Dark Matter Searches in the Crescendo of the 21st Century Experiments

Leung, (John) Shing Chau

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in
Physics

The Brown University
May 2020

This dissertation by John Shing Chau Leung is accepted in its present form by the Department of Physics as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date _____ _____

Prof. JiJi Fan, Advisor

Recommended to the Graduate Council

Date _____ _____

Prof. Ian Dell'Antonio, Reader

Date _____ _____

Prof. Savvas Koushiappas, Reader

Approved by the Graduate Council

Date _____ _____

Prof. Andrew Campbell,
Dean of the Graduate School

Vitae

John Shing Chau Leung obtained his Bachelor of Science in Physics from the Chinese University of Hong Kong in May, 2012. During his undergraduate years, John worked on cold atom physics with Chi Kwong Law and on particle physics experiment with Ming Chung Chu. He also participated in high energy experiment in the CMS experiment of CERN as a summer student, advised by Armando Lanaro. John obtained his Master of Philosophy in Physics from the Chinese University of Hong Kong in May, 2014. His thesis was on detector physics advised by Ming Chung Chu, and with collaboration with the ATLAS experiment of CERN advised by Maurice Sciveres. John joined the Brown University in August, 2014 and carried out research with JiJi Fan.

Acknowledgments

I would like to thank my advisor JiJi Fan, who gave me the dream opportunity to pursue physics research. I also thank Savvas Koushiappas and Ian Dell'Antonio for being my thesis committee members and, in general, being good friends and mentors. I thank Antal Jevicki for his care about me during my graduate study. And I thank Stephon Alexander for providing me research opportunity on cosmology. I also thank Evan McDonough for the collaborative research and helping me out with my postdoctoral position search.

I am very grateful for Robert Sims for working together to develop several insightful projects. His company across the years of my doctoral study was also invaluable. I also give my thanks to Adrian Lam for his support and useful discussion about machine learning. I thank my officemates Jatan Buch and Kyriakos Vattis for the numerous delightful and informative physics and statistics discussions.

Lastly, I would like to thank my undergraduate and master mentor Ming Chung Chu for his care and effort to help me develop my academic career. His dedication to students and their careers is something that I will not forget.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Let us start by brushing over dark matter (DM) in the historical context. The history of the conceptualization of DM to explain astrophysical observations dates back at least to the 18th century when Johann Heinrich Lambert attempted to explain the then-conjectured disk-like, rotational Milky Way structure[1] using "immense *dark* bodies" [8, 9]. In the framework of this Lambert's hypothesis, Newcomb [10] qualitatively investigated the possibility to explain the boundedness of a high proper motion star, Groombridge 1830 [11], which exceeded the Milky Way's escape velocity calculated from the visible stars back then. At the the dawn of the 20th century, Lord Kelvin [12] purposed the existence of dark celestial bodies when estimating the amount of matter around our sun, a sense that was close to the modern sense of dark matter. In one of the analyses of matter in the Milky Way, Poincaré [13] described a matière obscure (as opposed to matière brillante), which was later translated by Watson [14] who coined the term "*dark matter*".

Öpik [15] and Kapteyn [16] were the first to use vertical dynamics of stars to study the amount of gravitational mass, and thus investigated the possibility of the presence of dark matter in the Milky Way. Jeans [17], Lindblad [18] and Oort [19] similarly conjectured the need of DM in the galactic plane to account for the total mass in the solar neighborhood in conjunction with the luminous bodies. The language of that time-period was the prediction

[1] See also Lambert's contemporaries' attempts to describe the Milk Way [6, 7].

and measurement of stellar vertical acceleration, $K(z)$, that obeys

$$w = \sqrt{w_0^2 + 2\int_0^z K(z)\mathrm{d}z}, \tag{1.1}$$

where w is the z-component of the velocity of a star and w_0 is w at the galactic plane, $z = 0$. Eq. (1.1) can easily be derived by considering the acceleration per unit mass as the derivative on galactic gravitational potential

$$K(z) = -\frac{\mathrm{d}\Phi}{\mathrm{d}z} \tag{1.2}$$

and applying conservation of energy. We will apply the modern variation of Eq. (1.1) to study dissipative dark matter in chapter 3. The first concrete evidence of dark matter was from the observation of the extraordinary large velocity dispersion of the Coma Cluster [19], where Zwicky [20] pointed it out as a strong evidence for DM.

Of course the notion of DM developed in great leaps since the first half of the 20th century, departing from the thoughts of faint stars, dark stars, planets and nebulae as the source of DM. From the second half of the 20th century, starting from Rubin and Ford's pioneering work on Andromeda's rotational curve [20], evidence emerged strongly for the presence of an invisible matter. Evidence further piled up that DM indeed played an important role on the universe, from more observations on galaxy rotational curves [21] to cosmic microwave background [22,23], and DM has a more mysterious origin than other mundane astrophysical objects. For readers interested in the history of DM research, a more exhaustive review is available in Ref. [24].

Nowadays, physicists measure that DM occupies $\sim$26% of the total energy budget and $\sim$80% of total matter budget of the universe [23]. We often consider DM having a particle physics origin. While a very popular DM candidate in the first score of years of the 21st century is weakly interacting massive particles (WIMPs), among which a popular class of candidates is supersymmetric particles [25], there are plenty of other conjectures such as DM as sterile neutrinos [26], axions [27, 28], superheavy (WIMPzillas) DMs [29, 30], just to name a few. Of course, there are still conjectures of DM as astrophysical objects [31–35] and composite bound states [36–40].

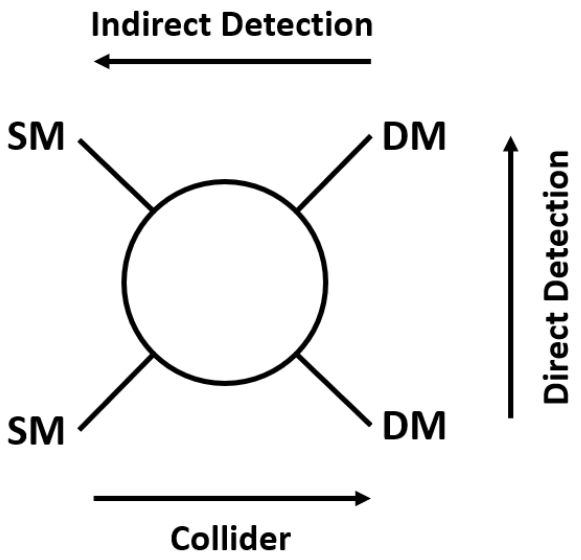


Figure 1.1: Methods of dark matter detection.

The common methods to detect dark matter are described by the famous diagram in Fig. 1.1. The limbs in the figure are incoming or outgoing particles, and the bubble in the middle represents some Standard Model-Dark Matter interaction we do not know yet. From the left to right is the indirect detection (ID) channel. Dark matter inhabits the Milky Way or other galaxies can annihilate among themselves to give off Standard Model (SM) signatures, such as photons, neutrinos and cosmic rays. While in this thesis we would not address the ID method, readers interested can refer to the reviews in Ref. [41, 42]. From the right to left is the collider production channel, where two SM particles are put in an accelerator and collide together to produce new particles, among which can be DM or DM-related particles. A review of collider terminology can be found in Ref. [43] for readers not familiar with the physics at a collider. In Chapter 2, we study new strategies to identity SUSY models in a future collider scenario. These models produce a stable lightest superpartner (LSP) that can act as a DM candidate. Lastly, from bottom to top (or top to bottom) is the direct detection (DD) channel. DM in the solar neighborhood of the Milk Way often passes through the Earth, possibly interacts with earthly materials. If a detector is built sensitive enough, it can detect the recoil of the detector material off such a collision event. Readers

interested in the method of DM direct detection can refer to reviews in Ref. [44, 45]. In chapter 4, we discuss our work on how DM direct detection interplays with astrophysics concerning local DM distribution. Another interpretation of the vertical arrow in Fig. 1.1 is that we replace the interaction bubble with a graviton as a force mediator. Gravity is the only known interaction DM participates in, both among themselves and to the SM. By observing kinematics of stars under gravitation potential of the Milky Way or other galaxies, we can infer how much DM there are and how they are distributed. In this sense, this kind of detection stands somewhere between DD in the interaction sense, and ID in the sense that it uses astronomical observations. Some methods are discussed in the reviews in Ref. [3, 46]. In Chapter 3, we will use this technique to study and constrain dissipative dark matter (DDM). Of course, these are not the only ways physicists can study DM. Other ways include observing galaxy rotational curves, large scale structures and the cosmic microwave background (CMB), etc. For more thorough reviews of DM physics, our evidence of it and the handles to study it, readers can refer to Ref. [47–50].

This thesis is organized as follow: In Chapter 2, we discuss a future proton-proton collider with center of mass energy around 100 TeV [51–53]. In this work we will study its sensitivity to two simplified SUSY models involving supersymmetric tops (stops): pair production of stops that decay to tops or bottoms and higgsinos; and stops that are either pair produced or produced together with a gluino and then cascade down through gluinos to the lightest superpartner (LSP). We devise new strategies to identify super-boosted tops or bottoms with transverse momentum of order TeV. In Chapter 3, we use stellar kinematics from the latest Gaia data release [54, 55] (DR2) to measure the local dark matter density and DM local substructure. By doing so we can set constraints on the surface density and scale height of a thin dark disk aligned with the baryonic disk and formed due to DM self-interaction. In Chapter 4, we study how the empirical DM velocity distribution inferred from Gaia-Sausage [56–61], a dominant substructure in the solar neighborhood, affects the interpretation of DD data. We survey different classes of operators in the non-relativistic effective field theory that could arise from several relativistic benchmark models and emphasize that the Gaia velocity distribution modifies both the total number of events as well as

the shape of the differential recoil spectra, the two primary observables in DD experiments. In Chapter 5, we summarize what we have learnt from this thesis.

Chapter 2

Jet Observables and Stops at 100 TeV Collider

2.1 Introduction

In this chapter, we discuss the prospect of identifying supersymmetry (SUSY) using a future proton-proton 100 TeV collider with jet substructure techniques [62,63]. In this introduction, we motivate the study of SUSY model.

It is known that the Standard Model (SM) Higgs scalar mass is unstable to quantum corrections, which in general are proportional to the highest mass scale of the system. One way to resolve the problem is to introduce weak scale SUSY to cancel out the quantum corrections, thus acquire a "natural Higgs". One of the most important motivations for SUSY is that for some SUSY models, there is an symmetry that governs the change in number of SUSY particles in an interaction. It is a $\mathbb{Z}_2$ symmetry known as the R-parity. The consequence is that lightest supersymmetric particle (LSP) will be stable. And it is a good candidate for weakly interacting massive particle (WIMP) DM[1] .

One other reason to motivate weak scale SUSY is that it also helps gauge coupling unification. A review of SUSY can be found in Ref. [25].

[1]The LSP of a weak scale SUSY happens to be an ideal DM candidate because of the WIMP miracle. We briefly demonstrate the reason in Appendix A

This chapter is organized as follows: In Section 2.2, we motivate the study of 100 TeV collider and introduce the basics of SUSY models. In Section 2.3, we present details of the two stop simplified models. In Section 2.4, we discuss the jet finding algorithms and demonstrate the discriminating powers of several jet observables we used in the analyses. In Section 2.5, we present the analysis for the $\tilde{t} - \tilde{H}$ model and its results. In Section 2.6, we present the analysis for the $\tilde{t} - \tilde{g} - \tilde{\chi}_1^0$ model and its results. We will conclude and discuss possible future directions in Section 2.7. This chapter is based on the work with JiJi Fan and Preit Jaiswal in Ref. [64].

2.2 Basic of SUSY in a 100 TeV collider

Collider experiments have been the most powerful probe to reveal the nature of the smallest possible distance scale in particle physics. While currently the Large Hadron Collider (LHC) is still busy exploring the TeV scale, there has been a growing effort in planning for future hadron colliders to take the baton from the LHC in hunting for new physics beyond the Standard Model (SM) [51–53, 65–71]. So far the most discussed future hadron collider scenario is a circular 100 TeV proton-proton machine. It has been demonstrated that such a machine can push the testable energy frontier by roughly one order of magnitude and could discover colored particles with masses near 10 TeV [72–80] as well as electroweak particles with masses near 1 TeV [81–92].

One of the most-motivated new physics targets at hadron colliders is the top partners, for example, stops in the supersymmetric (SUSY) scenarios. The mass scale of stops is an indication of the fine-tuning level of electroweak symmetry breaking in SUSY [93–100]. So far only the simplest possible stop decay, $\tilde{t} \to t + \tilde{\chi}_1^0$ with $\tilde{\chi}_1^0$ being the lightest neutralino has been studied at a 100 TeV collider [76].

In this chapter, we will investigate reach of a 100 TeV collider for stops in two new stop simplified models with more complicated stop decay chains and final state topologies. In the stop-higgsino model[2] ($\tilde{t} - \tilde{H}$ model), the higgsino multiplet is at the bottom of the SUSY

[2]This model is also considered in Ref. [101] in the context of the LHC.

spectrum. Right-handed stops will decay to both neutral and charged higgsinos, which are nearly degenerate in mass, with about equal probabilities. In the $\tilde{t}-\tilde{g}-\tilde{\chi}_1^0$ model, the gluino is lighter than the stops. The stops will cascade down to the the lightest neutralino (which we take to be bino) through the gluino and produce multiple tops. In this case, stop-gluino associated production could be as important as stop pair production.

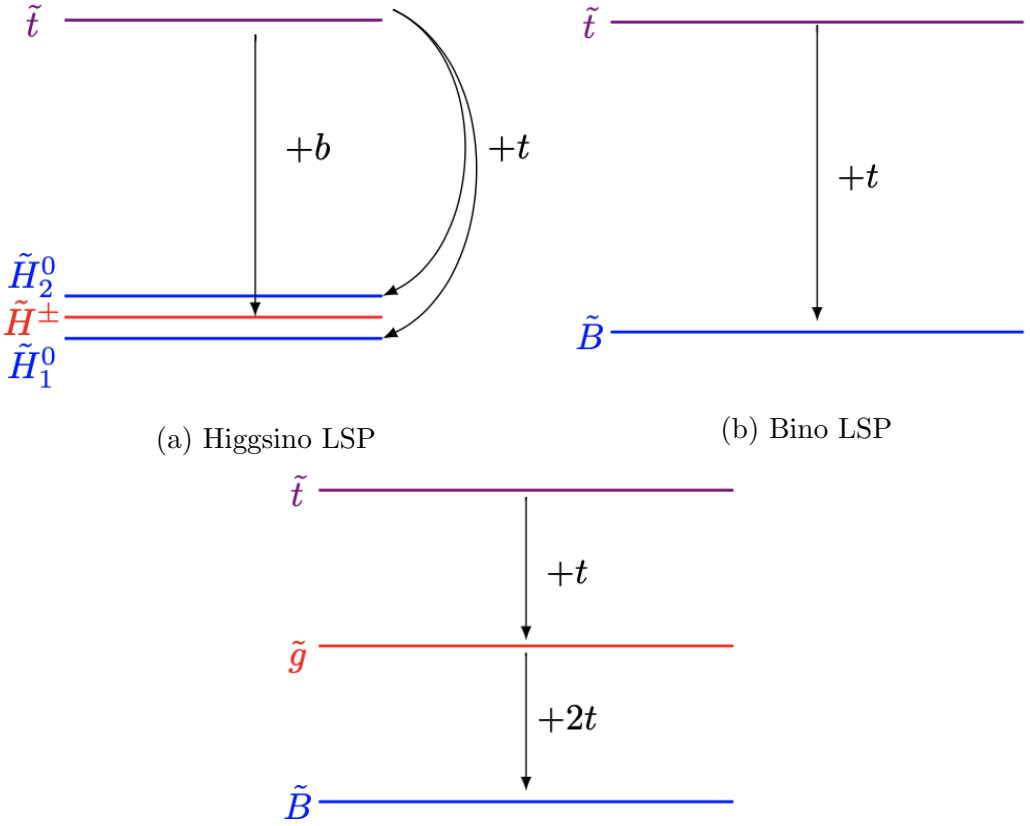
In the stop searches, one generic challenge is that SM particles, in particular, tops produced from decays of the heavy stops would be hyper-boosted with transverse momentum of order TeV and above. Their subsequent decay products would be collimated into a small cone with angular size comparable to or even smaller than a calorimeter cell. This makes the standard tagging procedure used at the LHC not directly applicable. In Ref. [76], it is suggested that leptonic-decaying tops could be identified by tagging a hard muon inside the jet at a 100 TeV collider. To study the more complicated stop decay topologies, we need to go beyond the simple muon tagging strategy and tag hadronic-decaying tops to improve the reach. We will develop boosted top and b jet tagging strategies based on several jet observables such as the track-based observables discussed in Ref. [102] to suppress both the SM and SUSY backgrounds.

2.3 Simplified Models

We consider two new simplified models: $\tilde{t}-\tilde{H}$ and $\tilde{t}-\tilde{g}-\tilde{\chi}_1^0$, which will be described in detail below. For simplicity, we only consider right-handed stops in the simplified models.

2.3.1 $\tilde{t}-\tilde{H}$ Simplified Model

In the $\tilde{t}-\tilde{H}$ simplified model, the higgsino multiplet is at the bottom of the SUSY mass spectrum and O(TeV) lighter than the stops while the remaining SUSY particles are assumed to be decoupled. The neutral and charged higgsino masses are nearly degenerate, separated by only O(GeV), with the neutral higgsino $\tilde{H}_1^0$ being the LSP (fig. 2.1a). In addition to studying the reach of $\tilde{t}-\tilde{H}$ model at a 100 TeV collider, we will also discuss how to distinguish it from the simplest stop simplified model with bino being the LSP (fig. 2.1b).

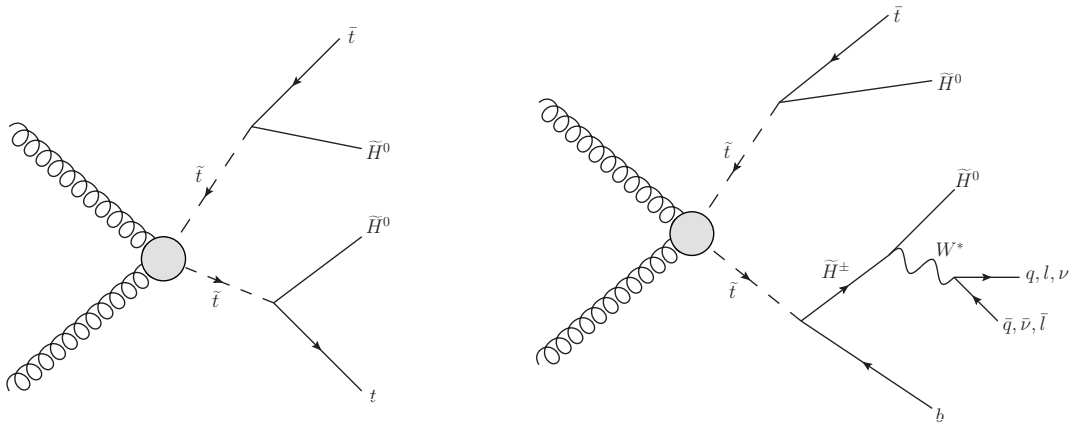


(a) Higgsino LSP

(b) Bino LSP

(c) Bino LSP with gluino NLSP

Figure 2.1: Stop simplified models.



(a) Stops decaying to $\tilde{H}^0$

(b) Stops decaying to either $\tilde{H}^0$ or $\tilde{H}^{\pm}$

Figure 2.2: Sample Feynman diagrams for signal in the $\tilde{t} - \tilde{H}$ simplified model.

There are two decay channels for stops in the $\tilde{t} - \tilde{H}$ model, each of which is equally likely:

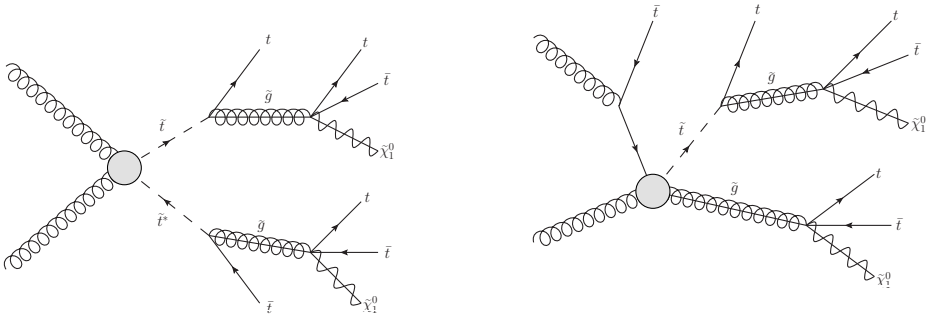
- The first channel is the stop decaying to neutral higgsinos, i.e. $\tilde{t} \to t\tilde{H}_2^0 \to tZ^* + \tilde{H}_1^0$, or $\tilde{t} \to t + \tilde{H}_1^0$ (fig. 2.2a). The decay emits a boosted top and the LSP, which may also be accompanied by soft particles if the stop decays to $\tilde{H}_2^0$ first. The particles from off-shell Z^* decays are soft, $E \sim O(\text{GeV})$, making their measurement difficult at a hadron collider. We do not consider tagging them in this piece of work.

- The other stop decay channel is $\tilde{t} \to b\tilde{H}^{\pm} \to bW^* + \tilde{H}_1^0$ (fig. 2.2b). The $\tilde{H}^{\pm}$ from stop decay promptly decays to the LSP and an off-shell W^*. Similar to the previous decay channel, SM particles resulting from W^* decay are too soft to be tagged.

The signal events will then be a mixture of b's and t's accompanied by missing energy. In comparison, in the $\tilde{t} - \tilde{B}$ simplified model, $\tilde{t} \to t\tilde{B}$ and the signal events contain only boosted t's.

2.3.2 $\tilde{t} - \tilde{g} - \tilde{\chi}_1^0$ Simplified Model

In this simplified model, we assume the three lightest SUSY particles to be stops, gluino and bino (LSP) while the remaining SUSY particles are decoupled (fig. 2.1c). Similar simplified



(a) Stop-pair production

(b) Stop-gluino associated production

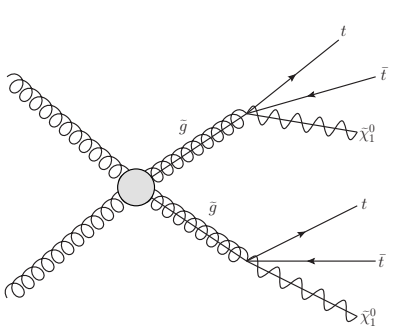
Figure 2.3: Sample Feynman diagrams for signal in the $\tilde{t} - \tilde{g} - \tilde{\chi}_1^0$ simplified model.

models have been considered before in the literature for future collider searches with one major difference: previous studies assume a mass hierarchy between stops and gluino so that one of them can be decoupled from the other [76]. In this study, however, we assume that stops and gluinos are both $\mathcal{O}(1 - 10\,\text{TeV})$ so that they can not be decoupled. We further assume that gluinos are lighter than stops so that the relevant decay channels are $\tilde{t} \to \tilde{g}t$ and $\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$ (fig. 2.3). Henceforth we refer to this simplified model as the $\tilde{t} - \tilde{g} - \tilde{\chi}_1^0$ model. Although this simplified model was considered in [74], the focus of that study was to estimate gluino reach at future colliders. Our goal instead is to estimate the reach of heavy stops in the context of $\tilde{t} - \tilde{g} - \tilde{\chi}_1^0$ model at future 100 TeV collider. There are two stop production channels in this model, each characterized by 6 top quarks and missing energy in final state :

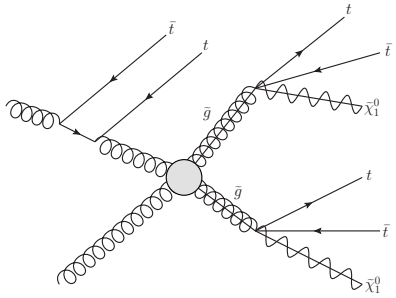
- Stop-pair production : $pp \to \tilde{t}\tilde{t}^* \to t\bar{t}t\bar{t}t\bar{t} + \not{E}_T$ (fig. 2.3a).
- Stop-gluino associated production : $pp \to t\tilde{t}^*\tilde{g} \to t\bar{t}t\bar{t}t\bar{t} + \not{E}_T$ (fig. 2.3b).

Besides SM backgrounds, there is an additional important SUSY background to be considered. The mass hierarchy of $\tilde{t} - \tilde{g} - \tilde{\chi}_1^0$ model implies that the gluino must have already been discovered before the stops. Therefore, we must also consider the following gluino-pair production channels for background :

- Gluino-pair production : $pp \to \tilde{g}\tilde{g} \to t\bar{t}t\bar{t} + \not{E}_T$ (fig. 2.4a).



(a) Gluino-pair production



(b) Gluino-pair production with tops

Figure 2.4: Sample Feynman diagrams for SUSY background in the $\tilde{t} - \tilde{g} - \tilde{\chi}_1^0$ simplified model.

- Gluino-pair production with tops : $pp \to \tilde{g}\tilde{g}t\bar{t} \to t\bar{t}t\bar{t}t\bar{t} + \not{E}_T$ (fig. 2.4b).

2.4 Event Generation and Jet Observables

2.4.1 Event Generation

Parton-level events were generated using `MadGraph5` [103], split into four bins : $H_T \in (1.5, 3]$ TeV, $(3, 5.5]$ TeV, $(5.5, 8.5]$ TeV and $(8.5, 100]$ TeV, followed by parton-showering and hadronization in `Pythia8` [104] and detector simulation in `Delphes` [105]. For SM background samples, additional jets were included at the parton-level[3] and then matched to parton shower using the MLM matching scheme [106]. H_T and $\not{E}_T$ distributions for SM and SUSY processes are shown in fig. 2.5 and fig. 2.6 respectively. These distributions serve as a consistency check for correct normalization of H_T bins as well as matching for SM processes. As pre-selection cuts, events are required to have $H_T > 2$ TeV and $\not{E}_T > 200$ GeV.

[3]Two additional jets were included for all SM processes except $t\bar{t} + W/Z$ for which only one additional jet was included.

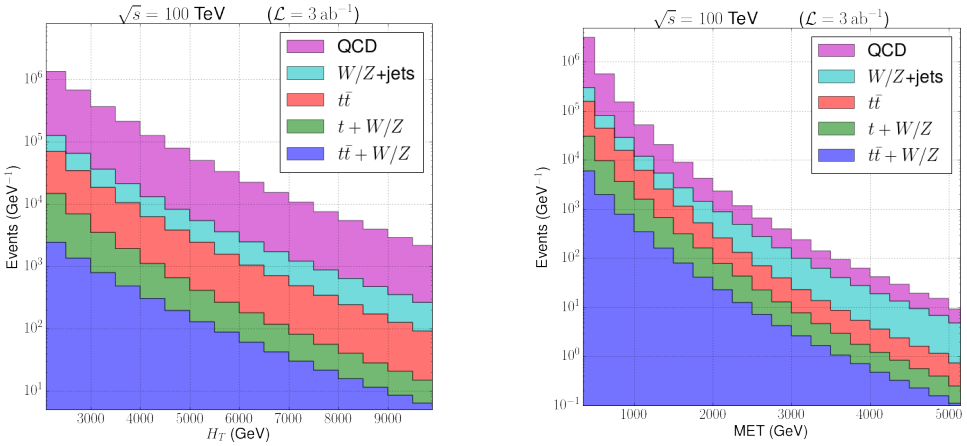


Figure 2.5: H_T (left) and $\not{E}_T$ (right) distributions for SM processes.

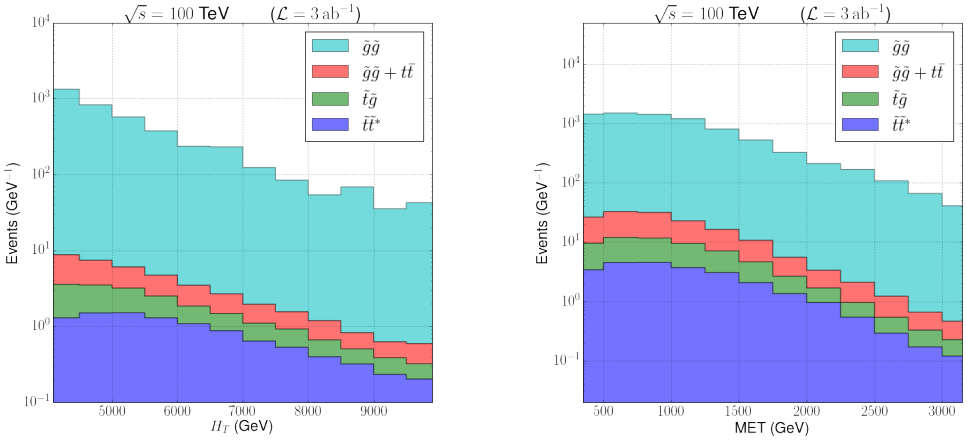


Figure 2.6: H_T (left) and $\not{E}_T$ (right) distributions for SUSY processes for $m_{\tilde{t}} = 4$ TeV, $m_{\tilde{g}} = 2$ TeV and $m_{\tilde{\chi}_1^0} = 200$ GeV.

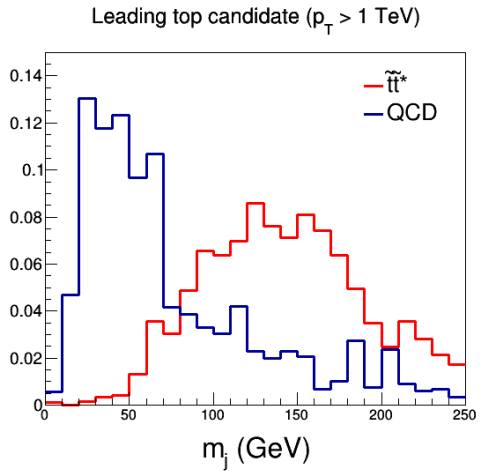
2.4.2 Jet Clustering

Final state hadrons and non-isolated leptons are clustered into jets using `FastJet` [107] with jet radius parameter $R = 0.5$ and using the anti-k_T algorithm [108]. Given that both simplified models are characterized by boosted top quarks in the final state, jet substructure is a valuable tool for identifying tops. To this end, we additionally cluster fat jets with $p_T > 200$ GeV using the Cambridge/Aachen (C/A) algorithm [109, 110] and jet radius parameter $R = 1$, with the idea being that fat jets adequately capture top decay products. A well-known issue with fat jets is that the presence of final state radiation (FSR) from top quarks and initial-state radiation (ISR)/underlying event can adversely affect the jet mass and other jet substructure properties. To mitigate this problem, Ref. [102] proposed scaling down the fat jet radius to $R = Cm_{\rm top}/p_T$ where C is $\mathcal{O}(1)$ number. The basic idea behind using dynamic radius is that the top decay products are confined to angular size of $m_{\rm top}/p_T$ while ISR/FSR outside this cone-size is excluded.

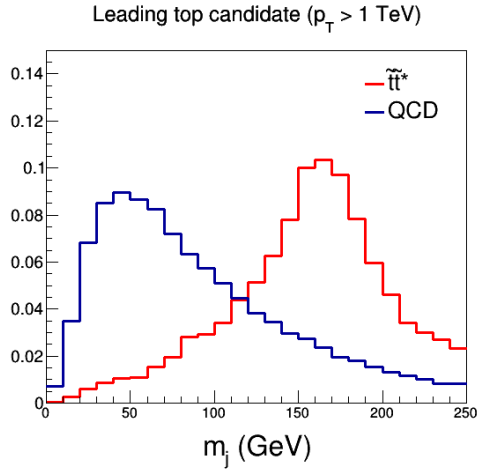
In our analyses, we recluster the C/A jets using the anti-k_T algorithm and winner-take-all (WTA) recombination scheme [111]. In the analysis of the $\tilde{t} - \tilde{H}$ simplified model, we recluster the C/A jets with $R = 600$ GeV$/p_T \approx 3.5 m_{\rm top}/p_T$. In the $\tilde{t} - \tilde{g} - \tilde{\chi}_1^0$ simplified model, there are 6 top quarks in the final state. Occasionally, multiple top quarks are clustered into a single fat jet. To resolve this issue, we perform a two-step scaling down procedure. In the first step, we recluster the C/A jets with $R = (1\,\text{TeV})/p_T$ to separate multiple top quarks if any. All subjets with $p_T > 500$ GeV are retained as top candidates. In the second step, each of the resulting jets are further reclustered with a smaller radius of $R = (600\,\text{GeV})/p_T$ to remove ISR/FSR. We compute jet observables which we will discuss below based on the reclustered final jets.

2.4.3 Jet Mass

We calculate the jet mass, m_j, in two ways depending on the p_T of the jet. For jets with $p_T < 1$ TeV, we calculate m_j using the energy-momentum information from both the the tracker and the colorimeters, which is the same way as is done at the LHC. For jets with



(a) Leptonically decaying top candidate



(b) Hadronically decaying top candidate

Figure 2.7: Jet mass distributions for $\tilde{t}\tilde{t}^*$ and QCD samples.

$p_T > 1$ TeV, the cone size of the jet is so small that calorimeter cells in the future collider may not provide enough spatial resolution to resolve the jet constituents. Therefore, we will use the method described in Ref. [102, 112], i.e. using only the tracker energy-momentum information to calculate $m_j^{(\text{track})}$. Then the jet mass is rescaled to remove the tracker's bias for charged particles,

$$m_j = m_j^{(\text{track})} \frac{p_T^{(\text{track+calorimeter})}}{p_T^{(\text{track})}} . \tag{2.1}$$

In fig. 2.7, we present the jet mass distributions for boosted top candidate jets with $p_T > 1$ TeV in $\tilde{t}\tilde{t}^*$ and QCD light flavor samples. Leptonically decaying top candidate jets characterized by the presence of a hard muon ($p_T > 200$ GeV) inside the jet are shown in the top panel while hadronically decaying top candidate jets are shown in the bottom panel. In the $\tilde{t}\tilde{t}^*$ sample, the leading top candidate jets are likely from boosted top quarks produced from stop decays. This is reflected in the jet mass distribution of the leading jet in $\tilde{t}\tilde{t}^*$ which peak at $\sim m_{\text{top}}$ while QCD jets peak at much lower values as shown in fig. 2.7 (b). A similar trend is observed for leptonically decaying top candidates, as shown in fig. 2.7 (a), with a minor difference that the jet mass distribution peaks at slightly lower values ~ 145 GeV due

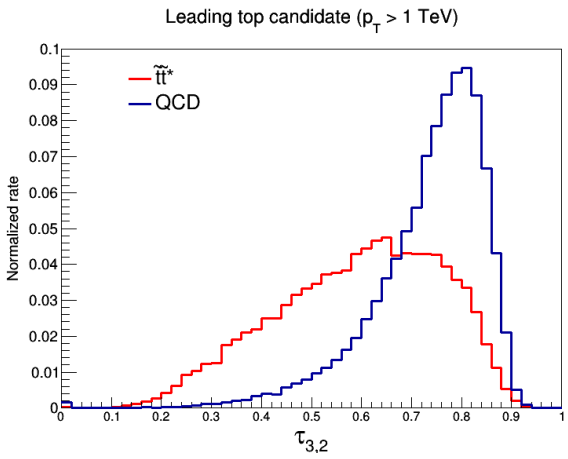


Figure 2.8: $\tau_{3,2}$ distribution of leading top candidate jet in $\tilde{t}\tilde{t}^*$ and QCD samples.

to missing energy.

2.4.4 N-subjettiness

A boosted top quark decaying hadronically has a three-prong substructure unlike a QCD jet. One of the jet observables that exploits this N-prong substructure of boosted particles is N-subjettiness $\tau_N^{(\beta)}$ which is defined as in Ref. [113] :

$$\tau_N^{(\beta)} = \sum_i p_{T,i} \min\left\{(\Delta R_{1,i})^\beta, (\Delta R_{2,i})^\beta, \cdots, (\Delta R_{N,i})^\beta\right\} \tag{2.2}$$

where the sum runs over all the constituent particles of the jet, $p_{T,i}$ is the p_T of the i^{th} constituent particle, $\Delta R_{J,i}$ is the angular separation[4] between the i^{th} constituent and subjet axis J and the β parameter is an angular weighting exponent. The N subjet axes are defined using the exclusive k_T-algorithm with WTA recombination scheme. For the case of top quark which has a 3-prong decay, $\tau_3^{(\beta)}$ is the relevant observable. However, it has been shown in Ref. [113] that following variable is a better discriminator between top jets and QCD jets:

$$\tau_{3,2}^{(\beta)} = \frac{\tau_3^{(\beta)}}{\tau_2^{(\beta)}} \tag{2.3}$$

From here on, we will set $\beta = 1$. For top candidate jets with $p_T < 1$ TeV, we use both tracker and calorimeter information to compute $\tau_{3,2}$ while for jets with $p_T > 1$ TeV, we only

[4]Angular separation is defined as $\Delta R_{J,i} = \sqrt{(\Delta\eta_{J,i})^2 + (\Delta\phi_{J,i})^2}$.

use tracker information. In fig. 2.8, the $\tau_{3,2}$ distribution of the leading top candidate jet is shown for $\tilde{t}\tilde{t}^*$ and QCD samples. For the figure, only boosted top candidate jets with $p_T > 1$ TeV are selected. The $\tilde{t}\tilde{t}^*$ sample was generated for $\tilde{t} - \tilde{g} - \tilde{\chi}_1^0$ simplified model with $m_{\tilde{t}} = 4$ TeV, $m_{\tilde{g}} = 2$ TeV and $m_{\tilde{\chi}_1^0} = 200$ GeV. Due to 3-prong substructure of top quark decays, $\tau_{3,2}$ for boosted tops peaks at smaller values compared to that for QCD jets.

2.4.5 Mass-drop

For boosted jets containing a hard muon $p_T > 200$ GeV, mass-drop x_μ is defined as follows [114, 115]:

$$x_\mu \equiv 1 - \frac{m^2_{j\not{\mu}}}{m^2_j}, \tag{2.4}$$

where m_j is the jet mass calculated as in Sec. 2.4.3 and $m_{j\not{\mu}}$ is the mass of the jet excluding the hard muon. The observable measures how much of the jet invariant mass is carried by hadronic activity. In a boosted top jet with W decaying to a muon, $m_{j\not{\mu}}$ is approximately the invariant mass of the b jet, which is only a small fraction of $m_j \sim m_{\text{top}}$. Thus we expect x_μ to be close to 1. On the other hand, for heavy flavor jets such as b jets, the muon only carries a small fraction of energy and a large jet invariant mass should come from hadronic activity, resulting in $x_\mu \to 0$.

The distributions of x_μ for different samples are presented in fig. 2.9. The left panel shows the x_μ distributions of leading jets in the SM $t\bar{t}$ and SUSY $\tilde{t} \to t\tilde{B}$ samples. The distributions are similar and both peak at $x_\mu \approx 1$ since most of the leading jets in both samples are top jets. There is a smaller bump at lower x_μ, which comes from tops with leptonic-decaying b's[5] and hadronic decaying W bosons. The right panel shows the distributions of leading jets in the SM QCD and SUSY $\tilde{t} \to b\tilde{H}^\pm$ samples. In these two samples, the leading jets are mostly b jets. Thus their distributions are comparable and both peak at smaller values of x_μ close to 0. Notice that the non-zero peak value of x_μ is due to the requirement that muon inside the jet satisfy $p_T > 200$ GeV. The mass drop of the leading b jets peaks at zero when the muon p_T cut is removed, consistent with the results in Ref. [114, 115].

[5]Leptonic decaying b is defined as a b jet with a muon in it.

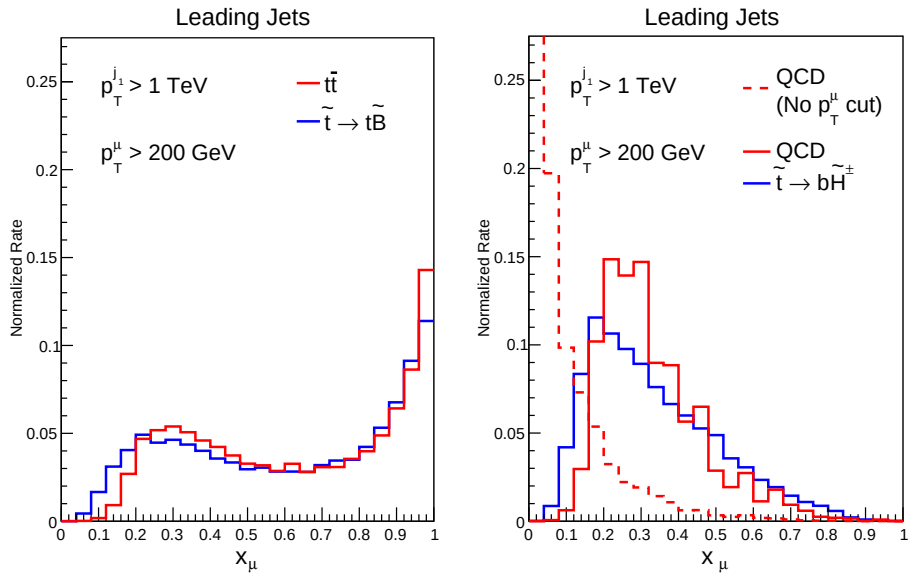


Figure 2.9: Mass-drop distributions for the leading jet with $p_T > 1$ TeV in signals and backgrounds. Left: mass drop distributions of $t\bar{t}$ background and $\tilde{t} \to t\tilde{B}$ SUSY sample. Right: mass drop distributions of QCD background and $\tilde{t} \to b\tilde{H}^{\pm}$ SUSY sample. We also require the muon to have $p_T > 200$ GeV except for the dashed QCD distribution, which is obtained without any muon p_T cut.

2.5 Analysis: $\tilde{t} - \tilde{H}$ Simplified Model

In this and next sections, we will present analyses and results for the two stop simplified models. NLO+NNLL cross-sections were used for stop-pair and gluino-pair processes [72] while LO cross-sections from `MadGraph` were used for the remaining processes[6] .

2.5.1 Boosted Top and Bottom Tagging

In analyzing the $\tilde{t} - \tilde{H}$ simplified model and distinguishing it from the $\tilde{t} - \tilde{B}$ model, we will combine jet mass m_j and mass drop x_μ variables to tag boosted top and bottom jets. We define a *boosted top* jet as a jet (clustered using the method in Sec. 2.4.2 with $p_T > 500$ GeV) with a $p_T > 200$ GeV muon inside and $x_\mu > 0.5$ **or** $m_j > 120$ GeV. A *boosted b* jet, on the other hand, is required to have a $p_T > 200$ GeV muon inside and satisfy $x_\mu < 0.5$ **and** $m_j < 120$ GeV. The tagging efficiencies for both the SUSY and the SM samples are shown in fig. 2.10.

In both taggings, the muon-in-the-jet requirement is because the decay of either a boosted bottom or top could give a hard muon close to the hadronic jet axis with a certain branching fraction. This is the same strategy as in Ref. [74]. Yet to further distinguish between $\tilde{t} - \tilde{H}$ and $\tilde{t} - \tilde{B}$ simplified models, we need to tell apart a boosted b and a top jet using a combination of m_j and x_μ observables. Tops with leptonic W's are likely to have x_μ close to 1 but smaller m_j while tops with hadronic W's but leptonic b's have smaller x_μ but larger m_j. To tag both cases and keep most of the SUSY signals after kinematic cuts, we require top jets to satisfy either $x_\mu > 0.5$ or $m_j > 120$ GeV. On the other hand, a b jet has a small jet mass as well as mass drop. Thus a tagged b jet is required to have $x_\mu < 0.5$ and $m_j < 120$ GeV simultaneously.

The efficiency to tag a top quark produced in the SUSY decay process $\tilde{t} \to t\tilde{\chi}^0$ ($\tilde{\chi}^0$ could be either a bino or a neutral higgsino) is around 10%. The efficiencies of SM background events containing top pairs is at around $1\% - 3\%$. The $t\bar{t}$ sample has a smaller efficiency

[6] We follow Ref. [116] and treat tops as final state particles instead of using top parton distribution function in evaluating SUSY production associated with tops such as $\tilde{t}\tilde{g}$ associated production.

than the stop-pair sample because the leading jet in the top-pair sample is occasionally ISR. On the other hand, QCD jets are mistagged as top jets at a rate of at most $\sim 0.1\%$ with the mistag rate even lower in the low p_T bins. The high background suppression is achieved due to a hard muon-in-jet requirement. b jets from $\tilde{t} \to b\tilde{H}^{\pm}$ are mistagged as top jets at a mere percentage level.

