Journal of Cosmology and Astroparticle Physics 2023.21 (2023): 004

DOI: https://doi.org/10.1088/1475-7516/2023/02/004

Prepared for submission to JCAP
IFT-UAM/CSIC-XXX, LUPM:22-006, ULB-TH/22-06

Analytical insight into dark matter subhalo boost factors for Sommerfeld-enhanced s- and p-wave γ-ray signals

Gaétan Facchinetti,[a] Martin Stref,[b] Thomas Lacroix,[c] Julien Lavalle,[d] Judit Pérez-Romero,[c] David Maurin,[e] and Miguel A. Sánchez-Conde[c]

[a]Service de Physique Théorique, Université Libre de Bruxelles, Boulevard du Triomphe, CP225, 1050 Brussels, Belgium

[b]Laboratoire d'Annecy-le-Vieux de Physique Théorique (LAPTh), Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, F-74000 Annecy, France

[c]Instituto de Física Teórica (IFT) UAM/CSIC, Universidad Autónoma de Madrid, Calle Nicolás Cabrera, 13-15, 28049 Madrid, Spain

[d]Laboratoire Univers et Particules de Montpellier (LUPM), Université de Montpellier & CNRS, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France

[e]Laboratoire de Physique Subatomique & Cosmologie (LPSC), Université Grenoble Alpes, CNRS/IN2P3, 53 avenue des Martyrs, 38026 Grenoble, France

E-mail: gaetan.facchinetti@ulb.be, martin.stref@univ-smb.fr, thomas.lacroix@uam.es, lavalle@in2p3.fr, judit.perez@uam.es, david.maurin@lpsc.in2p3.fr, miguel.sanchezconde@uam.es

Abstract. As searches for thermal and self-annihilating dark matter (DM) intensify, it becomes crucial to include as many relevant physical processes and ingredients as possible to refine signal predictions, in particular those which directly relate to the intimate properties of DM. We investigate the combined impact of DM subhalos and of the (velocity-dependent) Sommerfeld enhancement of the annihilation cross section. Both features are expected to play an important role in searches for thermal DM particle candidates with masses around or beyond TeV, or in scenarios with a light dark sector. We provide a detailed analytical description of the phenomena at play, and show how they scale with the subhalo masses and the main Sommerfeld parameters. We derive approximate analytical expressions that can be used to estimate the overall boost factors resulting from these combined effects, from which the intricate phenomenology can be better understood. DM subhalos lead to an increase of the Sommerfeld effect by several orders of magnitude (for both s- and p-wave annihilation processes), especially on resonances, which makes them critical to get sensible gamma-ray signal predictions for typical targets of different masses (from dwarf galaxies to galaxy clusters).

Keywords: Dark matter searches, Gamma rays, structure formation

ArXiv ePrint: 2203.16491

Contents

1 Introduction

Gamma-ray astronomy, and more generally multimessenger astronomy, provides powerful probes of thermally produced non-asymmetric particle dark matter (DM), in particular scenarios in which DM self-annihilation proceeds through s-wave processes[1] [2–5] — we generically refer to this kind of scenarios as the weakly-interacting massive particle (WIMP) paradigm [6–9]. Searches in the local universe are nicely complemented by early-universe probes, for instance those deriving from analyses of the cosmic-microwave background (CMB) radiation [10]. Current constraints disfavor DM particle masses below ~ 50 GeV with canonical cross sections annihilating into a variety of standard-model final states [11]. Indirect searches, namely the searches for DM annihilation or decay signals in astrophysical probes, are sensitive to parts of the parameter space usually hidden to direct searches (and vice versa), i.e. searches for DM particle collisions onto nuclear targets in underground detectors,

[1] In the non-relativistic regime that prevails since before chemical decoupling, the DM annihilation cross section can usually be expanded in powers of $v^2 = v_{\rm rel}^2/4$, which directly relates to an expansion in partial waves [1]. Accordingly, we generically refer to s-wave processes as those giving a velocity-independent annihilation rate, while p-wave processes come with a $\langle v^2\rangle$-dependent annihilation rate.

due to the different velocity dependencies arising when Feynman diagrams are rotated from annihilation to elastic collision — for instance, a p-wave annihilation into quarks can usually be efficiently probed by direct searches. Hence, deepening the exploration on both fronts is the best way to validate or exclude the thermal DM scenario.

As observations and experiments have now entered the thermal DM parameter space, and as theoretical modeling improves, it becomes important to refine predictions in order to explore further non-trivial, though very interesting, corners of theory space. One of such corners implies particle models in which DM self-interacts through long-range forces, which may lead to what is called the Sommerfeld enhancement of the annihilation cross section [12–20]. It is basically triggered when the interaction range becomes larger than the spread of a DM-particle pair wavefunction. This typically occurs when there is a large mass hierarchy between the DM particle and interaction mediators, which is rather generic for multi-TeV DM, but can also be present on more general grounds in case of relatively light dark sectors. Recent examples can be found in, e.g., refs. [9, 21].

The Sommerfeld enhancement of the annihilation cross section is similar to gravitational focusing as it depends on inverse powers of the relative speed, $v_{\rm rel}^{-n}$, where n is an integer that will be specified later. Since the DM dispersion velocity relates to the mass of the self-gravitating halo, one can naturally expect that DM subhalos, the tiniest DM structures expected in the universe [22–25], could be the sites for the largest enhancements. Of course, as we will see, that enhancement may saturate below some characteristic velocity inherent to the properties of DM particles and self-interactions, but subhalos still play a very important role in setting the overall annihilation signal predictions. In this paper, we restrict to DM annihilation into gamma rays, and our calculations are made in such a way that they can be applied to a diversity of DM targets in gamma-ray astronomy, from dwarf galaxies to galaxy clusters.

Several references have actually already addressed the Sommerfeld enhancement in specific DM systems with subhalos, e.g. [26–33] (see also refs. [34–38]). The goal of this paper is rather to expand upon these works and to provide a more complete and generic analytical understanding of the intricate processes at play, applicable to all targets and covering the full (though simplified) parameter space relevant to the Sommerfeld effect. In particular, we will see that the role of subhalos is critical both on Sommerfeld resonances, and in the case of very large mass hierarchy between the DM particle and the interaction mediator. In some cases, the exploratory power of different gamma-ray targets (e.g., dwarf galaxies vs. galaxy clusters) could even be inverted, which points to new interesting complementary ways to constrain the Sommerfeld parameter space.

The paper develops as follows. In Sec. 2, we present our general reasoning in simple physical terms, which will pave the way to our more technical discussion in the following sections, and already unveil some of the main results. In Sec. 3, we introduce the velocity dependencies of the general problem in more technical terms. We start by characterizing the Sommerfeld enhancement in Sec. 3.1, and then introduce the velocity-dependent J-factor in Sec. 3.2, which defines the amplitude of the gamma-ray signal for a given DM halo target. We further recall how one can get reasonable description of the phase-space distribution functions in self-gravitating DM halos. In Sec. 4, we extensively discuss how subhalos enter the game and affect the overall predictions, before concluding in Sec. 5. A companion paper [39] explores in details the consequences of Sommerfeld enhancement and subhalo boosts on specific gamma-ray targets from dwarf galaxies to galaxy clusters, based on the full numerical calculation.

2 A glimpse of the main results

Before digging into the technical aspects of the work, it is useful to summarize them in more simple terms. DM subhalos are well-known boosters of DM annihilation signals; see e.g. refs. [40–42] for their effects in indirect DM searches with different messengers, and e.g. ref. [43] for a review. For velocity-independent s-wave processes, this is a mere consequence of $\langle\rho^2\rangle \geq \langle\rho\rangle^2$, where ρ is the local DM density in a target host halo, which turns to a full inequality thanks to DM inhomogeneities such as subhalos. In that case, given the mass dependence of the signal for one object, the overall contribution is simply obtained by convolving this mass-dependent signal with the subhalo mass function. The latter derives from structure formation theory, and can be approached from both analytical considerations and cosmological simulations. If the signal associated with a single halo of mass m is scale invariant and proportional to m^β (which, we will show, is a reasonable approximation), and if the subhalo mass function scales like $m^{-\alpha}$, with both $\alpha, \beta > 0$, then the signal integrated over the subhalo mass range is simply $\phi_{\rm sub} \propto m^{-\alpha_{\rm eff}}$, with an effective index $\alpha_{\rm eff} = \alpha - \beta - 1$. One readily sees that depending on the sign of $\alpha_{\rm eff}$, the signal will be dominated either by the light mass boundary of the integral (i.e., many small objects), or the heavy one (i.e., fewer massive objects). The former case generically leads to a stronger subhalo *boost factor*, which is a measure of the ratio $\langle\rho^2\rangle_V/\langle\rho\rangle_V^2$ in the relevant volume V, and which characterizes the annihilation signal enhancement due to DM inhomogeneities. The amplitude of this boost factor is linked to that of $\alpha_{\rm eff}$, and to the mass hierarchy between the host halo mass and the minimal subhalo mass, the latter being linked to the interaction properties of DM. This holds for s-wave annihilation processes.

For p-wave processes, no subhalo enhancement is expected because the p-wave suppression factor, proportional to $\langle v^2\rangle$, is even more severe in subhalos inside which bound DM particles must have a smaller dispersion velocity not to escape. This argument makes it straightforward to guess that, in scenarios in which subhalos would represent a significant fraction of the total mass, the overall p-wave signal could actually even be further subhalo-suppressed. On the other hand, it is also obvious that bigger halos, with larger dispersion velocities, will lead to larger global annihilation rates.

The Sommerfeld enhancement of the annihilation cross section strongly affects the above statements, because it is itself velocity dependent, with a different dependence between s- and p-wave processes. Then two questions arise: (i) Since the Sommerfeld effect is local by nature, how does it scale at the level of a full object? (ii) How does it propagate over a population of objects? The main complication comes from the fact that the Sommerfeld effect behaves differently depending on whether the relative DM de Broglie wavelength is greater (saturation regime) or lower (Coulomb regime) than the DM self-interaction range, a transition which therefore depends on DM velocity. That specific transition is actually fixed by particle physics independently of any astrophysics, and can therefore be predicted rather accurately (at least in simplified particle DM models). Moreover, the reasoning made just above for Sommerfeld-free p-wave processes indicates a possible way: although (sub)halos are featured by spatially-dependent velocity distribution functions, which need to be integrated over to properly describe the Sommerfeld distortion of the annihilation cross section, one could still hope to capture the net effect from a typical velocity for each halo, which would then be related to its mass. If this typical velocity is a scale-invariant function of the halo mass, for instance $\overline{v} \propto m^\nu$, then we can apply the same recipe as above. The transition between the two Sommerfeld regimes occurs at a specific velocity, $\tilde{v}_{\rm sat}$, entirely fixed by particle physics, which can itself be translated into a specific halo mass, which we denote $\tilde{m}_{\rm sat}$. If the Sommerfeld enhancement scales locally like v^{-s_1} in one of its regime, and like v^{-s_2} in the other one, then this converts into a global scaling like $\overline{v}^{-s_1} \propto m^{-\nu\, s_1}$, say for $m > \tilde{m}_{\rm sat}$ (Coulomb regime), and $\overline{v}^{-s_2} \propto m^{-\nu\, s_2}$ for $m < \tilde{m}_{\rm sat}$ (saturation regime) — s_1 and s_2 may take different values for s- and p-wave processes. In the same vein, the velocity dependence associated with the p-wave "bare" cross section scales like $\overline{v}^2 \propto m^{2\nu}$. To figure out whether subhalos are susceptible to increase the signal, one needs to determine the overall velocity dependence of the Sommerfeld-corrected cross section (including the p-wave suppression). Basically, if $s_i > 0$ (or $s_i - 2 > 0$ for p-wave processes), where $i \in \{1, 2\}$, then the Sommerfeld-corrected cross section will be larger with decreasing subhalo mass, which will make them increase the signal.

By integrating this Sommerfeld-corrected cross section times the "bare" mass-dependent signal ($\propto m^{\beta}$) over a power-law subhalo mass function of index α, and assuming that the transition mass $\tilde{m}_{\rm sat}$ lies within the subhalo mass range defined by $[m_{\rm min}, m_{\rm max}]$, then we can readily infer two different contributions: one scaling like $m^{-\alpha_1}$, with $\alpha_1 = \alpha + s_1 - \beta - 1$, and the other one scaling like $m^{-\alpha_2}$, with $\alpha_2 = \alpha + s_2 - \beta - 1$ (with an additional factor of $m^{2\nu}$ for p-wave processes, and the corresponding change in the associated α_1 and α_2). The most contributing boundary will be either $\tilde{m}_{\rm sat}$ or $m_{\rm max}$ in the first regime, and either $\tilde{m}_{\rm sat}$ or $m_{\rm min}$ in the second regime, depending on the signs of the α's.

As a first important insight, we see that for p-wave processes, the mass hierarchy induced by the p-wave suppression factor can be fully compensated in the Sommerfeld-corrected case because globally, that suppression factor will simply disappear. The consequence in terms of target hierarchy is expected to be quite significant, as we shall see in more details below.

This paper will enter the technical details of this general program which turns out to work reasonably well, and which allows to derive fully analytical results — our main results for the Sommerfeld-enhanced subhalo contribution to the signal are given in Eqs. (4.57)-(4.58). These (i) can help understand the different scaling with the different parameters at play (Sommerfeld particle physics parameters vs. cosmological or astrophysical subhalo parameters), and (ii) provide decent quantitative approximations to the more involved numerical calculations used in the companion paper [39].

3 Velocity-dependent annihilation: theoretical ingredients

Calculations of Sommerfeld-enhanced signal predictions consist in scaling the velocity dependence of a single annihilation process up to an ensemble of particle annihilations proceeding over an entire halo. Here, we first introduce the velocity dependence of the Sommerfeld enhancement at the level of a pair of DM particles, before describing its integration over a DM halo along the line of sight.

3.1 Dark matter annihilation and self-interaction: the Sommerfeld effect

3.1.1 Conventional formulation

We start by shortly introducing the most important features of the Sommerfeld enhancement to the DM annihilation cross section that we are going to use throughout this paper. A slightly more extended introduction can be found in App. A. We assume a simplified model in which DM particles χ can self-interact through multiple exchanges of a single light mediator ϕ of mass m_ϕ, with a coupling $g_\chi = \sqrt{4\pi\alpha_\chi}$, where α_χ plays the role of a dark fine structure constant. If the interaction range, $1/m_\phi$, is larger than the DM Bohr radius, $1/(\alpha_\chi m_\chi)$ (close to the Compton length $1/m_\chi$), then the wave function of the two DM-particle system can be distorted. This rather generically leads to a non-perturbative enhancement of the annihilation cross section called the Sommerfeld effect (we restrict ourselves to attractive interactions), which can often be effectively described by means of a Yukawa potential. This enhancement depends on the relative velocity between the DM particles and saturates when the DM de Broglie wavelength roughly exceeds the interaction range. More detailed descriptions of this phenomenon can be found in, e.g., refs. [12–20].

The Sommerfeld enhancement factor $\hat{\mathcal{S}}$ allows one to correct for this effect and applies to the nominal annihilation cross section as follows [44]:

$$\sigma v_{\rm rel} = (\sigma v_{\rm rel})_0 \times \hat{\mathcal{S}}\,, \tag{3.1}$$

where the subscript 0 refers to the cross section as commonly computed from perturbation theory. Working in natural units ($\hbar = c = 1$) and expressing velocities in units of the speed of light c from now on, it turns useful to introduce the following dimensionless parameters,

$$\begin{aligned} \epsilon_v &\equiv \frac{v}{\alpha_\chi} \\ \epsilon_\phi &\equiv \frac{m_\phi}{\alpha_\chi m_\chi}\,, \end{aligned} \tag{3.2}$$

where ϵ_ϕ roughly expresses the ratio of the Bohr radius of a pair of DM particles, $2/(\alpha_\chi m_\chi)$, to the interaction range, $1/m_\phi$, with $\epsilon_\phi \lesssim 1$ indicating the possible onset of the Sommerfeld enhancement. On the other hand, the ratio ϵ_v/ϵ_ϕ roughly characterizes the ratio of the interaction range to the DM de Broglie wavelength. When this ratio gets < 1, then the long-range interaction is seen as finite again by the quantum system and the Sommerfeld effects saturates. The DM particle speed v featuring above stands for half the relative speed of the pair, $v_{\rm rel}/2$, and c is the speed of light. Parameters ϵ_ϕ and ϵ_v fully characterize the Sommerfeld parameter space in our simple model, and encode the relevant properties of particles and interactions in the dark sector.

Here we focus on s-wave and p-wave annihilation processes. We recall the more tractable expressions of the Sommerfeld enhancement factor for these two cases, which can be derived from the general analytical solution obtained for the Hulthén potential approximation—see App. A. For an s-wave annihilation process, the enhancement factor can be written in a simple form as [17]

$$\begin{aligned} \hat{\mathcal{S}}_s(v,\epsilon_\phi) &\underset{[\text{Yukawa}\to\text{Hulthén}]}{\simeq} \frac{\pi}{\epsilon_v}\frac{\sinh\left(\frac{2\pi\epsilon_v}{\epsilon_\phi^*}\right)}{\cosh\left(\frac{2\pi\epsilon_v}{\epsilon_\phi^*}\right)-\cos\left(2\pi\sqrt{\frac{1}{\epsilon_\phi^*}-\frac{\epsilon_v^2}{\epsilon_\phi^{*2}}}\right)} \\ &\xrightarrow[\left[\epsilon_v\gg\sqrt{\epsilon_\phi^*}\right]]{[\text{Coulomb limit}]} \frac{\pi/\epsilon_v}{1-e^{-\pi/\epsilon_v}}\,, \end{aligned} \tag{3.3}$$

where

$$\epsilon_\phi^* \equiv \pi^2\epsilon_\phi/6\,. \tag{3.4}$$

The first line in the expression of the Sommerfeld factor in Eq. (3.3), i.e. the standard result for the Hulthén potential in the literature, provides a good approximation to the exact result (Yukawa potential) all over the parameter space characterized by $\epsilon_v, \epsilon_\phi \lesssim 1$ (taking cos $\to$ cosh when $\epsilon_v > \sqrt{\epsilon_\phi^*}$) [17, 19]. On the other hand, when $\epsilon_v \gg \sqrt{\epsilon_\phi^*}$ (second line), it boils down to the result obtained assuming the Coulomb potential, $V_C(r) = -\alpha_\chi/r$. The condition $\epsilon_v \gg \sqrt{\epsilon_\phi^*}$ is sufficient but not necessary to ensure that the Coulomb-limit expression in Eq. (3.3) is accurate; for example, in the intermediate case where $\epsilon_\phi^* < \epsilon_v < \sqrt{\epsilon_\phi^*}$ and $\epsilon_v, \epsilon_\phi^* \ll 1$, the enhancement is also well-approximated by the Coulomb-limit expression [17, 28].

For a p-wave annihilation process, the enhancement factor reads instead

$$\hat{\mathcal{S}}_p(v,\epsilon_\phi) = \frac{\left(\frac{1}{\epsilon_\phi^*} - 1\right)^2 + 4\frac{\epsilon_v^2}{\epsilon_\phi^{*2}}}{1 + 4\frac{\epsilon_v^2}{\epsilon_\phi^{*2}}} \times \hat{\mathcal{S}}_s(v,\epsilon_\phi). \tag{3.5}$$

Different regimes arise according to the values of the dimensionless parameters ϵ_v and ϵ_ϕ:

- **Large velocity,** $\epsilon_v \gg 1$ or **heavy mediator,** $\epsilon_\phi \gg 1$
  There is no enhancement in that case: $\hat{\mathcal{S}}_s \approx 1$ and $\hat{\mathcal{S}}_p \approx 1$.

- **Intermediate velocities,** $\epsilon_\phi \ll \epsilon_v \ll 1$
  Here, we have $\hat{\mathcal{S}}_s \approx \pi/\epsilon_v \propto 1/v$ and $\hat{\mathcal{S}}_p \approx \pi/(4\epsilon_v^3) \propto 1/v^3$. This contains the regime in which the Yukawa potential tends to a Coulomb potential (for $\epsilon_v \gg \sqrt{\epsilon_\phi}$) but spans a broader range of values of ϵ_v.

- **Small velocities,** $\epsilon_v \ll \epsilon_\phi \ll 1$
  This corresponds to the saturation regime of the Sommerfeld effect for which

$$\hat{\mathcal{S}}_s(v,\epsilon_\phi) \approx \frac{12}{\epsilon_\phi}\frac{1}{1 + \frac{2\pi^2\epsilon_v^2}{\epsilon_\phi^{*2}} - \cos\left(2\pi\sqrt{1/\epsilon_\phi^*}\right)} \tag{3.6}$$
$$\text{and} \quad \hat{\mathcal{S}}_p(v,\epsilon_\phi) \approx \left(\frac{1}{\epsilon_\phi^*} - 1\right)^2 \hat{\mathcal{S}}_s(v,\epsilon_\phi),$$

  which is almost independent of the velocity of the DM particles, except at a series of resonances for

$$\epsilon_\phi = \epsilon_\phi^{\mathrm{res},n} \equiv 6/(\pi^2 n^2), \tag{3.7}$$

  with n an integer, for which $\hat{\mathcal{S}}_s \approx 1/(n^2\epsilon_v^2) \propto \epsilon_\phi^{\mathrm{res}}/v^2$ and $\hat{\mathcal{S}}_p \approx \left(n^2-1\right)^2/(n^2\epsilon_v^2) \propto 1/(\epsilon_\phi^{\mathrm{res}}v^2)$.

These analytical formulations match with the full numerical results within 10%, except on resonances where larger differences are found. This comes from the fact that the Hulthén potential approximation slightly offsets the solution from the one obtained with the Yukawa potential [18, 34]. However, for the purpose of this work, the features of the solution, comprising resonances, are sufficiently well accounted for by the analytical solution.

We note that the Sommerfeld factor neglects bound-state decay in the low-velocity regime, which leads to nonphysically large enhancements on resonances, where DM bound states can form, that violate the partial-wave unitarity limit. Consequently, the factorization in Eq. (3.1) is expected to fail at vanishing velocities, $\epsilon_v \ll \epsilon_\phi$. Actually, DM bound states have a finite lifetime, which induces a saturation of the enhancement at $v \approx \alpha_\chi^3 m_\phi/m_\chi$ [13, 18, 20], corresponding to $v \approx \alpha_\chi^4$ at resonances. As a result, a slightly modified version of Eq. (3.1) holds, with the nonphysical divergences regularized by replacing v by $v+\alpha_\chi^4$ [18]. We emphasize that this is only an approximate parametric regularization expected to capture reasonably well the relevant physical effects in the current study — for a more detailed description and discussion, see ref. [20]. If $\alpha_\chi \ll 1$, bound-state effects mostly restrict to resonances [34]. We can therefore consider a benchmark value of $\alpha_\chi = 10^{-2}$, though generalized Sommerfeld corrections can be easily rescaled simply by shifting the values of ϵ_ϕ and ϵ_v.

3.1.2 A practical ansatz

The practical reasoning we develop in this part is general and will turn useful when expressing the Sommerfeld enhancement at the level of a full DM halo. The important aspect is to correctly describe the velocity dependence of the Sommerfeld enhancement. In contrast to our formal definition of the Sommerfeld factor in Sec. 3.1.1, here we absorb the v^2 dependence of the p-wave annihilation cross section into our effective definition of the Sommerfeld enhancement factor. To be specific, we introduce the following effective Sommerfeld enhancement factor:

$$\mathcal{S}(v,\epsilon_\phi) \cong \left(\frac{v}{\tilde{v}_{\rm max}}\right)^p \times \hat{\mathcal{S}}(v,\epsilon_\phi)\,, \tag{3.8}$$

where $\hat{\mathcal{S}}$ is the exact Sommerfeld factor introduced in the previous paragraph, v is still half of the relative speed of the pair of DM particles, $\tilde{v}_{\rm max}$ is a reference speed that will be defined later, and

$$p = \begin{cases} 0 & \text{for } s\text{-wave annihilation} \\ 2 & \text{for } p\text{-wave annihilation} \end{cases}. \tag{3.9}$$

Accordingly, the p-wave annihilation cross section can be expressed as:

$$(\sigma\, v)_{p\text{-wave}} = \sigma_p^0\,(2v)^2 \times \hat{\mathcal{S}}(v,\epsilon_\phi) \simeq \sigma_p^0\,(2\tilde{v}_{\rm max})^2 \times \mathcal{S}(v,\epsilon_\phi)\,,$$

where σ_p^0 is the amplitude of the p-wave cross section. The extra factor of 2 is due to the fact that $v = v_{\rm rel}/2$ in our convention. This alternative form implies that there is no longer any velocity dependence in the reference cross section associated with the p-wave case. It is fully transferred to the effective Sommerfeld factor. Consequently, the effective p-wave Sommerfeld factor now scales like $1/v$ in the Coulomb regime (exactly like in the s-wave case), exhibits no speed dependence on resonances, and scales like v^2 between resonances in the saturation regime, as we shall review below. This redefinition allows us to introduce a unique ansatz for both s- and p-wave annihilation processes, and will further make the analytical estimate of the Sommerfeld-corrected subhalo boost factor much simpler to derive and to understand.

We now introduce a simplifying ansatz that captures the main features of the effective Sommerfeld enhancement in asymptotic regimes at the level of local interactions of test particles in a (sub)halo. This ansatz provides a good approximation to the exact result. We can write it as follows (disregarding resonances for the moment):

$$\mathcal{S}_{\text{no-res}}(v,\epsilon_\phi) = S_0 \left(\frac{v}{\tilde{v}_{\max}}\right)^{-1} \left[1 + S_1^{-\frac{\bar{s}_{v,c}}{(1+p)}} \left(\frac{v}{\tilde{v}_{\text{sat}}(\epsilon_\phi)}\right)^{-\bar{s}_{v,c}}\right]^{-\frac{(1+p)}{\bar{s}_{v,c}}} \tag{3.10}$$

$$\Rightarrow \mathcal{S}_{\text{no-res}}(v,\epsilon_\phi) \xrightarrow{\text{limits}} \begin{cases} S_0 \left(\frac{v}{\tilde{v}_{\max}}\right)^{-1} & \forall \tilde{v}_{\text{sat}}(\epsilon_\phi) \ll v \ll \tilde{v}_{\max} \\ S_0\, S_1 \left(\frac{\tilde{v}_{\max}}{\tilde{v}_{\text{sat}}(\epsilon_\phi)}\right) \left(\frac{v}{\tilde{v}_{\text{sat}}(\epsilon_\phi)}\right)^{p} \propto v^p\, \epsilon_\phi^{-(1+p)} & \forall v \ll \tilde{v}_{\text{sat}}(\epsilon_\phi) \end{cases},$$

where $\tilde{v}_{\text{sat}}$, which will be defined later on, marks the transition between the Coulomb and the saturation regimes, and S_0 and S_1 are calibration constants which can be calculated explicitly:

$$S_0 = (2\pi)^{-p}\,;\; S_1 = (6/\pi)\,(12/\pi^2)^p\,. \tag{3.11}$$

The ansatz of Eq. (3.10) is valid only when the Sommerfeld effect becomes effective, which corresponds to velocities $v \leqslant \tilde{v}_{\max}$, where $\tilde{v}_{\max}$ is defined just below. Two power-law indices appear, p and $\bar{s}_{v,c}$, all positive definite (p is defined in Eq. (3.9)). Index $\bar{s}_{v,c}$ has no asymptotic relevance, and simply indicates how fast one transits from the Coulomb regime to the saturation regime. The Sommerfeld power-law index in the Coulomb regime is explicitly set to -1 for both s- and p-wave annihilation, as a consequence of absorbing the velocity dependence of the cross section in the definition of the Sommerfeld correction for the latter. Parameter $\tilde{v}_{\max}$ stands for the velocity beyond which the Sommerfeld effect roughly turns off, and $\tilde{v}_{\text{sat}}(\epsilon_\phi)$, which does explicitly depend on ϵ_ϕ, is the velocity below which the Sommerfeld effect saturates and resonances may appear. That transition occurs when the interaction range becomes shorter than the DM de Broglie wavelength. In between $\tilde{v}_{\text{sat}}$ and $\tilde{v}_{\max}$, we are in the Coulomb regime (infinite interaction-range limit). These critical velocities can actually be related to the coupling strength α_χ and to the reduced Bohr radius ϵ_ϕ, which both characterize the parameter space of our minimal Sommerfeld-enhancement setup. The appropriate definitions read:

$$\begin{cases} \tilde{v}_{\max} & \equiv \pi\,\alpha_\chi \\ \tilde{v}_{\text{sat}}(\epsilon_\phi) & \equiv \epsilon_\phi\,\frac{\tilde{v}_{\max}}{\pi} \\ \tilde{v}_{\text{unit}} & \equiv \alpha_\chi^4 \end{cases}. \tag{3.12}$$

Here, we only consider situations in which the DM Bohr radius ($\propto$ Compton wavelength) is shorter than the interaction range ($\epsilon_\phi \lesssim 1$), a condition to trigger the Sommerfeld enhancement. The saturation velocity $\tilde{v}_{\text{sat}}$ delineates a transition in velocity dependence, fixed by $\epsilon_v = v/\alpha_X = \epsilon_\phi$, at which the DM self-interaction range and the de Broglie wavelength are similar, and below which the finite range of self-interactions becomes manifest again. Then, the Sommerfeld enhancement saturates and its velocity dependence is frozen, except on resonances. A resonance of order n can efficiently pop up if $v < \tilde{v}_{\text{sat}}(\epsilon_\phi = \epsilon_\phi^{\text{res},n})$, where the saturation velocity is evaluated at the corresponding resonant value of the reduced Bohr radius, $\epsilon_\phi^{\text{res},n}$. Finally, parameter $\tilde{v}_{\text{unit}}$ is meant to account for the unitarity constraint on Sommerfeld resonances.

Properties of resonances: We highlight the discussion of resonances, which will lead to non-trivial features throughout the paper and be specific zones in parameter space of gigantic signal enhancements. In the same spirit as above, we can write a simplifying ansatz to describe the enhancement on resonances, which we deliberately separate from the non-resonant ansatz of Eq. (3.10) for clarity:

$$
\begin{aligned}
\mathcal{S}_{\mathrm{res},n}(v,\epsilon_\phi) \overset{n\geqslant 1+\frac{p}{2}}{=} & \; S_0^{\mathrm{res}} \left(\frac{\tilde{v}_{\max}}{\tilde{v}_{\mathrm{sat}}(\epsilon_\phi)}\right)\left(\frac{v}{\tilde{v}_{\mathrm{sat}}(\epsilon_\phi)}\right)^{(p-2)}\left(1+\frac{\tilde{v}_{\mathrm{unit}}}{v}\right)^{-2} \\
& \times\theta\left(\tilde{v}_{\mathrm{sat}}(\epsilon_\phi)-v\right)\delta_{\epsilon_\phi/\{\epsilon_\phi^{\mathrm{res},n}\}} \\
\propto & \quad \epsilon_\phi^{(1-p)}\, v^{(p-2)}\,,
\end{aligned}
\tag{3.13}
$$

where $\tilde{v}_{\mathrm{unit}}$ has been defined in Eq. (3.12), and saturates the amplitudes of resonant peaks when $v<\tilde{v}_{\mathrm{unit}}$, which allows us to effectively prevent any violation of the unitarity constraint (see discussion at the very end of Sec. 3.1.1). We have introduced

$$
\delta_{\epsilon_\phi/\{\epsilon_\phi^{\mathrm{res},n}\}} \equiv \begin{cases} 1 & \text{if } \epsilon_\phi \in \{\epsilon_\phi^{\mathrm{res},n}\} \\ 0 & \text{otherwise} \end{cases}
\tag{3.14}
$$

where again $p=0\,/\,2$ for s/p-wave annihilation (and for which the first resonance is at $n=1\,/\,2$, respectively). The constant reads:

$$
S_0^{\mathrm{res}} = (\pi/6)\,(6/\pi^3)^p\,. \tag{3.15}
$$

It is important to recall the generic features of resonances, which occur at $\epsilon_\phi\sim\epsilon_\phi^{\mathrm{res},n}$, and can be triggered only if $v\leqslant\tilde{v}_{\mathrm{sat}}(\epsilon_\phi^{\mathrm{res},n})$ — in the above equations, for simplicity, we adopt an extreme simplification by means of a discrete measure, which triggers resonances only when ϵ_ϕ sits exactly on one of its resonant values (to avoid numerical discontinuities, this can be replaced by an extremely thin unnormalized Gaussian function, or even a Cauchy function if one fancies better capturing the actual shapes of resonances).

In the s-wave case, resonances are boosted at low velocity $\propto 1/(n\,v)^2 \propto \epsilon_\phi/v^2$, with decreasing amplitudes for higher-order resonances (in fact, linearly with ϵ_ϕ [or $\tilde{v}_{\mathrm{sat}}$], as the latter jumps to smaller and smaller resonant values $\epsilon_\phi^{\mathrm{res},n}$)—see Eq. (3.6) and Eq. (3.7). Note also the unitarity limit that saturates peak amplitudes to $\widetilde{\propto}\;\epsilon_\phi/(n\,\tilde{v}_{\mathrm{unit}})^2$ when $v<\tilde{v}_{\mathrm{unit}}$, which will have some impact when inspecting the translation in terms of subhalo masses. As for the inter-resonance baseline (saturation regime), it scales like $\propto 1/\tilde{v}_{\mathrm{sat}}\propto 1/\epsilon_\phi$ and remains velocity independent—see Eq. (3.10) and Eq. (3.12).

