## *The dynamics of self-interacting dark matter in galaxy clusters*

SIRKS, ELLEN,LAURA

# The dynamics of self-interacting dark matter in galaxy clusters

**Ellen Laura Sirks**

A thesis presented for the degree of

Doctor of Philosophy


Institute for Computational Cosmology

Department of Physics

The University of Durham

September 2022

# The dynamics of self-interacting dark matter in galaxy clusters

Ellen Laura Sirks

## Abstract

This thesis presents three different but connected projects related to the study of the nature of dark matter (DM) using galaxy clusters. In particular, in the first two projects I use cosmological simulations to investigate how DM particles that interact through forces other than gravity affect galaxy clusters as a whole as well as the galaxies that reside inside them.

First, I compared the mass loss of galaxies accreted unto simulated clusters ran with both cold dark matter (CDM) and self-interacting dark matter (SIDM) physics. Due to the additional interactions between the DM haloes of the galaxies and of the clusters, we expect there to be additional mass loss in SIDM galaxies on top of the tidal mass loss due to the gravitational field from the cluster. Indeed, I find that on average not only do SIDM galaxies lose more mass, they are also more susceptible to total disruption.

Second, I investigated the effects of SIDM on major mergers of galaxy clusters. In such events, the gas is offset from the collisionless galaxies due to ram pressure. If the SIDM cross-section is non-zero, the DM can be offset from the galaxies as well. By comparing the offsets of the gas, DM, and stars in simulations ran with different SIDM cross-sections, I found that the DM offset increases with cross-section as expected from analytical models.

The third project was undertaken for the upcoming balloon-borne telescope SuperBIT, whose main science goal will be to map out the DM in and surrounding galaxy clusters. To keep up with SuperBIT's (and any possible successor's) relatively high data rate, we have developed a toolkit of hardware and software that would allow us to physically downlink data mid-flight. I wrote software predicting the trajectories of the system, given the location and time of the release. The system was successfully tested from beginning to end during the SuperBIT 2019 test flight.

In essence, all three projects are based around simulations to predict the trajectories of some form of matter falling into some other form of matter, i.e. DM into clusters, or parachutes into the Earth's atmosphere. The intention was to bring the three projects together and use the SuperBIT hardware that I have helped develop to measure the behaviour of DM and calibrate it against the cosmological simulations. Unfortunately SuperBIT's first science flight was delayed due to the COVID-19 pandemic, and I did not get to measure the DM effects on real astronomical data. I intend to do so in the future.

Supervisors: Richard Massey and Carlos Frenk

# Acknowledgements

Undertaking a PhD has been no easy task, and I could not have done it without the support and help of many wonderful people. I hope I can convey here even a small amount of the gratitude I feel for you all.

First and foremost I want to thank my supervisors Richard Massey and Carlos Frenk, without whom I would not have been able to do any of the work presented here. Because of them, I have learned so much, seen so many places, and met so many great people. I want to thank them for their support, their time, their willingness to share their expertise with me, and finally for making my PhD the wonderful time it was.

I want to thank Kyle Oman and Andrew Robertson for always being ready to help, no matter how silly my questions were. Especially during the pandemic, when my brain seemed to work at half capacity.

Many thanks to Paul Clark, for letting me join his DRS project and getting me involved in the fascinating world of electrical engineering. Thanks to David Harvey for letting me join his project and work with him. Thanks to Anna Niemiec and David Lagatutta for all the helpful feedback over the years. Thanks to Lorraine Coghill for all the wonderful outreach I got to do. Big thanks to Shufei Rowe and Lindsay Borrero for all their help with arranging events, and with travel to conferences and summer schools. Thanks to Lourdes and Mike for the lovely dinners and gifts.

I have made some wonderful friends during my four years here. Vicky, Amy and Alice, thank you for all the lunches, dinners and chats. Being able to vent to you has made all the difference in keeping my sanity. Many thanks to Carol, Arnau, and Jack, who have not only been great friends, but also great housemates. I will miss chatting with you in the kitchen. Thanks to Elly, Scott, Connor, and Stephen, for all the laughs. I am lucky to have met you. Enormous thanks to Joaquin, Jake, Aidan, Victor, James, Nicole, Matteo, Sasha, Scott, Emmy, Martina, and Naomi.

Stepping away from Durham, a big thank you to my friends from Edinburgh, Mark, Ailsa, Fraser, Sarah, Anna, Luciana, Alex, and Adrianne. I miss living in the same city as you. I am also immensely grateful to my friends from home. To the furby's: Chris, Tim, Yasmine, Sterre, Isidor, Jesper, Laura, Merel, and Naomi, who I do not get to see as often as I want to, but who I know will always be there. To Maxime and Vicky, who have been with me for over two decades now, and who have been like sisters to me.

I have to thank my parents Jacqueline and André for their unwavering support throughout the years, and for patiently listening to my rambling explanations of my work that I am sure made no sense most of the time. They never gave me any idea that I could not do whatever I wanted to do. I am proud to be your daughter, I hope I have made you proud as well.

Finally, I want to thank Miguel for always being there for me. Even when we are half way across the world from each other, I feel your love and support. Durham has given me many wonderful things, but I never could have dreamed it would have given me the love of my life.

*This thesis was supported by the Royal Society grant RGF\EA\180026.*

*Toegewijd aan*

Mama en papa

*Dedicado a*

Miguel

# Contents

# Declaration

The work in this thesis is based on research carried out at the Institute for Computational Cosmology, Department of Physics, University of Durham, England between October 2018 and September 2022. No part of this thesis has been submitted elsewhere for any other degree or qualification, and it is the sole work of the author unless referenced to the contrary in the text.

## Publications

Chapter 2 has been published as a paper in the Monthly Notices of the Royal Astronomical Society (MNRAS):

*The effects of self-interacting dark matter on the stripping of galaxies that fall into clusters*
**Ellen L. Sirks**, Kyle A. Oman, Andrew Robertson, Richard Massey, Carlos Frenk
*MNRAS* 2022, Volume 511, Issue 4, pp. 5927-5935, arXiv:2109.03257

Chapter 5 has been published as a paper in the Journal of Instrumentation (JINST):

*Download by Parachute: Retrieval of Assets from High Altitude Balloons*
**E. L. Sirks**, P. Clark, R. J. Massey et al. *JINST* 2020, Volume 15, Issue 05, pp. P05014, arXiv:2004.10764

We hope to submit a modified version of chapter 3 for publication.

# List of Figures

# List of Tables

# Nomenclature

**BAHAMAS** BAryons and HAloes of MAssive Systems

**BIT** Balloon-borne Imaging Testbed

**BCG** brightest cluster galaxy

**CSA** Canadian Space Agency

**CDM** cold dark matter

**C-EAGLE** Cluster-EAGLE

**CNES** Centre National d'études Spatiales

**CMB** cosmic microwave background

**CSBF** Columbia Scientific Balloon Facility

**DM** dark matter

**DRS** Data Recovery System

**EAGLE** Evolution and Assembly of GaLaxies and their Environments

**GNSS** global navigation satellite system

**GR** General Theory of Relativity

**HAB** high altitude balloon

**HST** Hubble Space Telecope

**JWST** James Webb Space Telescope

**LSST** Legacy Survey of Space and Time

**MW** Milky Way

**NASA** National Aeronautics and Space Administration

**NGRST** Nancy Grace Roman Space Telescope

**NFW** Navarro-Frenk-White

**SHMR** stellar-to-halo mass relation

**SIDM** self-interacting dark matter

**SPB** super pressure balloon

**SuperBIT** Super-Pressure Balloon-borne Imaging Telescope

**WIMP** weakly interacting massive particle

Chapter 1

# Introduction

*"If you wish to make an apple pie from scratch, you must first invent the universe."*

— Carl Sagan, *Cosmos*

The word cosmology is derived from the ancient Greek κοσμος, meaning 'world' or 'order' and the suffix -λογια, meaning 'discourse' or 'study of'. Putting this together, the word cosmology roughly translates to 'study of the world'. Of course what is considered the world or the cosmos has changed over time, but nowadays cosmology refers to the branch of astronomy that studies the origin and evolution of the entire Universe, from its very beginning until the present and into the future.

## 1.1 A brief history of cosmology

While the word cosmology may have been derived from ancient Greek, every culture on Earth has partaken in the study of the Universe. Cosmology is as old as humankind itself. For as long as we have existed, we have asked ourselves questions like 'Why am I here?' and 'What is going on?'.

Around 964 AD, Persian astronomer Abd al-Rahman al-Sufi noted the presence of a 'little cloud' in the night sky (Hafez, 2010). We now know he was observing the Andromeda galaxy (M31). This is the first known mention in writing of any galaxy apart from our own. Over the course of history more of such clouds or *nebulae*

were discovered, however, it was not until centuries later that these nebulae were first suggested to be structures outside of our own Milky Way (MW) galaxy. The German Philosopher Immanuel Kant is often cited as the originator of this idea, but it is now believed that it was actually County Durham's own Thomas Wright who first speculated that faint nebulae were distant galaxies (Wright, 2014). He wrote in reference to these nebulae that '...those in all likelihood may be external creation, bordering upon the known one, too remote for even our telescopes to reach.'. The idea was popularised by Kant who termed these clouds *island universes*.

It was only a little over a hundred years ago that the first conclusive evidence was found proving the existence of galaxies outside of the MW. The American astronomer Vesto Melvin Slipher performed the first measurement of the radial velocity of M31. He observed a Doppler shift in its spectral lines, which revealed that M31 is moving towards us (Slipher, 1913). He also discovered Doppler shifts in the spectral lines of other nebulae, showing that they were actually moving away from us (Slipher, 1915, 1917). At that time, Slipher did not realise quite how significant his observations were, as they provided us with the first observational basis for the expansion of the Universe.

The existence of galaxies beyond our own remained a matter of debate for some time after Slipher's observations. In 1920, 'the Great Debate' was held at the Smithsonian Museum of Natural History between the astronomers Harlow Shapley and Heber Curtis. Shapley believed that distant nebulae were relatively small and lay within the outskirts of the MW galaxy, while Curtis held the believe that they were in fact independent galaxies, implying that they were exceedingly large and distant (for a review of the debate see, e.g., Hoskin, 1976).

The debate was finally settled once and for all in 1925. It was then that the American astronomer Edwin Hubble used Cepheid variables, a type of star with a definite relationship between its period and luminosity, to calculate the distance to the nebulae M31 and the Triangulum galaxy (M33). He found that they were much too distant to be part of the MW (Hubble, 1925). A few years later, having

studied the distances and radial velocities of 24 more galaxies, he found that the recessional velocity of these galaxies was proportional to their distance (Hubble, 1929), which is encapsulated in Hubble's law:

$$v = H_0 r, \tag{1.1}$$

where $v$ and $r$ are the recessional velocity and distance of a galaxy respectively, and $H_0$ is the Hubble constant $H(t)$ at present time ($t = t_0$). The Hubble constant is time-dependent and describes the rate of expansion of the Universe at a given time. We will revisit the Hubble constant and the Hubble law in the following sections. To determine the velocities of galaxies, both Hubble and Slipher used a phenomenon know as *redshift* ($z$). Redshift is defined as the fractional change in a photon's wavelength from when it was emitted to when it was received, i.e.

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{em}}}{\lambda_{\text{em}}} = \frac{\lambda_{\text{obs}}}{\lambda_{\text{em}}} - 1, \tag{1.2}$$

where $\lambda_{\text{obs}}$ and $\lambda_{\text{em}}$ are the observed and emitted wavelength respectively. When a light-emitting object moves towards us, the wavelength of the light is displaced towards the bluer end of the spectrum. Vice versa, if the object is moving away from us the object appears redder than it is. Redshift refers to the shift towards the red end of the spectrum. A shift towards the bluer end of the spectrum is called 'blueshift', but is generally referred to as having negative redshift. The magnitude of the redshift is related to how fast an object is moving, and can be used to derive an object's velocity $v$ using the Fizeau-Doppler formula:

$$z = \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}} \approx \frac{v}{c}, \tag{1.3}$$

where $c$ is the speed of light, and the second relation is for velocities $v \ll c$. We will discuss redshift in more detail in section 1.2.2.

## 1.2 Modern cosmology

Our understanding of modern cosmology rests on two theoretical pillars. The first is that Einstein's General Theory of Relativity (GR), first introduced in his 1917 paper 'Cosmological Considerations in the General Theory of Relativity', is the correct description of gravity. This work introduced the first relativistic model of the Universe. At the time, the Universe was thought to be *static*, i.e. it is infinite in both time and space, and neither contracts nor expands. In order to satisfy this assumption, Einstein added a cosmological constant ($\Lambda$) to his field equations which provided a repulsive force to counteract the effects of gravity. After Hubble's work showing the expansion of the Universe, he soon removed this constant from his equations. However, with the discovery in 1988 that the Universe is expanding at an *accelerated* rate, $\Lambda$ was reintroduced and reinterpreted as the energy density of space. We will revisit the expansion of the Universe and the cosmological constant in more detail later.