For boosted b tagging, the efficiency to tag a b jet from $\tilde{t} \to b\tilde{H}^{\pm}$ is around $4\% - 5\%$. The SM backgrounds with tops in the final state are suppressed with efficiencies $\lesssim 0.5\%$. The efficiency of the QCD background is even smaller at $\sim 0.2\%$. Top jets from $\tilde{t} \to t\tilde{H}^0/\tilde{B}$ are mistagged as bottom jets at 1% level, similar to the mistag rate of b jets using the top tagging strategy.

2.5.2 Event Selection

We require the events to satisfy the following requirements:

- At least two $R = 0.5$ anti-k_T jets each with $p_T > 1$ TeV;
- No isolated lepton with $p_T > 35$ GeV.
- $|\Delta\phi(j, \not{E}_T)| > 1.0$ for any anti-k_T jet with $p_T > 500$ GeV;
- $\not{E}_T > 3.0$ TeV;
- At least one top-tagged or bottom-tagged jet with the tagging described in Sec. 2.5.1.

The lepton isolation criteria is that the total sum of p_T of all the charged particles inside a cone with $R = 0.5$ around the lepton is less than 10% of the lepton's p_T.

For the $\tilde{t} - \tilde{B}$ simplified model, $\tilde{t}$ only decays to $t\tilde{B}$. Given the efficiencies shown in fig. 2.10, we expect 10% of the signal events to be top-tagged and a negligible fraction of the events to be b-tagged. On the other hand, in the $\tilde{t} - \tilde{H}$ simplified model, $\tilde{t}$ decays to both $t\tilde{H}^0$ and $b\tilde{H}^{\pm}$, each with 50% branching fraction. A SUSY signal event could contain either pure decays where both $\tilde{t}$'s decay though the same channel or mixed decays where one $\tilde{t}$ decays to $t\tilde{H}^0$ with the other one to $b\tilde{H}^{\pm}$. Since the signal efficiencies for tagging a boosted

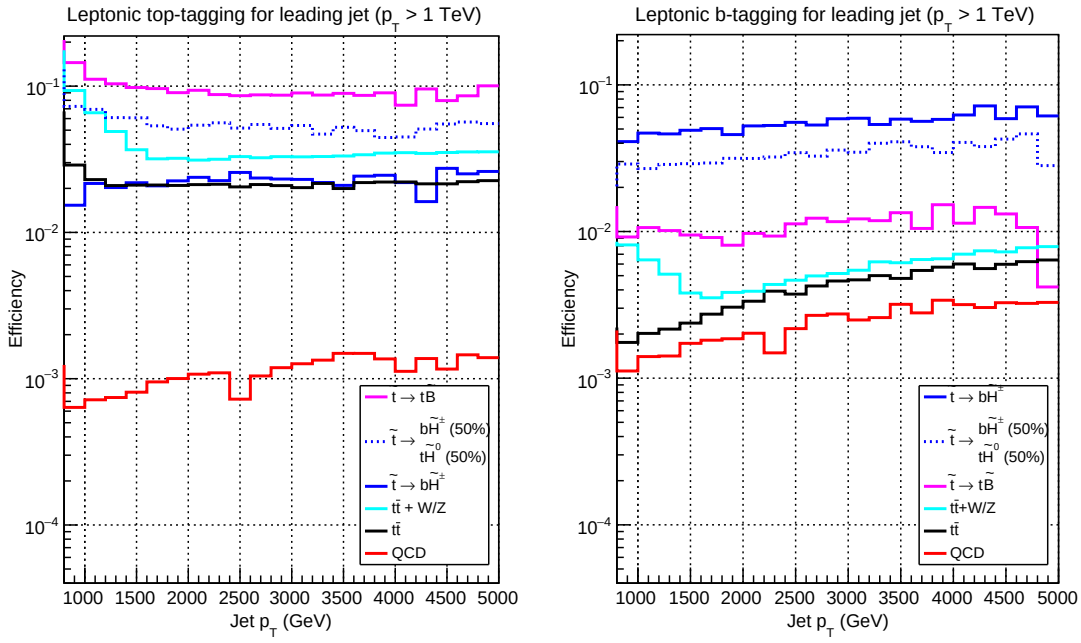
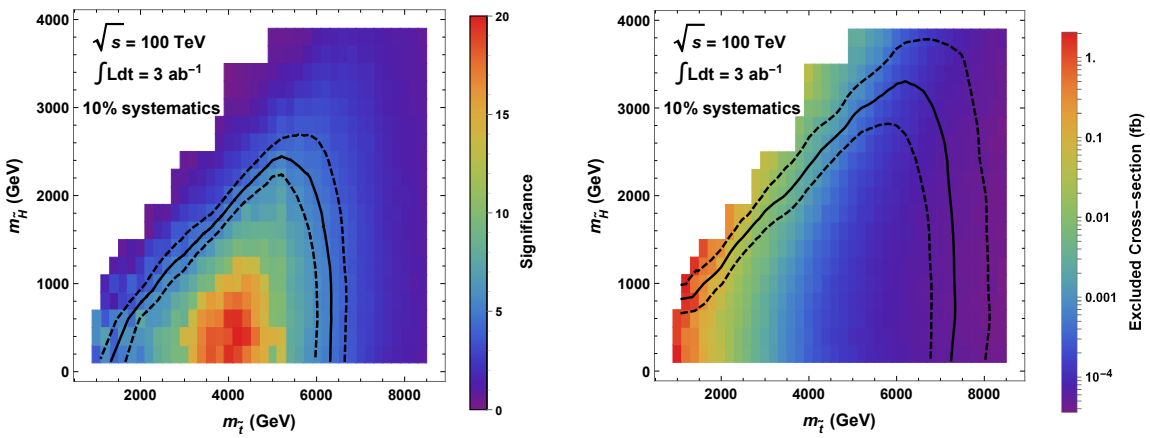


Figure 2.10: Boosted top and b tagging efficiencies for the leading jet (fraction of total events with the leading jet tagged) in different event samples as a function of jet p_T: SUSY events with stops decaying only to $b\tilde{H}^{\pm}$ (blue) or $t\tilde{B}$ (pink), SUSY events with stops decaying to either neutral or charged $\tilde{H}$ in the full $\tilde{t}-\tilde{H}$ simplified model (blue dotted), SM $t\bar{t}$ background (black), SM QCD background (red) and SM $t\bar{t}+W/Z$ background (light blue). We assumed $m_{\tilde{t}} = 4$ TeV and $m_{\tilde{H}}$, $m_{\tilde{B}} = 500$ GeV.

jet are $\lesssim 10\%$, the chance of tagging both t's or b's in the pure decay case or tagging both t and b in the mixed decay case is very low (typically less than 1 event after all the kinematic cuts for 3 ab^{-1} of data). The signal events are then a mixture with some events t-tagged and the rest b-tagged. We will use the number of t and b-tagged events to pin down the identity of LSP and differentiate the two simplified models in the next section.

2.5.3 Exclusion and Discovery

We use N_b to denote the number of b-tagged signal events after all the cuts and N_t for number of t-tagged signal events. The total number of signal events used to set the reach is then $N_+ = N_b + N_t$. We scan the $(m_{\tilde{t}}, m_{\tilde{H}})$ plane and apply CL$_s$ statistics [117] to



(a) Discovery

(b) Exclusion

Figure 2.11: The discovery and exclusion contours for the $\tilde{t}-\tilde{H}$ simplified model at a 100 TeV collider with an integrated luminosity of 3 ab^{-1}. We assume a 10% systematic uncertainty for both the signal and the background. The solid lines are $5\,\sigma$ discovery contour (left) and exclusion at 95% C.L.(right). The dashed lines are the $\pm 1\,\sigma$ boundaries.

compute exclusion and discovery contours. Both the signal and backgrounds are modelled by Poisson statistics. A point in the mass plane is excluded if its CL$_s$ < 0.05. A point could be discovered when the background only hypothesis is rejected with a p-value less than 3×10^{-7}. We also require at least 10 total signal events for a point to be excluded or discovered. This conservative requirement does not affect our results when using the CL$_s$ method. It will make the physics reach estimate more robust when using the simpler approximate $S/\sqrt{B}$ estimate in the analysis in Sec. 2.6.

From fig. 2.11a, stops with mass up to 5 - 6 TeV could be discovered for higgsino mass up to 2 TeV, assuming an integrated luminosity 3 ab^{-1}. At 95% C.L., stops with mass up to 7 TeV could be excluded. All the results shown here are based on the simple cut-flows in Sec. 2.5.2. We expect further optimization (e.g., through boosted decision tree) can improve the results. In addition, we do not try to perform a dedicated analysis for the compressed region where the stop mass gets closer to the higgsino mass.

The total number of signal events, N_+, will be the same for $\tilde{t}-\tilde{H}$ and $\tilde{t}-\tilde{B}$ simplified

models if the $\tilde{H}$ and $\tilde{B}$ have the same masses. Assuming a discovery of stops, we proceed to distinguish between the two simplified models using the difference between N_b and N_t. The observable we will use is a ratio

$$r_- = \frac{N_b - N_t}{N_b + N_t}\,. \tag{2.5}$$

The advantage of a ratio observable is that systematic uncertainties contributing to individual observables are likely to cancel out. The distributions of r_- for $\tilde{t}-\tilde{H}$ and $\tilde{t}-\tilde{B}$ simplified models are demonstrated in fig. 2.12. In the figure, r_- peaks at $\sim -(0.2-0.3)$ in the $\tilde{t}-\tilde{H}$ model. This can be understood as follows: the t-tagging efficiency of the signal is $\epsilon^t_{\rm sig} \approx 10\%$ while that for b-tagging efficiency is about $\epsilon^b_{\rm sig} \approx 5\%$, as in Sec. 2.5.1. In a $\tilde{t}-\tilde{H}$ sample, 1/4 of the events contain two b jets, 1/4 of the events contain two t jets while the rest half contains one b and one t jet. Thus $N_b \approx \epsilon^b_{\rm sig} N_+$ and $N_t \approx \epsilon^t_{\rm sig} N_+$, leading to $r_- \approx -0.3$. In contrast, r_- peaks at ~ -0.6 in the $\tilde{t}-\tilde{B}$ model. This is consistent with that almost all the events in a $\tilde{t}-\tilde{B}$ sample contain two boosted t jets. Ignoring the rate of mistagging a t jet as a b jet, $r_- \approx -1$. Including the mistag rate shifts the central value to -0.6.

Finally we show the 95% C.L. exclusion of $\tilde{t}-\tilde{B}$ model based on r_- assuming that the signal comes from the $\tilde{t}-\tilde{H}$ simplified model in fig. 2.13. From the figure, one could see that the 95% C.L. contour overlaps with the 5σ discovery reach in fig. 2.11a. Thus using N_+ and r_-, we could not only discover stops up to 6 TeV but also determine whether the LSP is a higgsino or bino.

2.6 Analysis: $\tilde{t}-\tilde{g}-\tilde{\chi}^0_1$ Simplified Model

2.6.1 Top-tagging

The final state in the $\tilde{t}-\tilde{g}-\tilde{\chi}^0_1$ simplified model is characterized by 6 top quarks, several of which may be boosted. Therefore, we rely on multiple top tags to discriminate signal from background. Anti-k_T jets with $R = 0.5$ and C/A jets with $R = 1.0$ are identified using the procedure outlined in Sec. 2.4.2. Two separate top tagging strategies are used for hadronic and leptonic top decays. If an energetic muon with $p_T > 200$ GeV is among

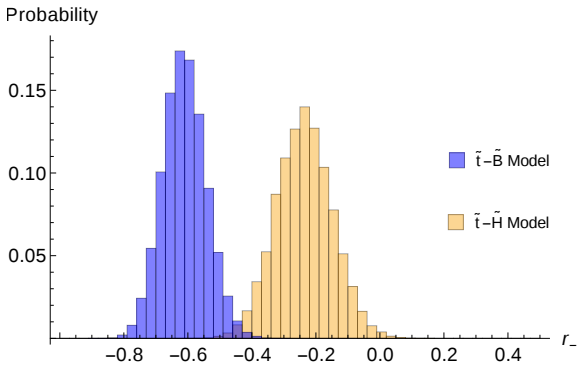


Figure 2.12: Distributions of r_- for stop-bino and stop-higgsino simplified models given $m_{\tilde{t}} = 4$ TeV, $m_{\tilde{H}}, m_{\tilde{B}} = 500$ GeV.

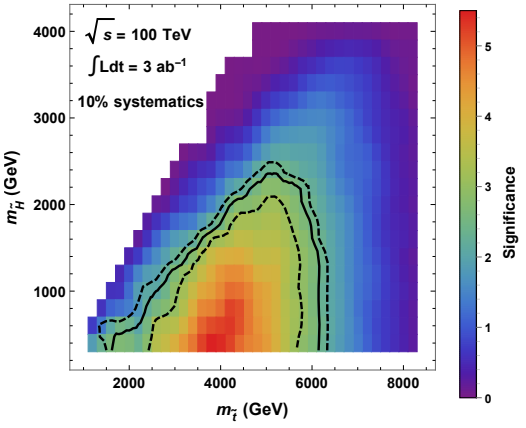


Figure 2.13: The 2σ exclusion contour of $\tilde{t} - \tilde{B}$ simplified model based on r_- assuming the signal events are from $\tilde{t} - \tilde{H}$ model. The dashed contours are $\pm 1\,\sigma$ boundaries.

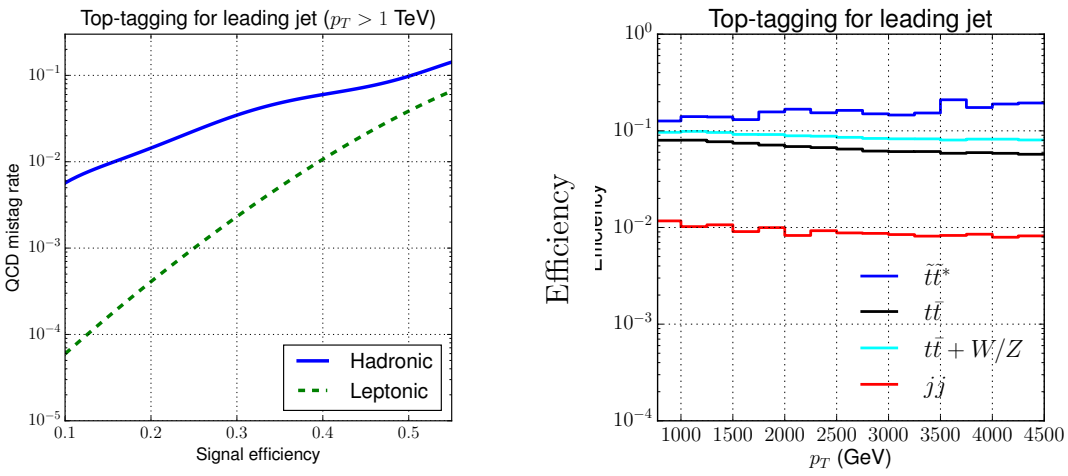
the constituents of a C/A jet, that jet is identified as a leptonically decaying top candidate. Otherwise, the C/A jets are reclustered using p_T dependent radius following the two-step scaling down procedure described in Section 2.4.2. The resulting subjets are identified as hadronically decaying top candidates. For hadronic top candidates, jet mass is required to lie in the top mass window of 140 GeV $< m_j <$ 240 GeV to reject QCD jets as shown in fig 2.7b. For leptonic top candidates, top mass reconstruction is not possible due to missing energy. Nevertheless, requiring $m_j > 75$ GeV provides a good discrimination between top jets and QCD jets as shown in fig. 2.7a.

To further improve top-tagging, we use the N-subjettiness variable $\tau_{3,2}$ (see Section 2.4.4) for hadronic top decays and the mass drop variable x_μ (see Section 2.4.5) for leptonic top decays. By imposing cuts on these two parameters, it is possible to obtain the desired signal efficiency. In fig 2.14a, the QCD mistag rate is plotted against signal efficiency for the leading top candidate jet. Cuts on the jet mass for both leptonic and hadronic channels are already imposed and included in the efficiency and mistag rates. Note that in computing the rates in the top panel of fig. 2.14, we used slightly different definitions from those for fig. 10 and the bottom panel of fig. 2.14: the efficiencies of hadronic (leptonic) top tagging are the fractions of events with a hadronic (leptonic) top candidate satisfying the tagging requirements. While top-tagging is more efficient in the leptonic channel, it suffers from a low branching ratio. Therefore, using both leptonic and hadronic tagging is beneficial. We will choose $0.1 < \tau_{3,2} < 0.45$ and $x_\mu > 0.7$ which corresponds to QCD mistag rate of $\sim 1\%$. Using these cuts, the jet-p_T dependence of the combined top-tagging efficiency for the leading top candidate is plotted in fig. 2.14b. The combined signal efficiency is $\sim 10 - 20\%$ compared to $\sim 1\%$ for QCD jets.

2.6.2 Event Selection and Results

The following cuts are used to discriminate between signal and background :

- $H_T > 4$ TeV and $\not{E}_T > 250$ GeV;
- No isolated leptons with $p_T > 50$ GeV;



(a) (b)

Figure 2.14: (a) QCD mistag rate vs signal efficiency for top-tagging the leading top candidate jet. (b) Top-tagging efficiency for signal and SM processes as a function of jet p_T.

Cuts	$\tilde{t}\tilde{t}^*$	$\tilde{t}\tilde{g}$	$\tilde{g}\tilde{g}$	$\tilde{g}\tilde{g} + t\bar{t}$
$H_T > 4$ TeV, $\not{E}_T > 250$ GeV	8809	12415	8.94×10^6	34990
No leptons	7687	10723	8.08×10^6	30312
$n_j \geq 7$ and ISR cuts	3574	4435	1.13×10^6	10517
$\lvert\Delta\phi(j, \not{E}_T)\rvert > 0.5$	2788	3589	901151	8294
1 top-tag	490	630	131816	1412
2 top-tags	228	233	27910	500
3 top-tags	52	48	3555	111
$H_T, \not{E}_T$ cuts	8.6	2.1	0	1.5

Table 2.1: Cut flow for SUSY processes at $\sqrt{s} = 100$ TeV and $\mathcal{L} = 30$ ab^{-1}. SUSY masses are $m_{\tilde{t}} = 5.5$ TeV, $m_{\tilde{g}} = 2.75$ TeV and $m_{\tilde{\chi}_1^0} = 200$ GeV.

Cuts	$t\bar{t}$	$t\bar{t}+W/Z$	QCD	$t+W/Z$	W/Z+jets
$H_T > 4$ TeV, $\not{E}_T > 250$ GeV	5.96×10^7	3.94×10^6	1.24×10^9	8.32×10^6	8.65×10^7
No leptons	5.72×10^7	3.76×10^6	1.24×10^9	8.06×10^6	8.34×10^7
$n_j \geq 7$ and ISR cuts	3.16×10^6	1.67×10^5	3.64×10^7	1.18×10^5	1.89×10^6
$\lvert\Delta\phi(j,\not{E}_T)\rvert > 0.5$	1.53×10^6	81624	1.49×10^7	52675	8.50×10^5
1 top-tag	70782	5698	193858	2522	9589
2 top-tags	9520	690	479	99	701
3 top-tags	132	18	0	0	0
$H_T, \not{E}_T$ cuts	0	0.1	0	0	0

Table 2.2: Cut flow for SM processes at $\sqrt{s} = 100$ TeV and $\mathcal{L} = 30$ ab^{-1}.

- At least 7 jets (anti-k_T with $R = 0.5$ and $p_T > 200$ GeV);
- At most one ISR jet among the leading 6 jets (see below);
- $|\Delta\phi(j, \not{E}_T)| > 0.5$ for the leading two jets;
- At least 3 top tagged jets with the top tagging described in Section 2.6.1;
- Optimized H_T and $\not{E}_T$ cuts (see below).

At 100 TeV collider, imposing jet multiplicity cut is not sufficient to distinguish hard jets from ISR. To resolve this issue, ISR jets are identified by one of the two criteria [118]:

- high rapidity : $|\eta| > 2$
- a big hierarchy in successive jet p_T's: for p_T-ordered jets, every ratio of successive jet p_Ts less than 0.2 is counted as an ISR.

In the last step, harder H_T and $\not{E}_T$ cuts are imposed and optimized so as to maximize the reach σ defined as :

$$\sigma = \frac{S}{\sqrt{B + \gamma^2(S^2 + B^2)}} \tag{2.6}$$

where S is the number of signal event, B is the number of background events and γ is the systematic uncertainty for both signal and background.

The cut flow for SUSY and SM processes at $\sqrt{s} = 100$ TeV and luminosity $\mathcal{L} = 30\,\text{ab}^{-1}$ is shown in Tables 2.1 and 2.2 respectively. The SUSY mass spectrum is chosen to be $m_{\tilde{t}} = 5.5$ TeV, $m_{\tilde{g}} = 2.75$ TeV and $m_{\tilde{\chi}_1^0} = 200$ GeV. Preliminary H_T and $\not{E}_T$ cuts are designed to suppress SM backgrounds which have very large cross-sections. The signal processes $\tilde{t}\tilde{t}^*$ and $\tilde{t}\tilde{g}$ have up to 6 top quarks in the final state while the SM backgrounds and the $\tilde{g}\tilde{g}$ background have fewer hard partons in the final state. This justifies the requirement for 7 hard jets. Nevertheless, the preliminary H_T cut inadvertently selects background events with ISR jets which can mimic hard jets. Therefore, the hard jet-multiplicity cut has to be supplemented by vetoing ISR jets. To this end, we require that at most one ISR jet be present among the 6 hardest jets. The $|\Delta\phi(j, \not{E}_T)|$ cut is designed to suppress SM backgrounds such as $W/Z+$ jets where the missing energy from W/Z decay is mostly aligned with the leading jets due to collinear emission of W/Z bosons from jets. In addition, it could suppress QCD mismeasurement backgrounds.

At this stage, several background processes still have 3 orders of magnitude more events than the signal with QCD being the dominant background. Next, we make use of the high top-quark multiplicity in the signal processes unlike SM backgrounds that have at most 2 top quarks in the final state. By requiring 3 top tags, the largest QCD background is completely eliminated while also suppressing other backgrounds. After top-tagging, the dominant background is $\tilde{g}\tilde{g}$ along with sub-dominant contributions from $t\bar{t}$, $\tilde{g}\tilde{g} + t\bar{t}$ and $t\bar{t} + W/Z$ processes. In the last step, we maximize the stop reach significance by performing a scan over H_T-$\not{E}_T$ cuts.

The stop reach for $\tilde{t} - \tilde{g} - \tilde{\chi}_1^0$ simplified model at $\sqrt{s} = 100$ TeV and luminosity of 30 ab^{-1} is shown in Table 2.3. The NLL+NLO gluino-pair cross-section is 1.33 pb for 2.75 TeV gluinos at $\sqrt{s} = 100$ TeV while stop-pair cross-sections are shown in Table 2.3. For $m_{\tilde{t}} = 5.5\,(6.0)$ TeV and $m_{\tilde{g}} = 2.75$ TeV, we were able to obtain a reach of $6.3\,(3.5)\,\sigma$ for a systematic uncertainty $\gamma = 0.1$. The optimal H_T-$\not{E}_T$ cuts were found to be $H_T > 9.5$ TeV and $\not{E}_T \gtrsim 1.5$ TeV (1.25 TeV for $m_{\tilde{t}} = 6.0$ TeV).

$m_{\tilde{t}}$ (TeV)	$\sigma^{\rm NLO+NLL}_{pp\to\tilde{t}\tilde{t}^*}$ (fb)	S	B	σ
5.5	0.40	10.7	1.7	6.3
6.0	0.23	10.0	6.7	3.5

Table 2.3: Stop reach for $\tilde{t} - \tilde{g} - \tilde{\chi}^0_1$ simplified model at $\sqrt{s} = 100$ TeV with luminosity $\mathcal{L} = 30$ ab^{-1} and systematic uncertainty $\gamma = 0.10$. Here, $m_{\tilde{g}} = 2.75$ TeV and $m_{\tilde{\chi}^0_1} = 200$ GeV.

Cuts	$\tilde{g}\tilde{g}$	$t\bar{t}$	$t\bar{t}+W/Z$	QCD	$t+W/Z$	W/Z+jets
$H_T > 4$ TeV, $\not{E}_T > 250$ GeV	802	5.96×10^6	3.94×10^5	1.24×10^8	8.32×10^5	8.65×10^6
No leptons	764	5.72×10^6	3.76×10^5	1.24×10^8	8.06×10^5	8.34×10^6
$n_j \geq 5$, ISR cuts	528	2.19×10^6	1.38×10^5	3.13×10^7	1.00×10^5	2.02×10^6
$\lvert\Delta\phi(j,\not{E}_T)\rvert > 0.5$	447	8.97×10^5	57806	9.74×10^6	38576	7.69×10^5
1 top-tag	88	49343	4804	87361	1951	10789
2 top-tags	34	5342	632	1352	98	351
$H_T, \not{E}_T$ cuts	12.4	0.57	0.23	0	0	0

Table 2.4: Cut flow for gluino-pair and SM processes at $\sqrt{s} = 100$ TeV and $\mathcal{L} = 3$ ab^{-1}. SUSY masses are $m_{\tilde{g}} = 10$ TeV and $m_{\tilde{\chi}^0_1} = 200$ GeV.

2.6.3 Improvement in Gluino Search

It should be noted that the jet observables presented so far can also be used to improve gluino reach at future hadron colliders. In Table 2.4, a cut flow analysis is presented for gluino-pair and SM processes at $\sqrt{s} = 100$ TeV and luminosity $\mathcal{L} = 3\,\text{ab}^{-1}$. The SUSY mass spectrum is chosen to be $m_{\tilde{g}} = 10$ TeV and $m_{\tilde{\chi}^0_1} = 200$ GeV. The only differences compared to the stop cut flow analysis is that the minimum number of jets requirement is relaxed to 5, up to two ISR jets are allowed and at most two top tags are required. In Table 2.5, the gluino reach at 100 TeV collider and luminosity of 3 ab^{-1} is presented. The final

$m_{\tilde{g}}$ (TeV)	$\sigma^{\rm NLO+NLL}_{pp\to\tilde{g}\tilde{g}}$ (fb)	Top tags	S	B	σ
10.0	0.31	2	12.4	0.8	8.1
11.0	0.13	1	13.8	9.5	3.9

Table 2.5: Gluino reach for $\tilde{t}-\tilde{g}-\tilde{\chi}^0_1$ simplified model at $\sqrt{s}=100$ TeV with luminosity $\mathcal{L}=3$ ab^{-1} and systematic uncertainty $\gamma=0.10$. Here, $m_{\tilde{\chi}^0_1}=200$ GeV while $m_{\tilde{t}}\gg m_{\tilde{g}}$.

H_T-$\not{E}_T$ optimized cuts were chosen to be $H_T>11$ TeV and $\not{E}_T>3$ TeV yielding a reach of 8.1 (3.9) σ for $m_{\tilde{g}}=10\,(11)$ TeV assuming systematic uncertainty $\gamma=0.1$ for both signal and background. Two top tags are used for $m_{\tilde{g}}=10$ TeV while only one top tag is used for $m_{\tilde{g}}=11$ TeV. Compared to the same-sign di-lepton search in [74], which could reach ~ 9 TeV gluino assuming zero pile-up, our strategy could be sensitive to smaller production cross section and higher gluino mass.

2.7 Conclusions and Outlook

The discovery of a 125 GeV Higgs boson at the LHC is a milestone in particle physics. Yet the absence of new physics signals at the LHC so far makes the existence of such a light scalar confusing. A new energy frontier is needed to resolve mysteries related to electroweak symmetry breaking and to obtain a more definite answer to whether the weak scale is tuned. Understanding the physics cases and search challenges at a future collider serve as first steps to construct this next-generation machine.

In this chapter, we focus on reach of two stop simplified models at a future 100 TeV collider. Stops are key ingredients of low-energy SUSY and their mass scale directly tells us the degree of electroweak fine-tuning. In the first simplified model we study, stops are pair-produced and decay to top or bottom plus higgsinos. In the other model with gluino lighter than the stops, stops could be produced either in pairs or associated with a gluino. They will subsequently decay through gluinos to tops plus bino. The main new features of these simplified models are that the final states contain a lot of highly boosted top or

bottom jets with p_T above a TeV. To suppress the SM top backgrounds and for the second simplified model, SUSY backgrounds, we study and apply several simple jet observables such as track-based jet mass, N-subjettiness and mass drop. Combining these jet observables gives us effective tagging strategies for boosted tops and bottoms. We find that assuming 10% systematic uncertainties, the future 100 TeV collider can discover (exclude) stops with masses up to 6 (7) TeV with 3 ab^{-1} of integrated luminosity if the stops decay to higgsinos. In the second simplified model with light gluinos and the stops decay through gluinos, due to additional SUSY backgrounds from gluino pair production, a higher luminosity of about 30 ab^{-1} is needed to discover stops up to 6 TeV. We could use jet observables to tell apart simplified models with different LSPs, for instance, $\tilde{t}-\tilde{H}$ model and $\tilde{t}-\tilde{B}$ model. In addition, the top tagging allows us to improve the gluino reach close to 11 TeV with 3 ab^{-1} data.

This work is the first one to apply jet substructure techniques at a 100 TeV collider to study (supersymmetric) top partners, which indicates the level of electroweak fine-tuning, one of the major physics questions that a future hadron collider hopefully can give a qualitative answer. Studies on applying jet substructure to search for other possible new particles at a 100 TeV collider could be found in Ref. [119–122]. Jet substructure techniques provide us a powerful way to discriminate intricate new physics final states containing many hyper-boosted objects from messy SM and SUSY backgrounds, for which the traditional search strategies may not work. The jet tools could also help us distinguish between different new physics models and improve their reach significantly, exploring further the power of the future energy frontier.

While we focus on the study of mass reach of stops, the jet observables we study could be applied to search for other new particles such as fermionic top partners, which suffer from similar issues from hyper-boosted SM objects. They may also be used in exploring new mechanisms at future colliders such as measuring the gluino decays to test whether the minimal supersymmetric SM explains the Higgs mass [123]. In addition, the hyper-boosted top or bottom tagging may be further improved as discussed in Ref. [124].

Chapter 3

Using Gaia DR2 to Constrain Local Dark Matter Density and Thin Dark Disk

3.1 Introduction

Self-interacting dark matter (SIDM) is a kind dark matter (DM) that scatter amount themselves. In general, the self-interaction in SIDM causes dragging (Bullet Cluster) or a change in halo shapes that lead to constraints to their interaction strengths [125–132] (see also some model constraints in Refs. [38, 39, 133, 134]). Nonetheless, strongly self-interacting DM can exist as a subdominant fraction of all the DM. Sometimes this multispecies DM model is called partially interacting dark matter (PIDM). In particular, we are interested in a PIDM scenario that the SIDM species is dissipative. In general, SIDM does not necessarily lead to dissipative dark matter (DDM). Dissipation can be achieved, for example, by having the SIDM be composed of a heavy and light species. A recently studied partially interacting DDM scenario is the one that DM dissipates energy to form a (possibly thin) dark disk (DD), co-rotating with the baryonic disk [135, 136]. Let the total mass ratio of this subdominant

DD-forming DM to all of the DM in the Milky Way be

$$\epsilon_{\rm DD} = \frac{M_{\rm DDDM}}{M_{\rm DM}} \tag{3.1}$$

Possible effects of such a DD and variants of dissipative DM scenarios have been studied further in Refs. [5,39,137–154]. The formation of the disk from DM self-interactions is highly debatable [155] and numerical simulations using a cooling prescription (as in Ref. [156]) are still absent. In spite of these uncertainties regarding its formation, it is still worthwhile to use the stellar data to test the simplest possibility of a thin DD aligned with the baryonic disk and parametrized by only two parameters: the surface density, Σ_{DD} and a scale height, h_{DD}. This has been carried out using *Hipparcos* data in Ref. [146] and *Tycho–Gaia* Astrometric Solution (TGAS), a joint solution combining *Tycho*-2 catalog with early *Gaia* data, in Ref. [1].

While there is no known DM-Standard Model matter interaction, gravity is the only interaction known between DM and visible matter. Gravitational effect of DM can be measured by observing the phase-space distributions of stars visible from telescopes. These stars are known as "tracers" as they trace the gravitational potential generated by the combine visible and dark sector. Observing tracers can also probe dark matter astrophysical substructures, such as the DD, as they often generate unique shapes of gravitational potential. As a result, stellar phase space distributions are probes of the properties of DM at the particle level and are of high phenomenological interest. There are many methods to measure gravitational potential from stellar phase space distribution, for example, the K_z force law discussed at Chapter 1 in Eq. (1.1) and (1.2). Another popular method is to integrate the collisionless Boltzmann equation[1] over all the velocities to obtain a set of moment equations that are the Jeans equations [157]. For example, we can use cylindrical coordinates (r, ϕ, z) to describe the Milky Way galactic plane and focus on the Jeans equation in the z direction,

$$\frac{1}{r\nu}\frac{\partial}{\partial r}\left(r\nu\sigma_{rz}^2\right) + \frac{1}{r\nu}\frac{\partial}{\partial \phi}\left(\nu\sigma_{\phi z}^2\right) + \frac{1}{\nu}\frac{\partial}{\partial z}\left(\nu\sigma_z^2\right) = -\frac{\partial \Phi}{\partial z}, \tag{3.2}$$

where ν is the number density of tracer stars, $\sigma_{rz}^2(\sigma_{\phi z}^2)$ are the off-diagonal entries in the velocity dispersion tensor that couple radial (axial) and vertical motions, and σ_z^2 is the

[1]To be introduced in Eq. (3.10) in Section 3.4.1.

vertical velocity dispersion (the diagonal zz component of the velocity dispersion tensor) and Φ is the gravitational potential. Properties of stellar phase space distribution can be studied order by order in the moments using this method. The first term, usually referred to as the "tilt" term, is sometimes negligible for study limited at small height z from the galactic plane [158] (as in the case of this work). The second term, the so-called "axial" term, is also sometimes negligible when assuming approximate axisymmetric of the system.

In this chapter, we use the distribution function method to study the stellar distributions. Also from the Boltzmann equation, one can find the partial differential equation the distribution function ($f(\boldsymbol{r}, \boldsymbol{v})$) obeys. For example, along the z-direction of the Milky Way,

$$w \frac{\partial f}{\partial z} - \frac{\partial \Phi}{\partial z} \frac{\partial f}{\partial w} = 0, \tag{3.3}$$

where w is the vertical velocity of the stars which we will define in a more clear manner in Section 3.3.2. The stellar distribution's relation to the gravitational potential can be studied by directly integrating Eq. (3.3),

$$\nu(z) \ = \int_{-\infty}^{\infty} dw \, f(z, w). \tag{3.4}$$

We will develop the details in the main text in Section 3.4.1. More different analysis methods and recent related literature can be found in Section 3.2.

In this chapter, we work with the second *Gaia* data release (DR2) [54] to estimate the local DM density as well as constrain thin DD models assuming that the DD is aligned with the baryonic disk. We follow the method in Refs. [1, 146, 159] and use A, F and early G dwarf stars in the *Gaia* catalog as the tracers. In Section 3.2, we briefly discuss the *Gaia* telescope and data release, and list some of the literature attempts to measure Milky Way gravitational potential. In Section 3.3, we discuss the details of *Gaia* DR2 and the empirically determined survey selection function (Section 3.3.1), which we use to construct the vertical number density profile and midplane velocity distribution for each tracer population in Sections 3.3.2 and 3.3.3 respectively. Our fiducial analysis is described in Section 3.4. We use the 1D distribution function method summarized in Section 3.4.1 to construct the equilibrium number density for the parameters of our mass model described in Section 3.4.2.

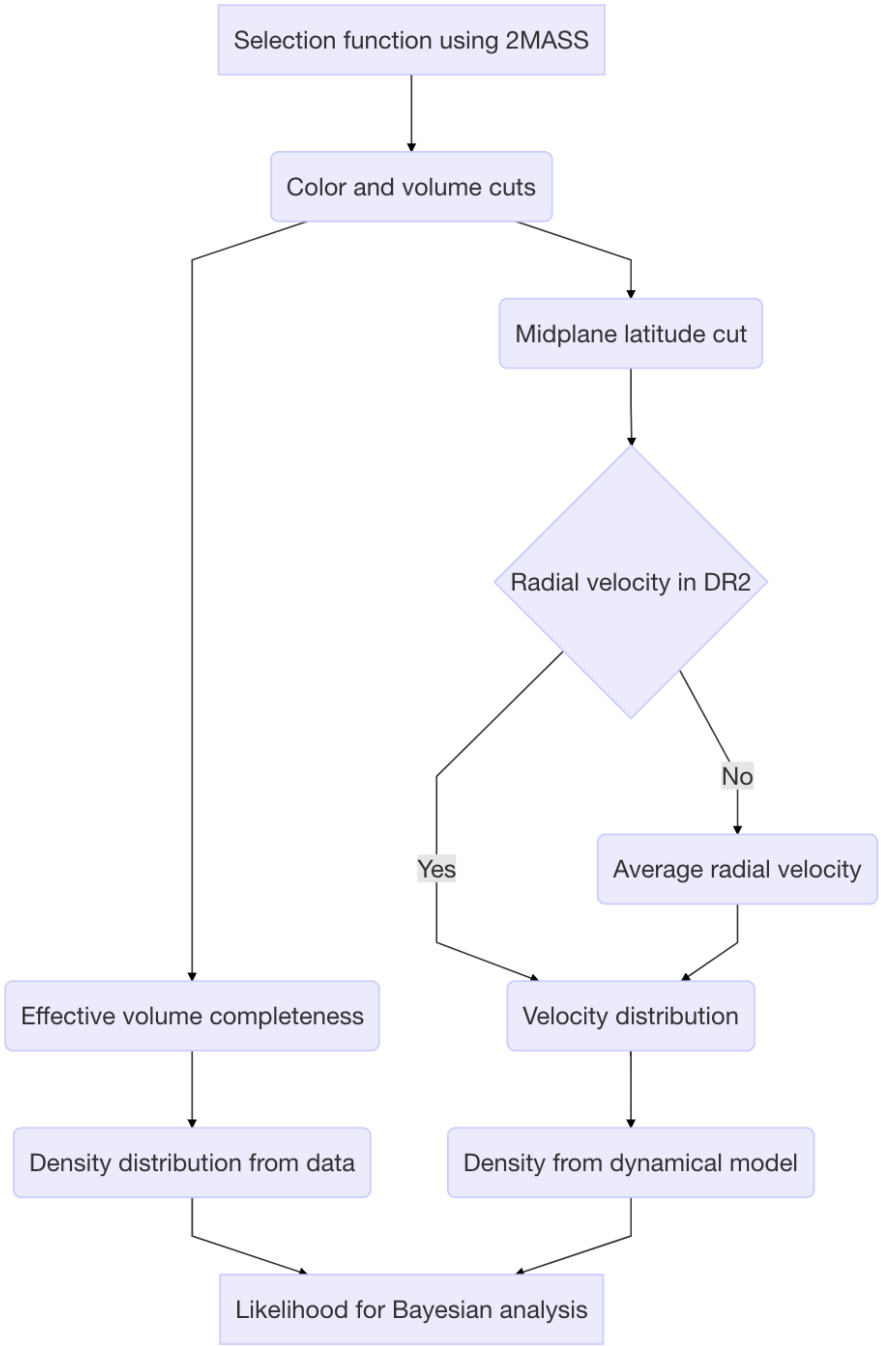


Figure 3.1: Flowchart of our analysis.

In Sections 3.4.3 and 3.4.4, we introduce a Bayesian framework for comparing our predicted density with data for each tracer population while taking into account the uncertainties due to potential non-equilibrium effects. The important steps of our analysis are outlined as a flowchart in Fig. 3.1. While our method is not new, we obtain interesting results, some of which are quite different from those based on TGAS. We present our results for the local DM content using *Gaia* DR2 in Section 3.5.1 and 3.5.2, and list various sources of systematic uncertainties in the context of our method in Sections 3.6.1-3.6.3. Differences between constraints derived using DR2 and TGAS are discussed in Section 3.6.4. We conclude and comment on future directions in Section 3.7.

This chapter is based on the work with Jatan Buch and JiJi Fan in Ref. [160].

3.2 Gaia Telescope and Measurement of Local Dark Matter Density

The second release of data collected by the European Space Agency's *Gaia* telescope provides the positions and proper motions, with unprecedented precision, of more than one billion sources in the Milky Way (MW) [54, 161–168]. With the release of line-of-sight velocities for about seven million stars, DR2 also allows, for the first time, a dynamical analysis with a self-consistent measurement of the 6D phase space for a stellar population.

DR2 presents an exciting opportunity to use the vertical velocity and number density distributions of different populations of stars that trace the gravitational potential for precisely determining the total matter density, including baryons and dark matter (DM), in the local solar neighborhood. Significant progress has been made in modeling the local baryon budget (interstellar gas, stars, stellar remnants) and its uncertainties [2, 4, 5, 169] since Oort's early estimate [19] of the baryon density. In this work, we primarily focus on the 1D distribution function method developed by Refs. [170–174] and used by Refs. [159,175] to constrain the local DM density with data from the *Hipparcos* satellite [176]. However, the approximations of isothermality and decoupling of radial and vertical motions in this

method are only valid up to scale height $z \sim 1$ kpc. Therefore, for using tracer data at high z, Refs. [177, 178] adopt the more general moment-based method to estimate the DM density. A non-parametric formulation of the moment-based method, described by Ref. [179] and implemented in Ref. [180], uses SDSS/SEGUE G stars in a heliocentric cylinder with $R \sim 1$ kpc and 0.5 kpc $\lesssim |z| \lesssim 2.5$ kpc, grouped by age, namely α-young and α-old stars, as tracers. Ref. [169] also uses SDSS/SEGUE G star data between 4 kpc $\lesssim R \lesssim 9$ kpc and 0.3 kpc $\lesssim |z| \lesssim 3$ kpc to constrain the stellar and DM density through action-based distribution function modeling [181]. Their analysis incorporated the age information of tracers in a more sophisticated manner by constructing mono-abundance populations [182] that consist of stars with similar elemental abundances. The above discussion is by no means an exhaustive review of the different attempts at measuring the local DM density (most notably, it doesn't address dynamical measurements made by, for example, Refs. [183, 184]); instead, we refer interested readers to Refs. [3, 185].

3.3 Data Selection

The *Gaia* DR2 catalog contains ∼1.7 billion sources, among which ∼1.3 billion sources have a five-parameter astrometric solution: $(\alpha, \delta, \mu_{\tilde{\alpha}}, \mu_\delta, \varpi)$, representing positions and proper motions along the right ascension and declination, and parallax respectively. We emphasize that for DR2, the parallaxes and proper motions are based solely on *Gaia* measurements, unlike DR1 which depends on the *Tycho*-2 catalog [54, 162]. DR2 also provides photometric data in three passbands, G, $G_{\rm BP}$, and $G_{\rm RP}$, for a majority of the sources in the range $3 \lesssim G \lesssim 21$ [163, 164]. Another new feature in DR2 is the line-of-sight radial velocities, v_r, for ∼7.2 million stars brighter than $G_{\rm RVS} \sim 12$ [165–167].