In contrast, as a consequence of absorbing the v^2 suppression factor in the ansatz, p-wave resonances $\propto n^2 \propto 1/\epsilon_\phi$ are velocity independent and have their amplitudes increasing for higher-order resonances, i.e. lower resonant values of ϵ_ϕ. That feature is actually very important because it implies that the annihilation signal is then only set by the full DM squared density on p-wave resonant peaks, except again when approaching the unitarity limit, $v\sim\tilde{v}_{\mathrm{unit}}$. Indeed, at velocities lower than $\tilde{v}_{\mathrm{unit}}$, the p-wave suppression re-appears as $\propto(v/\tilde{v}_{\mathrm{unit}})^2$, which bounds from below the phase-space distribution available to amplify resonances. On the other hand, the inter-resonance baseline remains

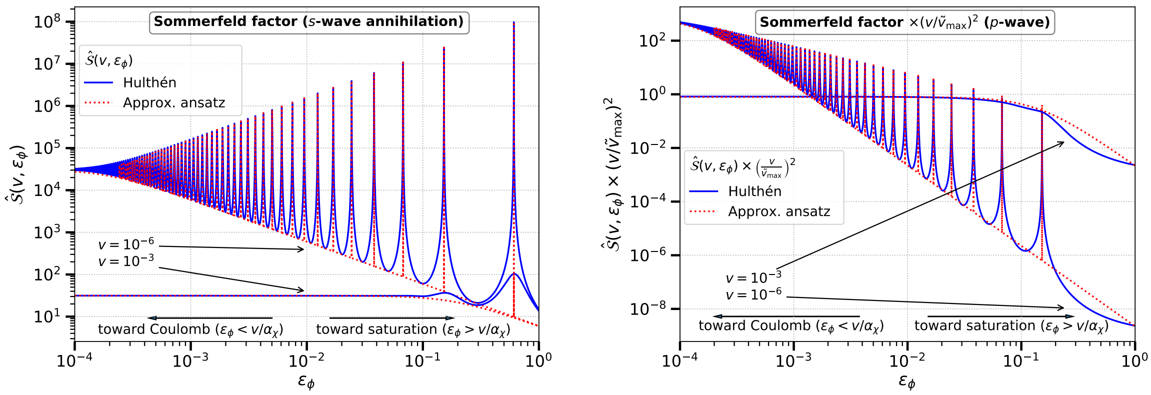

Figure 1. **Left panel:** Comparison between the Sommerfeld enhancement factor obtained for an s-wave annihilation process from Eq. (3.3) and the ansatz formulated in Eq. (3.17), for two different values of the speed v. **Right panel:** Same for a p-wave annihilation process, but then the actual Sommerfeld factor of Eq. (3.5) is multiplied by a factor of $(v/\tilde{v}_{\rm max})^2$ to carry the full velocity dependence of the cross section.

fully velocity suppressed $\propto v^2/\tilde{v}_{\rm sat}^3 \propto v^2/\epsilon_\phi^3$. Therefore, the surge of resonances in the p-wave case is due to the relative suppression of the baseline. Actually, the amplitude ratio $\mathcal{R}$ between resonances and baseline scales exactly the same for both the s-wave and p-wave cases in this formulation, and reduces to:

$$\mathcal{R}(v, \epsilon_\phi = \epsilon_\phi^{{\rm res},n}) = \left(\frac{\pi}{6}\right)^2 \left(\frac{v}{\tilde{v}_{\rm sat}}\right)^{-2} \left(1 + \frac{\tilde{v}_{\rm unit}}{v}\right)^{-2} \mathrel{\tilde{\propto}} (\epsilon_\phi/v)^2 \,. \tag{3.16}$$

The dependence of the resonant amplitudes on the reduced Bohr radius ϵ_ϕ is shown in Fig. 1, while their dependence on v is shown in Fig. 2, which will be discussed further below.

All this can be wrapped up in a more synthetic form,

$$\mathcal{S}(v, \epsilon_\phi) = \mathcal{S}_{\text{no-res}}(v, \epsilon_\phi) \left(1 - \sum_{n=1+\frac{p}{2}} \delta_{\epsilon_\phi/\{\epsilon_\phi^{{\rm res},n}\}}\right) + \sum_{n=1+\frac{p}{2}} \mathcal{S}_{{\rm res},n}(v, \epsilon_\phi) \tag{3.17}$$
$$\propto (v/v_0)^{-s_v} \,,$$

where the generic index s_v takes different values according to the different Sommerfeld regimes:

$$s_v = \begin{cases} 1 & \text{(Coulomb regime)} \\ -p & \text{(non-resonant saturation regime)} \\ (2 - p) & \text{(resonances)} \longrightarrow -p \;(\text{if } v \lesssim \tilde{v}_{\rm unit}) \end{cases} \,, \tag{3.18}$$

where $p = 0/2$ for s/p-wave annihilation. We stress that on resonances, the scaling of peak amplitudes becomes $\propto (v/\tilde{v}_{\rm unit})^p$ as soon as $v \lesssim \tilde{v}_{\rm unit}$, as a consequence of the unitarity limit. This translates into a transition of s_v from $(2 - p)$ to $-p$ on resonances at the unitarity boundary.

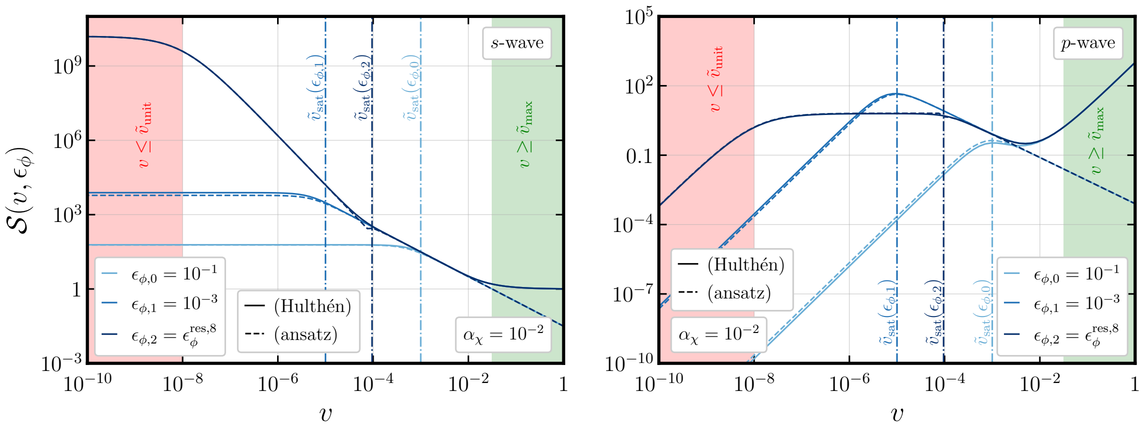

Figure 2. Effective Sommerfeld enhancement factor as a function of DM speed, for different values of the reduced Bohr radius ϵ_ϕ: a large value 0.1, a small value of 10^{-3}, and an intermediate value of $\sim 10^{-2}$ sitting on the $n = 8$ resonance. The enhancement factor is valid up to $\tilde{v}_{\rm max}$, and saturates below $\tilde{v}_{\rm unit}$ on s-wave resonances (not on p-wave ones). Transition from Coulomb to saturation regimes occurs at $\tilde{v}_{\rm sat}(\epsilon_\phi)$, reported as vertical dash-dotted lines. **Left panel:** s-wave case. **Right panel:** p-wave case.

Our general ansatz of Eq. (3.17) fully parameterizes the Sommerfeld enhancement factor at the level of local interactions in DM halos. It will serve as a basis to integrate the Sommerfeld effect over an entire (sub)halo. A comparison of this ansatz with the exact solution of the Sommerfeld enhancement factor is provided in Fig. 1 for both the s-wave and p-wave cases, assuming two values (high and low) of the relative DM speed. We see that this form closely matches with the exact result, except when ϵ_ϕ approaches 1, as expected. In the p-wave case, the change of hierarchy in the Sommerfeld enhancement between the low and high velocity curves (with respect to the s-wave case) is simply due to the fact that we have absorbed the v^2 suppression factor in the definition of the effective Sommerfeld factor. The virtue of this is that we directly see the true hierarchy of full cross sections as function of velocity from this effective definition. In particular, we see that even though there is a relative p-wave suppression of 10^{-6} between $v = 10^{-3}$ and $v = 10^{-6}$, the Sommerfeld-enhanced cross sections have similar amplitudes at $\epsilon_\phi \sim 1.5 \times 10^{-3}$, with a net and increasing advantage to smaller velocities for smaller values of ϵ_ϕ. Already, this helps understand the fundamental role to be played by DM structures with small dispersion velocities in the following.

To further illustrate the velocity dependency of the effective Sommerfeld factor, we explicitly show $\mathcal{S}$ as a function of DM speed in Fig. 2, for three different values of the reduced Bohr radius ϵ_ϕ: a relatively "large" value of 0.1, which implies a moderate hierarchy between the DM particle mass and that of the force carrier (moderate Sommerfeld enhancement); a small value of 10^{-3}, hence a stronger hierarchy (significant enhancement); and an intermediate value of $\sim 10^{-2}$, but sitting exactly on the $n = 8$ resonance (strong enhancement). The left (right) panel shows the dependence for an s-wave (p-wave) annihilation process. The saturation velocities $\tilde{v}_{\rm sat}(\epsilon_\phi)$ associated with the different choices of ϵ_ϕ are displayed as vertical dashed lines, which delineate the transition between the saturation (to the left thereof) and the Coulomb (to the right) regimes. This figure illustrates the reasonably good match between our Sommerfeld ansatz of Eq. (3.17) (dashed curves) and the exact formulation (solid curves). Following curves from right to left (decreasing velocity), for the s-wave (left panel), the enhancement $\propto 1/v$ in the Coulomb regime saturates as $v \leq \tilde{v}_{\rm sat}$, except on the resonance for which it further increases $\propto 1/v^2$ down to the unitarity limit characterized by $\tilde{v}_{\rm unit}(\alpha_\chi)$, at which it finally saturates. For the p-wave case (right panel), we actually see the product of the net Sommerfeld factor with the p-wave suppression factor $\propto v^2$ [i.e. the effective Sommerfeld factor as defined in Eq. (3.8)], which slightly delays the onset of the enhancement as v decreases below $\tilde{v}_{\rm max}$. Then, as for the s-wave case, the Coulomb regime ($v > \tilde{v}_{\rm sat}$) exhibits a $1/v$ scaling down to $\tilde{v}_{\rm sat}$ (which hardly compensates for p-wave suppression for large reduced Bohr radii ~ 0.1, leading to a small net enhancement). Transitioning to the saturation regime, the effective Sommerfeld effect saturates to its maximal value for the p-wave case for $v \sim \tilde{v}_{\rm sat}$, before the p-wave suppression factor takes over at velocities smaller than $\tilde{v}_{\rm sat}$. On the resonance, however, the maximal saturation value is further maintained independent of the velocity all the way down to the unitary limit characterized by $\tilde{v}_{\rm unit}$ (the actual Sommerfeld enhancement compensates for the p-wave suppression), below which p-wave suppression ends up taking over. All this explains the important role played by $\tilde{v}_{\rm sat}(\epsilon_\phi)$ in the p-wave case, as well as the one of $\tilde{v}_{\rm unit}$ on resonances. Fig. 2 will later help better understand the mass-velocity dependencies at fixed values of ϵ_ϕ.

3.2 Gamma-ray signals: astrophysical factors and DM phase-space modeling

The DM-induced γ-ray flux integrated over a sky region of solid angle $\Delta\Omega$ about a target halo center reads [45][2]

$$\frac{\mathrm{d}\Phi_\gamma}{\mathrm{d}E_\gamma} = \frac{1}{4\pi}\frac{(\sigma v_{\rm rel})_0}{\eta m_\chi^2}\frac{\mathrm{d}N}{\mathrm{d}E_\gamma} J_{\mathcal{S}}(\Delta\Omega) \tag{3.19}$$

where $\mathrm{d}N/\mathrm{d}E_\gamma$ is the γ-ray spectrum per annihilation, and $\eta = 2$ for self-conjugate DM ($\eta = 4$ for non-self-conjugate DM). In the case of a velocity-dependent annihilation cross section that can be expressed as $(\sigma v_{\rm rel})_0 \times \mathcal{S}(v)$, like in the effective formulation of the Sommerfeld enhancement above, the astrophysical factor $J_{\mathcal{S}}$ encodes the information on both the DM spatial and velocity distributions, and reads

$$J_{\mathcal{S}}(\Delta\Omega) = \iint_{\Delta\Omega} \mathrm{d}\Omega \int \mathrm{d}s \int \mathrm{d}^3\vec{v}_1 \iint \mathrm{d}^3\vec{v}_2\, f(r(s,\ ),\vec{v}_1)\, f(r(s,\ ),\vec{v}_2)\, \mathcal{S}\left(\frac{v_{\rm rel}}{2}\right), \tag{3.20}$$

where $\vec{v}_{\rm rel} = \vec{v}_2 - \vec{v}_1$ is the relative velocity with $v_{\rm rel} = |\vec{v}_{\rm rel}|$, and $f(r,\vec{v})$ is the phase-space distribution function (PSDF) of the DM (assuming spherical symmetry). Here, the PSDF is normalized to the total mass of the gravitational system of interest, such that at halocentric radius r

$$\rho_\chi(r) = \iint \mathrm{d}^3\vec{v}\, f(r,\vec{v}). \tag{3.21}$$

Note that if one trades the effective Sommerfeld factor $\mathcal{S}$ for its exact form $\hat{\mathcal{S}}$, one should add an additional factor of $(v/c)^2$ in the expression of the J-factor for p-wave annihilation — see Eq. (3.8). Our effective form allows to write a unique form for both s- and p-wave processes

[2] For easier comparison with the majority of previous works in the literature, we do not include in the definition of the J-factor the $1/(4\pi)$ factor that appears in the derivation of an intensity from a volume emissivity, and is a prefactor in Eq. (3.19). As a result, the J-factors given in this work are expressed in $\mathrm{GeV}^2\,\mathrm{cm}^{-5}\,\mathrm{sr}$.

by absorbing the full velocity dependence in $\mathcal{S}$. Eq. (3.20) is a generalization of the standard velocity-independent J-factor

$$J(\Delta\Omega) = \int_{\Delta\Omega} \mathrm{d}\Omega \int \mathrm{d}s\, \rho_\chi^2(r(s, \ )) \,, \tag{3.22}$$

which is valid for s-wave annihilation without Sommerfeld enhancement.

Assuming spherical symmetry of the DM halo, the integral over solid angle becomes an integral over the angular distance θ from the center of the object, with $\mathrm{d}\Omega = 2\pi \sin\theta\, \mathrm{d}\theta$ and $r(s, \ ) \equiv r(s,\theta) = \sqrt{s^2 + D^2 - 2sD\cos\theta}$, where D is the distance of the observer to the center of the object. The integral is usually performed over an angular size $\theta_{\rm int}$ that depends on the target and the γ-ray detection technique. In this study, we will assume the distances of target halos to be sufficiently large to integrate the signals over angular extents exceeding the virial sizes of halos (point-like approximation).

In practice, Eq. (3.20) can be rewritten in terms of a J-factor for an effective squared density profile $\rho_{\chi,\rm eff}$ as

$$J_{\mathcal{S}}(\theta_{\rm int}) = 2\pi \int_0^{\theta_{\rm int}} \mathrm{d}\theta\, \sin\theta \int \mathrm{d}s\, \rho_{\chi,\rm eff}^2(r(s,\theta)), \tag{3.23}$$

assuming that the telescope points to the center of the target halo (this is easily generalized to any direction, see, e.g., [46]), with a resolution angle of $\theta_{\rm int}$. Correspondingly, we introduce

$$\rho_{\chi,\rm eff}^2(r) \equiv \left\langle \mathcal{S}\left(\frac{v_{\rm rel}}{2}\right)\right\rangle_v (r) \times \rho_\chi^2(r), \tag{3.24}$$

where $\langle\rangle_v$ denotes an average over the DM relative velocity distribution. The average of an observable $\mathcal{O}(v_{\rm rel})$ that depends on the relative velocity is conventionally given by

$$\langle \mathcal{O}(v_{\rm rel})\rangle_v (r) = \int \mathrm{d}^3\vec{v}_{\rm rel}\, \mathcal{O}(v_{\rm rel}) F_{\rm rel}(r, \vec{v}_{\rm rel}), \tag{3.25}$$

where the relative velocity distribution reads

$$F_{\rm rel}(r, \vec{v}_{\rm rel}) = \int \mathrm{d}^3\vec{v}_{\rm c}\, f_{\vec{v}}(r, \vec{v}_1) f_{\vec{v}}(r, \vec{v}_2), \tag{3.26}$$

with $\vec{v}_{\rm c} = (\vec{v}_1 + \vec{v}_2)/2$ the center-of-mass velocity and $f_{\vec{v}}(r, \vec{v}) \equiv f(r, \vec{v})/\rho_\chi(r)$ the DM velocity distribution, defined as a probability density function (PDF), i.e. normalized to 1 over the relevant phase space.

The accurate numerical results of this work are based on the Eddington inversion formalism [47, 48], which, assuming an isotropic velocity distribution for DM particles and a spherically symmetric halo in dynamical equilibrium, predicts the full DM PSDF $f(\vec{r}, \vec{v})$. For a detailed discussion of the applicability of the Eddington inversion to different classes of DM halos, see ref. [49]. Note that the predictive power of this formalism has been tested against cosmological simulations in ref. [50], and has been shown to predict the velocity moments of DM within $\sim$ 15% accuracy. Interestingly, such isotropic PSDF models have a similar predictive power as more elaborate models including anisotropy in the velocity field [51].

Starting from the PSDFs of DM halos (in a large range of masses) as predicted from the Eddington inversion method, we found that a very good estimate ($\leqslant$ 30% of error) of the averaged effective Sommerfeld factor could be obtained by picking the non-averaged Sommerfeld factor at some averaged values of the relative speed:

$$\left\langle \mathcal{S}\left(\frac{v_{\rm rel}}{2}\right)\right\rangle_v (r) \simeq \mathcal{S}\left(\frac{\langle v_{\rm rel}^{\frac{2}{(p-1)}}\rangle^{\frac{(p-1)}{2}}(r)}{2}\right), \tag{3.27}$$

where the relative velocity moments $\langle v_{\rm rel}^{\pm n}\rangle_v(r)$ are calculated from the Eddington PSDF inferred for the considered halo. This is roughly valid for an extended range of halo masses, from very small subhalo to galaxy cluster masses. In the next section, we carefully inspect the impact of DM subhalos on the overall Sommerfeld-enhanced signal predictions.

4 Subhalo boost factor for velocity-dependent annihilation

DM subhalos, which are a generic prediction of the theory of structure formation within CDM [22, 52, 53] and thus also characterize the WIMP class of models [23, 24, 54–56], are known to increase the s-wave annihilation rate, which is referred to as *subhalo boost factor* in the frame of indirect DM searches [40–42, 57, 58]. Predictions for subhalo boost factors have been mostly derived for vanilla s-wave annihilation processes, for which the annihilation rate is velocity-independent. The impact of subhalos is also expected to be important when the annihilation rate depends on (inverse powers) relative speed, but it is then slightly more difficult to calculate. Indeed, for the broad picture, since the internal average velocity dispersion of DM in subhalos decreases as their masses decrease, then the mass function of subhalos should translate into a non-trivial relative speed function. Since the Sommerfeld enhancement scales like powers of $1/v$, it is clear that the presence of small subhalos can significantly amplify predictions of the annihilation rate in target objects. We shall see below that ϵ_ϕ, the DM Bohr radius ($\sim$ Compton wavelength) in units of the interaction range, is actually the key parameter that determines the most relevant subhalo mass range. We shall also see that the related additional boost factor amounts to orders of magnitude. Before going into more details, we recall that the impact of subhalos was already studied in several references, e.g. [26–30, 59], though with different perspectives.

Here we improve over past studies on several aspects. First, we rely on an analytical subhalo population model mostly built from constrained and controlled theoretical inputs, which self-consistently obeys the global kinematic and dynamical constraints on the host halo, and which includes subhalo tidal stripping. This means that given an observationally constrained global mass model for the host halo, we can self-consistently translate it into a halo model that comprises both a smooth distribution of DM and a substructure component. The bases and features of this model were proposed in ref. [60], and further explored in, e.g., refs. [61–64]. This analytical subhalo population model can be easily applied to any host halo configuration. The complete model is used to get our more accurate numerical results on the Sommerfeld-enhanced subhalo contribution to J-factors, while further analytical approximations are used to derive fully analytical results. Second, similar to other recent studies (e.g. [32, 34–36]), we take advantage of the phase-space distribution studies performed in refs. [49, 50]. The latter rigorously determine the regimes where the application of the Eddington inversion method [47] can lead to a reliable description of the PSDF of DM in structures, which can be used to compute any velocity-dependent DM signal (see direct applications of these studies in, e.g., ref. [65] for p-wave annihilation, or in ref. [66] for

DM capture by stars). Although hardly scalable to a full population of objects, the Eddington inversion applied to a reduced subhalo mass range can be used to calibrate analytical approximations and get accurate results.

In the following, we first give in Sec. 4.1 a description of the subhalo population model, before formalizing the general calculation of the induced boost factor in Sec. 4.2. Finally, we turn to an approximate analytical derivation of the boost factor in Sec. 4.3, which will allow us to make a detailed physical interpretation of the more accurate numerical results.

4.1 Subhalo population model

Here, we introduce the main properties of the subhalo population model proposed in ref. [60] (SL17 henceforth). The philosophy behind this model is to think of a DM halo as an assembly of smaller-scale pre-existing halos, consistently with the prescriptions of excursion set theory and merger-tree studies (see [67–69] and, e.g., [70, 71] for more recent approaches). Should these subhalos be hard spheres with negligible subhalo-subhalo encounter rate, they would simply track the global gravitational potential of the global host halo, which they are part of. However, they actually experience tidal mass loss and may even be disrupted in some cases. These phenomena depend on the time spent in the host and on their pericenter (deepest position in the host gravitational potential), and possibly encounters with stellar disks and individual stars in spiral galaxies. Sticking to a spherically symmetric description of both a smooth halo (which comprises both the originally diffuse DM and the DM stripped away from subhalos) and a subhalo population, one can write down a constrained *smoothed* mass density profile $\langle\rho_{\rm host}\rangle$, where $\langle\rangle$ denotes an average in spherical shells here, for the host in terms of two components:

$$\langle\rho_{\rm host}\rangle(R) = \rho_{\rm sm}(R) + \rho_{\rm sub}(R)\,, \tag{4.1}$$

where R is the distance to the host center, $\rho_{\rm sm}$ the smooth density profile and $\rho_{\rm sub}$ the averaged subhalo population density profile.[3] The coarse-grained global host mass density profile $\langle\rho_{\rm host}\rangle$ is constrained from structure formation to be close to a Navarro-Frenk-White (NFW) profile [72–78], which is also consistent with observational constraints on different scales [79–81], pending ongoing debates about possible core-cusp issues [25, 82, 83]. An important point is that $\rho_{\rm host}$ is also the one global density profile constrained by kinematic or dynamical studies of specific objects, should they be dwarf galaxies, galaxies, or galaxy clusters. A consistent subhalo population model should then be such that the sum of the smooth halo profile and the overall subhalo profile matches with observational constraints on the global host halo, whenever available.

Rigorously, $\rho_{\rm sub}$ should be described as a discrete sum over all subhalos mapping all inhomogeneities, but the smoothed limit (i.e., an average within spherical shells) can be considered to describe the overall subhalo density profile:

$$\rho_{\rm sub}(R) = \left\langle \sum_i^{N_{\rm tot}} \rho_i(|\vec{R}-\vec{r}_i|) \right\rangle_{|\vec{R}|}(R) = \int {\rm d}m\, \frac{{\rm d}n_{\rm sub}(R,m)}{{\rm d}m}\, \langle m_{\rm t}\rangle_c(m,R)\,, \tag{4.2}$$

where $\vec{r}_i$ is the position vector (center) of the $i^{\rm th}$ subhalo in the host frame and ρ_i its spherical density profile, $m = m_{200}$ is the canonical virial mass a subhalo would have in a homogeneous

[3]In practice, the smooth component is deduced from the subhalo population model and the global host profile, according to $\rho_{\rm sm}(R) = \langle\rho_{\rm host}\rangle(R) - \rho_{\rm sub}(R)$, and must obey the condition $\rho_{\rm sm}(R) \geqslant 0$.

background, $\langle m_t \rangle_c(R)$ is the subhalo physical tidal mass $m_t \leq m$ averaged over concentration at radius R, and $\mathrm{d}n_{\rm sub}/\mathrm{d}m$ is the differential number density of subhalos per unit mass, in the continuous limit. The implementation of tidal effects is hidden in the way the tidal mass m_t is predicted, given a fictitious virial mass m, a concentration c, and a prescription for the density profile, which will be specified later.

Therefore, designing a subhalo population model implies defining this continuous limit in terms of a subhalo number density consistent with Eq. (4.1) while carrying imprints of initial cosmological conditions distorted by environmental effects (gravitational tides). Our model defines this number density in terms of PDFs describing the mass function $\mathrm{d}\mathcal{P}_m(m)/\mathrm{d}m$, the concentration function $\mathrm{d}\mathcal{P}_c(c,m)/\mathrm{d}c$, the *driving* spatial distribution $\mathrm{d}\overline{\mathcal{P}}_V/\mathrm{d}V$ (the meaning of *driving* is made clear in the appendix), and the total number of subhalos $N_{\rm tot}$ orbiting the host halo:

$$\frac{\mathrm{d}n_{\rm sub}(R,m)}{\mathrm{d}m} = \frac{\mathrm{d}^2 N_{\rm sub}}{\mathrm{d}m\,\mathrm{d}V} = \frac{N_{\rm tot}}{K_{\rm tidal}}\,\frac{\mathrm{d}\overline{\mathcal{P}}_V(R)}{\mathrm{d}V}\int \mathrm{d}c\,\frac{\mathrm{d}^2\mathcal{P}_{c,m}(c,m,R)}{\mathrm{d}c\,\mathrm{d}m}\,. \tag{4.3}$$

In this equation, $N_{\rm tot}$ is the total number of surviving subhalos, and $K_{\rm tidal} \leq 1$ is a normalization constant that ensures the whole PDF to be normalized to unity (said differently, it accounts for the fact that the nominal concentration and mass PDFs can be cut off by tidal effects). The concentration and mass PDFs are intricate as a result of tidal effects, which is explained in App. B. This comes from the fact that gravitational tides are more efficient in pruning less concentrated objects, which induces a selection of halos in concentration (hence on mass) depending on their averaged orbital distance to the host halo center. At this stage, it is therefore important to introduce two other parameters of the model: the minimal and maximal subhalo *virial* masses, $m_{\rm min}$ and $m_{\rm max}$, respectively (keeping in mind that the actual smallest masses in the population model can be much smaller than $m_{\rm min}$, due to tidal stripping). The former relates to the interaction properties of DM particles, and is in most cases fixed by the free-streaming length of DM at matter-radiation equality [23, 54, 56]—for WIMPs, it may take values in the range 10^{-12}-10^{-4} $M_\odot$. The latter obviously depends on the host halo mass, and will be fixed to $m_{\rm max} = 0.01\,M_{\rm host}$ throughout this work, similar to what is found in cosmological simulations [84–86].

While we include all the details in the full numerical calculations of J-factors, it is interesting to write down an approximation of the expected subhalo distribution as follows:

$$\frac{\mathrm{d}^2 n_{\rm sub}(R,m,c)}{\mathrm{d}c\,\mathrm{d}m} \approx N_{\rm tot}\,\frac{\mathrm{d}\overline{\mathcal{P}}_V(R)}{\mathrm{d}V}\,\frac{\mathrm{d}\overline{\mathcal{P}}_m(m)}{\mathrm{d}m}\,\frac{\mathrm{d}\overline{\mathcal{P}}_c(c)}{\mathrm{d}c}\,. \tag{4.4}$$

If subhalos were hard spheres insensitive to tides, this equation would give a decent description of the subhalo population, with concentration and mass PDFs being close to those of field subhalos. In such a hard-sphere approximation, the spatial distribution would be merely

$$\frac{\mathrm{d}\overline{\mathcal{P}}_V(R)}{\mathrm{d}V} = \frac{\langle \rho_{\rm host}\rangle(R)}{M_{\rm host}}\,. \tag{4.5}$$

In fact, departures from the spatial matching between the total halo mass profile and the total subhalo mass profile are mostly observed in the inner parts of host halos in simulations, where tidal effects are strong [84, 85]. Although the abundance of subhalos in the very central regions of host halos can hardly be measured reliably in simulations, due to resolution issues, this flattening of the subhalo spatial distribution in the central parts of host halos can actually still be predicted analytically by considering tidal disruption on top of tidal stripping, as detailed in App. B. The above approximation will still turn useful when trying to get analytical estimates of the Sommerfeld-enhanced subhalo boost factor.

4.2 Subhalo boost factor: generalities

Here we establish the complete expressions that are used to perform generic computations of the subhalo boost factor — see more detailed discussions in, e.g., [5, 41, 45, 46, 60, 64, 87–90]. We have introduced the astrophysical J-factor in Sec. 3.2, which is proportional to the integral of the DM squared density profile ρ_χ^2 along the line of sight, with the replacement $\rho_\chi^2 \leftrightarrow \rho_{\chi,\mathrm{eff}}^2$ given in Eq. (3.24) to account for any velocity dependence in the annihilation signals. Neglecting subhalos amounts to setting $\rho_\chi^2(r) = \rho_{\mathrm{host}}^2(r)$, where the total DM profile of the host object is given in Eq. (4.1) and includes both a smooth DM component and a subhalo population. A definition of the subhalo boost factor is straightforward and may readily be expressed in terms of the relevant J-factors:

$$\mathcal{B} \equiv \frac{J_{\mathrm{tot}}}{J_{\mathrm{smooth\ approx}}}\,, \tag{4.6}$$

where J_{tot} is the J-factor including rigorously both the smooth and subhalo contributions, while $J_{\mathrm{smooth\ approx}}$ simply consider the contribution of the whole system after smoothing out all inhomogeneities. In this form, $\mathcal{B}$ is merely the multiplicative factor to apply to the smooth approximation of the J-factor to get the one accounting for subhalos. The fact that $\mathcal{B} \geqslant 1$ is a rather generic[4] consequence of that $\langle \rho_{\mathrm{host}}^2\rangle(r) \geqslant \langle \rho_{\mathrm{host}}\rangle^2(r)$ [40].

Using Eq. (3.22), we can already express the smooth approximation of the total J-factor as

$$J_{\mathrm{smooth\ approx}} = \int_{\Delta\Omega} \mathrm{d}\Omega \int \mathrm{d}s\, \langle \rho_{\mathrm{host}}\rangle^2 (R(s,\,))\,, \tag{4.7}$$

which is the integral of the squared global density profile of the host halo along the line of sight, neglecting any inhomogeneous component. Assuming that subhalos contribute as point-like sources, the actual total J-factor should rather be expressed as

$$J_{\mathrm{tot}} = \int_{\Delta\Omega} \mathrm{d}\Omega \int \mathrm{d}s\, \langle \rho_{\mathrm{host}}^2\rangle (R(s,\,))\,, \tag{4.8}$$

with

$$\langle \rho_{\mathrm{host}}^2\rangle (R) \simeq \rho_{\mathrm{sm}}^2(R) + \underline{\rho_{\mathrm{sub}}}^2(R) + 2\,\rho_{\mathrm{sm}}(R)\,\rho_{\mathrm{sub}}(R)\,. \tag{4.9}$$

We have introduced $\underline{\rho_{\mathrm{sub}}}^2 \neq \rho_{\mathrm{sub}}^2$ to account for the fact that if subhalos contribute as point-like sources, their contribution to the annihilation flux is not proportional to their smooth mass density profile squared ρ_{sub}^2, but rather to

$$\underline{\rho_{\mathrm{sub}}}^2(R) \equiv \rho_\circledast^2 \int \mathrm{d}c \int \mathrm{d}m\, \xi_{\mathrm{t}}(m,c,R)\, \frac{\mathrm{d}^2 n_{\mathrm{sub}}(c,m,R)}{\mathrm{d}c\,\mathrm{d}m}\,, \tag{4.10}$$

where the subhalo number density in mass-concentration phase-space $\mathrm{d}^2 n_{\mathrm{sub}}/\mathrm{d}c\,\mathrm{d}m$ can be inferred from Eq. (4.3), and is given an approximation in Eq. (4.4). We have introduced the tidal annihilation volume $\xi_{\mathrm{t}}(m,c,R)$ for a subhalo of virial mass m, concentration c, and radial position R in the host halo, defined as

$$\xi_{\mathrm{t}}(m,c,R) \equiv 4\pi \int^{r_{\mathrm{t}}(m,c,R)} \mathrm{d}r\, r^2 \left\{\frac{\rho(r,m,c,R)}{\rho_\circledast}\right\}^2\,. \tag{4.11}$$

[4] Note that for nominal velocity-suppressed p-wave annihilation, we could actually have $\mathcal{B} \leqslant 1$.

This is the integral of the inner subhalo density profile $\rho(r, m, c)$ performed over the assumed spherically symmetric subhalo tidal volume $\delta V_{\rm t}$ delineated by the tidal radius $r_{\rm t}(m, c, R)$, whose parametric dependencies are explicit. The constant parameter $\rho_{\circledast}$ is an arbitrary normalization density, which allows $\xi_{\rm t}$ to be interpreted as the effective volume a (sub)halo would need to reach the same annihilation rate as if it had a constant density of $\rho_{\circledast}$ [42, 58].

Following Eq. (4.9), the total J-factor can be rewritten as

$$J_{\rm tot} = J_{\rm sm} + J_{\rm sub} + J_{\rm cross} \simeq J_{\rm sm} + J_{\rm sub}\,, \tag{4.12}$$

where the definition of each term is now obvious, and where it is assumed that we can neglect the cross term to a very good approximation [60, 64, 88]. We still include it, though, in our numerical calculations.