The second pillar supporting modern cosmology is the Cosmological Principle, which states that the Universe is both homogeneous and isotropic on a sufficiently large enough scale (above $\sim$100 Mpc). An isotropic Universe has no preferred direction, it looks the same no matter in *what direction* an observer points their telescope. An homogeneous Universe has no preferred locations. It looks the same no matter *where* an observer places their telescope. In other words, Earth does not hold a particularly special location within the Universe. Hubble's law seemingly violates this principle, as it appears to put us in a special location or a centre, from which everything else moves away. This is not the case. Consider a distribution of galaxies that is made to expand uniformly. Mathematically, this means that all position vectors $\mathbf{x}$ at time $t$ are scaled versions of their values at a reference time $t_0$:

$$\mathbf{x}(t) = R(t)\mathbf{x}(t_0), \tag{1.4}$$

where $R(t)$ is the cosmic scale factor, which represents the expansion of the Universe. Differentiating the above equation with respect to time gives

$$\dot{\mathbf{x}}(t) = \dot{R}(t)\mathbf{x}(t_0) = \left(\frac{\dot{R}(t)}{R(t)}\right)\mathbf{x}(t), \tag{1.5}$$

where we have plugged equation 1.4 back in. The above equation gives a velocity proportional to distance similar to Hubble's law. Writing this relation for two points 1 and 2 and subtracting shows that this expansion appears the same for any choice of origin. In other words, everywhere is the centre of the Universe:

$$\dot{\mathbf{x}}_2(t) - \dot{\mathbf{x}}_1(t) = \left(\frac{\dot{R}(t)}{R(t)}\right)[\mathbf{x}_2(t) - \mathbf{x}_1(t)]. \tag{1.6}$$

From this relation we can relate the Hubble parameter $H(t)$ with $\dot{R}(t)/R(t)$, and we can see that in general it is not a constant. Moreover, the above explanation will always yield Hubble's law, regardless of what the rate of the expansion is. The scale factor is often made dimensionless as

$$a(t) = \frac{R(t)}{R_0}, \tag{1.7}$$

where $R_0$ is the scale factor at some time $t_0$, generally taken to be present time, such that $a_0 = 1$. As such, the Hubble parameter can also be written as $H(t) = \dot{a}(t)/a(t)$. The scale factor relates the proper distance between two objects $d(t)$, which changes due to their relatives velocities and the expansion of the Universe, to their comoving distance $r(t)$, which only changes due to the objects' relative velocities (the expansion of the Universe has been 'filtered out'), i.e. $d(t) = a(t)r(t)$.

### 1.2.1 The Hot Big Bang model

As stated earlier, the Hubble constant describes the rate of expansion of the Universe at a given time; it is time-dependent. If at present time galaxies are moving

away from each other, it follows that they were closer together in the past. Let us consider a pair of galaxies which are currently separated by a distance $r$ and have a velocity $v$ relative to each other given by Hubble's Law (equation 1.1). If there are no forces acting on the galaxies to accelerate or decelerate their relative motion, then their velocity is constant, and we can calculate the time since they were in contact:

$$t = \frac{r}{v} = \frac{r}{H_0 r} = H_0^{-1}, \tag{1.8}$$

which is independent of the present separation $r$. The time $H_0^{-1}$ is referred to as the Hubble time. In other words, if the relative velocities of galaxies is constant, they must have all been crammed close together in a small volume a time $t = H_0^{-1}$ ago. The observation that the Universe is expanding naturally leads to a *Big Bang* model for the evolution of the Universe. Generally speaking, a Big Bang model is a model in which the Universe expands from an initially highly dense state to its current low-density state.

However, the rate of expansion of the Universe is not constant. We can calculate $H(t)$ from its current value and the contents of the Universe, using solutions to the equations of GR for an isotropic and homogeneous universe. This leads to a Universe that had an *infinite* temperature and density at a *finite* time in the past. This model for the early Universe is known as the *Hot Big Bang* model.

### Observational evidence

Many cosmological observations can be explained by the Hot Big Bang model. In 1965, an isotropic microwave radiation filling all space was discovered, corresponding to what would be emitted by a body with a temperature of ∼3.5K (Penzias & Wilson, 1965). This radiation is generally referred to as the cosmic microwave background (CMB) radiation. If the evolution of the Universe can be described by the Hot Big Bang model, the CMB can be explained as follows. At the early

stages of the Universe, it was filled with a fog of protons and electrons. At this time any radiation was efficiently scattered by the free electrons, and the Universe was opaque to photons. The plasma cooled down with the expansion of the Universe and the radiation was redshifted to longer wavelengths. When the temperature had dropped enough, protons and electrons were now able to combine into neutral hydrogen atoms. This epoch was named *recombination* and occurred at a redshift of $z \approx 1100$ (see end of section 1.1). Note that the name is a misnomer as these particles had not combined before this time. Unlike the free electrons, these new atoms could not scatter the thermal radiation, and so the Universe became transparent to photons. These photons have been propagating ever since (decreasing in energy due to their wavelengths being redshifted) and are the source of the CMB. This natural explanation for the remnant radiation provides strong evidence for the Hot Big Bang model.

Another cosmological observation that could be explained by the Hot Big Bang model is the nucleosynthesis of light elements. The first version of the Big Bang nucleosynthesis theory was proposed by George Gamow and Ralph Alpher in the 1940s* (Alpher, Bethe & Gamow, 1948). As the Universe cooled until the temperature was lower than the nuclear binding energy, protons and neutrons were able to combine into atomic nuclei. At first the protons and neutrons were in thermal equilibrium, taking part in reactions like $n + \nu_e \leftrightarrow p + e$ and $n + e^+ \leftrightarrow p + \overline{\nu}_e$, mediated by the weak interaction. As the temperature dropped further, the equilibrium shifted in favour of the protons due to their slightly lower mass, causing the proton to neutron ratio to increase. These reactions continued until the decreasing temperature and density caused the reactions to become too slow and the abundance of the particles to remain 'frozen' at their last values. This process is generally referred to as *freeze out*. After freeze out, the proton to neutron ratio was approximately 6:1.

*They thought it to be the origin of *all* chemical elements. We now know that elements heavier than Helium are formed in the interior of stars, i.e. through *stellar* nucleosynthesis.

Baryons and light elements then fused to form heavier nuclei, with most fusion chains ultimately ending with Helium-4, while 'incomplete' reaction chains lead to small amounts of left-over Deuterium or Helium-3. The amount of these decreases with increasing baryon to photon ratio, which is proportional to the baryon density $\Omega_b$. The larger the baryon to photon ratio the more reactions there will be and the more efficiently Deuterium will be eventually transformed into Helium-4. The abundances of the various elements depend on $\Omega_b$ in different ways (as the reaction rates do), and it is therefore not immediately obvious that Big Bang nucleosynthesis would predict the observed values. The fact that a single value for $\Omega_b$ can simultaneously reproduce all the observed values, strongly supports the Hot Big Bang model.

### 1.2.2 The geometry and expansion history of the Universe

The Friedmann-Lemaître-Robertson-Walker (FLRW) is a metric based on the exact solution of Einstein's field equations of GR. The FLRW model describes a homogeneous, isotropic, expanding universe. The mathematician Alexander Friedmann first derived the main results of the FLRW model in 1922 (Friedmann, 1922). After Friedmann's death in 1925, George Lemaître independently developed a similar model in 1927 (Lemaître, 1927), and was one of the first people to suggest that the Universe began with a Big Bang. Howard P. Robertson and Arthur Geoffrey Walker modified and developed the model further during the 1930s, resulting in what we know today as the FLRW metric. Mathematically, this metric is described by the space-time line element $ds$

$$(ds)^2 = (cdt)^2 - a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta^2 d\phi^2)\right), \tag{1.9}$$

where $c$ is the speed of light, and $a(t)$ is the dimensionless scale factor of the Universe defined in section 1.4. The constant $k$ in equation 1.9 describes the geometry of the Universe, often referred to as the 'curvature'. $k < 0$ corresponds

to an open/hyperbolic Universe, $k = 0$ to a flat Universe, and finally $k > 0$ to a closed/spherical Universe. Note that equation 1.9 is written in comoving (spherical) coordinates.

In the limit of small velocities, GR needs to reduce to Newtonian mechanics. As such, we can study the expansion history of the Universe on a small scale where Newtonian mechanics should apply and then from homogeneity we can say that the results must apply to larger scales and on the scale of the Universe as a whole. The equations used to describe the evolution of the Universe are derived in the context of GR. However, the results are nearly identical to when taking a Newtonian approach. Let us consider a sphere of matter at time $t$ of radius $R$ and with mass $M$. The surface of the sphere expands within a homogeneous Universe at rate $\dot{R}$, where the dot refers to a derivative in time. Due to the gravitational pull of the matter inside the shell, the edge of the shell is decelerated while any symmetric mass distribution outside of the sphere produces zero acceleration from Newton's shell theorem*. This implies that the shell decelerates at a rate

$$\ddot{R} = -\frac{GM}{R^2} = -\frac{4\pi G R^3 \rho}{3R^2} = -\frac{4\pi G \rho R}{3}. \tag{1.10}$$

As mass inside the sphere is conserved, the density scales with $1/R(t)^3$, and so we must have the relation

$$\rho = \rho_0 \left(\frac{R_0}{R}\right)^3, \tag{1.11}$$

where $\rho_0$ is the density when $R = R_0$. Multiplying equation 1.10 by $2\dot{R}$ and plugging in equation 1.11, we get

$$2\dot{R}\ddot{R} = -\frac{8\pi G \rho_0 \dot{R} R_0^3}{3R^2}. \tag{1.12}$$

Integrating with respect to time in turn gets us

*In GR, the corresponding theorem is Birkhoff's theorem.

$$\dot{R}^2 = \frac{8\pi G\rho_0 R_0^3}{3R} + \text{constant}, \tag{1.13}$$

and with some more rearranging, we find

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3} + \frac{\text{constant}}{R^2}. \tag{1.14}$$

When considering the expansion of the Universe, is is useful to replace the radius of the sphere $R(t)$ with the cosmic (dimensionless) scale factor using equation 1.7. Had we gone through the proper GR calculations, we would have ended up with an additional constant term in equation 1.14, known as the cosmological constant, $\Lambda$. Including all this, we therefore arrive at the first *Friedmann* equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}, \tag{1.15}$$

where $k$ is the curvature constant we encountered earlier in this section. Often, the density $\rho$ is replaced with $\rho - \frac{\Lambda c^2}{8\pi G}$ such that the Friedmann equation simplifies to

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{kc^2}{a^2}. \tag{1.16}$$

Noting that $\dot{a}/a$ is the Hubble parameter, from this equation we can identify the *critical* density of the Universe, which is the density required for the Universe to have an exactly flat geometry with $k = 0$:

$$\rho_c = \frac{3H^2}{8\pi G}. \tag{1.17}$$

As space expands, distance increases as $\propto a(t)$, and so in turn volume increases as $\propto a^3(t)$. For ordinary matter then, density $\rho_m$ decreases as $\propto 1/a^3(t)$. As such we can write the evolution of matter density with the expansion of the Universe as

$$\rho_m = \rho_{m,0}/a^3, \tag{1.18}$$

where $\rho_{m,0}$ is the value of the matter density at present time. In the case of radiation density, we need an extra factor of $1/a(t)$ as in addition to the volume changing with the expansion, the wavelength of light is redshifted as $\propto a(t)$ and the energy is reduced as $\propto 1/a(t)$ from $E = hc/\lambda$. The evolution of radiation density with the expansion of the Universe is therefore

$$\rho_r = \rho_{r,0}/a^4, \tag{1.19}$$

where $\rho_{r,0}$ is the present-day value of the radiation density. Considering that these two expression have different dependencies on $a(t)$, we can see that at a point in the past the Universe must have switched from being radiation-dominated to being matter-dominated. The cosmological constant is related to the energy density of space, or vacuum density $\rho_v$. As the name suggests, it is a constant and is not dependent on $a(t)$ ($\rho_v = \rho_{v,0}$). We can relate the curvature to an energy density as well, which we can see from equation 1.15 must have a dependency on $1/a(t)^2$. With this in mind, we can rewrite the Friedmann equation (equation 1.15) as

$$H^2(a) = \frac{8\pi G}{3}\left(\rho_{m,0}/a^3 + \rho_{r,0}/a^4 + \rho_{k,0}/a^2 + \rho_{v,0}\right). \tag{1.20}$$

At present time, the critical density (equation 1.17) equals to $\rho_{c,0} = 3H_0^2/8\pi G$, and so equation 1.20 can also be written as

$$H^2(a) = \frac{H_0^2}{\rho_{c,0}}\left(\rho_{m,0}/a^3 + \rho_{r,0}/a^4 + \rho_{k,0}/a^2 + \rho_{v,0}\right). \tag{1.21}$$

We define the density parameter as the ratio of density to critical density, $\Omega_{\times,0} = \rho_{\times,0}/\rho_{c,0}$, where $\times$ represents $m$, $r$, $k$, or $v$*. We can now write the Friedmann equation as

$$H^2 = H_0^2\left(\Omega_{m,0}/a^3 + \Omega_{r,0}/a^4 + \Omega_{k,0}/a^2 + \Omega_{v,0}\right). \tag{1.22}$$

*Note that $\Omega_{k,0} + \Omega_{m,0} + \Omega_{r,0} + \Omega_{v,0} = 1$

We can solve the Friedmann equation in a few simple cases. Consider a matter-dominated flat universe with no cosmological constant ($\Omega_{m,0} = 1$, $\Omega_{r,0} = \Omega_{k,0} = \Omega_{v,0} = 0$). In this case the equation 1.22 reduces to

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{H_0^2}{a^3}. \tag{1.23}$$

And so we find $\dot{a} \propto a^{-\frac{1}{2}}$. Integrating this relation gives us $a(t) = (t/t_0)^{\frac{2}{3}}$, where $t_0 = \frac{2}{3H_0}$ is the age of the Universe when $a(t) = 1$. This special case is called an Einstein deSitter universe, named after Einstein and the astronomer Willem de Sitter who together proposed the model in 1932 (Einstein & de Sitter, 1932).

