Early kinetic decoupling effect on the forbidden dark matter annihilations into standard model particles

Yu Liu,* Xuewen Liu,† and Bin Zhu‡

Department of Physics, Yantai University, Yantai 264005, China

(Dated: June 14, 2023)

The early kinetic decoupling (eKD) effect is an inevitable ingredient in calculating the relic density of dark matter (DM) for various well-motivated scenarios. It appears naturally in forbidden dark matter annihilation, the main focus of this work, which contains fermionic DM and a light singlet scalar that connects the DM and standard model (SM) leptons. The strong suppression of the scattering between DM and SM particles happens quite early in the DM depletion history, where the DM temperature drops away from the thermal equilibrium, \( T_\chi < T_{SM} \), leading to the decreased kinetic energy of DM. The forbidden annihilation thus becomes inefficient since small kinetic energy cannot help exceed the annihilation threshold, naturally leading to a larger abundance. To show the eKD discrepancy, we numerically solve the coupled Boltzmann equations that govern the evolution of DM number density and temperature. It is found that eKD significantly affects the DM abundance, resulting in almost an order of magnitude higher than that by the traditional calculation. We also discuss the constraints from experimental searches on the model parameters, where the viable parameter space shrinks when considering the eKD effect.

I. INTRODUCTION

Relic density is an essential topic for dark matter (DM) physics. The classical scenario that explains the observed abundance in the present Universe is thermal particle production in the early Universe, which is the so-called freeze-out mechanism for the weakly interacting massive particle (WIMP) [1–5]. The DM particles were initially in thermal equilibrium with the heat bath via the intense interactions among them. The number density dilutes along with the expansion of the Universe and finally freezes out of the heat bath once the annihilation rate falls behind the cosmic expansion rate, resulting in a comoving constant.

Gondolo et al. [6, 7] have developed the renowned treatment of calculating the DM relic density by solving the Boltzmann equation of the number density with high accuracy, called the “standard” method. One hypothesis entering this treatment is that DM keeps in local kinetic equilibrium with the thermal plasma during or even after the freeze-out process. The scatterings with the standard model (SM) particles have been at a much more intense level [8].

However, this is not always true for many well-motivated mechanisms, where kinetic equilibrium might decouple earlier than assumed, leading to the early kinetic decoupling (eKD) around the freeze-out period. The eKD effect has been extensively studied in the literature [9–20]. The most influential scenarios include 1) resonant annihilation of dark matter [10, 13, 14], 2) Sommerfeld-enhanced annihilation [14], and 3) subthreshold annihilation (also known as the forbidden annihilation) [14]. In these regimes, the eKD occurs because the elastic scattering processes are suppressed; DM and the SM particles experience different temperatures. All these cases demonstrated the actual DM abundance can be affected by up to 1 order of magnitude compared to the traditional method, at least in some parts of the parameter space.

Here, we study the eKD in the forbidden DM scenario [21]. In such a scenario, DM dominantly annihilates into heavier final states, which can proceed at finite temperatures in the early Universe, relying on the thermal tail with the high velocity of DM. Many works were devoted to studying the forbidden annihilations with different theoretical models and a variety of phenomenological topics [22–28]. In this work, we employ the model where forbidden DM annihilations into the SM leptons are mediated by a singlet scalar. Such channels were studied in Ref.[23], which is experimentally viable and predicts a very narrow mass range for DM, that can be tested at future beam dump experiments. However, the eKD effect has not been studied in the context of this model.

Actually, the scatterings of DM against SM particles are strongly suppressed in the early times because of the mass splitting between the DM and SM leptons. The DM temperature drops away from the thermal bath temperature, \( T_\chi < T_{SM} \), which leads to the decreased kinetic energy of DM. The forbidden annihilations thus become inefficient since small kinetic energy cannot help exceed the annihilation threshold. So this naturally leads to a larger abundance and the eKD effects in such a model should not be neglected. On the contrary, eKD will cause significant impacts on DM abundance. We investigate the relic density beyond the standard treatment used in Ref. [23] by considering the coupled Boltzmann differential equations, where the temperature evolution of the dark sector could be taken into account. We use the public code DRAKE [14] to perform the numerical calculations. We find a DM relic density that differs by up to an order of magnitude from the standard treatment and leaves a reduced feasible parameter space under the various experimental constraints.

The rest of this paper is structured as follows. In Sec. II, we start with a general description of the coupled Boltzmann equations that govern the evolution of DM number density and temperature. In addition, we discuss the DM model for forbidden annihilations and analyze the occurrence of early kinetic decoupling. Section III is devoted to a thorough study

---

*Electronic address: liuyu2020@s.ytu.edu.cn
†Electronic address: xuewenliu@ytu.edu.cn
‡Electronic address: zhubin@mail.nankai.edu.cn
of DM relic density for the forbidden channels, and makes a
detailed comparison between the new treatment and the tra-
tional one. We further discuss various constraints from col-
lider searches and astrophysical observations on the parameter
space in Sec. IV. Finally, we conclude in Sec. V.