Despite a significant increase in statistics, the DR2 catalog is still incomplete for stars with $G < 12$. We calculate the completeness fraction for the DR2 catalog by comparing it with the *Two Micron All-Sky Survey* (*2MASS*) Point Source Catalog [186] that is 99% complete down to its faint magnitude limit $J = 15.8$ over almost the entire sky. The *2MASS* catalog also provides the angular positions, (α, δ), for each source along with its J and K_s

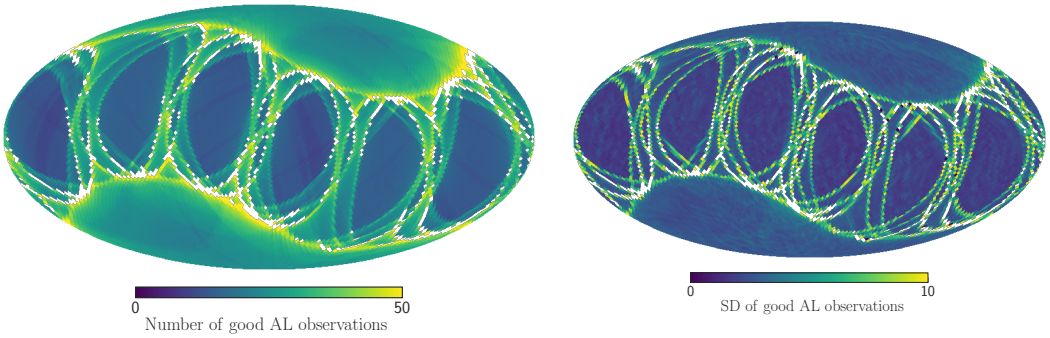


Figure 3.2: Skymaps showing the number (left) and variance (right) of good AL observations in 3.36 deg^2 ($N_{\rm side} = 2^5$) HEALPix pixels. The white regions are the parts of the sky which do not pass our selection cuts defined in the main text.

magnitudes, which we use to categorize stars in DR2 by stellar type.

We query the *Gaia* archive[2] for DR2 cross-matched with the *2MASS* catalog, requiring the apparent magnitude $J < 13.5$, thereby cutting away stars that are either too dim for the main sequence, or too distant from the Sun.[3] As we discuss in the following section, the star counts in the *2MASS* catalog with $J < 13.5$ are then used to determine the selection function, $S(J, J-K_s, \alpha, \delta)$, defined as the fraction of stars at a given $(J, J-K_s, \alpha, \delta)$ in the sky that are contained in the DR2 catalog.

3.3.1 Preliminaries: selection function, color and volume cuts

Since there has been no official release by the *Gaia* collaboration, we follow the procedure introduced in Ref. [187] to determine the selection function for the cross-matched data set. As a first step, using the `gaia_tools` package,[4] we identify parts of the sky that satisfy,

(i) Mean number of along-scan (AL) observations ≥ 8.5;

(ii) Spread in the number of AL observations ≤ 10.

[2]https://gea.esac.esa.int/archive/
[3]In our selected volume, the apparent magnitude of all tracer stars satisfy $J < 12$.
[4]https://github.com/jobovy/gaia_tools

After these cuts, 95.6% of the sky remains as shown in Fig. 3.2. The overall completeness in these parts of the sky is largely isotropic, implying that our selection function only has a negligible dependence on sky position. Thus, for the rest of our analysis, we use the approximation $S(J, J-K_s, \alpha, \delta) \approx S(J, J-K_s)$. A similar approximation has been adopted for analyzing TGAS data in Ref. [187], in which stronger cuts were used to identify the 'good' parts of the sky.

Although the uncertainties of astrometric parameters in DR2 are an order of magnitude improvement over TGAS, there are still several systematic effects that we need to include in our data analysis. As suggested by Ref. [188], we add 0.03 mas to the reported parallax to account for the global offset. Following Ref. [168], we also add in quadrature a systematic uncertainty of ± 0.1 mas and ± 0.1 mas/yr respectively to the reported values of parallax and proper motion uncertainties. We check that our analysis is not sensitive to the exact values of these numbers, only their order-of-magnitude estimate.

Next, we identify stellar populations that: a) are tracers of the local galactic potential (see Sec. 3.6 of Ref. [3] for a recent overview), and b) show a reasonable change in their number densities within the solar neighborhood (defined below). The most important factor in selecting tracer stars is their sensitivity to disequilibria, which could result in incompatible $\rho_{\rm DM}$ measurements [189]. While there is some disagreement in the literature about which stars, old [159, 190] or young [180], are more suitable for an equilibrium analysis, we follow Ref. [180] in choosing younger A (A0-A9), F (F0-F9), and early G (G0-G3) dwarf stars (simply stars henceforth) which have lower scale heights and consequently shorter equilibration timescales, instead of older stars, in our analysis.

We use the mean dwarf stellar locus from Ref. [191] to define each stellar type t based on its absolute magnitude $\overline{M}_{J^t}$ and color $\left(\bar{J}^t - \bar{K}_s^t\right)$,

$$J^t + 5\log_{10}(\varpi^t\,(\text{mas})) - 10 \in \{\overline{M}_{J^t,\,\min}, \overline{M}_{J^t,\,\max}\} \tag{3.5}$$

$$\left(J^t - K_s^t\right) \in \{\left(\bar{J}^t - \bar{K}_s^t\right)_{\min}, \left(\bar{J}^t - \bar{K}_s^t\right)_{\max}\} \tag{3.6}$$

and include all A, F, and early G stars in a heliocentric cylinder with radius $R = 150$ pc

and half-height $z = 200$ pc,

$$\left| \left(\frac{1}{\varpi} \right) \cos b \right| < 150\,\mathrm{pc}, \qquad \left| \left(\frac{1}{\varpi} \right) \sin b \right| < 200\,\mathrm{pc} \tag{3.7}$$

where all the un-barred quantities are data from our cross-matched catalog. In Eq. 3.6 and 3.7, we use $(1/\varpi)$ as an estimator of the distance, where ϖ is the noisy parallax reported in the DR2 catalog. We treat it as an unbiased estimator since our tracer sample contains nearby stars within a distance, $(1/\varpi) < 250$ pc, and a low parallax uncertainty, $\sigma_\varpi \lesssim 0.1$ mas. We discuss the uncertainties in our results due to the choice of cuts in Section 3.6.1.

3.3.2 Vertical number density distribution

To make any meaningful inferences about stellar dynamics, we require the volume complete number density of stars. Since *Gaia* DR2 is incomplete, the number density constructed from data must be corrected for the survey selection function (determined in Sec. 3.3.1) to reflect the underlying *true* Poisson-distributed density. Ref. [187] accomplishes this by deriving a bin-by-bin normalization for each stellar population, defined as the *effective volume completeness*, under the assumption that all observed stars are samples from an inhomogeneous Poisson process. We follow the same procedure and summarize its important details in Appendix B.1.2.

We calculate the effective volume completeness for our DR2 tracer populations using the `gaia_tools` package and plot it as a function of scale height in Fig. 3.3. We also include the effective volume completeness for TGAS data as a reference, and note that the DR2 sample is significantly more complete.

There are 4445 A, 37707 F, and 43332 early G stars in the solar neighborhood defined by our heliocentric cylinder. The volume complete vertical number density for the i^{th} tracer population, ν_i^{data} shown in Fig. 3.4, is obtained by dividing the number counts with the effective volume completeness in each z bin. We choose 20 pc as the bin size based on parallax uncertainties as discussed in Appendix B.2. However, varying the bin size doesn't significantly affect the results of our analysis. We also present a comparison of star counts in

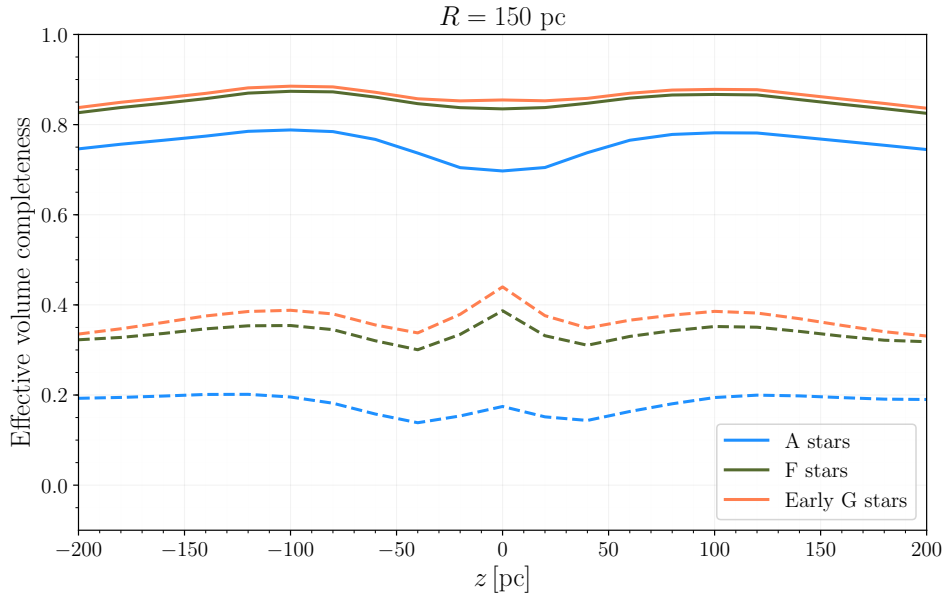


Figure 3.3: Effective volume completeness for each stellar type. The completeness of DR2 (solid) is significantly improved as compared to TGAS (dashed) for A (blue), F (green), and early G (orange) stars.

Data set		*Gaia* DR2		TGAS	
Type	Subtype	Total	Midplane	Total	Midplane
A	A0-A9	4445	321	1729	182
F	F0-F9	37707	2253	16789	1308
Early G	G0-G3	43332	2188	18653	1205

Table 3.1: Star counts in DR2 and TGAS catalogs for the heliocentric cylinder and the midplane region ($|b| < 5^{\circ}$) inside it.

the full volume and in the midplane (defined to be the region with $|b| < 5^{\circ}$ in the cylinder) between DR2 and TGAS in Table 3.1.

The number density uncertainty for the k^{th} z bin, $\left(\sigma^2_{\ln \nu_i}\right)^{\text{data}}$, is obtained by adding in quadrature the statistical uncertainty, $\sqrt{N_k}$ (see Eq. B.12), and a 3% systematic uncertainty due to dust extinction. We expect the dust extinction to be important in the visible spectrum

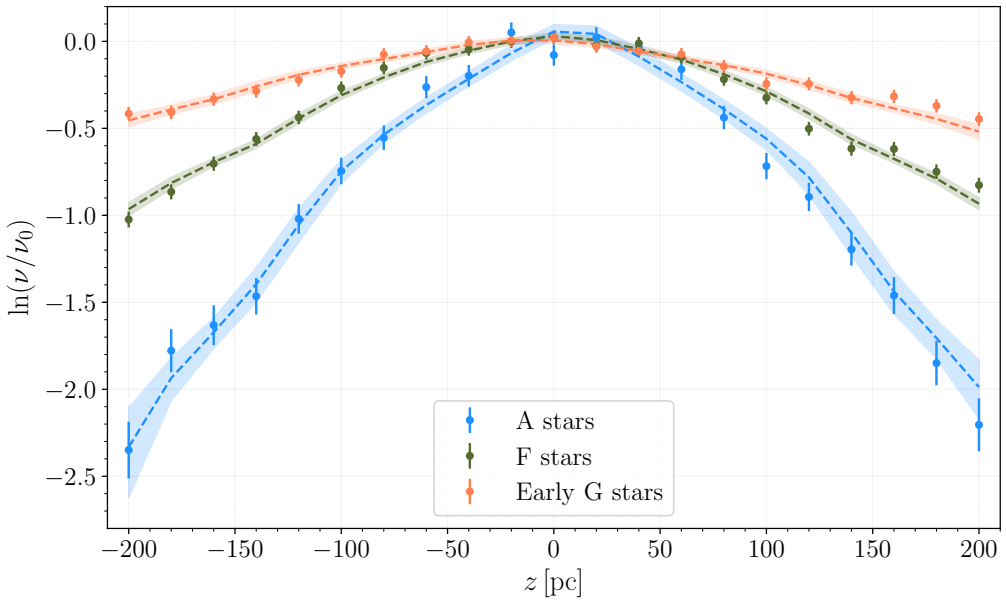


Figure 3.4: Binned vertical number density profiles for A, F, and early G stars. Also shown is the median predicted density (dotted line) and the 68% confidence interval (shaded region) for each tracer population obtained from our dynamical analysis described in Sec. 3.4.

such as the B and V colors used in *Hipparcos* catalog, or the $G_{\rm BP}$ and $G_{\rm RP}$ used in DR2. However, colors in the infrared spectrum, *i.e.* the J and K_s colors used in our cross-matched DR2-*2MASS* catalog, are associated with longer wavelengths and therefore much less affected by galactic dust. Ref. [187] finds that the effect of dust reddening on the number density of stars in the solar neighborhood defined using J and K_s is only $\lesssim 3\%$ and mostly affects the overall normalization. Thus, instead of doing a full dust analysis, we conservatively adopt an overall 3% correction for all three types of stars. We also find that this is only a sub-dominant uncertainty and doesn't significantly affect our constraints.

3.3.3 Midplane velocity distribution

The last ingredient we need from the data is the vertical velocity distribution in the midplane, *i.e.* in the region near $z = 0$. The vertical velocity of a star is approximated by,

$$w = w_\odot + \frac{\kappa \mu_b}{\varpi} \cos b + v_r \sin b, \tag{3.8}$$

where $w_\odot$ is the Sun's vertical velocity that we determine by fitting a Gaussian distribution to the data, $\kappa = 4.74$ km yr s $^{-1}$ is a unit conversion constant, μ_b is the proper motion along the galactic latitude b in mas/yr, ϖ is the parallax in mas, and v_r is the heliocentric radial velocity in km/s. Following the discussion below Eq. 3.7, we note that for regions near the midplane, $(1/\varpi)$ is an unbiased estimator for distance in the formula for vertical velocity.

The 'midplane region' can be defined by imposing a cut on either the: a) galactic latitude $|b|$, or b) height $|z|$.[5] In our analysis, we follow Refs. [1, 146] in choosing $|b| < 5^\circ$ as our midplane cut, and are left with 310, 2213 and 2166 A, F and early G stars respectively. We use radial velocities reported in DR2, when available, substituting v_r by its mean value otherwise,

$$\langle v_R \rangle = -u_\odot \cos l \cos b - v_\odot \sin l \sin b - w_\odot \sin b, \tag{3.9}$$

where $u_\odot = 11.1 \pm 0.7^{\text{stat}} \pm 1.0^{\text{sys}}$ km/s and $v_\odot = 12.24 \pm 0.47^{\text{stat}} \pm 2.0^{\text{sys}}$ km/s as the x- and y-component of the Sun's peculiar velocity [192]. We also note that as $\sin b \ll 1$ (in radians) in our midplane region, v_r only has a subdominant contribution to w in Eq. 3.8, and the velocity uncertainties have a negligible effect on our results. We determine the Sun's vertical velocity, $w_\odot$, by considering the weighted mean of the best-fit Gaussian distribution to each tracer population's vertical velocity data in the midplane region. Our estimate, $\overline{w}_\odot = 6.9 \pm 0.2$ km/s, is consistent within 1σ of the value quoted in Ref. [192].

Subtracting $\overline{w}_\odot$ from the stars' vertical velocity, we find the distributions are roughly symmetric about $w = 0$. The resultant normalized midplane vertical velocity distribution $f_0(w)$ with a w-bin size of 1.5 km/s (see Appendix B.2 for more details about this choice) is plotted in the left panel of Fig. 3.5. We consider the asymmetry between star counts

[5] At larger b and consequently larger z, the kinematically hotter stars broaden the distribution [159]. Meanwhile, simply choosing stars with $z = 0$ yields poor statistics.

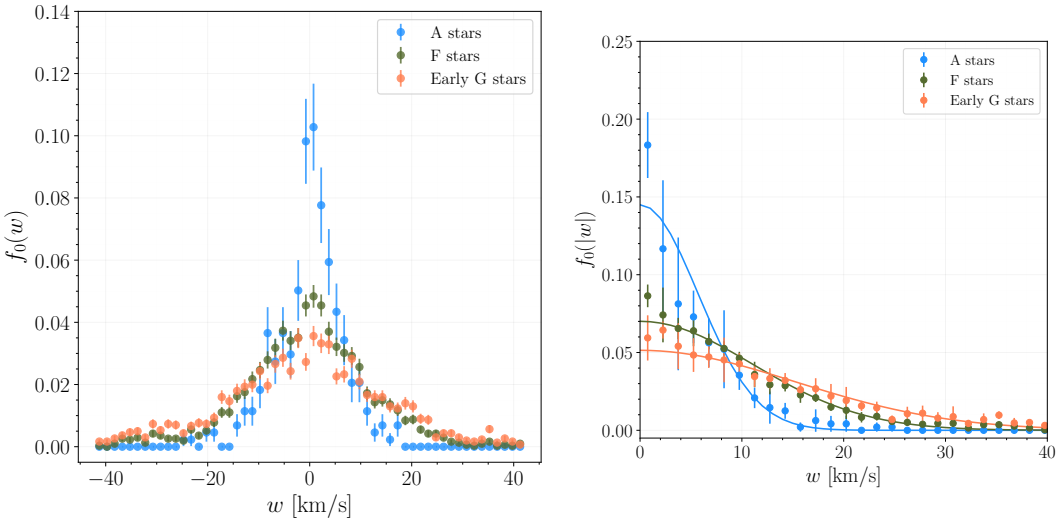


Figure 3.5: Midplane velocity distributions of A, F, and early G stars after subtracting $w_\odot$ (left). The best-fit Gaussian distribution to $f_0(|w|)$ with error bars that include contributions from the statistical uncertainty due to Poisson error and the asymmetry in $-|w|$ and $+|w|$ bins (right).

in $-|w|$ and $+|w|$ bins to be the systematic uncertainty, which may be attributed to non-equilibrium effects. We illustrate the magnitude of this uncertainty in the right panel of Fig. 3.5 by adding it in quadrature with the statistical error for every w bin. In practice, however, we propagate these errors into the uncertainty of the prediction density as described in Sec. 3.4.3.

We also explore the possibility of using the z-cut [193] in Appendix B.3 by including the radial velocities reported in DR2. Unfortunately, DR2 only contains radial velocities for approximately 2% of A stars, 53% of F stars, and 62% of early G stars within $|z| < 20$ pc. We check that the percentage of tracers with radial velocity doesn't change significantly for values of $|z| \lesssim 50$ pc. In that case, only including stars with available v_r could potentially introduce a selection bias, while approximating v_r by its mean value might result in large errors at higher b (even at low z). Thus, defining the midplane region using a z-cut isn't a viable option currently, but that could change with future data releases.

3.4 Fiducial Analysis

The main objective of our analysis is to infer the matter content of the solar neighborhood by fitting data observed by *Gaia* to a dynamical model. Sec. 3.4.1 summarizes the 1D distribution function method [170–174] we adopt for constructing the equilibrium number density of a tracer population, while the parameters of our gravitational potential model are described in Sec. 3.4.2. We outline the Bayesian statistical framework used for constraining these parameters in Secs. 3.4.3 and 3.4.4.

3.4.1 Equilibrium density modeling

A self-gravitating stellar population with phase space distribution function (DF), $f(\mathbf{x}, \mathbf{v})$, satisfies the collisionless Boltzmann equation (CBE),

$$\frac{Df}{Dt} \equiv \frac{\partial f}{\partial t} + (\nabla_{\mathbf{x}} f) \,.\, \mathbf{v} - (\nabla_{\mathbf{x}} \Phi) \,.\, (\nabla_{\mathbf{v}} f) = 0. \tag{3.10}$$

where $\mathbf{x}$ and $\mathbf{v}$ are the positions and velocities respectively, and Φ is the gravitational potential. Assuming axisymmetry for the local solar neighborhood, we use cylindrical polar coordinates for the rest of our analysis.

At this stage, we make two critical assumptions that allow us to drop the partial time derivative term in Eq. 3.10: a) populations of all tracer stars are in equilibrium,[6] , and b) the potential is time-independent, since the timescale for galactic processes is much longer than *Gaia*'s mission lifetime. Moreover, since we're interested in the dynamics of stars very close to the galactic plane ($|z| \lesssim 0.5$ kpc), we approximate the DF to be of the form,[7]

$$f(\mathbf{x}, \mathbf{v}) \equiv f_{r,\phi}(r, v_r, \phi, v_\phi)\, f_z(z, v_z). \tag{3.11}$$

Separability of the DF implies that the motion of a stellar population i in the z-direction is independent of $\{R, \phi\}$, and follows the 1D CBE,

$$w \frac{\partial f_{z,i}}{\partial z} - \frac{\partial \Phi}{\partial z} \frac{\partial f_{z,i}}{\partial w} = 0. \tag{3.12}$$

[6] This assumption is central to *all* dynamical analyses of stars in the MW. We discuss its validity at some length in Sec. 3.6.

[7] In the language of Jeans modeling, this follows from the observed smallness of the so-called 'tilt term' that couples the vertical and radial motions.

which has a general solution of the form, $f_{z,i}(z,w) = F_z(E_z) \equiv F_z\left(w^2/2 + \Phi(z)\right)$, where the vertical energy is defined as, $E_z = \frac{1}{2}w^2 + \Phi(z)$.

We integrate the DF over vertical velocity (or equivalently, energy) to obtain the stellar number density [173],

$$\begin{aligned}\nu_i(z) &= \int_{-\infty}^{\infty} dw\, f_{z,i}(z,w) = \int_{-\infty}^{\infty} dw\, f_{z,i}(E_z) \\ &= 2\int_0^{\infty} dw f_{z=0,\,i}(0, \sqrt{w^2 + 2\Phi(z)}) \\ &= 2\int_{\sqrt{2\Phi(z)}}^{\infty} \frac{f_{0,\,i}(|w|)\; w\; dw}{\sqrt{w^2 - 2\Phi(z)}},\end{aligned} \tag{3.13}$$

where $f_{0,\,i}(|w|)$ is the midplane (absolute) velocity distribution of stars of a tracer population i, which we determine from *Gaia* data using the procedure described in Sec. 3.3.3.

3.4.2 Local matter content: baryons, halo DM, and a thin DD

We calculate the gravitational potential due to the total mass density, $\rho_{\rm tot}$, in the solar neighborhood through the Poisson equation,

$$\nabla^2\Phi = \frac{\partial^2\Phi}{\partial z^2} + \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\Phi}{\partial r}\right) = 4\pi G\rho_{\rm tot}(z), \tag{3.14}$$

where the radial term, $\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\Phi}{\partial r}\right)$, effectively contributes a constant mass density[8] with a value $(3.4 \pm 0.6) \times 10^{-3}$ $\mathrm{M}_\odot/\mathrm{pc}^3$ determined from the TGAS data [194]. We have also assumed that the R and z components of the potential can be decoupled such that, $\Phi(R,z) = \Phi(R) + \Phi(z)$.

The total mass density contains contributions from N_b baryon components, DM in the halo, and other gravitational sources such as a thin DD. The baryon mass density, ρ_b, is given by the Bahcall model that consists of a set of isothermal components for gas, stars, and star remnants [195–197],

$$\rho_b(z) = \sum_{i=1}^{N_b} \rho_i(0) e^{-\Phi(z)/\sigma_{z;i}^2} \tag{3.15}$$

[8]For an axisymmetric system, the radial term is related to Oort's constants. Strictly speaking, the Oort's constants and consequently the radial term also depend on z. However, since our tracers only explore a small volume close to the midplane, the variation is smaller than the measurement uncertainty [178].

Baryonic components	$\rho(0)$ [$M_\odot/pc^3$]	σ_z [km/s]
Molecular gas (H_2)	0.0104 ± 0.00312	3.7 ± 0.2
Cold atomic gas ($H_I(1)$)	0.0277 ± 0.00554	7.1 ± 0.5
Warm atomic gas ($H_I(2)$)	0.0073 ± 0.0007	22.1 ± 2.4
Hot ionized gas (H_{II})	0.0005 ± 0.00003	39.0 ± 4.0
Giant stars	0.0006 ± 0.00006	15.5 ± 1.6
$M_V < 3$	0.0018 ± 0.00018	7.5 ± 2.0
$3 < M_V < 4$	0.0018 ± 0.00018	12.0 ± 2.4
$4 < M_V < 5$	0.0029 ± 0.00029	18.0 ± 1.8
$5 < M_V < 8$	0.0072 ± 0.00072	18.5 ± 1.9
$M_V > 8$ (M dwarfs)	0.0216 ± 0.0028	18.5 ± 4.0
White dwarfs	0.0056 ± 0.001	20.0 ± 5.0
Brown dwarfs	0.0015 ± 0.0005	20.0 ± 5.0

Table 3.2: Bahcall model consisting of midplane densities and velocity dispersions for N_b baryonic components adapted from Ref. [1]. The values and uncertainties, both observational and estimated, for all components have been compiled from Refs. [2–5]

where each isothermal component is characterized by its midplane density, $\rho(0)$, and vertical velocity dispersion, σ_z as shown in Table 3.2.

We approximate the contribution of DM density from the smooth halo near the disk, $\rho_{\rm DM}$, to be constant. As shown by Eq. (28) in Ref. [178], the DM density at or below 200 pc is equal to that in the midplane up to a 2% correction.

In models with a thin DD, we assume that the DD is isothermal, axisymmetric, and perfectly aligned with the baryonic disk. Following Ref. [198], we choose the parametrization of the thin DD density to be,

$$\rho_{DD}(z) = \frac{\Sigma_{DD}}{4h_{DD}} \mathrm{sech}^2\left(\frac{z}{2\,h_{DD}}\right), \tag{3.16}$$

where Σ_{DD} is the surface density and h_{DD} is the disk height. A thin DD aligned with the baryonic disk contributes an additional source of attractive potential, which pulls matter

towards the midplane (see Section 2.2 of Ref. [146] for an example with a toy model). This results in a narrowed *pinched* density profiles of tracers, as illustrated in Fig. 3.6.

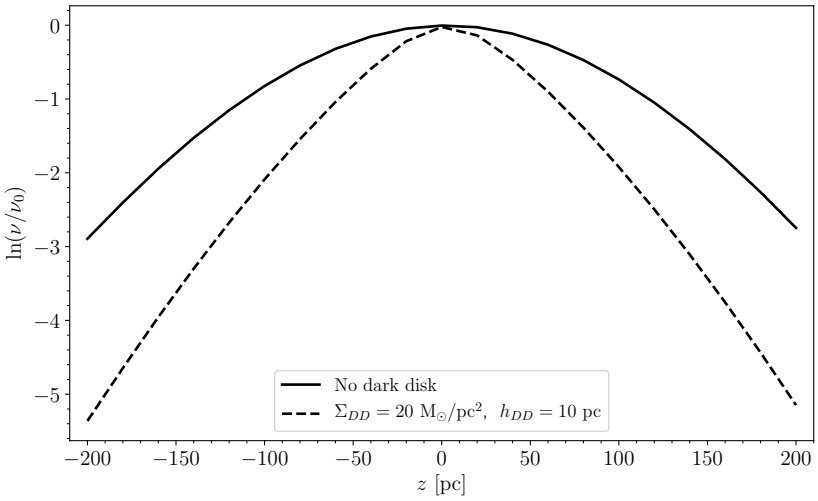


Figure 3.6: The predicted number density of a tracer in a model containing a thin DD with surface density $\Sigma_{DD} = 20$ M$_\odot$/pc^2 and scale height $h_{DD} = 10$ pc (dashed). For comparison, we also plot the prediction of a model with the same matter content but without the thin DD (solid).

Thus, to obtain the local gravitational potential, we plug the total matter density, $\rho_{\rm tot}$, given by,

$$\rho_{\rm tot}(z) = \sum_{i=1}^{N_b} \rho_i(0) e^{-\Phi(z)/\sigma_{z;i}^2} + \rho_{\rm DM} + \rho_{DD}(z). \tag{3.17}$$

into Eq. 3.14 and solve the resulting nonlinear, second-order differential equation numerically with `scipy.integrate.ODEint`. We also explicitly check that our results agree with those obtained by Refs. [1, 146] using the iterative solver method.

To summarize, we calculate the equilibrium number density in the solar neighborhood for each tracer population, *i.e.* A, F, and early G stars, by integrating its midplane velocity distribution (constructed from data) as a function of the parameters of our mass model using Eq. 3.13. We also apply a Gaussian kernel smoothing to the result to approximate the effect of parallax uncertainties that smear the exact positions of stars. However, since the parallax

uncertainties in DR2 are significantly reduced as compared to TGAS, this procedure only has a negligible effect on the predicted density.

3.4.3 Likelihood, model uncertainties, and priors

Our model $\mathcal{M}$ is characterized by $\boldsymbol{\theta} = \{\boldsymbol{\psi}, \boldsymbol{\xi}\}$, such that $\boldsymbol{\psi} = \{\rho_{\rm DM}, \Sigma_{DD}, h_{DD}\}$ are our parameters of interest, and $\boldsymbol{\xi}$ are the nuisance parameters which include: midplane densities, $\rho_k(0)$, and velocity dispersion, $\sigma_{z;k}$, for each baryonic component in the Bahcall model; overall normalization for each stellar population, N_ν; height of sun above the midplane, $z_\odot$.

For each tracer population, we use the likelihood, $p_\nu(d|\mathcal{M}, \boldsymbol{\theta})$, to fit the number density constructed from data with our model prediction in the presence of statistical and systematic uncertainties. Since each number density bin contains a large number of stars ($\mathcal{O}(1000)$ for F and early G stars), the likelihood can be reasonably approximated by a Gaussian distribution,[9]

$$p_\nu(d|\mathcal{M}, \boldsymbol{\theta}) = \prod_{i=1}^{N_z} \frac{1}{\sqrt{2\pi\sigma^2_{\ln\nu_i}}} \exp\left(-\frac{(\ln(N_\nu\,\nu_i^{\rm mod}(\boldsymbol{\theta})) - \ln\nu_i^{\rm data})^2}{2\,\sigma^2_{\ln\nu_i}(\boldsymbol{\theta})}\right), \tag{3.18}$$

where N_z is the number of z bins, $\nu_i^{\rm mod}$ is the prediction of a model with parameters $\boldsymbol{\theta}$, and $\nu_i^{\rm data}$ is *volume complete* number density constructed from data, as described in Sec. 3.3.2. Unlike Ref. [1], we do *not* multiply the likelihood functions for different stellar populations in our analysis since doing so assumes all populations are similar and trace the same galactic potential independently. This is a rather simplified assumption which ignores the evolution history of different stellar types. We comment more on this in Section 3.5.1.

The total uncertainty, $\sigma^2_{\ln\nu_i}$, is obtained by adding in quadrature the data and the prediction uncertainties,

$$\sigma^2_{\ln\nu_i}(\boldsymbol{\theta}) = \left(\sigma^2_{\ln\nu_i}(\boldsymbol{\theta})\right)^{\rm mod} + \left(\sigma^2_{\ln\nu_i}\right)^{\rm data}. \tag{3.19}$$

[9] We note that surveys like *Gaia* (*SDSS-SEGUE*) measure astrometric (spectrophotometric) parameters of individual stars. Thus, an ideal likelihood analysis should involve star-by-star predictions for these parameters drawn from a generative process that accounts for the survey selection function. The forward modeling approach of Refs. [169,182,199], for example, accomplishes this for several different scenarios. For DR2, such an analysis has been carried out by Ref. [200] after this work appeared on arXiv, and their results agree with ours.

The data uncertainty is discussed in Sec. 3.3.2, whereas the model uncertainty originates from uncertainties in the velocity profile $f_{z=0}(|w|)$. It consists of two sources: *a)* the statistical uncertainty due to the finite sample size, and *b)* the systematic uncertainty due to possible non-equilibrium effects, which we characterize by the difference between $f_{z=0}(w > 0)$ and $f_{z=0}(w < 0)$ following the treatment in Ref. [1].

Direct error propagation from uncertainties of $f_{z=0}(|w|)$ is difficult due to the large number of parameters and their correlations involved. Instead, we estimate the errors by bootstrap resampling. The bootstrap is a technique that extracts estimates for the mean and standard deviation of a given data set by repeated random sampling with replacement. For each stellar type, the raw midplane star data sets are bootstrapped multiple times to generate a suite of velocity distributions. For every velocity distribution, we use Eq. (3.13) to derive a density distribution, and estimate the statistical uncertainty, $\left(\sigma^2_{\nu_i}(z)\right)^{\rm mod,\,stat}$, as the bin-by-bin variance in the suite of density distributions. More details of the procedure are deferred to Appendix B.4.

We approximate the systematic uncertainty, $\left(\sigma^2_{\nu_i}(z)\right)^{\rm mod,\,sys}$, by computing the difference between number densities predicted using the velocity distributions $f^{(w>0)}_{z=0}$ and $f^{(w<0)}_{z=0}$ for every unique value of the gravitational potential,

$$(\sigma_\nu)^{\rm mod,\,sys} \approx |\ln \nu^{(w>0)}(z) - \ln \nu^{(w<0)}(z)|. \tag{3.20}$$

The total uncertainty for the predicted number density, in every z bin, is then given by,

$$\left(\sigma^2_{\nu_i}(z)\right)^{\rm mod} = \left(\sigma^2_{\nu_i}(z)\right)^{\rm mod,\,stat} + \left(\sigma^2_{\nu_i}(z)\right)^{\rm mod,\,sys}. \tag{3.21}$$

We find that the systematic uncertainties dominate over statistical ones in our analysis. The various sources of uncertainties in our analysis and their corresponding treatment are summarized in Table 3.3.

Our statistical analysis closely follows that of Ref. [1] with one major difference: the treatment of velocity uncertainties. In Ref. [1], normalization of each velocity bin is also treated as a nuisance parameter, which adds an additional 20-30 parameters to the analysis. In our approach, we propagate the velocity uncertainties, both statistical, estimated using

Type	Source	Treatement
$\nu^{\rm data}$	Poisson	$\sqrt{N_k}$ in the k-th bin
	3% dust extinction	$0.03 \times \nu^{\rm data}$
	Gaia systematic uncertainty	± 0.1 mas in ϖ; ± 0.1 mas/yr in $\mu_{\tilde{\alpha}}, \mu_\delta$
$\nu^{\rm mod}$	statistical errors of $f_{z=0}(\lvert w\rvert)$	bootstrap resampling
	$f_{z=0}(w>0) - f_{z=0}(w<0)$	$\lvert \ln \nu^{(+)}(z) - \ln \nu^{(-)}(z)\rvert$
	parallax uncertainty	Gaussian kernel smoothing

Table 3.3: Various sources of uncertainties and their treatment.

bootstrap resampling, and systematic, into the prediction uncertainties. We check that these two methods yield similar results for TGAS data.

Finally, to obtain the posterior distribution, we assume uniform prior distributions for all parameters except the baryonic ones; their priors follow a Gaussian distribution,

$$p_b(\boldsymbol{\zeta}|\mathcal{M}) = \prod_{k=1}^{N_b} \left(\frac{1}{\sqrt{2\pi\sigma^2_{\rho_k}}} \exp\left(-\frac{(\rho_k - \bar{\rho}_k)^2}{2\,\sigma^2_{\rho_k}} \right) \right) \left(\frac{1}{\sqrt{2\pi\sigma^2_{\sigma_{z;k}}}} \exp\left(-\frac{(\sigma_{z,k} - \bar{\sigma}_{z,k})^2}{2\,\sigma^2_{\sigma_{z,k}}} \right) \right), \tag{3.22}$$

where the mean and variance for each component are taken from Table 3.2. We summarize the details and ranges of assumed prior distributions for all parameters, $\boldsymbol{\theta}$, used in our analysis in Table 3.4.

3.4.4 Sampling the posterior

The posterior probability density function (simply the posterior henceforth) of the parameters can be defined using Bayes' theorem,

$$p(\boldsymbol{\theta}|\mathcal{M}, d) = \frac{p(d|\mathcal{M}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{M})}{p(d|\mathcal{M})}, \tag{3.23}$$

where the numerator is given by Eqs. (3.22) and (3.18) and the denominator, referred to in the literature as 'marginal likelihood' or 'evidence', is defined as

Parameters	Prior type	Range	Total
$\rho_k(0)$, $\sigma_{z;k}$	Gaussian	Eq. (3.22)	24
N_ν	Uniform	$[0.9, 2.0]$	3
$z_\odot$	Uniform	$[-30.0, 30.0]$ pc	1
h_{DD}	Uniform	$[0.0, 100.0]$ pc	1
$\rho_{\rm DM}$	Uniform	$[0.0, 0.06]$ $\mathrm{M}_\odot/\mathrm{pc}^3$	1
Σ_{DD}	Uniform	$[0.0, 30.0]$ $\mathrm{M}_\odot/\mathrm{pc}^2$	1

Table 3.4: Prior distributions of model parameters.

$$p(d|\mathcal{M}) = \int p(d|\mathcal{M}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{M})\, d\boldsymbol{\theta}. \tag{3.24}$$

We sample the posterior in Eq. (3.23) with the Markov Chain Monte Carlo (MCMC) sampler `emcee`[10] for estimating values of parameters and determining correlations between them. To draw samples from a d-dimensional parameter space, `emcee` implements the affine-invariant ensemble sampling algorithm of Ref. [201] that is based on simultaneously evolving an ensemble of N walkers. Since each walker in the ensemble independently samples the posterior, `emcee` is naturally suited for parallel computing on multicore systems (see Ref. [202] for more details).

In our implementation, we let (100-300) walkers run for (15000-25000) steps depending on the stellar type and components ($\rho_{\rm DM}$ or $\rho_{\rm DM}$ + thin DD) of the local DM content. These numbers are chosen to achieve an acceptance fraction $a_f \approx 0.3$ [203] for each walker. After accounting for the 'warm-up' time, $\sim$4000 steps, of the ensemble, we obtain $\gtrsim 2\times10^6$ samples on average for each iteration of our analysis.

[10] http://dfm.io/emcee/current/

3.5 Results

3.5.1 Local DM density

We summarize the results from the posterior sampling for the analysis with baryons and a constant halo DM density $\rho_{\rm DM}$ in Table 3.5. The median value of $\rho_{\rm DM}$ obtained through our kinematic analysis of A and early G stars are similar to each other, while using F stars yields a significantly higher value. We also note that our value of $\rho_{\rm DM}$ determined using A and early G stars is consistent with previous measurements made using SDSS/SEGUE G star data [204], $\rho_{\rm DM} = 0.012^{+0.001}_{-0.002}$ $\mathrm{M}_\odot/\mathrm{pc}^3$ (within 1σ) and $\rho_{\rm DM} = 0.008^{+0.025}_{-0.025}$ $\mathrm{M}_\odot/\mathrm{pc}^3$ (within 2σ), by Refs. [180] and [169] respectively.

While the 95% credible region (CR) for measurements of $\rho_{\rm DM}$ with A, F, and early G stars in Fig. B.4 overlap and seem consistent with each other at the 2σ level, we emphasize that each tracer population doesn't necessarily probe the same galactic environment due to differences in age and star formation history.[11] Consequently, without appropriate modeling of all prior information in a (hierarchical) Bayesian framework, results derived from different tracers should be compared with caution.

Stellar type	$\rho_{\rm DM}$ [$\mathrm{M}_\odot/\mathrm{pc}^3$]	$\rho_{\rm DM}$ [$\mathrm{GeV/cm}^3$]	ρ_b [$\mathrm{M}_\odot/\mathrm{pc}^3$]	$z_\odot$ [pc]
A stars	$0.016^{+0.010}_{-0.010}$	$0.608^{+0.380}_{-0.380}$	$0.088^{+0.007}_{-0.007}$	$8.80^{+3.74}_{-4.23}$
F stars	$0.039^{+0.008}_{-0.008}$	$1.482^{+0.304}_{-0.304}$	$0.089^{+0.007}_{-0.007}$	$2.04^{+2.84}_{-3.13}$
G stars	$0.011^{+0.010}_{-0.009}$	$0.418^{+0.380}_{-0.342}$	$0.087^{+0.007}_{-0.007}$	$-8.82^{+5.32}_{-4.64}$

Table 3.5: Median posterior values with 1σ errors for the local densities of baryons ρ_b and halo DM $\rho_{\rm DM}$, and height of the sun above the midplane $z_\odot$. The halo DM density $\rho_{\rm DM}$ is expressed in both $\mathrm{M}_\odot/\mathrm{pc}^3$ (astronomical unit) and $\mathrm{GeV/cm}^3$ (particle physics unit), where $1\ \mathrm{M}_\odot/\mathrm{pc}^3 \approx 38\ \mathrm{GeV/cm}^3$.

[11] For instance, each tracer type could have different sensitivity to non-equilibrium features of the MW [189]. Propagating these uncertainties to our estimates for the baryon and DM densities requires a detailed study using simulations, which is beyond the scope of this work.

3.5.2 Constraints on a thin DD

We perform a full MCMC scan of the posterior after including a thin DD component along with local density of halo DM $\rho_{\rm DM}$, and plot the marginalized posteriors for thin DD parameters, $\rho_{\rm DM}$, and the total midplane baryon density ρ_b in Figs. B.5, B.6, B.7. On the other hand, Fig. 3.7 gives the constraints on thin DD parameters after marginalizing over the uncertainties of the baryon mass model and asymmetries in velocity distribution. Given the exploratory nature of our analysis, this may be interpreted, at best, as *an approximate upper bound on the thin DD parameters*.

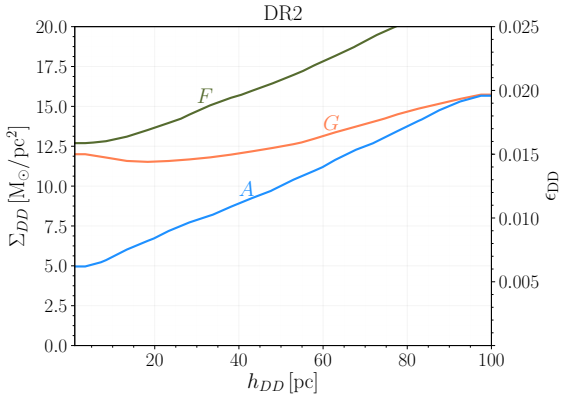


Figure 3.7: 95% CR upper limit contours for surface density Σ_{DD} and scale height h_{DD} of a thin DD for A (blue), F(green), and G (orange) stars using data from DR2 (left panel) and TGAS (right). The upper bound for the fraction of the total DM mass in the MW that could exist in a DD, ϵ_{DD}, is also shown on the right side of each plot for reference.

We also use the constraint on Σ_{DD} to estimate the amount of dissipative DM in the galaxy. Following Ref. [135], we relate the locally measured surface density of a thin DD to ϵ_{DD}, the fraction of the total DM mass in the MW that has dissipative self-interactions and forms a DD,

$$\Sigma_{DD}(R_\odot) = \frac{\epsilon_{DD}\, M_{\rm DM}^{\rm gal}}{2\pi R_{DD}^2}\, \exp\left(-R_\odot/R_{DD}\right) \tag{3.25}$$

where $M_{\rm DM}^{\rm gal} \sim 10^{12} M_\odot$ is the total DM mass in the MW, $R_\odot \sim 8.1$ kpc is the Sun's distance from the galactic center, and the scale radius of the thin DD is assumed to be equal to that

of baryons, $R_{DD} = 2.15$ kpc [169]. As indicated in Fig. 3.7, only $\sim$1% of the total DM mass could reside in a thin DD.

3.6 Discussion

Our main results from the MCMC sampling of the posterior, e.g. for A stars, imply that the local DM content can accommodate a constant density $\rho_{\rm DM} = 0.016 \pm 0.010$ $\rm M_\odot/pc^3$, or $\rho_{\rm DM} = 0.008^{+0.011}_{-0.008}$ $\rm M_\odot/pc^3$ and a thin DD with $\Sigma_{DD} = 2.99^{+3.75}_{-2.177}$ $\rm M_\odot/pc^2$, the precise value depending on h_{DD}. We observe that the 1σ uncertainties are fairly large in both cases and suggest high systematic noise in our determination. We discuss different sources of the uncertainties in Secs. 3.6.1-3.6.3 and comment on the robustness of our dynamical analysis. Lastly, we cross-validate our statistical setup by repeating our analysis with TGAS data in the same galactic volume, and comparing the results with those from DR2 in Sec. 3.6.4.