Similarly for a radiation-dominated flat universe with no cosmological constant ($\Omega_{r,0} = 1$, $\Omega_{m,0} = \Omega_{k,0} = \Omega_{v,0} = 0$), we have

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{H_0^2}{a^4}. \tag{1.24}$$

Resulting in $a(t) = (t/t_0)^{\frac{1}{2}}$ with $t_0 = \frac{1}{2H_0}$. Finally for a flat universe with only a cosmological constant, we find from equation 1.15

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{\Lambda c^2}{3} \tag{1.25}$$

and $a(t) \propto \exp(Ht) = \exp(\sqrt{\Lambda/3}t)$. This last result is particularly interesting as $a$ approaches zero when $t$ approaches minus infinity. In other words there is no Big Bang singularity. This model, proposed in 1917, is referred to as a deSitter universe, again named after Willem de Sitter.

### Cosmological Redshift

An observational consequence of the expansion of the Universe is that light waves are 'redshifted'; the *observed* wavelength has increased with respect to the *emitted* wavelength. To derive the mathematical expression for cosmological redshift, let

us consider two photons, representing two wave crests of a light-ray. The first is emitted from a source at time $t_{\rm em}$, the second at $t_{\rm em} + \Delta t_{\rm em}$. These two photons arrive at an observer at time $t_{\rm obs}$ and $t_{\rm obs} + \Delta t_{\rm obs}$ respectively. To simplify our calculations we can choose our axes such that the photons move radially with $d\theta = 0$ and $d\phi = 0$. Then the line element defined by equation 1.9 reduces to $(ds)^2 = (cdt)^2 - a^2(t)dr^2/(1 - kr^2)$. Additionally, we recount that photons move along null geodesics, i.e. $ds = 0$, and so we find $cdt = a(t)dr/\sqrt{1 - kr^2}$.

$r_{\rm obs}$ is the total (comoving) distance travelled by the first photon when it is observed at time $t_{\rm obs}$, similarly $r_{\rm em}$ is the distance travelled at time of emission $t_{\rm em}$, i.e. $r_{\rm em} = 0$. Let $r_{\rm obs} = r_1$ for the first photon. We can integrate along the light ray, to find

$$c\int_{t_{\rm em}}^{t_{\rm obs}} \frac{dt}{a(t)} = \int_0^{r_1} \frac{dr}{\sqrt{1 - kr^2}}. \tag{1.26}$$

Photons that are emitted at later times will be received at later times, but changes in $t_{\rm em}$ and $t_{\rm obs}$ cannot alter the integral on the right hand side of equation 1.26, since $r$ is a comoving quantity. So for the second photon we must have $r_{\rm obs} = r_1$ as well, and we find

$$c\int_{t_{\rm em}+\Delta t_{\rm em}}^{t_{\rm obs}+\Delta t_{\rm obs}} \frac{dt}{a(t)} = \int_0^{r_1} \frac{dr}{\sqrt{1 - kr^2}}. \tag{1.27}$$

Noting that the right hand sides of the two above equations are equal and given the relation

$$\int_{t_{\rm em}+\Delta t_{\rm em}}^{t_{\rm obs}+\Delta t_{\rm obs}} = \int_{t_{\rm em}}^{t_{\rm obs}} + \int_{t_{\rm obs}}^{t_{\rm obs}+\Delta t_{\rm obs}} - \int_{t_{\rm em}}^{t_{\rm em}+\Delta t_{\rm em}}, \tag{1.28}$$

we find

$$\int_{t_{\rm obs}}^{t_{\rm obs}+\Delta t_{\rm obs}} \frac{dt}{a(t)} = \int_{t_{\rm em}}^{t_{\rm em}+\Delta t_{\rm em}} \frac{dt}{a(t)}. \tag{1.29}$$

For sufficiently small $\Delta t_{\rm obs}$ and $\Delta t_{\rm em}$, we can assume $a(t)$ does not change significantly and treat it as a constant. Solving the integral, we find

$$\frac{\Delta t_{\rm obs}}{a(t_{\rm obs})} = \frac{\Delta t_{\rm em}}{a(t_{\rm em})} \Longleftrightarrow \frac{a(t_{\rm obs})}{a(t_{\rm em})} = \frac{\Delta t_{\rm obs}}{\Delta t_{\rm em}}. \tag{1.30}$$

$\Delta t_{\rm em}$ and $\Delta t_{\rm obs}$ are the time between the emitted and observed wave crests, and so they can be rewritten in terms of the wavelengths of the photons, i.e. $\Delta t_{\rm em} = \lambda_{\rm em}/c$ and $\Delta t_{\rm obs} = \lambda_{\rm obs}/c$. Plugging this into equation 1.30, we find

$$\frac{a(t_{\rm obs})}{a(t_{\rm em})} = \frac{\lambda_{\rm obs}}{\lambda_{\rm em}}. \tag{1.31}$$

Using the definition of redshift as the fractional change in a photon's wavelength (equation 1.2) and equation 1.31 together and taking $t_{\rm obs}$ to be present time ($a(t_{\rm obs}) = 1$), we find the definition for cosmological redshift

$$1 + z = \frac{1}{a(t_{\rm em})}. \tag{1.32}$$

The redshift of a light emitting source can be determined, e.g., by comparing the wavelengths of the spectral lines of the source to their known values in the laboratory.

### 1.2.3 Structure formation

While on sufficiently large scales the Universe is homogeneous and isotropic, it is quite clear that on small scales this is not the case: matter clusters in galaxies, which in turn cluster in groups, clusters and super-clusters. It is thought that all structure at present time grew gravitationally from quantum fluctuations in the early Universe to the macroscopic fluctuations we see today.

To derive how perturbations grow in a self-gravitating fluid, we need to solve the continuity, Euler, and Poisson equations. They are respectively

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \tag{1.33}$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{1}{\rho}\nabla p - \nabla \Phi, \tag{1.34}$$

and

$$\nabla^2 \Phi = 4\pi G \rho, \tag{1.35}$$

where $\rho(\mathbf{r}, t)$, $\mathbf{v}(\mathbf{r}, t)$, $p(\mathbf{r}, t)$, $\Phi(\mathbf{r}, t)$ are the density, velocity, pressure and gravitational potential of the fluid respectively. Quantities in bold represent vectors. The continuity and Euler equation describe the conservation of mass and momentum, while the Poisson equation describes how matter is the source of gravitational fields.

Equations 1.33, 1.34, and 1.35 were written in physical coordinates $\mathbf{r}$. If we wish to take into account the expansion of the Universe, we can rewrite these equations in comoving coordinates by introducing the co-moving position $\mathbf{x}$, the co-moving velocity $\mathbf{v}$, and the peculiar velocity $\mathbf{v}_p$ as

$$\mathbf{r} = a(t)\mathbf{x}, \tag{1.36}$$

and

$$\mathbf{v} = \dot{\mathbf{r}} = \dot{a}\mathbf{x} + a\dot{\mathbf{x}} \equiv \dot{a}\mathbf{x} + \mathbf{v}_\mathrm{p}. \tag{1.37}$$

The partial derivative $\partial/\partial t$ in equations equations 1.33, 1.34, and 1.35 are derivatives with respect to $t$ at constant $\mathbf{r}$. To rewrite the equations in terms of $\mathbf{x}$, the time derivatives should be expressed at constant $\mathbf{x}$, and so we make the following substitution*

*Consider density $\rho$. In time interval $\Delta t$, $\rho$ changes by: $\Delta\rho|_\mathbf{r} = \Delta\rho|_\mathbf{x} + \frac{\Delta\rho}{\Delta\mathbf{x}}\Delta\mathbf{x}$. $\mathbf{r}$ is constant, so $\Delta\mathbf{r} = \Delta(a\mathbf{x}) = \mathbf{x}\Delta a + a\Delta\mathbf{x} = 0$, giving us $\Delta\mathbf{x} = -(\Delta a/a)\mathbf{x}$. Taking $\Delta t$ to be an infinitesimal change in time $dt$, we then find $\partial\rho/\partial t|_\mathbf{r} = \partial\rho/\partial t|_\mathbf{x} - (da/adt)(\mathbf{x}\cdot\nabla_\mathbf{x})\rho = \partial\rho/\partial t|_\mathbf{x} - (H\mathbf{x}\cdot\nabla_\mathbf{x})\rho$.

$$\left.\frac{\partial}{\partial t}\right|_{\mathbf{r}} \rightarrow \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}} - (H\mathbf{x}\cdot\nabla_{\mathbf{x}}), \tag{1.38}$$

where $\nabla_{\mathbf{x}} = \partial/\partial\mathbf{x}$. Hence $\nabla_{\mathbf{r}} = (1/a)\nabla_{\mathbf{x}}$, and $\nabla^2_{\mathbf{r}} = (1/a^2)\nabla^2_{\mathbf{x}}$. Let us also define the *over-density* field, $\delta(\mathbf{x}, t)$, which reflects the deviations from the average (homogeneous) density of the Universe $\overline{\rho}(t)$, such that

$$\rho(\mathbf{x}, t) = \overline{\rho}(t)[1 + \delta(\mathbf{x}, t)]. \tag{1.39}$$

Making all these substitutions, we find the following comoving equivalents of equations 1.33, 1.34, and 1.35 respectively

$$\left.\frac{\partial\delta}{\partial t}\right|_{\mathbf{x}} + \frac{1}{a}\nabla_{\mathbf{x}}\cdot[(1+\delta)\mathbf{v}_p] = 0, \tag{1.40}$$

$$\left.\frac{\partial\mathbf{v}_p}{\partial t}\right|_{\mathbf{x}} + \frac{1}{a}(\mathbf{v}_p\cdot\nabla_{\mathbf{x}})\mathbf{v}_p + H\mathbf{v}_p = -\frac{1}{a}\nabla_{\mathbf{x}}\Phi_{\mathrm{p}} - \frac{c_s^2}{a}\frac{\nabla_{\mathbf{x}}\rho}{\rho}, \tag{1.41}$$

and

$$\frac{1}{a^2}\nabla^2_{\mathbf{x}}\Phi_{\mathrm{p}} = 4\pi G\overline{\rho}\delta. \tag{1.42}$$

We have introduced two new variables in the above equations, the adiabatic sound speed $c_s$ given by $c_s = \left(\frac{\partial p}{\partial\rho}\right)^{1/2}$ with $p$ being the pressure of the fluid, and the *peculiar gravitational potential* $\Phi_{\mathrm{p}}$, given by $\Phi_{\mathrm{p}} = \Phi - \Phi_0$. $\Phi_{\mathrm{p}}$ reflects the fluctuations in the potential about the homogeneous solution $\Phi_0$, i.e. when $\delta(\mathbf{x}, t) = 0$ and $\rho(t) = \overline{\rho}(t)$.

Let us look at the behaviour of small (linear) perturbations with $\delta \ll 1$ and $\mathbf{v}_p \ll 1$. The continuity and Euler equations then reduce to

$$\left.\frac{\partial\delta}{\partial t}\right|_{\mathbf{x}} + \frac{1}{a}\nabla_{\mathbf{x}}\cdot\mathbf{v}_p = 0, \tag{1.43}$$

and

$$\left.\frac{\partial \mathbf{v}_p}{\partial t}\right|_{\mathbf{x}} + H\mathbf{v}_p = -\frac{1}{a}\nabla_{\mathbf{x}}\Phi_{\mathrm{p}} - \frac{c_s^2}{a}\nabla_{\mathbf{x}}\delta, \tag{1.44}$$

where have dropped any terms that are second order in $\delta$ or $\mathbf{v}_p$. Taking the time derivative of equation 1.43 and multiplying equation 1.44 with $\frac{1}{a}\nabla_{\mathbf{x}}\cdot$ gives us

$$\left.\frac{\partial^2 \delta}{\partial t^2}\right|_{\mathbf{x}} - \frac{\dot{a}}{a^2}\nabla_{\mathbf{x}}\cdot\mathbf{v}_p + \frac{1}{a}\nabla_{\mathbf{x}}\cdot\frac{\partial \mathbf{v}_p}{\partial t} = 0, \tag{1.45}$$

and

$$\frac{1}{a}\nabla_{\mathbf{x}}\cdot\frac{\partial \mathbf{v}_p}{\partial t} + \frac{H}{a}\nabla_{\mathbf{x}}\cdot\mathbf{v}_p = -\frac{1}{a^2}\nabla_{\mathbf{x}}^2\Phi_{\mathrm{p}} - \frac{c_s^2}{a^2}\nabla_{\mathbf{x}}^2\delta. \tag{1.46}$$

Noting that $H = \dot{a}/a$, adding the above equations and substituting equations 1.42 and 1.43 finally gives us

$$\left.\frac{\partial^2 \delta}{\partial t^2}\right|_{\mathbf{x}} + 2H\left.\frac{\partial \delta}{\partial t}\right|_{\mathbf{x}} = (4\pi G\overline{\rho} + \frac{c_s^2}{a^2}\nabla_{\mathbf{x}}^2)\delta. \tag{1.47}$$

Let us look at the Einstein de Sitter universe (see end of section 1.2.2) with only pressure-less ($c_s = 0$) matter and $a \propto t^{2/3}$, a useful counterfactual but not a realistic reflection of the Universe. In this case, we find $H = 2/3t$. Then, from equation 1.15, we must have $8\pi G\overline{\rho}/3 = 4/9t^2$ giving us $4\pi G\overline{\rho} = 2/3t^2$. All this together gives us a differential equation that we can solve easily

$$\left.\frac{\partial^2 \delta}{\partial t^2}\right|_{\mathbf{x}} + \frac{4}{3t}\left.\frac{\partial \delta}{\partial t}\right|_{\mathbf{x}} - \frac{2}{3t^2}\delta = 0. \tag{1.48}$$

Starting with the *Ansatz* $\delta = At^n$, we find $n = 2/3$ or $n = -1$. We ignore the second solution as it corresponds to a decaying mode, and we are interested in structure growth. The first solution, however, gives us $\delta \propto t^{2/3} \propto a$. As long as the perturbations are linear, they grow proportional to the scale factor.