Therefore, one can fully compute the subhalo boost factor once $\mathrm{d}^2 n_{\rm sub}/\mathrm{d}c\,\mathrm{d}m$ and $\rho_{\rm sub}$ are determined (assuming an universal shape for the subhalo density profile). When a velocity dependence of the annihilation cross section is considered, the above expressions change only by the substitutions already introduced in Sec. 3.2 (we specialize to the case of the Sommerfeld enhancement, though this statement is more general):

$$\begin{cases} \rho^2_{\rm sm}(R) & \longrightarrow\ \rho^2_{\rm sm,eff}(R) \equiv \left\langle \mathcal{S}\left(\frac{v_{\rm rel}}{2}\right)\right\rangle_v (R) \times \rho^2_{\rm sm}(R)\,, \\ \xi_{\rm t}(m,c,R) & \longrightarrow\ \xi_{\rm t,eff}(m,c,R) \equiv 4\pi \int_0^{r_{\rm t}(m,c,R)} \mathrm{d}r\, r^2 \left\langle \mathcal{S}\left(\frac{v_{\rm rel}}{2}\right)\right\rangle_v (r) \left\{\frac{\rho(r,m,c)}{\rho_{\circledast}}\right\}^2 . \end{cases} \tag{4.13}$$

All full numerical calculations presented in this paper will be based on these equations. However, since the main goal is to get analytical insights of the results, we shall try to extract the simplest description that still allows to capture the correct orders of magnitude.

An additional simplification can be used if (i) the telescope is pointed to the center of the target host halo and (ii) if the smooth component dominates over the subhalo component there (which is expected as tidal stripping is very efficient in the central parts of host halos). In that case, we have $J_{\rm smooth\ approx} \simeq J_{\rm sm}$ to an excellent approximation [60]. Moreover, if the target host halo is sufficiently far away from the observer, at a distance $D \gg R_{\rm host}$, and appears (at least almost) as a point-like source, then we can further simplify the expressions of the J-factors, and thereby that of the subhalo boost factor. By defining

$$J(m,c,D) \equiv \frac{\rho^2_{\circledast}\, \xi_{\rm t}(m,c)}{D^2}\,, \tag{4.14}$$

where we now neglect tidal stripping and simply identify a (sub)halo with its conventional virial mass m and concentration c, and where it is assumed that $\xi_{\rm t}$ is integrated up to the virial radius; then we get

$$\begin{cases} J_{\rm sm} & \simeq\ J_{\rm smooth\ approx} = J_{\rm host} \equiv J(M_{\rm host}, c_{\rm host}, D) \\ J_{\rm sub} & \approx\ N_{\rm tot}\, \langle J(m,c,D)\rangle_{m,c} \end{cases}, \tag{4.15}$$

where from now on, $J_{\rm host}$ characterizes the smooth approximation of the J-factor for the host halo, and where $\langle\rangle_{m,c}$ denotes an average over mass and concentration phase space. For the latter, one can use the mass and concentration functions of field subhalos for decent order-of-magnitude estimates, because most of subhalos lie away from the central parts of the host halo, where tidal effects can be neglected. Consistently, the subhalo boost factor can be approximated by

$$\mathcal{B} \approx 1 + N_{\rm tot}\, \frac{\langle J(m,c,D)\rangle_{m,c}}{J_{\rm host}}\,, \tag{4.16}$$

where it clearly appears that both the subhalo mass-concentration relation and the mass function will play decisive roles. To further account for any velocity dependence of the annihilation cross section, one has to trade $\xi_{\rm t}$ for $\xi_{\rm t,eff}$ in Eq. (4.14) [see Eq. (**??**)].

We are now equipped to investigate analytically how Sommerfeld effects act on the subhalo boost factor. In the next paragraph, we first review the Sommerfeld-free case before moving to the more complex and intricate velocity-dependent cases induced by the Sommerfeld enhancement.

4.3 Subhalo boost factor: analytical insights

Here, we derive analytical approximations that will allow us to interpret our full results in terms of the driving physical parameters in the calculation, which remain to be determined.

Throughout this part, without so much loss of generality, we will assume that subhalos have NFW inner mass density profiles:

$$\rho(x \equiv r/r_{\rm s}) = \rho_0\, \left\{ f_{\rm nfw}(x) \equiv x^{-1}(1+x)^{-2} \right\} \theta(x_{\rm t} - x)\,, \tag{4.17}$$

where we have introduced the shape function $f_{\rm nfw}(x)$, the dimensionless radius x, and tidal radius $x_{\rm t}$, and where the structural properties such as the scale radius $r_{\rm s}$ and scale density ρ_0 are conventionally fixed by the mass and the mass-concentration relation. We will neglect tidal effects in the following discussion, as they are not critical to develop a good physical understanding of the Sommerfeld effect (we do account for them in the full numerical calculations). Thus, we can first assume that whatever their positions in the host halo, subhalos keep their virial mass, hence $x_{\rm t} = x_{200} = c$. Rigorously, we should also take into account the fact that these structural properties are described by non-trivial PDFs when tidal effects are considered (we do so in the full numerical calculations). For simplicity here, we assume that both the mass function and the mass-concentration relation follow power laws:

$$\begin{cases} \frac{dN_{\rm sub}(m,M_{\rm host})}{dm} & \approx \frac{N_0(M_{\rm host})}{m_0} \left\{ \mu \equiv \frac{m}{m_0} \right\}^{-\alpha} \\ c(m) & \approx c_0\, \mu^{-\varepsilon} \end{cases} \tag{4.18}$$

where m_0 is an arbitrary reference mass, N_0 is the normalization of the number of subhalos, which depends on the host halo mass $M_{\rm host}$, and where we have introduced the dimensionless reduced mass μ. That form of the mass function finds strong theoretical support, as discussed in App. B — see Eq. (B.10). We can further neglect the concentration PDF, and assume that all subhalos of a given mass m have the same concentration $c(m)$. These are good approximations to the more precise description used in our full numerical treatment — our detailed subhalo population model is described in App. B, where we find that

$$\begin{cases} \alpha & \approx 1.96 \\ \varepsilon & \approx 0.05 \end{cases} \tag{4.19}$$

provide a decent matching to the numerical results over a significant subhalo mass range.[5]

Given Eq. (4.18), we can relate the total number of subhalos $N_{\rm tot}$ to the subhalo mass fraction in the host $f_{\rm sub}$ through the minimal and maximal reduced subhalo masses $\mu_{\rm min}$, and the averaged subhalo mass $\langle\mu\rangle_m$ as follows:

$$f_{\rm sub}\,\mu_{\rm host} = N_{\rm tot}\,\langle\mu\rangle_m = \frac{N_0}{(2-\alpha)}\,\mu_{\rm max}^{2-\alpha}\left\{1 - \left[\frac{\mu_{\rm min}}{\mu_{\rm max}}\right]^{2-\alpha}\right\}\,, \tag{4.20}$$

where $\mu_{\rm host} \equiv M_{\rm host}/m_0$ is the reduced dimensionless host halo mass. In the limit $\mu_{\rm max} \gg \mu_{\rm min}$ and if $\alpha < 2$, then we have

$$\begin{cases} N_{\rm tot} & \simeq \frac{N_0}{(\alpha-1)}\,\mu_{\rm min}^{1-\alpha} \simeq \frac{\gamma}{(\alpha-1)}\left\{\frac{\mu_{\rm min}}{\mu_{\rm host}}\right\}^{1-\alpha} \\ \langle\mu\rangle_m & \simeq 10^{-2}\,\frac{(\alpha-1)}{(2-\alpha)}\,\mu_{\rm host}\left\{\frac{\mu_{\rm max}}{\mu_{\rm min}}\right\}^{1-\alpha} \\ f_{\rm sub} & \simeq \frac{N_0}{(2-\alpha)}\,\frac{\mu_{\rm max}^{2-\alpha}}{\mu_{\rm host}} = \frac{\gamma}{(2-\alpha)}\left\{\frac{\mu_{\rm max}}{\mu_{\rm host}}\right\}^{2-\alpha} \simeq \frac{10^{2(\alpha-2)}\gamma}{(2-\alpha)} \sim 34\% \end{cases}. \tag{4.21}$$

We have assumed $\mu_{\rm max} = 10^{-2}\mu_{\rm host}$, and $N_0 = \gamma\mu_{\rm host}^{\alpha-1}$, and except for the approximately universal subhalo mass fraction (before tidal stripping effects), the above values depend on $M_{\rm host} \leftrightarrow \mu_{\rm host}$ — see details in App. B.10.

4.3.1 Subhalo boost factor without Sommerfeld enhancement

Let us consider first annihilation through an s-wave process and a subhalo of virial mass m, concentration c, and located at a radius R in a host halo. The intrinsic annihilation luminosity can be expressed in terms of the effective annihilation volume $\xi_{\rm t}$ introduced in Eq. (4.11), that we can rewrite as

$$\xi_{\rm t}(m,c,R) = \frac{4\pi}{3}r_{\rm s}^3\,\frac{\rho_0^2}{\rho_\circledast^2}\left\{\eta_{\rm t} \equiv 3\int_0^{x_{\rm t}} {\rm d}x\,x^2\,f_{\rm nfw}^2(x)\right\}, \tag{4.22}$$

where $r_{\rm s}$, ρ_0, and $f_{\rm nfw}(x)$ were introduced in Eq. (4.17). We see here that the luminosity is computed within the dimensionless *tidal* radius $x_{\rm t}$ of the subhalo, which, in principle, implies a spatial dependence of the luminosity even for a given virial mass. For an NFW profile, the tidal cut reads

$$\eta_{\rm t} = \eta_{\rm t}(m,c,R) = 1 - (1 + x_{\rm t}(m,c,R))^{-3} \lesssim 1\,. \tag{4.23}$$

From the definition of the virial mass $m = m_{200}$, and assuming that the mass-concentration relation $c(m) = c_{200}(m) = r_{200}(m)/r_{\rm s}(m)$ obeys the power-law function in mass given in Eq. (4.18), we get

$$\xi_{\rm t}(m,c,R) \simeq \xi_{\rm t}(m) = \xi_{\rm t}(\mu = m/m_0) = \xi_0\,\mu^{1-3\varepsilon}\,, \tag{4.24}$$

where

$$\xi_0 \equiv \frac{200\,\rho_{\rm c}\,m_0}{\rho_\circledast^2}\,\frac{c_0^3}{A^2}\,\eta_{\rm t}\,, \tag{4.25}$$

[5] We find that $c_0 \simeq 12.9$ for $m_0 = 10^{10}\,{\rm M}$ [87] allows to get close to the parametric concentration function of field halos provided in ref. [91], which characterizes the initial concentration function of subhalos in most of the mass range of interest in our study.

ρ_c being the critical density today, and

$$A = A(c) = 3\left[\ln(c+1) - \frac{c}{(c+1)}\right] \approx \text{constant } \mathcal{O}(1-10)\,. \tag{4.26}$$

As already mentioned above, we can at first order neglect the position of a subhalo in its host, characterized here by the radial coordinate R. This is because subhalos that will dominantly contribute to the γ-ray flux are typically those located beyond the scale radius of the host (often called "field" subhalos), for which tidal stripping effects are not so important. These field subhalos actually constitute the bulk of the subhalo population (in an NFW host halo, subhalos located beyond the scale radius represent $\gtrsim 90\%$ of the whole subhalo population). This holds true while the γ-ray flux is integrated over a volume bigger than the one encompassing the scale radius of the host (a more involved description is necessary for hosts much more extended than the angular resolution of the telescope, or for hosts that have experienced significant tidal stripping and have sizes of the order or less than their scale radii).

From the annihilation volume ξ_t, we get an analytical expression for the point-like J-factor given in Eq. (4.14):

$$J(m) = J(\mu) = J_0 \times (2\,\overline{v}_0)^p \times \mu^{1-3\varepsilon+p\nu}\,, \tag{4.27}$$

where

$$J_0 \equiv \frac{\rho_\circledast^2\,\xi_0}{D^2} = \frac{200\,\rho_c\,m_0}{D^2}\,\frac{c_0^3}{A^2}\,\eta_t\,, \tag{4.28}$$

with ξ_0 given in Eq. (4.25), and $\rho_\circledast$ the arbitrary reference mass density introduced in Eq. (4.11)—it is then clear that neither $J(m)$ nor J_0 depend on $\rho_\circledast$, as shown explicitly in the above equations.

Note that the p-wave annihilation case is actually included in the previous two expressions, by setting $p = 2$ (we remind that $p = 0$ stands for the s-wave case). In fact, we have implicitly assumed that the radial profile of $\langle v^2\rangle_v(r)\,\rho^2(r) \approx \langle v^2\rangle_{v,V_h}\,\rho^2(r)$, where $\langle v^2\rangle_{v,V_h}$ is taken constant over the whole halo volume V_h. Assuming $\langle v^2\rangle_{v,V_h} = \overline{v}_0^2\,\mu^{2\nu} \ll 1$ allows us to characterize the p-wave suppression factor in terms of halo mass. The spectral index ν will be specified later, and $\overline{v}_0$ is an arbitrary reference velocity associated with a halo of arbitrary reference mass m_0.

From this, we can predict the ratio of J-factors of two point-like halos of different masses, m_1 and m_2, and respectively located at distances D_1 and D_2 from the observer:

$$\frac{J(m_1)}{J(m_2)} = \left\{\frac{D_2}{D_1}\right\}^2 \left\{\frac{m_1}{m_2}\right\}^{1-3\varepsilon+p\,\nu}\,. \tag{4.29}$$

This will be helpful to understand forthcoming results.

Let us now come back to our main working equation, Eq. (4.27). As we shall see just below, it will allow us to estimate the total contribution of subhalos to the γ-ray J-factor, assuming that the whole population of subhalos is contained within the field of view of the instrument. Indeed, this simply amounts to convolving Eq. (4.27) with the subhalo mass function, which we take as the power law of index α introduced in Eq. (4.18). Then, the total J-factor associated with the contribution of the whole subhalo population reads:

$$J_{\rm sub} = J_{\rm sub,0}\,\mu_{\rm min}^{-\alpha_{\rm boost}} \left[1 - \left(\frac{\mu_{\rm max}}{\mu_{\rm min}}\right)^{-\alpha_{\rm boost}}\right], \tag{4.30}$$

where we have introduced the effective boost index $\alpha_{\rm boost}$. We have also used

$$J_{\rm sub,0}(\mu_{\rm host}) \equiv \frac{N_0(\mu_{\rm host})}{\alpha_{\rm boost}} \times J_0 \times (2 v_0)^p \,. \tag{4.31}$$

The critical parameter in the above result is the effective boost index,

$$\alpha_{\rm boost} \equiv \alpha + 3\varepsilon - 2 - p\nu\,, \tag{4.32}$$

which fully characterizes the part of the mass function that sets the overall subhalo population luminosity. Indeed, three different regimes arise:

$$\begin{cases} \alpha_{\rm boost} & > 0 \implies m_{\rm min}\text{-dominated regime (strong boost)} \\ \alpha_{\rm boost} & = 0 \implies \text{democratic regime} \\ \alpha_{\rm boost} & < 0 \implies m_{\rm max}\text{-dominated regime (weak boost)}\,. \end{cases} \tag{4.33}$$

The positive sign convention has been chosen such that the boost is strong if $\alpha_{\rm boost} > 0$, which means that the smallest, most numerous, and most concentrated subhalos carry the dominant contribution to the annihilation rate. The democratic regime corresponds to a logarithmic dependence in the subhalo masses, $\propto \ln(m_{\rm max}/m_{\rm min})$, in Eq. (4.30). The sign of $\alpha_{\rm boost}$ is therefore crucial here, as already known from past studies. From this very simple equation, since $\varepsilon \approx 0.05$, we understand that changing α from 1.9 to 2 amounts to going from an $m_{\rm max}$-dominated regime to an $m_{\rm min}$-dominated regime for an s-wave annihilation. In the latter case, the overall subhalo population luminosity becomes very sensitive to the subhalo minimal mass cutoff $m_{\rm min}$, as is well known. This is reinforced by the fact that the total number of subhalos $N_{\rm tot} \overset{\sim}{\propto} \mu_{\rm min}^{1-\alpha}$. We shall see later on that this effective power-law index $\alpha_{\rm boost}$ can also be expressed analytically in the Sommerfeld-enhanced case, which will allow us to use a reasoning very similar to the one presented here. Before moving to the Sommerfeld-enhanced case, let us just introduce an analytical expression for the Sommerfeld-free boost factor $\mathcal{B}$:

$$\mathcal{B} - 1 \simeq \frac{J_{\rm tot}^{\rm sub}}{J_{\rm host}} = \frac{J_{\rm sub,0}}{J_{\rm host}}\,\mu_{\rm min}^{-\alpha_{\rm boost}} \left\{1 - \left[\frac{\mu_{\rm max}}{\mu_{\rm min}}\right]^{-\alpha_{\rm boost}}\right\}, \tag{4.34}$$

where $J_{\rm host}$ is the J-factor calculated assuming a fully smooth density profile for the host halo ($\propto \int {\rm d}s\,\langle\rho_{\rm host}\rangle^2$) — we call this the smooth approximation. The "-1" on the left-hand side implicitly assumes that the contribution to the J-factor of the smooth part of the actual inhomogeneous host halo, $J_{\rm sm}$, equals the smooth approximation, but one should keep in mind that formally $J_{\rm host} \gtrsim J_{\rm sm}$. Given the analytical expressions introduced above, we finally get

$$\mathcal{B}-1\simeq \frac{N_0\,A_{\rm host}^2}{\alpha_{\rm boost}\,A_{\rm sub}^2\,\mu_{\rm host}^{1-3\varepsilon}}\,\mu_{\rm min}^{-\alpha_{\rm boost}}\left\{1-\left[\frac{\mu_{\rm max}}{\mu_{\rm min}}\right]^{-\alpha_{\rm boost}}\right\} \tag{4.35}$$
$$\simeq \frac{\gamma}{\alpha_{\rm boost}}\,\frac{A_{\rm host}^2}{A_{\rm sub}^2}\left\{\frac{\mu_{\rm min}}{\mu_{\rm host}}\right\}^{-\alpha_{\rm boost}}\left\{1-\left[\frac{\mu_{\rm max}}{\mu_{\rm min}}\right]^{-\alpha_{\rm boost}}\right\},$$

where $A_{\rm host}$ and $A_{\rm sub}$, introduced in Eq. (4.26), are shown explicitly for definiteness because their ratio is not strictly 1. Note that the power-law dependence in the bracket on the right-hand-side becomes logarithmic, $\ln(\mu_{\rm max}/\mu_{\rm min})$, when $\alpha_{\rm boost}=0$ (democratic regime, for which each decade of subhalo mass contributes the same signal). In the latest equation line above, we have traded N_0 for its dependence in $\mu_{\rm host}$ according to Eq. (B.10), with γ a constant predicted from a merger-tree calculations, which provides a very compact expression that depends only on the subhalo-to-host mass ratio and on the subhalo mass index.

From the numerical values introduced in Eq. (4.19), we get $\alpha_{\rm boost}(p=0)\simeq 0.11>0$ for s-wave annihilation processes, hence a significant boost factor dominated by the contribution of the lightest subhalos to the annihilation rate (the last term in brackets in the right-hand-side of the above equation simplifies to 1):

$$\mathcal{B}_{s\text{-wave}}-1\simeq \frac{N_0\,A_{\rm host}^2}{\alpha_{\rm boost}\,A_{\rm sub}^2\,\mu_{\rm host}^{1-3\varepsilon}}\,\mu_{\rm min}^{-\alpha_{\rm boost}}\left\{1-\left[\frac{\mu_{\rm max}}{\mu_{\rm min}}\right]^{-\alpha_{\rm boost}}\right\} \tag{4.36}$$
$$\simeq \frac{\gamma}{\alpha_{\rm boost}}\,\frac{A_{\rm host}^2}{A_{\rm sub}^2}\left\{\frac{\mu_{\rm min}}{\mu_{\rm host}}\right\}^{-\alpha_{\rm boost}}.$$

In that case, the boost factor is fixed by the hierarchy between the host halo mass and the minimal subhalo mass, and modulated by the amplitude of the effective boost index $\alpha_{\rm boost}$.

There is no boost factor in the p-wave annihilation case because since the cross section is proportional to v^2 and the internal dispersion velocity decreases with the mass of a structure, the signal contributed by subhalos is strongly reduced with respect to that contributed by the host halo. Still under the assumption of $\langle v^2\rangle\propto m^{2\,\nu}$, where ν will be evaluated later to be $\sim 1/3$, and that for a halo of density profile $\rho(r)$, $\langle v^2\,\rho^2\rangle(r)\approx\langle v^2\rangle\rho^2(r)\propto m^{2\,\nu}\,\rho^2(r)$, then it is easy to show from Eq. (4.30) that

$$\mathcal{B}_{p\text{-wave}}-1\approx-\frac{\gamma}{\alpha_{\rm boost}}\,\frac{A_{\rm host}^2}{A_{\rm sub}^2}\left\{\frac{\mu_{\rm max}}{\mu_{\rm host}}\right\}^{-\alpha_{\rm boost}}\approx-\frac{\gamma}{\alpha_{\rm boost}}\left\{\frac{1}{100}\right\}^{-\alpha_{\rm boost}}\ll 1\,. \tag{4.37}$$

We have used the fact that $m_{\rm max}\simeq M_{\rm host}/100$, that the boost mass index for the p-wave annihilation $\alpha_{\rm boost}(p=2)=\alpha+3\varepsilon-2(1+\nu)\approx-0.56<0$ [see Eq. (4.30)], and that $A_{\rm host}^2\approx A_{\rm sub}^2$. In that approximation, valid as long as $J_{\rm sm}\simeq J_{\rm host}$, then clearly $\mathcal{B}_{p\text{-wave}}\simeq 1$. If $J_{\rm sm}<J_{\rm host}$, which can be the case if the mass fraction in subhalos is significant within the scale radius of the host halo, then we could even have $\mathcal{B}_{p\text{-wave}}<1$, which would imply that subhalos would no longer act as a boost factor, but rather as a damping factor to the signal. We will see just below that this picture changes radically when Sommerfeld effects kick in.

4.3.2 Sommerfeld enhancement at the level of one (sub)halo.

In this part, we initiate the derivation of an analytical expression for the subhalo boost factor further subject to Sommerfeld-enhancement effects. The derivation proceeds in three steps. We first develop an analytical understanding of the Sommerfeld effect at the level of a single structure (this paragraph). This is a crucial step before generalizing to a population of structures in the next paragraph, where we derive a full analytical expression for the Sommerfeld-enhanced J-factor associated with a subhalo population. Finally, we determine the overall boost factor by calculating the ratio between the Sommerfeld enhanced J-factor for the subhalo population and that of the host halo. This series of analytical developments is helpful to reach a clearer physical understanding of the intricate phenomena at play in terms of the specific particle physics model parameters, here characterized by the reduced DM Bohr radius, ϵ_ϕ. We recall that a Sommerfeld configuration is entirely fixed by ϵ_ϕ and the coupling strength α_χ in our simplified model. Decreasing ϵ_ϕ roughly amounts to increasing the DM particle mass or the interaction coupling constants, or decreasing the mediator mass, assuming all of the other parameters are fixed.

We start by examining the overall Sommerfeld enhancement for one halo. The particle velocity dependent ansätze introduced in Sec. 3.1.2 suggest the possibility of formulating an effective Sommerfeld enhancement factor at the level of an entire (sub)halo. This can be done by picking the most representative value of the particle velocity in a DM structure, which depends on the structure mass (a mere consequence of the virial theorem for systems in dynamical equilibrium). If such a characteristic velocity in a (sub)halo can be estimated (e.g. from its averaged velocity dispersion), then one can effectively relate an average Sommerfeld enhancement to the (sub)halo mass. We can actually expect the characteristic velocity of a structure of mass m, $\overline{v}(m)$, to scale like

$$\overline{v}(m) \sim \sqrt{\langle v^2\rangle} \sim \sqrt{\frac{G_{\rm N}\, m(r_{\rm c})}{r_{\rm c}}}\,, \tag{4.38}$$

where $r_{\rm c}$ is some characteristic radius to be determined. For the sake of generality, considering that we can also relate that characteristic radius to the virial mass, we shall assume

$$\overline{v} = \overline{v}_0\, \mu^\nu\,, \tag{4.39}$$

where ν is the power-law index that relates the characteristic dimensionless velocity $\overline{v}/\overline{v}_0$ to the dimensionless (sub)halo mass $\mu = m/m_0$. Parameter $\overline{v}_0$ is the characteristic velocity associated with the arbitrary reference virial mass $m_0 = m_0(r_{200,0})$ of an NFW halo, obeying the general relation

$$\overline{v}(m) \equiv \omega_0 \sqrt{\frac{G_{\rm N}\, m(r_{\rm s})}{r_{\rm s}}}\,, \tag{4.40}$$

where $\omega_0 \sim 1$ is a tuning parameter, $r_{\rm s}$ is the scale radius associated with some halo of virial mass m, and $m(r_{\rm s})$ is the mass contained within $r_{\rm s}$. The speed parameter ω_0 is meant to optimize the estimates of the speed moments relevant to the Sommerfeld enhancement over a given structure with a single value of $\overline{v}$, which should capture different regimes at the same time ($\propto \langle 1/v\rangle$ or $\langle 1/v^2\rangle$). By picking the subhalo characteristic mass and size at the scale radius of an NFW halo, it is easy to show that

$$\nu \simeq \frac{1}{2}\left\{\frac{2}{3} - \varepsilon\right\} \approx \frac{1}{3}\,, \tag{4.41}$$

where ε is the power-law index of the approximate concentration-mass relation given in Eq. (4.18). More concretely, in numbers, this gives

$$\overline{v} \simeq 6 \times 10^{-6}\, \omega_0 \left\{ \frac{m}{10^6\,\mathrm{M}_\odot} \right\}^{1/3} , \tag{4.42}$$

with $\omega_0 \sim 1$.

Even though $\overline{v}$ can be used as a characteristic velocity where most of the phase-space distribution is supposed to concentrate, one should still not forget that the speed of any DM particle bound to a (sub)halo can actually take any value between 0 and the escape speed. The Sommerfeld factor should therefore be integrated over the full available range—hence different parts of the (sub)halo phase-space distribution may contribute to different Sommerfeld regimes, not necessarily to a single regime. However, the characteristic velocity-mass relation written above is meant to reflect the typical velocity at which the bulk of annihilations in a (sub)halo of mass m proceeds, which turns out to be a good approximation.

We can now opportunely reformulate the ansatz of Eq. (3.10) by replacing the dependence on velocity v by a dependence on the characteristic (sub)halo velocity $\overline{v}(m)$, and then by a dependence in (sub)halo mass m. This gives

$$\begin{aligned} \overline{\mathcal{S}}_{\text{no-res}}(m, \epsilon_\phi) &= \mathcal{S}_{\text{no-res}}(\overline{v}(m), \epsilon_\phi) \\ &= S_0 \left(\frac{m}{\tilde{m}_{\max}} \right)^{-\nu} \left[1 + S_1^{-\frac{\bar{s}_{v,c}}{(1+p)}} \left(\frac{m}{\tilde{m}_{\text{sat}}} \right)^{-\nu \bar{s}_{v,c}} \right]^{-\frac{(1+p)}{\bar{s}_{v,c}}} , \end{aligned} \tag{4.43}$$

with m the (sub)halo mass, and the constants S_0 and S_1 given in Eq. (3.11). This ansatz is essentially valid for $\overline{v}(m) \leqslant \tilde{v}_{\max}$, or equivalently $m \leqslant \tilde{m}_{\max} = m(\tilde{v}_{\max})$, for which the Sommerfeld effect starts being operative. This maximal mass $\tilde{m}_{\max}$ should not be confused with the maximal subhalo mass $m_{\max}$ in a given host halo; it is really the (sub)halo mass beyond which the characteristic velocity of DM is too large for the Sommerfeld enhancement to be turned on efficiently. The power-law indices have been introduced in Eq. (3.10) up to a correction by the speed-to-mass index ν, introduced in Eq. (4.39), and evaluated in Eq. (4.41). Switching from velocity to mass dependence, the power-law index in the Coulomb regime becomes $-\nu$. We have also introduced $\tilde{m}_{\text{sat}} = \tilde{m}_{\text{sat}}(\epsilon_\phi) = m(\tilde{v}_{\text{sat}}(\epsilon_\phi))$, the halo mass below which most of the halo phase-space volume is in the Sommerfeld saturation regime and resonances may appear. The different velocity dependencies in the different regimes are summarized below Eq. (3.10).

Similarly to the corresponding velocities, the transition masses introduced above can be expressed in terms of the main Sommerfeld parameters:

$$\begin{cases} \tilde{m}_{\max} &= m_0 \left(\frac{\tilde{v}_{\max}}{\overline{v}_0} \right)^{\frac{1}{\nu}} = m_0 \left(\frac{\pi\, \alpha_X}{\overline{v}_0} \right)^{\frac{1}{\nu}} \approx 9.6 \times 10^{17}\,\mathrm{M}_\odot \left(\frac{\alpha_X}{0.01} \right)^3 \\ \tilde{m}_{\text{sat}}(\epsilon_\phi) &= m_0 \left(\frac{\tilde{v}_{\text{sat}}}{\overline{v}_0} \right)^{\frac{1}{\nu}} = m_0 \left(\frac{\alpha_X\, \epsilon_\phi}{\overline{v}_0} \right)^{\frac{1}{\nu}} \approx 7.6 \times 10^{9}\,\mathrm{M}_\odot \left(\frac{\alpha_X}{0.01} \times \frac{\epsilon_\phi}{0.01} \right)^3 \\ \tilde{m}_{\text{unit}} &= m_0 \left(\frac{\tilde{v}_{\text{unit}}}{\overline{v}_0} \right)^{\frac{1}{\nu}} \approx 8 \times 10^{-4}\,\mathrm{M}_\odot \left(\frac{\alpha_X}{0.01} \right)^{12} \end{cases} . \tag{4.44}$$

We stress that $\tilde{m}_{\text{sat}}$ is a smooth function of ϵ_ϕ even on resonances. This saturation mass defines a threshold in phase space: halos with masses below $\tilde{m}_{\text{sat}}$ will have most of their phase-space distribution in the saturation regime. Resonant saturation masses are simply characterized by $\tilde{m}_{\rm sat} = \tilde{m}_{\rm sat}(\epsilon_\phi = \epsilon_\phi^{{\rm res},n})$. The power-law dependence of $\tilde{m}_{\rm sat} \propto \epsilon_\phi^{1/\nu}$ can be predicted from Eq. (4.41) to be close to $\tilde{m}_{\rm sat} \propto \epsilon_\phi^3$. This is actually recovered from a numerical calculation of Eq. (4.40), as shown in the bottom right panel of Fig. 4, which will be discussed more thoroughly later on. Finally, the numerical estimate of $\tilde{m}_{\rm unit}$ given above can make us anticipate the important role it will play in the determination of the resonant peak amplitudes, and then the intrinsic limit set in the potential of the latter to probe the minimal (sub)halo masses if $\tilde{m}_{\rm unit} > m_{\rm min}$. We stress that this mass boundary $\tilde{m}_{\rm unit}$ is extremely sensitive to the DM fine structure constant, as it scales like $\sim \alpha_\chi^{12}$ in our approximate parametric regularization (but see the discussion at the end of Sec. 3.1.1).

We can get a similar formulation for the halo mass dependent Sommerfeld factor on resonances by inserting $v = \overline{v}(m)$ in Eq. (3.13):

$$\begin{aligned} \overline{\mathcal{S}}_{{\rm res},n}(m,\epsilon_\phi) &= \mathcal{S}_{{\rm res},n}(\overline{v}(m),\epsilon_\phi) \\ &\overset{n\geqslant 1+\frac{p}{2}}{=} S_0^{\rm res}\left(\frac{\tilde{m}_{\rm sat}(\epsilon_\phi)}{\tilde{m}_{\rm max}}\right)^{-\nu}\left(\frac{m}{\tilde{m}_{\rm sat}(\epsilon_\phi)}\right)^{-\nu(2-p)}\left(1+\frac{\tilde{m}_{\rm unit}^\nu}{m^\nu}\right)^{-2} \\ &\quad\times\theta\left(\tilde{m}_{\rm sat}(\epsilon_\phi)-m\right)\delta_{\epsilon_\phi/\{\epsilon_\phi^{{\rm res},n}\}} \end{aligned} \tag{4.45}$$

where $\delta_{\epsilon_\phi/\{\epsilon_\phi^{{\rm res},n}\}}$ was defined in Eq. (3.14), and where we see that a halo can efficiently trigger resonances provided its mass $m \ll \tilde{m}_{\rm sat}(\epsilon_\phi^{{\rm res},n})$.

A full understanding of the mass dependence in the resonant regime actually follows from that of the velocity dependence discussed around Eq. (3.13) and illustrated in Fig. 2, keeping in mind that $\overline{v} \propto m^\nu$—focus on dark blue curves in both panels. Indeed, beside the step function responsible for turning resonances on or off, the only direct dependence of the above resonant Sommerfeld factor on the halo mass m shows up in the s-wave case, down to the unitarity limit characterized by $\tilde{m}_{\rm unit}$. In contrast, the amplitudes of p-wave resonant peaks ($p = 2$) do not, essentially, depend on m, which is reminiscent from the fact that the effective Sommerfeld enhancement (which includes the v^2 p-wave suppression factor as well) is velocity independent on p-wave peaks, as shown in Eq. (3.13) and in the right panel of Fig. 2. The p-wave suppression factor re-appears once the unitarity bound is reached, and translates into a mass-dependent suppression factor of $(1+\tilde{m}_{\rm unit}/m)^{-2\nu}$ that becomes operative when $m \lesssim \tilde{m}_{\rm unit}$. Hence, the potential numerical error made by converting a local velocity into a global velocity is significantly reduced on p-wave resonances (except close to the step function threshold, $m \lesssim \tilde{m}_{\rm sat}$, where only part of the phase-space distribution lies in the saturation regime, or close to the unitarity bound).