II. EARLY KINETIC DECOUPLING EFFECTS ON
FORBIDDEN ANNIHILATIONS

A. Basic formulas

Keeping kinetic equilibrium during and even after the
freeze-out epoch is one underlying assumption for traditional
relic density calculations. However, this is not always the
case for various scenarios, where kinetic decoupling happens
earlier than the chemical decoupling process. To study the
DM relic density by taking into account the early kinetic de-
coupling effect, we should consider the following Boltzmann
equation for DM phase-space distribution [10, 14]

\[ E \left( \frac{\partial}{\partial t} - H \hat{p} \frac{\partial}{\partial \hat{p}} \right) f_{\chi}(t, \hat{p}) = C_{\text{ann}}[f_{\chi}] + C_{\text{el}}[f_{\chi}], \]  

where \( E \) is the energy of the DM, \( H \) is the Hubble constant,
\( \hat{p} \) is the momentum of DM, and \( f_{\chi} \) is the DM phase-space
density. The collision term \( C_{\text{ann}} \) represents the annihilation
of DM particles into thermal bath particles, and \( C_{\text{el}} \) is for elastic
scattering processes between DM and SM scattering partners.
For two-body processes,

\[
C_{\text{ann}} = \frac{1}{2g_{\chi}} \sum B' \int \frac{d^{3}p'}{(2\pi)^{3}2E_{p'}} \int \frac{d^{3}k}{(2\pi)^{3}2E_{k}} \int \frac{d^{3}k'}{(2\pi)^{3}2E_{k'}} \times (2\pi)^{4} \delta^{4}(p + p' - k - k') \times (-|M_{\chi\chi \rightarrow B\phi}|^{2}f_{\chi}(\hat{p})f_{\chi}(\hat{p}')(1 \pm f_{\text{eq}}^{\chi}(\vec{k}))(1 \pm f_{\text{eq}}^{\phi}(\vec{k}') - |M_{\phi\phi \rightarrow B\chi}|^{2}f_{\phi}(\hat{p}'\bar{\phi})(1 \pm f_{\text{eq}}^{\chi}(\vec{k}))(1 \pm f_{\text{eq}}^{\phi}(\vec{k}))),
\]

\[
C_{\text{el}} = \frac{1}{2g_{\chi}} \sum B \int \frac{d^{3}p'}{(2\pi)^{3}2E_{p'}} \int \frac{d^{3}k}{(2\pi)^{3}2E_{k}} \int \frac{d^{3}k'}{(2\pi)^{3}2E_{k'}} \times (2\pi)^{4} \delta^{4}(p + p' - k - k') \times (-|M_{\chi\phi \rightarrow B\chi}|^{2}f_{\chi}(\hat{p})f_{\phi}(\hat{p}')(1 \pm f_{\text{eq}}^{\chi}(\vec{k}))(1 \pm f_{\text{eq}}^{\phi}(\vec{k}')) + |M_{\phi\phi \rightarrow B\chi}|^{2}f_{\phi}(\hat{p}'\bar{\phi})(1 \pm f_{\text{eq}}^{\chi}(\vec{k}))(1 \pm f_{\text{eq}}^{\phi}(\vec{k}))),
\]

where \( B \) and \( B' \) stand for particles in the thermal bath such as
SM leptons, \( g_{\chi} \) is the number of internal degrees of freedom
of DM, and \( f_{\text{eq}} \) is given by the Fermi-Dirac or Bose-Einstein
distribution depending on the spin of \( B \). The summation should
be taken for all the internal degrees of freedom for all the par-
ticles. For the nonrelativistic DM, \( C_{\text{el}} \) can be simplified as the
Fokker-Planck operator [8, 29–31]:

\[
C_{\text{el}} \approx \frac{E}{2} \gamma(T) \left[ T E \partial_{\hat{p}}^{2} + (2TE + p + pT E) \partial_{\hat{p}} + 3 \right] f_{\chi}. \tag{4}
\]

In the above, the momentum transfer rate \( \gamma(T) \) is given by (see
also Ref. [32])

\[
\gamma = \frac{1}{3g_{\chi}m_{\chi} T} \int \frac{d^{3}k}{(2\pi)^{3}} f_{\text{eq}}^{\chi}(E_{k}) \left[ 1 \mp f_{\text{eq}}^{\chi}(E_{k}) \right] \int_{-4\pi^{2}E_{k}}^{0} \frac{d\tau}{d\tau} v, \tag{5}
\]

where the differential cross section can be expressed as
\( (d\sigma/d\tau)v \equiv \left| M_{i\rightarrow f,\chi} \right|^{2} / (64\pi k E m_{\chi}^{2}) \), and \( k_{\text{cm}}^{2} \) is given by

\[
k_{\text{cm}}^{2} = \frac{m_{\chi}^{2} - m_{Y}^{2}}{m_{\chi}^{2} + m_{\phi}^{2} + 2m_{\phi} E_{k}}. \tag{6}
\]

Here \( E_{k} \) is the energy of heat bath particle \( B \). Note that \( k_{\text{cm}}^{2} \not= E_{k}^{2} - m_{B}^{2} = |k|^2 \).

During the chemical decoupling, the scattering processes
may not be frequent enough to maintain the kinetic equilib-
rium, which means that DM particles own a dynamical degree of
freedom. By integrating Eq. (1) with \( g_{\chi} \int \frac{d^{3}p}{(2\pi)^{3}2E} = g_{\chi} \int \frac{d^{3}p}{(2\pi)^{3}2E} \), one obtains the zeroth and second
moments of the Boltzmann equation, respectively. This
leads to a relatively simple coupled system of Boltzmann differen-
tial equations (denoted as the cBE method hereafter),

\[
\begin{align*}
\frac{Y'}{Y} &= \frac{X}{Y} \frac{Y_{\text{eq}}^{2}}{X_{\text{eq}}^{2}} \left( \langle \sigma v \rangle_{T_{1}} - \langle \sigma v \rangle_{T_{2}} \right), \\
\frac{Y'}{Y} &= \frac{1}{xH} \langle C_{\text{el}} \rangle_{2} + \frac{X}{Y} \frac{Y_{\text{eq}}^{2}}{X_{\text{eq}}^{2}} \left( \langle \sigma v \rangle_{T_{1}} - \langle \sigma v \rangle_{T_{2}} \right),
\end{align*}
\]

where \( x \) is defined as usual \( x = m_{\chi} / T \) and \( Y_{\text{eq}}(x) \equiv n_{\text{eq}}(T) / s. \)