Going through similar calculations for an open universe without a cosmological constant (again just a counterfactual), we find

$$\left.\frac{\partial^2\delta}{\partial t^2}\right|_{\mathbf{x}} + \frac{2}{t}\left.\frac{\partial\delta}{\partial t}\right|_{\mathbf{x}} = 0. \tag{1.49}$$

With the same Ansatz as before, we now get $n = 0$ or $n = -1$. The 'growing' mode is constant in time, i.e. due to the low matter density perturbations have stopped growing.

The CMB radiation is highly isotropic, with the amplitude of the typical density fluctuations being of the order of $\delta \sim 10^{-5}$ (Smoot et al., 1992; Bennett et al., 1996). Since radiation decoupled from matter at the epoch of recombination, the CMB anisotropy reflects the level of inhomogeneities in that early period of Universe. Using equation 1.32, at $z = 1100$ the scale factor was of the order of $a \approx 0.001$, or about a thousand times smaller than its current value of 1. Considering that for a matter dominated universe perturbations grow with the scale factor, by present time these perturbations should now be of the order of $10^{-5} \times 1000 = 10^{-2}$.

While we made some simplifications in calculating how $\delta$ grows, the above results seem to suggest that we should not expect any non-linear structures with $\delta \gg 1$ in the Universe today. This is clearly not the case. A possible solution to this discrepancy is that we are missing a component of matter that does not couple to radiation or ordinary matter. Therefore, its density perturbations can grow before those of the ordinary matter. As a consequence, its gravitational potential can act as an potential well for the ordinary matter when it collapses later, thus speeding up the structure formation process and allowing for large structures at present time. Before the fluctuations in the CMB were measured, other cosmological observations had already seemed to suggest the existence of another matter component. This new type of matter is what we now refer to as *dark matter* (DM).

## 1.3 Dark matter

Currently, it is thought that most of the matter in the Universe is not baryonic, with approximately 85% of the Universe's matter content being DM (Planck Collaboration et al., 2020). The simplest model of DM is the cold dark matter (CDM) model, where the *cold* refers to the fact that the DM moved slowly relative to the speed of light in the early Universe. The CDM model asserts that DM does not* emit, absorb, or reflect light (hence 'dark'), and so the only way to study it is through its gravitational influence.

### 1.3.1 Observational evidence for dark matter

Some of the first observational evidence for the existence of DM was obtained by astronomer Fritz Zwicky in 1933. Using the virial theorem, he calculated the mass of the galaxies in the Coma cluster based on the observed rotational velocities of the galaxies. He obtained a value almost 400 times higher than the mass inferred from just the luminous matter (Zwicky, 1933). He named this discrepancy in mass dunkele materie (German for dark matter). While his calculations were not entirely correct, present-day calculations agree that the majority of the mass in the Coma cluster is indeed made up of DM.

Another key piece of observational evidence for DM is that many spiral galaxies show flat rotation curves. A rotation curve is the radial velocity of matter in galaxies as a function of its distance from the galactic centre. Assuming a Newtonian spherically symmetric model, the rotation speed for circular orbits $V$ is given by

$$V = \frac{GM(< r)}{r}, \tag{1.50}$$

where $M(< r)$ is the mass enclosed in radius $r$, and $G$ is the gravitational constant. Looking only at the luminous matter in a galaxy, there is a radius $r$ beyond which

*Or very rarely...

there are no more stars to be found. In other words, $M(< r)$ should be constant at radii larger than $r$, and the velocity curve should drop as $1/r$. In the 1970s, however, Vera Rubin and Kent Ford obtained velocity curves for various edge-on spiral galaxies, and found that the curves remained flat as the radius increased (Rubin & Ford, 1970), see figure 1.1. Assuming that Newtonian mechanics is correct (which is true on small scales), the obvious way to resolve this discrepancy is to conclude that there is a large amount of non-luminous matter, i.e. DM, in the outskirts of the galaxies.

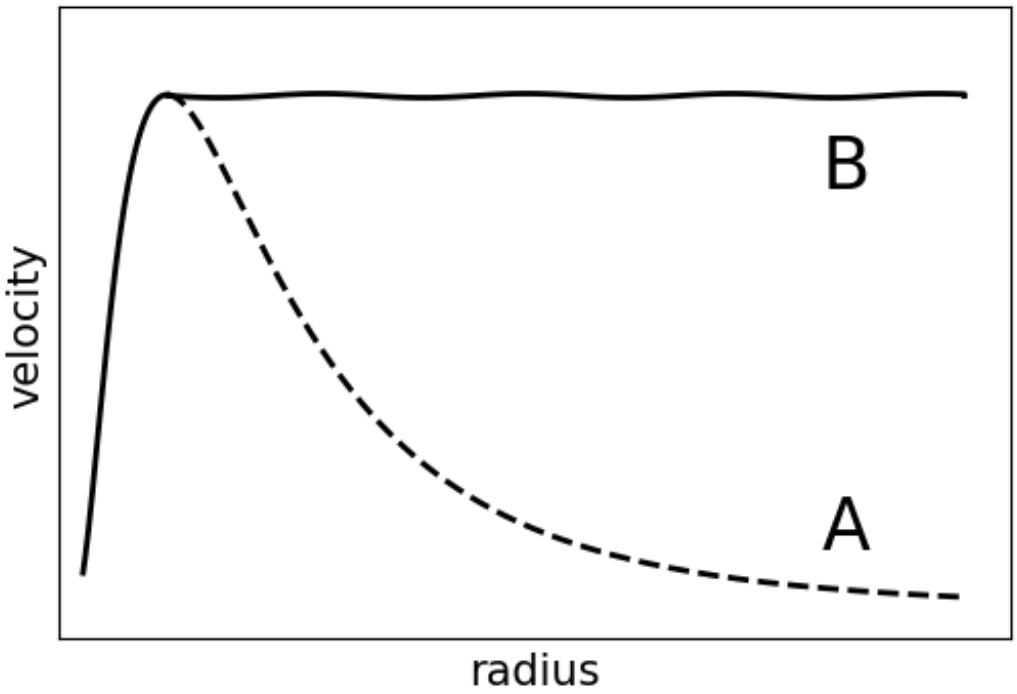


Figure 1.1: The predicted (A) and observed (B) rotation curve of a typical spiral galaxy. Credit: E. Sirks

Other evidence for DM comes from studying the images of distant galaxies. GR predicts that the presence of mass density bends, or 'lenses', rays of light*. This phenomenon is called *gravitational lensing*. As a result of gravitational lensing, the images of sources behind the intervening mass appear distorted. If the lensing causes visible distortions such as multiple images, arcs or Einstein rings, it is

*Gravitational lensing is a prediction of classical mechanics and Newtonian gravity as well, however, the magnitudes of the deflections are about twice as large in GR.

referred to as *strong* gravitational lensing. When this lensing effect is not strong enough to produce multiple images, it is referred to as *weak* gravitational lensing. In the case of weak lensing, the presence of mass can still be inferred due to the statistical alignment of multiple background sources. Gravitational lensing is a powerful tool in studying DM as it does not depend on the dynamical state of the matter acting as the lens, but only on the distribution of the total mass (for a review of gravitational lensing see Bartelmann, 2010). Lensing measurements confirm the existence of large amounts of DM in galaxies as well as in clusters of galaxies (Massey et al., 2010).

Not just gravitationally bound structures lens the light from background sources. The large scale structure of the Universe itself acts as a gravitational lens. The distortion of the images of background galaxies due to the (weak) gravitational lensing from the large-scale structure is called *cosmic shear*. Cosmic shear essentially measures the clustering of galaxies in the Universe. The observed large-scale structure (e.g. Rodríguez-Torres et al., 2016) is described extremely well by the structure predicted when assuming that the matter in the Universe is dominated by collisionless DM (e.g. Springel et al., 2006).

Instead of the existence of a new type of matter to explain the various observations mentioned above, another possibility is that our understanding of gravity is not correct. Modified Newtonian dynamics (MOND) is an alternative to the hypothesis of DM in terms of explaining why galaxies do not appear to obey the currently understood laws of physics (Milgrom, 1983). MOND was initially proposed as a way of explaining the flat rotation curves of galaxies by proposing a new effective gravitational force law. Essentially, at high accelerations this force law reduces to the Newtonian version, but at low accelerations MOND leads to different behaviours.

### 1.3.2 Dark matter detection experiments

It is often assumed that the constituents of DM could be new elementary particles. The particle DM hypothesis can be tested via three processes: directly via scattering on target nuclei, indirectly by searching for signals from DM annihilation or decay products, or through production at particle accelerators, where DM (occasionally) breaks the assumptions of CDM.

Weakly interacting massive particles (WIMPs) in the GeV-$\mathcal{O}(100\,\mathrm{TeV})$ mass range are hypothetical particles that are one of the proposed candidates for DM, and, as the name suggests, weakly interact. Generally detection experiments focus on finding signatures of WIMPs. The main reason for the popularity of the WIMP is that thermal WIMP production in the big bang, whose processes are well gauged by the observations of light elements (see section 1.2.1), predict a global DM abundance within one dex of the observed one (e.g. Jungman et al., 1996). This result is often called the 'WIMP miracle'. There are searches for other DM particles that are not WIMPs, such as axions.

#### Direct detection

A variety of experiments have been developed over the past decades aiming to detect DM particles via their scattering in particle detectors, i.e. reactions of the type $\chi P \rightarrow \chi P$, where $\chi$ is a DM particle and $P$ a standard model particle. Essentially, direct detection DM experiments aim to measure the energy deposited when WIMPs interact with nuclei in a detector, transferring some of their energy to the nuclei. Cooled crystals can be used to detect elastic collisions between detector nuclei and DM particles as minute crystal lattice vibrations (phonons) and ionisation (charge). In noble liquid detectors, interactions of the DM with the liquid lead to scintillation.

Cosmic rays, $\alpha$-particles, electrons, and photons can scatter from atomic nuclei mimicking the DM signal. Usually, experiments operate deep underground in order

to reduce background to the signal from cosmic rays as they are efficiently stopped by the ground above. In order to prevent the other secondary interactions, various materials are used to encase the detectors and stop the particles from entering.

While there has been no detection as of yet, experiments have set upper limits to the mass of WIMPs. Some of the tightest constraints to the cross-section of WIMP-nucleon interactions come from XENON1T (Aprile et al., 2017) and LUX (Akerib et al., 2013), both of which use liquid xenon as their detector material. Successors to each experiment are currently in development, named LUX-ZEPLIN (The LZ Collaboration et al., 2019) and XENONnT (Aprile et al., 2020) respectively.

### Indirect detection

Indirection detection experiments aim to detect DM particles through their annihilation or decay to standard model particles, and in particular gamma rays, charged leptons and neutrinos. DM annihilations are reactions of the type $\chi\overline{\chi} \rightarrow xx$, where $\overline{\chi}$ is a DM anti-particle, and $xx$ could be a pair of quarks, or $W/Z$ bosons, the gauge bosons that mediate the weak interaction. Subsequent hadronisation and pion decay of these particles could then yield gamma-rays*. There are a number of channels along which a DM particle theoretically could decay, which include (but are not limited to) $\chi \rightarrow \ell^+\ell^-\nu, Z^0\nu, W^\pm\ell^\mp$, where $\ell$ is a lepton, and $\nu$ a neutrino.

In the energy range between about ∼100 MeV and several 100 GeV, gamma rays can be observed by pair-conversion telescopes on satellites. Such telescopes can detect gamma rays through the generation of electron-positron pairs in the material of the instrument. One of the larger still operational pair-conversion telescopes is the Fermi Large Area Telescope (FermiLAT, Atwood et al., 2009). As annihilation is proportional to DM density squared, the best chances of observing a signal would be to point telescopes at regions where we expect a high DM density. Above 100 GeV, Imaging Air Cherenkov Telescopes (IACT) become more sensitive, such

*Direct annihilation to gamma rays is also possible.

as e.g. VERITAS (Holder et al., 2008). IACTs work by imaging the very short flash of Cherenkov radiation generated by the cascade of relativistic charged particles produced when high energy gamma rays strike the atmosphere.

