible in Fig. 2, as a sharp decrease in the temperature at which the DM velocity dispersion deviates from its equilibrium value. The origin of this feature is *not* an actual delay of kinetic decoupling, but that DM annihilation now starts to deplete $f_\chi(q)$ *below* the peak of the would-be

Maxwellian distribution.⁴ This leads to an increase of the velocity dispersion, once equilibrium is left, rather than a decrease as in the subresonant regime. This effect is very clearly seen in Figs. 5 and 6, up to DM masses at the higher end of this regime, where the influence of the resonance starts to become less important because we have $\sqrt{s} \lesssim m_h + \Gamma_h$ only for DM momenta well below the peak of the phase-space distribution.

In the *superresonant regime* with $m_S \gtrsim m_h/2$, finally, we have necessarily $\sqrt{s} \gtrsim m_h$. A resonantly enhanced annihilation rate is thus only possible for a very small portion of phase space, with almost vanishing relative DM momenta. This implies not only that we always have $\langle\sigma v\rangle > \langle\sigma v\rangle_2$ in this regime, but also that the effect of the resonance rapidly becomes negligible.

Lastly, it is interesting to note that for $\sqrt{s} \gtrsim m_h$ the annihilation rate effectively features a $1/v^2$ velocity dependence. This is similar to resonant Sommerfeld-enhanced annihilation, which leads to a suppressed relic density after a prolonged freeze-out phase [30]. This can clearly be seen in the evolution of $Y(x)$ in Fig. 6, for $m_S \sim m_h/2$, where the differences between the numerical and the coupled Boltzmann approach are mostly due to the late-time differences in $y(x)$ —which in turn come about because of the rather significant differences in $f_\chi(q)$ at large values of x (cf. Fig. 5).

⁴In a similar way, the sharp rise around $m_S \sim 54 \text{ GeV}$ in Fig. 2 should not be interpreted as a feature in the momentum exchange rate $\gamma(T)$. Rather, it can be understood as the point where the shape of the $y(x)$ evolution starts to develop from something close to the one in the top left panel in Fig. 6 into something that is much closer to the one in the top center panel (which in turn is driven by the annihilation terms, as explained in the text). As a result, the temperature at which y departs from y_{eq} increases very quickly as the mass increases beyond this transition point.

APPENDIX B: SEMIRELATIVISTIC KINETIC THEORY

In this Appendix, we discuss how to generalize the highly nonrelativistic elastic scattering term in Eq. (5) to incorporate the most important relativistic corrections needed for the numerical implementation of the full Boltzmann equation. Throughout, we refer to this result as “semirelativistic” scattering.

The starting point is to expand the full collision term C_{el} in small momentum transfer compared to the typical DM momentum—similar to what is done in order to arrive at Eq. (5), but not only keeping lowest-order terms in $\mathbf{p}^2/m_\chi^2 \sim T/m_\chi$. From this, we can derive a Fokker-Planck scattering operator in a relativistic form (for details, see [21]):

$$C_{\text{el}} \simeq \frac{E}{2} \nabla_{\mathbf{p}} \cdot [\gamma(T, \mathbf{p})(E T \nabla_{\mathbf{p}} + \mathbf{p}) f_\chi]. \quad (\text{B1})$$

Being a total divergence, this scattering operator manifestly respects number conservation, as it should. Another important property, which one can directly read off from the part inside the brackets, is that it features a stationary point given by the relativistic Maxwell-Boltzmann distribution,

$$f_\chi^{\text{eq}} \propto e^{-E/T}. \quad (\text{B2})$$

The nonrelativistic limit of Eq. (B1) gives the scattering operator (5), but in this limit the stationary point would instead be the nonrelativistic version $f_\chi^{\text{eq}} \propto \exp[-p^2/(2m_\chi T)]$ —which would cause a problem in the full BE as this does not correspond to the *actual* equilibrium distribution fed into the annihilation term of Eq. (37).