All this allows us to translate the velocity-dependent ansatz of Eq. (3.17) in terms of a halo-mass-dependent and generic effective Sommerfeld factor,

$$\begin{aligned} \overline{\mathcal{S}}(m,\epsilon_\phi) &= \overline{\mathcal{S}}_{\text{no-res}}(m,\epsilon_\phi)\left(1-\sum_{n=1+\frac{p}{2}}\delta_{\epsilon_\phi/\{\epsilon_\phi^{{\rm res},n}\}}\right)+\sum_{n=1+\frac{p}{2}}\overline{\mathcal{S}}_{{\rm res},n}(m,\epsilon_\phi) \\ &\propto \mu^{-s_m}\,, \end{aligned} \tag{4.46}$$

where $\overline{\mathcal{S}}_{\rm res}(m)$ is the transcript of $\mathcal{S}_{\rm res}(v)$ of Eq. (3.13) in terms of mass m (evaluated at $v = \overline{v}(m)$). At the level of a (sub)halo, the Sommerfeld enhancement can be written as power-law in mass, whose effective index $s_m = \nu\, s_v$ can be readily inferred from the possible values of s_v listed in Eq. (3.18):

$$s_m = \begin{cases} \nu & \text{(Coulomb regime)} \\ -\nu\, p & \text{(non-resonant saturation regime)} \\ \nu\,(2-p) & \text{(resonances)} \longrightarrow -p\,\nu \ (\text{if } m \lesssim \tilde{m}_{\rm unit}) \end{cases} . \tag{4.47}$$

We stress that the correspondence between the characteristic speed and the (sub)halo mass in the Sommerfeld factor has only a global meaning—we shall refer to $\overline{v}$ as the characteristic speed in a (sub)halo, and to v as an arbitrary or local speed from now on. Indeed, as mentioned above, the DM speed v in a subhalo can take any value between $\sim 0 \ll \overline{v}$ and the escape speed $v_{\rm e} \gtrsim \overline{v}$. The whole (sub)halo lies in the Sommerfeld-enhancement regime typically when $v_{\rm e} \leq \tilde{v}_{\rm max}$. Then, the part of the phase-space distribution located between $\tilde{v}_{\rm sat}(\epsilon_\phi)$ and $\tilde{v}_{\rm max}$ will essentially participate in the Coulomb enhancement ($\propto 1/v$), while the part of the phase-space distribution below $\tilde{v}_{\rm sat}(\epsilon_\phi)$ will instead contribute in the saturation regime. In the latter case, the speed dependence saturates, except at resonances, which are triggered at special values of $\epsilon_\phi = \epsilon_\phi^{{\rm res},n}$ at which all of the phase-space distribution located below $\tilde{v}_{\rm sat}$ participates in the enhancement ($\propto 1/v^2$ for s-wave annihilation). The very fact that different parts of the phase-space distribution of a (sub)halo feed different Sommerfeld regimes implies that the ansatz of Eq. (4.46) cannot lead to accurate predictions. However, we shall see below that it is still very powerful to capture the main phenomenological features of the intricate phenomena at play.

It is instructive to further inspect the relative amplitudes of resonant peaks when the Sommerfeld factor is applied over an entire halo. To do so, let us briefly convert the ratio of resonance-to-baseline enhancement in the saturation regime, introduced in Eq. (3.16), in terms of an overall mass-dependent ratio:

$$\overline{\mathcal{R}}(\mu,\epsilon_\phi) = \left(\frac{\pi}{6}\right)^2 \left\{\frac{\mu}{\tilde{\mu}_{\rm sat}}\right\}^{-2\,\nu} \left\{1+\left(\frac{\tilde{\mu}_{\rm unit}}{\mu}\right)^{\nu}\right\}^{-2} \stackrel{\sim}{\propto} \epsilon_\phi^2\,\mu^{-2/3} . \tag{4.48}$$

Therefore, the relative amplitudes of peaks scale like $\overline{v}^{-2}(m) \leftrightarrow m^{-2/3}$ for both s- and p-wave processes, which means that resonances are more pronounced for less massive halos (though saturating when $m \lesssim \tilde{m}_{\rm unit}$), while still suppressed like $(\epsilon_\phi \sim \epsilon_\phi^{{\rm res},n})^2$ at higher and higher resonances, with respect to the saturation baseline.

We pursue by writing down the analytical expression obtained for the J-factor corrected for the Sommerfeld enhancement at the level of one structure of mass m (or dimensionless reduced mass $\mu = m/m_0$), combining Eq. (4.27) and Eq. (4.46):

$$\begin{aligned} J_{\mathcal{S}}(m,\epsilon_\phi) &= J_{\mathcal{S}}(\mu,\epsilon_\phi) = J(\mu) \times \left\{\frac{\mu}{\tilde{\mu}_{\rm max}}\right\}^{-p\nu} \times \overline{\mathcal{S}}(\mu,\epsilon_\phi) \\ &= J_0 \times (2\tilde{v}_{\rm max})^p \times \mu^{1-3\varepsilon}\,\overline{\mathcal{S}}(\mu,\epsilon_\phi) \\ &= J_{0,\mathcal{S}}(\epsilon_\phi)\,\mu^{1-3\varepsilon-s_m} , \end{aligned} \tag{4.49}$$

which gives an implicit definition to the factor $J_{0,\mathcal{S}}(\epsilon_\phi)$, and where the Sommerfeld mass index s_m was introduced in Eq. (4.47). Note that the factor $(\mu/\tilde{\mu}_{\rm max})^{-p\nu} = (\overline{v}/\tilde{v}_{\rm max})^{-p}$ is simply there not to double count the v^2 dependence of the p-wave cross section that is included in the definition of the nominal J-factor $J(m)$ [see Eq. (4.27)], which we have also conveniently absorbed in the definition of the effective Sommerfeld enhancement factor [see Eq. (3.8)].

Like in the Sommerfeld-free case, it is interesting to determine the ratio of J-factors for halos of different masses, say $m_1 \leftrightarrow \mu_1$ and $m_2 \leftrightarrow \mu_2$ (assumed here to be located at di erent distances, D_1 and D_2, from the observer):

$$\frac{J_{\mathcal{S}}(m_1, \epsilon_\phi)}{J_{\mathcal{S}}(m_2, \epsilon_\phi)} = \left\{\frac{D_2}{D_1}\right\}^2 \left\{\frac{m_1}{m_2}\right\}^{1-3\varepsilon} \left\{\frac{m_1^{-s_{m_1}}}{m_2^{-s_{m_2}}}\right\}, \tag{4.50}$$

where s_{m_i} refers to the Sommerfeld mass index of the halo of index i. In the absence of Sommerfeld enhancement, $J_1/J_2 \sim (m_1/m_2)^{(1-3\varepsilon+p\,\nu)}$ according to Eq. (4.29), where the $\overline{v}^p(m)$ factor relevant to the p-wave case remains (contributing νp in the power-law mass index, a contribution hidden in the definition of the index s_m in the Sommerfeld-enhanced case).

Now, we determine the asymptotic expressions for the different Sommerfeld regimes, which will turn useful later because subhalos are not necessarily all in the same Sommerfeld regime, nor necessarily in the same Sommerfeld regime as the host halo itself. This gives:

- **Coulomb regime** ($\mu \geqslant \tilde{\mu}_{\rm sat}(\epsilon_\phi)$)**:**

$$\begin{aligned} J_{\mathcal{S}}(\mu, \epsilon_\phi) &\xrightarrow[\text{Coulomb}]{\mu \geqslant \tilde{\mu}_{\rm sat}} J_0\,(2\,\tilde{v}_{\rm max})^p\, S_0\, \tilde{\mu}_{\rm max}^{1-3\varepsilon} \left\{\frac{\mu}{\tilde{\mu}_{\rm max}}\right\}^{1-3\varepsilon-\nu} \\ &\propto \quad \mu^{1-3\varepsilon-\nu}\,. \end{aligned} \tag{4.51}$$

- **Saturation** ($\mu \leqslant \tilde{\mu}_{\rm sat}(\epsilon_\phi)$)**:**

$$\begin{aligned} J_{\mathcal{S}}(\mu, \epsilon_\phi) &\xrightarrow[\text{saturation}]{\mu \leqslant \tilde{\mu}_{\rm sat}} J_0\,(2\,\tilde{v}_{\rm max})^p\, S_0\, S_1 \left\{\frac{\tilde{\mu}_{\rm max}}{\tilde{\mu}_{\rm sat}}\right\}^{\nu} \tilde{\mu}_{\rm sat}^{1-3\varepsilon} \left\{\frac{\mu}{\tilde{\mu}_{\rm sat}}\right\}^{1-3\varepsilon+\nu p} \\ &\propto \quad \epsilon_\phi^{-(1+p)}\, \mu^{1-3\varepsilon+\nu p}\,. \end{aligned} \tag{4.52}$$

- **Resonances** ($\mu \ll \tilde{\mu}_{\rm sat}(\epsilon_\phi^{{\rm res},n})$)**:**

$$\begin{aligned} J_{\mathcal{S}}(\mu, \epsilon_\phi = \epsilon_\phi^{{\rm res},n}) &\xrightarrow[\text{resonance/s}]{\mu \ll \tilde{\mu}_{\rm sat}} J_0\,(2\tilde{v}_{\rm max})^p\, S_0^{\rm res} \left\{\frac{\tilde{\mu}_{\rm max}}{\tilde{\mu}_{\rm sat}}\right\}^{\nu} \tilde{\mu}_{\rm sat}^{1-3\varepsilon} \left\{\frac{\mu}{\tilde{\mu}_{\rm sat}}\right\}^{1-3\varepsilon-\nu(2-p)} \\ &\quad\times \left\{1 + \left(\frac{\tilde{\mu}_{\rm unit}}{\mu}\right)^{\nu}\right\}^{-2} \\ &\propto \quad \epsilon_\phi^{(1-p)}\, \mu^{1-3\varepsilon-\nu(2-p)}\,. \end{aligned} \tag{4.53}$$

The full mass dependence (which derives from velocity dependence) of the overall Sommerfeld-enhanced $J_{\mathcal{S}}$ factor at the level of a single halo is shown in Fig. 3, where the power-law scalings derived just above are illustrated for different values of the reduced Bohr radius ϵ_ϕ, hence for different particle physics configurations. We actually took the same reference cases as in Fig. 2, $\epsilon_\phi = 0.1$ (moderate enhancement), 10^{-3} (significant enhancement), and the $n = 8$ resonance popping up at an intermediate value of $\epsilon_\phi = \epsilon_\phi^{{\rm res},8} \simeq 10^{-2}$ (strong enhancement). The corresponding saturation masses $\tilde{m}_{\rm sat}(\epsilon_\phi)$ are reported as dashed vertical lines, which mark the transition between the Coulomb regime domination of the phase-space volume ($m > \tilde{m}_{\rm sat}$) and the saturation regime domination ($m < \tilde{m}_{\rm sat}$). We see that in both the s- (left panel) and p-wave (right panel) cases, this transition is characterized by a change

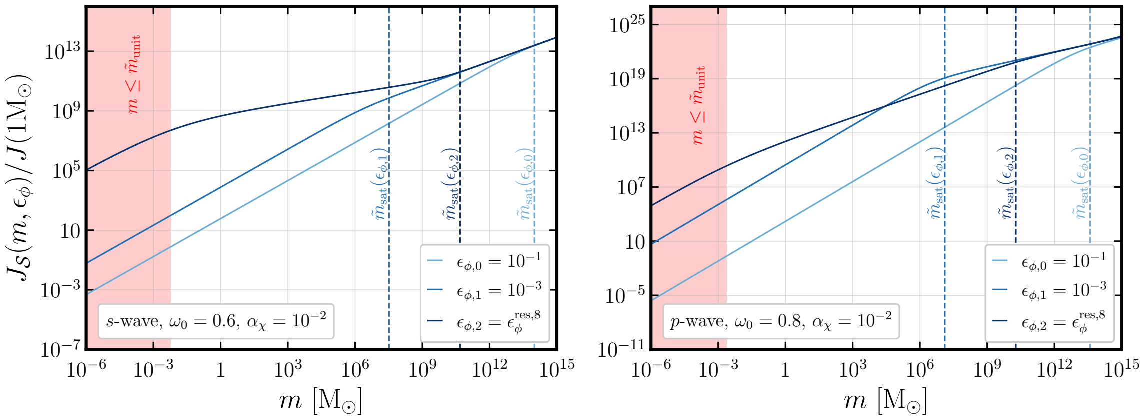

Figure 3. Sommerfeld-enhanced $J_{\mathcal{S}}$ factor as a function of halo mass, for different values of the reduced Bohr radius ϵ_ϕ: a large value 0.1, a small value of 10^{-3}, and an intermediate value of $\sim 10^{-2}$ sitting on the $n = 8$ resonance. This figure is somewhat a translation of the local effective Sommerfeld factor as a function of DM velocity shown in Fig. 2 (times the nominal J factor). Transition from Coulomb to saturation regimes occurs around $\tilde{m}_{\rm sat}(\epsilon_\phi)$, reported as vertical dashed lines. **Left panel:** s-wave case. **Right panel:** p-wave case.

of logarithmic slope for $J_{\mathcal{S}}$, as analytically predicted above. The important point to note here is that the slope gets generically steeper below the saturation mass (stronger decrease with decreasing mass), except on the resonance, where the mass dependence is much shallower toward low masses down to the unitary mass $\tilde{m}_{\rm unit}$ (from both panels, we see that the resonant curves maintain a higher level of $J_{\mathcal{S}}$ factor compared to non-resonant curves, as m decreases down to $\tilde{m}_{\rm unit}$). This will obviously have strong consequences when integrated over a subhalo population, whose power-law mass function will act as an extra weight in favor of low-mass subhalos.

We are now equipped with all necessary analytical results to understand the Sommerfeld enhancement at the level of an entire halo. A final characteristic ingredient to better identify the Sommerfeld regime a given a (sub)halo of mass m should fall in is the value of ϵ_ϕ for which that halo would transition from the Coulomb to the saturation regime. Since the saturation mass $\tilde{m}_{\rm sat}$ is defined from ϵ_ϕ, we can conversely assign a reference value $\epsilon_\phi^{\rm sat}(m)$ to a halo of virial mass m such that

$$\begin{aligned} \tilde{m}_{\rm sat}(\epsilon_\phi^{\rm sat}) &= m \quad (\text{definition of } \epsilon_\phi^{\rm sat}) \\ \Rightarrow \epsilon_\phi^{\rm sat}(m) &\simeq 0.01 \times \left\{\frac{\alpha_X}{0.01}\right\}^{-1} \times \left\{\frac{m}{6\times 10^9\,\mathrm{M}}\right\}^{\nu} \tilde{\propto}\, m^{1/3}\,. \end{aligned} \tag{4.54}$$

From this definition, we can have a better intuition of the Sommerfeld enhancement regime in which a (sub)halo sits: if $\epsilon_\phi > \epsilon_\phi^{\rm sat}(m)$ ($\epsilon_\phi < \epsilon_\phi^{\rm sat}(m)$), then most of the halo phase-space distribution is located in the saturation regime (Coulomb regime, respectively).

It is also helpful to understand the dependence on ϵ_ϕ. At fixed values of the coupling strength α_χ, decreasing ϵ_ϕ amounts to exploring different particle physics model configurations (increasing the DM particle mass, or equivalently decreasing the mediator mass). This also amounts to decreasing $\tilde{m}_{\rm sat}(\epsilon_\phi)$ accordingly, hence moving the halo phase-space distribution from the saturation regime domination ($m < \tilde{m}_{\rm sat}$) to the Coulomb regime domination

($m > \tilde{m}_{\rm sat}$). We have the following behaviors for the *effective* Sommerfeld enhancement as a function of (sub)halo mass m:

- $m > \tilde{m}_{\rm sat}(\epsilon_\phi) \leftrightarrow \epsilon_\phi^{\rm sat} > \epsilon_\phi$: The bulk of the phase-space volume is located in the Coulomb regime of the Sommerfeld factor, since $\tilde{v}_{\rm sat}(\epsilon_\phi) < \overline{v}$, so the enhancement is $\propto (m/\tilde{m}_{\rm max})^{-\nu} \propto (\overline{v}/\tilde{v}_{\rm max})^{-1}$. Hence, the enhancement factor does not depend on ϵ_ϕ, and it is fixed at a value $\propto \overline{v}^{-1}(m)$ even for decreasing (while non-resonant) ϵ_ϕ. This situation is typically encountered when ϵ_ϕ is very small, or when the halo is very massive.

- $m < \tilde{m}_{\rm sat}(\epsilon_\phi \neq \epsilon_\phi^{{\rm res},n}) \leftrightarrow \epsilon_\phi^{\rm sat} < \epsilon_\phi \neq \epsilon_\phi^{{\rm res},n}$: In this configuration, most of the halo phase space lies in the non-resonant saturation regime ($\overline{v} < \tilde{v}_{\rm sat}(\epsilon_\phi)$), and the Sommerfeld factor scales at $\propto (\tilde{v}_{\rm max}/\tilde{v}_{\rm sat}(\epsilon_\phi))^{-1}(v/\tilde{v}_{\rm sat}(\epsilon_\phi))^p \propto (\tilde{m}_{\rm max}/\tilde{m}_{\rm sat}(\epsilon_\phi))^{-\nu}(m/\tilde{m}_{\rm sat}(\epsilon_\phi))^{p\nu}\epsilon_\phi^{-(p+1)}$, independent of mass only in the s-wave case. The overall Sommerfeld factor is therefore entirely set by ϵ_ϕ (and is $\propto 1/\epsilon_\phi$ or $1/\epsilon_\phi^3$ for s- or p-wave annihilation).

- $m < \tilde{m}_{\rm sat}(\epsilon_\phi \sim \epsilon_\phi^{{\rm res},n}) \leftrightarrow \epsilon_\phi^{\rm sat} < \epsilon_\phi \sim \epsilon_\phi^{{\rm res},n}$: Here, we sit on the $n^{\rm th}$ resonance, and since $\overline{v}(m) < \tilde{v}_{\rm sat}^{{\rm res},n}$, the bulk of the phase-space volume participates in the enhancement, whose amplitude is maximized when $\tilde{v}_{\rm unit} \lesssim \overline{v} \ll \tilde{v}_{\rm sat}^{{\rm res},n} \leftrightarrow \tilde{m}_{\rm unit} \lesssim m \ll \tilde{m}_{\rm sat}^{{\rm res},n}$. The amplitude of the resonance peak *relative to* the baseline enhancement is larger for smaller halos, as predicted from Eq. (4.48). It is also suppressed like ϵ_ϕ^2 at higher and higher resonances, which, combined with the $1/\epsilon_\phi$ scaling of the baseline, explains why the amplitude of the series of peaks globally decreases linearly with ϵ_ϕ as ϵ_ϕ decreases.

- $\tilde{m}_{\rm sat}(\epsilon_\phi \sim \epsilon_\phi^{{\rm res},n}) < m \leftrightarrow \epsilon_\phi \sim \epsilon_\phi^{{\rm res},n} < \epsilon_\phi^{\rm sat}$: Here, we also sit on the $n^{\rm th}$ resonance, but only the lower tail of the phase-space volume participates in the enhancement because $\overline{v} > \tilde{v}_{\rm sat}(\epsilon_\phi^{{\rm res},n})$. The remaining (higher) part of the phase-space volume is in the Coulomb regime. The amplitude of the resonance is therefore controlled by the reduced volume of available relevant phase space, and then suppressed if $\overline{v} \gg \tilde{v}_{\rm sat}^{{\rm res},n}$. In that case, only the Coulomb enhancement is active, and actually saturates at the characteristic velocity of the (sub)halo $\propto (\overline{v}/\tilde{v}_{\rm max})^{-1} \leftrightarrow (m/\tilde{m}_{\rm max})^{-\nu}$.

All this is illustrated in Fig. 4, where the top panels show the Sommerfeld-enhanced $J_{\mathcal{S}}$-factors for halos of different masses located at the same distance from the observer (normalized to the Sommerfeld-free J-factor of a reference halo of 10^{15}M). They are plotted as a function of the Bohr-to-interaction length ratio ϵ_ϕ. These enhanced $J_{\mathcal{S}}$-factors are calculated fully numerically assuming an s-wave (p-wave) annihilation in the top left (right) panel, with, from the top to bottom curves, predictions for halos of masses from 10^{15} M (typical of galaxy clusters) down to 10^{-6} M (typical of the cutoff mass in the matter power spectrum for WIMPs). This corresponds to characteristic speeds $\overline{v}$ spanning a range from $\sim 10^{-3}$ down to $\sim 10^{-9}$ (in natural units), hence of $\overline{\epsilon}_v = \overline{v}/\alpha_X$ of $\sim 10^{-1}$, down to $\sim 10^{-7}$.

As a practical toolkit to better understand these results, we also trace in the bottom left panel the key relation between the saturation velocity $\tilde{v}_{\rm sat}$ and ϵ_ϕ (solid black curve), with the Sommerfeld enhancement factor represented in a third dimension as a function of both the velocity v and ϵ_ϕ (gray color contrast code). In the same panel, we report the characteristic DM velocity $\overline{v}(m)$ for halos of masses 10^{15}, 10^{12}, and 10^6 M , as inferred from Eq. (4.38). To each of them, we associate along the right vertical axis the full (unnormalized) DM velocity distribution calculated from the Eddington inversion at the scale radii of these halos. This plot is particularly helpful to illustrate how the phase-space distribution of DM

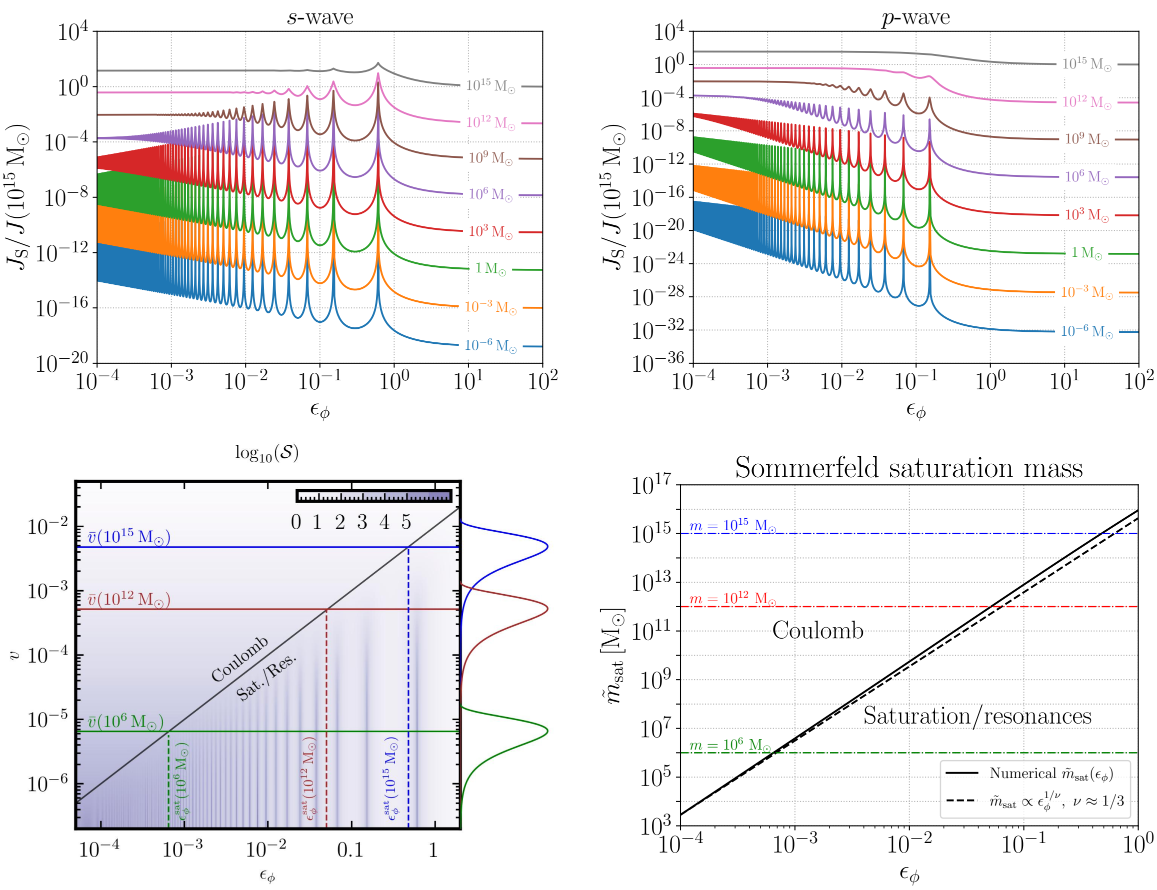

Figure 4. Sommerfeld-enhanced $J_{\mathcal{S}}$-factors as a function of ϵ_ϕ for DM halos of different masses located at the same distance, normalized to the Sommerfeld-free J-factor of a 10^{15}M halo. **Top left**: The s-wave annihilation case. **Top right**: Same as left panel, but for a p-wave annihilation. **Bottom left**: Saturation velocity as a function of ϵ_ϕ (solid black curve), delineating the transition between the Coulomb and saturation regimes. The ($\log_{10}$ of the) Sommerfeld factor is represented as the third dimension (gray color scale), as a function of velocity v and ϵ_ϕ. The characteristic speeds of 10^{15}, 10^{12}, and 10^6 M halos are indicated, with the corresponding (unnormalized) full Eddington velocity distribution taken at the scale radius, along the right vertical axis for illustration. **Bottom right**: Saturation mass of the Sommerfeld effect as a function of ϵ_ϕ.

concentrates around $\overline{v}(m)$, though with a lower and a higher tail. The bottom right panel of Fig. 4 further shows the scaling of the saturation mass $\tilde{m}_{\rm sat}$ with ϵ_ϕ, namely $\tilde{m}_{\rm sat} \propto \epsilon_\phi^{1/\nu} \approx \epsilon_\phi^3$, which follows from Eq. (4.44) and from Eq. (4.41). This approximation matches pretty well with the exact numerical result.

Let us now describe the top left panel of Fig. 4, which shows the Sommerfeld-enhanced $J_{\mathcal{S}}$-factors in the s-wave case for different halo masses. Large values of ϵ_ϕ imply that most halos have the bulk of their phase-space volume in the saturation regime, as long as m is smaller than $\tilde{m}_{\rm sat}(\epsilon_\phi)$ (equivalently $\overline{v}(m) < \tilde{v}_{\rm sat}(\epsilon_\phi)$). The transition to the Coulomb regime occurs as ϵ_ϕ decreases below a specific value, $\epsilon_\phi = \epsilon_\phi^{\rm sat}(m)$, defined in Eq. (4.54) (equivalently $\tilde{v}_{\rm sat}(\epsilon_\phi^{\rm sat}) = \overline{v}(m)$). We can read values of $\epsilon_\phi^{\rm sat}(m)$ off the plot in the bottom right panel of Fig. 4, and can also use the bottom left panel to translate the halo masses in terms of characteristic velocities and velocity distributions. For illustration, let us focus on three specific virial halo masses, 10^{15}, 10^{12}, and 10^{6} M . All these masses have characteristic velocities smaller than $\sim \alpha_\chi = 0.01$, and are therefore subject to Sommerfeld enhancement ($m < \tilde{m}_{\rm max}$). Let us follow their $J_{\mathcal{S}}$-curves on the top left panel from the right to left (large to small ϵ_ϕ or $\tilde{m}_{\rm sat}$), while keeping in mind Eq. (4.51), which describes the Coulomb regime, Eq. (4.52) the saturation regime, and Eq. (4.53) resonances:

- $m = 10^{15}$ M $\leftrightarrow \overline{v} \sim 6 \times 10^{-3}$: The transition from saturation to the Coulomb regime occurs at $\epsilon_\phi^{\rm sat} \sim 0.5$, a value just below the first resonance. For $\epsilon_\phi > \epsilon_\phi^{\rm sat}$, we are in the saturation regime ($m < \tilde{m}_{\rm sat}(\epsilon_\phi)$): the Sommerfeld factor saturates at $\propto (\tilde{m}_{\rm sat}/\tilde{m}_{\rm max})^{-\nu} \propto (\tilde{v}_{\rm sat}/\tilde{v}_{\rm max})^{-1} \propto 1/\epsilon_\phi$, but takes a small value because $\tilde{m}_{\rm sat} \lesssim \tilde{m}_{\rm max}$. When ϵ_ϕ hits the first resonance, $\epsilon_\phi^{\rm res,1} \simeq 2/3$, only a tiny part of the phase-space volume can participate in the enhancement $\propto 1/v^2$, because the bulk of the velocity distribution lies around $\overline{v} \sim \tilde{v}_{\rm sat}^{\rm res,1}$. Consequently, the amplitude of the first resonance is suppressed. The transition from the saturation regime to the Coulomb regime occurs when $\epsilon_\phi < \epsilon_\phi^{\rm sat}$, below which the bulk of the Sommerfeld boost becomes $\propto 1/\overline{v}(m)$. Once the whole phase-space distribution finds itself in the Coulomb regime, the Sommerfeld boost factor remains fixed at a constant value $\propto 1/\overline{v}$ determined by the characteristic velocity of the halo. For the same reason, higher-order resonances are suppressed (no phase-space volume left below $\tilde{v}_{\rm sat}(\epsilon_\phi)$). Therefore, for further decreasing values of ϵ_ϕ, even though in the Coulomb regime, the Sommerfeld enhancement factor remains constant, fixed by the characteristic velocity $\propto 1/\overline{v}$: the J-factor stops evolving accordingly and remains flat.

- $m = 10^{12}$ M $\leftrightarrow \overline{v} \sim 6 \times 10^{-4}$: The transition from the saturation to the Coulomb regime occurs at $\epsilon_\phi^{\rm sat} \sim 0.05$, which is located between the fourth ($\epsilon_\phi^{\rm res,4} \simeq 1/24$) and third ($\epsilon_\phi^{\rm res,3} \simeq 2/27$) resonances. For $\epsilon_\phi > \epsilon_\phi^{\rm sat}$, we are in the regime $\tilde{m}_{\rm sat} > m$, hence in the saturation regime for which the enhancement is $\propto 1/\epsilon_\phi$. When ϵ_ϕ hits resonant values of order $n < 3$, a significant part of the phase-space volume can participate in the $(\tilde{v}_{\rm sat}/v)^2$ enhancement, which cannot exceed $\sim (\tilde{v}_{\rm sat}/\overline{v})^2$ because most of the phase-space volume concentrates around $\overline{v}$. Therefore, even though the first resonances are turned on, their amplitudes are phase-space limited. When ϵ_ϕ further decreases below $\epsilon_\phi^{\rm sat}$, the bulk of the phase-space volume switches to the Coulomb regime, but with a Sommerfeld factor asymptoting to a constant $\propto \tilde{v}_{\rm max}/\overline{v}$. There is no longer enough phase-space volume available below $\tilde{v}_{\rm sat}(\epsilon_\phi)$ to trigger higher-order resonances, and the $J_{\mathcal{S}}$-factor stops evolving and remains flat.

- $m = 10^{6}$ M $\leftrightarrow \overline{v} \sim 6 \times 10^{-6}$: The saturation-Coulomb transition occurs at $\epsilon_\phi^{\rm sat} \sim 5 \times 10^{-4}$, which is located in the resonance forest. Like in the previous case, as long as $\epsilon_\phi > \epsilon_\phi^{\rm sat}$ (equivalently $\tilde{m}_{\rm sat} > m$), we are in the saturation regime, and the Sommerfeld factor is $\propto 1/\tilde{v}_{\rm sat} \propto 1/\epsilon_\phi$. All resonances encountered by decreasing ϵ_ϕ down to $\epsilon_\phi^{\rm sat}$ are turned on and have their amplitudes roughly set by $1/(n\,\overline{v})^2 \propto \epsilon_\phi/\overline{v}^2$ —the amplitudes decrease linearly with ϵ_ϕ as the latter decreases. When ϵ_ϕ becomes smaller than $\epsilon_\phi^{\rm sat}$, we switch to the Coulomb regime, and the enhancement is frozen to $\propto 1/\overline{v}$, and there is not enough phase-space volume available in the lower tail to turn the remaining resonances on. Hence, the Sommerfeld enhancement remains frozen and no longer evolves as ϵ_ϕ keeps on decreasing below $\epsilon_\phi^{\rm sat}$.

The ratio of J-factors for two halos of masses m_1 and m_2 can easily be estimated from Eq. (4.29) and Eq. (4.50). In the absence of Sommerfeld enhancement, i.e. $\epsilon_\phi \gtrsim 1$, then $J_1/J_2 \sim (m_1/m_2)^{1-3\,\varepsilon+p\,\nu}$. From this rough scaling relation, we can predict a factor of $\sim 3.5\times 10^2$ between each successive curve in the top left panel of Fig. 4 for the s-wave case, and $\sim 3.5\times 10^4$ for the p-wave case in the top right panel. This is reasonably close to the exact results, $\sim 5\times 10^2$ and 4×10^4, respectively. When the Sommerfeld enhancement kicks in, this ratio is corrected by an additional factor $\sim (m_1^{-s_{m_1}-\nu\,p}/m_2^{-s_{m_2}-\nu\,p})$, where s_{m_1} and s_{m_2} refer to the Sommerfeld mass indices of the halo of mass m_1 and the halo of mass m_2, respectively [see Eq. (4.47)]—indeed, the two halos can be in different Sommerfeld regimes. We can still verify from the top panels of Fig. 4 that, for instance, when halos are in the Coulomb regime (asymptotic values on the very left parts of the panels), then successive curves should be asymptotically split by a factor of ~ 35 for both the s- and p-wave cases, according to Eq. (4.29). This is again close to the accurate numerical evaluation.