The DM density should be high near the centre of the MW, and there have been observed excesses in the FermiLAT data around this region (e.g. Bringmann et al., 2012). However, as of yet it is not clear if these are gamma rays from DM annihilation/decay, or other astrophysical sources, e.g. pulsars.

#### Collider searches

Theoretically, DM particles could be produced in a laboratory. DM particles produced in collisions of proton beams in particle accelerators, such as the Large Hadron Collider (LHC), may be detected. These reactions are of the type $pp \rightarrow \chi\chi + x$, where $p$ is a proton, and $x$ represents either a jet of hadrons, a photon, or a $Z/W$ boson decaying into leptons. While DM particles do not interact with ordinary matter, its presence could be inferred from the missing energy and momentum in these detectors when other collision products have been detected. To proof that the unaccounted for energy and momentum are in fact DM particles, discoveries from direct or indirect detection experiments are required. These searches are mostly sensitive to WIMPs with masses of the order ∼GeV.

As of yet, there has been no suggestion of DM particles having been produced in collider experiments, and only upper limits have been placed on the cross-section of the interactions of DM with ordinary matter.

### 1.3.3 The standard model of cosmology: ΛCDM

The current standard model of cosmology is referred to ΛCDM, and provides the current best description of the expansion history and the large-scale structure features of the Universe. In this model, the Universe contains three major components:

dark energy, CDM, and ordinary matter. In this model, the DM is *collisionless*: it only interacts with other DM particles through the gravitational force.

### Dark energy

As mentioned at the beginning of section 1.2, in 1988 it was discovered that not only is the Universe expanding, it is expanding at an *accelerated* rate. The first piece of evidence supporting the accelerated expansion came from observations of type Ia supernovae (SNe Ia) by Riess et al. (1998). The observed SNe Ia are at larger distances than predicted based on the assumption of a universe with a constant expansion rate.

In a universe dominated by matter, the net gravitational pull should in principle slow the expansion down instead of accelerating it. Therefore, it was theorised that there must be another type of energetic component (aside from radiation) that exerts a pressure opposing gravity and affects the Universe on the largest scales. The nature of this energy is unknown and is referred to as *dark energy*. There are various proposed forms of dark energy.

We encountered the cosmological constant before in sections 1.2 and 1.2.2. The cosmological constant represents a constant energy density filling space homogeneously. This constant energy is a property of space itself, and as such it would not be diluted as the Universe expands. Unlike in classical mechanics, in quantum mechanics the lowest possible energy state of a vacuum is non-zero. The cosmological constant is assumed to be equivalent to the zero-point energy of space, referred to as the vacuum energy (however, see section 1.3.4).

The existence of dark energy has also been confirmed via various other means including observations of the CMB by Planck Collaboration et al. (2020) which estimates the content of the Universe to be made up of 68.3% dark energy, 26.8% DM and 4.9% ordinary matter. This is in good agreement with measurements from the eBOSS collaboration using the Baryon Acoustic Oscillations (BAO) feature in

the clustering of galaxies and quasars (e.g. Alam et al., 2021).

### Hierarchical clustering

Under the CDM assumption, structure formation is hierarchical: small gravitionally bound structures form first, and their continuous merging creates larger structures, from massive galaxies to galaxy clusters and super-clusters. The highest level of this hierarchy is represented by galaxy clusters, which are the largest gravitationally bound objects in the Universe*. Galaxy clusters do not collapse as they are but grow through minor mergers by accreting smaller galaxy groups or isolated field galaxies. Clusters do correspond to the densest patches of the Universe at $z = 6$ ('proto clusters'), but while they start forming early they virialise late.

Assuming a simple model of spherical collapse, we can express the condition for the virialisation of structures as a function of their velocity dispersion $\sigma$ and total halo mass $M_{\rm halo}$, giving an estimation of the redshift at which a given structure formed:

$$1 + z_{\rm vir} \leq 0.93 \left( \frac{\sigma}{100\,{\rm km\,s^{-1}}} \right)^2 \left( \frac{M_{\rm halo}}{10^{12} M_\odot} \right)^2 \tag{1.51}$$

A MW sized halo has a velocity dispersion of the order $\sigma \sim 300\,{\rm km\,s^{-1}}$ and a halo mass of $M_{\rm halo} \sim 10^{12} M_\odot$. Plugging this into equation 1.51, we find a formation time of $z_{\rm vir} \leq 7.37$. For a galaxy cluster sized halo, with $\sigma \sim 1000\,{\rm km\,s^{-1}}$ and a halo mass of $M_{\rm halo} \sim 10^{14} M_\odot$, we find $z_{\rm vir} \leq 1.93$.

While the equation 1.51 was derived using simple and inaccurate models, it does provide an idea about the time scale on which structures of different sizes form. In other words, it would be unusual to find massive galaxies with redshifts $z > 10$, and if we wish to observe a galaxy cluster we need to observe relatively recently formed structures. Note that galaxies at $z > 10$ do exist. There may even be plenty, but

*Super-clusters are *not* gravitationally bound.

they are just faint and hard to see. However, if they exist, they are not necessarily virialised or have finished forming, which is what equation 1.51 really calculates.

### 1.3.4 Challenges to the ΛCDM paradigm

While the ΛCDM model has been able to successfully explain observables at large scales, there are a number of current (and past) challenges to the ΛCDM model at small scales. Below follows an incomplete list of some of these remaining tensions.

#### The Hubble tension

As observational surveys increase in sensitivity, tensions between predictions from ΛCDM and observations have become apparent. One such tension is the difference between the value of $H_0$ inferred from late- and early-Universe measurements of $H_0$. The 'Supernovae, $H_0$, for the Equation of State of Dark Energy' (SH0ES) collaboration extracted a value from late-Universe supernovae data for the Hubble constant of $H_0 = 74.0^{+1.4}_{-1.4}\,\mathrm{km/s}$. Similarly, the '$H_0$ Lenses in COSMOGRAIL's Wellspring' (H0LiCOW) collaboration derived an independent constraint based on observations of lensed quasars of $73.3^{+1.7}_{-1.8}\,\mathrm{km/s}$ (Wong et al., 2020). Combining these two measurements results in a $5.3\sigma$ tension with the value of $67.4^{+0.5}_{-0.5}$ derived from early-Universe Planck observations of the CMB (Planck Collaboration et al., 2020).

Systematics in either the late-Universe/local or early-Universe/CMB measurements of $H_0$ could be the cause for the tension. However, other studies (with independent systematics) produce similar values of $H_0$ in the early- and late-Universe. Instead, physics beyond the standard *flat* ΛCDM model could be required to explain this tension.

### The cosmological constant problem

Another challenge to ΛCDM is the *cosmological constant problem*. This tension refers to the fact that the observed value of Λ is approximately 120 times smaller than the predicted zero-point energy of quantum field theory. If Λ were slightly larger, the negative pressure from dark energy would dominate over the gravitational attraction from matter and the Universe would fly apart. On the other hand, if Λ were slightly smaller, gravity would dominate and the Universe would collapse. The perfect balance between vacuum and matter is often deemed unnatural. As such, the cosmological constant problem is often referred to as *fine-tuning problem* as well. There are many proposed solutions, including (but not limited to) dynamic forms of dark energy (e.g. Copeland et al., 2006), modified gravity models (e.g. Clifton et al., 2012), physics beyond the standard model (e.g. Marsh, 2016), or simply the anthropic principle.

### Missing satellite problem

There are two frequently discussed problems found in galaxy statistics related to the apparent under-abundance of faint, low mass galaxies in local groups. One of these problems, the 'missing satellite problem', is generally considered to have already been solved. However, because it was a critical challenge to the CDM paradigm at the time, we will briefly describe the problem here. The missing satellite problem notes that the mass function for galaxies at the faint end is significantly less steep than the mass function expected for DM halos. Originally, the problem referred to the discrepancy between the number of satellites predicted in CDM-based simulations and the number observed in the MW. Numerical simulations (Klypin et al., 1999; Moore et al., 1999)* and Monte Carlo realisations of the merging paths of DM haloes (Kauffmann et al., 1993) predicted the number of companions for the MW to be of the order of $\mathcal{O}(100)$. This was in steep contrast with the observed count of

*The missing satellite problems is named after the title of the first paper: 'Where are the Missing Galactic Satellites?'

approximately 10 (e.g. Mateo, 1998). Simulations with improved resolution have since confirmed that a large number of subhaloes should be present in MW-like haloes (e.g. Springel et al., 2008).

The proposed solutions to this problem can be divided into essentially two categories. Either CDM produces too many low mass (sub)haloes or the efficiency with which galaxies form in these haloes decreases with halo mass. The prevailing view favours the second solution. Galaxy evolution models (Bullock et al., 2000; Somerville, 2002; Sobacchi & Mesinger, 2013) and star-formation histories of ultra-faint dwarfs (Brown et al., 2014) have shown that gas accretion is suppressed by the photoionising background. At the same time, stellar feedback can also inhibit further star formation (e.g. Hopkins et al., 2014; Trujillo-Gomez et al., 2014). As such, subhaloes below $\sim 10^9 M_\odot$ are inefficient in forming a luminous component (Wheeler et al., 2014; Shen et al., 2014). In addition, the observed satellite count has been pushed to approximately 50 with the discovery of new ultra-faint ($L \lesssim 50,000 L_\odot$) dwarfs (Bechtol et al., 2015; Drlica-Wagner et al., 2015), and more companions are predicted to be discovered in future surveys (Hargis et al., 2014), reducing the discrepancy further.

Considering all this, the missing satellite problem is considered to be 'solved'. However, satellites of MW size galaxies can still be used the test DM models beyond the standard CDM. For example models that erase too much substructure* could be constrained using the number of satellites (e.g. Bose et al., 2017; Dekker et al., 2021).

### Too-big-to-fail

Related to the missing satellite problem discussed in the previous section is the 'too-big-to-fail' problem. The paper that introduced the issue focuses mainly on the

*E.g. *warm* DM, which moves faster than CDM in the early Universe, where it was created with kinetic energy; not today. A lighter and faster DM particle can travel farther in a given time, and smooth out existing structure along the way.

satellites of the MW (Boylan-Kolchin et al., 2011). In this paper and in subsequent works (Boylan-Kolchin et al., 2012), they demonstrate that the bright satellites of the MW have internal kinematics that are inconsistent with predictions from CDM-based simulations. Specifically, the most massive subhaloes in simulations have masses systematically larger than those measured in the brightest dwarf Spheroidal (dSph) satellites of the MW. The potential wells of massive satellite haloes are deep and so it is unlikely that photoionising feedback can inhibit gas accretion and suppress galaxy formation. It is surprising that these theoretically expected massive satellites are not observed, as these substructures should be 'too big to fail' to form galaxies. A similar problem is present for isolated galaxies in the field (Ferrero et al., 2012; Garrison-Kimmel et al., 2014; Klypin et al., 2015; Papastergis et al., 2015).

There are numerous solutions to this problem that do not require physics beyond the standard model. A possibility is that the MW is less massive than currently thought, and should therefore host a smaller number of massive subhaloes Wang et al. (2012). It has also been argued that density profiles with flat centres (cores, see next section) could solve the too-big-to-fail problem (Brooks et al., 2013), where baryonic processes, such as supernova feedback, could flatten the inner DM density distribution (Navarro et al., 1996; Pontzen & Governato, 2012; Oñorbe et al., 2015; Faucher-Giguére, 2017). However, several studies have shown that feedback is possibly not energetic enough to remove the mass required to explain the too-big-to-fail problem (e.g. Peñarrubia et al., 2012). On the other hand, some studies have suggested cored haloes or feedback are not required at all (Fattahi et al., 2016; Sawala et al., 2016). Ram pressure can remove the galactic gas, while tidal forces can strip away a halo's DM, thus reducing the satellites' mass*.

*Field galaxies are not subject to environmental effects and so only internal baryonic effects can be invoked to reduce their halo masses. It remains to be seen if the various solutions mentioned here can solve the too-big-to-fail problem in the field as well.

### The core-cusp problem & the diversity of rotation curves

DM-only simulations predict that DM halo profiles can be described by a nearly universal profile across all masses and cosmologies (Navarro, Frenk & White, 1997). A common way to characterise this is via the Navarro-Frenk-White (NFW) profile:

$$\rho = \frac{4\rho_s}{\frac{r}{R_s}(1+\frac{r}{R_s})^2}, \tag{1.52}$$

where the scale radius $R_s$ and $\rho_s$ (the density at the scale radius) are parameters that vary from halo to halo. These profiles rise steeply at smaller radii, i.e. $\rho(r) \propto r^{-\gamma}$ with $\gamma > 0$. This is in direct tension with observations of halos which prefer fits with more flattened density profiles in the inner regions, i.e. $\rho(r) \propto r^{0-0.5}$. The density profiles of the simulated halos tend to be 'cuspy', while the profiles of observed halos are 'cored', hence this discrepancy is often referred to as the 'core-cusp' problem. The core-cusp problem first emerged when Flores & Primack (1994) and Moore (1994) studied the rotation curves of low mass dwarf galaxies, and was later identified by several other studies (e.g. de Blok et al., 2003; Oh et al., 2011). It has also been found to be present in galaxy clusters sized halos (Sand et al., 2002, 2004; Newman et al., 2009, 2011, 2013a,b).