In general, the momentum transfer rate $\gamma(T, \mathbf{p})$ in Eq. (B1) depends on the DM momentum \mathbf{p} . However, the stationary point is independent of γ , which motivates us to restrict

ourselves to the leading order term $\gamma(T) \equiv \gamma(T, \mathbf{0})$, neglecting any momentum dependence, and use the nonrelativistic limit in Eq. (B1) *only* to evaluate the momentum transfer rate $\gamma(T)$ as it appears in Eq. (6). To this order, we could thus also replace the leading E in Eq. (B1) by m_χ ; here, we choose to still keep it as it leads to a much more compact analytical form of the equation governing the DM temperature (see below). Explicitly performing the first partial derivative in C_{el} then leads to the final form of our semirelativistic Fokker-Planck operator as given by Eq. (8). This operator is our default choice for the numerical implementation of the full Boltzmann equation.

As already pointed out in Sec. II C, it is mandatory for the full phase-space calculation to have a scattering operator with a fixpoint that matches the equilibrium distribution of Eq. (B2) assumed in the annihilation term. For the coupled integrated Boltzmann system, on the other hand, this issue is fully addressed by using the relativistic temperature definition of Eq. (21)—rather than its nonrelativistic version typically adopted in the literature in the context of kinetic decoupling—because this automatically leads to the correct fixpoint $T_\chi = T$ for both the semirelativistic Eq. (8) and, to the lowest order, for the nonrelativistic version Eq. (5); see the discussion in Sec. II B.

Another advantage of our semirelativistic Fokker-Planck operator is that the differential equation for T_χ , often quoted when discussing kinetic decoupling, takes a very simple form even beyond the highly nonrelativistic limit. To see this, let us for the moment ignore the impact of annihilations, and take the second moment of the Boltzmann equation with this operator (using the relativistic definition of T_χ). This leads to

$$\dot{T}_\chi + 2 \left(1 - \frac{\langle p^4/E^3 \rangle}{6T_\chi} \right) H T_\chi = \gamma \left[T \left(1 - \frac{5}{6} \langle p^2/E^2 \rangle + \frac{2}{6} \langle p^4/E^4 \rangle \right) - T_\chi \left(1 - \frac{\langle p^4/E^3 \rangle}{6T_\chi} \right) \right], \quad (\text{B3})$$

which of course is equivalent to Eq. (28) in the main text, when neglecting the annihilation terms and implementing the replacement given in Eq. (35). Let us repeat that the rhs of the above equation only takes this particular form with our default choice of the semirelativistic Fokker-Planck term, whereas the moment appearing on the left-hand side is an exact result. This equation is in general not closed in terms of T_χ . However, if we make the ansatz of a Maxwellian DM phase-space distribution, cf. Eq. (34), we get a relation between the different momentum moments,

$$5 \langle p^2/E^2 \rangle - 2 \langle p^4/E^4 \rangle = \langle p^4/E^3 \rangle / T_\chi, \quad (\text{B4})$$

such that the differential equation closes in terms of T_χ . Indeed, introducing

$$2(1-w) \equiv \frac{g_\chi}{3T_\chi n_\chi} \int \frac{d^3 p}{(2\pi)^3} \frac{\mathbf{p}^4}{E^3} f_\chi(\mathbf{p}) = \frac{\langle p^4/E^3 \rangle}{3T_\chi}, \quad (\text{B5})$$

it takes a very simple form:

$$\dot{T}_\chi + 2w(T_\chi) H T_\chi = w(T_\chi) \gamma(T) (T - T_\chi). \quad (\text{B6})$$

This generalizes the highly nonrelativistic result [18], for which $w \rightarrow 1$ and we hence find the familiar scaling $T_\chi \propto T^2$ after kinetic decoupling (i.e. when $\gamma \ll H$). In the ultra-relativistic limit, on the other hand, we have $w(T_\chi) \rightarrow 1/2$ and the likewise familiar scaling of $T_\chi \propto T$ for relativistic particles. We note that in the region $x \gtrsim 10$ relevant for early kinetic decoupling, the correction to the nonrelativistic limit is already sizable; e.g., $w(x=10) \approx 0.8$.