3.6.1 Effect of volume cuts

We vary the cylinder radius R and find that the tracers' vertical density distributions do not vary much for $R \lesssim 200$ pc. Increasing R from 150 pc to 250 pc, though, results in an overall broadening of the density distributions. Ref. [1] attributed a similar trend in TGAS data to the so-called 'Eddington' bias, *i.e:* higher parallax uncertainties of distant stars could lead to a smearing of the density distributions at large $|z|$. However, as shown in Fig. B.2, the parallax uncertainties are significantly reduced in DR2 and remain small at large $|z|$ even when R is increased to 250 pc. Thus, it seems unlikely that the broadening of the density distributions is due to the 'Eddington' bias. A more plausible option is the presence of local disequilibrium effects as we discuss in the following section. We note that our procedure would result in a lower local DM density estimate for a broader density distribution, since additional matter tends to pinch the predicted density profile as shown in Fig. 3.6.

3.6.2 Disequilibria in the solar neighborhood?

An implicit assumption in our modeling of the tracer density profile is that the local neighborhood is axisymmetric and the stellar disk is in dynamic equilibrium. However, growing evidence in DR2 data for: asymmetry in the vertical number counts [205, 206]; vertical waves in the disk at Sun's position [207–209]; kinematic substructure [60, 210, 211], warrants a closer look at sources of disequilibria in the solar neighborhood. We defer searches of local disequilibria and the corresponding revision of our traditional kinematic method outlined in Sec. 3.4.1 to future work. Presently, we only approximate the effect of non-equilibrium behavior by propagating asymmetries in the midplane velocity distribution to the errors in the predicted density.

3.6.3 Degeneracy between $\rho_{\rm DM}$ and ρ_b

The marginalized posterior for each tracer in Fig. B.4 indicates a strong degeneracy between measurements of ρ_b and $\rho_{\rm DM}$. As proposed by Ref. [195], and recently implemented on simulated data by Ref. [193], this degeneracy can only be broken if any kinematic analysis includes the density falloff at larger $|z|$ away from the midplane. Since most of the baryonic matter is confined to the stellar disk with a scale height $\mathcal{O}$(kpc), any excess matter that causes the falloff can be attributed to (at least to leading order) to DM, allowing a more precise measurement of $\rho_{\rm DM}$ with smaller error bars. On the other hand, this introduces another layer of complexity as the coupling between the radial and vertical motions is no longer negligible at $|z| \gtrsim 0.5$ kpc and must be modeled by simultaneously fitting to the velocity data [158, 179].

Meanwhile, the highly diagonal posterior in the $\rho_{\rm DM}$–Σ_{DD} plane combined with identically flat posterior in the $\rho_{\rm DM}$–ρ_b and Σ_{DD}–ρ_b planes of Figs. B.5, B.6, B.7 implies that introducing a thin DD in our analysis merely shifts some of the DM density from $\rho_{\rm DM}$ while increasing its relative error. Thus, to set realistic constraints on, or seek evidence for, DM density in the thin DD (or equivalently some form of extended substructure near the midplane) using our procedure, we would need more physical insight to break the degeneracy between different

distributions of DM.

As the discussion above indicates, our results are dominated by systematic errors stemming from an approximate modeling of non-equilibrium behavior and a strong degeneracy between different matter components near the midplane. *We note that these errors, in the context of the 1D distribution function method, may not be reduced significantly in future Gaia data releases.*

3.6.4 Comparison of constraints between DR2 and TGAS

We plot the 95% CR upper limit contours for the thin DD parameters using data from DR2 and TGAS in Fig. 3.7 and Fig. 3.8 respectively. Both sets of exclusion curves are significantly stronger than previous results based on the *Hipparcos* catalog [146]. However, there are obvious differences between our results derived using DR2 and TGAS data.[12]

Using TGAS data, early G stars exclude $\Sigma_{DD} \approx 5\,\mathrm{M}_\odot/\mathrm{pc}^3$ depending on h_{DD} while A stars set the weakest constraint. On the other hand, using DR2 data, A stars exclude $\Sigma_{DD} \gtrsim (5-12)\,\mathrm{M}_\odot/\mathrm{pc}^3$ while the weakest constraint is due to F stars.

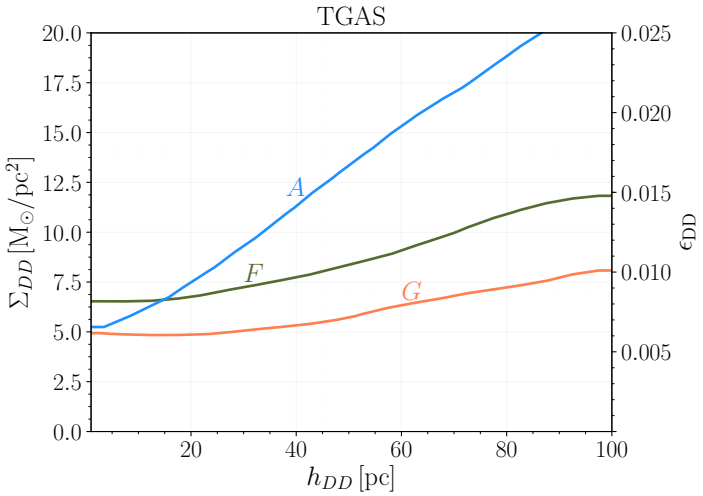


Figure 3.8: Same as Fig. 3.7 but using TGAS data.

[12]Our TGAS results derived using a fully Bayesian analysis roughly agree with those of Ref. [1]; see left panel of their Fig. S17 in particular.

Naively, we expect that there would be an (modest) improvement in the constraints from DR2 data compared to those from TGAS due to increased statistics (about a factor of ∼2.5) and decreased parallax uncertainties (due to our choice of binning, these only affect the high z bins). We check numerically that if we take central values from TGAS and uncertainties from DR2 to generate mock distributions for the tracers, the derived constraints on thin DD are indeed similar to those from TGAS data with minor improvements. Given this expectation, it seems counterintuitive that our DR2 constraints are different from the TGAS ones.

Before discussing possible origins of the differences for each tracer population, we note that adding more matter pinches the density profile of tracer stars, such as the effect of thin DD discussed in Sec. 3.4.1. Thus, the narrower the profile from data or broader the predicted density is, the more matter that can be included, and weaker the constraint on local DM content.

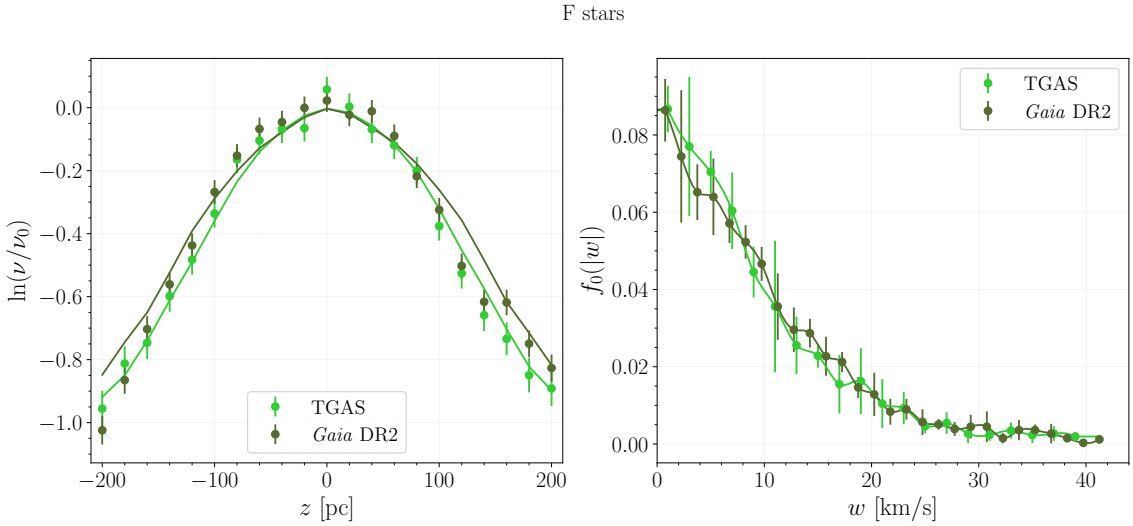


Figure 3.9: F stars: (left) volume complete number density profiles overlaid with the predicted density derived using the mean TGAS and DR2 velocity distributions assuming fiducial values for baryons and $\rho_{\rm DM} = 0.02$ $M_\odot/pc^3$; (right) midplane velocity distributions with interpolated fits to the data. Note that the TGAS velocity distribution has a bin size of 2 km/s while DR2 bin size is 1.5 km/s.

The significant weakening of constraints for F stars stems from small differences in the midplane velocity distributions, as shown in the right panel of Fig. 3.9. The DR2 velocity distribution is slightly broader. We verify that this trend in the velocity distribution is not an artifact of our choice of the midplane latitude cut or the binning of the velocity data. Although velocity (and vertical density) profiles from TGAS and DR2 are consistent with each other within uncertainties, the predicted density distribution with DR2 data is broader than that with TGAS data with fixed model parameters (one example is shown in the left panel of Fig. 3.9). As a result, a higher density in DM components is required to fit the predicted density of F stars to the DR2 number density profile for a fiducial baryon mass model.

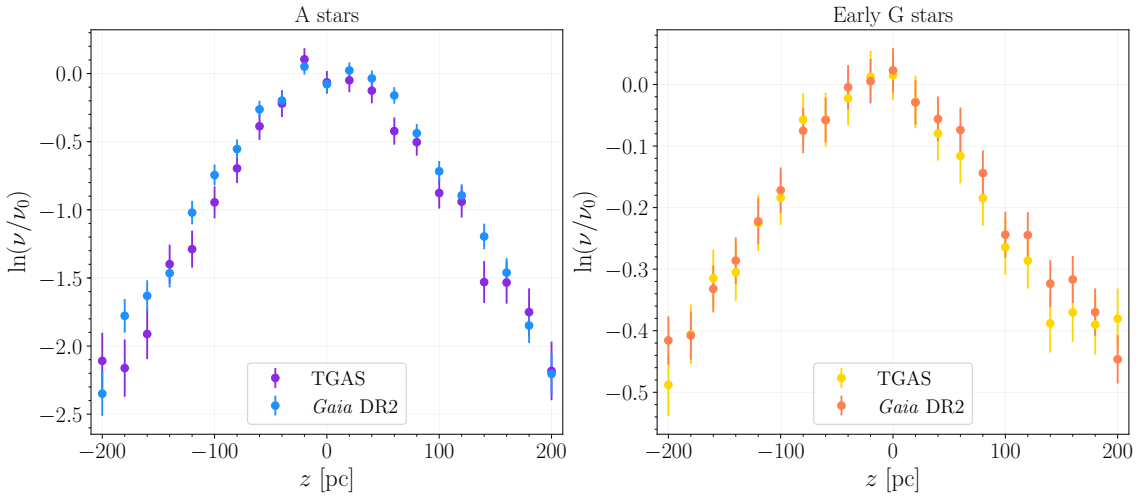


Figure 3.10: Comparison of volume complete number density profiles in TGAS and DR2 data for A (left) and G (right) stars.

We also present the volume complete number density profiles and midplane velocity distributions for A and early G stars in Fig. 3.10 and Fig. 3.11. From the plots, we note that all the distributions based on TGAS and DR2 data for both these tracers are also consistent within uncertainties, yet there are subtle differences. The velocity distributions using DR2 data are smoother compared to the TGAS ones with smaller systematic uncertainties from asymmetry between negative and positive velocity data.

The constraint from early G stars in the DR2 data set gets weaker due to both a slightly

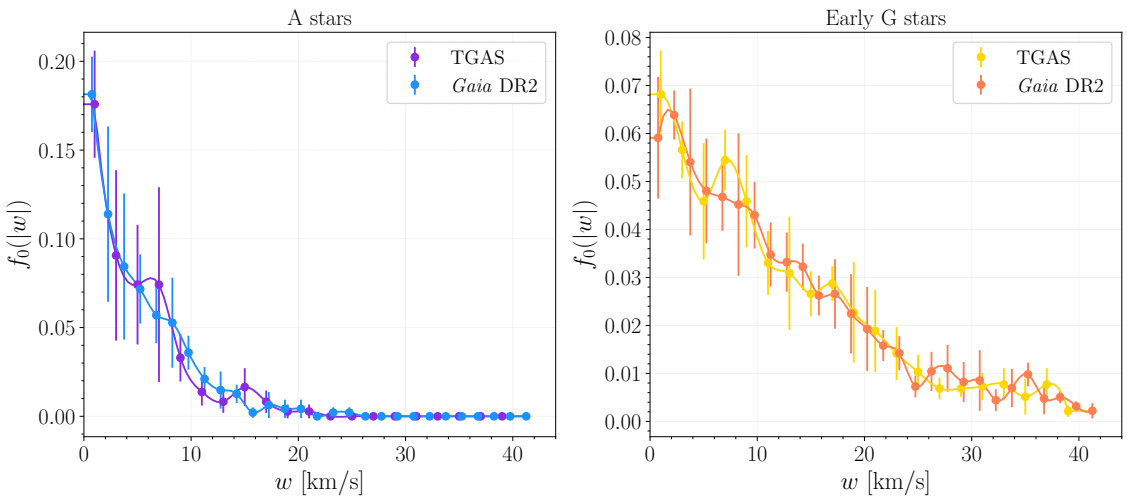


Figure 3.11: Comparison of midplane velocity distributions in TGAS and DR2 data for A (left) and G (right) stars. Note that the TGAS velocity distribution has a bin size of 2 km/s.

narrower density profile, and a slightly broader predicted density. However, in the case of A stars, the constraint gets considerably stronger at high h_{DD} due to the reduction in the systematic errors from the asymmetry in the midplane velocity distribution.

We reiterate that *Gaia* DR2 should be regarded as a different data catalog from TGAS, rather than just a statistical improvement over it [54]. DR1 incorporated positions from the Tycho-2 catalog to generate the five-parameter astrometric solution in the TGAS catalog, whereas, the DR2 catalog is independent from any other external catalogs with its own self-consistent astrometric solution. Any comparison between the constraints on local DM content from TGAS and DR2 should be made bearing this difference in mind.

3.7 Conclusions and Outlook

We apply the 1D distribution function method to *Gaia* DR2 and use stellar kinematics in the solar neighborhood to constrain the local DM density and properties of a thin DD aligned with the baryonic disk by performing our analysis within a Bayesian framework. We adopt young A, F, and early G stars as tracers as they have shorter equilibration timescales and consequently are expected not to be strongly affected by disequilibria. Using A stars gives

an estimate of $\rho_{\rm DM} = 0.016 \pm 0.010$ $M_\odot/pc^3$ and sets the strongest constraint on the thin DD, excluding $\Sigma_{DD} \gtrsim$ (5-12) $M_\odot/pc^2$ depending on the scale height with 95% confidence. This upper bound is used to constrain the amount of dissipative DM in the galaxy: a thin DD with $\Sigma_{DD} \lesssim 12$ $M_\odot/pc^2$ and a scale radius $\sim$3 kpc contains $\lesssim 1\%$ of the total DM mass in the Milky Way [135]. While we obtain similar results from early G stars, F stars seem to prefer a higher value of the local DM content. Even though the distributions derived from DR2 are consistent with those from TGAS data within uncertainties, the allowed DM density and parameters of DD model are quite different for all tracers. In light of these results, we address the origins of the differences and discuss the robustness of our kinematic analysis.

Our results also suggest that we need a better understanding of the physical origin of the systematic uncertainties, which we include in our analysis to account for the asymmetry in the midplane velocity distributions of tracers. One possibility is that with complete data for radial velocities, we could define the midplane region using the z-cut instead of the b-cut and obtain a more precise determination of the velocity distribution. Another possibility is to take a closer look at local disequilibria and their effects on traditional kinematic methods. Although we do not find any statistically significant evidence for non-equilibrium in the vertical density and velocity distributions in our samples, several analyses based on DR2 seem to suggest various sources of disequilibria at distances larger than the heliocentric cylinder we consider. In terms of baryon modeling, it could be useful to find a self-consistent, data-driven approach to determine the baryon distributions instead of assuming the isothermal Bahcall model. One way to achieve this would be to construct the mass density for stars directly from the data rather than treating it as an isothermal disk.

For a more precise determination of the local DM density, a dynamical analysis could be performed using tracers at heights greater than the scale height of the stellar disk to minimize the latent degeneracy between baryons and DM. However, besides modeling effects of disequilibria, an analysis at larger scale height has to go beyond the 1D method and must include terms that couple the motions of tracers in different directions. We also see a degeneracy between parameters of ordinary DM and thin DD in the marginalized posteriors

obtained through MCMC sampling. To break the degeneracy, we would need to distinguish between their effects on tracers by developing new observables and modeling priors that reflect these differences.

Chapter 4

The Gaia Sausage for Dark Matter Nuclear Interactions

4.1 Introduction

Confirming the existence of dark matter (DM) through a variety of cosmological and astrophysical observations has been one of the major successes of 20^{th} century physics. Simultaneously, questions regarding the particle nature of DM and its interactions with ordinary matter beyond gravity remain unresolved. Fortunately, there is a vibrant research program that seeks to answer these questions on the experimental and observational frontiers. A leading probe in the hunt for DM is direct detection (DD) experiments, which look for signals from DM particles scattering in underground detectors. Although there have been no statistically significant detections of non-background events so far, next-generation experiments such as LUX-ZEPLIN (LZ) [212], XENON-nT [213], PANDAX-xT [214], SuperCDMS SNOLAB [215], DAMIC-M [216], DARWIN [217] and DARKSIDE-20K [218] serve as promising avenues not just for DM discovery, but, as we argue below, for reconstructing its astroparticle properties as well.

The main physical observable in a DD experiment is the differential recoil spectrum, typically quoted as a function of the primary scintillation signal. Interestingly, modeling

the DM recoil spectra at DD experiments relies on several independent aspects of its phenomenology. More specifically, DD experiments probe a combination of three important DM properties: its mass, interaction type (with nucleus and/or electrons) which we will refer to as the model, and its astrophysical distribution, namely through its density and velocity distribution in the solar neighborhood. From the perspective of statistical inference, this results in a three-fold degeneracy depicted in Fig. 4.1. As corollary, determining the local astrophysical properties of DM precisely will be crucial in reconstructing its particle physics properties.

The Standard Halo Model (SHM) of DM velocity distribution [219] has been the cornerstone of DD analyses since it was proposed nearly three decades ago. The SHM follows from modeling the Milky Way (MW) as an isotropic, isothermal halo in equilibrium formed through the virialization of multiple subhalo merger residue. However, observations [220–223] and results from N-body simulations [224–229] have suggested the presence of diverse stellar and DM substructures from recent mergers, challenging the MW's steady state characterization. There have also been attempts to semi-analytically model the local DM velocity distribution using kinematic data [230–236], but these rely on additional, potentially restrictive, assumptions about the structure of the MW.

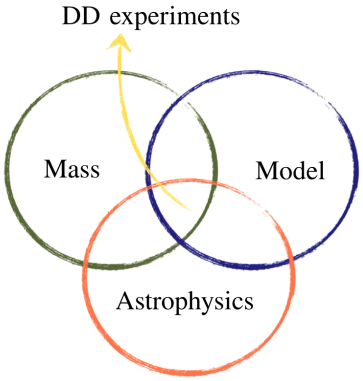


Figure 4.1: Three-fold degeneracy in DM direct detection.

On the other hand, astrometric data released by the European Space Agency's *Gaia* mission [54,55] presents a unique opportunity to study the MW's accretion history, and take

first steps toward an empirical determination of the DM phase-space distribution. Even with a subset of the full data, a few groups [56–59] have reported evidence of tidal debris from a dominant merger in the solar neighborhood, the so-called *Gaia* Sausage or *Gaia* Enceladus, with very different kinematics compared to the old, virialized stellar population. While the full implication of stellar data for the astrophysical properties of DM will take decades to analyse, pioneering work by ref. [60] used the kinematics of MW halo stars in the cross-matched Sloan Digital Sky Survey (SDSS)-*Gaia* data set as tracers for the DM velocity, and validated their analysis [61] with the FIRE-2 cosmological zoom-in simulation. Their analysis used a mixture model to characterize the local DM velocity distribution as a sum of the smooth halo and a radially aniostropic substructure components, which we refer to henceforth as the *Gaia* velocity distribution.

In this chapter, we perform a systematic study of how the DM velocity distribution affects the reconstruction of DM model parameters at current or near-future DD experiments. Our analysis expands the existing DM direct detection literature in two important ways:

- We consider the effect of DM velocity distribution for DM models that encompass a diverse set of operators in non-relativistic effective field theory (NREFT) with different recoil energy and velocity dependences. We show that the *Gaia* velocity distribution could significantly change not just the overall rate, but also the shape of the DM recoil spectrum at DD experiments.

- Adopting the euclideanized signal (ES) method developed by refs. [237,238], we forecast the ability of next-generation experiments to resolve DM model parameters using SHM and *Gaia* velocity distributions as two representative cases. We find that both the DM mass and the recoil energy dependence of the model could enhance or suppress the effect of the *Gaia* velocity distribution vis-á-vis the SHM while inferring the DM model parameters. This is another example of the curious interplay between different DM properties illustrated in Fig. 4.1.

Our approach is quite different from both traditional forecasting approaches for DM direct detection that used benchmark-dependent mock data sets [239–243], or analyses which

studied the effect of uncertainties in SHM on constraining the DM particle physics properties [244–247].

This chapter is organized as follows. We review the basic ingredients to compute the recoil spectra at DD in section 4.2: we summarize and review the velocity distribution taking into account of the *Gaia* Sausage in section 4.2.1, and review the NREFT formalism along with several benchmark DM models in section 4.2.2. In section 4.3, we demonstrate how *Gaia* velocity distribution could modify both the overall rate and the shape of the recoil spectra in the NREFT framework. In Section 4.4, we introduce the recently developed ES statistical framework, which allows us to make forecasts without running MC simulations. In section 4.5, we present our results on the effects of *Gaia* distribution on reconstructing DM model parameters and distinguishing between different models at future DD experiments as well as setting constraints using current data. We conclude in section 4.6.

This chapter is based on the work with Jatan Buch and JiJi Fan in Ref. [248].

4.2 Phenomenology of dark matter-nuclear Interaction

In this section, we briefly review the key ingredients to compute the rate of DM scattering off nuclei in a DD experiment. In particular, we summarize possible new DM velocity distributions inferred from the *Gaia* data. We also review both the model-independent framework, the non-relativistic effective theory and some specific benchmark models to study different types of DM-nucleus interactions. We emphasize that this section is a review of the literature, which are most relevant to our studies. Readers who are familiar with the subject could skip this section. A more extensive recent review on DD can be found in ref. [44].

4.2.1 DM velocity distributions

Our key experimental observable, the differential recoil rate (the full formula is provided in Appendix C.1), is sensitive to the DM velocity distribution. In this section, we first review the SHM and the *Gaia* distribution which we use in our analysis and then compare the velocity moments from different possible velocity distributions.

The local velocity distribution in the GC frame for the SHM, *i.e.* for an isotropic, isothermal DM halo in equilibrium, is well-modeled by a truncated Maxwell-Boltzmann distribution [219],

$$f_{\rm SHM}(\boldsymbol{v}) = \frac{1}{N_{\rm esc}} \frac{1}{(2\pi\sigma_v^2)^{3/2}} e^{-v^2/2\sigma_v^2}\, \Theta(v_{\rm esc} - v), \tag{4.1}$$

where we take the velocity dispersion and escape velocity of the DM halo to be $\sigma_v \approx 160$ km/s and $v_{\rm esc} \approx 540$ km/s respectively [249], and the normalization constant $N_{\rm esc}$ is given by,

$$N_{\rm esc} = \text{erf}\left[\frac{v_{\rm esc}}{\sqrt{2}\sigma_v}\right] - \frac{2}{\sqrt{\pi}}\left(\frac{v_{\rm esc}}{\sqrt{2}\sigma_v}\right) e^{-v_{\rm esc}^2/2\sigma_v^2}\,. \tag{4.2}$$

The DM velocity distrubution in the Earth frame, $\tilde{f}(\boldsymbol{v})$, can be obtained by boosting GC frame distribution, $f(\boldsymbol{v})$, with the Earth's velocity, $\boldsymbol{v}_{\rm obs}(t)$,

$$\tilde{f}(\boldsymbol{v}) = f\left(\boldsymbol{v}_{\rm obs}(t) + \boldsymbol{v}\right). \tag{4.3}$$

where the t dependence arises due to Earth's orbital motion around the Sun and is commonly referred to as annual modulation [250, 251]. Ignoring modulation effects, we assume a time averaged value for the Earth's velocity $v_{\rm obs} \approx 230\,\text{km/s}$ in our analysis.

A crucial assumption in the formulation of the SHM is the condition of local equilibrium. However, if the MW has undergone one or more recent mergers, the equilibrium condition is then invalid, and we would need a method to empirically determine the DM velocity distribution. Assuming the CDM paradigm, hierarchical structure formation implies that the MW halo should primarily consist of virialized tidal debris from old subhalo mergers with other spatial and kinematic substructure sourced by more recent ones. Ref. [252] used the `Eris` N-body simulation to show the correlation between velocities of old, metal-poor stars and the virialized DM component of a MW-like halo. Along the same lines, ref. [226] argued that the velocity distribution of stars in a class of substructure called *debris flow*[1] is a good kinematic tracer of its accreted DM counterpart based on the `Via Lactea` simulation.

[1]Debris flow consists of tidal debris of an accreted dwarf galaxy that has made several orbits such that it is spatially mixed on large scales while retaining a unique signature in velocity space [253].

In unarguably the golden-age of data-driven astrophysics, we can now obtain 7D information[2] for main-sequence stars in the MW halo by cross-matching the *Gaia* data releases [54, 55] with the Sloan Digital Sky Survey (SDSS). Using this cross-matched catalog, refs. [56, 57] found signatures of a debris flow in the solar neighborhood (within ~ 4 kpc of the Sun)– the so-called *Gaia*-Sausage – that consists of metal-rich halo stars with high radial anisotropy. Further investigation of other phase space substructures [56, 58, 254] indicates that, in fact, the MW might have experienced at least two different accretion events, namely those of the *Sausage* [56] and *Sequoia* [255] dwarf galaxies. The *Gaia*-Enceladus structure [57] hints at a possible third event, although it appears to partially consist of debris from the other two mergers [256].

For our analysis, we only focus on the effects of the *Gaia*-Sausage, since it is the dominant merger in the solar neighborhood contributing $\sim 70\%$ of all accreted low-metallicity stars [56]. Ref. [60] used the SDSS-*Gaia* DR2 data set with a subsample of $\sim 190,000$ stars to make the first empirical determination of the local DM velocity distribution. They performed a Gaussian mixture model (GMM) analysis on the joint distribution of stellar velocities and metallicities to classify stars in three populations with distinct kinematic properties: metal-rich young disk stars formed *in-situ*; accreted stars which include metal-poor stars in the smooth isotropic halo, and intermediate metallicity stars with a high radial anisotropy that constitute the *Gaia*-Sausage substructure. The total DM velocity can then be written as a linear combination of the substructure and halo velocities, weighted by the the fraction of DM in each component,

$$f(v) = (1 - \eta_{\rm sub}) f_{\rm halo}(v) + \eta_{\rm sub} f_{\rm sub}(v), \tag{4.4}$$

where $\eta_{\rm sub}$ parametrizes the DM fraction in substructure. By sampling the GMM model with the SDSS-*Gaia* DR2 subsample, ref. [60] inferred the best-fit velocity distribution for each component[3] as well as a posterior distribution of their stellar fractions. Subsequently, ref. [61] derived an approximate relation between the mass-to-light ratio and metallicity

[2]Including parallax, sky positions and proper motions on the celestial sphere, and radial velocity measurements from *Gaia*, along with metallicity data from SDSS. The stellar metallicity, given by the iron-to-hydrogen abundance ratio, [Fe/H], is used as a proxy for the star age.

[3]Available publicly here: https://github.com/linoush/DM_Velocity_Distribution.

for the MW [257, 258] for using the stellar fraction posterior to estimate the DM fraction distribution, $p(\eta_{\rm sub})$, shown in the left panel of Fig. 4.2.[4] Importantly, however, the fraction of stars in substructure relative to total accreted stars is typically a poor tracer for $\eta_{\rm sub}$ as simulations show that the halo DM component may contain significant contributions from accretion of diffuse DM and DM in non-luminous subhalos. In fact, as demonstrated by ref. [61, 259, 260], late-time accretion from the latter component may affect both the shape and the velocity distribution of all accreted DM. We postpone a detailed study of these effects to future work.

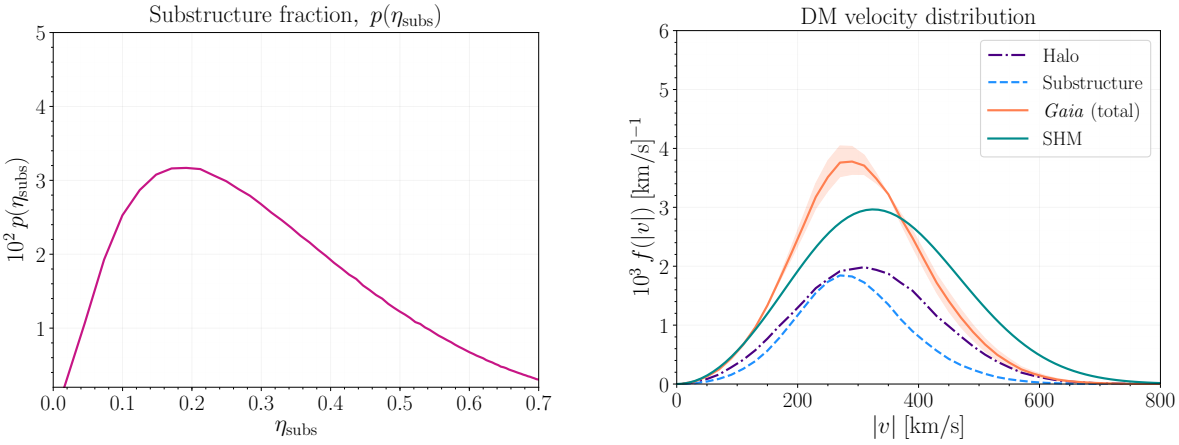


Figure 4.2: Left: distribution of the DM fraction in substructure $\eta_{\rm sub}$ determined using an empirical relation between the mass-to-light ratio and metallicities for the MW (provided by L. Necib in private communication). Right: the normalized *Gaia* DM velocity distributions for the halo (indigo dash dotted) and substructure (blue dashed) components weighted by the median DM substructure fraction, *i.e.* $\eta_{\rm sub} = 0.42$. Also shown for reference are the total velocity distributions in the heliocentric frame for *Gaia* for the median value of DM fraction (orange solid) and the SHM (cyan solid). The band of the *Gaia* distribution is obtained by varying $\eta_{\rm sub}$ in the 1σ range around its median value.

We plot the *Gaia* DM velocity distribution in the heliocentric frame for each component from ref. [60] in the right panel of Fig. 4.2. We also note that there is no smooth interpolation

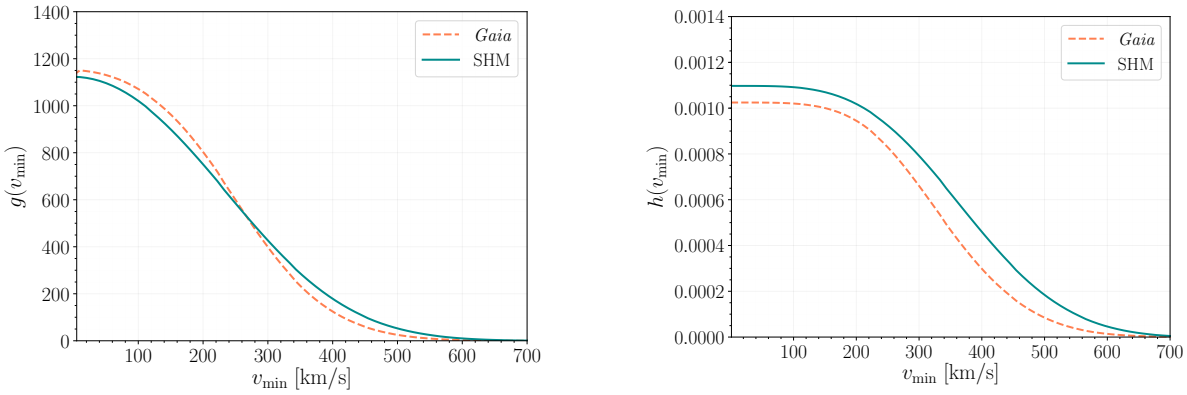
[4]While ref. [61] only provides the median value of $\eta_{\rm sub}$ with its 1σ error bar, the full distribution was provided to us by one of the authors, L. Necib, in private communication.

in $\eta_{\rm sub}$ between the SHM and the *Gaia* velocity distribution, *i.e.* $\eta_{\rm sub} = 0$ does not yield the SHM.[5] Heuristically, the differences between the two velocity distributions can be attributed to the *Gaia* one being inferred through a better statistical modeling of the same stellar population.

There has also been growing interest in studying the effect of other phase-space substructures in the solar neighborhood discovered using *Gaia* data, in particular for the retrograde S1 [254], and the prograde Nyx [263, 264] streams. Stellar streams appear as a coherently moving group of stars resulting from the tidal debris of a galaxy localized in both position and velocity space, and presence of a significant DM fraction in stream(s) could result in a very different annual modulation signature compared to the SHM. With this motivation, ref. [265] revisited the DM interpretation of the latest DAMA data with the S1 stream, and found that absence of a DM signal at other experiments rules out the preferred DM parameter space of DAMA even if 100% of the local DM was present in such a stream. Another promising avenue to look for interesting signatures of DM substructure are axion searches and directional detection experiments as discussed in refs. [261,266,267]. The main drawback of the aforementioned analyses is the underlying assumption of a near perfect stellar-DM velocity correlation. As illustrated in ref. [61] (see top panel of Fig. 7 for example), stellar streams turn out to be poor tracers of the DM velocity, since the tidal debris in the stream hasn't had enough time to completely mix with the halo. In addition, compared to the Sausage substructure, streams are expected to contribute only a subdominant fraction of DM in the solar neighborhood. For example, using the results of ref. [268], ref. [267] argued that the progenitors of the S1 and S2 stream could contribute $\sim$1-10% of the local DM fraction. Thus, we only focus on the DM velocity distribution associated with the *Gaia* Sausage based on refs [60, 61] in our work, and ignore the effect of streams in our analysis.

Dark matter interactions with the nucleus beyond the simplest contact one could introduce additional factors of incident DM velocity squared v^2 into the velocity integration. We

[5] We note here that ref. [261] uses the sphericity constraint of the DM halo [262] to argue that $\eta_{\rm sub} \lesssim 20\%$. However, this constraint relies on the assumption that the observed stellar density of the Sausage is a reasonably good tracer for the density of its DM component. In any case, as we show in fig. 4.6, the qualitative behavior of our result holds for a wide range of substructure fractions.



(a) $g(v_{\rm min})$ (b) $h(v_{\rm min})$

Figure 4.3: Moments of the empirical *Gaia* velocity distribution (orange, dashed) and the SHM distribution (cyan, solid) relevant in our analysis, $g(v_{\rm min})$ (left) and $h(v_{\rm min})$ (right), as functions of $v_{\rm min}$. The value of $g(v_{\rm min})$ ($h(v_{\rm min})$) has been multiplied (divided) by the speed of light to be made dimensionless, for a better illustration of their relative magnitudes.

will consider two velocity moments:

$$g(v_{\rm min}) = \int_{v_{\rm min}}^{v_{\rm esc}} {\rm d}^3 v \frac{\tilde{f}(\boldsymbol{v})}{v}, \tag{4.5}$$

$$h(v_{\rm min}) = \int_{v_{\rm min}}^{v_{\rm esc}} {\rm d}^3 v \, v \tilde{f}(\boldsymbol{v}), \tag{4.6}$$

where $v_{\rm min} = \sqrt{\frac{m_T E_R}{2\mu_T^2}}$ is the minimum DM velocity for a given recoil energy E_R, target mass m_T, DM-nucleus reduced mass μ_T. The velocity integrals are bounded from above by the escape speed, $v_{\rm esc}$, which we take to be ∞ for the *Gaia* velocity distribution following ref. [60].

We plot $g(v_{\rm min})$ and $h(v_{\rm min})$ as functions of $v_{\rm min}$ in Fig. 4.3. From the left panel of Fig. 4.2, we note that f_{Gaia} peaks at a lower v and has a smaller high velocity tail compared to $f_{\rm SHM}$. Integrating it with $1/v$ results in $g(v_{\rm min})_{Gaia} > g(v_{\rm min})_{\rm SHM}$ for small $v_{\rm min}$, whereas for $v_{\rm min} \gtrsim 250$ km/s, $g(v_{\rm min})_{Gaia} < g(v_{\rm min})_{\rm SHM}$. Meanwhile in case of $h(v_{\rm min})$, the preference for large velocities due to an additional factor of v results in $h(v_{\rm min})_{Gaia} < h(v_{\rm min})_{\rm SHM}$ for all values of $v_{\rm min}$. These features are crucial for us to understand how the *Gaia* distribution affects the recoil spectra at the qualitative level in Sec. 4.3.

4.2.2 DM-nucleus scattering theory

In this section, we will briefly review a model-independent framework and some concrete models to study different types of DM scattering off nucleus in DD experiments. There is a huge literature on possible DM scattering in DD and we do not intend to provide an exhaustive review here. We will only refer the reader to the original papers and papers we actually use for our analysis.

The typical DM velocity in our galaxy is $v \lesssim 10^{-3}c$. The incident DM kinetic energy is around $\mathcal{O}(10)$ keV for a DM particle with mass ~ 10 GeV. Given that DM scattering in DD experiments is non-relativistic (NR), a simple model-independent way to parametrize different types of DM-nucleon interaction is the Galilean-invariant NR effective field theory (NREFT), first proposed and developed in refs. [269] and [270]. The core result is that in the NR limit, DM-nucleon interactions could be encoded in 16 NREFT operators, 15 of which are linearly independent. These operators are expressed in terms of four three-vectors: DM spin $\boldsymbol{S}_\chi$, nuclear spin $\boldsymbol{S}_N$, the momentum transfer $\boldsymbol{q} = \boldsymbol{p}' - \boldsymbol{p}$ with $\boldsymbol{p}$ ($\boldsymbol{p}'$) the incoming (outgoing) DM three-momentum and the transverse velocity

$$\boldsymbol{v}^\perp = \boldsymbol{v} + \frac{\boldsymbol{q}}{2\mu_T}, \quad \text{where } \boldsymbol{v}^\perp \cdot \boldsymbol{q} = 0. \tag{4.7}$$

We will use only 12 linearly independent operators, $\mathcal{O}_i$'s, listed in Table 4.1. These 12 operators are usually sufficient to describe the NR limit of many relativistic operators that appear in simple models with spin-0 or spin-1 mediators. In this work, we only focus on spin-independent scattering.

In the NR limit, a relativistic operator in the field theory can be mapped onto a linear combination of NR operators. Thus, a relativistic Lagrangian for a particular DM scattering model, which could contain several relativistic operators for DM-nucleon interaction, can be written in terms of the NREFT operators as,

$$\mathcal{L}_{\text{NREFT}} = \sum_{N=n,p} \sum_{i=1}^{12} \frac{c_i^{(N)} \mathcal{O}_i^{(N)}}{q^2 + m_{\text{med};i}^2}, \tag{4.8}$$

where $N = n, p$ labels the type of nucleon DM interacts with, which could be either neutron or proton and i labels NREFT operators. The coefficients, c_i's, depend on the coupling

Operators	Form	Spin-Dependence
$\mathcal{O}_1$	$\mathbb{1}$	✗
$\mathcal{O}_2$	$(\boldsymbol{v}^\perp)^2$	✗
$\mathcal{O}_3$	$i\boldsymbol{S}_N \cdot (\boldsymbol{q} \times \boldsymbol{v}^\perp)$	✓
$\mathcal{O}_4$	$\boldsymbol{S}_\chi \cdot \boldsymbol{S}_N$	✓
$\mathcal{O}_5$	$i\boldsymbol{S}_\chi \cdot (\boldsymbol{q} \times \boldsymbol{v}^\perp)$	✗
$\mathcal{O}_6$	$(\boldsymbol{S}_N \cdot \boldsymbol{q})(\boldsymbol{S}_\chi \cdot \boldsymbol{q})$	✓
$\mathcal{O}_7$	$\boldsymbol{S}_N \cdot \boldsymbol{v}^\perp$	✓
$\mathcal{O}_8$	$\boldsymbol{S}_\chi \cdot \boldsymbol{v}^\perp$	✗
$\mathcal{O}_9$	$i\boldsymbol{S}_\chi \cdot (\boldsymbol{S}_N \times \boldsymbol{q})$	✓
$\mathcal{O}_{10}$	$i\boldsymbol{S}_N \cdot \boldsymbol{q}$	✓
$\mathcal{O}_{11}$	$i\boldsymbol{S}_\chi \cdot \boldsymbol{q}$	✗
$\mathcal{O}_{12}$	$\boldsymbol{v}^\perp \cdot (\boldsymbol{S}_\chi \times \boldsymbol{S}_N)$	✗

Table 4.1: Summary of the NREFT operators. The second column indicates the operators' nuclear spin dependence.

coefficients in the relativistic theory and the Wilson coefficients obtained by mapping the relativistic operators to the NR ones. Compared to the standard literature (e.g., refs [270–272]), we also take the mediator propagator out of c_is and explicitly write it out, where $m_{\mathrm{med};i}$ is the mediator mass for the ith interaction. Strictly speaking, the formula above holds when the mediator is light with mass below GeV. When the mediator is heavy with mass above GeV, it could be integrated out and the propagator in eq. (4.8) is reduced to m_{med}^{-2}.

In the work, we only consider the leading order spin-independent (SI) elastic scattering in which DM scatters off the entire nucleus coherently. To go from DM scattering with individual nucleons to scattering with nucleus, one needs to take into account of the nuclear

response which is encoded in the form factor

$$\begin{aligned}\mathcal{F}_{i,j}^{(N,N')} &= \left\langle \text{ Nucleus } \middle| \, \mathcal{O}_i^{(N)} \mathcal{O}_j^{(N')} \middle| \text{ Nucleus } \right\rangle, \\ &= \sum_{n=0}^{\infty} F_{i,j}^{(n)}(q^2) v^{2n}, \end{aligned} \tag{4.9}$$

where in the second line, we expand it as a power series of v^2 and $F^{(n)}$ are associated coefficients. Further development of effective field theory from quarks to nucleons could be found in refs [273,274]. In NREFT, the differential scattering rate (full formula in Eq. (C.1)) in terms of the c_is and the form factors is,

$$\begin{aligned}\frac{\mathrm{d}R}{\mathrm{d}E_R} &\propto \sum_{\substack{i,j,\\ N,N'}} \int_{v_{\min}}^{\infty} \mathrm{d}^3 v \frac{\tilde{f}(\boldsymbol{v})}{v} \frac{c_i^{(N)} c_j^{(N')}}{(q^2 + m_{\mathrm{med};i}^2)(q^2 + m_{\mathrm{med};j}^2)} \mathcal{F}_{i,j}^{(N,N')}(q^2, v^2), \\ &= \sum_{i,j} \frac{c_i c_j}{\left(q^2 + m_{\mathrm{med};i}^2\right)\left(q^2 + m_{\mathrm{med};j}^2\right)} \left[g(v_{\min}) F_{i,j}^{(0)}(q^2) + h(v_{\min}) F_{i,j}^{(1)}(q^2) + \cdots \right]. \end{aligned} \tag{4.10}$$

Note that our normalization of the form factors $\mathcal{F}$'s differs from that in the literature (e.g., refs [270, 271]) by a factor of $(4 m_\chi m_N)^2$ with m_N the nucleon mass. The form factors used in our analysis are provided in Appendix C.2.

A DM model could contain several different DM-nucleon interactions and a relativistic operator between DM and nucleons could map onto multiple NR operators. Moreover, the constraint on the coupling of the relativistic operator could be a complicated sum of constraints on each NR operator it corresponds to with different weights due to interferences between different NR operators. Thus in analyzing how the new velocity distribution affects interpretation of DD data, we also consider a few simple representative models. Again, we do not intend to be exhaustive; we only select and compare a few models with different E_R and velocity dependences, which could be affected by SHM and *Gaia* velocity distributions in different ways.