These tensions first arose between observations and DM-only simulations. As such, there has been extensive debate whether this discrepancy can be alleviated by the inclusion of baryonic physics in simulations, which could alter the inner regions of DM haloes. As stated in the previous section, supernova feedback could flatten density profiles in low mass galaxies. On the other hand observations could be biased such that cored profiles are inferred when in fact a cusp is present (e.g. Dalcanton & Stilp, 2010; Pineda et al., 2016).

However, not all dwarf galaxies have cored density profiles. In fact, there is a surprising amount of diversity in the density profiles of galaxies (Oman et al., 2015) considering the prediction from simulations that their should be a universal density

profile. The parameters of the NFW profile, $R_s$ and $\rho_s$, correlate tightly, meaning only one is needed to specify the profile of a halo. If a given halo parameter, such as e.g. $V_{\max}$*, is fixed, the halo density profile is completely determined at all radii. However, galaxies with similar $V_{\max}$ can have quite different central densities.

A possible explanation for the core-cusp problem and the diversity of rotation curves is that the CDM paradigm breaks down at sub-galactic scales. Instead DM particles could interact with other DM particles through some force besides gravity. This DM is not collisionless, but *self-interacting*, and hence it is called self-interacting dark matter (SIDM). We will discuss this alternative to CDM in more detail in the next section.

### 1.3.5 Self-interacting dark matter

DM self-interactions were first suggested by Spergel & Steinhardt (2000) as a solution to the core-cusp problem discussed in section 1.3.4. In their original model for SIDM, the particles scatter elastically and isotropically with each other through $2 \to 2$ interactions, i.e. interactions where both the initial and final state are two DM particles.

The scattering rate of a DM particle is dependent on the local DM density, its relative velocity with respect to other DM particles, and the interaction cross-section $\sigma/m$, where $m$ is the DM particle mass. The cross-section is a measure of the probability of an interaction occurring. If this cross-section is large enough, the interactions could affect the internal structure of halos. With a mean free path ranging from 1 kpc to 1 Mpc† DM self-interactions would preserve the large scale success of ΛCDM and could resolve the tensions discussed in section 1.3.4 (Spergel & Steinhardt 2000, and for a review see Bullock & Boylan-Kolchin 2017).

*$V_{\max}$ is the maximum value of the circular velocity, see equation 1.50. It is a proxy for halo mass.

†At densities characteristic of the MW's DM halo (0.4 Gev/cm$^3$; Read 2014), leading to cross sections of $400 \gtrsim \sigma/m \gtrsim 0.4\,\mathrm{cm}^2/\mathrm{g}$.

The mechanism through which SIDM can induce core formation is thermalisation: particle collisions redistribute energy and consequently heat the inner regions of the halo. The heated particles move to orbits which greater apocentres, depleting the centre of mass (Spergel & Steinhardt, 2000; Burkert, 2000; Yoshida et al., 2000). This is illustrated in figure 1.2, which shows the DM density profile of a simulated cluster-sized halo at various times after DM self-interactions have been 'turned on'. After a few Gyr, the cuspy profile has been turned into a cored profile.

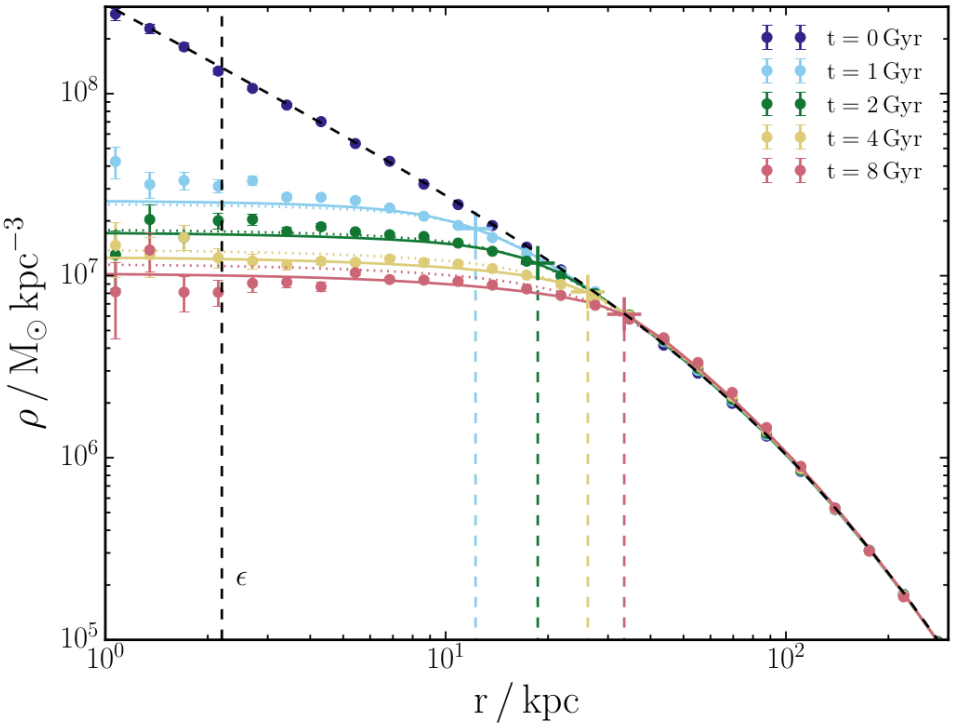


Figure 1.2: Figure 4.9 from Robertson (2017). DM halos can form cores through the process of thermalisation: the redistribution of energy due to particle collisions heats the inner regions and consequently mass flows out leaving a depleted centre. The density profile of the simulated halo is initially cuspy, but 1 Gyr after self-interactions have been 'turned on', the profile has become cored.

In relation to galaxy rotation curves, several studies have investigated to what extent DM self-interactions could produce their striking diversity (e.g. Elbert et al. 2018 with $N$-body simulations, and Kaplinghat et al. 2014 with analytical models). For example, the interaction between a baryonic disk and the SIDM halo in which it resides could lead to small changes in the baryonic component of a dwarf galaxy

producing large changes in the total density profile. The studies mentioned above showed that SIDM can both increase and decrease the central density of DM in the presence of baryons, depending on how centrally concentrated the baryonic component is. SIDM could possibly explain the diversity of rotation curves even with the inclusion of baryonic physics.

The cross-section necessary to produce cores in galaxy sized halos is of the order of $\sigma/m = 0.1\,\mathrm{cm}^2/\mathrm{g}$ (Newman et al., 2013a,b). However, SIDM models that alleviate the too-big-to-fail problem require the cross-section to be larger than $\sigma/m > 1\,\mathrm{cm}^2/\mathrm{g}$ (Zavala et al., 2013). If DM behaves as a collisional fluid on small scales while it is essentially collisionless over large scales, SIDM models could simultaneously reproduce the cores of dwarf galaxies as well as the galaxy clusters' shapes. Since the average DM particle velocity increases with halo mass, such that studies at different astrophysical scales probe $\sigma/m$ as a function of scattering velocity, SIDM with a velocity-*dependent* cross-section could produce this behaviour.

### Constraints to the SIDM cross-section

The velocity dispersion of DM particles is of the order of 1000 km/s in galaxy clusters, approximately two orders of magnitude higher than in dwarf galaxies. Therefore, combined with the high density of these environments, the scattering rate is expected to be highest in clusters*. As such many studies investigating SIDM have focused on galaxy clusters.

Meneghetti et al. (2001) investigated how the ability of galaxy clusters to produce giant gravitationally lensed arcs is influenced by DM self-interactions. Internal scattering changes the structure of a halo, reducing the number of substructures and making the halo less centrally concentrated. The morphology of long arcs observed in and around the cluster depends on the core density, and the location of the radial arcs can put strong constraints on the size and compactness of the

*This is not true if the cross-section decreases with velocity, which is the case for some SIDM models with a velocity-*dependent* cross-section.

core of the cluster acting as the lens. Using high resolution simulations of galaxy clusters, Meneghetti et al. (2001) constrained the cross-section to be no larger than $\sigma/m = 0.1\,\mathrm{cm}^2/\mathrm{g}$. However, this study was recently revisited by Vega-Ferrero et al. (2021), and they conclude that it is not possible to rule out a cross-section of $\sigma/m \lesssim 1\,\mathrm{cm}^2/\mathrm{g}$ based on the formation of radial arcs by simulated galaxy clusters, suggesting the relatively low redshift of Meneghetti et al. (2001)'s simulated cluster was the cause of the stringent constraints.

The interactions between DM particles can cause their orbits to be changed. This in turn causes the velocity distribution of a set of DM particles to become more isotropic, leading to more spherical spatial distributions. By comparing the shapes of the DM halos of galaxies and galaxy clusters to simulated counterparts run with various DM physics, one can put a constraint on the strength of self-interactions. Initial studies of cluster ellipticities put strong constraints to the cross-section of $\sigma/m < 0.02\,\mathrm{cm}^2/\mathrm{g}$ (Miralda-Escudé, 2002). Such a small cross-section would render SIDM essentially useless for the initial reason it was proposed, i.e. relieving the tension on small scales between observations and theoretical predictions

However, Peter et al. (2013) showed that these constraints were off by more than an order of magnitude. The found that the core set by scatterings* retains more of its triaxial nature than estimated before, and that the triaxial mass distribution outside this region contributes to the ellipticity of the core. As such, they allow for a DM self-interaction cross-section at least as large as $\sigma/m < 1\,\mathrm{cm}^2/\mathrm{g}$. Rocha et al. (2013) compared the central densities of observed and simulated clusters using the same set of N-body simulations as Peter et al. (2013), and found similar observational constraints.

In case of cluster mergers, where the interaction rate is expected to be even higher than in isolated clusters, DM self-interactions can induce an offset between the DM and the collisionless galaxies. Observations of merging systems have placed a limit

*The region within a radius where DM particles on average have interacted only once. Outside of this region, the (average) scattering rate drops off.

to the cross-section of $\sigma/m \leq 0.47\,\mathrm{cm}^2/\mathrm{g}$ (95%) (Harvey et al., 2015). However, after correcting for an underestimation of uncertainty in the offset measurements, recent studies have relaxed this upper limit on $\sigma/m$ to $2\,\mathrm{cm}^2/\mathrm{g}$ (Wittman et al., 2018). We will investigate the effects of DM self-interactions on merging clusters in chapter 3.

There have also been various studies of SIDM at smaller scales. As well as the shapes of galaxy clusters, the shapes of galaxies are also affected by DM self-interactions. Using cosmological baryonic zoom simulations of MW-mass galaxies, Vargya et al. (2021) found that the assembly history of galaxies had a greater effect on the shape of the halos than any variation in $\sigma/m$*. However, the radius where the shape of the total mass distribution begins to differ from that of the stellar mass distribution is dependent on $\sigma/m$. This transition could potentially be used to set limits on the SIDM cross-section in the MW.

Di Cintio et al. (2017) found that the dynamics of supermassive black holes differ in their hydrodynamical simulations run with SIDM and CDM physics. Due to the increased dynamical friction time-scale caused by the lower DM density in SIDM galaxies, a large fraction of the black hole population is off-centred from the centre of their host galaxy in less massive galaxies. This could indicate another possible test of SIDM at scales smaller than galaxy clusters.

Similarly, an observable consequence of cored density profiles is oscillations of the brightest cluster galaxy (BCG) in otherwise relaxed galaxy clusters (Harvey et al., 2017; Kim et al., 2017). This phenomenon is absent with CDM where a cuspy density profile keeps a BCG tightly bound at the centre. BCG 'wobbles' therefore represent another avenue for constraining SIDM. We discuss this phenomenon in more detail in chapter 3.

All of these possible tests of SIDM have been proposed, yet of none of them have help up or have been followed up by a dedicated observing campaign. In this thesis

*Robertson et al. (2018) found similar results for galaxy clusters.

we aim to develop a robust test of SIDM, and (some of) the observational tools needed to carry it out.

## 1.4 Thesis Outline

This thesis mainly aims to constrain the self-interacting DM cross-section using simulated galaxy clusters. For this, we use data from cosmological hydrodynamic simulations run with CDM and SIDM physics. The particular simulations are described in chapter 2 and chapter 3.

In chapter 2, we exploit the effects of dark matter self-interactions on the mass loss of galaxies accreted unto galaxy clusters. In chapter 3, we introduce an ongoing project that aims to constrain the DM self-interaction cross-section by comparing offsets between DM and stellar matter in simulations of clusters run with different DM physics.

In chapter 4, we introduce an upcoming balloon-borne telescope named SuperBIT. SuperBIT's main science goal is to map out DM around galaxy clusters. Then, in chapter 5, we describe the SuperBIT Data Recovery System (DRS), a toolkit of flight-proven hardware and software to retrieve data from a stratospheric balloon platform, which was conceived to retrieve data mid-flight from the SuperBIT telescope.

Finally, chapter 6 summarises and discusses all the work presented here, and explores possible directions for future work.