For the p-wave case illustrated in the top right panel of Fig. 4, the main differences with the s-wave case are the following. (i) In the saturation regime, the baseline enhancement is $\propto 1/\epsilon_\phi^3$ (instead of $1/\epsilon_\phi$). (ii) The amplitude of the resonance peak scales like $n^2 \propto 1/\epsilon_\phi$ (instead of $1/n^2 \propto \epsilon_\phi$), and therefore increases with the order of the resonance (linearly with $1/\epsilon_\phi$, as ϵ_ϕ decreases). (iii) The overall $\propto \overline{v}^2$ suppression factor in the cross section (effectively captured in our ansatz for the Sommerfeld boost factor above), is compensated for by the Sommerfeld enhancement, except on the baseline of the saturation regime where it contributes an additional splitting factor $\propto m^{2\nu}$ (very right part of the plot), which then disappears in the Coulomb regime (very left part of the plot). A full description of resonance properties can be found around Eq. (3.13).

Finally, before moving to the study of the global contribution of a subhalo population to the Sommerfeld enhancement, it is interesting to compare the accurate numerical results at the level of single halos with those derived from our approximate ansatz of Eq. (4.46). We report such a comparison in Fig. 5 in terms of both the Sommerfeld-enhanced $J_{\mathcal{S}}$ factors (top panels) and the ratios $J_{\mathcal{S}}/J$ (bottom panels) for halos of masses 10^8, 10^{12}, and 10^{15} M , typical of dwarf galaxies, spiral galaxies, and galaxy clusters, placed at distances 0.1, 1, 100 Mpc, respectively. An averaged Sommerfeld enhancement factor at the level of an entire halo can be formulated from the ratio $J_{\mathcal{S}}/J$ — J and $J_{\mathcal{S}}$ are calculated from a full phase-space integration in the accurate numerical results shown in the plots. We actually compare these accurate results (dotted curves) with our analytical approximations (dashed and dot-dashed curves) — see Eq. (4.27) for the J-factor, and Eq. (4.49) for the Sommerfeld-enhanced $J_{\mathcal{S}}$-factor. The left (right, respectively) panels show the comparisons for an s-(p-)wave annihilation. We have also reported solid curves that correspond to the exact Sommerfeld factors of Eq. (3.3) and Eq. (3.5) evaluated at a single characteristic velocity $\overline{v}(m)$ for each halo. We used Eq. (4.40) for the latter, and tuned the constant ω_0 to 0.6 (0.8, respectively) for s-(p-)wave annihilation. The baselines and the peaks envelopes (dashed and dot-dashed curves) are instead calculated from our analytical ansatz of Eq. (4.46), evaluated at the same characteristic velocities. We see that the analytical approximations capture the exact behaviors reasonably well. The peaks amplitudes are slightly underestimated because $\overline{v}(m)$ overestimates the typical speed at the very center of objects. As expected from our analytical approximations, the Sommerfeld enhancement at the level of a full halo is quite similar to the local velocity-dependent effective Sommerfeld enhancement depicted in Fig. 1 (except for the v^2 p-wave correction absorbed in the definition of $\mathcal{S}$ in the latter case, which does not change the scaling in ϵ_ϕ but rescales the Sommerfeld enhancement by a factor of

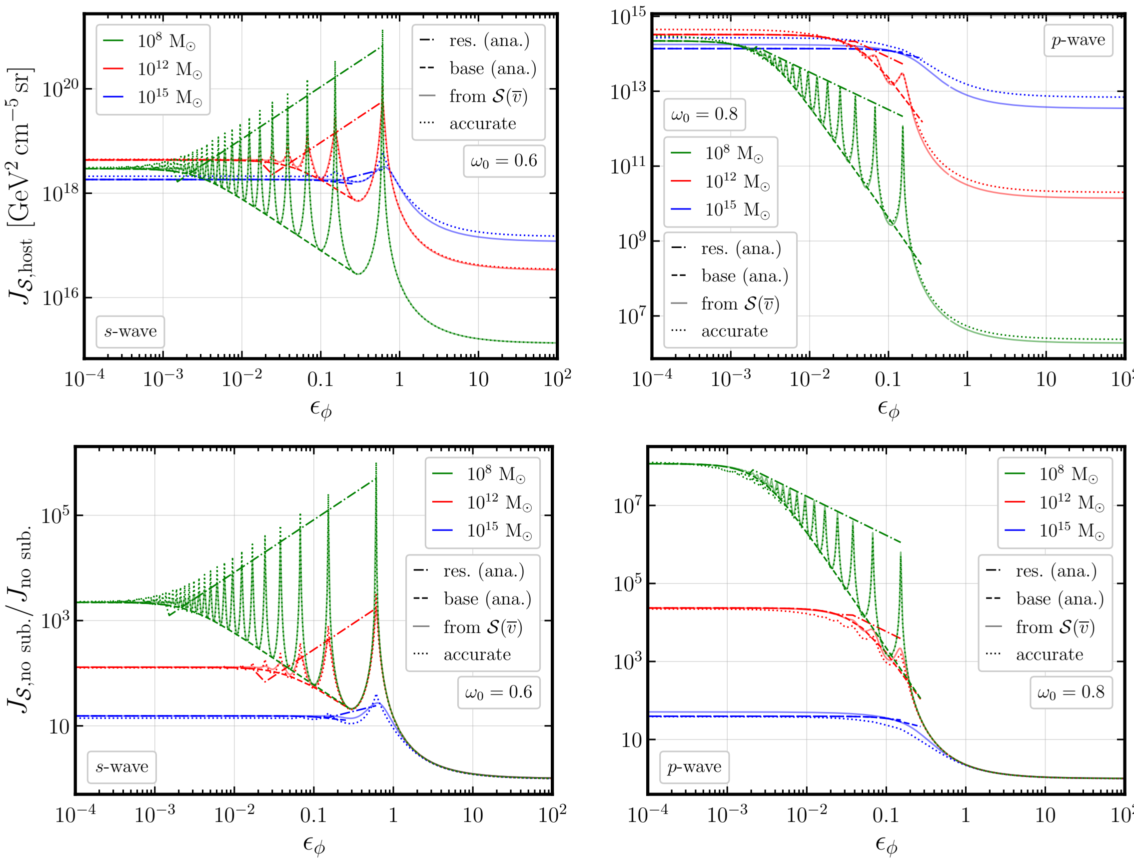

Figure 5. **Top left panel:** Sommerfeld-enhanced $J_{\mathcal{S}}$ factors over entire halos obtained for an s-wave annihilation process calculated (i) from the full phase-space numerical integral of exact expressions (dotted curve), (ii) from the analytical approximation of the J-factor in Eq. (4.27) times the Sommerfeld factor of Eq. (3.3) and Eq. (3.5) evaluated at speeds $\overline{v}(m)$, and (iii) from the full analytical approximation of Eq. (4.49) (for the Coulomb and saturation baseline, and for the peaks envelope) — we consider three halos of masses $m = 10^8$, 10^{12}, and 10^{15} M . **Bottom left panel:** Corresponding effective Sommerfeld enhancement factors over entire halos ratio expressed as $J_{\mathcal{S}}/J$, neglecting subhalos. **Right panels:** Same as left panels for a p-wave annihilation process.

$\mu^{2\nu}$: this benefits more massive halos but at the same time is more representative of the scaling of the true cross section).

4.3.3 Sommerfeld enhancement for a population of subhalos

To understand the global Sommerfeld enhancement arising from a population of subhalos, it is convenient to combine the results obtained in the previous paragraph, where we have defined an ansatz for the Sommerfeld enhancement in terms of the (sub)halo virial mass m, with the analytical results obtained for the subhalo boost factor in Sec. 4.3.1. We warn the reader that the analytical results derived from now on will be much less precise when compared to the numerical results (generically much more precise for s-wave than for p-wave processes). Still, they turn very useful to really understand the different features of the numerical results.

Given a (sub)halo of virial mass m, we can define a $J_{\mathcal{S}}$-factor corrected for the overall

Sommerfeld effect according to Eq. (4.49), which scales like $\propto \mu^{1-3\varepsilon-s_m}$. Exponent s_m is the effective power-law mass index introduced in Eq. (4.46), which takes different values for the different Sommerfeld regimes. From this, assuming $\tilde{m}_{\rm unit} < m_{\rm min} < \tilde{m}_{\rm sat}(\epsilon_\phi) < m_{\rm max} < \tilde{m}_{\rm max}$ (which is not always the case[6]), it is easy to express the total J-factor for a population of subhalos:

$$
\begin{aligned}
J_{\mathcal{S},\rm sub}(\epsilon_\phi) &= \int_{m_{\rm min}}^{m_{\rm max}} {\rm d}m\, \frac{{\rm d}N_{\rm sub}}{{\rm d}m}\, J_{\mathcal{S}}(m,\epsilon_\phi) \\
&= \underbrace{\int_{m_{\rm min}}^{\tilde{m}_{\rm sat}(\epsilon_\phi)} {\rm d}m\, \frac{{\rm d}N_{\rm sub}}{{\rm d}m}\, J_{\mathcal{S}}(m,\epsilon_\phi)}_{\text{saturation+resonances}} + \underbrace{\int_{\tilde{m}_{\rm sat}(\epsilon_\phi)}^{m_{\rm max}} {\rm d}m\, \frac{{\rm d}N_{\rm sub}}{{\rm d}m}\, J_{\mathcal{S}}(m,\epsilon_\phi)}_{\text{Coulomb}} \\
&\,\tilde{\propto}\, \left(\sum_{s\in\text{Somm. regimes}} \left(\frac{1}{\alpha_s}\, \mu^{-\alpha_s} \Big|^{\text{lower relevant mass bound}}_{\text{upper relevant mass bound}} \right) \right).
\end{aligned}
\tag{4.55}
$$

Our sign convention for the Sommerfeld-enhanced subhalo boost factor mass index α_s is such that a positive value gives large values of the total subhalo J-factor, hence of the boost factor. This occurs when lighter subhalos contribute the most to the annihilation rate, hence when the integral above is dominated by contributions at the lower mass boundary [see discussion around Eq. (4.33)]. The important features of this total luminosity are therefore (i) the mass boundaries of the integral, and (ii) the effective subhalo mass index α_s (and its sign), which depends on the mass and mass-concentration indices α and ε, as well as on the Sommerfeld-enhancement mass index s_m introduced earlier for different regimes. This effective index can easily be derived by integrating Eqs. (4.51)-(4.53) over the subhalo mass function. It generically reads:

$$\alpha_s = \alpha - 2 + 3\varepsilon + s_m\,, \tag{4.56}$$

where s_m is the Sommerfeld mass index that depends on the Sommerfeld regime—it is given in Eq. (4.47).

In fact, both α_s and the mass boundaries depend on the relevant Sommerfeld regime, which is itself fixed by ϵ_ϕ or, equivalently, by $\tilde{m}_{\rm sat}(\epsilon_\phi)$ [hence the splitting of the integral as a sum of different pieces in Eq. (4.55)]. Consequently, for a given ϵ_ϕ, there can be two different contributions, assuming $m_{\rm min} < \tilde{m}_{\rm sat}(\epsilon_\phi) < m_{\rm max} < \tilde{m}_{\rm max}$: one from the Coulomb regime, involving subhalo masses between $\tilde{m}_{\rm sat}(\epsilon_\phi)$ and $m_{\rm max}$, and another one from the saturation regime, involving subhalo masses lighter than $\tilde{m}_{\rm sat}(\epsilon_\phi)$. For resonant values of ϵ_ϕ (i.e. still in the saturation regime), subhalos lighter than $\tilde{m}_{\rm sat}(\epsilon_\phi^{{\rm res},n})$, for any order n, are also the ones most involved in the enhancement. Since the minimal subhalo mass $m_{\rm min}$ is fixed for all host halos (it depends on the DM particle scenario itself), and the maximal subhalo mass $m_{\rm max} \simeq M_{\rm host}/100$ is always smaller than $\tilde{m}_{\rm max}$ in the configurations studied in this paper, we can advantageously split the subhalo population yield to the J-factor by taking the asymptotic form of the Sommerfeld enhancement factor relevant to each part of the integral of Eq. (4.55). This provides us with a one of our main fully analytical results:

[6] Among the possible departures from this assumed mass hierarchy, we indicate three generic variants: (i) $m_{\rm min} < \tilde{m}_{\rm unit}$, which is actually the case for the template parameters used in this paper; (ii) $\tilde{m}_{\rm sat} < m_{\rm min}$, which is generic in the Coulomb massless-mediator limit $\epsilon_\phi \to 0$, in which case there is no saturation regime; (iii) $m_{\rm max} < \tilde{m}_{\rm sat}$, which can happen for small host halos (typically dwarf galaxies) and moderate values of ϵ_ϕ. In all of those cases, one needs to recast the splitting of the mass integral accordingly, which leads to different mass boundaries in the asymptotic regimes.

$$
\begin{aligned}
J_{\mathcal{S},\mathrm{sub}}(\epsilon_\phi) = \; & \theta(m_{\max} - \tilde{m}_{\mathrm{sat}}(\epsilon_\phi))\, J^{\mathrm{Coul}}_{\mathcal{S},\mathrm{sub}}(\epsilon_\phi) \\
& + \theta(\tilde{m}_{\mathrm{sat}}(\epsilon_\phi) - m_{\min})\, J^{\mathrm{sat}}_{\mathcal{S},\mathrm{sub}}(\epsilon_\phi) \\
& + \sum_{n=1+\frac{p}{2}} \left(\delta_{\epsilon_\phi/\{\epsilon_\phi^{\mathrm{res},n}\}}\, \theta(\tilde{m}_{\mathrm{sat}}(\epsilon_\phi) - m_{\min})\, J^{\mathrm{res}}_{\mathcal{S},\mathrm{sub}}(\epsilon_\phi) \right).
\end{aligned}
\tag{4.57}
$$

We have deliberately separated the different Sommerfeld regimes for clarity. We have respectively for the Coulomb, saturation, and resonance regimes:

$$
J^{\mathrm{Coul}}_{\mathcal{S},\mathrm{sub}}(\epsilon_\phi) = \frac{N_0(\mu_{\mathrm{host}})\, J_0}{\alpha_{\mathrm{Coul}}} (2 v_0)^p\, S_0\, \tilde{\mu}_{\max}^{\nu(1+p)} \left[\mu^{-\alpha_{\mathrm{Coul}}} \right]_{\mu_{\max}}^{\max(\tilde{\mu}_{\mathrm{sat}},\mu_{\min})} \tag{4.58a}
$$

$$
J^{\mathrm{sat}}_{\mathcal{S},\mathrm{sub}}(\epsilon_\phi) = \frac{N_0(\mu_{\mathrm{host}})\, J_0}{\alpha_{\mathrm{sat}}} (2 v_0)^p\, S_0\, S_1 \left\{ \frac{\tilde{\mu}_{\max}}{\tilde{\mu}_{\mathrm{sat}}} \right\}^{\nu(1+p)} \left[\mu^{-\alpha_{\mathrm{sat}}} \right]_{\min(\tilde{\mu}_{\mathrm{sat}},\mu_{\max})}^{\mu_{\min}} \tag{4.58b}
$$

$$
\begin{aligned}
J^{\mathrm{res}}_{\mathcal{S},\mathrm{sub}}(\epsilon_\phi) = \; & \frac{N_0(\mu_{\mathrm{host}})\, J_0}{\alpha_{\mathrm{res}}} (2 v_0)^p\, S_0^{\mathrm{res}} \left\{ \frac{\tilde{\mu}_{\max}}{\tilde{\mu}_{\mathrm{sat}}} \right\}^{\nu(1+p)} \tilde{\mu}_{\mathrm{sat}}^{2\nu} \\
& \times \left\{ \left[\mu^{-\alpha_{\mathrm{res}}} \right]_{\min(\tilde{\mu}_{\mathrm{sat}},\mu_{\max})}^{\max(\mu_{\min},\tilde{\mu}_{\mathrm{unit}})} + \frac{\theta(\tilde{\mu}_{\mathrm{unit}} - \mu_{\min})}{\tilde{\mu}_{\mathrm{unit}}^{2\nu}} \frac{\alpha_{\mathrm{res}}}{\alpha_{\mathrm{res}}^{\mathrm{unit}}} \left[\mu^{-\alpha_{\mathrm{res}}^{\mathrm{unit}}} \right]_{\tilde{\mu}_{\mathrm{unit}}}^{\mu_{\min}} \right\},
\end{aligned}
\tag{4.58c}
$$

where $\mu_{\min}$ and $\mu_{\max}$ are the subhalo reduced minimal and maximal masses, given a host halo of mass M_{host} and a DM particle scenario. The boost mass index Eq. (4.56) allows us to determine the different indices, α_{Coul}, α_{sat}, and α_{res}, from the Sommerfeld mass index s_m of Eq. (4.47) given for the three Sommerfeld regimes. This separation is rather artificial though, because the Sommerfeld enhancement factor smoothly transits between these regimes. That, together with the fact that we approximate phase-space integrals by evaluating the relevant functions at characteristic velocities, which induces nonphysical thresholds between the saturation/resonant and Coulomb regimes, will be the main source of numerical errors with respect (i) to the full mass integral of the Sommerfeld factor, and (ii) a fortiori also to the exact numerical integration over both mass and phase space. However, this division has the virtue of providing fully analytical scaling relations and a fine understanding of parameter dependencies, despite the significant cost in precision.

As an additional detail, mind the last term of the result obtained for resonances, which features $\tilde{\mu}_{\mathrm{unit}}$ and includes the possibility of having $\mu_{\min} < \tilde{\mu}_{\mathrm{unit}}$, in which case we have to account for the unitarity saturation of resonances. In this small corner of the parameter space, the Sommerfeld-corrected index α_{res} changes, which we write explicitly by using $\alpha_{\mathrm{res}}^{\mathrm{unit}}$ (this term is not crucial, so we will mostly neglect it in forthcoming discussion). As the frames indicate, these are still very insightful results which allow us to fully understand how the Sommerfeld enhancement propagates over a full population of subhalos.

Assuming a value for ϵ_ϕ such that $m_{\min} < \tilde{m}_{\mathrm{sat}}(\epsilon_\phi) < m_{\max}$, the upper subhalo mass range $\in [\tilde{m}_{\mathrm{sat}}(\epsilon_\phi), m_{\max}]$ lies in the Coulomb regime, while the lower one $\in [m_{\min}, \tilde{m}_{\mathrm{sat}}(\epsilon_\phi)]$ lies in the saturation regime, for which the asymptotic mass slopes associated with $J_{\mathcal{S},\mathrm{sub}}$ take different values—resonances further show up in the saturation regime. Each regime is featured by its own index α_s, whose generic form is given in Eq. (4.56). The dominant boundary term of each piece in Eq. (4.58) will be selected according to the sign of each index:

as explained above, a positive sign implies a dominant contribution from lighter and therefore more numerous and more concentrated subhalos. Let us inspect these indices in more detail, by combining Eq. (4.47) and Eq. (4.56):

$$\alpha_s = \begin{cases} \alpha_{\rm Coul} & \equiv \alpha - 2 + 3\varepsilon + \nu \\ \alpha_{\rm sat} & \equiv \alpha - 2 + 3\varepsilon - \nu\, p \\ \alpha_{\rm res} & \equiv \alpha - 2 + 3\varepsilon + \nu\,(2 - p) \\ \alpha_{\rm res}^{\rm unit} & \equiv \alpha - 2 + 3\varepsilon - \nu\, p = \alpha_{\rm sat} \end{cases} \xrightarrow[\text{evaluation}]{\text{numerical}} \begin{cases} \alpha_{\rm Coul} \approx 0.43 \\ \alpha_{\rm sat} \approx \begin{cases} 0.10 \text{ (for } p = 0) \\ -0.57 \text{ (for } p = 2) \end{cases} \\ \alpha_{\rm res} \approx \begin{cases} 0.76 \text{ (for } p = 0) \\ 0.10 \text{ (for } p = 2) \end{cases} \end{cases} \tag{4.59}$$

From Eq. (4.41), we have the characteristic-speed-to-mass index $\nu \simeq 1/3$. Parameter $p = 0/2$ for an s/p-wave annihilation. Since $\alpha \approx 1.95$ and $\varepsilon \approx 0.05$, we see that $\alpha_{\rm Coul} > 0$ quite generically. Therefore, the Coulomb regime is $\tilde{m}_{\rm sat}$-dominated (provided $\tilde{m}_{\rm sat} < m_{\rm max}$, which is not always the case in particular if the host halo is a dwarf galaxy). On the other hand, in the saturation regime, $\alpha_{\rm sat} > 0$ for the s-wave case ($p = 0$), while it gets negative for the p-wave case ($p = 2$). Therefore, the saturation regime is either $m_{\rm min}$-dominated (s-wave) or $\tilde{m}_{\rm sat}$-dominated (p-wave). In the latter case, this means that mostly subhalo masses down to $\tilde{m}_{\rm sat}(\epsilon_\phi)$ participate in an extra-Sommerfeld enhancement, whereas the lower part of the mass function does not add up a significant yield—this actually comes from the v^2 p-wave suppression factor absorbed in our effective Sommerfeld ansatz, which remains in the saturation regime of p-wave annihilation. In contrast, on resonances, we see that $\alpha_{\rm res}$ is positive for both s-wave and p-wave annihilation processes. Therefore, all subhalos down to the cutoff mass $m_{\rm min}$ participate in the extra-enhancement on resonances in both cases. There is still a fundamental different between s- and p-wave resonances that needs to be emphasized: there is formally a velocity dependence in the s-wave case, which can be seen from the 2ν contribution to $\alpha_{\rm res}$, while p-wave resonance amplitudes do not depend on velocity—see detailed discussion around Eq. (3.13). Finally, it is important to stress that values of α_s close to 0 are subject to uncertainties. A small change in the mass function slope α, for instance, could change the hierarchy in the contributing masses, hence in the global enhancement. This concerns mostly the saturation regime of the s-wave annihilation and resonances of the p-wave annihilation.

The previous discussion is illustrated in Fig. 6, where we have actually calculated a dimensionless quantity proportional to the product of the integrated number of subhalos more massive than m, $N(> m) \propto m^{1-\alpha}$, with the $J_{\mathcal{S}}(m)$ factor for a single (sub)halo of mass m (divided by $J(1\,{\rm M}_\odot)$ to get a dimensionless quantity). This is meant to capture the dominant scaling of the global $J_{\mathcal{S},\rm sub}$ factor given in Eq. (4.57) with the lower mass bound m, which also gives insight on the most contributing mass range in $J_{\mathcal{S},\rm sub}$ [Eq. (4.57)]. Again, we take the three different Sommerfeld configurations used before: $\epsilon_\phi = 0.1$ (moderate enhancement), 10^{-3} (significant enhancement), and the $n = 8$ resonance ($\epsilon_\phi \sim 10^{-2}$, strong enhancement). The corresponding saturating masses $\tilde{m}_{\rm sat}(\epsilon_\phi)$ are shown as vertical dashed lines, marking the transition between subhalos mostly in the Coulomb regime ($m > \tilde{m}_{\rm sat}$) or mostly in the saturation regime ($m < \tilde{m}_{\rm sat}$). For the s-wave case (left panel), we see that the lower bound is always the most contributing one (curves increase as the mass boundary m decreases in all Sommerfeld regimes), consistently with the positive values of α_s found in Eq. (4.59). In contrast, the p-wave curves (right panel) only increase down to the saturation mass, below which contributions become negligible; except of course on the resonance, where the contribution increases as the boundary mass m decreases down to the unitary limit. This

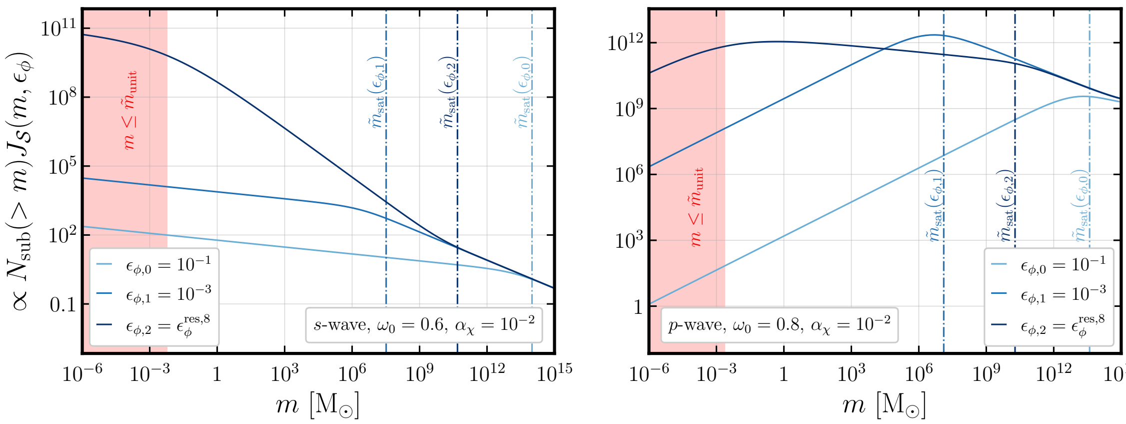

Figure 6. Estimate of the contribution of the subhalo population to the Sommerfeld-enhanced $J_{\mathcal{S},\rm sub}$ factor above some mass m, for different values of the reduced Bohr radius ϵ_ϕ: a large value 0.1 (moderate effect), a small value of 10^{-3} (significant effect), and an intermediate value of $\sim 10^{-2}$ sitting on the $n=8$ resonance (strong effect). A maximum in the curves show the subhalo mass range that contributes the most to the annihilation signal. This recasts most of the information already included in Fig. 2 (effective Sommerfeld enhancement as a function of velocity), and in Fig. 3. Transition from Coulomb to saturation regimes occurs around $\tilde{m}_{\rm sat}(\epsilon_\phi)$, reported as vertical dash-dotted lines. **Left panel:** s-wave case. **Right panel:** p-wave case.

is again consistent with the fact that $\alpha_{\rm sat}<0$ while $\alpha_{\rm res}>0$ with our choice of parameters for the p-wave case.

The peak-to-baseline ratio in the saturation regime for a full subhalo population, $\overline{\mathcal{R}}_{\rm sub}$, is given by:

$$\begin{aligned}\overline{\mathcal{R}}_{\rm sub}(\epsilon_\phi,\mu_{\rm min},\alpha) &= \frac{S_0^{\rm res}}{S_0 S_1}\,\tilde{\mu}_{\rm sat}^{2\nu+\frac{p}{2}\alpha_{\rm sat}}\,\mu_{\rm min}{}^{-\alpha_{\rm res}}\,\mu_{\rm min}^{\frac{(2-p)}{2}\alpha_{\rm sat}} \\ &= \frac{\pi^2}{6}\,\tilde{\mu}_{\rm sat}^{2\nu+\frac{p}{2}\alpha_{\rm sat}}\,\mu_{\rm min}{}^{-\alpha_{\rm res}}\,\mu_{\rm min}^{\frac{(2-p)}{2}\alpha_{\rm sat}} \\ &\propto \epsilon_\phi^{2+\frac{p}{2}\frac{\alpha_{\rm sat}}{\nu}}\,\mu_{\rm min}{}^{-\alpha_{\rm res}}\,\mu_{\rm min}^{\frac{(2-p)}{2}\alpha_{\rm sat}}\,.\end{aligned} \tag{4.60}$$

For s-wave processes, we have $\overline{\mathcal{R}}_{\rm sub}\propto\epsilon_\phi^2$, while for p-wave processes,[7] $\overline{\mathcal{R}}_{\rm sub}\propto\epsilon_\phi^{0.3}$. This result predicts that the peak-to-baseline ratio should decrease much faster as ϵ_ϕ decreases in the s-wave case than in the p-wave case. Note that the above ratio assumes $\tilde{\mu}_{\rm sat}<\mu_{\rm max}$, which is not always verified (notably for light host halos). If instead $\tilde{\mu}_{\rm sat}>\mu_{\rm max}$, then the dependence in ϵ_ϕ becomes $\propto\epsilon_\phi^2$ in both cases, and the ratio decreases fast with ϵ_ϕ (the baseline increases fast) until $\tilde{\mu}_{\rm sat}$ enters the subhalo mass range, whence the dependence becomes much weaker.

[7] The indices or parameters in blue featuring factors of p are tricks to account for the change of sign of $\alpha_{\rm sat}$ here between the s- and p-wave processes. Indeed, the sign of the index decides whether one picks only the lower or the upper bound of the integral, as generically illustrated in Eq. (4.55), so as to write a simplified approximate results in the limit $m_{\rm lower}\ll m_{\rm upper}$. It turns out that with our choice of parameters, the sign of the saturation mass index $\alpha_{\rm sat}$ changes from the s- to the p-wave case, hence the associated final results scale with different boundary masses (for instance $\propto m_{\rm lower}^{-\alpha_{\rm sat}}$ in one case, and $\propto m_{\rm upper}^{-\alpha_{\rm sat}}$). Therefore, while Eq. (4.58) is generic, Eq. (4.60) is not and is only valid for our choice of reference parameters. In the same vein, other equations with blue indices are not generic. All generic results can formally be expressed from Eq. (4.58), but would lead to rather tedious expressions. Parameter $\mu_{\rm min}$ also appears in blue whenever it could be traded for $\tilde{\mu}_{\rm unit}$, i.e., when $\mu_{\rm min}<\tilde{\mu}_{\rm unit}$.

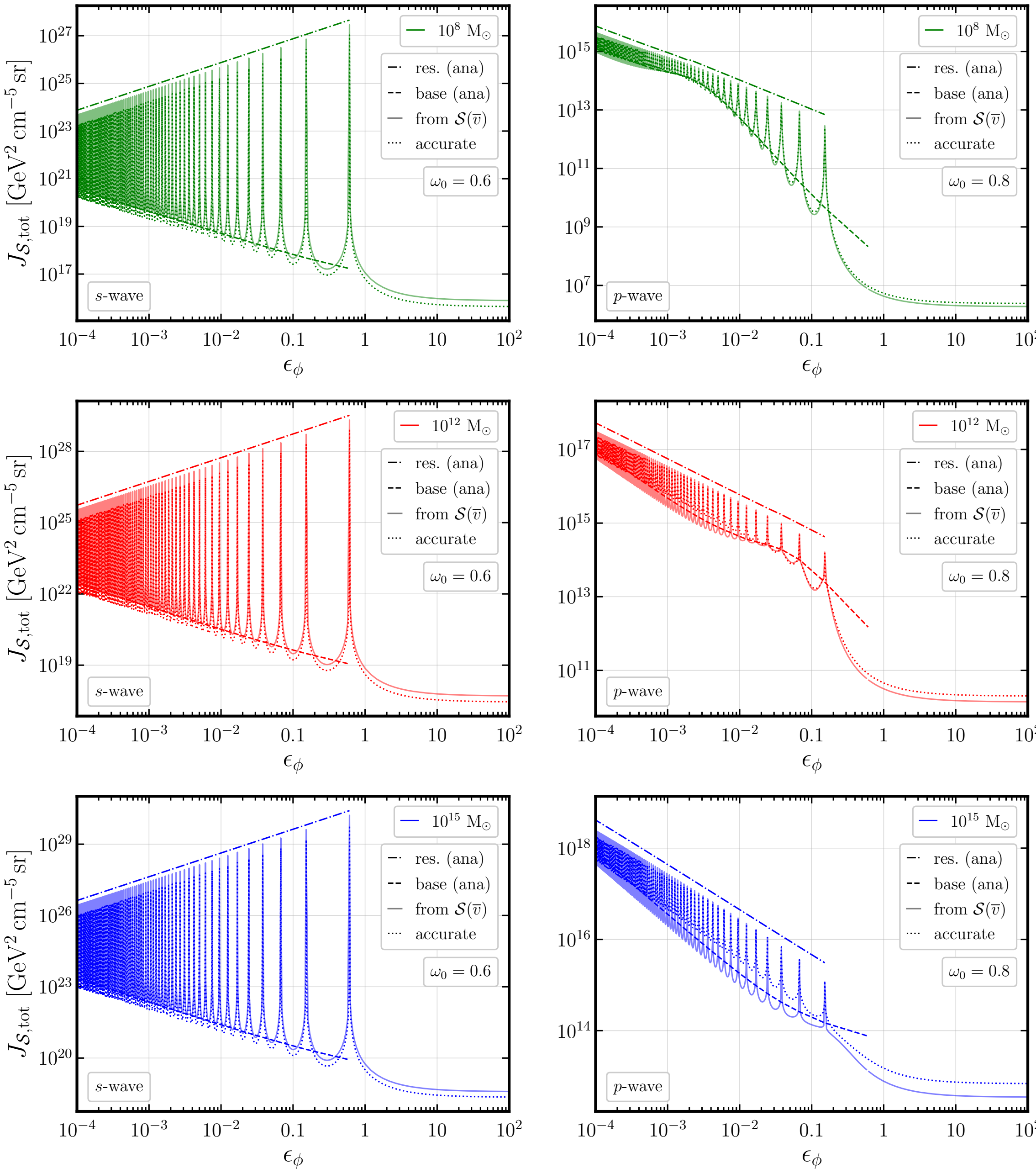

Figure 7. **Left column:** Global Sommerfeld-enhanced subhalo J-factors for s-wave annihilation, after integration of the whole subhalo population for three different host halo masses, 10^8, 10^{12}, and 10^{15} M$_\odot$, from top to bottom panels. **Right column:** Same for p-wave annihilation.