We assume for simplicity that the DM is fermionic, and analyze six SI DM models: DM interacting through heavy scalar mediator (leading to the simplest contact interaction), millicharged DM (mC), DM with magnetic dipole moment (MD) with either heavy or light mediators, DM with electric dipole moment (ED) with light mediator and anapole DM with

a heavy mediator. All the models are summarized in table 4.2. More details of the models could be found in Appendix C.3.

Model	Relativistic Operator	NREFT Operator	E_R and DM velocity moment
Contact	$g_c\bar{\chi}\gamma_\mu\chi\bar{N}\gamma^\mu N$	$g_c\mathcal{O}_1^{(N)}$	$g(v_{\min})$
Millicharge (mC) light mediator	$e\epsilon_\chi\bar{\chi}\gamma^\mu\chi A_\mu$	$e^2\epsilon_\chi\frac{1}{q^2}\mathcal{O}_1^{(p)}$	$E_R^{-2}\,g(v_{\min})$
Magnetic dipole (MD) heavy mediator	$\frac{\mu_\chi}{2}\bar{\chi}\sigma^{\mu\nu}\partial_\mu\chi\partial^\alpha F_{\alpha\nu}$	$\frac{e\mu_\chi}{2}\big(\frac{q^2}{m_\chi}\mathcal{O}_1^{(p)}-4\mathcal{O}_5^{(p)}\big)$	$E_R^2\,g(v_{\min})+E_R\,h(v_{\min})$
Magnetic dipole (MD) light mediator	$\frac{\mu_\chi}{2}\bar{\chi}\sigma^{\mu\nu}\chi F_{\mu\nu}$	$\frac{e\mu_\chi}{2}\big(\frac{1}{m_\chi}\mathcal{O}_1^{(p)}-\frac{4}{q^2}\mathcal{O}_5^{(p)}\big)$	$g(v_{\min})+E_R^{-1}h(v_{\min})$
Electric dipole (ED) light mediator	$i\frac{d_\chi}{2}\bar{\chi}\sigma^{\mu\nu}\gamma^5\chi F_{\mu\nu}$	$2ed_\chi\frac{1}{q^2}\,\mathcal{O}_{11}^{(p)}$	$E_R^{-1}\,g(v_{\min})$
Anapole heavy mediator	$ig_{\rm ana}\bar{\chi}\gamma^\mu\gamma_5\chi\partial^\nu F_{\mu\nu}$	$2eg_{\rm ana}\mathcal{O}_8^{(p)}$	$E_R\,g(v_{\min})+h(v_{\min})$

Table 4.2: Summary of the representative DM models used in our analysis. For brevity, only the SI NR operators of each model are shown here. We also show the leading order E_R and velocity moment dependences of the corresponding spectra. The coupling constant in front of the operator defines the model parameter to be constrained, e.g. coupling strength g_c for contact interaction, charge fraction ϵ_χ for millicharged DM, and the dipole moment μ_χ for magnetic dipole DM.

4.3 Effect of the DM velocity distribution on DD recoil spectrum

Before presenting the statistical method for DD forecasting and the results, we want to discuss how different DM velocity distributions could affect the differential recoil spectrum, the key quantity in DD experiments. Here we present results using NREFT operators. This will help us obtain some qualitative ideas and physical intuition of the effects of DM velocity distribution on interpreting current and future DD data in the context of a full-fledged model, which we present in section 4.5.

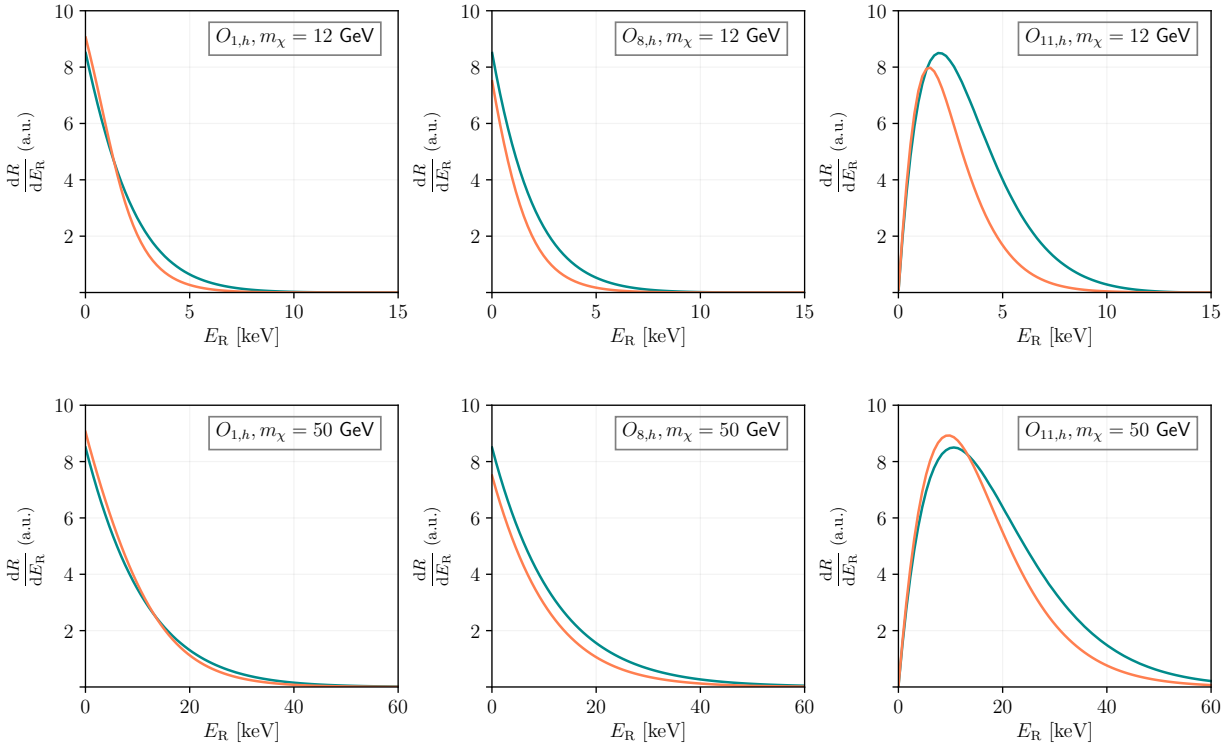


Figure 4.4: Top row: recoil spectra of $\mathcal{O}_1$, $\mathcal{O}_8$ and $\mathcal{O}_{11}$ with heavy mediators for DM with $m_\chi = 12$ GeV. Bottom row: recoil spectra of the same set of operators for $m_\chi = 50$ GeV. The cyan (orange) curve assumes SHM (*Gaia*) velocity distributions. In this figure and the next one, the differential recoiling rate is in arbitrary unit.

The recoil spectrum depends on not only the velocity moments but also the coefficients

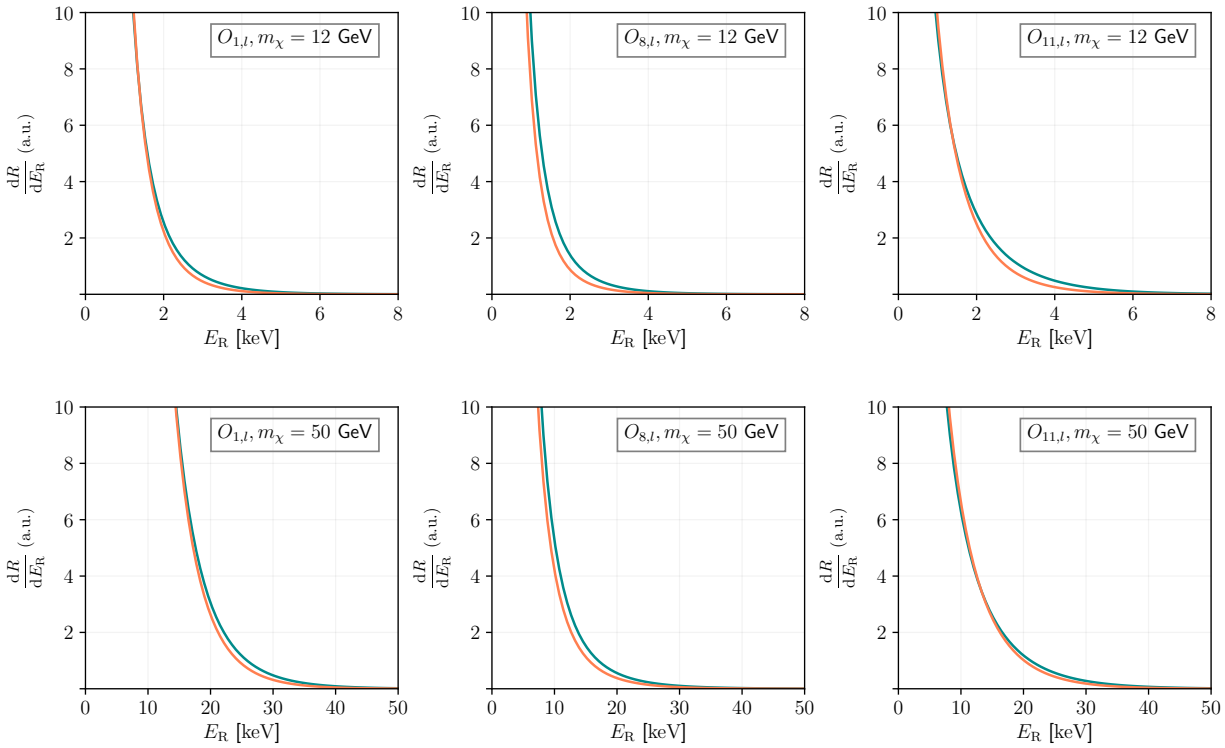


Figure 4.5: Recoil spectra of $\mathcal{O}_1$, $\mathcal{O}_8$ and $\mathcal{O}_{11}$ for DM with mass $m_\chi = 12$ (top) and 50 (bottom) GeV. Here we consider operators with light mediators. The cyan (orange) curve assumes SHM (*Gaia*) velocity distributions.

$F_{i,j}^{(n)}(q^2)/(q^2 + m_{\rm med}^2)^2$. The form factor function $F_{i,j}^{(n)}(q^2)$ is, in general, a polynomial of q^2 multiplying an exponent suppression factor $\sim e^{-q^2 s^2}$, in which the effective nuclear radius is $s \approx 0.9$ fm. Since the exponential factor is common for all nuclei, we will focus on the polynomial part of $F_{i,j}^{(n)}(q^2)$ that depends on the type of DM-nucleon interaction.

The momentum transfer is related to the recoiling energy E_R and $v_{\rm min}$ as,

$$q^2 = 2m_T E_R = 4\mu_T^2 v_{\rm min}^2. \tag{4.11}$$

We could express q^2 in terms of $v_{\rm min}$ and the differential recoil rate as a function of $v_{\rm min}$ only. To illustrate how the recoil spectra vary with velocity distributions, we select three NR operators as examples

$$\mathcal{O}_1 = \mathbb{1}, \qquad \mathcal{O}_8 = \boldsymbol{S}_\chi \cdot \boldsymbol{v}^\perp, \qquad \mathcal{O}_{11} = i\boldsymbol{S}_\chi \cdot \boldsymbol{q}, \tag{4.12}$$

and compute their recoil spectra assuming a xenon target. For simplicity, we consider only one NR operator at a time with a single type of mediator. Let's consider two limits to illustrate the effects of different DM velocity distributions: heavy mediator with $m_{\rm med}^2 \gg q^2$ so that the propagator is approximately a constant, and light mediator with $m_{\rm med}^2 \ll q^2$ so that the propagator is approximately $1/q^2$. It is not difficult to generalize the discussion to cases in between the two limits with $m_{\rm med}^2 \sim q^2$. In the two limits, we have the dependence of the recoil spectrum on $v_{\rm min}$ for each operator schematically as

$$\begin{aligned} &\mathcal{O}_1 : g(v_{\rm min}), \qquad \mathcal{O}_8 : v_{\rm min}^2 g(v_{\rm min}) + h(v_{\rm min}), \qquad \mathcal{O}_{11} : v_{\rm min}^2 g(v_{\rm min}), \quad \text{heavy mediator;} \\ &\mathcal{O}_1 : \frac{g(v_{\rm min})}{v_{\rm min}^4}, \qquad \mathcal{O}_8 : \frac{g(v_{\rm min})}{v_{\rm min}^2} + \frac{h(v_{\rm min})}{v_{\rm min}^4}, \qquad \mathcal{O}_{11} : \frac{g(v_{\rm min})}{v_{\rm min}^2}, \qquad \text{light mediator.} \end{aligned} \tag{4.13}$$

To derive the spectrum's dependence on $v_{\rm min}$ for $\mathcal{O}_8$, we use $\boldsymbol{v}^\perp = \boldsymbol{v} + \frac{\boldsymbol{q}}{2\mu_T}$ and $\boldsymbol{v}^\perp \cdot \boldsymbol{q} = 0$.

We could understand some general features of the spectra independent of the velocity distributions. Given the exponential factor in the nuclear response function, all spectra fall off at large recoiling energies. In the heavy mediator limit, the spectrum of $\mathcal{O}_1$ peaks at low recoil energy (or equivalently, small $v_{\rm min}$) since $g(v_{\rm min})$ is a monotonic decreasing function of $v_{\rm min}$. The differential rate of $\mathcal{O}_{11}$ contains an additional factor, $v_{\rm min}^2$, which prefers larger E_R. Thus the spectrum peaks at a higher E_R, away from the detection threshold, due to a balance between $v_{\rm min}^2$ and $g(v_{\rm min})$. For $\mathcal{O}_8$, the differential rate is a sum of two terms with opposite behavior: $v_{\rm min}^2 g(v_{\rm min})$ peaks at large recoiling energy while $h(v_{\rm min})$ peaks at threshold. Numerically it turns out that $h(v_{\rm min})$ dominates over $v_{\rm min}^2 g(v_{\rm min})$ so that the spectrum still peaks at low recoil energy. The situation is much simpler in the light mediator limit. The spectra for all operators peak at threshold due to the mediator propagator $1/q^2$.

We present the spectra for both light and heavy DM with masses at 12 and 50 GeV respectively in Fig. 4.4 (in the heavy mediator limit) and Fig. 4.5 (in the light mediator limit). In the plots, the cyan curves are based on SHM while the orange curves are based on the *Gaia* distribution with $\eta_{\rm sub} = 0.42$. Based on the discussion above, we could understand further some details of the spectra for each velocity distribution.

In the heavy mediator case,

- For both $\mathcal{O}_1$ and $\mathcal{O}_{11}$, spectra based on *Gaia* peaks at a higher value at low recoiling energy and falls off faster compared to the one from SHM given the shapes of $g(v_{\rm min})$ shown in Fig. 4.3a.

- For $\mathcal{O}_1$, the spectrum based on *Gaia* distribution is steeper for light DM than that for heavy DM. More specifically, the recoiling energy at which the differential rate from SHM becomes larger than that from *Gaia* for heavy DM (the cross-over of the two curves) is larger than that for light DM. This is because at $v_{\rm min} \sim 250$ km/s, the relative sizes of $g(v_{\rm min})$'s for *Gaia* distribution and SHM switches (Fig. 4.3a). For a given $v_{\rm min}$, the more heavy DM is, the larger E_R it corresponds to, as one could see from Eq. (4.11). Similar argument could be used to explain the differences for spectral of light and heavy DM scattering through $\mathcal{O}_{11}$.

- For $\mathcal{O}_8$, the spectra based on *Gaia* are always below those based on SHM since the scattering is mostly determined by $h(v_{\rm min})$.

Similar features are present in the light mediator case (though they are less evident in Fig. 4.5):

- For $\mathcal{O}_1$ and $\mathcal{O}_{11}$, the *Gaia* spectra are steeper than those of SHM. At low E_R, the differential rate based on the *Gaia* distribution is greater than that based on SHM while the opposite is true at higher E_R. Analogous to the heavy mediator case, the cross-over happens at a larger E_R for heavy DM compared to light DM.

- For $\mathcal{O}_8$, *Gaia*'s differential rate is always below the SHM's mainly due to the dominance of $h(v_{\rm min})$.

In general, the overall rate of light DM with mass $\lesssim 10$ GeV could be suppressed with a *Gaia* distribution compared to that of SHM. This leads to weaker constraints and poorer determinations of the light DM parameters using *Gaia* velocity distribution. On the other hand, for heavy DM, the relative sizes of scattering rate with either *Gaia* or SHM distributions depends on the type of the interaction. The scattering rate with *Gaia* could be

enhanced when the associated velocity moment is $g(v_{\rm min})$ and the scattering rate is proportional to a non-negative power of $v_{\rm min}$ (or equivalently, q). This could lead to stronger constraints and better determinations of the DM parameters, e.g., when the model maps onto $\mathcal{O}_1$ and $\mathcal{O}_{11}$ with heavy mediators. For interactions associated with $h(v_{\rm min})$, however, the recoil rates assuming *Gaia* distribution are suppressed compared to that of SHM, resulting in a weaker constraint on the coupling. Lastly, the differences of the spectral shapes in the light mediator case, which depend on negative powers of $v_{\rm min}$ (q), with either *Gaia* or SHM distributions are small. These qualitative discussions based on the recoil spectra will indeed be confirmed with numerical computations in Sec. 4.5.

4.4 Statistical framework

In this section, we introduce the statistical framework to study the effect of uncertainties in the DM velocity distribution while reconstructing particle physics parameters with DD data, followed by a short discussion on how to interpret its results with a concrete example.. As discussed previously in the introduction, for a given DD experiment, there is a three-fold degeneracy between the different classes of DM parameters. At the same time, from a statistical inference perspective, we can only access (a combination of) these parameters through experimental observables. For DM-nuclear interactions, these observables are simply the *overall rate* and the number of events per bin or the *shape* of the recoil spectrum (which may also include background events). The qualitative relationships between these observables and DM parameters of interest are summarized in table 4.3.

However, no statistically significant number of DM events have been detected at PICO [275], LUX [276], SUPERCDMS [277], PANDAX-II [278], and XENON-1T [279]. These null results in turn can be used to obtain constaints on particle physics parameters, traditionally expressed through 90% confidence level (CL) upper limit in the plane of dark matter mass and scattering cross section per nucleon. While studying upper limits is the most straightforward way to assess the impact of qualitatively different velocity distributions on extracting particle physics parameters with DD experiments [60, 245, 261, 265, 266], we could gain more

Type	Observables / Signal parameters	Total rate	Shape
Particle and nuclear physics	DM mass [m_χ]	✓	✓
	Couplings [c's]	✓	✗*
	Mediator mass [$m_{\rm med}$]	✓	✓**
	Form factor [$\mathcal{F}$'s]	✓	✓
Astrophysics	Local DM density [ρ_χ]	✓	✗
	DM velocity distribution [$f(v)$]	✓	✓

* Exceptions occur when multiple relativisitc interactions are relevant, each giving rise to a different shape. Then varying couplings (non-uniformly) change the weights of each interaction.

** Applicable only for light mediators, *i.e* when $m_{\rm med}/q \lesssim 1$ as well as the case with multiple mediators.

Table 4.3: Schematic summary of the relationship between DM signal parameters and experimental observables in a typical DD experiment.

information from forecasting the ability of next-generation experiments to reconstruct model parameters through pairwise comparison of neighboring points. We also note that several analyses in the literature have forecasted degeneracies between DM model parameter by analyzing mock data sets generated for various experimental configurations [239–244], but these methods rely on the choice of several pre-defined benchmark points and some of them could be computationally expensive.

A faster alternative for forecasting are techniques that rely on the so-called Asimov data set [280], an artificial data with no statistical fluctuations generated using the true parameter values of a model. Consider a d-dimensional parameter space, $\boldsymbol{\theta} \in \Omega_{\mathcal{M}} \in \mathbb{R}^d$ for a given DM model $\mathcal{M}$, and the associated Asimov data set $D_A(\boldsymbol{\theta})$ for each point. Given two model parameter points $\boldsymbol{\theta}_{1,2} \in \Omega_{\mathcal{M}}$, we can construct a likelihood ratio test statistic (TS) [281],

$$\mathrm{TS} = -2\ln\frac{\mathcal{L}_X(\mathcal{D}_A(\boldsymbol{\theta}_2)|\boldsymbol{\theta}_1)}{\mathcal{L}_X(\mathcal{D}_A(\boldsymbol{\theta}_2)|\boldsymbol{\theta}_2)} \approx \sum_{i,j=1}^{n_b}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)_i\,\widetilde{I}_{ij}\,(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)_j \sim \chi^2_d\,, \tag{4.14}$$

which asymptotically has a χ^2 distribution with d degrees of freedom [282]. Here, $\tilde{I}$ is the profiled Fisher information matrix, and the summation runs from 1 to n_b, the number of data bins. While the TS is used to reject the null hypothesis that $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are indistinguishable at the $(1-\alpha)\%$ confidence level, Eq. 4.14 can also be suitably modified to obtain sensitivity projections for future experiments in terms of 90/95% CL upper limits.

Despite using the Asmiov data set, calculating the TS in Eq. (4.14) for model comparison of N points in the parameter space can be expensive when N is large. Thus, to facilitate fast, *benchmark-free* model comparison, refs. [237, 238] introduced a novel method based on information theoretic techniques.[6] Noting that the profiled Fisher information $\widetilde{I}$ transforms as a *metric* on the parameter space $\boldsymbol{\theta}$, ref. [283] mapped the parameter space into a higher-dimensional signal space and expressed the TS as a euclidean distance between two signals. More concretely, they used the embedding, $\boldsymbol{\theta} \mapsto x(\boldsymbol{\theta}) \in \mathbb{R}^{n_b}$, to transform the parameter space to the n_b-dimensional signal space with unit Fisher information matrix. After this transformation the TS can be written in terms of the appropriately named euclidean signal x_i,

$$TS \approx ||\boldsymbol{x}(\boldsymbol{\theta}_1) - \boldsymbol{x}(\boldsymbol{\theta}_2)||^2. \tag{4.15}$$

Eq. (4.15) is the main ingredient of our benchmark-free forcasting approach. In the language of this method, as long as the parameter space is sufficiently sampled, signal discrimination is only possible at the $(1-\alpha)\%$ CL if the signals from two parameter points are at least a distance $r_\alpha(\mathcal{M})$ apart in the projected signal space. The distance, in turn, is related to the sampling distribution of the TS,

$$r_\alpha(\mathcal{M})^2 \leq P^{-1}_{\chi^2_d}(1-\alpha), \tag{4.16}$$

where $P_{\chi^2_d}$ is the cumulative distribution function (CDF) of the χ^2 distribution with d degrees of freedom.

The above procedure can be understood very loosely as comparing two distributions, albeit incorporating the fact that they arise from the same likelihood function. We emphasize

[6] A `python` implementation of their results is available in the open source `swordfish` code: https://github.com/cweniger/swordfish, and a proof-of-concept application to future DD searches has been demonstrated in ref. [283].

that there is no mock data generated at any stage of our analysis. The use of Asimov data set in eq. (4.15) implies that the sampling distribution gives the median significance for two hypothetical data sets which have the parameter points $\boldsymbol{\theta}_{1,2}$ as their maximum likelihood estimates. We illustrate the efficacy of this method in Fig. 4.6, the left panel of which shows the constraints in DM mass-coupling space for a model with contact interaction mediated by a heavy vector particle. The closed ellipses represent the usual 68% CL contours in parameter space for arbitrary benchmark points at a next-generation DARWIN-like liquid Xenon (LXe) experiment [217] assuming SHM. These are obtained by constructing hyperspheres of radius, r_α, in the Euclidean signal space, and back-projecting them to the parameter space using a lookup table for the embedding map. For a χ^2_d distribution with $d = 2$, eq. (4.16) implies that 68% CL corresponds to a threshold value of $r_{0.32} = 1.52$.

4.5 Results

In this section, we combine the formulae and methodology presented in previous sections for studying how DM velocity distribution inferred from *Gaia* sausage could affect reconstruction of various DM particle physics parameters at next-generation DD experiments. For concreteness, we only consider a DARWIN-like liquid Xenon (LXe) experiment and a complementary DARKSIDE-20K-like Argon experiment with both high [218] and low mass [284] search programs (see Appendix C.4 for more details). Unlike Sec. 4.3, we present our results for the *benchmark* DM models listed in Table 4.2 instead of individual operator in NREFT. While examining recoil spectrum of each operator is insightful, concrete models, especially those with well-motivated UV completions and/or distinct phenomenologies, enable an easy comparison of our results with those in the literature. In addition, since there could be non-trivial mapping between a model and NREFT, it may not be straightforward to find the sensitivity of DD to a full-fledged model by combining the sensitivity to individual NR operators. Yet we will still find the qualitative understanding developed in Sec. 4.3 a useful starting point for the results discussed here.

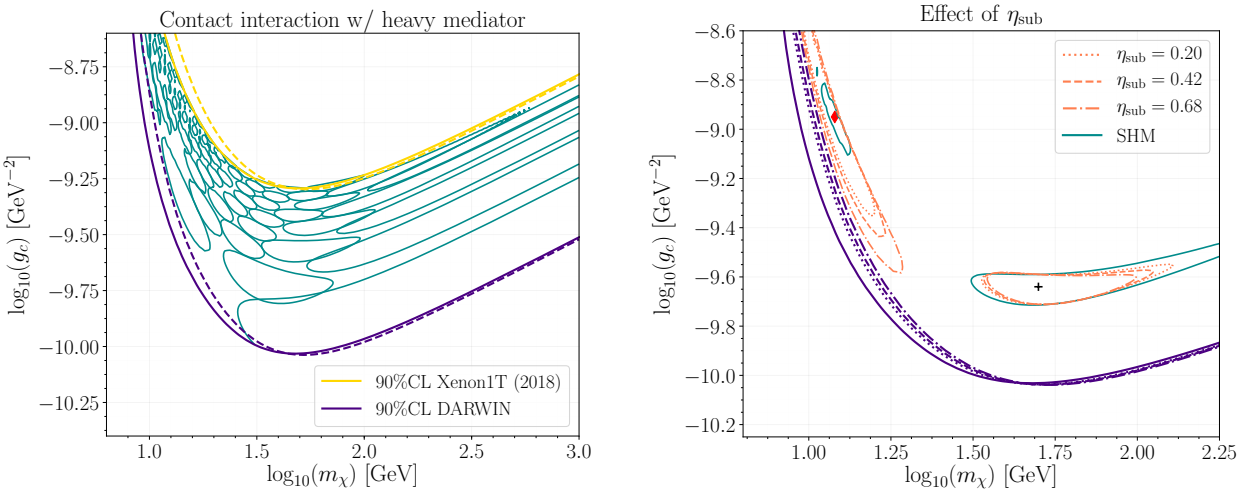


Figure 4.6: Left: the 'fishnet' plot obtained using the ES method is capable of illustrating the degeneracies between various pairs of signal parameters without recourse to computationally expensive MC simulations. The closed ellipses represent the usual 68% CL contours in parameter space for arbitrary benchmark points at a DARWIN-like experiment. Also indicated for reference are 90% CL upper limits following the latest XENON-1T results (yellow) and projected upper limits for a DARWIN-like experiment (indigo) assuming SHM (solid) and *Gaia* (dashed) velocity distributions. Right: 68% CL contours and 90% CL upper limits assuming a *Gaia* velocity distribution with different DM substructure fractions (dotted, dashed, and dot-dashed). These are shown alongside the SHM (solid) constraints to demonstrate that the dominant uncertainty is due to differences in the velocity distributions, and not the DM substructure fraction.

4.5.1 DM Mass - coupling

We forecast the sensitivity of a next-generation LXe, DARWIN-like, experiment to simultaneously resolve the DM mass and coupling for SHM and *Gaia* velocity distributions. Our first step is to investigate the effect of uncertainty in the DM substructure fraction $\eta_{\rm sub}$, estimated by ref. [61] to be $\eta_{\rm sub} = 0.42^{+0.26}_{-0.22}$. We show, in the right panel of Fig. 4.6, the 68% CL contours and 90% CL upper limits corresponding to a *Gaia* velocity distribution with the median and $\pm 1\sigma$ DM substructure fractions. Comparing them with constraints for the

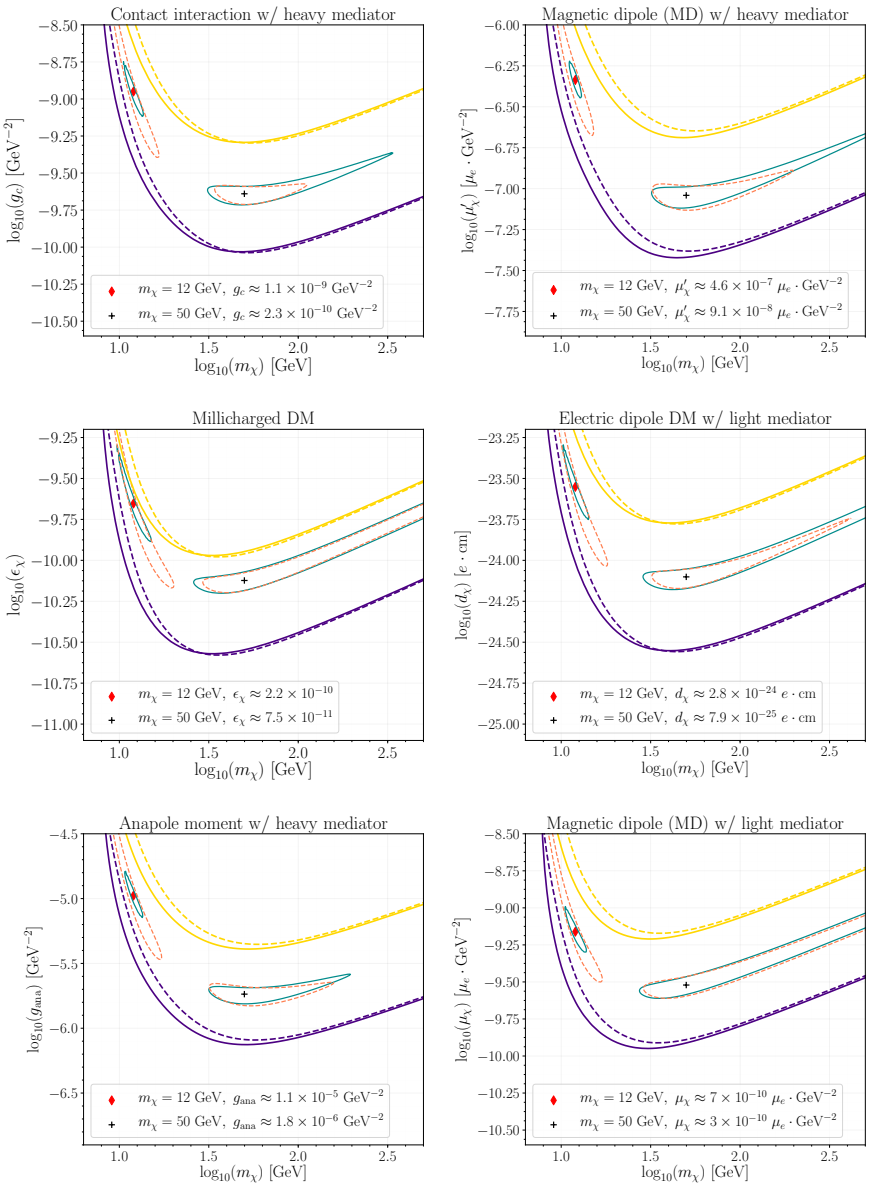


Figure 4.7: Constraints and forecasts in the DM coupling-mass plane for all the benchmark models in Table 4.2 with varying q^2 and v^2 dependence. The 68% CL forecast contours for SHM (cyan, solid) and *Gaia* (orange, dashed) velocity distributions are shown for both light (red diamond, $m_\chi = 12$ GeV) and heavy (black cross, $m_\chi = 50$ GeV) DM. Also indicated for reference are 90% CL upper limits following the latest XENON-1T results (yellow) and projected upper limits for a DARWIN-like experiment (indigo) assuming SHM (solid) and *Gaia* (dashed) velocity distributions. The constraints for MD with heavy mediator are quoted in units of electron Bohr magneton, $\mu_e = \frac{e}{2m_e}$.

SHM leads us to conclude: *the primary effect in reconstructing DM model parameters arises due to the qualitative differences between the SHM and Gaia velocity distributions, while the variation of the substructure fraction is only a subdominant effect beyond it.* Thus, for the rest of our analysis, we fix the DM substructure fraction to its median value, $\eta_{\rm sub} = 0.42$.[7] We have also checked that varying $\eta_{\rm sub}$ in its 1σ range does not affect our discussion for DM interactions beyond the minimal contact interaction.

In Fig. 4.7, we show the 68% CL contours obtained using the ES method for four DM models with characteristic E_R and velocity moment dependences (outlined in table 4.2), namely contact interaction, millicharged DM, DM with electric and magnetic dipole moments, anapole DM with a heavy mediator and magnetic dipole DM with a light mediator for SHM and *Gaia* velocity distributions. For reference, we also plot the 90% CL upper limits following the latest XENON-1T results [279] and projected upper limits for future LXe experiments using the *equivalent counts* method [237, 238]. We constrain couplings instead of cross-sections (cf. [242, 283]) as we don't integrate over the entire E_R range to obtain the respective cross-sections for the DM models we consider here. Although, as illustrated in the left panel of Fig. 4.6, the ES method allows us to plot the degeneracy contour for any point in the parameter space, we show our results at two benchmark points corresponding to light ($m_\chi = 12$ GeV) and heavy ($m_\chi = 50$ GeV) DM for easier interpretability of our results. To ensure we are making an apples-to-apples comparison when studying the changes in constraints across models, we choose couplings such that the number of events is the same for each benchmark point with SHM.

Before discussing the effect of DM velocity distribution, we explain the general behavior of constraints in DM coupling-mass space given in Fig. 4.7 in terms of the recoil spectra shape and the total event rate. For a given DM model, the recoil spectra for low mass DM peaks closer to threshold than for heavy DM. Moreover, the shape of the recoil spectra is degenerate only for a narrow range of masses, whereas a change in the total rate can be compensated by a wide range of couplings. As a result, in the light DM regime, we observe

[7]The ES method also provides a straightforward way to marginalize over the uncertainty in nuisance parameters, such as $\eta_{\rm sub}$, through the inclusion of a penalisation term; see Appendix A of ref. [283] for more details.

a large degeneracy in the coupling but a reasonable DM mass resolution. Conversely, for heavier DM, a DD experiment is more sensitive to the couplings but suffers from poor mass resolution, because the recoil spectra shape is degenerate for a wide range of DM mass. Thus, given a similar number of events, we can heuristically treat the light and heavy DM regimes as shape- and total rate-limited respectively.

While these interpretations hold generally, certain qualitative details like the shapes and sizes of the contours can vary significantly between the SHM and *Gaia* velocity distributions. For instance, in the light DM regime, we observe that the contour size increases for all models with *Gaia* velocities, implying that LXe experiments have reduced sensitivity to *Gaia* distribution as compared to the SHM. This observation is consistent with the weakening of upper limits on SI cross-section for light DM first reported by ref. [60] with the latest XENON-1T data. Our results also indicate an interesting effect of *Gaia* velocity distribution that has not been previously discussed in the literature: *depending on the model, there is a marginal improvement in the sensitivity of LXe experiments to heavy DM for Gaia when contrasted with the SHM.* More concretely, in the case of contact interaction, when dark matter mass is at 12 GeV, the resolution of coupling assuming *Gaia* distribution is reduced by a factor of 7 compared to SHM. On the other hand, when dark matter mass is at 50 GeV, the mass resolution is improved by a factor of 3 compared to SHM. We also note that for models with light mediators, the experimental sensitivity becomes poorer across the entire mass range irrespective of the DM velocity distribution.

These results can be understood, at least to leading order, in terms of the E_R and velocity moment dependences of each model. We start with the observation that contours of the *Gaia* velocity distribution are less constraining than those of SHM at low DM masses for all models we consider here. These models have a leading order DM velocity moment that scales as $g(v_{\rm min})$ and/or $h(v_{\rm min})$ suppressed or enhanced by additional powers of E_R. Since light DM corresponds to a high $v_{\rm min}$ for a heavy target like xenon, only the tail of the $g(v_{\rm min})$ and/or $h(v_{\rm min})$ distribution (Fig. 4.3) contributes to the recoil rate, where the SHM curves always dominate over the *Gaia* ones.

For heavy DM, on the other hand, varying DM mass could lead to sharper changes in the

recoil spectra shapes with the *Gaia* distribution as compared to the SHM. Thus, there is an improvement in the sensitivity to DM mass as evidenced by the shrinking of 68% CL forecast contours in the mass direction in Fig. 4.7. Moreover, as shown in the top row, this effect is most apparent for models with a heavy mediator, or equivalently for non-negative powers of E_R. We also note that there could be some subtle difference, e.g., between models with contact interaction and magnetic dipole interaction. For contact interaction, the upper limits for DM with mass above 50 GeV are slightly tightened with the *Gaia* distribution. This is due to an enhanced recoil rate contributed by *Gaia*'s larger $g(v_{\rm min})$ at $v_{\rm min}\sim$150-200 km/s as compared to SHM. Whereas, in the case of magnetic dipole DM, the *Gaia* upper limits for DM with mass above 50 GeV are slightly weakened, since the scattering rate of magnetic dipole DM scales as $E_R^2 g(v_{\rm min}) + E_R h(v_{\rm min})$. The second term proportional to $h(v_{\rm min})$ leads to a small reduction in the overall recoil rate with the *Gaia* distribution, as compared to the SHM one. Meanwhile, the positive powers of E_R for magnetic dipole interaction result in an enhanced sensitivity of forecasts using the *Gaia* distribution. This is an interesting example where despite the reduction in the total number of events, the sensitivity actually improves with the *Gaia* distribution!

In case of light mediators, the inverse powers of E_R make the recoil spectra peak sharply as $E_R \to 0$. Yet for a finite threshold $E_R \approx 5$ keV, the DARWIN-like experiment is only sensitive to the tail resulting in highly degenerate recoil spectra for different velocity distributions. This leads to poorer experimental sensitivity for both SHM and *Gaia* velocity distributions across all DM masses for models with light mediators, compared to the contact interaction model with a heavy mediator. In this case, lowering the detection threshold could improve the sensitivity to light mediators, making another physical case for the low threshold frontier.

4.5.2 Mediator - DM mass

DD experiments are best suited to constrain the mediator mass $m_{\rm med}$ when it is at the same order of the momentum exchange, i.e. $m_{\rm med} \sim q \sim \mathcal{O}(10\,{\rm MeV})$ [244, 285, 286]. We use the ES method to simultaneously constrain the mediator and DM masses at a DARWIN-like

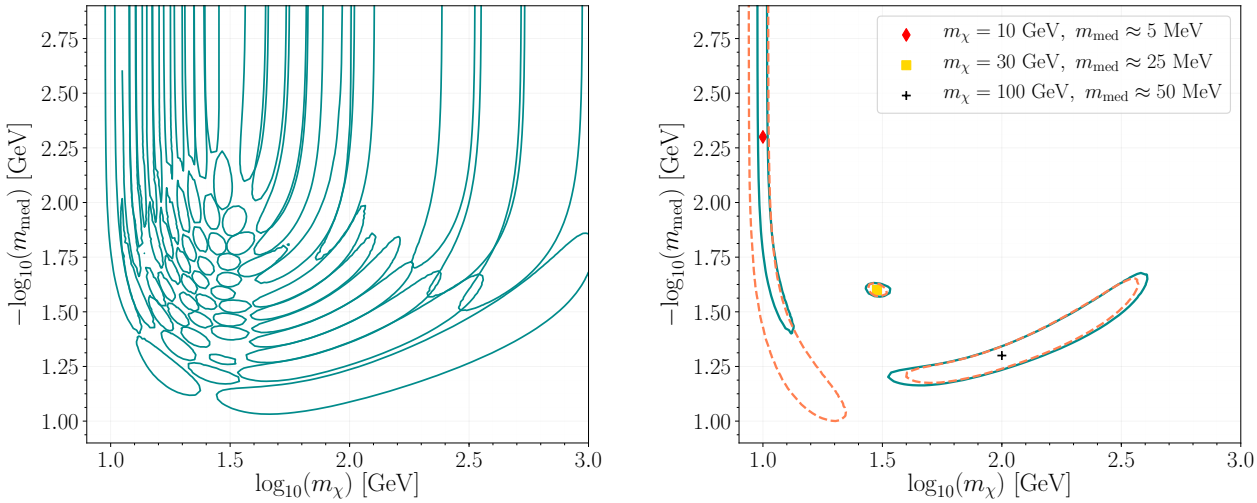


Figure 4.8: Forecasts in the $m_{\rm med}^{-1} - m_\chi$ plane for a DM with contact interaction mediated by a light scalar particle. Left: lattice of non-overlapping 68% CL forecast contours for the SHM. Right: 68% CL contours for SHM (cyan, solid) and *Gaia* (orange, dashed) velocity distributions shown for three benchmark points corresponding to light (red diamond), intermediate (yellow square), and heavy (black cross) DM with mediator masses in the 1-50 MeV range.

experiment for a fixed coupling. We also study the effect of the DM velocity distribution on the forecast. For simplicity, we only consider DM with contact interaction mediated by a light scalar particle and fix the DM coupling to be the same for all benchmark points.

In the left panel of Fig. 4.8, we show the 68% CL forecast contours for SHM in the $m_{\rm med}^{-1} - m_\chi$ plane. Broadly, the structure of these contours resembles the fishnet plot in the DM mass-coupling plane shown in the left panel of Fig. 4.6. On closer inspection, however, we can roughly delineate three regimes of sensitivity in this parameter space. The upper part of the plots is the light mediator ($m_{\rm med} \ll q$) regime where the propagator squared simply scales as $\sim 1/q^4$. In this case, the $m_{\rm med}$ dependence drops out, and DD experiment is insensitive to the mediator mass. Next, we consider the heavy DM regime in the lower right part of the plots, where we have chosen the mediator mass such that $m_{\rm med} \lesssim q$. We find that the degeneracy in the recoil spectra in this limit is due to the DM mass, and any

change in the mediator mass effectively acts as a rescaling of the overall coupling. Lastly, for intermediate DM mass and $m_{\rm med} \lesssim q$, our DARWIN-like experiment can precisely reconstruct both mediator and DM masses primarily due to the high number of signal events (a factor of a few greater than the other benchmark points) in this regime.

We also illustrate the differences between SHM and *Gaia* velocity distributions in the right panel of Fig. 4.8 by plotting the 68% CL contours for three benchmark points: light ($m_\chi = 10$ GeV), intermediate ($m_\chi = 30$ GeV), and heavy ($m_\chi = 100$ GeV) DM with mediator masses in the 1-50 MeV range. The constraints for SHM and *Gaia* velocity distributions at the light and heavy DM benchmark points broadly follow the trend discussed for Fig. 4.7 in the previous section, while the difference between them is negligible in the intermediate DM mass regime.

4.5.3 Model discrimination

In presence of a positive signal at a future DD experiment, one of the most important goals is to determine the type of DM-nuclear interaction and discriminate between different model candidates. To demonstrate the model selection, we postulate a scenario in which there are two candidate models of interest. We parameterize our model as the sum of a pair of interactions,

$$\mathcal{L}_{\rm int} = \frac{c_a \mathcal{O}_a}{q^2 + m^2_{{\rm med};a}} + \frac{c_b \mathcal{O}_b}{q^2 + m^2_{{\rm med};b}}. \tag{4.17}$$

While holding m_χ and $m_{\rm med}$ fixed, we sample different values of (c_a, c_b) and test how well a given recoil spectrum shape can determine the model parameters at 68% confidence level. We test the two pairs of models: 1) a, b = contact interaction, millicharge, and 2) a, b = magnetic DM with light mediator and millicharge respectively. We present the results in Fig. 4.9 for light DM with mass at 12 GeV and Fig. 4.10 for heavy DM with mass at 50 GeV. These results are all based on a DARWIN-like LXe experiment.