Chapter 2

# The effects of self-interactions on dark matter stripping of galaxies falling into clusters

*"Do not go gentle into that good night.*

*Rage, rage against the dying of the light."*

— Dylan Thomas

## 2.1 Introduction

As galaxies fall into clusters, they are transformed, morphologically and spectroscopically. Their gas content, hitting the intra-cluster gas, is shocked. Turbulence causes a sudden, final burst of star formation — before ram pressure and gravitational tides strip it away, quenching star formation thereafter (e.g. McCarthy et al., 2008; van den Bosch et al., 2008; Oman et al., 2021). The galaxies' DM is also eventually stripped by tidal gravity and gradually becomes incorporated into the (now slightly larger) cluster. This is the key mechanism for the growth of structure in the Universe; yet, the timescale for DM stripping and virialisation remains poorly understood.

In the standard ΛCDM model of cosmology, DM particles interact with each other

only through gravity. The model successfully explains all observables at large scales, such as the galaxy clustering signal (for a review see Frenk & White, 2012) and the CMB anisotropy (e.g. Planck Collaboration et al., 2016). However, there is no a priori reason why DM particles should not interact with each other (Burkert, 2000; Spergel & Steinhardt, 2000), and weak self-interactions are a natural consequence of some particle physics theories for the origin of DM (for a review see, e.g., Tulin & Yu, 2018). With a mean free path ranging from 1 kpc to 1 Mpc (see section 1.3.5), DM self-interactions would preserve the large scale success of ΛCDM, and could resolve tensions between the results of DM-only simulations and observations of dwarf and low-mass galaxies (for a review see Bullock & Boylan-Kolchin, 2017).

Massive galaxy clusters are a promising environment to search for DM-DM interactions, because the interaction rate would be proportional to the local DM density and to the local velocity dispersion of DM particles (for a review see Massey et al., 2010). Observations have placed several limits on the strength of the SIDM cross-section per unit mass ($\sigma/m$) at the typical velocities encountered in clusters, including $\sigma/m \lesssim 1\,\mathrm{cm^2\,g^{-1}}$ (Peter et al., 2013, from cluster halo shapes), $\sigma/m < 1\,\mathrm{cm^2\,g^{-1}}$ (Rocha et al., 2013, from cluster core sizes), $\sigma/m < 0.1\,\mathrm{cm^2\,g^{-1}}$ (Meneghetti et al. 2001, from strong lensing arc statistics, but see also Vega-Ferrero et al. 2021), and $\sigma/m < 0.47\,\mathrm{cm^2\,g^{-1}}$ (Harvey et al. 2015, from DM-galaxy offsets in merging clusters, but see also Wittman et al. 2018). Merging clusters are sufficiently rare that interpretation of them tends to be limited by uncertainty in their orientation with respect to the line-of-sight (Clowe et al., 2006; Bradač et al., 2008; Dawson et al., 2012). However, the promising prospects revealed by Robertson et al. (2017a)'s detailed study of high-velocity DM collisions motivates a search for more ubiquitous examples of objects falling into clusters.

Whenever a galaxy falls into an SIDM cluster, interactions between its DM particles and those in the cluster could scatter DM out of the galaxy. This 'evaporation' acts in addition to tidal stripping, and accelerates overall mass loss. The orbits of infalling galaxies might also be changed. Galaxies spiral toward the centre of a

cluster due to dynamical friction, which has strength proportional to the galaxy's mass (Binney & Tremaine, 2008, chapter 8). If galaxies lose additional mass, they might sink less far or more slowly into the cluster. On the other hand, drag due to the DM self-interactions (which may be positive or negative Robertson et al., 2017a) could increase the rate of decay; or inhibit the formation of trailing density wakes in the first place (Di Cintio et al., 2017).

The aims of this work are to study the differences in DM mass loss and orbital dynamics of cluster galaxies, using hydrodynamical simulations with CDM and SIDM physics — and to investigate whether the differences would be observable. The only previous study of such effects used DM-only simulations (Bhattacharyya et al., 2022).

This paper is organised as follows: in section 2.2, we present the simulation suite used in this work; in section 2.3 we study the effects of self-interactions by matching galaxies between our CDM and SIDM simulations; and in section 2.4 we investigate the effects on observables using the population of galaxies at $z = 0$. Finally, we discuss our results and present our conclusions in section 2.5.

## 2.2 Data

### 2.2.1 The EAGLE and Cluster-EAGLE simulations

We use the 50 Mpc Evolution and Assembly of GaLaxies and their Environments (EAGLE) cosmological simulation (Schaye et al., 2015) and the Cluster-EAGLE (C-EAGLE) zoom cosmological simulations of smaller volumes centred on $\gtrsim 10^{14} M_{\odot}$ galaxy clusters (Bahé et al., 2017). Both were run with a modified version of the GADGET-3 code that includes radiative cooling, star formation, chemical evolution, and stellar and AGN feedback (with the 'AGNdT9' feedback model Schaye et al. 2015; Crain et al. 2015). The DM particle mass is $9.7 \times 10^6$ M$_{\odot}$, the initial gas particle mass is $1.8 \times 10^6$ M$_{\odot}$, and the gravitational softening length was set to 2.66

comoving kpc before $z = 2.8$, and then kept fixed at 0.7 physical kpc at $z < 2.8$. The simulations assume cosmological parameters from Planck Collaboration et al. (2014).

The EAGLE volume and two of the C-EAGLE clusters, CE-05 and CE-12, have been re-simulated from identical initial conditions in a ΛSIDM universe (see table 2.1 and Robertson et al. 2018 for more details). These two particular C-EAGLE clusters are 'relaxed', based on their gas properties at $z = 0.1$ (Barnes et al., 2017). Since CE-12 is slightly more massive, and has more member galaxies, we shall quote the higher signal-to-noise statistics from that cluster whenever we study the differences between CDM and SIDM at $z = 0$. However, no data are available for that cluster at higher redshift, so we shall use CE-05 whenever we trace the evolution of DM through time. Note that the central galaxy in CE-05 happened to form early, and the central density cusp has been retained in both CDM and SIDM. The central galaxy of CE-12 formed later, and SIDM interactions created a ∼100 kpc constant density core by $z = 0$. In the inner ∼100 kpc few satellites enter, and if they do they stay for a short time, and so we expect the effect from the constant density core in CE-12 to be negligible compared to the cluster being more massive.

Our implementation of SIDM assumes an isotropic, velocity-independent interaction cross-section, $\sigma/m = 1\,\mathrm{cm}^2\,\mathrm{g}^{-1}$. This is around the upper limit of values compatible with current measurements, and therefore maximises the observable consequences. During each simulation timestep, $\Delta t$, DM particles scatter elastically off neighbours within radius $h_{\mathrm{SI}} = 2.66$ kpc (comoving) with probability

$$P_{\mathrm{scat}} = \frac{(\sigma/m)\; m_{\mathrm{DM}}\, v\, \Delta t}{\frac{4}{3}\pi h_{\mathrm{SI}}^3}\,, \tag{2.1}$$

where $v$ is the particles' relative velocity and $m_{\mathrm{DM}}$ the DM particle mass (for more details see Robertson et al., 2017b). We log the time and particle IDs of all DM scattering events. This enables us to distinguish between: DM particles that have not scattered; those that have scattered with other DM particles from their own (sub)halo; and those that have scattered with DM particles from elsewhere in the

cluster.

### 2.2.2 Finding and tracking individual galaxies

We detect groups of particles in the simulations using a FRIENDS-OF-FRIENDS (FoF, Davis et al., 1985) algorithm with linking length 0.2, and identify individual subhaloes (in all 30 simulation snapshots from $z = 14$ to $z = 0$) using the SUBFIND (Springel et al., 2001; Dolag et al., 2009) algorithm. For SUBFIND to identify a galaxy it must have at least 20 particles. We track subhaloes between snapshots, and construct their merger trees using the D-TREES algorithm (Jiang et al., 2014). This identifies each subhalo's $N_{\rm link}$ most bound particles of any species, with $N_{\rm link} = \min(100, \max(0.1N_{\rm gal}, 10))$, where $N_{\rm gal}$ is the total number of particles in the subhalo in each snapshot. The descendant of a subhalo is the object that contains most of these $N_{\rm link}$ particles in the next snapshot. A subhalo can have multiple progenitors in the previous snapshot, but we define the main progenitor as that for which the mass summed across all earlier snapshots is the largest. The main branch of a subhalo is comprised of its main progenitors and descendants. We use the main branches of subhaloes to trace their properties through time.

We identify as 'field galaxies' all SUBFIND central halos (rank 0 in a given FoF group) in EAGLE that contain at least one star particle. We identify as 'cluster member galaxies' all SUBFIND subhaloes in C-EAGLE that contain at least one star particle and are within radius $2R_{200}$. We define their time of infall as the first snapshot after they enter that radius for the first time. By keeping all galaxies within $2R_{200}$ we include those galaxies which have already passed through the cluster once (and thus have felt its effects) and have passed beyond $R_{200}$ again, i.e. the splashback population. Additionally, by keeping galaxies within $2R_{200}$ we end up with a larger number of total and high mass galaxies.

The mass of every galaxy is defined as the total mass, $M_{\rm tot}$, of all particles gravit-

Table 2.1: Properties of the CDM and SIDM versions of the two C-EAGLE clusters at redshift $z = 0$. The mass $M_{200}$ is that enclosed within the sphere of physical radius $R_{200}$ whose mean density is 200 times the critical density of the Universe. Cluster member galaxies are the $N_{\rm gal}$ subhaloes in the FoF group of the cluster that are within $2R_{200}$ of the cluster centre and contain one or more star particles.

| Simulation | DM Type | $M_{200}/\mathrm{M}_\odot$ | $R_{200}/\mathrm{Mpc}$ | $N_{\rm tot}$ |
|---|---|---|---|---|
| CE-05 | CDM | $1.38 \times 10^{14}$ | 1.09 | 1442 |
| | SIDM | $1.36 \times 10^{14}$ | 1.09 | 1183 |
| CE-12 | CDM | $3.96 \times 10^{14}$ | 1.55 | 3893 |
| | SIDM | $3.91 \times 10^{14}$ | 1.54 | 2938 |

ationally bound to it (i.e. the mass $M_{\rm SUB}$ assigned to the subhalo by the SUBFIND algorithm). Its stellar mass, $M_\star$, is defined as the total mass of stars within twice its half light radius. Its location is defined by the location of its constituent particle with the lowest gravitational potential energy.

### 2.2.3 The stellar-to-halo mass relation

Below we will compare the stellar-to-halo mass relation (SHMR) of galaxies in our SIDM and CDM models. We fit the SHMR to a population of simulated galaxies using the form of the Moster et al. (2013) relation derived from abundance matching,

$$M_\star(M_{\rm tot}) = 2NM_{\rm tot}\left[\left(\frac{M_{\rm tot}}{M_1}\right)^{-\beta} + \left(\frac{M_{\rm tot}}{M_1}\right)^{\gamma}\right]^{-1}. \tag{2.2}$$

By numerically inverting equation (2.2), we also fit $M_{\rm tot}(M_\star)$, which can be measured observationally.

We use the Markov Chain Monte Carlo (MCMC) sampler EMCEE (Foreman-Mackey et al., 2013) to obtain the best-fit values and posterior PDFs of the free parameters, $M_1$, $N$, $\beta$, $\gamma$, as well as the free parameter, $\sigma_{\rm M}$, the scatter in stellar mass (or in total mass for the inverse fit), which we assume to be constant. The latter enters the fit through the log likelihood function,

$$\log\mathcal{L} = -\frac{1}{2}\sum_{i=1}^{N}\left(\frac{\log M_i - \log M_i^{\rm mod}}{\sigma_{\rm M}}\right)^2 - \frac{N}{2}\log\left(2\pi\sigma_{\rm M}^2\right), \tag{2.3}$$

where the summation is over the total number of galaxies, $N$; $M_i$ is the stellar/total mass of galaxy $i$, and $M_i^{\rm mod}$ is the modelled stellar/total mass of galaxy $i$, for a

given set of parameters. When fitting the SHMR, we truncate fits at the mass where each galaxy includes at least 10 star particles.

### 2.2.4 Matching galaxies between simulations

We match galaxies between the CDM and SIDM simulations, so their evolution can be individually compared. In the snapshot after each galaxy crosses $2R_{200}$, we identify its counterpart in the other simulation as that which contains the highest fraction, $f_{\rm match}$, of shared particle IDs

$$f_{\rm match} = \frac{N_{\rm shared}^2}{N_{\rm CDM,tot} N_{\rm SIDM,tot}}, \tag{2.4}$$

where $N_{\rm shared}$ is the number of DM particles the CDM galaxy and a possible matching SIDM galaxy have in common, $N_{\rm CDM,TOT}$ the total number of DM particles in the CDM galaxy, and $N_{\rm SIDM,TOT}$ the total number of DM particles in the SIDM galaxy. To complete an association, we require a bijective match: i.e. the CDM galaxy points to an SIDM galaxy that points back to it. The paired CDM and SIDM galaxies inevitably have slightly different infall masses and infall times. When we bin by these, we use the CDM values. This is an arbitrary choice, but none of our results change qualitatively when using either SIDM or common bins (with logarithmic bins of 1 dex in mass, only 10 per cent of galaxies are binned differently).