- [1] G. Hinshaw *et al.* (WMAP Collaboration), *Astrophys. J. Suppl. Ser.* **208**, 19 (2013); arXiv:1212.5226.
- [2] P. A. R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **594**, A13 (2016); arXiv:1502.01589.
- [3] G. Bertone, D. Hooper, and J. Silk, *Phys. Rep.* **405**, 279 (2005); arXiv:hep-ph/0404175.
- [4] B. W. Lee and S. Weinberg, *Phys. Rev. Lett.* **39**, 165 (1977).
- [5] P. Hut, *Phys. Lett. B* **69**, 85 (1977).
- [6] K. Sato and M. Kobayashi, *Prog. Theor. Phys.* **58**, 1775 (1977).
- [7] D. A. Dicus, E. W. Kolb, and V. L. Teplitz, *Phys. Rev. Lett.* **39**, 168 (1977); **39**, 973(E) (1977).
- [8] S. Wolfram, *Phys. Lett. B* **82**, 65 (1979).
- [9] P. Gondolo and G. Gelmini, *Nucl. Phys. B* **360**, 145 (1991).
- [10] J. Edsjö and P. Gondolo, *Phys. Rev. D* **56**, 1879 (1997); arXiv:hep-ph/9704361.
- [11] K. Griest and D. Seckel, *Phys. Rev. D* **43**, 3191 (1991).
- [12] T. Bringmann, *New J. Phys.* **11**, 105027 (2009); arXiv:0903.0189.
- [13] R. T. D’Agnolo, D. Pappadopulo, and J. T. Ruderman, *Phys. Rev. Lett.* **119**, 061102 (2017); arXiv:1705.08450.
- [14] V. Silveira and A. Zee, *Phys. Lett. B* **161**, 136 (1985).
- [15] J. McDonald, *Phys. Rev. D* **50**, 3637 (1994); arXiv:hep-ph/0702143.
- [16] C. P. Burgess, M. Pospelov, and T. ter Veldhuis, *Nucl. Phys. B* **619**, 709 (2001); arXiv:hep-ph/0011335.
- [17] E. W. Kolb and M. S. Turner, *Front. Phys.* **69**, 1 (1990).
- [18] T. Bringmann and S. Hofmann, *J. Cosmol. Astropart. Phys.* **04** (2007) 016; **03** (2016) E02; arXiv:hep-ph/0612238.
- [19] J. Kasahara, Neutralino dark matter: The mass of the smallest halo and the golden region, Ph.D. thesis, The University of Utah, 2009.
- [20] P. Gondolo, J. Hisano, and K. Kadota, *Phys. Rev. D* **86**, 083523 (2012); arXiv:1205.1914.
- [21] T. Binder, L. Covi, A. Kamada, H. Murayama, T. Takahashi, and N. Yoshida, *J. Cosmol. Astropart. Phys.* **11** (2016) 043; arXiv:1602.07624.
- [22] T. Bringmann, H. T. Ihle, J. Kersten, and P. Walia, *Phys. Rev. D* **94**, 103529 (2016); arXiv:1603.04884.
- [23] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, *Comput. Phys. Commun.* **149**, 103 (2002); arXiv:hep-ph/0112278.
- [24] P. Gondolo, J. Edsjö, P. Ullio, L. Bergström, M. Schelke, and E. A. Baltz, *J. Cosmol. Astropart. Phys.* **07** (2004) 008; arXiv:astro-ph/0406204.
- [25] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, *Comput. Phys. Commun.* **176**, 367 (2007); arXiv:hep-ph/0607059.
- [26] A. Arbey and F. Mahmoudi, *Comput. Phys. Commun.* **181**, 1277 (2010); arXiv:0906.0369.
- [27] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, *Comput. Phys. Commun.* **185**, 960 (2014); arXiv:1305.0237.
- [28] T. Bringmann, J. Edsjö, P. Gondolo, P. Ullio, and L. Bergström, DarkSUSY 6: An Advanced Tool to Compute Dark Matter Properties Numerically (to be published).
- [29] T. Bringmann *et al.* (GAMBIT Dark Matter Workgroup), arXiv:1705.07920.
- [30] L. G. van den Aarssen, T. Bringmann, and Y. C. Goedecke, *Phys. Rev. D* **85**, 123512 (2012); arXiv:1202.5456.