In Fig. 4.9, we present 68% CL forecast contours in the coupling-coupling space for a 12 GeV DM particle. In the top row, we have a millicharged DM giving rise to an experimental signal and we want to test whether we would confuse it with the simplest contact interaction

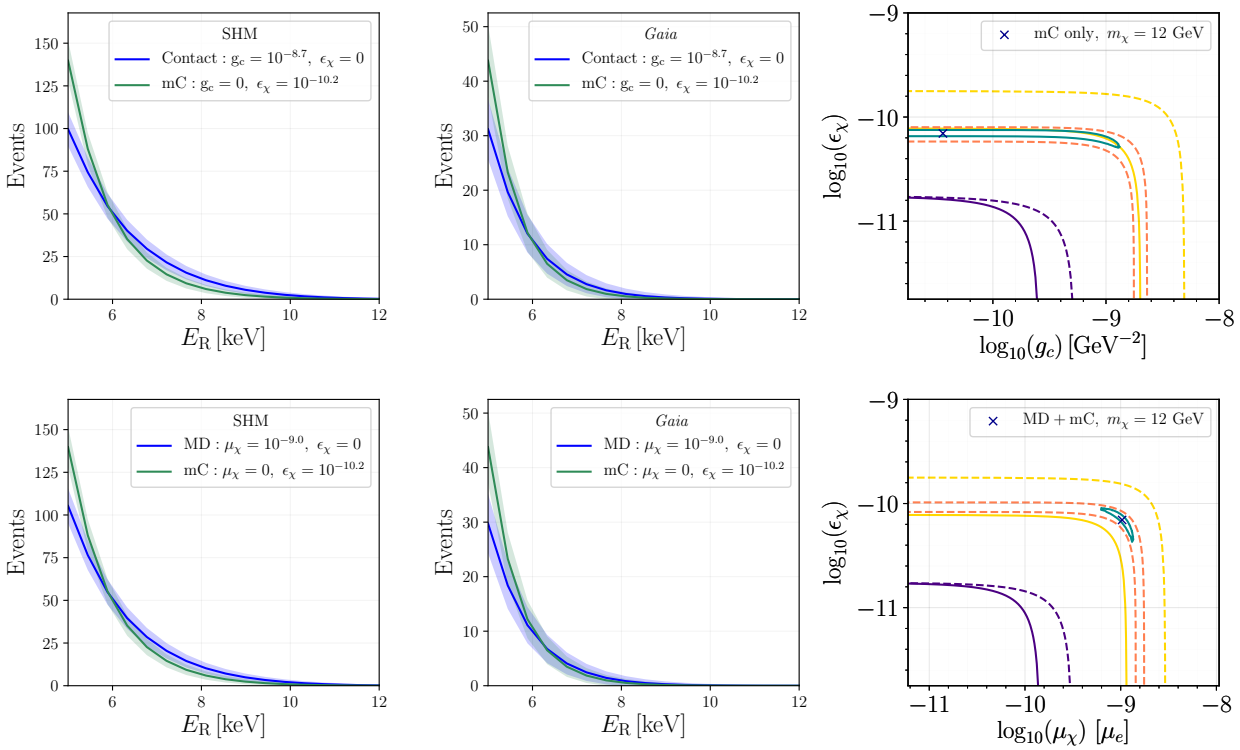


Figure 4.9: Forecasts for model discrimination in case of a 12 GeV DM at a DARWIN-like experiment. The first two columns are recoil spectra for the two models under comparison with either SHM or *Gaia* velocity distributions, where the light-colored bands indicate the 1σ Poisson uncertainties. The third column shows the 68% CL forecast contours for SHM (cyan) and *Gaia* (orange) in the coupling-coupling space for two comparison models: (first row) millicharge with a light mediator and contact interaction with a heavy mediator, (second row) millicharge and magnetic dipole both with a light mediator (SM photon). We also include 90% CL upper limits from the latest XENON-1T results (yellow) and projected upper limits for a DARWIN-like experiment (indigo) assuming SHM (solid) and *Gaia* (dashed) velocity distributions.

as both interactions lead to spectra peaking at experimental threshold. This scenario is equivalent to setting $c_a \neq 0, c_b = 0$ in eq. (4.17), where $c_a = \epsilon_\chi e$ for the millicharge model and $c_b = g_c$ for the contact interaction. In the bottom row, we have a DM particle with both magnetic dipole moment and millicharge interacting with the nucleus through the SM

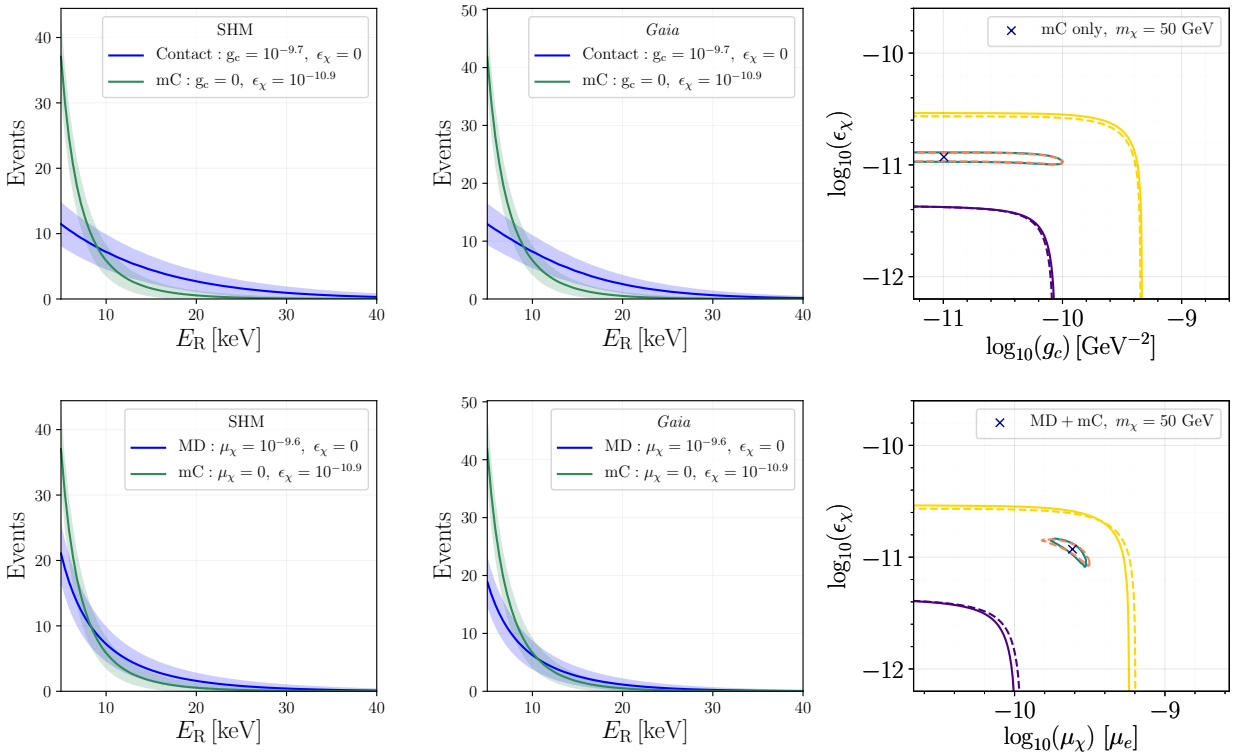


Figure 4.10: A similar plot as Fig. 4.9, showing the same set of model pairs comparison, except that the DM mass is 50 GeV.

photon contributing with comparable rates, where $c_a = e\mu_\chi/2$ and $c_b = \epsilon_\chi e$. We want to test how well we could constrain the two relevant electromagnetic moments.

From the third column, one could see that while it is possible to reconstruct the model parameters (with large uncertainties) assuming SHM, the discrimination power is entirely lost with the *Gaia* distribution. The millicharged light DM could be misidentified as a light DM with simple contact interaction at 68% C.L, as shown in the last plot in the top row. DM with both millicharge and magnetic dipole moment could not be distinguished from DM with only one of them, as shown in the last plot in the bottom row. This result could be understood from the recoil spectra shown in the first two columns, in which we fix the couplings of different DM models to give the same event numbers with SHM. Comparing the spectra based on the *Gaia* velocity distribution with those from SHM, one find that with *Gaia* distribution: *i)* the total number of events is significantly lower, which, in turn,

increases the Poisson uncertainty; *ii)* the spectral shapes of different models, especially the tails of the distributions, are more degenerate. This is consistent with what we find in Sec. 4.3 using NREFT.

Results of model discrimination for heavy DM with mass at 50 GeV are presented in Fig. 4.10. In this case, we find that reconstructing the couplings is more accurate as compared to the light DM case, independent of the model combinations. We could almost distinguish different models or different combinations of model parameters equally well for both SHM and the *Gaia* distribution. The main reason is that the total number of events and the spectral shape do not change much when the velocity distribution varies.

4.5.4 Combining forecasts for different targets

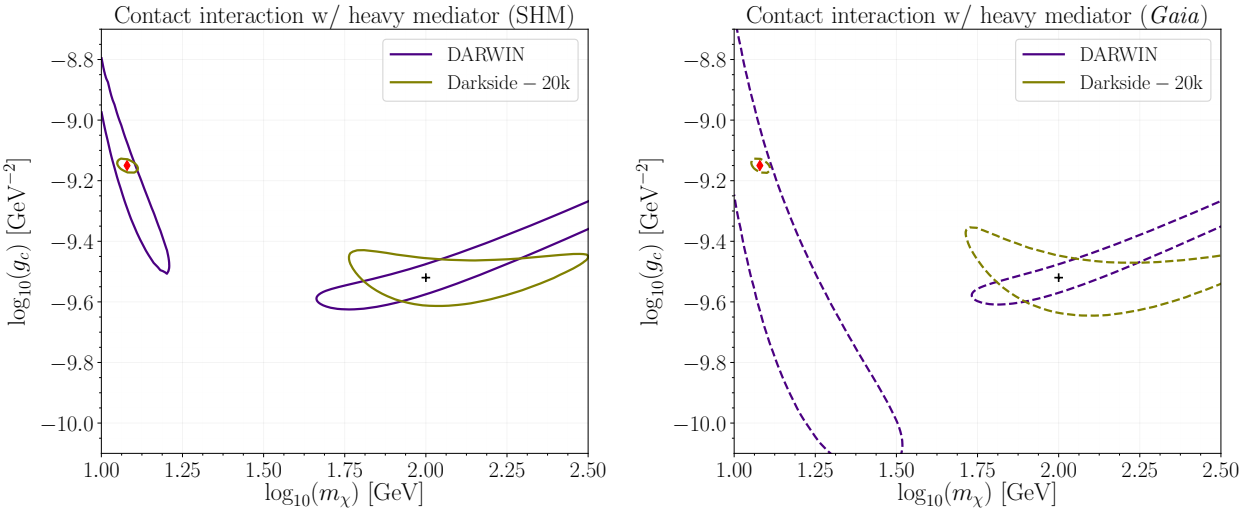


Figure 4.11: Forecasts in the DM coupling-mass plane for DM contact interaction assuming SHM (left) and *Gaia* (right) velocity distributions. Unlike Fig. 4.7, the 68% CL contours here represent forecasts for two complementary next-generation experiments: a DARWIN-like experiment (indigo) and a DARKSIDE-20K-like experiment (olive) with xenon and argon targets respectively. Also indicated for reference are two benchmark points for light (red diamond, $m_\chi = 12$ GeV) and heavy (black cross, $m_\chi = 100$ GeV) DM.

The idea of combining different targets for a more accurate identification of momentum

dependence of DM interactions, or for precise reconstruction of DM mass and model parameters is well-documented in the literature [239, 242, 287–289]. We revisit this idea by using the ES method to forecast the sensitivity of two complementary next-generation experiments with different targets. In particular, we focus on a DARWIN-like and a DARKSIDE-20K-like experiment with xenon and argon targets respectively. Besides forecasting for high mass DM as officially proposed by the DARKSIDE-20K collaboration [218], we also include results for a dedicated low DM mass search with a low threshold configuration similar to ref. [284].

For simplicity, we consider the contact interaction as an example. Fig. 4.11 shows the 68% CL contours in the DM mass-coupling space for each of these experiments at two benchmark points corresponding to light ($m_\chi = 12$ GeV) and heavy ($m_\chi = 50$ GeV) DM. We find that combining forecasts from both xenon and argon targets could dramatically improve the sensitivity for low mass DM and, to a lesser extent, high mass DM. This result is independent of the model for DM velocity distribution. The extremely high resolution for low mass DM is due to the low threshold version of a DARKSIDE-20K-like experiment as listed in table C.1 of Appendix C.4. An advantage of using a lighter target like argon is that low DM masses correspond to much lower values of $v_{\rm min}$ as compared to xenon. This implies that the sensitivity forecast for a DARKSIDE-20K-like experiment are largely unaffected by the suppressed *Gaia* $g(v_{\rm min})$ distribution at large $v_{\rm min}$.

4.6 Conclusions and Outlook

The new insight into the substructure of MW's DM distribution provided by the *Gaia* survey forces us to move away from the simplest SHM and to re-evaluate astrophysical uncertainties in DD experiments. In this chapter, we investigate the effect of *Gaia* Sausage, one of the most established and representative substructures, on interpreting DD data for different DM models. We demonstrate that the new *Gaia* velocity distribution could result in potentially large modifications of both the overall scattering rate and the recoil spectral shape. Given the limited information from existing data sets, we focus on how the *Gaia* velocity distribution could affect forecasting at the next generation DD experiments with the euclideanized signal

method.[8]

We study the sensitivity of DD experiments to different combinations of model parameters for representative DM benchmark models and its potential to distinguish different DM models given the *Gaia* distribution. We summarize our main findings below:

- While there is still uncertainty in the fraction of DM in the substructure, the primary effect in reconstructing DM model parameters is due to the qualitative differences between the shapes of *Gaia* and SHM distributions, which are independent of the precisie value of $\eta_{\rm sub}$.

- For light DM with mass at or below 10 GeV, the *Gaia* velocity distribution leads to a significantly weakened constraint for all the models we consider. Moreover, it poses a serious challenge for identifying the DM interaction and determining its strength assuming discovery at a DARWIN-like experiment with threshold $E_R \approx 5$ keV.

- On the contrary, for heavy DM with mass above ~ 30 GeV, there could be a (moderate) improvement in the sensitivity of the next generation DD experiments when the mediator is heavy.

- Moreover, our results show that for positive (negative) powers of E_R in the DM model, the sensitivity of a DARWIN-like experiment is improved (worsened) irrespective of the DM mass and velocity distribution.

- The additional challenge in probing light DM due to the *Gaia* distribution could be overcome using complementary experiments with lighter targets and lower thresholds.

As mentioned in the chapter's introduction, our work, along with several others, consists the early stages of a larger program to determine the DM phase space distribution, and assess its impact on various terrestrial experiments searching for DM in the *Gaia* era. In this chapter, we restrict ourselves to models with leading order SI elastic scattering for DM with

[8]Current experiments perform the full event-by-event likelihood analysis, e.g., as described in refs. [290, 291], which cannot be easily reproduced by researchers outside the collaboration. We hope that highlighting the importance of spectral shape information could motivate the experimental collaborations to release the full three dimensional likelihood function for reconstructed DM events in case of a positive detection.

mass above GeV. It will also be of interest to extend the work to models with leading order spin-dependent and/or inelastic scattering. We intend to explore how the substructures discovered using *Gaia* data affect the new DD experiments probing DM-electron scattering in a future publication.

Chapter 5

Concluding Remarks

In this thesis, we study three methods to understand the nature of DM: 1. Collider production. 2. Local gravitational potential. 3. Direct detection. Each method probes the dark sector in a different manner that is complementary to the others. The scientific community has continued effort to develop new methods, strategies and experiments to push forward all these fronts.

In the context of collider production, we study a set of simplified SUSY models in a future $\sqrt{s} = 100$ TeV collider using jet substructure techniques. As the LHC reaches its maximum discovery potential, a collider of higher center-of-mass energy is needed to explore new territories. A 100 TeV proton-proton collider is a good benchmark that it gives a final verdict on the naturalness of the SM Higgs and in general, is a great discovery machine at the TeV scale. Nonetheless, such a high energy machine does not come without its challenges. Protons colliding at 100 TeV generate hyper-boosted SM objects that are difficult to identify, for example different decay objects can fall into the same jet cone and can make discrimination difficult. We have developed strategies the combine the jet substructure technologies with conventional observables such as MET and HT, to improve our ability to constrain, discover and discriminate SUSY models with non-minimal mass spectra.

Concerning utilizing Milky Way gravitational potential, we measure local DM density and constrain thin dark disk using the DR2 of *Gaia* telescope. The *Gaia* provide billions of light sources in the sky with five dimensional phase-space coordinates (positions, parallax

and proper motions) and photometric information (brightness and color). Among them 100 million of stars will have also the 6th phase-space coordinate, radial motion (in *Gaia*'s ultimate scientific goal). We use the phase space distribution function method, combined with other literature estimations of the densities of baryonic matter, to estimate the amount of dark matter in the solar neighbor and the shape of the density profile in the vertical direction. We are able to constrain the density of a thin dark disk as a function of the scale height of the disk.

Finally, we forecast the implication of a recently discovered astrophysical substructure known as the *Gaia*-Sausage on future direct detection experiments, its ability to constrain, discover and discriminate various DM models. Different astrophysical substructures give different DM velocity profiles, from the shapes of which give different DD recoil spectra. Another factor that generates different recoil spectra is the DM model parameters, *e.g.* the mass and the type of interaction with SM matter. We examine how a future DARWIN-like liquid xenon experiment can identify DM model parameters under *Gaia*-Sausage. We also propose a DARKSIDE-20K-like liquid argon experiment can perform complementarily to DARWIN and provide better discrimination on DM model parameters regardless of the presence of astrophysical substructures.

Dark matter remains one of the most elusive and difficult problems for the physics community this century. Ongoing developments of the next-generation experiments provide us with the hope that the hints we need to solve the DM mystery is just around the corner. In terms of collider at the end of the High Luminosity-LHC life cycle, there are proposed high-energy circular discovery machines from a modest 27 TeV HE-LHC [292] to 100 TeV proposals [53, 293, 294]. On the astrometry front, *Gaia* [161] is currently carrying out the heavy labor, which gives us kinematic data of stars we need to find out the Milk Way gravitational potential. Far down the line, a proposed *Theia* satellite [295] will further advance astrometric observation essential to DM study, from identifying dwarf spheroidal galaxy halo profiles, the Milky Way halo shape and galactic disk gravitational perturbations, to searching for ultra-compact minihalos. On direct detection side, while many of the Gen-2 detectors are underway or planned, the upcoming Gen-3 [216–218] detectors will soon close the allowed

DM parameter window between the current sensitivity and the neutrino floor, beyond which DD experiments will be significantly less sensitive to DM detection.

In short, there are a lot of exciting upcoming experiments that will bring critical understandings to the nature of DM in the following decades. A lot more diligent work is needed to explore the potentials and challenges in the future ahead. New ideas are needed to bring all these opportunities to their full fruition.

Bibliography

[1] K. Schutz, T. Lin, B. R. Safdi and C.-L. Wu, *Constraining a Thin Dark Matter Disk with Gaia*, *Phys. Rev. Lett.* **121** (2018) 081101, [1711.03103].

[2] C. Flynn, J. Holmberg, L. Portinari, B. Fuchs and H. Jahreiss, *On the mass-to-light ratio of the local Galactic disc and the optical luminosity of the Galaxy*, *Mon. Not. Roy. Astron. Soc.* **372** (2006) 1149–1160, [astro-ph/0608193].

[3] J. I. Read, *The Local Dark Matter Density*, *J. Phys.* **G41** (2014) 063101, [1404.1938].

[4] C. F. McKee, A. Parravano and D. J. Hollenbach, *Stars, Gas, and Dark Matter in the Solar Neighborhood*, *Astrophys. J.* **814** (Nov., 2015) 13, [1509.05334].

[5] E. D. Kramer and L. Randall, *Interstellar Gas and a Dark Disk*, *Astrophys. J.* **829** (2016) 126, [1603.03058].

[6] T. Wright, *An original theory or new hypothesis of the universe : founded upon general phaenomena of the visible creation; and particularly the Via the laws of nature, and solving by mathematical principles : the Lactea ...compris'd in nine familiar letters from the author to his friendand : illustrated with upward of thirty graven and mezzotinto plates ...* 1750, 10.3931/e-rara-28672.

[7] I. Kant, *Allgemeine Naturgeschichte und Theorie des Himmels*. 1755.

[8] J. H. Lambert, *Cosmologische Briefe über die Einrichtung des Weltbaues*. 1761, 10.3931/e-rara-515.

[9] M. Hoskin, *Cosmological Letters on the Arrangement of the World-edifice, by J. H. Lambert, trans. by Stanley L. Jaki, and Lettres cosmologiques sur l'Organisation de l'Univers, by J. H. Lambert, with preface by Jacques Merleau-Ponty*, *Journal for the History of Astronomy* **9** (Jan., 1978) 134.

[10] S. Newcomb, *Popular astronomy*. 1883.

[11] G. B. Airy and S. Groombridge, *Catalogue of Circumpolar Stars, deduced from the observations of Stephen Groombridge, Esq. F. R. S. S. R. A Nap;, Fellow of the Royal Astronomical Society of London, &C, Reduced to January 1, 1810*. 1838.

[12] W. Thomson, Baron Kelvin, *LECTURE XVI*, p. 260–278. Cambridge Library Collection - Physical Sciences. Cambridge University Press, 2010. 10.1017/CBO9780511694523.020.

[13] H. Poincaré, *La Voie lactée et la théorie des gaz*, *Bulletin de la société astronomique de France* **20** (1906) 153–165.

[14] H. Poincare, *The Milky Way and the Theory of Gases*, *Popular Astronomy* **14** (Oct., 1906) 475–488.

[15] E. J. Öpik, *Absorption of Light in Space as Viewed from the Standpoint of Galactic Dynamics*, *Bulletin of the Russian Astronomical Society* (Nov., 1915) 10.

[16] J. C. Kapteyn, *First Attempt at a Theory of the Arrangement and Motion of the Sidereal System*, ApJ **55** (May, 1922) 302.

[17] J. H. Jeans, *The Motions of Stars in a Kapteyn Universe*, MNRAS **82** (Jan., 1922) 122–132.

[18] B. Lindblad, *Researches based on determinations of stellar luminosities*. Meddelanden från Astr Observatorium Upsala No 11. Almqvist & W. Boktryckeri, Uppsala, 1926.

[19] J. H. Oort, *The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems*, Bull. Astron. Inst. Netherlands **6** (Aug., 1932) 249.

[20] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, *Helvetica Physica Acta* **6** (Jan., 1933) 110–127.

[21] D. H. Rogstad and G. S. Shostak, *Gross Properties of Five Scd Galaxies as Determined from 21-CENTIMETER Observations*, ApJ **176** (Sept., 1972) 315.

[22] N. Jarosik, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, M. Halpern et al., *Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results*, ApJS **192** (Feb., 2011) 14, [1001.4744].

[23] Planck collaboration, N. Aghanim et al., *Planck 2018 results. VI. Cosmological parameters*, 1807.06209.

[24] G. Bertone and D. Hooper, *History of dark matter*, *Rev. Mod. Phys.* **90** (2018) 045002, [1605.04909].

[25] S. P. Martin, *A Supersymmetry primer*, hep-ph/9709356.

[26] M. Drewes, *The Phenomenology of Right Handed Neutrinos*, *Int. J. Mod. Phys.* **E22** (2013) 1330019, [1303.6912].

[27] W. Hu, R. Barkana and A. Gruzinov, *Cold and fuzzy dark matter*, *Phys. Rev. Lett.* **85** (2000) 1158–1161, [astro-ph/0003365].

[28] L. Hui, J. P. Ostriker, S. Tremaine and E. Witten, *Ultralight scalars as cosmological dark matter*, *Phys. Rev.* **D95** (2017) 043541, [1610.08297].

[29] D. J. H. Chung, E. W. Kolb and A. Riotto, *Superheavy dark matter*, *Phys. Rev.* **D59** (1998) 023501, [hep-ph/9802238].

[30] E. W. Kolb, D. J. H. Chung and A. Riotto, *WIMPzillas!*, *AIP Conf. Proc.* **484** (1999) 91–105, [hep-ph/9810361].

[31] B. J. Carr, K. Kohri, Y. Sendouda and J. Yokoyama, *New cosmological constraints on primordial black holes*, *Phys. Rev.* **D81** (2010) 104019, [0912.5297].

[32] B. Carr, F. Kuhnel and M. Sandstad, *Primordial Black Holes as Dark Matter*, *Phys. Rev.* **D94** (2016) 083504, [1607.06077].

[33] F. E. Schunck and E. W. Mielke, *General relativistic boson stars*, *Class. Quant. Grav.* **20** (2003) R301–R356, [0801.0307].

[34] D. Spolyar, K. Freese and P. Gondolo, *Dark matter and the first stars: a new phase of stellar evolution*, *Phys. Rev. Lett.* **100** (2008) 051101, [0705.0521].

[35] M. Yu. Khlopov, G. M. Beskin, N. E. Bochkarev, L. A. Pustylnik and S. A. Pustylnik, *Observational Physics of Mirror World*, *Sov. Astron.* **35** (1991) 21.

[36] J. Madsen, *Physics and astrophysics of strange quark matter*, *Lect. Notes Phys.* **516** (1999) 162–203, [astro-ph/9809032].

[37] F. Weber, *Strange quark matter and compact stars*, *Prog. Part. Nucl. Phys.* **54** (2005) 193–288, [astro-ph/0407155].

[38] D. E. Kaplan, G. Z. Krnjaic, K. R. Rehermann and C. M. Wells, *Atomic Dark Matter*, *JCAP* **1005** (2010) 021, [0909.0753].

[39] F.-Y. Cyr-Racine and K. Sigurdson, *Cosmology of atomic dark matter*, *Phys. Rev.* **D87** (2013) 103515, [1209.5752].

[40] Z. G. Berezhiani, A. D. Dolgov and R. N. Mohapatra, *Asymmetric inflationary reheating and the nature of mirror universe*, *Phys. Lett.* **B375** (1996) 26–36, [hep-ph/9511221].

[41] J. M. Gaskins, *A review of indirect searches for particle dark matter*, *Contemp. Phys.* **57** (2016) 496–525, [1604.00014].

[42] T. R. Slatyer, *Indirect Detection of Dark Matter*, in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics : Anticipating the Next Discoveries in Particle Physics (TASI 2016): Boulder, CO, USA, June 6-July 1, 2016*, pp. 297–353, 2018, 1710.05137, DOI.

[43] M. D. Schwartz, *TASI Lectures on Collider Physics*, in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics : Anticipating the Next Discoveries in Particle Physics (TASI 2016): Boulder, CO, USA, June 6-July 1, 2016*, pp. 65–100, 2018, 1709.04533, DOI.

[44] M. Schumann, *Direct Detection of WIMP Dark Matter: Concepts and Status*, *J. Phys.* **G46** (2019) 103003, [1903.03026].

[45] T. Marrodán Undagoitia and L. Rauch, *Dark matter direct-detection experiments*, *J. Phys.* **G43** (2016) 013001, [1509.08767].

[46] L. E. Strigari, *Galactic Searches for Dark Matter*, *Phys. Rept.* **531** (2013) 1–88, [1211.7090].

[47] M. Lisanti, *Lectures on Dark Matter Physics*, in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings (TASI 2015): Boulder, CO, USA, June 1-26, 2015*, pp. 399–446, 2017, 1603.03797, DOI.

[48] K. Garrett and G. Duda, *Dark Matter: A Primer*, *Adv. Astron.* **2011** (2011) 968283, [1006.2483].

[49] J. L. Feng, *Dark Matter Candidates from Particle Physics and Methods of Detection*, *Ann. Rev. Astron. Astrophys.* **48** (2010) 495–545, [1003.0904].

[50] G. Bertone, D. Hooper and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, *Phys. Rept.* **405** (2005) 279–390, [hep-ph/0404175].

[51] B. Benedikt, B. Goddard, D. Schulte, F. Zimmermann and M. J. Syphers, *FCC-hh Hadron Collider - Parameter Scenarios and Staging Options*, in *6th International Particle Accelerator Conference 2015 (IPAC 2015) Richmond, Virginia, USA, May 3-8, 2015*, p. TUPTY062, 2015, http://accelconf.web.cern.ch/AccelConf/IPAC2015/papers/tupty062.pdf.

[52] J. Tang et al., *Concept for a Future Super Proton-Proton Collider*, 1507.03224.

[53] CEPC-SPPC Study Group, *CEPC-SPPC Preliminary Conceptual Design Report. 1. Physics and Detector*, .

[54] Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux et al., *Gaia Data Release 2. Summary of the contents and survey properties*, *ArXiv e-prints* (Apr., 2018) , [1804.09365].

[55] Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, F. Mignard et al., *Gaia Data Release 1. Summary of the astrometric, photometric, and survey properties*, *A. & A.* **595** (Nov, 2016) A2, [1609.04172].

[56] G. C. Myeong, N. W. Evans, V. Belokurov, J. L. Sand ers and S. E. Koposov, *The Sausage Globular Clusters*, *Astrophys. J. Lett.* **863** (Aug, 2018) L28, [1805.00453].

[57] A. Helmi, C. Babusiaux, H. H. Koppelman, D. Massari, J. Veljanoski and A. G. A. Brown, *The merger that led to the formation of the Milky Way's inner stellar halo and thick disk*, Nature **563** (Nov., 2018) 85–88, [1806.06038].

[58] G. C. Myeong, N. W. Evans, V. Belokurov, J. L. Sanders and S. E. Koposov, *The Milky Way Halo in Action Space*, *Astrophys. J. Lett.* **856** (Apr., 2018) L26, [1802.03351].

[59] A. J. Deason, V. Belokurov, S. E. Koposov and L. Lancaster, *Apocenter Pile-up: Origin of the Stellar Halo Density Break*, *Astrophys. J. Lett.* **862** (Jul, 2018) L1, [1805.10288].

[60] L. Necib, M. Lisanti and V. Belokurov, *Inferred Evidence for Dark Matter Kinematic Substructure with SDSS-Gaia*, *Astrophys. J.* **874** (Mar, 2019) 3, [1807.02519].

[61] L. Necib, M. Lisanti, S. Garrison-Kimmel, A. Wetzel, R. Sanderson, P. F. Hopkins et al., *Under the Firelight: Stellar Tracers of the Local Dark Matter Velocity Distribution in the Milky Way*, 1810.12301.

[62] R. Kogler et al., *Jet Substructure at the Large Hadron Collider: Experimental Review*, *Rev. Mod. Phys.* **91** (2019) 045003, [1803.06991].

[63] A. Altheimer et al., *Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks*, *J. Phys.* **G39** (2012) 063001, [1201.0008].

[64] J. Fan, P. Jaiswal and S. C. Leung, *Jet Observables and Stops at 100 TeV Collider*, *Phys. Rev.* **D96** (2017) 036017, [1704.03014].

[65] A. Avetisyan et al., *Methods and Results for Standard Model Event Generation at $\sqrt{s}$ = 14 TeV, 33 TeV and 100 TeV Proton Colliders (A Snowmass Whitepaper)*, in *Community Summer Study 2013: Snowmass on the Mississippi (CSS2013) Minneapolis, MN, USA, July 29-August 6, 2013*, 2013, 1308.1636, http://lss.fnal.gov/archive/test-fn/0000/fermilab-fn-0965-t.pdf.

[66] B. Richter, *High Energy Colliding Beams; What Is Their Future?*, *Rev. Accel. Sci. Tech.* **7** (2014) 1–8, [1409.1196].

[67] T. G. Rizzo, *Mass Reach Scaling for Future Hadron Colliders*, *Eur. Phys. J.* **C75** (2015) 161, [1501.05583].

[68] N. Arkani-Hamed, T. Han, M. Mangano and L.-T. Wang, *Physics Opportunities of a 100 TeV Proton-Proton Collider*, *Phys. Rept.* **652** (2016) 1–49, [1511.06495].

[69] M. L. Mangano et al., *Physics at a 100 TeV pp collider: Standard Model processes*, 1607.01831.

[70] T. Golling et al., *Physics at a 100 TeV pp collider: beyond the Standard Model phenomena*, 1606.00947.

[71] R. Contino et al., *Physics at a 100 TeV pp collider: Higgs and EW symmetry breaking studies*, 1606.09408.

[72] C. Borschensky, M. Krämer, A. Kulesza, M. Mangano, S. Padhi, T. Plehn et al., *Squark and gluino production cross sections in pp collisions at $\sqrt{s}$ = 13, 14, 33 and 100 TeV*, *Eur. Phys. J.* **C74** (2014) 3174, [1407.5066].

[73] D. Stolarski, *Reach in All Hadronic Stop Decays: A Snowmass White Paper*, in *Community Summer Study 2013: Snowmass on the Mississippi (CSS2013) Minneapolis, MN, USA, July 29-August 6, 2013*, 2013, 1309.1514, http://inspirehep.net/record/1253108/files/arXiv:1309.1514.pdf.

[74] T. Cohen, T. Golling, M. Hance, A. Henrichs, K. Howe, J. Loyal et al., *SUSY Simplified Models at 14, 33, and 100 TeV Proton Colliders*, *JHEP* **04** (2014) 117, [1311.6480].

[75] S. Jung and J. D. Wells, *Gaugino physics of split supersymmetry spectra at the LHC and future proton colliders*, *Phys. Rev.* **D89** (2014) 075004, [1312.1802].

[76] T. Cohen, R. T. D'Agnolo, M. Hance, H. K. Lou and J. G. Wacker, *Boosting Stop Searches with a 100 TeV Proton Collider*, *JHEP* **11** (2014) 021, [1406.4512].

[77] A. Fowlie and M. Raidal, *Prospects for constrained supersymmetry at $\sqrt{s} = 33$ TeV and $\sqrt{s} = 100$ TeV proton-proton super-colliders*, *Eur. Phys. J.* **C74** (2014) 2948, [1402.5419].

[78] S. A. R. Ellis and B. Zheng, *Reaching for squarks and gauginos at a 100 TeV p-p collider*, *Phys. Rev.* **D92** (2015) 075034, [1506.02644].

[79] S. Antusch and C. Sluka, *Testable SUSY spectra from GUTs at a 100 TeV pp collider*, *Int. J. Mod. Phys.* **A31** (2016) 1644011, [1604.00212].

[80] G. Grilli di Cortona, E. Hardy and A. J. Powell, *Dirac vs Majorana gauginos at a 100 TeV collider*, *JHEP* **08** (2016) 014, [1606.07090].

[81] M. Low and L.-T. Wang, *Neutralino dark matter at 14 TeV and 100 TeV*, *JHEP* **08** (2014) 161, [1404.0682].

[82] M. Cirelli, F. Sala and M. Taoso, *Wino-like Minimal Dark Matter and future colliders*, *JHEP* **10** (2014) 033, [1407.7058].

[83] A. Hook and A. Katz, *Unbroken $SU(2)$ at a 100 TeV collider*, *JHEP* **09** (2014) 175, [1407.2607].

[84] S. Gori, S. Jung, L.-T. Wang and J. D. Wells, *Prospects for Electroweakino Discovery at a 100 TeV Hadron Collider*, *JHEP* **12** (2014) 108, [1410.6287].

[85] J. Bramante, P. J. Fox, A. Martin, B. Ostdiek, T. Plehn, T. Schell et al., *Relic neutralino surface at a 100 TeV collider*, *Phys. Rev.* **D91** (2015) 054015, [1412.4789].

[86] D. S. M. Alves, J. Galloway, J. T. Ruderman and J. R. Walsh, *Running Electroweak Couplings as a Probe of New Physics*, *JHEP* **02** (2015) 007, [1410.6810].

[87] J. Hajer, Y.-Y. Li, T. Liu and J. F. H. Shiu, *Heavy Higgs Bosons at 14 TeV and 100 TeV*, *JHEP* **11** (2015) 124, [1504.07617].

[88] J. Bramante, N. Desai, P. Fox, A. Martin, B. Ostdiek and T. Plehn, *Towards the Final Word on Neutralino Dark Matter*, *Phys. Rev.* **D93** (2016) 063525, [1510.03460].

[89] A. Ismail, E. Izaguirre and B. Shuve, *Illuminating New Electroweak States at Hadron Colliders*, *Phys. Rev.* **D94** (2016) 015001, [1605.00658].

[90] R. Mahbubani, P. Schwaller and J. Zurita, *Closing the window for compressed Dark Sectors with disappearing charged tracks*, 1703.05327.

[91] H. Fukuda, N. Nagata, H. Otono and S. Shirai, *Higgsino Dark Matter or Not: Role of Disappearing Track Searches at the LHC and Future Colliders*, 1703.09675.

[92] A. Kobakhidze, M. Talia and L. Wu, *Probing the MSSM explanation of the muon g-2 anomaly in dark matter experiments and at a 100 TeV pp collider*, *Phys. Rev.* **D95** (2017) 055023, [1608.03641].

[93] R. Barbieri and G. F. Giudice, *Upper Bounds on Supersymmetric Particle Masses*, *Nucl. Phys.* **B306** (1988) 63–76.

[94] S. Dimopoulos and G. F. Giudice, *Naturalness constraints in supersymmetric theories with nonuniversal soft terms*, *Phys. Lett.* **B357** (1995) 573–578, [hep-ph/9507282].

[95] A. Pomarol and D. Tommasini, *Horizontal symmetries for the supersymmetric flavor problem*, *Nucl. Phys.* **B466** (1996) 3–24, [hep-ph/9507462].

[96] A. G. Cohen, D. B. Kaplan and A. E. Nelson, *The More minimal supersymmetric standard model*, *Phys. Lett.* **B388** (1996) 588–598, [hep-ph/9607394].

[97] R. Kitano and Y. Nomura, *Supersymmetry, naturalness, and signatures at the LHC*, *Phys. Rev.* **D73** (2006) 095004, [hep-ph/0602096].

[98] M. Perelstein and C. Spethmann, *A Collider signature of the supersymmetric golden region*, *JHEP* **04** (2007) 070, [hep-ph/0702038].

[99] M. Papucci, J. T. Ruderman and A. Weiler, *Natural SUSY Endures*, *JHEP* **09** (2012) 035, [1110.6926].

[100] C. Brust, A. Katz, S. Lawrence and R. Sundrum, *SUSY, the Third Generation and the LHC*, *JHEP* **03** (2012) 103, [1110.6670].

[101] M. L. Graesser and J. Shelton, *Hunting Mixed Top Squark Decays*, *Phys. Rev. Lett.* **111** (2013) 121802, [1212.4495].

[102] A. J. Larkoski, F. Maltoni and M. Selvaggi, *Tracking down hyper-boosted top quarks*, *JHEP* **06** (2015) 032, [1503.03347].

[103] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, [1405.0301].

[104] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [0710.3820].

[105] DELPHES 3 collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens et al., *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057, [1307.6346].

[106] M. L. Mangano, M. Moretti, F. Piccinini and M. Treccani, *Matching matrix elements and shower evolution for top-quark production in hadronic collisions*, *JHEP* **01** (2007) 013, [hep-ph/0611129].

[107] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C72** (2012) 1896, [1111.6097].

[108] M. Cacciari, G. P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **04** (2008) 063, [0802.1189].

[109] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001, [hep-ph/9707323].

[110] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, in *Monte Carlo generators for HERA physics. Proceedings,*

Workshop, Hamburg, Germany, 1998-1999, pp. 270–279, 1998, hep-ph/9907280, https://inspirehep.net/record/484872/files/arXiv:hep-ph_9907280.pdf.

[111] A. J. Larkoski and J. Thaler, *Aspects of jets at 100 TeV*, *Phys. Rev.* **D90** (2014) 034010, [1406.7011].

[112] S. Schatzel and M. Spannowsky, *Tagging highly boosted top quarks*, *Phys. Rev.* **D89** (2014) 014007, [1308.0540].

[113] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015, [1011.2268].

[114] J. Thaler and L.-T. Wang, *Strategies to Identify Boosted Tops*, *JHEP* **07** (2008) 092, [0806.0023].

[115] K. Rehermann and B. Tweedie, *Efficient Identification of Boosted Semileptonic Top Quarks at the LHC*, *JHEP* **03** (2011) 059, [1007.2221].

[116] S. Dawson, A. Ismail and I. Low, *Redux on "When is the top quark a parton?"*, *Phys. Rev.* **D90** (2014) 014005, [1405.6211].

[117] A. L. Read, *Presentation of search results: The CL(s) technique*, *J. Phys.* **G28** (2002) 2693–2704.

[118] D. Krohn, L. Randall and L.-T. Wang, *On the Feasibility and Utility of ISR Tagging*, 1101.0810.

[119] M. Spannowsky and M. Stoll, *Tracking new physics at the lhc and beyond*, *Phys. Rev. D* **92** (Sep, 2015) 054033.

[120] S. M. Etesami and M. Mohammadi Najafabadi, *Double higgs boson production with a jet substructure analysis to probe extra dimensions*, *Phys. Rev. D* **92** (Oct, 2015) 073013.

[121] Z. Kang, P. Ko and J. Li, *New avenues to heavy right-handed neutrinos with pair production at hadronic colliders*, *Phys. Rev. D* **93** (Apr, 2016) 075037.

[122] M. Mitra, R. Ruiz, D. J. Scott and M. Spannowsky, *Neutrino jets from high-mass W_R gauge bosons in tev-scale left-right symmetric models*, *Phys. Rev. D* **94** (Nov, 2016) 095016.

[123] P. Agrawal, J. Fan, M. Reece and W. Xue, *Deciphering the MSSM Higgs Mass at Future Hadron Colliders*, 1702.05484.

[124] S. Bressler, T. Flacke, Y. Kats, S. J. Lee and G. Perez, *Hadronic Calorimeter Shower Size: Challenges and Opportunities for Jet Substructure in the Superboosted Regime*, *Phys. Lett.* **B756** (2016) 137–141, [1506.02656].

[125] S. W. Randall, M. Markevitch, D. Clowe, A. H. Gonzalez and M. Bradac, *Constraints on the Self-Interaction Cross-Section of Dark Matter from Numerical Simulations of the Merging Galaxy Cluster 1E 0657-56*, *Astrophys. J.* **679** (2008) 1173–1180, [0704.0261].

[126] M. Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, L. David, W. Forman et al., *Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56*, *Astrophys. J.* **606** (2004) 819–824, [astro-ph/0309303].

[127] A. H. G. Peter, M. Rocha, J. S. Bullock and M. Kaplinghat, *Cosmological Simulations with Self-Interacting Dark Matter II: Halo Shapes vs. Observations*, *Mon. Not. Roy. Astron. Soc.* **430** (2013) 105, [1208.3026].

[128] D. A. Buote, T. E. Jeltema, C. R. Canizares and G. P. Garmire, *Chandra evidence for a flattened, triaxial dark matter halo in the elliptical galaxy ngc 720*, *Astrophys. J.* **577** (2002) 183–196, [astro-ph/0205469].

[129] M. R. Buckley and P. J. Fox, *Dark Matter Self-Interactions and Light Force Carriers*, *Phys. Rev.* **D81** (2010) 083522, [0911.3898].

[130] B. D. Wandelt, R. Dave, G. R. Farrar, P. C. McGuire, D. N. Spergel and P. J. Steinhardt, *Selfinteracting dark matter*, in *Sources and detection of dark matter and dark energy in the universe. Proceedings, 4th International Symposium, DM 2000, Marina del Rey, USA, February 23-25, 2000*, pp. 263–274, 2000, astro-ph/0006344, http://www.slac.stanford.edu/spires/find/books/www?cl=QB461:I57:2000.

[131] D. N. Spergel and P. J. Steinhardt, *Observational evidence for selfinteracting cold dark matter*, *Phys. Rev. Lett.* **84** (2000) 3760–3763, [astro-ph/9909386].

[132] J. F. Navarro, C. S. Frenk and S. D. M. White, *The Structure of cold dark matter halos*, *Astrophys. J.* **462** (1996) 563–575, [astro-ph/9508025].

[133] L. Ackerman, M. R. Buckley, S. M. Carroll and M. Kamionkowski, *Dark Matter and Dark Radiation*, *Phys. Rev.* **D79** (2009) 023519, [0810.5126].

[134] J. L. Feng, M. Kaplinghat, H. Tu and H.-B. Yu, *Hidden Charged Dark Matter*, *JCAP* **0907** (2009) 004, [0905.3039].

[135] J. Fan, A. Katz, L. Randall and M. Reece, *Dark-Disk Universe*, *Phys. Rev. Lett.* **110** (2013) 211302, [1303.3271].