We can further determine the effective Sommerfeld enhancement at the level of a subhalo population, which helps understand how the Sommerfeld effect manifests itself on top of the subhalo boost factor. We define this global Sommerfeld enhancement as

$$\widetilde{\mathcal{S}}(M_{\rm host}, m_{\rm min}, \epsilon_\phi) \equiv \frac{J_{\mathcal{S},\rm sub}(\epsilon_\phi)}{J_{\rm sub}}\,, \tag{4.61}$$

where the Sommerfeld-enhanced subhalo contribution $J_{\mathcal{S},\rm sub}$ is given in Eqs. (4.58a)-(4.58c)

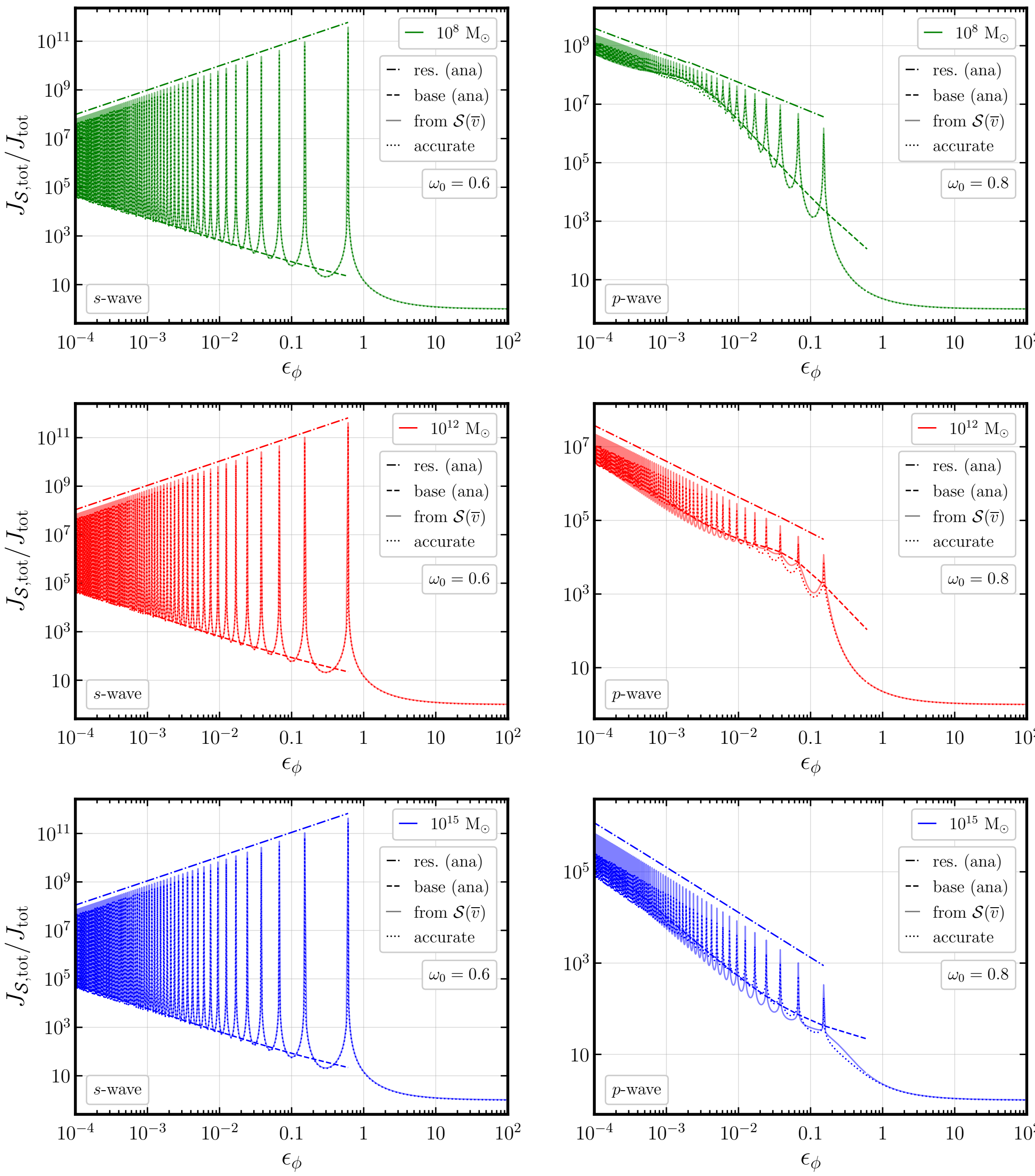

Figure 8. **Left column:** Global Sommerfeld enhancement for s-wave annihilation after integration of the whole subhalo population for three different host halo masses, 10^8, 10^{12}, and 10^{15} M , from top to bottom panels, expressed in terms of the ratio of the Sommerfeld-enhanced-to-Sommerfeld-free J-factors. **Right column:** Same for p-wave annihilation.

for the different regimes, while the Sommerfeld-free subhalo contribution J_{sub} is given in Eq. (4.30) for the s- and p-wave cases.

In Fig. 7 and Fig. 8, we compare the exact numerical calculations with the analytical approximations of respectively the total Sommerfeld-enhanced J-factors $J_{\mathcal{S},\mathrm{tot}} \equiv J_{\mathcal{S},\mathrm{host}} + J_{\mathcal{S},\mathrm{sub}} \simeq J_{\mathcal{S},\mathrm{sub}}$, and their ratios to the Sommerfeld-free cases, $J_{\mathcal{S},\mathrm{tot}}/J_{\mathrm{tot}}$, for the three template host halos introduced before. Note the resemblance with Fig. 5 (top and bottom panels, respectively), which compares calculations of the Sommerfeld enhancement at the level of a single halo—this is indicative of the fact that a few specific masses drive the overall enhancement in each of the Sommerfeld regimes. Again, the dotted curves represent the full integrated results, while the solid curves represent the subhalo mass integral performed over the exact Sommerfeld factor evaluated at subhalo mass-dependent characteristic velocities. The analytical baselines and the peaks envelopes (dot-dashed curves) are instead calculated from the ansatz of Eq. (4.46), integrated over the subhalo mass function. Panels in the left (respectively right) column display our results for the s-wave (p-wave) case. While the results for the J-factors can be understood from Eq. (4.58), we can better interpret the ratio $J_{\mathcal{S},\rm tot}/J_{\rm tot} \approx J_{\mathcal{S},\rm sub}/J_{\rm sub} \equiv \widetilde{\mathcal{S}}$, i.e. the overall effective Sommerfeld enhancement of the subhalo population, by inspecting the analytical expression of each asymptotic regime of the mass-integrated Sommerfeld enhancement $\widetilde{\mathcal{S}}$:[8]

$$\widetilde{\mathcal{S}}_{\rm Coul} = \frac{\alpha_{\rm boost}}{\alpha_{\rm Coul}}\, S_0\, \tilde{\mu}_{\max}^{(1+p)\,\nu}\, \frac{\left(\max(\tilde{\mu}_{\rm sat},\mu_{\min})\right)^{-\alpha_{\rm Coul}}}{\mu^{-\alpha_{\rm boost}}\Big|^{\mu_{\min}}_{\mu_{\max}}}\,, \tag{4.62a}$$

$$\widetilde{\mathcal{S}}_{\rm sat} = \frac{\alpha_{\rm boost}}{\alpha_{\rm sat}}\, S_0\, S_1\, \tilde{\mu}_{\max}^{\nu}\, \tilde{\mu}_{\rm sat}^{-\nu(p+1)}\, \frac{\mu^{-\alpha_{\rm sat}}\Big|^{\mu_{\min}}_{\min(\mu_{\max},\tilde{\mu}_{\rm sat})}}{\mu^{-\alpha_{\rm boost}}\Big|^{\mu_{\min}}_{\mu_{\max}}}\,, \tag{4.62b}$$

$$\widetilde{\mathcal{S}}_{\rm res} = \frac{\alpha_{\rm boost}}{\alpha_{\rm res}}\, S_0^{\rm res}\, \tilde{\mu}_{\max}^{\nu}\, \tilde{\mu}_{\rm sat}^{-\nu(p-1)}\, \frac{\mu^{-\alpha_{\rm res}}\Big|^{\max(\mu_{\min},\tilde{\mu}_{\rm unit})}_{\min(\mu_{\max},\tilde{\mu}_{\rm sat})}}{\mu^{-\alpha_{\rm boost}}\Big|^{\mu_{\min}}_{\mu_{\max}}}\,. \tag{4.62c}$$

We have removed the relevant step functions assuming $m_{\min} < \tilde{m}_{\rm sat} < m_{\max}$ for simplicity, but the general result can easily be derived from Eqs. (4.57)-(4.57) and Eq. (4.30)—a couple of footprints of the general result are still indicated with the min() and max() functions. The only parameter that depends on ϵ_ϕ is $\tilde{\mu}_{\rm sat}$, while the only parameter that depends on $M_{\rm host}$ is $\mu_{\max} \simeq \mu_{\rm host}/100$. The other masses, including the minimal dimensionless subhalo mass $\mu_{\min}$, are taken universal.[9] Since $\mu_{\min} \ll \tilde{\mu}_{\rm sat}(\epsilon_\phi)$ over a wide range in ϵ_ϕ, we can assume that most subhalos are in the saturation regime. Therefore, they contribute both to the baseline and to the resonance peaks of the overall Sommerfeld factor (we can disregard the Coulomb regime). Looking each term of the initial ratio expression, we see that the denominator $J_{\rm sub}$ is $\propto \mu_{\min}^{-\alpha_{\rm boost}}$ in the s-wave case, and therefore can be assumed constant for any host halo. This explains why all plots in the left column of Fig. 8, which are associated with different host halo masses, look the same. On the other hand, in the p-wave case, the denominator is $\propto \mu_{\max}^{-\alpha_{\rm boost}} \propto (\mu_{\rm host}/100)^{-\alpha_{\rm boost}}$, and therefore $\widetilde{\mathcal{S}}$ indirectly depends on the host halo mass, which is readily verified in the right panels.

It is now straightforward to further predict the scaling in $\epsilon_\phi \propto \tilde{\mu}_{\rm sat}^{\nu}$ [see Eq. (4.44)] and $\mu_{\rm host}$ from the previous analytical expressions, by accounting for the signs of the α's indices

[8]Note that for p-wave annihilation, the Sommerfeld-free J-factor $J_{\rm tot} \gg J_{\rm sub}$, so that $J_{\mathcal{S},\rm tot}/J_{\rm tot}$ as reported in Fig. 8 does not strictly measure the amplitude of the overall Sommerfeld boost factor $\widetilde{\mathcal{S}} \equiv J_{\mathcal{S},\rm sub}/J_{\rm sub}$ in that case, but rather the full combined boost factor and its scaling with ϵ_ϕ.

[9]Strictly speaking, $\mu_{\min}$ depends on the full underlying particle physics scenario, and may therefore depend on ϵ_ϕ [59]. Here, we assume that self-interactions play no role in setting the kinetic decoupling of DM particles in the early universe, and thereby that $\mu_{\min}$ does not depend on ϵ_ϕ.

given in Eq. (4.59):

$$\widetilde{\mathcal{S}}_{\rm Coul} \propto \mu_{\rm host}^{\frac{p}{2}\alpha_{\rm boost}}\,\mu_{\rm min}^{\frac{(2-p)}{2}\alpha_{\rm boost}}\,\epsilon_\phi^{-\frac{\alpha_{\rm Coul}}{\nu}} \tag{4.63a}$$

$$\widetilde{\mathcal{S}}_{\rm sat} \propto \mu_{\rm host}^{\frac{p}{2}\alpha_{\rm boost}}\,\mu_{\rm min}^{\frac{(2-p)}{2}(\alpha_{\rm boost}-\alpha_{\rm sat})}\,\epsilon_\phi^{-(p+1)} \times \begin{cases} \mu_{\rm host}^{-\frac{p}{2}\alpha_{\rm sat}} & \text{if } \mu_{\rm max} < \tilde{\mu}_{\rm sat} \\ \epsilon_\phi^{-\frac{p}{2}\frac{\alpha_{\rm sat}}{\nu}} & \text{else} \end{cases} \tag{4.63b}$$

$$\widetilde{\mathcal{S}}_{\rm res} \propto \mu_{\rm host}^{\frac{p}{2}\alpha_{\rm boost}}\,\mu_{\rm min}^{\frac{(2-p)}{2}\alpha_{\rm boost}}\,\{\max(\mu_{\rm min},\tilde{\mu}_{\rm unit})\}^{-\alpha_{\rm res}}\,\epsilon_\phi^{-(p-1)}\,. \tag{4.63c}$$

We recall that $p = 0$ (2) for an s-(p-)wave annihilation. From these expressions, we can understand why in the s-wave case (left-column panels of Fig. 7 and Fig. 8) the baseline of the saturation regime goes $\propto \epsilon_\phi^{-1}$, while the curve following the amplitude of the peaks is instead $\propto \epsilon_\phi$. We also understand why in the p-wave case (right-column panels), the baseline of the saturation regime experiences a change in the slope in ϵ_ϕ when $\tilde{\mu}_{\rm sat} < \mu_{\rm max} \sim \mu_{\rm host}/100$. This is due to the fact that the exponent $\alpha_{\rm sat}$ is negative in the p-wave case, which implies that it is the upper bound of the subhalo mass integral $\min(\mu_{\rm max}, \tilde{\mu}_{\rm sat})$ that matters. Indeed, as shown in the second equation above, the scaling in ϵ_ϕ goes from $\propto \epsilon_\phi^{-3}$ when $\mu_{\rm max} < \tilde{\mu}_{\rm sat}$ to a much more moderate $\propto \epsilon_\phi^{-3-\frac{\alpha_{\rm sat}}{\nu}} \sim \epsilon_\phi^{-1.3}$ when $\mu_{\rm max} \geqslant \tilde{\mu}_{\rm sat}$, which explains why the increase of the ratio is first very steep as ϵ_ϕ decreases from large values, and then changes of slope. This is particularly visible for the lightest host halo with a mass of $10^8\,M_\odot$ because then $m_{\rm max} \sim 10^6 M_\odot$, which corresponds to a scaling transition around $\epsilon_\phi^{\rm sat}(m_{\rm max}) \sim 6\times 10^{-3}$ (see bottom right panel of Fig. 4), above which no subhalo can participate in the saturation regime. The same transition is slightly less visible for the host halo of $10^{12}\,M_\odot$. On the other hand, the scaling of the resonance peaks does not feature any such transition, as expected from the analytical results. The peak-to-baseline ratio can be fully understood from Eq. (4.60), and associated discussion. We emphasize that in our template calculations, the peak amplitudes are fixed by $\tilde{m}_{\rm unit} \sim 8 \times 10^{-4}\,M_\odot$, not by $m_{\rm min} = 10^{-6}\,M_\odot < \tilde{m}_{\rm unit}$.

From these plots, we see that a quick integration of our simplified ansatz describes reasonably well the more accurate numerical results, slightly degrading from the s-wave to the p-wave case. This departure from the numerical results comes from the error made by changing the phase-space integral by an evaluation through a characteristic speed $\overline{v}(m)$, which needs to be adjusted by playing with the value of ω_0 in Eq. (4.39) (the tuning values are given in the plots, and are fixed once and for all for a given configuration). Further splitting the subhalo mass integral into analytical asymptotic pieces as done just above to get analytical approximations and insight on the full result would induce bigger numerical errors (a factor of a few for s-wave processes, up to an order of magnitude for p-wave processes), because the actual Sommerfeld-enhancement factor transits smoothly between regimes over the available mass range. Still, the full analytical prediction gets the scaling relations correct, which strongly helps in the interpretation. We also see from Fig. 7 that a very simple expression, like that in Eq. (4.57), can be used for quick signal predictions to a reasonable precision, without resorting to a complex numerical machinery.

To summarize, independently of the scaling relations, we see that the overall Sommerfeld effect induced by subhalos does not change the host target hierarchy in the s-wave annihilation case, because it is driven by the minimal subhalo mass $\mu_{\rm min}$, taken the same for all host halos and all values of ϵ_ϕ. Decreasing $\mu_{\rm min}$ would simply enhance the signal by the same amount for all host halos (though one should keep in mind the unitarity limit on resonances, set by $\tilde{\mu}_{\rm unit}$ if $\mu_{\rm min} < \tilde{\mu}_{\rm unit}$). In contrast, in the p-wave case, we see that the subhalo contribution to the global Sommerfeld enhancement is relatively stronger for lighter host halos (see the right panels of Fig. 8), and could therefore potentially invert the hierarchy of the Sommerfeld-free p-wave signal set by the (squared) dispersion velocities of the most massive subhalos. This is due to the fact that while further suppressing the overall Sommerfeld-free signal, subhalos now act as extra enhancement factors due to their smaller dispersion velocities, making the Sommerfeld-enhanced to Sommerfeld-free ratio much more contrasted than in the s-wave case.

4.3.4 Sommerfeld-enhanced subhalo boost factor

We can now determine the overall Sommerfeld-corrected subhalo boost factor for a host of mass $M_{\rm host}$. It may be written as [see Eq. (4.34)]

$$\mathcal{B}_{\mathcal{S}} \simeq 1 + \frac{J_{\mathcal{S},\rm sub}(\epsilon_\phi)}{J_{\mathcal{S},\rm host}(\epsilon_\phi)}, \tag{4.64}$$

where the Sommerfeld-enhanced contribution of the host in the denominator is given in Eq. (4.49), while the Sommerfeld-enhanced contribution of the subhalo population is given in Eq. (4.57). Here, the difficulty comes from the fact that different Sommerfeld regimes come about at different values of ϵ_ϕ depending on the (sub)halo masses (including the host halo). The clearest way to understand the net impact of subhalos in Sommerfeld-enhanced scenarios is to separate the discussion in terms of the different Sommerfeld regimes for the host halo.

We order the different Sommerfeld regimes of the host halo by varying the reduced Bohr radius ϵ_ϕ from large to small values. Therefore, we first discuss the saturation regime, and then the Coulomb regime. Note that for host halo masses of 10^{15}, 10^{12}, and 10^8 M , the transition between these regimes occurs around $\epsilon_\phi^{\rm sat} \sim 0.5, 5\times 10^{-2}, 3\times 10^{-3}$, respectively. A concrete illustration is given in Fig. 9, where we see the Sommerfeld-corrected subhalo boost factors computed for the s-wave (p-wave) annihilation case in the left column (respectively right column) panels, and for different host halo masses.

Saturation regime of the host halo The saturation regime of the host halo corresponds to values of $\epsilon_\phi > \epsilon_\phi^{\rm sat}(M_{\rm host})$, or equivalently $\tilde{m}_{\rm sat}(\epsilon_\phi) > M_{\rm host}$—see Eqs. (4.54) and (4.44). Since subhalos are all lighter than the host halo, they are also all in the saturation regime. The boost factor can then be written as, assuming that $m_{\rm min} < m_{\max} < M_{\rm host} < \tilde{m}_{\rm sat}(\epsilon_\phi)$:

$$\mathcal{B}_{\mathcal{S}}^{\rm sat} - 1 \simeq \frac{J^{\rm sat}_{\mathcal{S},\rm sub}(\epsilon_\phi)}{J^{\rm sat}_{\mathcal{S},\rm host}(\epsilon_\phi)} = \frac{\gamma}{(1-p)\alpha_{\rm sat}}\frac{A^2_{0,\rm host}}{A^2_{0,\rm sub}}\left\{\left(\frac{\mu_{\rm min}}{\mu_{\rm host}}\right)\right\}^{-\frac{(2-p)}{2}\alpha_{\rm sat}}\left\{\left(\frac{\mu_{\max}}{\mu_{\rm host}}\right)\right\}^{-\frac{p}{2}\alpha_{\rm sat}} \tag{4.65}$$

where the multiple appearance of p here is simply a trick to account for the change of sign in the boost mass spectral index $\alpha_{\rm sat}$ between the s- and p-wave cases in our specific choice of parameters, which makes either $\mu_{\rm min}$ or $\min(\mu_{\max}, \tilde{\mu}_{\rm sat}) = \mu_{\max} \propto \mu_{\rm host}$ dominate the mass function integral. From this equation, we clearly understand why in the saturation regime of the host halo (right parts of panels in Fig. 9), the baselines of both the s-wave and p-wave boost factors remain constant: this is due to the fact that they are independent of ϵ_ϕ. The s-wave one has its amplitude $\propto \mu_{\rm host}/\mu_{\rm min}$, though hindered by a small power index $\alpha_{\rm sat} \sim 0.1$. In contrast, the boost factor amplitude for p-wave annihilation is vanishingly small because

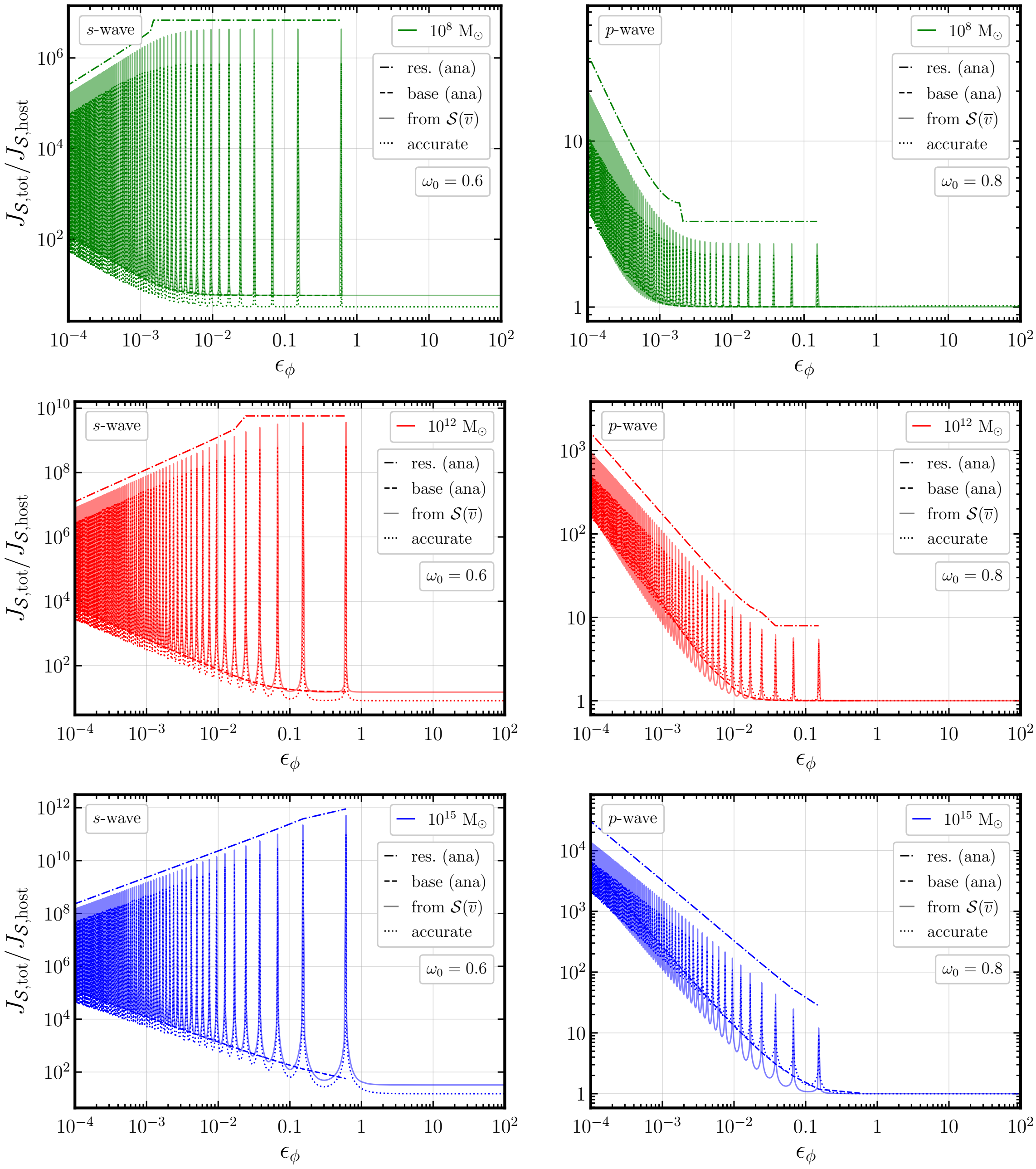

Figure 9. **Left column:** Subhalo boost factor for s-wave Sommerfeld-enhanced DM annihilation, calculated for three typical host halos of masses 10^8, 10^{12}, and 10^{15} M , from top to bottom panels. The different curves show (i) the exact numerical results (dotted lines), (ii) an approximation in which we calculate the mass integral numerically by taking the exact Sommerfeld formula but evaluated at the characteristic velocity of (sub)halos (plain curves), and (iii) the integrated analytical ansatz, which is reported for both the baseline and the peak amplitude (dot-dashed curves). **Right column:** Same for p-wave annihilation.

$(\mu_{\mathrm{max}}/\mu_{\mathrm{host}})^{-\alpha_{\mathrm{sat}}} < 1$, as a consequence of $\alpha_{\mathrm{sat}} < 0$ in that case (asymptotically similar to the Sommerfeld-free case).

On resonances, this becomes

$$\mathcal{B}_{\mathcal{S}}^{\rm res} - 1 \simeq \frac{J_{\mathcal{S},\rm sub}^{\rm res}(\epsilon_\phi)}{J_{\mathcal{S},\rm host}^{\rm res}(\epsilon_\phi)} = \frac{\gamma}{\alpha_{\rm res}} \frac{A_{0,\rm host}^2}{A_{0,\rm sub}^2} \left\{ \frac{\mu_{\rm min}}{\mu_{\rm host}} \right\}^{-\alpha_{\rm res}} . \tag{4.66}$$

The minimal subhalo mass $\mu_{\rm min}$ is featured in blue to keep in mind that it should be replaced by $\tilde{\mu}_{\rm unit}$ when $\mu_{\rm min} < \tilde{\mu}_{\rm unit}$ (this is the case in our template examples, but this is not generic). Interestingly, when both the host halo and its subhalos sit on resonances ($m < \tilde{m}_{\rm sat}$), the boost factor does not depend on Sommerfeld parameters, and remains flat as a function of ϵ_ϕ. This can be seen from all panels of Fig. 9 by inspecting the right-hand-side peaks (more peaks are concerned as the host halo mass decreases, as the latter remains longer in the saturation regime). Not visible in this formula but also theoretically important, there is a formal difference between predictions of the s- and p-wave boost factors on resonant peaks. In the latter case, the Sommerfeld-enhanced annihilation cross section does not depend on DM velocity (the v^2 suppression is canceled out by the $1/v^2$ enhancement), which in principle reduces the potential error associated with the approximation of trading the phase-space average of the effective Sommerfeld factor for a its local expression evaluated at an average characteristic velocity.

Coulomb regime of the host halo When the host halo is in the Coulomb regime, then subhalos can themselves be either in the Coulomb regime or in the saturation regime (which includes resonances). Therefore, we need to combine all possibilities, which depend on whether $\tilde{m}_{\rm sat}$ lies within the subhalo mass range $[m_{\rm min}, m_{\rm max}]$ or not. The corresponding expression for the boost factor is slightly more involved:

$$\begin{aligned}\mathcal{B}_{\mathcal{S}}^{\rm Coul} - 1 \simeq \frac{J_{\mathcal{S},\rm sub}(\epsilon_\phi)}{J_{\mathcal{S},\rm host}^{\rm Coul}(\epsilon_\phi)} = \gamma\, \mu_{\rm host}^{\alpha_{\rm Coul}} \Bigg\{ \Bigg(\frac{\theta\,(\mu_{\rm max} - \tilde{\mu}_{\rm sat})}{\alpha_{\rm Coul}} [\max{(\tilde{\mu}_{\rm sat}, \mu_{\rm min})}]^{-\alpha_{\rm Coul}} \\ + \theta(\tilde{\mu}_{\rm sat} - \mu_{\rm min}) \frac{S_1}{(1 - p)\alpha_{\rm sat}} \left\{ \frac{\tilde{\mu}_{\rm max}}{\tilde{\mu}_{\rm sat}} \right\}^{\nu(p+1)} \mu_{\rm min}^{-\frac{(2-p)}{2}\alpha_{\rm sat}} [\min(\tilde{\mu}_{\rm sat}, \mu_{\rm max})]^{-\frac{p}{2}\alpha_{\rm sat}} \\ + \sum_{n=1+\frac{p}{2}} \Bigg(\delta_{\epsilon_\phi/\{\epsilon_\phi^{{\rm res},n}\}}\, \theta(\tilde{\mu}_{\rm sat} - \mu_{\rm min}) \frac{S_0^{\rm res}}{\alpha_{\rm res} S_0} \left\{ \frac{\tilde{\mu}_{\rm max}}{\tilde{\mu}_{\rm sat}} \right\}^{\nu(p+1)} \mu_{\rm min}{}^{-\alpha_{\rm res}}\, \tilde{\mu}_{\rm sat}^{2\nu} \Bigg\} .\end{aligned} \tag{4.67}$$

Assuming $m_{\rm min} < \tilde{m}_{\rm sat} < m_{\rm max}$, this expression simplifies to:

$$\begin{aligned}\mathcal{B}_{\mathcal{S}}^{\rm Coul} - 1 \simeq \gamma\, \mu_{\rm host}^{\alpha_{\rm Coul}} \Bigg\{ \Bigg(\frac{\tilde{\mu}_{\rm sat}{}^{-\alpha_{\rm Coul}}}{\alpha_{\rm Coul}} \propto \epsilon_\phi^{-\frac{\alpha_{\rm Coul}}{\nu}} \\ + \frac{S_1}{(1 - p)\alpha_{\rm sat}} \left\{ \frac{\tilde{\mu}_{\rm max}}{\tilde{\mu}_{\rm sat}} \right\}^{\nu(p+1)} \mu_{\rm min}^{-\frac{(2-p)}{2}\alpha_{\rm sat}}\, \tilde{\mu}_{\rm sat}^{-\frac{p}{2}\alpha_{\rm sat}} \propto \epsilon_\phi^{-(p+1) - \frac{\alpha_{\rm sat} p}{2\nu}}\, \mu_{\rm min}^{-\frac{(2-p)}{2}\alpha_{\rm sat}} \\ + \sum_{n=1+\frac{p}{2}} \Bigg(\delta_{\epsilon_\phi/\{\epsilon_\phi^{{\rm res},n}\}} \frac{S_0^{\rm res}}{\alpha_{\rm res} S_0} \left\{ \frac{\tilde{\mu}_{\rm max}}{\tilde{\mu}_{\rm sat}} \right\}^{\nu(p+1)} \mu_{\rm min}{}^{-\alpha_{\rm res}}\, \tilde{\mu}_{\rm sat}^{2\nu} \propto \epsilon_\phi^{1-p}\, \mu_{\rm min}{}^{-\alpha_{\rm res}} \Bigg\} \Bigg(\end{aligned} \tag{4.68}$$

From the top to bottom lines, we have the Coulomb/Coulomb, saturation/Coulomb, and resonant peaks/Coulomb terms. The scaling in ϵ_ϕ and $\mu_{\rm min}$ is made explicit at the end of each line, for convenience.

Quite generically, we first see that when the host halo lies in the Coulomb regime, the subhalo boost factor is $\propto \mu_{\rm host}^{\alpha_{\rm Coul}}$, with $\alpha_{\rm Coul} \sim 0.57 > 0$ here. Consequently, the boost factor increases with the host halo mass for both s- and p-wave Sommerfeld-enhanced annihilation processes, a result similar to the Sommerfeld-free result for the s-wave annihilation [see Eq. (4.36)].

The Coulomb/Coulomb term is not visible Fig. 9, and would asymptotically take over in the extreme left parts of the panels at lower values of $\epsilon_\phi < \epsilon_\phi^{\rm sat}(\mu_{\rm min})$. It would then freeze in as $\tilde{\mu}_{\rm sat} \sim \mu_{\rm min}$, and remain constant, $\propto \mu_{\rm min}^{-\alpha_{\rm Coul}}$, similar to the Sommerfeld-free boost factor for s-wave annihilation processes.

The saturation/Coulomb term characterizes the baseline of the boost factor over a large range of $\epsilon_\phi \in [\epsilon_\phi^{\rm sat}(\mu_{\rm min}), \epsilon_\phi^{\rm sat}(\mu_{\rm host})]$. In the s-wave case, it scales like $\propto \mu_{\rm host}^{\alpha_{\rm Coul}}\, \epsilon_\phi^{-1}\, \mu_{\rm min}^{-\alpha_{\rm sat}}$. Except for the explicit host halo mass dependence, which sets an absolute hierarchy, we see that the scaling in $\mu_{\rm min}$ and ϵ_ϕ is the same for all halos. The only implicit difference is that the onset of the subhalo saturation regime at $\epsilon_\phi^{\rm sat}(\mu_{\rm host})$ shifts to lower values as $\mu_{\rm host}$ decreases. We can therefore understand why the boost factor behaves the same, i.e. it increases linearly $\propto \epsilon_\phi^{-1}$ as ϵ_ϕ decreases, while with some increasing delay as $\mu_{\rm host}$ decreases. Besides, the minimal subhalo mass $\mu_{\rm min}$ participates in setting the overall amplitude of the saturation baseline of the subhalo boost factor, which increases as $\mu_{\rm min}$ decreases. In the p-wave case, the baseline scales like $\propto \mu_{\rm host}^{\alpha_{\rm Coul}}\, \epsilon_\phi^{-3-\frac{\alpha_{\rm sat}}{\nu}} \sim \mu_{\rm host}^{\alpha_{\rm Coul}}\, \epsilon_\phi^{-1.3}$, which is independent of $\mu_{\rm min}$ (as long as $\mu_{\rm min} < \tilde{\mu}_{\rm sat}$). The slope in ϵ_ϕ is therefore slightly steeper than in the s-wave case, but the delay in the onset of the saturation regime as ϵ_ϕ decreases is the same.