- [31] M. Duch and B. Grzadkowski, *J. High Energy Phys.* **09** (2017) 159; arXiv:1705.10777.
- [32] J. L. Feng, M. Kaplinghat, and H.-B. Yu, *Phys. Rev. Lett.* **104**, 151301 (2010); arXiv:0911.0422.
- [33] M. R. Buckley and P. J. Fox, *Phys. Rev. D* **81**, 083522 (2010); arXiv:0911.0422.
- [34] J. L. Feng, M. Kaplinghat, and H.-B. Yu, *Phys. Rev. D* **82**, 083525 (2010); arXiv:1005.4678.
- [35] C. E. Yaguna, *J. Cosmol. Astropart. Phys.* **03** (2009) 003; arXiv:0810.4267.
- [36] S. Profumo, L. Ubaldi, and C. Wainwright, *Phys. Rev. D* **82**, 123514 (2010); arXiv:1009.5377.
- [37] C. Arina and M. H. G. Tytgat, *J. Cosmol. Astropart. Phys.* **01** (2011) 011; arXiv:1007.2765.
- [38] Y. Mambrini, *Phys. Rev. D* **84**, 115017 (2011); arXiv:1108.0671.
- [39] R. N. Lerner and J. McDonald, *Phys. Rev. D* **80**, 123507 (2009); arXiv:0909.0520.
- [40] M. Herranen, T. Markkanen, S. Nurmi, and A. Rajantie, *Phys. Rev. Lett.* **115**, 241301 (2015); arXiv:1506.04065.
- [41] F. Kahlhoefer and J. McDonald, *J. Cosmol. Astropart. Phys.* **11** (2015) 015; arXiv:1507.03600.
- [42] S. Profumo, M. J. Ramsey-Musolf, and G. Shaughnessy, *J. High Energy Phys.* **08** (2007) 010; arXiv:0705.2425.
- [43] V. Barger, P. Langacker, M. McCaskey, M. Ramsey-Musolf, and G. Shaughnessy, *Phys. Rev. D* **79**, 015018 (2009); arXiv:0811.0393.
- [44] J. M. Cline and K. Kainulainen, *J. Cosmol. Astropart. Phys.* **01** (2013) 012; arXiv:1210.4196.
- [45] A. Djouadi, O. Lebedev, Y. Mambrini, and J. Quevillon, *Phys. Lett. B* **709**, 65 (2012); arXiv:1112.3299.
- [46] K. Cheung, Y.-L. S. Tsai, P.-Y. Tseng, T.-C. Yuan, and A. Zee, *J. Cosmol. Astropart. Phys.* **10** (2012) 042; arXiv:1207.4930.
- [47] M. Endo and Y. Takaesu, *Phys. Lett. B* **743**, 228 (2015); arXiv:1407.6882.
- [48] A. Djouadi, A. Falkowski, Y. Mambrini, and J. Quevillon, *Eur. Phys. J. C* **73**, 2455 (2013); arXiv:1205.3169.
- [49] J. M. Cline, K. Kainulainen, P. Scott, and C. Weniger, *Phys. Rev. D* **88**, 055025 (2013); **92**, 039906(E) (2015); arXiv:1306.4710.
- [50] A. Urbano and W. Xue, *J. High Energy Phys.* **03** (2015) 133; arXiv:1412.3798.
- [51] X.-G. He and J. Tandean, *J. High Energy Phys.* **12** (2016) 074; arXiv:1609.03551.
- [52] M. Escudero, A. Berlin, D. Hooper, and M.-X. Lin, *J. Cosmol. Astropart. Phys.* **12** (2016) 029; arXiv:1609.09079.
- [53] A. Goudelis, Y. Mambrini, and C. Yaguna, *J. Cosmol. Astropart. Phys.* **12** (2009) 008; arXiv:0909.2799.
- [54] N. Craig, H. K. Lou, M. McCullough, and A. Thalailil, *J. High Energy Phys.* **02** (2016) 127; arXiv:1412.0258.
- [55] H. Han, J. M. Yang, Y. Zhang, and S. Zheng, *Phys. Lett. B* **756**, 109 (2016); arXiv:1601.06232.
- [56] P. Ko and H. Yokoya, *J. High Energy Phys.* **08** (2016) 109; arXiv:1603.04737.
- [57] A. Beniwal, F. Rajec, C. Savage, P. Scott, C. Weniger, M. White, and A. G. Williams, *Phys. Rev. D* **93**, 115016 (2016); arXiv:1512.06458.
- [58] A. Cuoco, B. Eiteneuer, J. Heisig, and M. Krämer, *J. Cosmol. Astropart. Phys.* **06** (2016) 050; arXiv:1603.08228.