\( H \equiv H \left[ 1 + (1/3) d(\log g_{\text{eff}}^{2}) / d(\log T) \right], \) with \( g_{\text{eff}}^{2} \)

\footnote{As pointed out by Ref. [14], for forbidden DM, the Fokker-Planck approx-
imation is not that accurate, but the dominant cKD effect on the relic den-
sity can in many cases still be fairly well captured by the Fokker-Planck approx-
imation. For more precise treatment, we leave it to future work.}
being the entropy degrees of freedom of the background plasma. \(w(T_x) \equiv 1 - \langle p^4 / E^3 \rangle r_x / (6 T_x)\), with \(\langle p^4 / E^3 \rangle = g_\nu \int d^3 p (\mathbf{p} \cdot \mathbf{p})^3 / E^3 e^{-E / T} \). Note that the elastic scattering term given in Eq. (4) does not contribute to the zeroth moment term. This is a natural consequence because the elastic scattering processes do not change the number density of DM.

The above compact form of the differential equations contains the following thermally averaged cross sections,

\[
(C_{el})_2 \equiv \frac{g_\nu}{3 n(T_x)} \int \frac{d^3 \mathbf{p} \, p^2}{(2 \pi)^3 E^2} C_{el},
\]

\[
(\sigma v)_T \equiv \frac{g_\nu^2}{(n_e^0)^2} \int \frac{d^3 \mathbf{p}}{(2 \pi)^3} \int \frac{d^3 \mathbf{q}}{(2 \pi)^3} \frac{\mathbf{p} \cdot \mathbf{p}}{3 E^2} (\sigma v)_{\chi \chi \rightarrow BB}
\]

\[
\times f_x^q(\beta, T_x) f_x^q(\mathbf{q}, T_x),
\]

The thermal average \((\sigma v)_T\) is a variant of the commonly used thermal average \((\sigma v)_T\), and is explicitly stated in Ref. [10] and introduced as

\[
(\sigma v)_T \equiv \langle \sigma v \rangle_{T_x} = \frac{g_\nu^2}{(n_e^0)^2} \int \frac{d^3 \mathbf{p}}{(2 \pi)^3} \int \frac{d^3 \mathbf{q}}{(2 \pi)^3} \frac{\mathbf{p} \cdot \mathbf{p}}{3 E^2} (\sigma v)_{\chi \chi \rightarrow BB}
\]

\[
\times f_x^q(\beta, T_x) f_x^q(\mathbf{q}, T_x). \quad (11)
\]

For \((\sigma v)_T\) and \((\sigma v)_T^2\), replace \(T_x\) by \(T\) in \((\sigma v)_T\) and \((\sigma v)_T^2\), respectively. And \(n_{eq}^0(T_x)\) is given by

\[
n_{eq}^0(T_x) = g_\nu \int \frac{d^3 \mathbf{p}}{(2 \pi)^3} f_x^q(\beta, T_x) = g_\nu \int \frac{d^3 \mathbf{p}}{(2 \pi)^3} e^{-E / T}. \quad (13)
\]

In this work, we use the numerical routine DRAKE to solve the coupled Boltzmann equations. The measured value of \(\Omega h^2\) by the Planck Collaboration is \(\Omega h^2 = 0.120 \pm 0.001 [33]\). The viable parameter space is determined by matching this value.

### B. Model and discussion on forbidden channels

We have adopted a simple model that only takes into account DM annihilations into SM leptons. The DM is a Dirac fermion coupled to the SM sector via the scalar portal \(\phi\). After the electroweak symmetry breaking, the effective Lagrangian can be written as

\[
-L \supset g_{ij} \bar{l}_i l_j + g_{ij}^A \bar{\tau}_i \gamma_5 \tau_j + g_{x} \bar{\chi} \phi \chi + g_{x}^A \bar{\chi} \gamma_5 \phi \chi. \quad (14)
\]

where the indices on the couplings \(i, j = e, \mu, \tau\). This model has been studied thoroughly for the forbidden mechanism (for detail, refer to Ref. [23]). The merits include the following:

1) DM mass is limited in quite a small window close to the masses of the SM leptons, which is a strong prediction that can be tested soon by colliders or beam-dump experiments.

2) Kinematically forbidden DM naturally evade the stringent constraints from the energy injections into the cosmic microwave background (CMB) [21]. In the forbidden DM scenario, the energy injection processes suffer the Boltzmann suppression at \(T \leq \text{eV}\), so that sub-GeV thermal relics are consistent with the experiment, making annihilations to SM leptons with DM masses \(m_\chi \ll 10 \text{ GeV}\) still viable.

In this scenario, the DM relic density should be carefully scrutinized, as the eKD effect appears generic. The actual relic density receives a significant correction compared with the conventional method, as shown in the following sections.

Following Refs. [21, 23], we consider a pair of DM particles dominantly annihilating into two SM particles \(2 \ell\) with mass \(m_\ell < m_{\chi}\). Cosmological constraints make forbidden annihilations into electrons unfeasible, including big bang nucleosynthesis (BBN) and CMB [34]. So, for simplicity, we only consider the \(\mu^+ \mu^-\) and\( \tau^+ \tau^-\) channels, with abbreviated couplings \(g_{\mu}^{(A)}, g_{\tau}^{(A)}, \) and \(g_{\chi}^{(A)}\), which allows us to explore all the relevant DM phenomenology systematically.