[136] J. Fan, A. Katz, L. Randall and M. Reece, *Double-Disk Dark Matter*, *Phys. Dark Univ.* **2** (2013) 139–156, [1303.1521].

[137] M. McCullough and L. Randall, *Exothermic Double-Disk Dark Matter*, *JCAP* **1310** (2013) 058, [1307.4095].

[138] J. Fan, A. Katz and J. Shelton, *Direct and indirect detection of dissipative dark matter*, *JCAP* **1406** (2014) 059, [1312.1336].

[139] L. Randall and M. Reece, *Dark Matter as a Trigger for Periodic Comet Impacts*, *Phys. Rev. Lett.* **112** (2014) 161301, [1403.0576].

[140] W. Fischler, D. Lorshbough and W. Tangarife, *Supersymmetric Partially Interacting Dark Matter*, *Phys. Rev.* **D91** (2015) 025010, [1405.7708].

[141] R. Foot and S. Vagnozzi, *Dissipative hidden sector dark matter*, *Phys. Rev.* **D91** (2015) 023512, [1409.7174].

[142] L. Randall and J. Scholtz, *Dissipative Dark Matter and the Andromeda Plane of Satellites*, *JCAP* **1509** (2015) 057, [1412.1839].

[143] M. Reece and T. Roxlo, *Nonthermal production of dark radiation and dark matter*, *JHEP* **09** (2016) 096, [1511.06768].

[144] R. Foot and S. Vagnozzi, *Solving the small-scale structure puzzles with dissipative dark matter*, *JCAP* **1607** (2016) 013, [1602.02467].

[145] N. J. Shaviv, *The Paleoclimatic evidence for Strongly Interacting Dark Matter Present in the Galactic Disk*, 1606.02851.

[146] E. D. Kramer and L. Randall, *Updated Kinematic Constraints on a Dark Disk*, *Astrophys. J.* **824** (2016) 116, [1604.01407].

[147] P. Agrawal, F.-Y. Cyr-Racine, L. Randall and J. Scholtz, *Dark Catalysis*, *JCAP* **1708** (2017) 021, [1702.05482].

[148] P. Agrawal and L. Randall, *Point Sources from Dissipative Dark Matter*, *JCAP* **1712** (2017) 019, [1706.04195].

[149] M. R. Buckley and A. DiFranzo, *Collapsed Dark Matter Structures*, *Phys. Rev. Lett.* **120** (2018) 051102, [1707.03829].

[150] G. D'Amico, P. Panci, A. Lupi, S. Bovino and J. Silk, *Massive Black Holes from Dissipative Dark Matter*, *Mon. Not. Roy. Astron. Soc.* **473** (2018) 328–335, [1707.03419].

[151] A. Caputo, J. Zavala and D. Blas, *Binary pulsars as probes of a Galactic dark matter disk*, *Phys. Dark Univ.* **19** (2018) 1–11, [1709.03991].

[152] K. Vattis and S. M. Koushiappas, *Self-interacting dark matter constraints in a thick dark disk scenario*, 1801.06556.

[153] N. J. Outmazgine, O. Slone, W. Tangarife, L. Ubaldi and T. Volansky, *Accretion of Dissipative Dark Matter onto Active Galactic Nuclei*, 1807.04750.

[154] S. Alexander and L. Smolin, *The Equivalence Principle and the Emergence of Flat Rotation Curves*, 1804.09573.

[155] A. Ghalsasi and M. McQuinn, *Exploring the astrophysics of dark atoms*, 1712.04779.

[156] E. Rosenberg and J. Fan, *Cooling in a Dissipative Dark Sector*, *Phys. Rev.* **D96** (2017) 123001, [1705.10341].

[157] J. Binney and S. Tremaine, *Galactic Dynamics: Second Edition*. Princeton University Press, 2008.

[158] A. Budenbender, G. van de Ven and L. L. Watkins, *The tilt of the velocity ellipsoid in the Milky Way disc*, *Mon. Not. Roy. Astron. Soc.* **452** (2015) 956–968, [1407.4808].

[159] J. Holmberg and C. Flynn, *The local density of matter mapped by hipparcos*, *Mon. Not. Roy. Astron. Soc.* **313** (2000) 209–216, [astro-ph/9812404].

[160] J. Buch, S. C. J. Leung and J. Fan, *Using Gaia DR2 to Constrain Local Dark Matter Density and Thin Dark Disk*, *JCAP* **1904** (2019) 026, [1808.05603].

[161] Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, A. G. A. Brown, A. Vallenari, C. Babusiaux et al., *The Gaia mission*, *A&A* **595** (Nov., 2016) A1, [1609.04153].

[162] L. Lindegren, J. Hernandez, A. Bombrun, S. Klioner, U. Bastian, M. Ramos-Lerate et al., *Gaia Data Release 2: The astrometric solution*, *ArXiv e-prints* (Apr., 2018) , [1804.09366].

[163] D. W. Evans, M. Riello, F. De Angeli, J. M. Carrasco, P. Montegriffo, C. Fabricius et al., *Gaia Data Release 2: Photometric content and validation*, *ArXiv e-prints* (Apr., 2018) , [1804.09368].

[164] M. Riello, F. De Angeli, D. W. Evans, G. Busso, N. C. Hambly, M. Davidson et al., *Gaia Data Release 2: processing of the photometric data*, *ArXiv e-prints* (Apr., 2018) , [1804.09367].

[165] P. Sartoretti, D. Katz, M. Cropper, P. Panuzzo, G. M. Seabroke, Y. Viala et al., *Gaia Data Release 2: Processing the spectroscopic data*, *ArXiv e-prints* (Apr., 2018) , [1804.09371].

[166] M. Cropper, D. Katz, P. Sartoretti, T. Prusti, J. H. J. de Bruijne, F. Chassat et al., *Gaia Radial Velocity Spectrometer*, *ArXiv e-prints* (Apr., 2018) , [1804.09369].

[167] D. Katz, P. Sartoretti, M. Cropper, P. Panuzzo, G. M. Seabroke, Y. Viala et al., *Gaia Data Release 2: Properties and validation of the radial velocities*, *ArXiv e-prints* (Apr., 2018) , [1804.09372].

[168] X. Luri, A. G. A. Brown, L. M. Sarro, F. Arenou, C. A. L. Bailer-Jones, A. Castro-Ginard et al., *Gaia Data Release 2: using Gaia parallaxes*, *ArXiv e-prints* (Apr., 2018) , [1804.09376].

[169] J. Bovy and H.-W. Rix, *A Direct Dynamical Measurement of the Milky Way's Disk Surface Density Profile, Disk Scale Length, and Dark Matter Profile at 4 kpc $\lesssim R \lesssim$ 9 kpc*, *Astrophys. J.* **779** (2013) 115, [1309.0809].

[170] K. Kuijken and G. Gilmore, *The mass distribution in the galactic disc. I - A technique to determine the integral surface mass density of the disc near the sun.*, *Mon. Not. Roy. Astron. Soc.* **239** (Aug., 1989) 571–603.

[171] K. Kuijken and G. Gilmore, *The Mass Distribution in the Galactic Disc - II - Determination of the Surface Mass Density of the Galactic Disc Near the Sun*, *Mon. Not. Roy. Astron. Soc.* **239** (Aug., 1989) 605–649.

[172] K. Kuijken and G. Gilmore, *The Mass Distribution in the Galactic Disc - Part III - the Local Volume Mass Density*, *Mon. Not. Roy. Astron. Soc.* **239** (Aug., 1989) 651–664.

[173] B. Fuchs and R. Wielen, *Kinematical constraints on the dynamically determined local mass density of the Galaxy*, in *Back to the Galaxy* (S. S. Holt and F. Verter, eds.), vol. 278 of *American Institute of Physics Conference Series*, pp. 580–583, 1993, DOI.

[174] C. Flynn and B. Fuchs, *Density of dark matter in the Galactic disk*, *Mon. Not. Roy. Astron. Soc.* **270** (Oct., 1994) .

[175] J. Holmberg and C. Flynn, *The local surface density of disc matter mapped by Hipparcos*, *Mon. Not. Roy. Astron. Soc.* **352** (Aug., 2004) 440–446, [astro-ph/0405155].

[176] F. van Leeuwen and E. Fantino, *A New reduction of the raw Hipparcos data*, *Astron. Astrophys.* **439** (2005) 791–803, [astro-ph/0505432].

[177] S. Garbari, C. Liu, J. I. Read and G. Lake, *A new determination of the local dark matter density from the kinematics of K dwarfs*, *Mon. Not. Roy. Astron. Soc.* **425** (Sept., 2012) 1445–1458, [1206.0015].

[178] J. Bovy and S. Tremaine, *On the local dark matter density*, *Astrophys. J.* **756** (2012) 89, [1205.4033].

[179] H. Silverwood, S. Sivertsson, P. Steger, J. I. Read and G. Bertone, *A non-parametric method for measuring the local dark matter density*, *Mon. Not. Roy. Astron. Soc.* **459** (July, 2016) 4191–4208, [1507.08581].

[180] S. Sivertsson, H. Silverwood, J. I. Read, G. Bertone and P. Steger, *The Local Dark Matter Density from SDSS-SEGUE G-dwarfs*, *Mon. Not. Roy. Astron. Soc.* (Apr., 2018) , [1708.07836].

[181] Y.-S. Ting, H.-W. Rix, J. Bovy and G. van de Ven, *Constraining the Galactic potential via action-based distribution functions for mono-abundance stellar populations*, *"Mon. Not. Roy. Astron. Soc."* **434** (Sept., 2013) 652–660, [1212.0006].

[182] J. Bovy, H.-W. Rix, C. Liu, D. W. Hogg, T. C. Beers and Y. S. Lee, *The Spatial Structure of Mono-abundance Sub-populations of the Milky Way Disk*, *Astrophys. J.* **753** (July, 2012) 148, [1111.1724].

[183] P. Salucci, F. Nesti, G. Gentile and C. F. Martins, *The dark matter density at the Sun's location*, *Astron. Astrophys.* **523** (2010) A83, [1003.3101].

[184] M. Pato, F. Iocco and G. Bertone, *Dynamical constraints on the dark matter distribution in the Milky Way*, *JCAP* **1512** (2015) 001, [1504.06324].

[185] L. E. Strigari, *Galactic searches for dark matter*, *Phys. Rep.* **531** (Oct., 2013) 1–88, [1211.7090].

[186] M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter et al., *The Two Micron All Sky Survey (2MASS)*, *AJ* **131** (Feb., 2006) 1163–1183.

[187] J. Bovy, *Stellar inventory of the solar neighbourhood using Gaia DR1*, *"Mon. Not. Roy. Astron. Soc."* **470** (Sept., 2017) 1360–1387, [1704.05063].

[188] J. C. Zinn, M. H. Pinsonneault, D. Huber and D. Stello, *Confirmation of the zero-point offset in Gaia Data Release 2 parallaxes using asteroseismology and APOGEE spectroscopy in the Kepler field*, *ArXiv e-prints* (May, 2018) , [1805.02650].

[189] N. Banik, L. M. Widrow and S. Dodelson, *Galactoseismology and the local density of dark matter*, *"Mon. Not. Roy. Astron. Soc."* **464** (Feb., 2017) 3775–3783, [1608.03338].

[190] J. N. Bahcall, C. Flynn and A. Gould, *Local dark matter from a carefully selected sample*, *Astrophys. J.* **389** (Apr., 1992) 234–250.

[191] M. J. Pecaut and E. E. Mamajek, *Intrinsic Colors, Temperatures, and Bolometric Corrections of Pre-main-sequence Stars*, *ApJS* **208** (Sept., 2013) 9, [1307.2657].

[192] R. Schönrich, J. Binney and W. Dehnen, *Local kinematics and the local standard of rest*, *Mon. Not. Roy. Astron. Soc.* **403** (Apr., 2010) 1829–1833, [0912.3693].

[193] S. Garbari, J. I. Read and G. Lake, *Limits on the local dark matter density*, *"Mon. Not. Roy. Astron. Soc."* **416** (Sept., 2011) 2318–2340, [1105.6339].

[194] J. Bovy, *Galactic rotation in Gaia DR1*, *Mon. Not. Roy. Astron. Soc.* **468** (June, 2017) L63–L67, [1610.07610].

[195] J. N. Bahcall, *Self-consistent determinations of the total amount of matter near the sun*, *Astrophys. J.* **276** (Jan., 1984) 169–181.

[196] J. N. Bahcall, *The distribution of stars perpendicular to galactic disk*, *Astrophys. J.* **276** (Jan., 1984) 156–168.

[197] J. N. Bahcall, *K giants and the total amount of matter near the sun*, *Astrophys. J.* **287** (Dec., 1984) 926–944.

[198] L. Spitzer, Jr., *The Dynamics of the Interstellar Medium. III. Galactic Distribution.*, *Astrophys. J.* **95** (May, 1942) 329.

[199] J. Bovy, H.-W. Rix, G. M. Green, E. F. Schlafly and D. P. Finkbeiner, *On Galactic Density Modeling in the Presence of Dust Extinction*, *ApJ* **818** (Feb., 2016) 130, [1509.06751].

[200] A. Widmark, *Measuring the local matter density using Gaia DR2*, 1811.07911.

[201] J. Goodman and J. Weare, *Ensemble samplers with affine invariance*, *Communications in Applied Mathematics and Computational Science, Vol. 5, No. 1, p. 65-80, 2010* **5** (2010) 65–80.

[202] D. Foreman-Mackey, D. W. Hogg, D. Lang and J. Goodman, *emcee: The MCMC Hammer*, *Publ. Astron. Soc. Pac.* **125** (2013) 306–312, [1202.3665].

[203] G. O. Roberts, A. Gelman and W. R. Gilks, *Weak convergence and optimal scaling of random walk metropolis algorithms*, *The Annals of Applied Probability* **7** (04, 1997) .

[204] B. Yanny, C. Rockosi, H. J. Newberg, G. R. Knapp, J. K. Adelman-McCarthy, B. Alcorn et al., *SEGUE: A Spectroscopic Survey of 240,000 Stars with g = 14-20*, *Astronomical. J.* **137** (May, 2009) 4377–4399, [0902.1781].

[205] L. M. Widrow, S. Gardner, B. Yanny, S. Dodelson and H.-Y. Chen, *Galactoseismology: Discovery of Vertical Waves in the Galactic Disk*, *Astrophys. J. Lett.* **750** (May, 2012) L41, [1203.6861].

[206] M. Bennett and J. Bovy, *Vertical waves in the solar neighbourhood in Gaia DR2*, *Mon. Not. Roy. Astron. Soc.* **482** (Jan, 2019) 1417–1425, [1809.03507].

[207] F. A. Gómez, I. Minchev, B. W. O'Shea, T. C. Beers, J. S. Bullock and C. W. Purcell, *Vertical density waves in the Milky Way disc induced by the Sagittarius dwarf galaxy*, *"Mon. Not. Roy. Astron. Soc."* **429** (Feb., 2013) 159–164, [1207.3083].

[208] J. L. Carlin, J. DeLaunay, H. J. Newberg, L. Deng, D. Gole, K. Grabowski et al., *Substructure in Bulk Velocities of Milky Way Disk Stars*, *Astrophys. J. Lett.* **777** (Nov., 2013) L5, [1309.6314].

[209] L. M. Widrow, J. Barber, M. H. Chequers and E. Cheng, *Bending and breathing modes of the Galactic disc*, *"Mon. Not. Roy. Astron. Soc."* **440** (May, 2014) 1971–1981, [1404.4069].

[210] T. Antoja, A. Helmi, M. Romero-Gomez, D. Katz, C. Babusiaux, R. Drimmel et al., *Wrinkles in the Gaia data unveil a dynamically young and perturbed Milky Way disk*, *ArXiv e-prints* (Apr., 2018) , [1804.10196].

[211] G. C. Myeong, N. W. Evans, V. Belokurov, J. L. Sanders and S. E. Koposov, *The Sausage Globular Clusters*, 1805.00453.

[212] LUX-ZEPLIN collaboration, D. S. Akerib et al., *Projected WIMP Sensitivity of the LUX-ZEPLIN (LZ) Dark Matter Experiment*, 1802.06039.

[213] XENON collaboration, G. Plante. https://conferences.pa.ucla.edu/dm16/talks/plante.pdf.

[214] PANDAX collaboration, H. Zhang et al., *Dark matter direct search sensitivity of the PandaX-4T experiment*, *Sci. China Phys. Mech. Astron.* **62** (2019) 31011, [1806.02229].

[215] SUPERCDMS collaboration, R. Agnese et al., *Projected Sensitivity of the SuperCDMS SNOLAB experiment*, *Phys. Rev.* **D95** (2017) 082002, [1610.00006].

[216] DAMIC collaboration, M. Settimo, *The DAMIC experiment at SNOLAB*, in *53rd Rencontres de Moriond on QCD and High Energy Interactions (Moriond QCD 2018) La Thuile, Italy, March 17-24, 2018*, 2018, 1805.10001.

[217] DARWIN collaboration, J. Aalbers et al., *DARWIN: towards the ultimate dark matter detector*, *JCAP* **1611** (2016) 017, [1606.07001].

[218] C. E. Aalseth et al., *DarkSide-20k: A 20 tonne two-phase LAr TPC for direct dark matter detection at LNGS*, *Eur. Phys. J. Plus* **133** (2018) 131, [1707.08145].

[219] A. K. Drukier, K. Freese and D. N. Spergel, *Detecting cold dark-matter candidates*, *Phys. Rev.* **D33** (Jun, 1986) 3495–3508.

[220] R. A. Ibata, G. Gilmore and M. J. Irwin, *A dwarf satellite galaxy in Sagittarius*, *Nature* **370** (July, 1994) 194–196.

[221] K. V. Johnston, D. N. Spergel and L. Hernquist, *The Disruption of the Sagittarius Dwarf Galaxy*, *Astrophys. J. Lett.* **451** (Oct, 1995) 598, [astro-ph/9502005].

[222] R. Ibata, G. F. Lewis, M. Irwin, E. Totten and T. Quinn, *Great Circle Tidal Streams: Evidence for a Nearly Spherical Massive Dark Halo around the Milky Way*, *Astrophys. J.* **551** (Apr, 2001) 294–311, [astro-ph/0004011].

[223] V. Belokurov et al., *The Field of Streams: Sagittarius and its Siblings*, *Astrophys. J.* **642** (2006) L137–L140, [astro-ph/0605025].

[224] M. Vogelsberger, A. Helmi, V. Springel, S. D. M. White, J. Wang, C. S. Frenk et al., *Phase-space structure in the local dark matter distribution and its signature in direct detection experiments*, *MNRAS* **395** (2009) 797–811, [0812.0362].

[225] M. Kuhlen, N. Weiner, J. Diemand, P. Madau, B. Moore, D. Potter et al., *Dark Matter Direct Detection with Non-Maxwellian Velocity Structure*, *JCAP* **1002** (2010) 030, [0912.2358].

[226] M. Kuhlen, M. Lisanti and D. N. Spergel, *Direct Detection of Dark Matter Debris Flows*, *Phys. Rev.* **D86** (2012) 063505, [1202.0007].

[227] M. Kuhlen, M. Vogelsberger and R. Angulo, *Numerical Simulations of the Dark Universe: State of the Art and the Next Decade*, *Phys. Dark Univ.* **1** (2012) 50–93, [1209.5745].

[228] Y.-Y. Mao, L. E. Strigari and R. H. Wechsler, *Connecting Direct Dark Matter Detection Experiments to Cosmologically Motivated Halo Models*, *Phys. Rev.* **D89** (2014) 063513, [1304.6401].

[229] N. Bozorgnia, F. Calore, M. Schaller, M. Lovell, G. Bertone, C. S. Frenk et al., *Simulated Milky Way analogues: implications for dark matter direct searches*, *JCAP* **1605** (2016) 024, [1601.04707].

[230] M. Lisanti, L. E. Strigari, J. G. Wacker and R. H. Wechsler, *The Dark Matter at the End of the Galaxy*, *Phys. Rev.* **D83** (2011) 023519, [1010.4300].

[231] R. Catena and P. Ullio, *The local dark matter phase-space density and impact on WIMP direct detection*, *JCAP* **1205** (2012) 005, [1111.3556].

[232] P. Bhattacharjee, S. Chaudhury, S. Kundu and S. Majumdar, *Sizing-up the WIMPs of Milky Way : Deriving the velocity distribution of Galactic Dark Matter particles from the rotation curve data*, *Phys. Rev.* **D87** (2013) 083525, [1210.2328].

[233] N. Bozorgnia, R. Catena and T. Schwetz, *Anisotropic dark matter distribution functions and impact on WIMP direct detection*, *JCAP* **1312** (2013) 050, [1310.0468].

[234] M. Fornasa and A. M. Green, *Self-consistent phase-space distribution function for the anisotropic dark matter halo of the Milky Way*, *Phys. Rev.* **D89** (2014) 063531, [1311.5477].

[235] S. Mandal, S. Majumdar, V. Rentala and R. Basu Thakur, *Observationally inferred dark matter phase-space distribution and direct detection experiments*, *Phys. Rev.* **D100** (2019) 023002, [1806.06872].

[236] M. Petac and P. Ullio, *Two-integral distribution functions in axisymmetric galaxies: implications for dark matter searches*, *Phys. Rev.* **D99** (2019) 043003, [1812.01531].

[237] T. D. P. Edwards and C. Weniger, *A Fresh Approach to Forecasting in Astroparticle Physics and Dark Matter Searches*, *JCAP* **1802** (2018) 021, [1704.05458].

[238] T. D. P. Edwards and C. Weniger, *swordfish: Efficient Forecasting of New Physics Searches without Monte Carlo*, 1712.05401.

[239] S. D. McDermott, H.-B. Yu and K. M. Zurek, *The Dark Matter Inverse Problem: Extracting Particle Physics from Scattering Events*, *Phys. Rev.* **D85** (2012) 123507, [1110.4281].

[240] R. Catena, *Prospects for direct detection of dark matter in an effective theory approach*, *JCAP* **1407** (2014) 055, [1406.0524].

[241] V. Gluscevic and A. H. G. Peter, *Understanding WIMP-baryon interactions with direct detection: A Roadmap*, *JCAP* **1409** (2014) 040, [1406.7008].

[242] V. Gluscevic, M. I. Gresham, S. D. McDermott, A. H. G. Peter and K. M. Zurek, *Identifying the Theory of Dark Matter with Direct Detection*, *JCAP* **1512** (2015) 057, [1506.04454].

[243] G. B. Gelmini, V. Takhistov and S. J. Witte, *Casting a Wide Signal Net with Future Direct Dark Matter Detection Experiments*, *JCAP* **1807** (2018) 009, [1804.01638].

[244] F. Kahlhoefer, S. Kulkarni and S. Wild, *Exploring light mediators with low-threshold direct detection experiments*, *JCAP* **1711** (2017) 016, [1707.08571].

[245] Y. Wu, K. Freese, C. Kelso, P. Stengel and M. Valluri, *Uncertainties in Direct Dark Matter Detection in Light of Gaia*, 1904.04781.

[246] G. Besla, A. Peter and N. Garavito-Camargo, *The highest-speed local dark matter particles come from the Large Magellanic Cloud*, 1909.04140.

[247] A. Hryczuk, E. Karukes, L. Roszkowski and M. Talia, *Impact of uncertainties in the halo velocity profile on direct detection of sub-GeV dark matter*, 2001.09156.

[248] J. Buch, J. Fan and J. S. C. Leung, *Implications of the Gaia Sausage for Dark Matter Nuclear Interactions*, *Phys. Rev. D* **101** (2020) 063026, [1910.06356].

[249] XENON collaboration, E. Aprile et al., *First Dark Matter Search Results from the XENON1T Experiment*, *Phys. Rev. Lett.* **119** (2017) 181301, [1705.06655].

[250] C. Savage, K. Freese and P. Gondolo, *Annual Modulation of Dark Matter in the Presence of Streams*, *Phys. Rev.* **D74** (2006) 043531, [astro-ph/0607121].

[251] S. K. Lee, M. Lisanti and B. R. Safdi, *Dark-Matter Harmonics Beyond Annual Modulation*, *JCAP* **1311** (2013) 033, [1307.5323].

[252] J. Herzog-Arbeitman, M. Lisanti, P. Madau and L. Necib, *Empirical Determination of Dark Matter Velocities using Metal-Poor Stars*, *Phys. Rev. Lett.* **120** (2018) 041102, [1704.04499].

[253] M. Lisanti and D. N. Spergel, *Dark Matter Debris Flows in the Milky Way*, *Phys. Dark Univ.* **1** (2012) 155–161, [1105.4166].

[254] G. C. Myeong, N. W. Evans, V. Belokurov, J. L. Sanders and S. E. Koposov, *Discovery of new retrograde substructures: the shards of Omega Centauri?*, *MNRAS* **478** (06, 2018) 5449–5459.

[255] R. H. Barbá, D. Minniti, D. Geisler, J. Alonso-García, M. Hempel, A. Monachesi et al., *A Sequoia in the Garden: FSR 1758—Dwarf Galaxy or Giant Globular Cluster?*, *Astrophys. J. Lett.* **870** (Jan, 2019) L24, [1812.04999].

[256] G. C. Myeong, E. Vasiliev, G. Iorio, N. W. Evans and V. Belokurov, *Evidence for two early accretion events that built the Milky Way stellar halo*, *MNRAS* **488** (Sep, 2019) 1235–1247, [1904.03185].

[257] E. N. Kirby, J. G. Cohen, P. Guhathakurta, L. Cheng, J. S. Bullock and A. Gallazzi, *The Universal Stellar Mass-Stellar Metallicity Relation for Dwarf Galaxies*, *Astrophys. J.* **779** (Dec., 2013) 102, [1310.0814].

[258] S. Garrison-Kimmel, J. S. Bullock, M. Boylan-Kolchin and E. Bardwell, *Organized Chaos: Scatter in the relation between stellar mass and halo mass in small galaxies*, *MNRAS* **464** (2017) 3108–3120, [1603.04855].

[259] N. Bozorgnia, A. Fattahi, D. G. Cerdeno, C. S. Frenk, F. A. Gomez, R. J. J. Grand et al., *On the correlation between the local dark matter and stellar velocities*, *JCAP* **1906** (2019) 045, [1811.11763].

[260] N. Bozorgnia, A. Fattahi, C. S. Frenk, A. Cheek, D. G. Cerdeno, F. A. Gómez et al., *The dark matter component of the Gaia radially anisotropic substructure*, 1910.07536.

[261] N. W. Evans, C. A. J. O'Hare and C. McCabe, *Refinement of the standard halo model for dark matter searches in light of the Gaia Sausage*, *Phys. Rev.* **D99** (2019) 023012, [1810.11468].

[262] C. Wegg, O. Gerhard and M. Bieth, *The gravitational force field of the Galaxy measured from the kinematics of RR Lyrae in Gaia*, *MNRAS* **485** (May, 2019) 3296–3316, [1806.09635].

[263] L. Necib, B. Ostdiek, M. Lisanti, T. Cohen, M. Freytsis and S. Garrison-Kimmel, *Chasing Accreted Structures within Gaia DR2 using Deep Learning*, 1907.07681.

[264] L. Necib, B. Ostdiek, M. Lisanti, T. Cohen, M. Freytsis, S. Garrison-Kimmel et al., *Evidence for a Vast Prograde Stellar Stream in the Solar Vicinity*, 1907.07190.

[265] M. R. Buckley, G. Mohlabeng and C. W. Murphy, *Direct Detection Anomalies in light of Gaia Data*, 1905.05189.

[266] C. A. J. O'Hare, C. McCabe, N. W. Evans, G. Myeong and V. Belokurov, *Dark matter hurricane: Measuring the S1 stream with dark matter detectors*, *Phys. Rev.* **D98** (2018) 103006, [1807.09004].

[267] C. A. J. O'Hare, N. W. Evans, C. McCabe, G. Myeong and V. Belokurov, *Dark Shards: velocity substructure from Gaia and direct searches for dark matter*, 1909.04684.

[268] N. C. Amorisco, *Contributions to the accreted stellar halo: an atlas of stellar deposition*, *MNRAS* **464** (Jan, 2017) 2882–2895, [1511.08806].

[269] J. Fan, M. Reece and L.-T. Wang, *Non-relativistic effective theory of dark matter direct detection*, *JCAP* **1011** (2010) 042, [1008.1591].

[270] A. L. Fitzpatrick, W. Haxton, E. Katz, N. Lubbers and Y. Xu, *The Effective Field Theory of Dark Matter Direct Detection*, *JCAP* **1302** (2013) 004, [1203.3542].

[271] M. Cirelli, E. Del Nobile and P. Panci, *Tools for model-independent bounds in direct dark matter searches*, *JCAP* **1310** (2013) 019, [1307.5955].

[272] E. Del Nobile, *Complete Lorentz-to-Galileo dictionary for direct dark matter detection*, *Phys. Rev.* **D98** (2018) 123003, [1806.01291].

[273] F. Bishara, J. Brod, B. Grinstein and J. Zupan, *Chiral Effective Theory of Dark Matter Direct Detection*, *JCAP* **1702** (2017) 009, [1611.00368].

[274] F. Bishara, J. Brod, B. Grinstein and J. Zupan, *From quarks to nucleons in dark matter direct detection*, *JHEP* **11** (2017) 059, [1707.06998].

[275] PICO collaboration, C. Amole et al., *Dark Matter Search Results from the PICO-2L C_3F_8 Bubble Chamber*, *Phys. Rev. Lett.* **114** (2015) 231302, [1503.00008].

[276] LUX collaboration, D. S. Akerib et al., *Results from a search for dark matter in the complete LUX exposure*, *Phys. Rev. Lett.* **118** (2017) 021303, [1608.07648].

[277] SuperCDMS collaboration, R. Agnese et al., *Results from the Super Cryogenic Dark Matter Search Experiment at Soudan*, *Phys. Rev. Lett.* **120** (2018) 061802, [1708.08869].

[278] PandaX-II collaboration, X. Cui et al., *Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment*, *Phys. Rev. Lett.* **119** (2017) 181302, [1708.06917].

[279] XENON collaboration, E. Aprile et al., *Dark Matter Search Results from a One Ton-Year Exposure of XENON1T*, *Phys. Rev. Lett.* **121** (2018) 111302, [1805.12562].

[280] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J.* **C71** (2011) 1554, [1007.1727].

[281] J. Neyman and E. S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, *Philosophical Transactions of the Royal Society of London Series A* **231** (Jan, 1933) 289–337.

[282] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, *Annals Math. Statist.* **9** (1938) 60–62.

[283] T. D. P. Edwards, B. J. Kavanagh and C. Weniger, *Assessing Near-Future Direct Dark Matter Searches with Benchmark-Free Forecasting*, *Phys. Rev. Lett.* **121** (2018) 181101, [1805.04117].

[284] DarkSide collaboration, P. Agnes et al., *Low-Mass Dark Matter Search with the DarkSide-50 Experiment*, *Phys. Rev. Lett.* **121** (2018) 081307, [1802.06994].

[285] E. Del Nobile, M. Kaplinghat and H.-B. Yu, *Direct Detection Signatures of Self-Interacting Dark Matter with a Light Mediator*, *JCAP* **1510** (2015) 055, [1507.04007].

[286] XENON collaboration, E. Aprile et al., *Constraining the spin-dependent WIMP-nucleon cross sections with XENON1T*, *Phys. Rev. Lett.* **122** (2019) 141301, [1902.03234].

[287] M. Pato, L. Baudis, G. Bertone, R. Ruiz de Austri, L. E. Strigari and R. Trotta, *Complementarity of Dark Matter Direct Detection Targets*, *Phys. Rev.* **D83** (2011) 083505, [1012.3458].

[288] J. F. Cherry, M. T. Frandsen and I. M. Shoemaker, *Halo Independent Direct Detection of Momentum-Dependent Dark Matter*, *JCAP* **1410** (2014) 022, [1405.1420].

[289] N. Bozorgnia, D. G. Cerdeño, A. Cheek and B. Penning, *Opening the energy window on direct dark matter detection*, *JCAP* **1812** (2018) 013, [1810.05576].

[290] XENON100 collaboration, E. Aprile et al., *Likelihood Approach to the First Dark Matter Results from XENON100*, *Phys. Rev.* **D84** (2011) 052003, [1103.0303].

[291] XENON collaboration, E. Aprile et al., *XENON1T dark matter data analysis: Signal and background models and statistical inference*, *Phys. Rev.* **D99** (2019) 112009, [1902.11297].

[292] FCC collaboration, A. Abada et al., *HE-LHC: The High-Energy Large Hadron Collider*, *Eur. Phys. J. ST* **228** (2019) 1109–1382.

[293] TLEP Design Study Working Group collaboration, M. Bicer et al., *First Look at the Physics Case of TLEP*, *JHEP* **01** (2014) 164, [1308.6176].

[294] M. Mangano, P. Azzi, M. D'Onofrio and M. P. Mccullough, *Physics at its limits*, *CERN Cour.* **57** (2017) 34–42.

[295] Theia collaboration, C. Boehm et al., *Theia: Faint objects in motion or the new astrometry frontier*, 1707.01348.

[296] Z. Ivezic, A. J. Connolly, J. T. VanderPlas and A. Gray, *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data*. Princeton University Press, Princeton, NJ, USA, 2014.

[297] P. F. de Salas, K. Malhan, K. Freese, K. Hattori and M. Valluri, *On the estimation of the local dark matter density using the rotation curve of the Milky Way*, *JCAP* **2019** (Oct, 2019) 037, [1906.06133].

[298] L. B. Okun, M. B. Voloshin and V. I. Zakharov, *ELECTRICAL NEUTRALITY OF ATOMS AND GRAND UNIFICATION MODELS*, *Phys. Lett.* **138B** (1984) 115–120.

[299] R. Foot, G. C. Joshi, H. Lew and R. R. Volkas, *Charge quantization in the standard model and some of its extensions*, *Mod. Phys. Lett.* **A5** (1990) 2721–2732.

[300] R. Foot, H. Lew and R. R. Volkas, *Electric charge quantization*, *J. Phys.* **G19** (1993) 361–372, [hep-ph/9209259].

[301] B. Holdom, *Two U(1)'s and Epsilon Charge Shifts*, *Phys. Lett.* **166B** (1986) 196–198.

[302] N. Weiner and I. Yavin, *UV completions of magnetic inelastic and Rayleigh dark matter for the Fermi Line(s)*, *Phys. Rev.* **D87** (2013) 023523, [1209.1093].

[303] E. E. Radescu, *On the electromagnetic properties of majorana fermions*, *Phys. Rev. D* **32** (Sep, 1985) 1266–1268.

[304] M. Pospelov and T. ter Veldhuis, *Direct and indirect limits on the electromagnetic form-factors of WIMPs*, *Phys. Lett.* **B480** (2000) 181–186, [hep-ph/0003010].

[305] A. L. Fitzpatrick and K. M. Zurek, *Dark Moments and the DAMA-CoGeNT Puzzle*, *Phys. Rev.* **D82** (2010) 075004, [1007.5325].

[306] C. Blanco, M. Escudero, D. Hooper and S. J. Witte, *Z' Mediated WIMPs: Dead, Dying, or Soon to be Detected?*, 1907.05893.

[307] DarkSide collaboration, P. Agnes et al., *First Results from the DarkSide-50 Dark Matter Experiment at Laboratori Nazionali del Gran Sasso*, *Phys. Lett.* **B743** (2015) 456–466, [1410.0653].

Appendix A

The WIMP Miracle

Weak scale physics is important for the dark matter (DM) problem because of the "WIMP Miracle". Let us briefly derive why the weak scale could be the correct scale for DM. Consider a Friedmann-Lemaître-Robertson-Walker (FLRW) universe equipped with a scale factor a, a DM particle species of density n undergoing self-annihilation to light species follows the Boltzmann equation

$$\frac{1}{a^3}\frac{\mathrm{d}na^3}{\mathrm{d}t} = -\langle\sigma v\rangle(n^2 - n_{(\mathrm{eq})}^2), \tag{A.1}$$

where $\langle\sigma v\rangle$ is the scattering cross-section and $n_{(\mathrm{eq})}$ is the equilibrium density

$$n_{(\mathrm{eq})} = \begin{cases} g\frac{1.202}{\pi^2}T^3 & T > m, \\ g\left(\frac{mT}{2\pi}\right)^{3/2} e^{-m/T} & T < m. \end{cases} \tag{A.2}$$

Here, m is the DM mass and g is the DM degree of freedom. We can rewrite the equation with $Y = n/s$ (abundance) and $x = m/T$. Dividing eq. (A.1) by the entropy of the universe ($s = g_*\frac{2\pi^2}{45}T^3$) and use the conservation of entropy $sa^3 =$ constant, we get

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = -\frac{\lambda}{x^2}(Y^2 - Y_{(\mathrm{eq})}^2), \quad \text{where } \lambda = g_*\frac{2\pi^2}{45}\frac{\langle\sigma v\rangle m}{H(T=m)}, \tag{A.3}$$

and $H(T = m)$ is the Hubble,

$$H(T)^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{T^4}{3M_{\mathrm{pl}}^2} \tag{A.4}$$

with temperature T replaced with m. $M_{\rm pl} = 1/\sqrt{8\pi G}$ is the planck mass. g_* is the relativistic degree of freedom of the whole universe, which is ~ 100 before electroweak symmetry breaking. We also convert ∂_t to ∂_x by

$$\partial_x(sa^3) = a^3\partial_x s + 3s\Big(\frac{\partial t}{\partial x}\Big)a^3 H = 0. \tag{A.5}$$

At $T \to 0$, *i.e.* $x \to \infty$, the abundance of the DM "freezes-out",

$$Y_\infty \to \frac{1}{\lambda} \tag{A.6}$$

The DM density at late time universe is evaluated as

$$\rho\Big|_{t=\infty} = mn\Big|_{t=\infty} = Y_\infty s_0, \tag{A.7}$$

where $s_0 = 2.3 \times 10^{-38}\,\mathrm{GeV}^3$ is the late-time entropy (evaluated from the CMB and CνB's entropies). Let's assume the heavy mediator for DM annihilation is of the mass scale of the DM,

$$\langle \sigma v \rangle = \frac{\alpha^2}{m^2}, \tag{A.8}$$

where α is the strength of the DM annihilation. Substituting Eqs. (A.6), (A.7), (A.8) into Eq. (A.3), let us evaluate the ratio of DM density to the measured cold dark matter (CDM) density, $\rho_{\rm CDM} = 9.8 \times 10^{-48}\,\mathrm{GeV}^4$,

$$\frac{\rho}{\rho_{\rm CDM}} = \frac{1}{\lambda}\frac{s_0}{\rho_{CDM}}. \tag{A.9}$$

Putting $\alpha \approx \alpha_{\rm em} \approx 1/137$, we find out $m \sim O(100\,\mathrm{GeV} - 1\,\mathrm{TeV})$ depending on the precise value of coupling α and the mediator mass.

Appendix B

Statistical verification of *Gaia*'s dark disk constraint

In this appendix, we carry more extensive statistical verification of the various results and assumptions in Chapter 3.

B.1 Constructing a Volume Complete Density

We use the `gaia_tools` package to determine the: a) selection function by comparing the number counts in *Gaia* DR2 to those in *2MASS*, and b) effective volume completeness in each z bin using the Poisson likelihood approach introduced by Ref. [187]. We modify the default color-magnitude modeling in `gaia_tools`, and discuss, in Appendix B.1.1, its effect on the completeness of the DR2 selection function. We highlight the important parts of the Poisson likelihood approach in Appendix B.1.2.

B.1.1 Color-magnitude modeling

The completeness for TGAS stars has a strong color dependence and drops off sharply at faint magnitudes, $J \gtrsim 12$. To account for this effect, `gaia_tools` calculates the completeness in each bin as a function of a color-dependent magnitude, J_G. However, as the faint end of

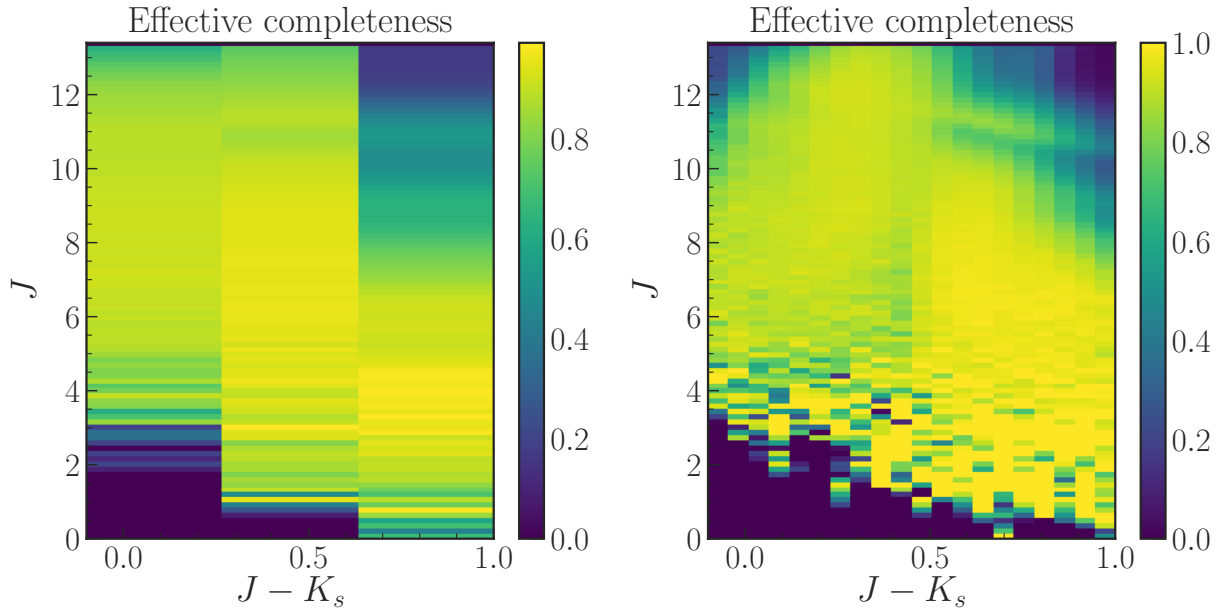


Figure B.1: The effective completeness in color-magnitude space. Left: 3 $J-K_s$ bins. Right: 20 $J-K_s$ bins.

DR2 extends well beyond $J{\sim}12$, we use the J magnitude instead of J_G in our computation.

As a consistency check, we also vary the $J-K_s$ color in the range $-0.05 < J-K_s < 1.05$ from the default 3 bins to 20 bins, as shown in Fig. B.1, and find that the variation on our number density profiles (through the effective volume completeness) is $\lesssim 2\%$. Thus, we conclude that the choice of binning has a negligible effect on the completeness of the selection function.

B.1.2 Poisson process likelihood

Given the the selection function of a survey, $S(J, J-K_s, \alpha, \delta)$, which indicates the fraction of stars observed at any $(J, J-K_s, \alpha, \delta)$, we are interested in determining the volume complete number density, $\nu_\star(X, Y, Z)$, for a particular stellar population. An important ingredient that relates the completeness in $(J, J-K_s, \alpha, \delta)$ space to real (X, Y, Z) space is the (dereddened) color-(absolute) magnitude density (CMD), $\rho_{\rm CMD}(M_J, [J-K_s]_0 | X, Y, Z)$. In case of *Gaia*, the CMD may be determined empirically using an external survey like *2MASS* that

is (almost) complete over the entire sky and a three-dimensional extinction map ($A_J, E(J-K_s)$)$[X,Y,Z]$.