- [59] P. Athron *et al.* (GAMBIT Collaboration), [arXiv:1705.07908](#).
- [60] P. Athron *et al.* (GAMBIT Collaboration), *Eur. Phys. J. C* **77**, 568 (2017); [arXiv:1705.07931](#).
- [61] C. Patrignani *et al.* (Particle Data Group), *Chin. Phys. C* **40**, 100001 (2016).
- [62] S. Dittmaier *et al.* (LHC Higgs Cross Section Working Group), [arXiv:1101.0593](#).
- [63] D. Boyanovsky, H. J. de Vega, and D. J. Schwarz, *Annu. Rev. Nucl. Part. Sci.* **56**, 441 (2006); [arXiv:hep-ph/0602002](#).
- [64] M. Drees, F. Hajkarim, and E. R. Schmitz, *J. Cosmol. Astropart. Phys.* **06** (2015) 025; [arXiv:1503.03513](#).
- [65] J. B. Dent, S. Dutta, and R. J. Scherrer, *Phys. Lett. B* **687**, 275 (2010); [arXiv:0909.4128](#).
- [66] J. Zavala, M. Vogelsberger, and S. D. M. White, *Phys. Rev. D* **81**, 083502 (2010); [arXiv:0910.5221](#).
- [67] J. L. Feng, M. Kaplinghat, H. Tu, and H.-B. Yu, *J. Cosmol. Astropart. Phys.* **07** (2009) 004; [arXiv:0905.3039](#).
- [68] B. von Harling and K. Petraki, *J. Cosmol. Astropart. Phys.* **12** (2014) 033; [arXiv:1407.7874](#).
- [69] F. D'Eramo and J. Thaler, *J. High Energy Phys.* **06** (2010) 109; [arXiv:1003.5912](#).
- [70] F. D'Eramo, N. Fernandez, and S. Profumo, *J. Cosmol. Astropart. Phys.* **05** (2017) 012; [arXiv:1703.04793](#).
- [71] K. Redmond and A. L. Erickcek, *Phys. Rev. D* **96**, 043511 (2017); [arXiv:1704.01056](#).
- [72] Y. Hochberg, E. Kuflik, T. Volansky, and J. G. Wacker, *Phys. Rev. Lett.* **113**, 171301 (2014); [arXiv:1402.5143](#).
- [73] Y. Hochberg, E. Kuflik, H. Murayama, T. Volansky, and J. G. Wacker, *Phys. Rev. Lett.* **115**, 021301 (2015); [arXiv:1411.3727](#).
- [74] E. Kuflik, M. Perelstein, N. R.-L. Lorier, and Y.-D. Tsai, *Phys. Rev. Lett.* **116**, 221302 (2016); [arXiv:1512.04545](#).