Finally, the resonant peaks/Coulomb term, which characterizes the amplitude of the subhalo boost factor on resonant peaks, scales like $\propto \mu_{\rm host}^{\alpha_{\rm Coul}}\, \epsilon_\phi^{1-p}\, \max(\mu_{\rm min}, \tilde{\mu}_{\rm unit})^{-\alpha_{\rm res}}$. In addition to the host-halo mass hierarchy set by $\mu_{\rm host}^{\alpha_{\rm Coul}}$, we first see that the amplitude of the boost is also affected by $\mu_{\rm min}$ (or $\tilde{\mu}_{\rm unit}$), but more for s-wave ($\alpha_{\rm res} \sim 0.76$) than for p-wave processes ($\alpha_{\rm res} \sim 0.1$). In contrast, we also see that the dependence in ϵ_ϕ is inverted from the s- to p-wave case, with a scaling $\propto \epsilon_\phi$ in the former case, but $\propto \epsilon_\phi^{-1}$ in the latter case. This explains why the relative amplitude of peaks with respect to the baseline of the subhalo boost decreases faster, $\propto \epsilon_\phi^2$, in the s-wave configuration than in the p-wave one, for which the relative decrease is accordingly $\propto \epsilon_\phi^{0.3}$. This behavior matches exactly the peak-to-baseline ratio of the subhalo signal derived in Eq. (4.60), which means that the signal itself is completely driven by subhalos.

From the plots of Fig. 9, we see that our semi-analytical approximations (numerical mass integrals of analytical expressions) come with significant errors, but get the scaling relations and the orders of magnitude correct. Note that the numerical errors are more exacerbated in the subhalo boost factor than in individual signals because it is a ratio that combines quite different mass scales ($\mu_{\rm host} \gg \mu_{\rm min}$).

4.3.5 Absolute Sommerfeld-enhanced subhalo boost factor

As a last useful result which may help re-evaluating the hierarchy of targets, we calculate the Sommerfeld-enhanced subhalo boost factor with respect to the Sommerfeld-free signal of the host halo. We recall that we keep the subhalo mass slope α "fixed by theory", which determines the signs of the Sommerfeld mass slopes $\alpha_{\rm Coul}$, $\alpha_{\rm sat}$, and $\alpha_{\rm res}$. In that case, we get:

$$\begin{aligned}\mathcal{B}_{\mathcal{S}/\text{no-}\mathcal{S}} - 1 \simeq & \frac{J_{\mathcal{S},\text{sub}}(\epsilon_\phi)}{J_\text{host}(\mu_\text{host})} \\ = & \,\gamma\, \mu_\text{host}^{\alpha_\text{sat}} \Bigg\{ \theta(\mu_\text{max} - \tilde{\mu}_\text{sat}) \frac{S_0}{\alpha_\text{Coul}} \tilde{\mu}_\text{max}^{\nu(p+1)} \left[\max(\tilde{\mu}_\text{sat}, \mu_\text{min})\right]^{-\alpha_\text{Coul}} \\ & + \theta(\tilde{\mu}_\text{sat} - \mu_\text{min}) \left\{\frac{\tilde{\mu}_\text{max}}{\tilde{\mu}_\text{sat}}\right\}^{\nu(p+1)} \times \\ & \left\{ \frac{S_0 S_1}{(1-p)\alpha_\text{sat}} \mu_\text{min}^{-\frac{(2-p)}{2}\alpha_\text{sat}} \left[\min(\tilde{\mu}_\text{sat}, \mu_\text{max})\right]^{-\frac{p}{2}\alpha_\text{sat}} \right. \\ & \left. + \sum_{n=1+\frac{p}{2}} \delta_{\epsilon_\phi/\{\epsilon_\phi^{\text{res},n}\}} \frac{S_0^\text{res}}{\alpha_\text{res}} \tilde{\mu}_\text{sat}^{2\nu} \mu_\text{min}^{-\alpha_\text{res}} \right\} \Bigg\} .\end{aligned} \tag{4.69}$$

Assuming $m_\text{min} < \tilde{m}_\text{sat} < m_\text{max}$, this simplifies as follows:

$$\begin{aligned}\mathcal{B}_{\mathcal{S}/\text{no-}\mathcal{S}} - 1 \simeq & \,\gamma\, \mu_\text{host}^{\alpha_\text{sat}} \Bigg\{ \frac{S_0}{\alpha_\text{Coul}} \tilde{\mu}_\text{max}^{\nu(p+1)} \tilde{\mu}_\text{sat}^{-\alpha_\text{Coul}} \\ & + \left\{\frac{\tilde{\mu}_\text{max}}{\tilde{\mu}_\text{sat}}\right\}^{\nu(p+1)} \times \left\{ \frac{S_0 S_1}{(1-p)\alpha_\text{sat}} \mu_\text{min}^{-\frac{(2-p)}{2}\alpha_\text{sat}} \tilde{\mu}_\text{sat}^{-\frac{p}{2}\alpha_\text{sat}} \right. \\ & \left. + \sum_{n=1+\frac{p}{2}} \delta_{\epsilon_\phi/\{\epsilon_\phi^{\text{res},n}\}} \frac{S_0^\text{res}}{\alpha_\text{res}} \tilde{\mu}_\text{sat}^{2\nu} \mu_\text{min}^{-\alpha_\text{res}} \right\} \Bigg\}\end{aligned} \tag{4.70}$$

From this equation, we see the crucial roles played by both m_min and ϵ_ϕ in the s-wave case ($p = 0$, $\alpha_\text{sat} \sim 0.1$) to set the boost amplitude. We also see that for s-wave processes, the global factor of $\mu_\text{host}^{\alpha_\text{sat}}$ makes the boost factor increase as μ_host increases, which exacerbates the signal hierarchy between targets as function of their mass. In contrast, the p-wave boost factor ($p = 2$, $\alpha_\text{sat} < 0$) is almost entirely fixed by ϵ_ϕ through $\tilde{\mu}_\text{sat}$, as long as $\tilde{\mu}_\text{sat} > \mu_\text{min}$. Besides, the global factor of $\mu_\text{host}^{\alpha_\text{sat}}$ makes the boost factor decrease as μ_host increases for p-wave processes ($\alpha_\text{sat} \sim -0.57$), which now tends to invert the signal hierarchy between targets as function of their masses. On resonance peaks, it is again μ_min (or $\tilde{\mu}_\text{unit}$ if $\mu_\text{min} < \tilde{\mu}_\text{unit}$) that sets the amplitude, with a stronger impact in the s-wave ($\alpha_\text{res} \sim 0.57$) than in the p-wave case ($\alpha_\text{res} \sim 0.1$).

In Fig. 10, we display our different results for the absolute boost factor introduced just above, for the different reference host halo masses. The full numerical calculation results appear as dotted curves, the mass integral performed over the exact Sommerfeld enhancement factor evaluated at the characteristic speeds of subhalos as solid curves, and the integrated ansatz envelope as dot-dashed curves. There is a reasonable agreement between the analytical approximation and the full numerical results. The left-column panels show the results for s-wave annihilation, while the right-column panels show our results for p-wave annihilation (the latter are quite similar to the right panels of Fig. 8, because $J_{\mathcal{S},\text{sub}}(\epsilon_\phi)/J_\text{host} \simeq J_{\mathcal{S},\text{sub}}(\epsilon_\phi)/J_\text{tot}$ in the p-wave case). We do not discuss longer the former, which exhibit no surprise, but we emphasize the inverted hierarchy now occurring in the latter, where it is evident that the absolute boost factor can then be much larger for less massive host halos (from bottom to

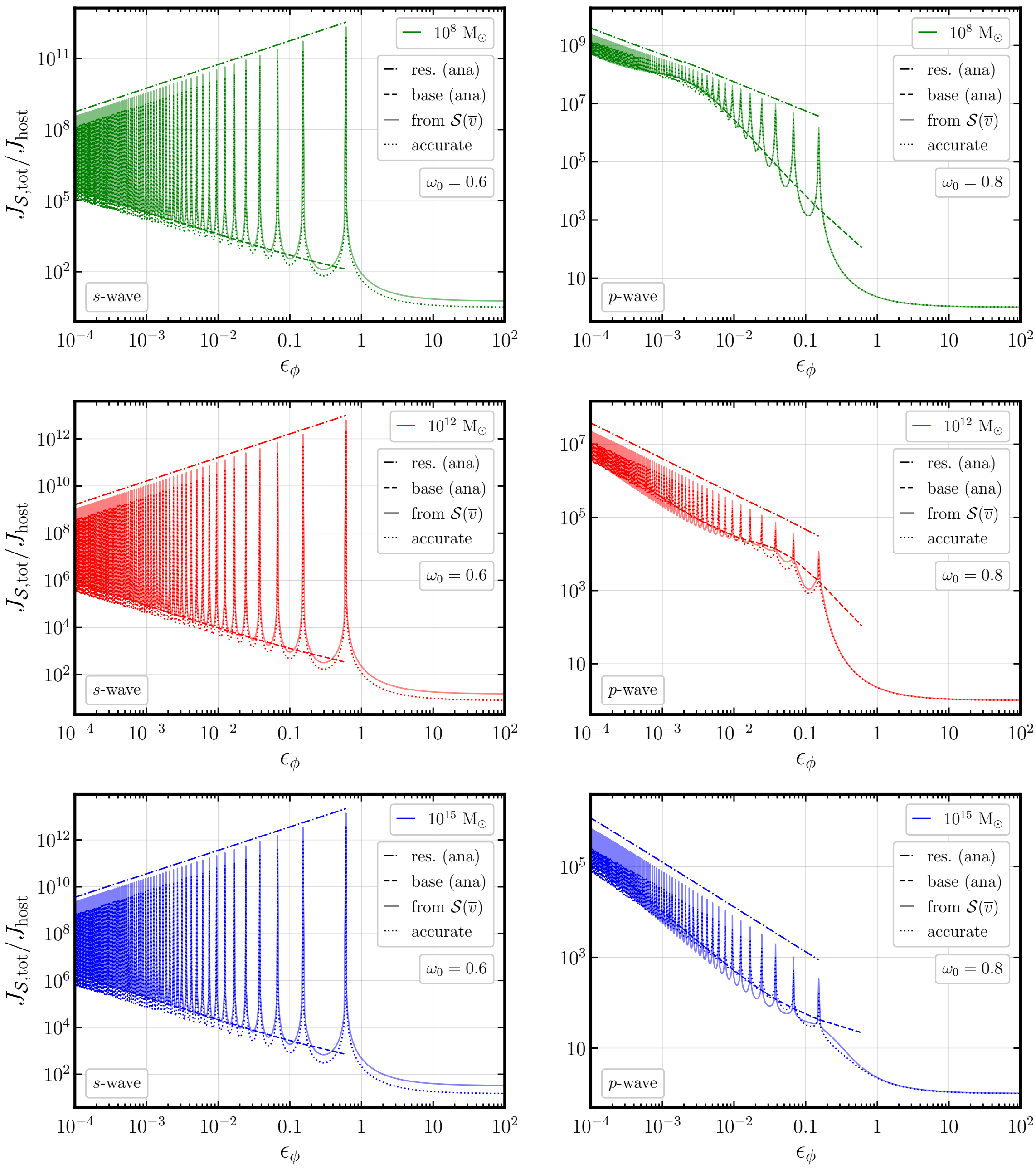

Figure 10. Left column: Absolute subhalo boost factor for s-wave Sommerfeld-enhanced DM annihilation, calculated for three typical host halos of masses 10^8, 10^{12}, and 10^{15} M , from top to bottom panels. The different curves show (i) the exact numerical results (dotted lines), (ii) an approximation in which we calculate the mass integral numerically by taking the exact Sommerfeld formula but evaluated at the characteristic velocity of (sub)halos (plain curves), and (iii) the integrated analytical ansatz, which is reported for both the baseline and the peak amplitude (dot-dashed curves). **Right column:** Same for p-wave annihilation.

top right panels). We have already explained why above, but these plots allow us to be slightly more quantitative. Let us for instance compare a dwarf galaxy-like host halo of mass $m_1 \sim 10^8\,\mathrm{M}_\odot$ and a galaxy-cluster-like host halo of mass $m_2 \sim 10^{15}\,\mathrm{M}_\odot$, both located at the same distance. The p-wave suppression factor induces an extra relative reduction of $\sim (m_1/m_2)^{2\nu} \sim (m_1/m_2)^{2/3} \sim 10^{-3.7}$, not favorable to the lightest host halo. As soon as both host halos have entered the Coulomb regime and have their subhalos contributing in the saturation regime, then a boost factor applies with an inverse balance, giving a boost factor ratio of $\sim (m_1/m_2)^{\alpha_{\rm sat}} \sim (m_1/m_2)^{-0.57} \sim 10^4$ (in perfect agreement with the numerical results in the plots), which fully compensates for the initial Sommerfeld-free p-wave penalty. Such a compensation might actually change the initial hierarchy between targets of different masses, depending on their respective distances to the observer.

4.3.6 Caveats

Before summarizing and concluding, it is useful to mention a few caveats.

(i) The calculations of the J-factors (enhanced or not) presented in this paper assume the integration of a subhalo population over an entire target object, i.e., within its virial radius. If the real target halo has its tidal radius significantly smaller than its virial radius, or if the angular size used to perform the signal analysis is significantly smaller than the angular extension of the target halo, then although the host J-factor may not change significantly, the subhalo contribution (hence the subhalo boost factor) can be more strongly affected because subhalos dominate the overall mass profile in the outskirts of their host halo (they are subject to gravitational tides in the central regions, where they can experience strong mass losses and even disruption). Our results likely overestimate the contribution of the subhalo population to the signal in that case. A possible way out is to rescale our results for subhalos by the tidal-to-virial (or contained-to-virial) mass ratio of the host halo, assuming the missing or lost mass is mostly made of subhalos. This may be particularly relevant to dwarf satellite galaxies that orbit within our Milky Way, and also to targets with large angular extensions in the sky such as galaxy clusters.

(ii) We have used fixed values for the subhalo mass slope α and the free-streaming cutoff mass $m_{\rm min}$. We have motivated the former from theoretical arguments in the framework of concordance cosmology (see App. B), and the latter by uncorrelating DM self-interactions from DM-baryons interactions, but all this does not come without uncertainties. If the primordial power spectrum departs from almost scale invariant and exhibits extra features on small scales, then our predictions would be strongly affected. It is a priori possible to adapt our analysis by starting from another subhalo mass function inferred from a modified primordial spectrum, but explaining what would happen in different specific scenarios goes beyond the scope of this paper. One can easily guess, though, the impact if the only change is in the mass slope α, or even if the mass function exhibits spectral breaks or bumps. On the other hand, if $m_{\rm min}$ becomes related to the intrinsic Sommerfeld parameters, then it is a priori easy to determine the consequences from our results.

(iii) We have limited our study to NFW profiles for both the host halo and the entire subhalo population. Changing the shape of profiles either for the host halo or its subhalos would mostly change our analytical approximation for the J-factor of Eq. (4.27), which would propagate in subhalo-to-host ratios and then affect our predictions. However, we do not expect a significant change in the overall subhalo signal, because subhalos are mostly characterized by cuspy profiles and are essentially not subject to baryonic feedback. Anyway, modifications of inner shapes are in principle not difficult to account for, even analytically. For changes to Einasto profiles [75–77], see, e.g., SL17. For other changes, one simply needs to feed our results with another J-factor-to-mass relation [see Eq. (4.27)].

(iv) We do not include sub-subhalos (nor any subsequent sublayers), which would a priori tend to further increase the Sommerfeld enhancement. Indeed, our subhalo mass function introduced in App. B only contains the first generation of objects (those accreted directly into the host halo). It is actually not difficult to incorporate all layers in the model, which would tend to sharpen the mass function by increasing the effective mass index α [92], and further proceed with the hard-sphere approximation discussed above or in the appendix. This goes beyond the scope of this paper, and would certainly add extra theoretical uncertainties related to tidal effects internal to the different layers of subhalos. A more detailed semi-analytical subhalo population study is in preparation [93], where it is shown that these additional layers mostly shape the lower part of the subhalo mass range (see also ref. [70] for another merger-tree inferred subhalo population example).

5 Summary and conclusion

In this paper, we have reviewed quite in detail how the presence of DM subhalos affects the gamma-ray signal amplitude predictions in a scenario in which DM self-interacts through long-range interactions, leading to the Sommerfeld enhancement of the annihilation cross section. We have proposed a simplifying analytical ansatz in Eq. (3.17) to incorporate the rather complex Sommerfeld enhancement factor in the signal predictions, showing that calculations can then be performed fully analytically. This helps better understand the dependencies of signal predictions in terms of the main physical parameters. These parameters are the Sommerfeld-enhancement parameters on the one hand, dictated by particle physics only, and the subhalo parameters on the other hand, dictated both by cosmology (DM power spectrum and structure formation) and particle physics (minimal subhalo mass). We adopted a simplifying description for the former by means of a DM fine-structure constant α_χ, that we have kept fixed to 0.01 throughout the paper, and of the reduced Bohr radius ϵ_ϕ, which then characterizes the mediator-to-DM mass ratio (the Sommerfeld-enhancement regime typically corresponds to $0 < \epsilon_\phi < 1$, triggered at velocities $v < \pi\alpha_\chi$). Although it is formally DM velocity-dependent and related to local interactions, we have shown that an averaged Sommerfeld factor could be expressed at the level of a full DM halo of virial mass m by means of the corresponding characteristic dispersion velocity $\overline{v}(m)$, and that ϵ_ϕ could be turned into a transition velocity $\tilde{v}_{\rm sat}(\epsilon_\phi)$ below which halos transit from the Coulomb regime to the saturation regime of the Sommerfeld enhancement. This global expression of the Sommerfeld effect allowed us to perform the whole chain of calculation fully analytically up to the gamma-ray signal amplitude in terms of J-factors. Our main results for the subhalo population signal, given a host halo mass, are summarized in Eqs. (4.57)-(4.58), to be compared with the Sommerfeld-free results in Eq. (4.30). They can also be expressed as a Sommerfeld-enhanced subhalo boost factor with respect to the Sommerfeld-enhanced smooth-halo approximation [see Eqs. (4.65)-(4.68)], or with respect to the Sommerfeld-free smooth-halo approximation [see Eqs. (4.69)-(4.70)]. We have shown that our analytical results are in reasonable agreement with the more accurate numerical calculations (but still in excellent qualitative agreement), by a factor of a few (with respect to amplitudes of several orders of magnitude), and can therefore be used for quick estimates associated with specific targets.

As a general conclusion, we see that the Sommerfeld enhancement exacerbates the subhalo boost factor, and vice versa. This is true not only for s-wave processes, for which a subhalo boost factor was already present in the Sommerfeld-free case, but also and more dramatically so for p-wave processes, for which subhalos tended to further suppress the signal in the Sommerfeld-free case. In the latter case, this comes from a full compensation of the v^2 p-wave suppression factor by the velocity-dependent Sommerfeld factor. This may lead to changes in the hierarchy of targets as a function of mass (assuming the same distance to the observer), initially more favorable to bigger host halos in the Sommerfeld-free case, but then conversely to less massive halos in the Sommerfeld-enhanced case. For both s- and p-wave processes, the enhancement at resonances is phenomenal. For s-wave processes, we have boost factors ranging from $\sim 10^8$ for dwarf-like host halos ($10^8\mathrm{M}$) up to $\sim 10^{13}$ for galaxy-cluster-like host halos ($10^{15}\mathrm{M}$), for the first peaks, decreasing like $1/n^2 \propto \epsilon_\phi$, where n is the order of the resonance. They are more moderate for p-wave annihilation, ranging from a few for a $10^8\mathrm{M}$ host halo up to a few tens for a $10^{15}\mathrm{M}$ host halo, but increasing like $n^2 \propto 1/\epsilon_\phi$ with the order of resonance — this still shows that subhalos provide the dominant contribution to the overall signal for Sommerfeld-enhanced p-wave annihilation processes, and therefore must be included in the predictions.

There has been lots of studies considering the Sommerfeld enhancement induced by subhalos, e.g. [26–33] (see also refs. [34–38]). Most of them address the s-wave case, and overall, our results are in qualitative agreement with these. There are quantitative differences coming from the different theoretical assumptions or parameters used, but our analytical results can be applied to a wide range of model configurations, and should allow to recover (or complete in overlooked regimes) those of past studies. We are not aware of such a full analytical derivation, especially for p-wave annihilation, thus we hope that our study will allow the reader to grasp the very details of the Sommerfeld-enhanced subhalo contributions to gamma-ray signals. These questions are further explored in a companion study [39], dedicated to a thorough analysis of the combined Sommerfeld and subhalo enhancement effects on concrete target examples.

Acknowledgments

We would like to thank the anonymous referee for her/his constructive and precise comments which helped us improve the presentation of our results. This work has been supported by funding from the ANR project ANR-18-CE31-0006 (*GaDaMa*), from the national CNRS-INSU programs PNHE and PNCG, and from European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement N° 860881-HIDDeN. JPR work is supported by grant SEV-2016-0597-17-2 funded by MCIN/AEI/10.13039/501100011033 and "ESF Investing in your future". MASC was also supported by the *Atracción de Talento* contracts no. 2016-T1/TIC-1542 and 2020-5A/TIC-19725 granted by the Comunidad de Madrid in Spain. The work of JPR and MASC was additionally supported by the grants PGC2018-095161-B-I00 and CEX2020-001007-S, both funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe". GF acknowledges support of the ARC program of the Federation Wallonie-Bruxelles and of the Excellence of Science (EoS) project No. 30820817 - be.h "The H boson gateway to physics beyond the Standard Model".

A Short review of the Sommerfeld enhancement

In this appendix section, we shortly review the impact of DM self-interaction on DM self-annihilation, which leads to the Sommerfeld enhancement. We consider a phenomenological scenario in which DM particles self-interact through the exchange of a (light) mediator ϕ of mass m_ϕ with coupling $g_\chi = \sqrt{4\pi\alpha_\chi}$, where α_χ plays the role of a dark fine structure constant. In this approach, attractive self-interactions between non-relativistic DM particles are described by an attractive Yukawa potential,

$$V_{\rm Y}(r) = -\alpha_\chi \frac{{\rm e}^{-m_\phi r}}{r}\,, \tag{A.1}$$

with r the relative distance between two annihilating DM particles. In the absence of self-interaction, i.e. for $\alpha_\chi = 0$, the annihilation cross section times relative velocity $(\sigma v_{\rm rel})_0$ is computed perturbatively from the short-range annihilation process. However, a long-range Yukawa potential, which encodes multiple exchanges of the light mediator between the two incoming DM particles, can distort the wave function of the corresponding two-body system in a non-perturbative way, leading to Sommerfeld enhancement of the annihilation cross section.[10] The Sommerfeld-enhanced cross section is then expressed as in Eq. (3.1), which we simply repeat here [44]:

$$\sigma v_{\rm rel} = (\sigma v_{\rm rel})_0 \times \mathcal{S}_\ell\,, \tag{A.2}$$

where $v_{\rm rel}$ is the relative speed of DM particles, and the enhancement factor $\mathcal{S}_\ell$ is computed by solving the Schrödinger equation for the radial part of the wave function $R_\ell(r)$ for the partial wave with angular momentum ℓ (e.g. [16, 17, 19]),

$$\left(-\frac{\hbar^2}{m_\chi}\partial_r^2 - m_\chi\frac{v^2}{c^2} + V_{\rm Y}(r) + \frac{\hbar^2\ell(\ell+1)}{m_\chi r^2}\right)\chi_\ell(r) = 0\,, \tag{A.3}$$

where $\chi_\ell(r) = rR_\ell(r)$ and $v = v_{\rm rel}/2$ the velocity of the incoming DM particles in the center-of-mass frame. Eq. (A.3) is solved with the boundary conditions that the interaction only leads to outgoing spherical plane waves at infinity, and with $R_\ell(r) \propto r^\ell$ as $r \to 0$. Then the Sommerfeld enhancement factor for partial wave ℓ reads [16, 17, 19]

$$\mathcal{S}_\ell = \left|\frac{(2\ell+1)!!\,\chi_\ell^{\ell+1}(0)}{(\ell+1)!\,k^{\ell+1}}\right|^2\,, \tag{A.4}$$

where $k = m_\chi v/\hbar$, and $(2\ell+1)!! \equiv (2\ell+1)!/(2^\ell \ell!)$. The radial function χ_ℓ can only be obtained numerically when assuming a Yukawa potential, but a good approximation of the latter is given by the Hulthén potential,

$$V_{\rm H}(r) = -\frac{\alpha_\chi m_* {\rm e}^{-m_* r}}{1 - {\rm e}^{-m_* r}}\,, \tag{A.5}$$

for $m_* = (\pi^2/6)m_\phi$. Note that strictly speaking, the above result is only valid for s-wave annihilation ($\ell = 0$), as an extra centrifugal term must be added to derive analytical expressions for larger partial-wave expansion modes [19]. Accounting for this generalization to

[10] We restrict ourselves to symmetric DM with attractive interactions, for which the Sommerfeld factor is effectively an enhancement factor.

$\ell \neq 0$, the radial Schrödinger equation can be solved analytically for the Hulthén potential, leading to a closed form of the Sommerfeld enhancement factor S_ℓ:

$$\mathcal{S}_\ell = \left| \frac{\Gamma(a^-)\Gamma(a^+)}{\Gamma(1+\ell+2\mathrm{i}\epsilon_v/\epsilon_\phi^*)} \frac{1}{\ell!} \right|^2 , \tag{A.6}$$

where ϵ_ϕ and ϵ_v have been defined in Eq. (3.2). Other parameters are: $\epsilon_\phi^* \equiv \pi^2\epsilon_\phi/6$, Γ is the Gamma function, and $a^\pm = 1+\ell+\mathrm{i}\epsilon_v/\epsilon_\phi^*\left(1 \pm \sqrt{1-\epsilon_\phi^*/\epsilon_v^2}\right)$, with a square root to be understood as a complex number.

From this equation, we may derive the relevant expressions for the s-wave and p-wave annihilation processes, corresponding to $\ell = 0$ and $\ell = 1$, respectively. They are given in Eq. (3.3) and Eq. (3.5). This already covers a broad variety of underlying particle-physics models.

B Building up a semi-analytical subhalo population model

Here we provide more technical details as for the modeling of subhalo populations in host halos. This theoretical modeling is improved from ref. [60] (SL17), to which we add a subhalo mass fraction normalization based on first-principle arguments rather than calibrated from cosmological simulation results. We start by rewriting Eq. (4.3) that describes the differential number density of subhalos,

$$\frac{\mathrm{d}n_{\rm sub}(m,R)}{\mathrm{d}m} = \frac{\mathrm{d}^2 N_{\rm sub}}{\mathrm{d}m\,\mathrm{d}V} = N_{\rm tot}\,\frac{1}{K_{\rm tidal}}\frac{\mathrm{d}\overline{\mathcal{P}}_V(R)}{\mathrm{d}V}\int \mathrm{d}c\,\frac{\mathrm{d}^2\mathcal{P}_{c,m}(c,m,R)}{\mathrm{d}c\,\mathrm{d}m}\,. \tag{B.1}$$

As we shall see below, in the above formulation, the concentration and mass pdfs are actually intricate as a result of tidal effects. Therefore, in contrast to many works, we see an explicit dependence of the mass-concentration pdf (consequently also of the mass function) on the position R, which makes the phase space fully intricate. This spatial dependence is induced by tidal stripping effects, which depend on the position of subhalos in the host halo and on its detailed gravitational potential (including all components, DM and baryons). We shall discuss tidal effects in more detail below. Note that in order to interpret $N_{\rm tot}$ as the total number of subhalos in the host, one must have the volume integral of the above equation over the host halo normalized to $N_{\rm tot}$, which constrains the full phase-space integral of the pdfs to be equal to the constant $K_{\rm tidal}$. All this will be more clearly defined below.

Before specifying the pdfs, we can already provide the link between the subhalo number density of Eq. (4.3) and the associated averaged density profile of Eq. (4.2), which is the tidal mass:

$$m_{\rm t}(m,c,R) = \frac{4\pi}{3}r_{\rm s}^3\rho_0\left\{\frac{1}{3}\int_0^{x_{\rm t}}\mathrm{d}x\,x^2\,f(x)\right\}, \tag{B.2}$$

where we define the subhalo profile shape $f(x)$ in terms of the dimensionless radius $x \equiv r/r_{\rm s}$ and subhalo scale density ρ_0 as follows:

$$f(x) \equiv \frac{\rho(r)}{\rho_0}\,. \tag{B.3}$$

The dependence of the tidal mass $m_{\rm t}$ on the virial mass m and concentration c appears indirectly as a dependence on the $r_{\rm s}$ and ρ_0. The dependence on the radial position R within the host halo is further hidden in the upper bound of the volume integral over $f(x)$, the dimensionless tidal radius $x_{\rm t} = x_{\rm t}(m,c,R) \equiv r_{\rm t}(m,c,R)/r_{\rm s}$—we impose $x_{\rm t} = \min(x_{\rm t}, x_{200})$, such that $m_{\rm t}(m,c,R) \leq m$. Here, $r_{\rm t}$ is the subhalo tidal radius and $r_{\rm s}$ its scale radius, given an inner density profile shape $f(x)$. In the following, we will only consider an NFW profile for subhalos,[11] such that

$$f(x) = f_{\rm nfw}(x) = x^{-1}(1+x)^{-2}\,. \tag{B.4}$$

The tidal radius further depends on the virial mass, concentration, position (somewhat related to accretion time), and can be predicted. Our model actually provides such a prediction, based on a detailed description of both the baryonic and global DM components within the host halo [60, 63, 92]. A simplification of the model is to consider that the density profile within $x_{\rm t}$ is not significantly affected by gravitational tides, which is a reasonable approximation [94–96] and can further be justified in some cases from adiabatic invariance arguments [97, 98]. Trying to describe more precisely the evolution of the inner profile would lead to very little change in our predictions, but would be prohibitive in terms of numerical convergence, since DM subhalos may cover up to $\sim$20 orders of magnitude in mass for galaxy clusters.

A related important ingredient of our subhalo population model is the *tidal disruption threshold*, $\epsilon_{\rm t} \geq 0$, which basically allows us to disrupt subhalos with $x_{\rm t} \leq \epsilon_{\rm t}$. This tidal disruption criterion is inspired from studies of tidal disruption performed with dedicated numerical simulations [99], but might be an oversimplified description of this complex process. Still, it allows us to effectively implement tidal disruption in a very efficient way, and study the impact of either aggressive disruption ($\epsilon_{\rm t} \sim 1$), or subhalos strongly resilient to tidal disruption ($\epsilon_{\rm t} \ll 1$). The recent literature tends to suggest that the latter case is more likely [100]. The calculation of the tidal radius $x_{\rm t}$ and the value taken for the disruption threshold $\epsilon_{\rm t}$ are actually key parameters at the origin of the spatial dependence of the mass and concentration pdfs introduced in Eq. (4.3).

We now specify the pdfs introduced in Eq. (4.3). For the initial spatial distribution, we adopt the hard-sphere approximation and simply assume that should subhalos be hard spheres with a negligible encounter rate, they would simply track the global host gravitational potential (like the bodies of N-body simulations), such that:

$$\frac{{\rm d}\overline{\mathcal{P}}_V(R)}{{\rm d}V} = \frac{\rho_{\rm host}(R)}{M_{\rm host}}\theta(R_{\rm host}-R)\,, \tag{B.5}$$

where $R_{\rm host}$ is the radial extent of the host halo, and $M_{\rm host}$ is the total DM mass within $R_{\rm host}$ and allows for normalization to unity over the volume of the host halo. We emphasize that this spatial pdf is *not* the actual spatial distribution of the subhalo population, which accounts for tidal stripping and can formally simply be inferred from Eq. (4.3) as:

$$\frac{{\rm d}\mathcal{P}_V^{\rm actual}(R)}{{\rm d}V} = \frac{n_{\rm sub}}{N_{\rm tot}} \neq \frac{{\rm d}\overline{\mathcal{P}}_V(R)}{{\rm d}V}\,. \tag{B.6}$$

The difference between the "initial" and "final" spatial pdf will become more striking after the impact of tidal stripping on the concentration and mass pdfs is discussed. This explains the term "*driving* pdf" used earlier. Note that if our tidal disruption parameter $\epsilon_{\rm t} \longrightarrow 0$, then the actual spatial distribution tends to the initial one (i.e. the host profile), a trend

[11] See [60] for discussion on the impact on changing the inner subhalo profile).

confirmed by recent work on the resilience of subhalos to tidal effects [101]. On the other hand, non-zero values of $\epsilon_{\rm t}$ up to ~ 0.1 allow us to recover antibiased spatial distributions found in several past analyses of cosmological simulations [84, 85, 102, 103], though very likely affected by spurious numerical effects [100, 104]—we will shortly come back to that below. For completeness, in our numerical study, we will use the following two values:

$$\epsilon_{\rm t} = \begin{cases} 1 & \text{(fragile subhalos)} \\ 0.01 & \text{(resilient subhalos)}, \end{cases} \tag{B.7}$$

with the former very conservatively limiting the number of subhalos, and the latter being more realistic according to recent literature.

We resort to the mass-concentration relation as fitted in ref. [91], to which we further assign a log-normal pdf of constant width $\sigma_c = 0.14 \log(10)$ (in natural logarithm basis), which stems from analyses of cosmological simulations and associated interpretations [91, 105–107]. This pdf, hidden in the mass pdf in Eq. (4.3) (this will appear explicitly below), is initially universal. We denote this universal initial pdf $\mathrm{d}\overline{\mathcal{P}}_c(c, c_0(m))/\mathrm{d}c$, where $c_0(m)$ carries the mass dependence and refers to the mass-concentration relation proposed in ref. [91] (this pdf is taken log-normal, normalized to unity within $1 \leq c < \infty$—see SL17 for details). The spatial dependence of the evolved concentration pdf is then fully induced by our tidal disruption criterion, according to:

$$\frac{\mathrm{d}\mathcal{P}_c(c, c_0(m), R)}{\mathrm{d}c} = \frac{\mathrm{d}\overline{\mathcal{P}}_c(c, c_0(m))}{\mathrm{d}c} \times \theta(x_{\rm t}(m, c, R) - \epsilon_t). \tag{B.8}$$

The main technical difficulty here concerns the calculation of the dimensionless tidal radius, which is detailed in ref. [60]. Note that in addition to being spatial-dependent, $\mathcal{P}_c$ is no longer normalized to unity because of tidal disruption (unless $\epsilon_{\rm t} = 0$), which will actually allow us to predict the total number of surviving subhalos after tidal disruption.