The number density, $\nu_\star(X,Y,Z)$, is determined by assuming that the observed stars are independent samples of an inhomogeneous Poisson process. This process is characterized by its rate function, $\lambda(O|\theta)$, that relates the observables, $O \equiv \{X,Y,Z,J,J-K_s\}$, measured by the survey to the model parameters θ,

$$\lambda(O|\theta) = \nu_\star(X,Y,Z|\theta) \times |J(X,Y,Z;\alpha,\delta,D)| \tag{B.1}$$

$$\rho_{\rm CMD}(M_J,[J-K_s]_0|X,Y,Z)\, S(J,J-K_s,\alpha,\delta), \tag{B.2}$$

where $|J(X,Y,Z;\alpha,\delta,D)| = D^2\cos\delta$ is the Jacobian for the coordinate transformation. Dropping all terms independent of θ, the likelihood of this process, $\mathcal{L}(\theta)$, can be written as,

$$\ln\mathcal{L}(\theta) = \sum_i \lambda(O_i|\theta) - \int dO\lambda(O|\theta) \tag{B.3}$$

$$= \sum_i \nu_\star(X_i,Y_i,Z_i|\theta) - \int dD\, D^2\, d\alpha\, d\delta\, \cos\delta\nu_\star(X,Y,Z|\theta)\, \mathfrak{S}(\alpha,\delta,D),] \tag{B.4}$$

where $\mathfrak{S}(\alpha,\delta,D)$ is the effective selection function as defined by Ref. [187],

$$\mathfrak{S}(\alpha,\delta,D) = \int dJ\, d(J-K_s)\, \rho_{\rm CMD}(M_J,[J-K_s]_0|X,Y,Z)\, S(J,J-K_s,\alpha,\delta). \tag{B.5}$$

We can interpret the effective selection function as the fraction of stars of a stellar population at a distance D and position (α,δ) observed by the survey.

With these ingredients in place, we can estimate the *true* underlying (binned) stellar density, n_k, from the observed number counts of stars, N_k, in non-overlapping bins, $\Pi_k(X,Y,Z)$. Thus, plugging in a parametric density law,

$$\nu_\star(X,Y,Z|\theta) = \sum_k n_k\, \Pi_k(X,Y,Z) \tag{B.6}$$

into the expression for the log-likelihood in Eq. B.4, we obtain,

$$\ln \mathcal{L}(\{n_k\}_k) \tag{B.7}$$

$$= \sum_i \ln \sum_k n_k \Pi_k(X,Y,Z) - \int dD\, D^2\, d\alpha\, d\delta\, \cos\delta \sum_k n_k\, \Pi_k(X,Y,Z)\, \mathfrak{S}(\alpha,\delta,D) \tag{B.8}$$

$$= \sum_k \left[N_k \ln n_k - n_k \int dD\, D^2\, d\alpha\, d\delta\, \cos\delta\, \Pi_k(X,Y,Z)\, \mathfrak{S}(\alpha,\delta,D) \right], \tag{B.9}$$

where the second equality follows from considering all possible combinations of i stars distributed in k identical bins. The maximum likelihood estimate (MLE) can be calculated analytically by differentiating the above equation with respect to n_k and setting the k derivatives to zero. We find that

$$\hat{n}_k = \frac{N_k}{\int dD\, D^2\, d\alpha\, d\delta\, \cos\delta\, \Pi_k(X,Y,Z)\, \mathfrak{S}(\alpha,\delta,D)}, \tag{B.10}$$

which can be written more compactly by defining $\Xi(\Pi_k)$ as the *effective volume completeness* per bin and $\Xi(\Pi_k)\, V(\Pi_k) = \int_{\Pi_k} d^3x\, \mathfrak{S}(\alpha,\delta,D)$ as the effective volume,

$$\hat{n}_k = \frac{N_k}{\Xi(\Pi_k)\, V(\Pi_k)}. \tag{B.11}$$

The uncertainty in our estimate is easily evaluated by calculating the Fisher information,

$$\sqrt{-\frac{\partial^2 \ln \mathcal{L}(\{n_k\}_k)}{\partial(\{n_k\}_k)^2}} \equiv \sigma_{\hat{n}_k} = \frac{\hat{n}_k}{\sqrt{N_k}}. \tag{B.12}$$

B.2 Uncertainty Analysis

In this section, we discuss our choices of bin sizes in the vertical height z and velocity w for constructing the number density and midplane velocity distribution respectively.

The uncertainty in z is given by,

$$\delta z^2\ (\mathrm{kpc}^2) = \left(\frac{\sin b}{\varpi^2}\right)^2 \sigma_\varpi^2 + \left(\frac{\cos b}{\varpi}\right)^2 \sigma_b^2 + \left(\frac{2 \sin b \cos b}{\varpi^3}\right) \sigma_{\varpi b}^2 \tag{B.13}$$

which is dominated by the parallax uncertainty due to the extra factor of ϖ in unit of mas $\approx 10^{-9}$ in the first term. We plot the uncertainty in z (at leading order) as a function of

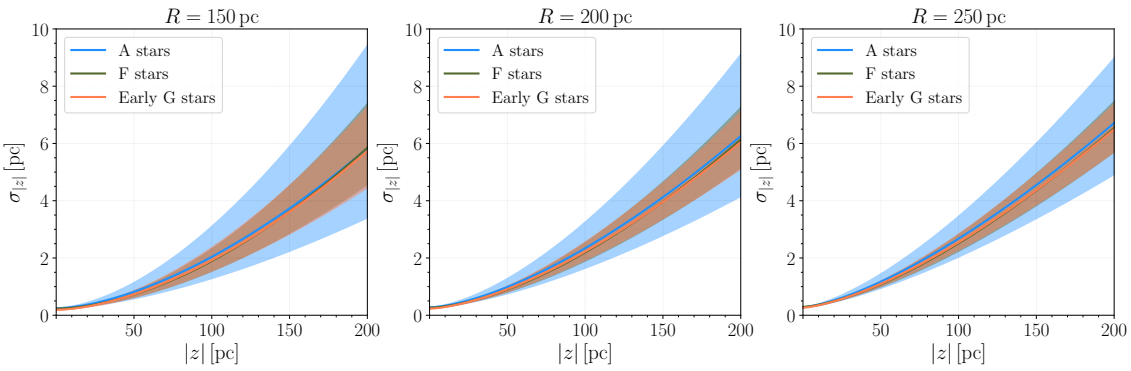


Figure B.2: 1σ spread in the uncertainty (at leading order) of z as a function of z for different radial cuts.

z for all tracers in Fig. B.2. Although the maximum uncertainty is ≈ 10 pc, we conservatively adopt 20 pc as the bin size to account for the underestimation of the reported uncertainties in DR2 [162].

Similarly, the uncertainty in w is

$$\left(\frac{\sigma_w}{w}\right)^2 = \left(\frac{\sigma_\varpi}{\varpi}\right)^2 + \left(\frac{\sigma_{\mu_b}}{\mu_b}\right)^2 + \text{ subleading terms.} \tag{B.14}$$

where the omitted terms are suppressed by 10^{-2} when $|b| < 5^\circ$. Around the midplane, $\sigma_{\mu_b}/\mu_b \lesssim 0.2$, which translates to $\sigma_w \approx 1.5$ km/s. Therefore, we pick 1.5 km/s as the bin size for obtaining the $f_0(w)$ profile.

B.3 Variation of Midplane Cut

The midplane velocity profile is required in Eq. (3.13) to predict the tracer density for a given mass model. With partial radial velocity measured by *Gaia*, we define the midplane in two ways: one is putting a cut on the galactic latitude $|b| < 5^\circ$ while the other is requiring $|z| < (20 - 50)$ pc [193]. For both samples, we approximate v_r by its mean value $\langle v_r \rangle$ in Eq. (3.9) when there is no v_r data available for a star. However, for the z-cut sample, we discard stars with $|b| > 5^\circ$ that do not have any v_r data.

The midplane velocity distributions of the z- and b-cut samples are presented in Fig. B.3

and agree with each other within 1σ uncertainties. We note that the uncertainties in the midplane velocity data using z-cut are smaller than those using the b-cut. The uncertainties are dominated by systematics due to differences between $f(w > 0)$ and $f(w < 0)$. It turns out that the z-cut data is more symmetric about $z = 0$ and thus has smaller uncertainties. In our analysis, we still use the b-cut sample, since there could be a potential selection bias in the z-cut sample, in which we discard a considerable fraction of stars with five-parameter astrometric solutions because we don't know their radial velocities.

We also check the isothermality of the tracers by fitting the midplane data with Gaussian distributions. From the fits, we find that the velocity dispersions σ_z are 5.7, 11.2, 15.0 km/s for A, F and early G stars respectively. The χ^2 of the fits are 11.9, 20.3 and 35.4 for 16, 28, and 28 degrees of freedom respectively. The Gaussian (isothermal) distributions give reasonable fits for A and early G stars, but not as good a fit for F stars. In our analysis, we always use the distributions from data and never their Gaussian fits.

B.4 Bootstrap Statistics

Bootstrap resampling is a standard statistical technique to acquire the mean and uncertainty when there is only one data set available and analytic propagation of uncertainty cannot be performed easily. The basic idea of the method is described below.

Suppose we have a set of N stars labelled as $S_N = \{X_1, X_2, \cdots, X_N\}$. Each star X_k is associated with 6 dimensional phase space coordinates denoted by θ_k. In bootstrap resampling, we make random draws *with replacement* star-by-star from the original set of stars S_N. This generates a new data set $\widetilde{S}_N$ of the same size N, with each star labeled as $\widetilde{X}_k$. Since the draws are with replacement, we expect (many) duplicated coordinate values in the new data set, such as $\widetilde{X}_k = \theta_k$ and $\widetilde{X}_{k+1} = \theta_k$, for large N. Therefore, $\widetilde{S}_N \neq S_N$ in general.

We resample B times the original data set S_N, labeling them as $\widetilde{S}_N^{(1)}$, $\widetilde{S}_N^{(2)}$, ..., $\widetilde{S}_N^{(B)}$. The variance of the underlying distribution in each z bin can be estimated as follows,

$$\widetilde{\sigma}_\nu^2(z) = \frac{1}{B}\sum_{k=1}^{B}(\widetilde{\nu}^{(k)}(z) - \overline{\widetilde{\nu}}(z))^2\,, \qquad \text{where} \quad \overline{\widetilde{\nu}}(z) = \frac{1}{B}\sum_{k=1}^{B}\widetilde{\nu}^{(k)}(z). \tag{B.15}$$

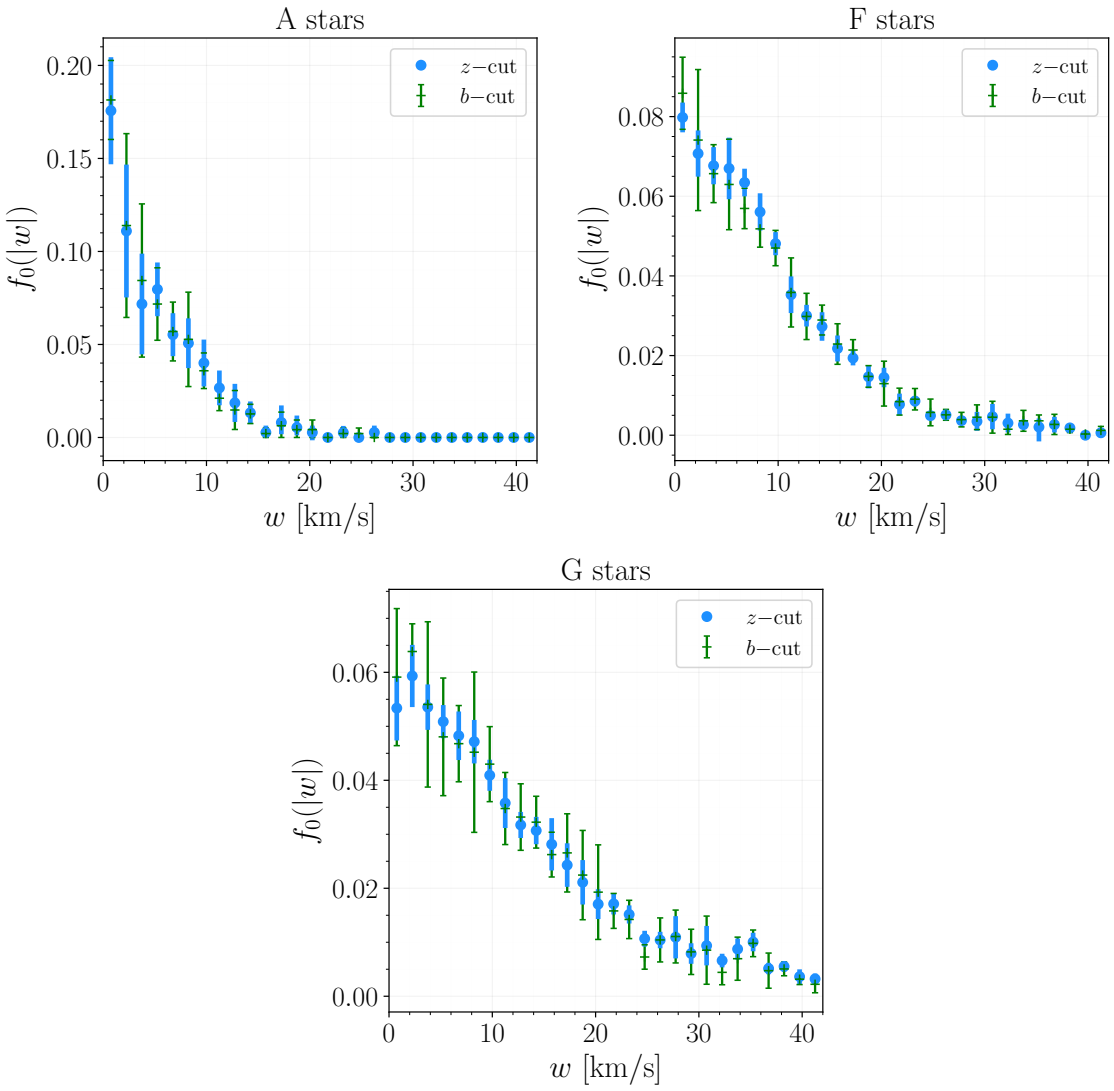


Figure B.3: Midplane velocity distribution $f_0(|w|)$ for A (left), F (middle) and early G (right) stars. The distributions obtained using the $|b| < 5°$ cut (green) and the $|z| < 20$ pc cut (blue) are consistent within error bars.

For sufficiently large B, it can be proven that $\widetilde{\sigma}_\nu^2 \to \sigma_\nu^2$ [296].

In propagating velocity uncertainties into prediction uncertainties via the bootstrap method, we choose $B = 1000$ as a compromise between computational time and statistical precision. We take the bin-by-bin variance of all the predictions based on the 1000 resampled velocity sets as the estimator for the statistical uncertainty, $(\sigma_\nu^2)^{\mathrm{mod,stat}}$, of the predicted profile ν^{mod}.

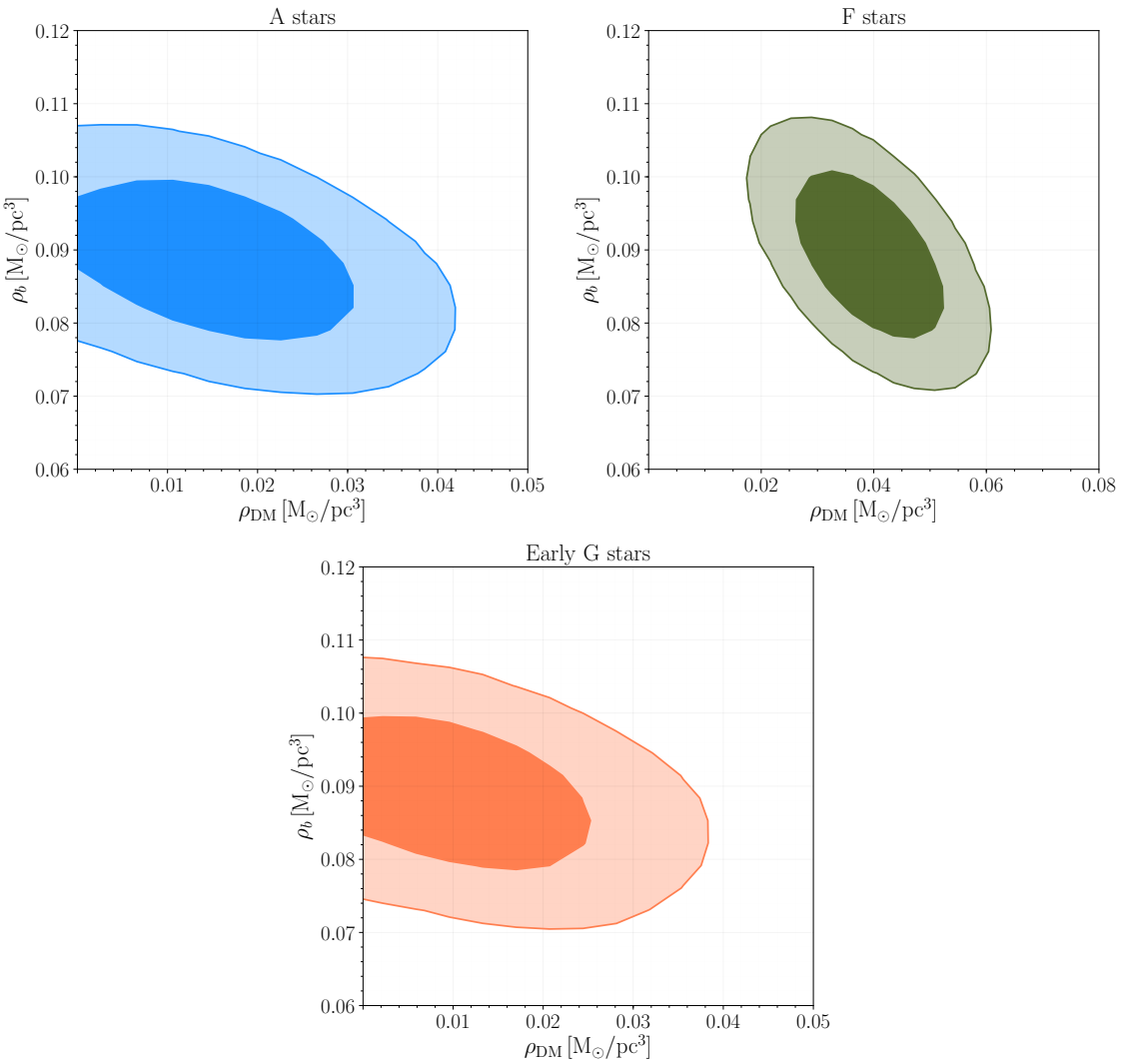


Figure B.4: Marginalized posteriors indicating the degeneracy between the local densities of baryons ρ_b and halo DM $\rho_{\rm DM}$.

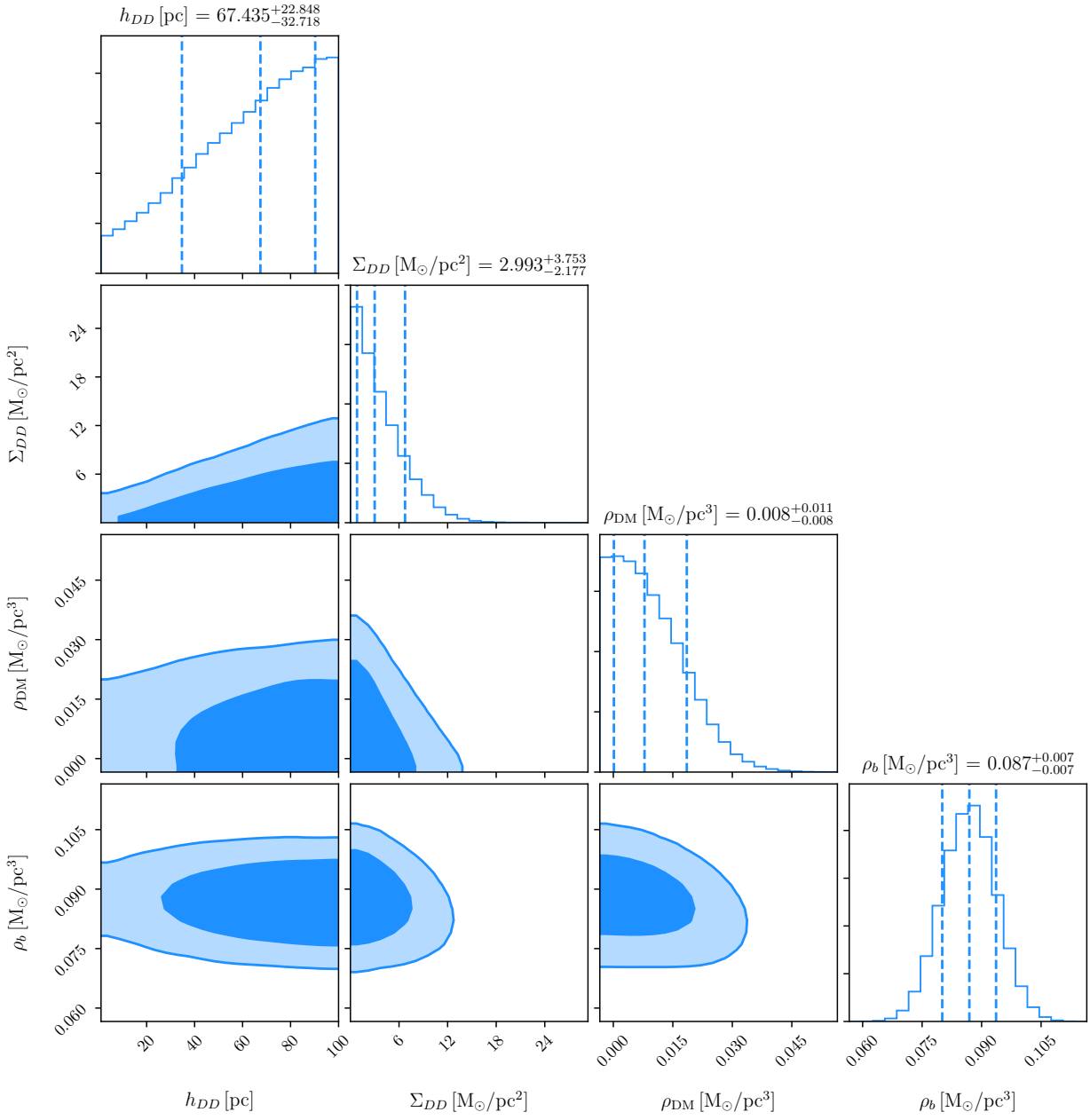


Figure B.5: Marginalized posterior distributions of thin DD parameters, local dark matter density ρ_{DM}, and the total baryon density in the midplane ρ_b for A stars. The dark (light) shaded regions indicate the 68% (95%) credible regions, whereas the dashed lines represent the 16th, 50th, and 84th percent quantile values of the posterior distribution.

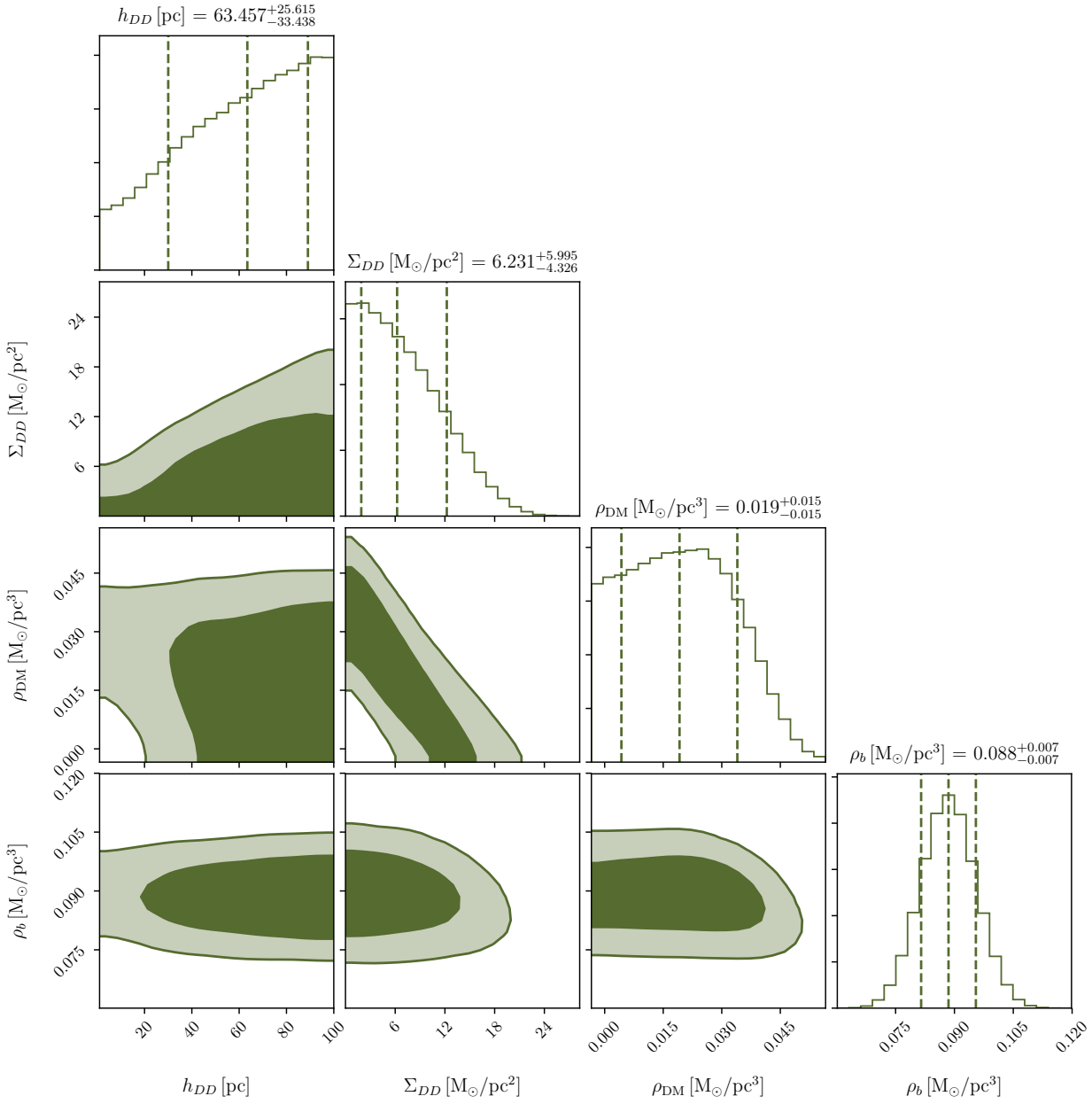


Figure B.6: Marginalized posterior distributions of thin DD parameters, local dark matter density ρ_{DM}, and the total baryon density in the midplane ρ_b for F stars. The dark (light) shaded regions indicate the 68% (95%) credible regions, whereas the dashed lines represent the 16th, 50th, and 84th percent quantile values of the posterior distribution.

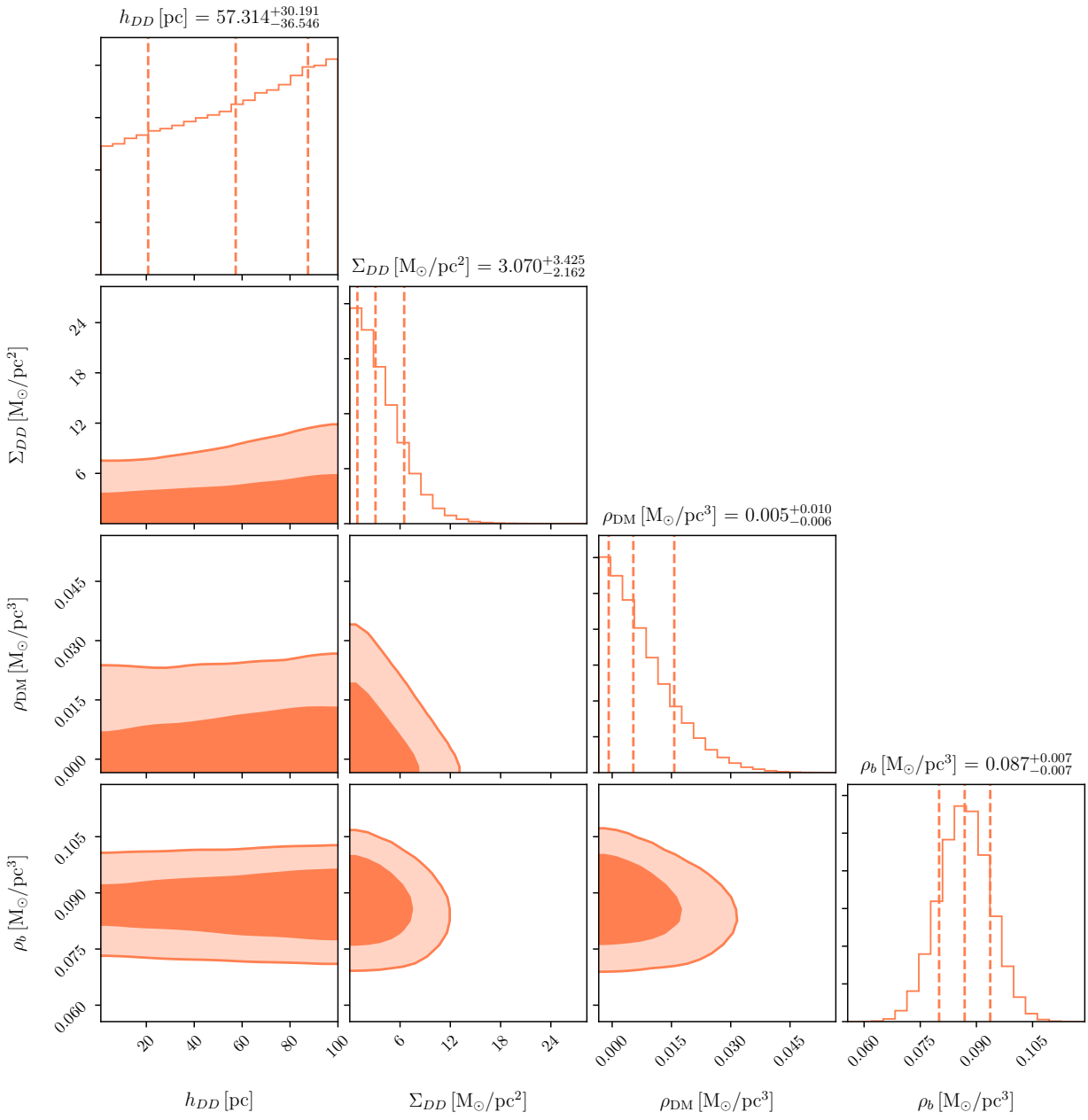


Figure B.7: Marginalized posterior distributions of thin DD parameters, local dark matter density ρ_{DM}, and the total baryon density in the midplane ρ_b for G stars. The dark (light) shaded regions indicate the 68% (95%) credible regions, whereas the dashed lines represent the 16th, 50th, and 84th percent quantile values of the posterior distribution.

Appendix C

Extended Materials for Dark Matter Direct Detection Forecast

In this appendix, we extend the material given in Chapter 4. We give the explanation for direct detection (DD) differential rate in Section C.1 and the detailed form factors for nuclear nonrelativistic effective field theory (NREFT) in Section C.2. In Section C.3, we provide some benchmark models for the NREFT operators. Lastly, we provide some details on the next generation experiments in Section C.4.

C.1 DD basics

The rate of DM scattering off nucleus in a DD experiment is given by $R = n_\chi \langle \sigma v_\chi \rangle$, where n_χ and v_χ are the local DM number density and speed relative to the Earth respectively, σ is the scattering cross section and $\langle \ldots \rangle$ indicates an average over the local DM velocity distribution. The differential recoil rate for a target T per unit recoiling energy can be

written as,

$$\frac{\mathrm{d}R}{\mathrm{d}E_R} = \xi_T N_T \frac{\rho_\chi}{m_\chi} \int_{v_{\min}}^{v_{\rm esc}} \mathrm{d}^3 v \, v \tilde{f}(\boldsymbol{v}) \frac{\mathrm{d}\sigma}{\mathrm{d}E_R}(v, E_R), \tag{C.1}$$

$$= \frac{\xi_T}{32\pi m_\chi^2 m_T^2} \times \underbrace{\frac{\rho_\chi}{m_\chi} \int_{v_{\min}}^{v_{\rm esc}} \mathrm{d}^3 v \, \frac{\tilde{f}(\boldsymbol{v})}{v}}_{\text{astrophysics}} \times \underbrace{\frac{1}{(2J+1)(2J_\chi+1)} \sum_{\text{spins}} |\mathcal{M}|^2}_{\text{particle/nuclear physics}}, \tag{C.2}$$

where in the second line, we use

$$\frac{d\sigma}{dE_R} = \frac{1}{32\pi v^2} \frac{\overline{|\mathcal{M}|^2}}{m_\chi^2 m_T}, \quad \text{with } \overline{|\mathcal{M}|^2} = \frac{1}{(2J+1)(2J_\chi+1)} \sum_{\text{spins}} |\mathcal{M}|^2 , \tag{C.3}$$

and group together different factors by the main type of physics they rely on. In the master equation for the differential rate, ξ_T is the mass fraction for each type of target nucleus T (the detector could be composed of different nuclides), m_T is the target nucleus mass and $N_T = 1/m_T$ is number of scattering centers per unit mass. ρ_χ is the local DM density in the solar system, which we take to be $\rho_\chi = 0.4$ GeV/cm^3 [160,180,297]. While there could be an $\mathcal{O}(1)$ uncertainty in the determination of ρ_χ due to non-equilibrium effects in the dynamical modeling of the MW, we can always absorb it into the overall normalization of the recoil rate. In other words, local DM density only affects the overall rate but not the recoil shape. Thus in our paper, we will ignore uncertainty in the local DM density, focusing instead on the more interesting effects from varying the velocity distributions. $\tilde{f}(\boldsymbol{v})$ is the local DM velocity distribution in the Earth frame. The velocity integration range is bounded from above by the escape velocity of DM particles $v_{\rm esc}$. The minimal velocity for DM to scatter with recoiling energy E_R, in the case of elastic scattering, is

$$v_{\min} = \sqrt{\frac{m_T E_R}{2\mu_T^2}}, \tag{C.4}$$

where μ_T is the DM-nucleus reduced mass. Lastly, in the part that depends on particle and nuclear physics, J and J_χ are nuclear and DM spins respectively. $\overline{|\mathcal{M}|^2}$ is the scattering matrix element squared averaged over $2J_\chi + 1$ and $2J + 1$ initial DM and nuclear spins, and summed over the final spins. Note that the recoiling spectrum depends on the detector material in multiple ways. For example, in the velocity integration, $v_{\min}$ depends on the

target nucleus mass and the matrix element depends on the nuclear form factor, which is determined by the type of the target.

In terms of the form factors, the spin-averaged amplitude squared is then

$$\overline{|\mathcal{M}|^2} = 16 m_T^2 m_\chi^2 \sum_{i,j=1}^{12} \sum_{N,N'=n,p} \frac{c_i^{(N)} c_j^{(N')}}{(q^2 + m_{\mathrm{med};i}^2)(q^2 + m_{\mathrm{med};j}^2)} \mathcal{F}_{i,j}^{(N,N')}(q^2, v^2). \tag{C.5}$$

C.2 Nuclear form factors in NREFT

An important ingredient for calculating the recoil rate is the form factor, $\mathcal{F}_{a,b}^{(N,N')} \sim \mathcal{O}_a^{(N)} \times \mathcal{O}_b^{(N')}$, that encodes nuclear response functions. Ref. [270] showed that the complete basis of NREFT operators corresponds to six different types of nuclear response functions $\widetilde{F}_i^{(N,N')}$, where $i \in \{M, \Delta, \Sigma', \Sigma'', \tilde{\Phi}', \Phi''\}$ and the superscripts $N, N' = n, p$ indicates the type of nucleon. Listed below are the nuclear form factors for the relevant operators in terms of the response functions,

$$\mathcal{F}_{1,1}^{(N,N')} = \widetilde{F}_M^{(N,N')}, \tag{C.6}$$

$$\mathcal{F}_{4,4}^{(N,N')} = \frac{1}{16}(\widetilde{F}_{\Sigma'}^{(N,N')} + \widetilde{F}_{\Sigma''}^{(N,N')}), \tag{C.7}$$

$$\mathcal{F}_{5,5}^{(N,N')} = \frac{q^2}{4}\Big(\big(v^2 - \frac{q^2}{4\mu_T^2}\big)\widetilde{F}_M^{(N,N')} + \frac{q^2}{m_N^2}\widetilde{F}_\Delta^{(N,N')}\Big), \tag{C.8}$$

$$\mathcal{F}_{6,6}^{(N,N')} = \frac{q^4}{16}\widetilde{F}_{\Sigma''}^{(N,N')}, \tag{C.9}$$

$$\mathcal{F}_{8,8}^{(N,N')} = \frac{1}{4}\Big(\big(v^2 - \frac{q^2}{4\mu_T^2}\big)\widetilde{F}_M^{(N,N')} + \frac{q^2}{m_N^2}\widetilde{F}_\Delta^{(N,N')}\Big), \tag{C.10}$$

$$\mathcal{F}_{9,9}^{(N,N')} = \frac{q^2}{16}\widetilde{F}_{\Sigma'}^{(N,N')}, \tag{C.11}$$

$$\mathcal{F}_{11,11}^{(N,N')} = \frac{q^2}{4}\widetilde{F}_{M}^{(N,N')}, \tag{C.12}$$

$$\mathcal{F}_{4,5}^{(N,N')} = \frac{q^2}{8m_N}\widetilde{F}_{\Sigma',\Delta}^{(N,N')}, \tag{C.13}$$

$$\mathcal{F}_{4,6}^{(N,N')} = \frac{q^2}{16}\widetilde{F}_{\Sigma''}^{(N,N')}, \tag{C.14}$$

$$\mathcal{F}_{8,9}^{(N,N')} = \frac{q^2}{8m_N}\widetilde{F}_{\Sigma',\Delta}^{(N,N')}, \tag{C.15}$$

where the subscripts indicate the NREFT operator(s). The response functions $\widetilde{F}_i^{(N,N')}$ for the nuclides relevant in our analysis (see Appendix C.4) have been adopted from Appendix A.3 of ref. [270].

C.3 Some benchmark models

In this appendix, we present a few more details of benchmark models listed in Table 4.2.

Heavy gauge boson mediator: In this case, quarks and DM are both charged under a broken gauge symmetry with a heavy gauge boson. The interactions are given by,

$$\mathcal{L} \supset g_\chi \bar{\chi}\gamma_\mu \chi Z'^\mu + g_q \bar{q}\gamma_\mu q Z'^\mu, \tag{C.16}$$

where Z' is the heavy gauge boson that could be integrated out, giving rise to a four fermion contact operator.

Millicharged DM: DM carries a small electric charge $\epsilon_\chi e$ and couples to the SM photon, A_μ,

$$\mathcal{L} \supset \epsilon_\chi e A_\mu \bar{\chi}\gamma^\mu \chi + A_\mu J^\mu. \tag{C.17}$$

χ particle can either carry a fractional electric charge directly [298–300] or charged under other $U(1)$ which kinetically mixes with the SM hypercharge [301]. A_μ couples to the SM current, which is given by

$$J^\mu = \bar{p}(k')\left(\frac{e(k+k')^\mu}{2m_N} + \frac{g_p}{2}\frac{i\sigma^{\mu\nu}q_\nu}{2m_N}\right)p(k) + \bar{n}(k')\left(\frac{g_n}{2}\frac{i\sigma^{\mu\nu}q_\nu}{2m_N}\right)n(k), \tag{C.18}$$

where g_p and g_n are the g-factors of protons and neutrons respectively.

Magnetic dipole DM with a light mediator: DM does not carry electric charge, but it could still couple to the SM photon through a loop of other charged species. This could generate an anomalous magnetic dipole moment of the DM particle [302]:

$$\mathcal{L} \supset \frac{i\mu_\chi}{2}\bar{\chi}\sigma^{\mu\nu}\chi F_{\mu\nu}. \tag{C.19}$$

Magnetic dipole DM with a heavy mediator: magnetic dipole DM can also interact with the SM electromagnetic current through a heavy mediator. One way is through the kinetic mixing of a broken dark $U(1)_D$ with the SM $U(1)_{\rm em}$

$$\mathcal{L} \supset \frac{i\mu_\chi^D}{2}\bar{\chi}\sigma^{\mu\nu}\chi F^D_{\mu\nu} + \varepsilon F^D_{\mu\nu}F^{\mu\nu}, \tag{C.20}$$

where the dark photon is heavy with a mass larger than the momentum transfer in DD, $m_D^2 \gg q^2$. Integrating out the heavy dark photon leads to a coupling between DM and the SM photon as listed in Table 4.2.

Electric dipole DM with a light mediator: if a dark sector has CP violating interaction, it is also possible that the DM carries an electric dipole moment which aligns with the DM spin.

Anapole DM with a heavy mediator: if DM is a majorana fermion, monopole and dipole interactions with electromagnetism is forbidden by CPT symmetry. The only allowed EM coupling is the anapole coupling. For example, the majorana DM could have the anapole moment under a dark broken $U(1)$, which kinetically mixes with the SM photon [303]. Integrating out the heavy dark photon, we have

$$\mathcal{L} \supset i g_{\rm ana}\bar{\chi}\gamma^\mu\gamma^5\chi\partial^\nu F_{\mu\nu}. \tag{C.21}$$

Early references on anapole DM could be found in [304, 305].

C.4 Next-generation experiments

We describe configurations of the next-generation experiments used in our forecasts in this Appendix. Although there is a rich experimental program underway with different targets,

Experiment	Target	Exposure [ton · year]	Energy window [keV]
DARWIN [217]	^{131}Xe	200	$[5-40]$
DARKSIDE-20K (High) [218]	^{40}Ar	200	$[32-200]$
DARKSIDE-20K (Low) [284]	^{40}Ar	200	$[0.6-15]$

Table C.1: Schematic outline of the next-generation experiments included in our results.

we only focus on experiments that use the dual-phase (liquid-gas) time projection chamber (TPC) technology with noble elements as targets. Table C.1 contains a schematic outline of three experiments: a DARWIN-like liquid Xenon (LXe) experiment, a DARKSIDE-20K-like liquid Argon (LAr) experiment with both a high and low DM mass program. Each of them are discussed in turn below:

1. DARWIN: The DARk matter WImp search with liquid xenoN (DARWIN) is Generation-3 LXe experiment proposed by the XENON collaboration. With a 40 ton active volume and 5 year observation time, it's projected 200 ton $\times$ year exposure will allow DM probes to reach the neutrino-floor for DD experiments. We model it based on the conceptual design report [217] and the latest XENON-1T configuration. We choose the observation window in recoil energy to be [5-40] keV (cf. [306]) and divide it into 19 equally spaced bins for a primary scintillation signal (S1)-only analysis. This assumes near-perfect electron recoil (ER) background subtraction that is ensured by focusing only on the events in the nuclear recoil (NR) region. In practice, we achieve this by convolving our theory recoil spectra with the efficiency curve from the latest XENON DM analysis (given in Fig. 1 of ref. [279]) and multiplying by a factor of 0.5. Following ref. [238], we conservatively adopt the rate for NR background components (including a 10% uncertainty) from ref. [249], rescaling them with the appropriate exposure factors. The precise choice of the background should not affect our results, since all our constraints are derived for the signal-limited region with $\mathcal{O}(100)$ total events.

2. DARKSIDE-20K (High): We follow the official proposal for a neutrino-floor LAr experiment with an integrated exposure of 200 ton×year achieved over a 10 year observation period. Our model detector has an S1-only search region with 19 linearly spaced bins in the range [32-200] keV that provides us a unique probe of $\mathcal{O}$(0.1-1) TeV DM candidates. An advantage of using Ar targets is their superior pulse-shape discrimination (PSD) of the S1 signal allowing for nearly background-free detection of any DM events. Thus, we only consider a background rate of 0.1 events over the entire observation period along with a 10% systematic uncertainty. Finally, we adopt the efficiency curve for NR detection from Fig. 6 of ref. [307].

3. DARKSIDE-20K (Low): Although there is no outlined for a low DM mass search using the Gen3 DARKSIDE-20K setup discussed above, we make the science case for one in Sec. 4.5.4 as a complementary probe to the DARWIN experiment. We borrow the configuration in ref. [284] that used the DARKSIDE-50 apparatus to perform an ionization (S2)-only analysis, albeit with a 200 ton×year exposure. There are two major differences compared to the high DM mass search: i) an S2-only analysis allows a far lower recoil energy threshold, ii) the PSD is no longer available and we need to contend with higher background rates. Thus, our fiducial analysis has a recoil energy observation window of [0.6-15] keV and an optimistic background rate of 1 event for the entire observation period. We also use a constant acceptance of 0.43 following the discussion below Fig.1 of ref. [284].