Applying the detailed balance condition for the DM number-changing process, the cross section of the forbidden channels is exponentially suppressed,

\[
(\sigma v) = (\sigma v)(n_{eq}^0)^2 / (\Omega h^2)^2 \approx (\sigma v)e^{-2\Delta x}, \quad (15)
\]

where \(\Delta \equiv (m_\ell - m_\chi) / m_{\chi}. \) \(\sigma v\) is the cross section for the inverse process. When the annihilation rate becomes slower than the Hubble expansion, DM is no longer in equilibrium with the SM thermal bath, resulting in chemical decoupling.

What about the scattering between DM and SM particles during this period? From Eq. (5), the momentum transfer rate \(\gamma\) is proportional to an exponential factor:

\[
\gamma(x) \propto e^{-\Delta x}. \quad (16)
\]

The full expression of the rates is listed in Appendix A. In the forbidden scenario, \(\Delta > 1\) implies the scattering frequency experiences a strong suppression at a much earlier period.

It is known that DM kinetically decouples out of the SM thermal bath as long as the momentum transfer rate \(\gamma\) is smaller than the Hubble expansion, \(\gamma < H\). For illustration, we show the momentum transfer rate \(\gamma\) for the \(\chi \mu^+ \rightarrow \chi \mu^-\) scattering process, in Fig. 1. In comparison, we plot the Hubble parameter \(H(x)\) as a function of \(x\). The red lines stand for the evolution of \(\gamma(x)\) and \(H(x)\) in the forbidden DM case where we take \(m_{\chi} = 0.1\text{GeV}\). To demonstrate the distinctiveness of the forbidden DM, we also provide the results of a nonforbidden case where DM mass is larger than that of the annihilations products \(m_\ell = 1\text{ GeV}\). The most remarkable finding is that \(\gamma\) of the forbidden case becoming comparable with \(H(x)\) happens much earlier than that of the nonforbidden case. The kinetic decoupling starts at around \(x = 20\), which is usually the same time as DM chemically decoupled from the thermal bath. The reason for this very early kinetic decoupling is straightforward to understand as the result of an exponential suppressed momentum transfer rate, as derived in Eq. (16). One can obtain similar results for the \(\chi \tau^+ \rightarrow \tau^+ \tau^-\) forbidden channel. We conclude that eKD exists in the forbidden DM scenario.

The next step is to find the effect of the eKD more concretely. We systematically study DM relic density in both eBE and traditional methods (denoted as nBE as in Ref. [14]) and then discuss the phenomenological possibilities.
our study to the same range of $m$. The corresponding DM mass is $0.0$. $0$ scattering, and comparing with Hubble constant $H(x)$. The red lines set a rough but straight comparison. The values of the rest annihilations and the momentum transfer rates for the forbidden cases of $\mu/\tau$ are known as the resonance region, where $2m_\chi < m_x$. Here, the eKD effect for the forbidden cases of $\chi \chi \rightarrow \mu^+\mu^-$ and $\chi \chi \rightarrow \tau^+\tau^-$, where we define the deviations of the relic density in $cBE$ and nBE approaches by Deviation $\equiv \Omega_{nBE} - \Omega_{cBE}/\Omega_{cBE}$. We display the results in the $(m_\chi, \mu/\tau)$ plane by the density plotting method, with setting $m_x = 0.1$GeV and $m_\chi = 1.48$GeV for illustration. The sizeable deviations appear in most parameter spaces from 20% up to almost 100%. The maximum deviations seem to emerge in the resonance region. However, it should be noted that eKD effects already exist in the resonant annihilations of DM [10, 13, 14]. So in this region, one should study the eKD for resonance and forbidden channels together. It is beyond the scope of our work, as we mainly focus on the forbidden annihilations.

To emphasize the importance of the improved treatment of the decoupling history near the mass threshold, we plot in Fig. 4 the ratio of the resulting relic density to that of the standard nBE approach. Here the parameters satisfy the requirement for $\Omega_{nBE}h^2 = 0.12$. The different choices of coupling $g_\phi^\mu$ correspond to different curves, as labeled in the plots. The mass ratio $m_\chi/m_\mu$ can be divided into three regions. At lower mass ratios, dark matter evolves as ordinary WIMPs where kinetic equilibrium is maintained. The relic density derived by the $cBE$ and nBE methods agrees with each other. The gray-shaded region is known as the resonance region, where $2m_\chi = m_x$. Here, the eKD effect arises due to the distinct cooling and heating effects of dark matter. For a more detailed discussion of the origin of these features, please refer to Ref.[10].

In the forbidden DM mass region (brown-shaded), we can see that the $cBE$ results are larger than nBE several times with the same parameters. For $\mu/\tau$ case, $3 \leq \Omega_{nBE}/\Omega_{cBE} \leq 10$, and $2 \leq \Omega_{nBE}/\Omega_{cBE} \leq 15$ for the $\tau^+\tau^-$ forbidden case.

As the mass ratio $m_\mu/m_\tau$ increases, the coupling $(g_{\mu/\tau})$, which is needed to obtain the correct relic density, rises.
TABLE I: Benchmark points, selected for $\chi\bar{\chi} \to \mu^+\mu^-$ and $\chi\bar{\chi} \to \tau^+\tau^-$ channels. We set $g_{\tau} = 0$ and $g_{\mu}^s = 1.121$ for all cases as in Ref. [23] which can make a direct comparison.