Finally, for the subhalo mass function, we significantly improve over the initial version of the subhalo population model of ref. [60], which was previously used either with power-law mass functions [61–63] or with the Sheth-Tormen mass function [64]. Here, instead, we fully resort to merger-tree techniques. This semi-analytical approach is still based on the extended Press-Schechter formalism [67, 68, 108], which allows us not only to self-consistently incorporate relevant cosmological information,[12] but also to predict the subhalo mass fraction in host halos of different sizes (from dwarf galaxies to galaxy clusters)—in ref. [60], the subhalo mass fraction was a tunable free parameter of the model. We perform a calculation similar to the one presented in refs. [70, 110], which compares very well with cosmological simulations when artificial tidal disruption is included [101, 110]. We have used the merger-tree algorithm introduced in ref. [111] on purpose, because it only depends on cosmological parameters and is not tuned on cosmological simulations—more details on the model upgrade will be given in subsequent papers [92, 93]. We find that the *unevolved* subhalo mass functions for several realizations of merger trees and for different host halos can be very well fitted by the parametric function proposed in ref. [70]:

$$\frac{\mathrm{d}N(m, M_{\rm host})}{\mathrm{d}m} = \frac{1}{M_{\rm host}} \left[\gamma_1 \left(\frac{m}{M_{\rm host}}\right)^{-\alpha_1} + \gamma_2 \left(\frac{m}{M_{\rm host}}\right)^{-\alpha_2} \right] \exp\left\{ -\beta \left(\frac{m}{M_{\rm host}}\right)^{\zeta} \right\}, \tag{B.9}$$

[12] We use the most recent Planck cosmological parameters [109].

with the best-fit parameters $\gamma_1 = 0.014, \gamma_2 = 0.41, \alpha_1 = 1.965, \alpha_2 = 1.57, \beta = 20, \zeta = 3.4$. These parameters very slightly differ from the parameters found in ref. [70], because they derive from different cosmological inputs and normalization procedure. However, this only leads to order percent differences in terms of global subhalo mass fraction. The mass function just above counts the average number of subhalos accumulated along the history of the host halo per "bin" of mass (here assumed to be hard spheres, i.e. keeping their virial masses after accretion). As in other studies, this leads to an *unevolved* effective subhalo mass fraction of $\sim 10\%$ in a mass range $m/M_{\rm host} \in [10^{-5}, 10^{-3}]$ [43, 71, 84, 85, 87, 110, 112] (this fraction is calculated by taking subhalos with their virial masses, not their actual tidal masses). Like for the concentration, the *unevolved* mass function is universal here, prior to any tidal stripping effect.

Note that despite the rather complex form of Eq. (B.9), the unevolved mass function remains rather close to a single power-law function $\propto (m/m_0)^{-\alpha}$, with m_0 an arbitrary normalization and $\alpha \approx \alpha_1 \simeq 1.96$. Therefore, introducing $\mu \equiv m/m_0$ and $\mu_{\rm host} \equiv M_{\rm host}/m_0$, a useful approximation is the following:

$$\frac{{\rm d}N(m, M_{\rm host})}{{\rm d}m} \simeq \frac{N_0}{m_0}\,\mu^{-\alpha} \tag{B.10}$$
$$\text{with } N_0 \equiv \gamma\,\mu_{\rm host}^{\alpha-1} = 1.4 \times 10^{12(\alpha-1)-2} \left\{\frac{\mu_{\rm host}}{10^{12}}\right\}^{\alpha-1} \approx 4.67 \times 10^9 \left\{\frac{\mu_{\rm host}}{10^{12}}\right\}^{0.96},$$

where we have used $\gamma = \gamma_1$. This is the approximation we use to get analytical understanding of our numerical results.

For completeness, it is useful to define the total number of subhalos prior to tidal stripping, $N^0_{\rm tot}$, from the considered subhalo mass range $[m_{\rm min}, m_{\rm max}]$:

$$N^0_{\rm tot} = N^0_{\rm tot}(m_{\rm min}, m_{\rm max}, M_{\rm host}) = \int_{m_{\rm min}}^{m_{\rm max}} {\rm d}m \frac{{\rm d}N(m, M_{\rm host})}{{\rm d}m}, \tag{B.11}$$

such that we can now fully define the unevolved mass pdf prior to tidal effects:

$$\frac{{\rm d}\overline{\mathcal{P}}_m(m)}{{\rm d}m} = \frac{1}{N^0_{\rm tot}} \frac{{\rm d}N(m, M_{\rm host})}{{\rm d}m}. \tag{B.12}$$

This unevolved pdf is then normalized to unity within the considered subhalo mass range.

The *evolved* subhalo mass function, in contrast, accounts for tidal stripping and as a consequence, as emphasized in ref. [60], becomes spatially dependent. This is again induced by the disruption parameter $\epsilon_{\rm t}$, that depletes the subhalo population according to position, mass, and concentration. In our model, the spatially dependent evolved mass pdf is somewhat entangled with the concentration pdf, but can be formally derived from (in terms of the *virial* mass m):

$$\begin{aligned} \frac{{\rm d}^2\mathcal{P}_{c,m}(c, m, R)}{{\rm d}c\,{\rm d}m} &= \frac{{\rm d}\overline{\mathcal{P}}_m(m)}{{\rm d}m} \times \frac{{\rm d}\mathcal{P}_c(c, c_0(m), R)}{{\rm d}c} \\ \Rightarrow \frac{{\rm d}\mathcal{P}_m(m, R)}{{\rm d}m} &= \frac{{\rm d}\overline{\mathcal{P}}_m(m)}{{\rm d}m} \times \int {\rm d}c \frac{{\rm d}\mathcal{P}_c(c, c_0(m), R)}{{\rm d}c}, \end{aligned} \tag{B.13}$$

where $\mathcal{P}_c$ is the evolved concentration pdf given in Eq. (B.8). Since $\mathcal{P}_c$ is not normalized to unity because of tidal disruption, neither is $\mathcal{P}_m$ (except if $\epsilon_{\rm t} = 0$). As a result, we see explicitly here how the concentration pdf is entangled with the mass function as a consequence of tidal disruption, in the formulation of Eq. (4.3).

We have now all the necessary ingredients to determine the total number of subhalos $N_{\rm tot}$ and the normalization constant $K_{\rm tidal}$ introduced in Eq. (4.3). Indeed, we have

$$K_{\rm tidal} = \int_{V_{\rm host}} \mathrm{d}V \int\!\!\Bigg(\frac{\mathrm{d}\overline{\mathcal{P}}_V(R)}{\mathrm{d}V} \mathrm{d}m \frac{\mathrm{d}\mathcal{P}_m(m,R)}{\mathrm{d}m} \leq 1\,, \tag{B.14}$$

and

$$N_{\rm tot} = N^0_{\rm tot} \times K_{\rm tidal} \leq N^0_{\rm tot}\,. \tag{B.15}$$

Note that we have $K_{\rm tidal} \to 1$ and $N_{\rm tot} \to N^0_{\rm tot}$ in the limit $\epsilon_{\rm t} \to 0$, i.e. in the absence of tidal disruption (which does not mean absence of tidal stripping).

For more physical insight on the *real* subhalo mass function, it might prove useful to have access to the tidal mass distribution instead of the virial mass distribution. Indeed, virial masses have no physical meaning for subhalos, for which the only true masses are the tidal ones. The actual spatial-dependent tidal mass function is simply given by:

$$\begin{aligned}\frac{\mathrm{d}\mathcal{P}_{m_{\rm t}}(m_{\rm t},R)}{\mathrm{d}m_{\rm t}} &= \int \mathrm{d}m \int\!\!\Bigg(\mathrm{d}c \frac{\mathrm{d}^2\mathcal{P}_{c,m}(c,m,R)}{\mathrm{d}c\,\mathrm{d}m} \delta(m_{\rm t} - m_{\rm t}^\star(m,c,R)) \\ &\neq \frac{\mathrm{d}\mathcal{P}_m(m,R)}{\mathrm{d}m}\,.\end{aligned} \tag{B.16}$$

In this equation $m_{\rm t}$ is a free variable and $m_{\rm t}^\star(m,c,R)$ is the tidal mass calculated from the model, given m, c, and R; the cross pdf $\mathrm{d}^2\mathcal{P}_{c,m}/\mathrm{d}c\,\mathrm{d}m$ was introduced in Eq. (B.13). An important remark to make here is that even in the absence of tidal disruption (i.e. $\epsilon_{\rm t} = 0$), the real tidal mass function still differs from the nonphysical virial mass function, simply as a consequence of tidal stripping. Furthermore, as already mentioned above, tidal effects are typically much stronger in the central region of the host, and much less important beyond the scale radius. The model effectively leads to a selection in concentration: more concentrated (i.e. denser) subhalos are more resilient to tidal effects, which explains why originally lighter subhalos (which formed earlier and are denser) survive more efficiently in the central regions of the host. Consequently, the mass function is strongly altered in the central parts of the host halo, and becomes much steeper than the *unevolved* one. In contrast, the mass function is very close to its initial shape in the outskirts of the host halo. All this is consistent with other works showing a spatial evolution of the mass-concentration relation [46, 87, 113]. Therefore, so long as we are not concerned with the inner subhalo population of the host halo, the relevant mass function should remain close to Eq. (B.9). This is actually the case for all targets considered in this paper, for which total luminosity of subhalos dominates over that of the host beyond the scale radius, typically.

As we have just seen, tidal stripping and tidal disruption slightly degrade the effective global unevolved subhalo mass fraction, but more importantly, they can strongly flatten the spatial distribution of subhalos in the host center. A damping of the subhalo population is even predicted in the very inner parts of the host for most of the model parameters used in our study (which is only moderately reflected by the mass fraction, which integrates subhalos over the whole host volume). The level of this flattening is obviously driven by the value taken for $\epsilon_{\rm t}$, the disruption efficiency parameter, with a population damping more severe for larger values. This is consistent with other analytical studies based on different approaches (see e.g. [101]).

References

[1] K. Griest and M. Kamionkowski, *Unitarity limits on the mass and radius of dark-matter particles*, *Phys. Rev. Lett.* **64** (Feb., 1990) 615–618.

[2] L. Bergström, *Non-baryonic dark matter: observational evidence and detection methods*, *Reports on Progress in Physics* **63** (May, 2000) 793–841, [hep-ph/0002126].

[3] T. Bringmann and C. Weniger, *Gamma ray signals from dark matter: Concepts, status and prospects*, *Physics of the Dark Universe* **1** (Nov., 2012) 194–217, [1208.5481].

[4] J. Lavalle and P. Salati, *Dark matter indirect signatures*, *Comptes Rendus Physique* **13** (July, 2012) 740–782, [1205.1004].

[5] M. Fornasa and M. A. Sánchez-Conde, *The nature of the Diffuse Gamma-Ray Background*, *Phys. Rep.* **598** (Oct., 2015) 1–58, [1502.02866].

[6] L. Bergström, *Dark matter candidates*, *New Journal of Physics* **11** (Oct., 2009) 105006, [0903.4849].

[7] J. L. Feng, *Dark Matter Candidates from Particle Physics and Methods of Detection*, *ARA&A* **48** (Sept., 2010) 495–545, [1003.0904].

[8] G. Arcadi, M. Dutra, P. Ghosh, M. Lindner, Y. Mambrini, M. Pierre et al., *The waning of the wimp? a review of models, searches, and constraints*, *Eur. Phys. J. C* **78** (Mar., 2018) 203, [1703.07364].

[9] S. Bottaro, D. Buttazzo, M. Costa, R. Franceschini, P. Panci, D. Redigolo et al., *Closing the window on wimp dark matter*, *Eur. Phys. J. C* **82** (2022) 31, [2107.09688].

[10] T. R. Slatyer, *Indirect dark matter signatures in the cosmic dark ages. I. Generalizing the bound on s -wave dark matter annihilation from Planck results*, *Phys. Rev. D* **93** (Jan., 2016) 023527, [1506.03811].

[11] T. R. Slatyer, *Les houches lectures on indirect detection of dark matter*, *arXiv e-prints* (Sept., 2021) arXiv:2109.02696, [2109.02696].

[12] J. Hisano, S. Matsumoto and M. M. Nojiri, *Explosive Dark Matter Annihilation*, *Phys. Rev. Lett.* **92** (Jan., 2004) 031303, [hep-ph/0307216].

[13] J. Hisano, S. Matsumoto, M. M. Nojiri and O. Saito, *Nonperturbative effect on dark matter annihilation and gamma ray signature from the galactic center*, *Phys. Rev. D* **71** (Mar., 2005) 063528, [hep-ph/0412403].

[14] M. Cirelli, A. Strumia and M. Tamburini, *Cosmology and astrophysics of minimal dark matter*, *Nuclear Physics B* **787** (Dec., 2007) 152–175, [0706.4071].

[15] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer and N. Weiner, *A theory of dark matter*, *Phys. Rev. D* **79** (Jan., 2009) 015014, [0810.0713].

[16] R. Iengo, *Sommerfeld enhancement: general results from field theory diagrams*, *Journal of High Energy Physics* **2009** (May, 2009) 024, [0902.0688].

[17] T. R. Slatyer, *The Sommerfeld enhancement for dark matter with an excited state*, *J. Cosmology Astropart. Phys.* **2010** (Feb., 2010) 028, [0910.5713].

[18] J. L. Feng, M. Kaplinghat and H.-B. Yu, *Sommerfeld enhancements for thermal relic dark matter*, *Phys. Rev. D* **82** (Oct., 2010) 083525, [1005.4678].

[19] S. Cassel, *Sommerfeld factor for arbitrary partial wave processes*, *Journal of Physics G Nuclear Physics* **37** (Oct., 2010) 105009, [0903.5307].

[20] K. Blum, R. Sato and T. R. Slatyer, *Self-consistent calculation of the Sommerfeld enhancement*, *J. Cosmology Astropart. Phys.* **2016** (June, 2016) 021, [1603.01383].

[21] T. Hambye and L. Vanderheyden, *Minimal self-interacting dark matter models with light mediator*, *J. Cosmology Astropart. Phys.* **2020** (May, 2020) 001, [1912.11708].

[22] P. J. E. Peebles, *Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations*, *ApJ* **263** (Dec., 1982) L1–L5.

[23] A. M. Green, S. Hofmann and D. J. Schwarz, *The power spectrum of SUSY-CDM on subgalactic scales*, *MNRAS* **353** (Sept., 2004) L23–L27, [astro-ph/0309621].

[24] V. S. Berezinsky, V. I. Dokuchaev and Y. N. Eroshenko, *Small-scale clumps of dark matter*, *Physics Uspekhi* **57** (Jan., 2014) 1–36, [1405.2204].

[25] J. Zavala and C. S. Frenk, *Dark matter haloes and subhaloes*, *Galaxies* **7** (Sept., 2019) 81, [1907.11775].

[26] M. Kuhlen and D. Malyshev, *Atic, pamela, hess, and fermi data and nearby dark matter subhalos*, *Phys. Rev. D* **79** (June, 2009) 123517, [0904.3378].

[27] J. Bovy, *Substructure boosts to dark matter annihilation from sommerfeld enhancement*, *Phys. Rev. D* **79** (Apr., 2009) 083539, [0903.0413].

[28] M. Lattanzi and J. Silk, *Can the WIMP annihilation boost factor be boosted by the Sommerfeld enhancement?*, *Phys. Rev. D* **79** (Apr., 2009) 083523, [0812.0360].

[29] L. Pieri, M. Lattanzi and J. Silk, *Constraining the dark matter annihilation cross-section with Cherenkov telescope observations of dwarf galaxies*, *MNRAS* **399** (Nov., 2009) 2033–2040, [0902.4330].

[30] T. R. Slatyer, N. Toro and N. Weiner, *Sommerfeld-enhanced annihilation in dark matter substructure: Consequences for constraints on cosmic-ray excesses*, *Phys. Rev. D* **86** (Oct., 2012) 083534, [1107.3546].

[31] J. Zavala and N. Afshordi, *Clustering in the phase space of dark matter haloes - II. Stable clustering and dark matter annihilation*, *MNRAS* **441** (June, 2014) 1329–1339, [1311.3296].

[32] K. K. Boddy, J. Kumar, J. Runburg and L. E. Strigari, *Angular distribution of gamma-ray emission from velocity-dependent dark matter annihilation in subhalos*, *Phys. Rev. D* **100** (Sept., 2019) 063019, [1905.03431].

[33] J. Runburg, E. J. Baxter and J. Kumar, *Constraining dark matter microphysics with the annihilation signal from subhalos*, *arXiv e-prints* (June, 2021) arXiv:2106.10399, [2106.10399].

[34] K. K. Boddy, J. Kumar, L. E. Strigari and M.-Y. Wang, *Sommerfeld-enhanced j -factors for dwarf spheroidal galaxies*, *Phys. Rev. D* **95** (June, 2017) 123008, [1702.00408].

[35] K. K. Boddy, J. Kumar and L. E. Strigari, *The effective j-factor of the galactic center for velocity-dependent dark matter annihilation*, *Phys. Rev. D* **98** (May, 2018) 063012, [1805.08379].

[36] B. Boucher, J. Kumar, V. B. Le and J. Runburg, *j-factors for velocity-dependent dark matter*, *arXiv e-prints* (Oct., 2021) arXiv:2110.09653, [2110.09653].

[37] S. Ando and K. Ishiwata, *Sommerfeld-enhanced dark matter searches with dwarf spheroidal galaxies*, *arXiv e-prints* (Mar., 2021) arXiv:2103.01446, [2103.01446].

[38] E. Board, N. Bozorgnia, L. E. Strigari, R. J. J. Grand, A. Fattahi, C. S. Frenk et al., *Velocity-dependent J-factors for annihilation radiation from cosmological simulations*, *arXiv e-prints* (Jan., 2021) arXiv:2101.06284, [2101.06284].

[39] T. Lacroix, G. Facchinetti, J. Pérez-Romero, M. Stref, J. Lavalle, D. Maurin et al., *Classification of gamma-ray targets for velocity-dependent and subhalo-boosted dark-matter annihilation*, *arXiv e-prints* (Mar., 2022) arXiv:2203.16440, [2203.16440].

[40] J. Silk and A. Stebbins, *Clumpy cold dark matter*, *ApJ* **411** (July, 1993) 439–449.

[41] L. Bergström, J. Edsjö, P. Gondolo and P. Ullio, *Clumpy neutralino dark matter*, *Phys. Rev. D* **59** (Feb., 1999) 043506, [astro-ph/9806072].

[42] J. Lavalle, Q. Yuan, D. Maurin and X.-J. Bi, *Full calculation of clumpiness boost factors for antimatter cosmic rays in the light of* ΛCDM *N-body simulation results. Abandoning hope in clumpiness enhancement?*, *A&A* **479** (Feb., 2008) 427–452, [0709.3634].

[43] S. Ando, T. Ishiyama and N. Hiroshima, *Halo substructure boosts to the signatures of dark matter annihilation*, *Galaxies* **7** (July, 2019) 68, [1903.11427].

[44] A. Sommerfeld, *Über die Beugung und Bremsung der Elektronen*, *Annalen der Physik* **403** (Jan., 1931) 257–330.

[45] L. Bergström, P. Ullio and J. H. Buckley, *Observability of* γ *rays from dark matter neutralino annihilations in the Milky Way halo*, *Astroparticle Physics* **9** (Aug., 1998) 137–162, [astro-ph/9712318].

[46] M. Kuhlen, J. Diemand and P. Madau, *The Dark Matter Annihilation Signal from Galactic Substructure: Predictions for GLAST*, *ApJ* **686** (Oct., 2008) 262–278, [0805.4416].

[47] A. S. Eddington, *The distribution of stars in globular clusters*, *MNRAS* **76** (May, 1916) 572–585.

[48] J. Binney and S. Tremaine, *Galactic Dynamics.* Princeton series in astrophysics. Princeton University Press, Princeton, NJ USA, 2008., 2nd ed ed., Jan., 2008.

[49] T. Lacroix, M. Stref and J. Lavalle, *Anatomy of eddington-like inversion methods in the context of dark matter searches*, *J. Cosmology Astropart. Phys.* **09** (Sept., 2018) 040, [1805.02403].

[50] T. Lacroix, A. Núñez-Castiñeyra, M. Stref, J. Lavalle and E. Nezri, *Predicting the dark matter velocity distribution in galactic structures: tests against hydrodynamic cosmological simulations*, *J. Cosmology Astropart. Phys.* **2020** (Oct., 2020) 031, [2005.03955].

[51] M. Petač, J. Lavalle, A. Núñez-Castiñeyra and E. Nezri, *Testing the predictions of axisymmetric distribution functions of galactic dark matter with hydrodynamical simulations*, *J. Cosmology Astropart. Phys.* **2021** (Aug., 2021) 031, [2106.01314].

[52] J. R. Bond, A. S. Szalay and M. S. Turner, *Formation of galaxies in a gravitino-dominated universe*, *Phys. Rev. Lett.* **48** (June, 1982) 1636–1639.

[53] G. R. Blumenthal, S. M. Faber, J. R. Primack and M. J. Rees, *Formation of galaxies and large-scale structure with cold dark matter*, *Nature* **311** (Oct., 1984) 517–525.

[54] S. Hofmann, D. J. Schwarz and H. Stöcker, *Damping scales of neutralino cold dark matter*, *Phys. Rev. D* **64** (Oct., 2001) 083507, [astro-ph/0104173].

[55] E. Bertschinger, *Effects of cold dark matter decoupling and pair annihilation on cosmological perturbations*, *Phys. Rev. D* **74** (Sept., 2006) 063509, [astro-ph/0607319].

[56] T. Bringmann and S. Hofmann, *Thermal decoupling of WIMPs from first principles*, *J. Cosmology Astropart. Phys.* **4** (Apr., 2007) 16, [hep-ph/0612238].

[57] P. Ullio, L. Bergström, J. Edsjö and C. Lacey, *Cosmological dark matter annihilations into* γ *rays: A closer look*, *Phys. Rev. D* **66** (Dec., 2002) 123502, [astro-ph/0207125].

[58] J. Lavalle, J. Pochon, P. Salati and R. Taillet, *Clumpiness of dark matter and the positron annihilation signal*, *A&A* **462** (Feb., 2007) 827–840, [arXiv:astro-ph/0603796].

[59] L. G. van den Aarssen, T. Bringmann and Y. C. Goedecke, *Thermal decoupling and the smallest subhalo mass in dark matter models with Sommerfeld-enhanced annihilation rates*, *Phys. Rev. D* **85** (June, 2012) 123512, [1202.5456].

[60] M. Stref and J. Lavalle, *Modeling dark matter subhalos in a constrained galaxy: Global mass and boosted annihilation profiles*, *Phys. Rev. D* **95** (Mar., 2017) 063003, [1610.02233].

[61] F. Calore, M. Hütten and M. Stref, *Gamma-ray sensitivity to dark matter subhalo modelling at high latitudes*, *Galaxies* **7** (Nov, 2019) 90, [1910.13722].

[62] M. Stref, T. Lacroix and J. Lavalle, *Remnants of galactic subhalos and their impact on indirect dark-matter searches*, *Galaxies* **7** (June, 2019) 65, [1905.02008].

[63] M. Hütten, M. Stref, C. Combet, J. Lavalle and D. Maurin, *γ-ray and ν searches for dark-matter subhalos in the milky way with a baryonic potential*, *Galaxies* **7** (May, 2019) 60, [1904.10935].

[64] G. Facchinetti, J. Lavalle and M. Stref, *Statistics for dark matter subhalo searches in gamma rays from a kinematically constrained population model. i: Fermi-lat-like telescopes*, *arXiv e-prints* (July, 2020) arXiv:2007.10392, [2007.10392].

[65] M. Boudaud, T. Lacroix, M. Stref and J. Lavalle, *Robust cosmic-ray constraints on p-wave annihilating mev dark matter*, *Phys. Rev. D* **99** (Mar., 2019) 061302, [1810.01680].

[66] J. Lopes, T. Lacroix and I. Lopes, *Towards a more rigorous treatment of uncertainties on the velocity distribution of dark matter particles for capture in stars*, *J. Cosmology Astropart. Phys.* **2021** (Jan., 2021) 073, [2007.15927].

[67] J. R. Bond, S. Cole, G. Efstathiou and N. Kaiser, *Excursion set mass functions for hierarchical Gaussian fluctuations*, *ApJ* **379** (Oct., 1991) 440–460.

[68] C. Lacey and S. Cole, *Merger rates in hierarchical models of galaxy formation*, *MNRAS* **262** (June, 1993) 627–649.

[69] F. C. van den Bosch, *The universal mass accretion history of cold dark matter haloes*, *MNRAS* **331** (Mar., 2002) 98–110, [astro-ph/0105158].

[70] F. Jiang and F. C. van den Bosch, *Statistics of Dark Matter Substructure: I. Model and Universal Fitting Functions*, *MNRAS* **458** (May, 2016) 2848–2869, [1403.6827].

[71] T. Ishiyama and S. Ando, *The abundance and structure of subhaloes near the free streaming scale and their impact on indirect dark matter searches*, *MNRAS* **492** (Mar., 2020) 3662–3671, [1907.03642].

[72] H. Zhao, *Analytical models for galactic nuclei*, *MNRAS* **278** (Jan., 1996) 488–496, [astro-ph/9509122].

[73] J. F. Navarro, C. S. Frenk and S. D. M. White, *The Structure of Cold Dark Matter Halos*, *ApJ* **462** (May, 1996) 563, [astro-ph/9508025].

[74] J. F. Navarro, C. S. Frenk and S. D. M. White, *A Universal Density Profile from Hierarchical Clustering*, *ApJ* **490** (Dec., 1997) 493–508, [astro-ph/9611107].

[75] J. Einasto, *On the Construction of a Composite Model for the Galaxy and on the Determination of the System of Galactic Parameters*, *Trudy Astrofizicheskogo Instituta Alma-Ata* **5** (1965) 87–100.

[76] J. F. Navarro, E. Hayashi, C. Power, A. R. Jenkins, C. S. Frenk, S. D. M. White et al., *The inner structure of ΛCDM haloes - III. Universality and asymptotic slopes*, *MNRAS* **349** (Apr., 2004) 1039–1051, [astro-ph/0311231].

[77] D. Merritt, A. W. Graham, B. Moore, J. Diemand and B. Terzić, *Empirical Models for Dark Matter Halos. I. Nonparametric Construction of Density Profiles and Comparison with Parametric Models*, *AJ* **132** (Dec., 2006) 2685–2700, [astro-ph/0509417].

[78] J. F. Navarro, A. Ludlow, V. Springel, J. Wang, M. Vogelsberger, S. D. M. White et al., *The diversity and similarity of simulated cold dark matter haloes*, *MNRAS* **402** (Feb., 2010) 21–34, [0810.1522].

[79] J. I. Read, M. G. Walker and P. Steger, *The case for a cold dark matter cusp in draco*, *MNRAS* **481** (Nov., 2018) 860–877, [1805.06934].

[80] P. J. McMillan, *The mass distribution and gravitational potential of the Milky Way*, *MNRAS* **465** (Feb., 2017) 76–94, [1608.00971].

[81] S. Ettori, V. Ghirardini, D. Eckert, E. Pointecouteau, F. Gastaldello, M. Sereno et al., *Hydrostatic mass profiles in x-cop galaxy clusters*, *A&A* **621** (Jan., 2019) A39, [1805.00035].

[82] W. J. G. de Blok, *The Core-Cusp Problem*, *Advances in Astronomy* **2010** (2010) 5, [0910.3538].

[83] J. S. Bullock and M. Boylan-Kolchin, *Small-Scale Challenges to the ΛCDM Paradigm*, *ARA&A* **55** (Aug., 2017) , [1707.04256].

[84] J. Diemand, M. Kuhlen and P. Madau, *Formation and Evolution of Galaxy Dark Matter Halos and Their Substructure*, *ApJ* **667** (Oct., 2007) 859–877, [astro-ph/0703337].

[85] V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi et al., *The Aquarius Project: the subhaloes of galactic haloes*, *MNRAS* **391** (Dec., 2008) 1685–1711, [0809.0898].

[86] A. A. Klypin, S. Trujillo-Gomez and J. Primack, *Dark matter halos in the standard cosmological model: Results from the bolshoi simulation*, *ApJ* **740** (Oct., 2011) 102, [1002.3660].

[87] L. Pieri, J. Lavalle, G. Bertone and E. Branchini, *Implications of High-Resolution Simulations on Indirect Dark Matter Searches*, *Phys. Rev. D* **83** (Jan., 2011) 023518, [0908.0195].

[88] A. Charbonnier, C. Combet and D. Maurin, *CLUMPY: A code for γ-ray signals from dark matter structures*, *Computer Physics Communications* **183** (Mar., 2012) 656–668, [1201.4728].

[89] F. Calore, V. De Romeri, M. Di Mauro, F. Donato and F. Marinacci, *Realistic estimation for the detectability of dark matter subhalos using fermi-lat catalogs*, *Phys. Rev. D* **96** (Sept., 2017) 063009, [1611.03503].

[90] M. Di Mauro, M. Stref and F. Calore, *Investigating the detection of dark matter subhalos as extended sources with fermi-lat*, *arXiv e-prints* (July, 2020) arXiv:2007.08535, [2007.08535].

[91] M. A. Sánchez-Conde and F. Prada, *The flattening of the concentration-mass relation towards low halo masses and its implications for the annihilation signal boost*, *MNRAS* **442** (Aug., 2014) 2271–2277, [1312.1729].

[92] G. Facchinetti, *Analytical study of particle dark matter structuring on small scales and implications for dark matter searches*, Ph.D. thesis, Université de Montpellier, 2021.

[93] G. Facchinetti et al., *Analytical dark matter subhalo mass functions*, *in preparation* (2022) .

[94] M. Sten Delos, *Evolution of dark matter microhalos through stellar encounters*, *Phys. Rev. D* **100** (Oct., 2019) 083529, [1907.13133].

[95] R. Errani and J. Peñarrubia, *Can tides disrupt cold dark matter subhaloes?*, *MNRAS* **491** (Feb., 2020) 4591–4601, [1906.01642].

[96] R. Errani and J. F. Navarro, *The asymptotic tidal remnants of cold dark matter subhalos*, 2011.07077.

[97] M. D. Weinberg, *Adiabatic invariants in stellar dynamics. 1: Basic concepts*, *AJ* **108** (Oct., 1994) 1398–1402, [astro-ph/9404015].

[98] O. Y. Gnedin, L. Hernquist and J. P. Ostriker, *Tidal Shocking by Extended Mass Distributions*, *ApJ* **514** (Mar., 1999) 109–118.

[99] E. Hayashi, J. F. Navarro, J. E. Taylor, J. Stadel and T. Quinn, *The Structural Evolution of Substructure*, *ApJ* **584** (Feb., 2003) 541–558, [astro-ph/0203004].

[100] F. C. van den Bosch and G. Ogiya, *Dark matter substructure in numerical simulations: A tale of discreteness noise, runaway instabilities, and artificial disruption*, *MNRAS* **475** (Apr., 2018) 4066–4087, [1801.05427].

[101] S. B. Green, F. C. van den Bosch and F. Jiang, *The tidal evolution of dark matter substructure - ii. the impact of artificial disruption on subhalo mass functions and radial profiles*, *MNRAS* (Mar., 2021) , [2103.01227].

[102] J. Diemand, B. Moore and J. Stadel, *Velocity and spatial biases in cold dark matter subhalo distributions*, *MNRAS* **352** (Aug., 2004) 535–546, [astro-ph/0402160].

[103] J. Han, S. Cole, C. S. Frenk and Y. Jing, *A unified model for the spatial and mass distribution of subhaloes*, *MNRAS* **457** (Apr., 2016) 1208–1223, [1509.02175].

[104] F. C. van den Bosch, G. Ogiya, O. Hahn and A. Burkert, *Disruption of dark matter substructure: fact or fiction?*, *MNRAS* **474** (Mar., 2018) 3043–3066, [1711.05276].

[105] A. V. Macciò, A. A. Dutton and F. C. van den Bosch, *Concentration, spin and shape of dark matter haloes as a function of the cosmological model: WMAP1, WMAP3 and WMAP5 results*, *MNRAS* **391** (Dec., 2008) 1940–1954, [0805.1926].

[106] F. Prada, A. A. Klypin, A. J. Cuesta, J. E. Betancort-Rijo and J. Primack, *Halo concentrations in the standard* Λ *cold dark matter cosmology*, *MNRAS* **423** (July, 2012) 3018–3030, [1104.5130].

[107] A. A. Dutton and A. V. Macciò, *Cold dark matter haloes in the Planck era: evolution of structural parameters for Einasto and NFW profiles*, *MNRAS* **441** (July, 2014) 3359–3374, [1402.7073].

[108] W. H. Press and P. Schechter, *Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation*, *ApJ* **187** (Feb., 1974) 425–438.

[109] Planck Collaboration, *Planck 2018 results. vi. cosmological parameters*, *A&A* **641** (Sept., 2020) A6, [1807.06209].

[110] F. C. van den Bosch and F. Jiang, *Statistics of Dark Matter Substructure: II. Comparison of Model with Simulation Results*, *MNRAS* **458** (May, 2016) 2870–2884, [1403.6835].

[111] S. Cole, C. G. Lacey, C. M. Baugh and C. S. Frenk, *Hierarchical galaxy formation*, *MNRAS* **319** (Nov., 2000) 168–204, [astro-ph/0007281].

[112] C. Giocoli, L. Pieri and G. Tormen, *Analytical approach to subhalo population in dark matter haloes*, *MNRAS* **387** (June, 2008) 689–697, [0712.1476].

[113] Á. Moliné, M. A. Sánchez-Conde, S. Palomares-Ruiz and F. Prada, *Characterization of subhalo structural properties and implications for dark matter annihilation signals*, *MNRAS* **466** (Apr., 2017) 4974–4990, [1603.04057].