| Benchmark ($\mu^+\mu^-$) | $m_{\chi}$ | $m_{\phi}$ | $g_{\mu}$ | $g_{\mu}^s$ | $\Omega_{\chi BE} h^2 / \Omega_{\chi nBE} h^2$ |
|--------------------------|-----------|------------|-----------|-----------|----------------------------------|
| BP1                      | 0.1 GeV   | 0.26 GeV   | 0         | 0         | 0.12/0.018                       |
| BP2                      | 0.1 GeV   | 0.26 GeV   | 0.00091757| 0         | 12/0.0186                       |
| BP3                      | 0.1055 GeV| 0.3 GeV    | 0.00145082| 0         | 0.12/0.036                      |
| BP4                      | 0.1055 GeV| 0.3 GeV    | 0.00026751| 0         | 0.12/0.064                      |

| Benchmark ($\tau^+\tau^-$) | $m_{\chi}$ | $m_{\phi}$ | $g_{\tau}$ | $g_{\tau}^s$ | $\Omega_{\chi BE} h^2 / \Omega_{\chi nBE} h^2$ |
|-----------------------------|------------|------------|------------|-------------|----------------------------------|
| BPR1                        | 1.6 GeV    | 5 GeV      | 1.09261    | 0           | 0.12/0.0088                     |
| BPR2                        | 1.6 GeV    | 5 GeV      | 0.268806   | $g_{\tau}$  | 0.12/0.0078                     |
| BPR3                        | 1.77 GeV   | 6 GeV      | 0.0565632  | 0           | 0.12/0.032                      |
| BPR4                        | 1.77 GeV   | 6 GeV      | 0.0096236  | $g_{\tau}$  | 0.12/0.058                      |

FIG. 3: Density distribution of the deviation between $c_{BE}$ and $n_{BE}$ approaches, which is defined as $(\Omega_{\chi cBE} h^2 - \Omega_{\chi nBE} h^2) / \Omega_{\chi cBE} h^2$. The left panel is for the $\chi\bar{\chi} \to \mu^+\mu^-$ channel and the right is for the $\chi\bar{\chi} \to \tau^+\tau^-$ channel; we take $m_{\chi} = 0.1$ GeV and $m_{\phi} = 1.48$ GeV for the left and right panels respectively.

FIG. 4: Relic density comparison between $\Omega_{\chi BE} h^2$ and $\Omega_{\chi nBE} h^2$ for the $\chi\bar{\chi} \to \mu^+\mu^-$ and $\chi\bar{\chi} \to \tau^+\tau^-$ channels. Here, we choose the parameters to match the observed abundance in the $n_{BE}$ approach. The left diagram corresponds to the $\mu^+\mu^-$ mode, and we take $m_{\phi} = 0.26$ GeV, and the right diagram presents the $\tau^+\tau^-$ mode, for which we take $m_{\phi} = 5$ GeV.
rapidly (see Fig. 5). We find the upper bounds of the mass ratio, beyond which the couplings become non-perturbative (i.e., $g_{\mu/m} < \sqrt{4\pi}$). Since a larger coupling is required in the cBE approach, a smaller mass ratios are allowed compared to the standard nBE treatment.

![FIG. 5: Correct relic density as function of mass ratio and coupling $g_{\mu/m}$ in both nBE and cBE approaches. We take $m_\phi = 0.26$ GeV for the $\mu^+\mu^-$ mode, and $m_\phi = 5$ GeV for the $\tau^+\tau^-$ mode.](image)

Speaking overall, DM relic density in the cBE method is larger than that by using the standard nBE method. When the elastic scattering is strongly suppressed, the temperature of DM particles drops below that of the thermal bath ($T_\chi < T_{\text{SM}}$), meaning DM particles do not have enough kinetic energy. The forbidden annihilation thus becomes ineffective due to the inability to overcome the annihilation threshold, leading to a larger abundance in cBE treatment.

IV. PARAMETER SPACE AND EXPERIMENTAL CONSTRAINTS

In this section, we will find the feasible parameter space for the cBE approach in the $(m_\phi, g_{\mu/m})$ plane, by requiring the correct DM relic density. The viable parameter space further displays more accurate results compared with nBE. Of course, there are numerous constraints on the model parameters that are imposed by the collider searches and the astrophysical observations across a wide range.

A. $\chi\chi \rightarrow \mu^+\mu^-$

We first study the $\mu^+\mu^-$ forbidden channel. The numerical results are shown in Fig. 6. We set $g_\phi^A = 0$ for the left panel, while $g_\phi^f = g_\phi$ for the right panel. The different choices of the couplings to muons do not strongly affect the phenomenology.

To elaborate on our findings, we start by searching the parameter boundary of the forbidden annihilation, in which $\Delta \rightarrow 0$ and the right relic density are fulfilled. The green line in the plots shows the boundary of the cBE approach; the green shaded region, denoted by the expression $\Delta(cBE) \leq 0$, represents the unforbidden space. In the remaining parameter space, where the forbidden annihilations dominate the DM depletion, $m_\chi$ is chosen at each point to match the correct relic density. The required coupling becomes larger when $m_\mu/m_\chi > 1$, leading to the corresponding curves lying inside the funnel area with a similar shape as the boundary. For comparison, we also repeated the parameter boundary of the standard nBE approach in the same figure, as shown by the gray lines. The gray regions stand for the non-forbidden regions.

The allowed parameter space of the two approaches is noticeably distinctive. The eKD effect reduces the parameter space, compared to that of nBE. Or we can say that to satisfy the relic density requirement, the larger coupling $g_\mu$ is required for the cBE scenario. The reason is straightforward as already pointed out in the last section. Around the freeze-out stage, the temperature of DM decreases resulting in reduced kinetic energy and then inefficient forbidden annihilations. To maintain the DM dilution process, a larger coupling is required for $T_\chi < T$.

In the small mediator mass region, the annihilation into pairs of mediators dominates the relic density. We depict it in gray with the label $\chi\chi \rightarrow \phi\phi$ as the non-forbidden DM region. For experimental constraints, the most crucial parameter is the mediator mass $m_\phi$. In the following, we show the constraints one by one:

- Planck (brown region).
- DM annihilations into SM electromagnetically interacting particles could modify the anisotropies of the CMB [47–49]. The measurements of the CMB by the Planck satellite[33] can thus put robust constraints on such energy injection processes [35]. In this model, the photon pairs can be produced in DM annihilations via a muon loop, leading to the injection into the CMB. We recast the corresponding constraints from Ref. [23] which are shown in brown.

As pointed out in Ref. [23], when $m_\phi$ near $2m_\chi$ which is also close to but smaller than $2m_\mu$, the annihilation cross section of $\chi\chi \rightarrow \gamma\gamma$ is enhanced due to $\sigma(\chi\chi \rightarrow \gamma\gamma) \sim 1/[(m_\gamma^2 - 4m_\chi^2)^2 + m_\chi^2\Gamma_\chi^2]$. So the CMB constraints exclude the most parameter space for $m_\phi < 2m_\mu$ and only leave a small part of the allowed room when $m_\phi > 2m_\mu$.

From Fig. 6, we find that the eKD effects narrow down the allowed parameter space that obtains the right amount of DM relic abundance. Especially the CMB constraints exclude the most parameter space with cBE in the pure scalar interaction scenario (i.e. $g_\mu^f = 0$).

- E137 (for orange region) and BDX (blue region).

Secondary muons are produced from the electron beam-dump experiments, such as SLAC E137 [36] and Jefferson Lab BDX experiments[40], which can be used to explore the signals of light scalar emission and the muon-scalar coupling via muon-nucleon scattering pro-
FIG. 6: Results and constraints on forbidden channel $\chi \bar{\chi} \rightarrow \mu^+ \mu^-$. The green line represents the parameter boundary of the forbidden DM region with $m_{\mu}/m_{\chi} = 1$; and the green shaded region stands for the non-forbidden space that $m_\tau > m_\mu$. The rest space is allowed for the forbidden annihilations to get the correct DM relic density. The gray region labeled $\Delta(nBE) \leq 0$ is obtained in the nBE approach. Also, the non-forbidden space that is dominated by the $\chi \bar{\chi} \rightarrow \phi \phi$ process is shown. The brown region displays the Planck bounds on the energy injection process [35]. The orange region is excluded by the E137 electron beam-dump experiment [36]. The purple region is excluded by energy loss process of SN1987A [37–39]. The projected sensitivities for future beam-dump experiments from BDX [40] (light blue), $M^1$ [41, 42] (dashed blue line), NA62 [43] (red), and NA64-$\mu$ [44–46] (dashed magenta line) are also shown here.

cess $\mu + N \rightarrow \mu + N + \phi$. The E137 experiment’s null result established exclusion limits on the parameter space, and the upcoming BDX experiment can likewise yield a predicted limit.

In Fig. 6, we display, as shaded areas, constraints from the E137 electron beam-dump experiment [50], and projections from BDX [40]. The constraints exclude part of the available parameter space in the $(m_\phi - g_\mu (g^1_\mu))$ plane.

- NA62 (red shaded region).

NA62 is a fixed-target experiment at the CERN Super Proton Synchrotron (SPS) that is dedicated to measurements of kaon rare decays, including projected searches on $K \rightarrow \mu \nu \phi$ [43]. Such a decay channel is an excellent probe of new light scalars that couples preferentially to muons. Ref. [51] has derived the probe sensitivity for this process, which can be used to test our parameter space, as shown in red.

- NA64-$\mu$ and $M^1$ (magenta and blue dashed lines).

Similar to the above beam-dump experiments, NA64-$\mu$ [44–46] and $M^1$ [41, 42] are designed to search light scalars in the muon-nucleon scattering process $\mu N \rightarrow \mu N \phi$, using muon beams. It’s worth noting that the allowed parameter space can be tested in the coming future.

- SN 1987A (purple region).

At last, for the parameter space involving new light scalars, we should also consider the constraints from the observation of supernovae cooling. The most famous constraints arise from the energy loss process via the Primakoff effect $\gamma p \rightarrow p \phi$ in SN1987A [37–39]. The excluded region is displayed in purple.

B. $\chi \bar{\chi} \rightarrow \tau^+ \tau^-$

The forbidden annihilation in the di-tau case is discussed here. Annihilation to $\tau^+ \tau^-$ shares several qualitative features with annihilations to muons. The same computations are performed, including the searches for the forbidden annihilation parameter region, the comparison of cBE and nBE treatments, and the variety of limitations from experimental searches. The results are displayed in Fig. 7 with different parameter settings. Note that the consideration of the experimental constraints is rather different from the $\chi \bar{\chi} \rightarrow \mu^+ \mu^-$ case due to the different mediator mass.

- BaBar (orange region) and Belle II (red region)

Searching for $e^+ e^- \rightarrow \gamma + \text{invisible}$ at the $e^+ e^-$ colliders, such as BaBar and the future Belle II experiments, sets existed and projected constraints on the parameter space. As depicted in Fig. 7, the orange line and region show the constraints arising from BaBar [39, 44, 52], and the red region shows the exquisite sensitivity from Belle II using $50 \text{ ab}^{-1}$ integrated luminosity [39, 54], which indicates that a large portion of the viable parameter space will be tested.

- LEP (brown region) and Tera Z (purple region)

Precision measurements of $Z$ boson decay width can place constraints on the model parameters, namely on the exotic $Z$ decays. In this scenario, $Z \rightarrow \tau^+ \tau^- \phi$ ($\phi \rightarrow \text{invisible}$) contributed to the measured $Z \rightarrow \tau^+ \tau^-$ width.
FIG. 7: Results and constraints on forbidden channel $\chi \bar{\chi} \rightarrow \tau^+ \tau^-$. The green and gray shaded regions have the same meaning as in Fig. 6, and we use mass ratio $m_\tau/m_\chi = 1$. We show constraints from searches at BaBar for $e^+e^- \rightarrow \phi \gamma$ [39, 44, 52] (orange region), searches at LEP for $Z \rightarrow \tau\tau + \text{MET}$ [44, 53] (brown region), and Planck constraints on DM annihilations (blue). We include projections for Belle II [39, 54] (red region) and future $Z$-factories [55–58] (purple region).

Large Electron Positron Collider (LEP) has set corresponding limits on such channels [44, 53] which are shown in brown.

Additionally, there have been several proposals for future $Z$-factories to search the same processes [55–58], based on Circular Electron Positron Collider (CEPC) and the Future Circular Collider $e^+e^-$ (FCC-ee) for instance. The projections of Tera-$Z$ options (with accumulated $10^{12} Z$’s ) provide leading sensitivities in the tens GeV range, shown in purple.

- Planck (blue region)

  The CMB constraints are similar with the $\chi \bar{\chi} \rightarrow \mu^+ \mu^-$ forbidden channel. The parameter space at $m_\phi < 2m_\tau$ is much constrained as always, as shown in Fig. 7, due to the enhancement of $\chi \bar{\chi} \rightarrow \gamma \gamma$ when $m_\phi \approx 2m_\tau$.

V. CONCLUSION

In the forbidden DM scenario, consideration of the early kinetic decoupling is not only a correction to the DM relic density but an indispensable ingredient. In this work, we investigate the early kinetic decoupling effect in forbidden channels, where DM annihilation to SM leptons is kinetically forbidden. Specifically we focus on the $\chi \bar{\chi} \rightarrow \mu^+ \mu^-$ and $\chi \bar{\chi} \rightarrow \tau^+ \tau^-$ modes. By analyzing the scattering momentum transfer rate, we found the kinetic equilibrium breaks at about $x \approx 20$, which is the same stage of chemical decoupling. So eKD should be taken seriously into account. With different benchmark points, we found that there is a cooling phase during the evolution that causes the DM temperature to deviate from the thermal bath and evolve solely. The decreased kinetic energy of DM particles suppresses the forbidden annihilation rate, which gives rise to larger abundances. The difference between the cBE and traditional nBE methods is significant, within part of the parameter space showing a deviation up to an order of magnitude larger, for both $\mu^+ \mu^-$ and $\tau^+ \tau^-$ forbidden channels. We also considered the experimental constraints from beam-dump experiments, collider searches, and astrophysical observations. The viable parameter space in the forbidden DM model has been reduced when using the cBE treatment. Most of the parameter space will be tested by the forthcoming experimental searches.

ACKNOWLEDGEMENTS

We thank Murat Abdughani for the helpful discussions. This work was supported by the National Natural Science Foundation of China under Grants No. 12005180 and No. 12275232 and by the Natural Science Foundation of Shandong Province under Grant No. ZR2020QA083.
Appendix A: Cross sections for annihilation and scattering processes

The annihilation cross section for $\chi\bar{\chi} \rightarrow t^+t^-$ is

$$\sigma = \left(\frac{\sqrt{s - 4m_t^2(g_t^A s + g_t^A(s - 4m_t^2))}(g_t^A s + g_t^2(s - 4m_t^2))}{16\pi s \sqrt{s - 4m_t^2((s - m_\phi^2)^2 + m_\phi^2\Gamma_\phi^2)}}\right),$$

(A1)

where the total decay rate of scalar $\phi$ is calculated as

$$\Gamma_\phi = \frac{1}{16\pi m_\phi}\sqrt{1 - \frac{4m_t^2}{m_\phi^2}} \times$$

$$((g_t - g_t^A)^2(m_\phi^2 - 2m_t^2) + (g_t^A + g_t)(m_\phi^2 - 2m_t^2) - 4m_t^2(g_t^A + g_t)(g_t - g_t^A))$$

(A2)

$$+ \frac{1}{16\pi m_\phi}\sqrt{1 - \frac{4m_t^2}{m_\phi^2}} \times$$

$$((-g_t^A - g_t)^2(m_\phi^2 - 2m_t^2) + (g_t^A - g_t)(m_\phi^2 - 2m_t^2) + 4m_t^2(g_t^A - g_t)(g_t^A + g_t))$$

For the elastic scattering process $\chi\mu^+ \rightarrow \chi\mu^+$, the amplitude is

$$|M|^2 = \frac{(g_t^A t + g_t^2(t - 4m_t^2))(g_t^A t + g_t^2(t - 4m_t^2))}{(m_\phi^2 - t)^2}.$$  

(A3)

And the momentum transfer rate can be written as,

$$\gamma = \frac{1}{3g_t^2m_\phi^2T} \int \frac{d^3k}{(2\pi)^3} f^2_{g_t}(E_k) \left[1 + f^2_{g_t}(E_k)\right] \int_{-4k_e}^{0} dt(-t) \frac{d\sigma}{dt},$$

(A4)

with

$$\int_{-4k_e}^{0} dt(-t) \frac{d\sigma}{dt} = \frac{1}{64\pi k E x m_\phi^2} \int_{-4k_e}^{0} dt(-t)|M|^2 = \frac{1}{64\pi k E x m_\phi^2} \times tAmp,$$

(A5)

and

$$tAmp = 4k^2m_\phi^2 \left(\frac{2(g_t^A + g_t^2)(g_t^A + g_t^2)k^2m_\phi^2}{(m_\phi^2 + m_t^2 + 2m_x E_k)^2} + \frac{4(g_t^A + g_t^2)g_t^2m_\phi^2 + 4g_t^2(g_t^A + g_t^2)m_\phi^2 - 2(g_t^A + g_t^2)(g_t^A + g_t^2)m_\phi^2}{m_\phi^2 + m_t^2 + 2m_x E_k} + \frac{4k^2m_\phi^2 - (g_t^A + g_t^2)m_\phi^2)(-4g_t^2m_\phi^2 + (g_t^A + g_t^2)m_\phi^2)}{4k^2m_\phi^2 + m_t^2 + 2m_x E_k} \right)$$

$$\left(16g_t^2g_t^2m_t^2m_\phi^2 - 8(g_t^A + g_t^2)g_t^2m_\phi^2 + g_t^2(g_t^A + g_t^2)m_\phi^2 m_\phi^2 + 3(g_t^A + g_t^2)(g_t^A + g_t^2)m_\phi^2 \left[\log m_\phi^2 - \log(m_\phi^2 + \frac{4k^2m_\phi^2}{m_\phi^2 + m_t^2 + 2m_x E_k})\right] \right).$$

(A6)

[1] B. W. Lee and S. Weinberg, Phys. Rev. Lett. 39, 165 (1977).
[2] P. Hut, Phys. Lett. B 69, 85 (1977).
[3] K. Sato and M. Kobayashi, Prog. Theor. Phys. 58, 1775 (1977).
[4] D. A. Dicus, E. W. Kolb, and V. L. Teplitz, Phys. Rev. Lett. 39,
[5] S. Wolfram, Phys. Lett. B 82, 65 (1979).
[6] P. Gondolo and G. Gelmini, Nucl. Phys. B 360, 145 (1991).
[7] J. Edsjo and P. Gondolo, Phys. Rev. D 56, 1879 (1997), arXiv:hep-ph/9704361.
[8] T. Bringmann, New J. Phys. 11, 105027 (2009), arXiv:0903.0189 [astro-ph.CO].
[9] L. G. van den Aarssen, T. Bringmann, and Y. C. Goedecke, Phys. Rev. D 85, 123512 (2012), arXiv:1202.5456 [hep-ph].
[10] T. Binder, T. Bringmann, M. Gustafsson, and A. Hryczuk, Phys. Rev. D 96, 115010 (2017), Erratum: Phys.Rev.D 101, 099901 (2020), arXiv:1706.07433 [astro-ph.CO].
[11] F. Brümmer, JHEP 01, 113 (2020), arXiv:1910.01549 [hep-ph].
[12] K. Ala-Mattinen and K. Kainulainen, JCAP 09, 040 (2020), arXiv:1912.08270 [hep-ph].
[13] T. Abe, Phys. Rev. D 102, 035018 (2020), arXiv:2004.10041 [hep-ph].
[14] T. Binder, T. Bringmann, M. Gustafsson, and A. Hryczuk, Eur. Phys. J. C 81, 577 (2021), arXiv:2103.01944 [hep-ph].
[15] B. Zhu and X. Liu, Sci. China Phys. Mech. Astron. 65, 231011 (2022), arXiv:2104.03238 [hep-ph].
[16] A. Hryczuk and M. Laletin, JHEP 06, 026 (2021), arXiv:2104.05684 [hep-ph].
[17] T. Abe, Phys. Rev. D 104, 035025 (2021), arXiv:2106.01956 [hep-ph].
[18] Y. Du, F. Huang, H.-L. Li, Y.-Z. Li, and J.-H. Yu, JCAP 04, 012 (2022), arXiv:2111.01267 [hep-ph].
[19] K. Ala-Mattinen, M. Heikinheimo, K. Kainulainen, and K. Tuominen, Phys. Rev. D 105, 123005 (2022), arXiv:2201.06456 [hep-ph].
[20] A. Hryczuk and M. Laletin, Phys. Rev. D 106, 023007 (2022), arXiv:2204.07078 [hep-ph].
[21] R. T. D’Agnolo and J. T. Ruderman, Phys. Rev. Lett. 115, 061301 (2015), arXiv:1505.07107 [hep-ph].
[22] A. Delgado, A. Martin, and N. Raj, Phys. Rev. D 95, 035002 (2017), arXiv:1608.05345 [hep-ph].
[23] R. T. D’Agnolo, D. Liu, J. T. Ruderman, and P.-J. Wang, JHEP 06, 103 (2021), arXiv:2012.11766 [hep-ph].
[24] G. N. Wojcik and T. G. Rizzo, JHEP 04, 033 (2022), arXiv:2109.07369 [hep-ph].
[25] K.-C. Yang, JHEP 11, 083 (2022), arXiv:2209.10827 [hep-ph].
[26] Y. Cheng, S.-F. Ge, X.-G. He, and J. Sheng, (2022), arXiv:2211.05643 [hep-ph].
[27] J. Herrms, S. Jana, V. P. K., and S. Saad, Phys. Rev. Lett. 129, 091803 (2022), arXiv:2203.05579 [hep-ph].
[28] J. Herrms, S. Jana, V. P. K., and S. Saad, in 14th International Workshop on the Identification of Dark Matter 2022 (2022) arxiv:2209.14882 [hep-ph].
[29] T. Binder, L. Covi, A. Kamada, H. Murayama, T. Takahashi, and N. Yoshida, JCAP 11, 043 (2016), arXiv:1602.07624 [hep-ph].
[30] E. Bertschinger, Phys. Rev. D 74, 063509 (2006), arXiv:astro-ph/0607319.
[31] T. Bringmann and S. Hofmann, JCAP 04, 016 (2007), Erratum: JCAP 03, E02 (2016), arXiv:hep-ph/0612238.
[32] P. Gondolo, J. Hisano, and K. Kadota, Phys. Rev. D 86, 083523 (2012), arXiv:1205.1914 [hep-ph].