When analysing matched galaxies, we ignore any cluster galaxies that were unmatched to cluster galaxies, and any field galaxies that were unmatched to a central galaxy. In cluster CE-12, 96 out of 889 CDM cluster galaxies were matched to the central galaxy of the SIDM cluster. For the field galaxies, 383 out of 9126 CDM galaxies were matched to a satellite galaxy in the SIDM simulation.

## 2.3 Evolution of DM since infall

In this section we examine the effect of self-interactions on the DM mass of galaxies after they fall into the clusters, by directly matching galaxies between the CDM

and SIDM simulations.

### 2.3.1 The behaviour of one example galaxy

To build intuition, we first track the DM halo of one galaxy in detail. We identify a typical galaxy that fell into the cluster CE-05 at $z = 1.99$ with mass $2.7 \times 10^{11} M_\odot$, and track the 6D phase space coordinates (cluster-centric radius and velocity) of all its DM particles to 2 Gyr ($z = 1.15$) and 10.5 Gyr ($z = 0$) after infall. The result is illustrated in figure 2.1 which shows that self-interactions increase the mass loss of the SIDM galaxy compared to its CDM counterpart, but the orbit is unaffected.

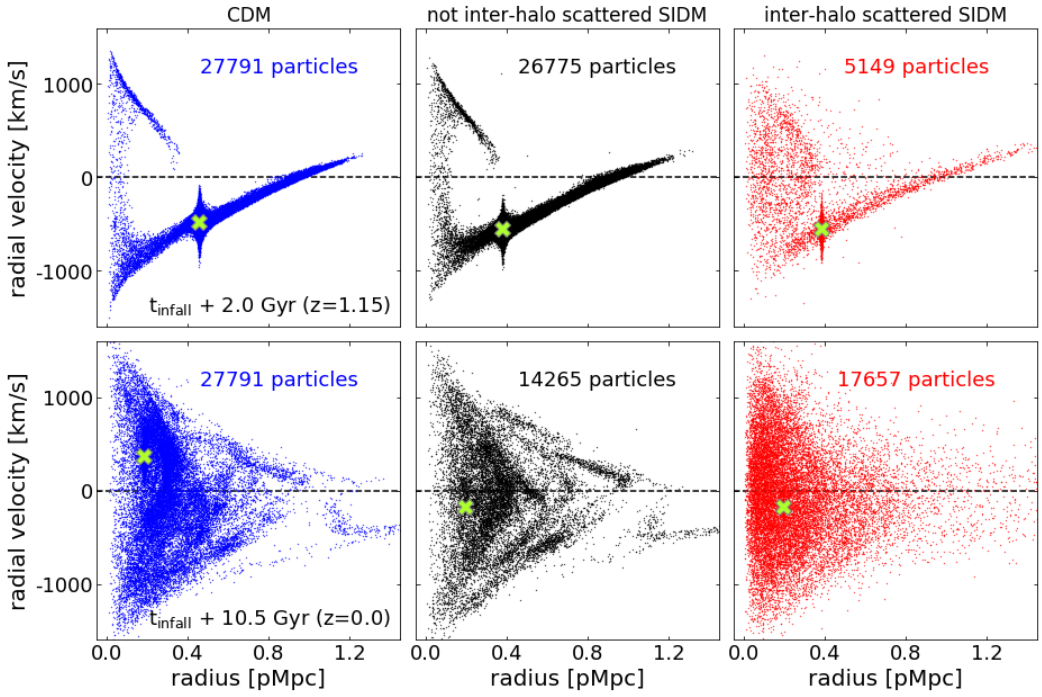


Figure 2.1: Cluster-centric radial velocity as a function of distance from the cluster centre for the DM in a CDM satellite of CE-05 and its SIDM counterpart, 2 Gyr (top) and 10.5 Gyr (bottom) after infall. Particles moving outwards from the centre of potential of the cluster have positive radial velocity. Plotted here is the DM that was in the satellite at infall. *Left column*: the phase space properties of the DM in the CDM galaxy. *Middle column*: the properties of the DM in the SIDM galaxy that has not scattered with the cluster halo DM in the time since infall. *Right column*: same as middle column, but for the SIDM that has scattered with the cluster halo since infall. The location of the galaxy itself is indicated by a green cross on each panel.

### Dark matter loss

The velocity dispersion of DM within the galaxy is reflected in the 'Fingers of God' extending towards positive and negative radial velocities. Tidally stripped DM extends both forwards and backwards along the galaxy's orbit: by 2 Gyr, some particles have already passed through pericentre and are now moving back out. On a phase-space diagram, tidally stripped material moves along the same path as the galaxy it has been removed from, both in the case of CDM and SIDM. However, the evaporated material should occupy a region distinct from the galaxy and tidally stripped material.

We separate the SIDM into particles that have scattered with DM particles in the cluster, and particles that have not (figure 2.1). Note that some scattering events result in very low exchange of momentum or merely swap particle trajectories, so the scattered particles include some that have barely been perturbed. However, we find many DM particles that do not follow the tidally stripped material and therefore must be evaporated DM. After 2 Gyr the CDM galaxy has lost roughly 54 per cent of its DM mass since infall, whereas the SIDM galaxy has lost approximately 76 per cent of its DM mass. By $z = 0$, these fractions have increased to 91 per cent and 99 per cent. Evaporation has increased the mass loss in the SIDM galaxy with respect to its CDM counterpart.

We find a much greater SIDM mass loss from galaxies in clusters, than Dooley et al. (2016, figure 9) found for dwarf galaxies in the MW (with the same SIDM cross-section, only a few per cent more than CDM, 10 Gyr after accretion). This striking difference is probably due to the much greater DM density and scattering rate in a cluster, but occurs despite the deeper potential wells.

### Orbital evolution

After 2 Gyr, the CDM galaxy has moved to a 3D cluster-centric radius of $\sim$0.4 physical Mpc (pMpc), with a mean radial velocity centred on about $-500\,\mathrm{km\,s^{-1}}$, i.e. the

galaxy is moving towards the centre of potential of the cluster (green cross on the top row of figure 2.1). Its SIDM counterpart is within ∼0.1 pMpc and has a similar mean radial velocity. By $z = 0$ the CDM galaxy has moved to a radius of ∼0.2 pMpc, with a mean velocity of about $+500\,\mathrm{km\,s^{-1}}$. Its SIDM counterpart has a mean velocity of about $-100\,\mathrm{km\,s^{-1}}$, but is located at about the same radius (green cross on the bottom row of figure 2.1). Indeed, we find that there is virtually no difference between the evolution of the 3D cluster-centric radius over time of the CDM and SIDM galaxy (not shown). Self-interactions increase the mass loss of the galaxy, but do not have a significant effect on its orbit.

### 2.3.2 The behaviour of a population of galaxies

The galaxy used to produce figure 2.1 is just one example of the many member galaxies of cluster CE-05. In this section, we investigate the effect of self-interactions on the evolution of DM particles for a large sample of infalling galaxies.

#### Dark matter loss

In figure 2.2 we plot the cumulative distribution at redshifts* $z = 1$ and $z = 0$ of the fraction of DM lost from all CDM and SIDM galaxies that were within their cluster in one or more of the 30 simulation snapshots of CE-05. We separate the galaxies into logarithmic bins of 1 dex in infall mass from $10^9$ to $10^{12}\mathrm{M}_\odot$. When a galaxy merges with the cluster central galaxy or into the main branch of some other galaxy, we consider it to have been completely disrupted and we set the fraction of DM lost to 1.

At both redshifts, we find that for a given fraction of DM lost, $f$, a greater fraction of the SIDM galaxies have lost a greater portion of their DM than $f$ compared with the CDM galaxies, reflecting the increased mass loss due to self-interactions. The biggest difference is between the low infall mass CDM and SIDM galaxies (dotted

*To be precise, there is no simulation snapshot at exactly $z = 1$. The snapshot used here is actually at $z = 1.02$.

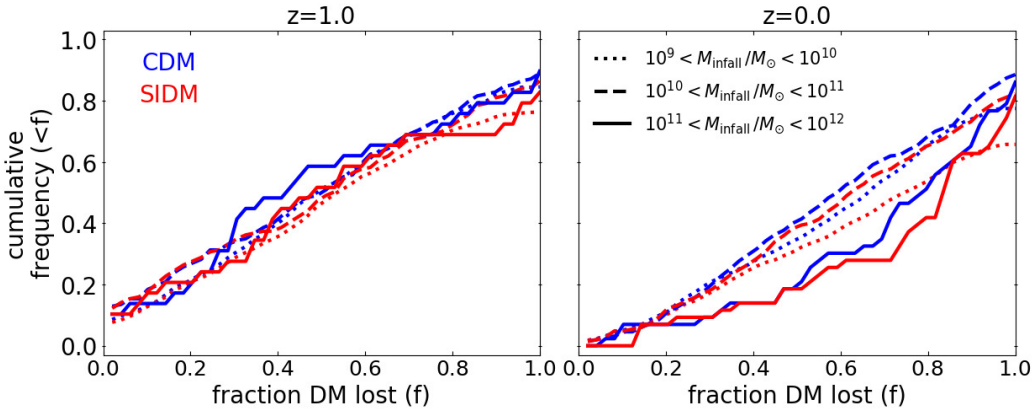


Figure 2.2: The cumulative frequency of galaxies with a fraction of DM lost smaller than $f$. Plotted in blue are the distributions for the CDM galaxies of CE-05, and in red their SIDM counterparts. The left panel shows the results at $z = 1$, the right panel the results at $z = 0$. The different line styles represent different bins of mass at infall, as shown in the legend. A galaxy that has been completely disrupted, i.e. merged with another galaxy or with the main cluster halo is assigned $f = 1$. The fraction of disrupted galaxies in each infall mass bin is given by 1 minus the cumulative frequency at $f = 1$, as the cumulative frequency is plotted for fractions smaller than $f$.

lines on figure 2.2). By $z = 0$, the mass loss and the difference between the SIDM and CDM galaxies have increased relative to $z = 1$. We find that a larger fraction of SIDM than CDM cluster galaxies have been disrupted across all mass bins and at both redshifts; see table 2.2. This is in line with our expectations, as increased mass loss from self-interactions should lead to more disrupted cluster galaxies.

The high mass galaxies (solid lines) have lost a greater fraction of their DM than the galaxies in the other infall mass bins (the solid lines have a different shape than the dotted and dashed lines). This is most likely a consequence of the high mass galaxies having sunk further into the cluster, where stripping becomes more efficient. The timescale for dynamical friction scales with the inverse of the velocity dispersion of the galaxy cubed (section 8.1.1 in Binney & Tremaine 2008), i.e. the time scale decreases as the (infall) mass of the cluster galaxy increases.

For SUBFIND to identify a galaxy it needs to have at least 20 particles. As a consequence a $10^8 M_\odot$ galaxy can only lose approximately 90 per cent of its mass before it is already considered disrupted, compared to approximately 99.9 per cent for a

Table 2.2: Fraction of disrupted cluster member galaxies of the CDM and SIDM version of CE-05, at $z = 1$ and $z = 0$ and separated into bins of 1 dex in mass at infall.

| | $z = 1$ | | $z = 0$ | |
|---|---|---|---|---|
| Mass range | $N_{\rm disrupted}/N_{\rm tot}$ | | $N_{\rm disrupted}/N_{\rm tot}$ | |
| $M_\odot$ | CDM | SIDM | CDM | SIDM |
| $10^9 - 10^{10}$ | 0.15 | 0.24 | 0.23 | 0.34 |
| $10^{10} - 10^{11}$ | 0.11 | 0.13 | 0.11 | 0.18 |
| $10^{11} - 10^{12}$ | 0.1 | 0.17 | 0.14 | 0.19 |
| $10^9 - 10^{12}$ | 0.15 | 0.22 | 0.2 | 0.31 |

$10^{11}\mathrm{M}_\odot$ galaxy. As a result, relatively fewer high mass galaxies disrupt compared to low mass galaxies, even though the high mass galaxies tend to lose a larger fraction of their DM overall.

The cluster galaxy used to produce figure 2.1 has an infall mass of $2.7 \times 10^{11}\,\mathrm{M}_\odot$, placing it in the high mass bin of figure 2.2. By $z = 0$, the CDM and SIDM version of this galaxy have lost approximately 91 per cent and 99 per cent of their DM mass at infall, corresponding to cumulative frequencies of approximately 0.7 and 0.8 respectively. While both have lost more of their DM than most galaxies of their (high) mass, the loss is not remarkable.

### Orbital evolution

We found that the CDM galaxy and its SIDM counterpart used to produce figure 2.1 followed nearly the same orbit. To determine whether galaxy orbits in general are unaffected by self-interactions, we now consider the median evolution since time of infall for a sample of galaxies orbiting in the cluster CE-05, in figure 2.3. We use a sample of 396 matched cluster member galaxies (see section 2.2.4) from CE-05 that have $M_\star \gtrsim 10^7\,\mathrm{M}_\odot$ at $z = 0$. Depending on their infall redshift, the galaxies have spent a different amount of time in the cluster, so a different number of galaxies contribute to each point of figure 2.3.

SIDM galaxies start losing more mass than their CDM counterparts about 2 Gyr after infall (bottom left panel of figure 2.3). By 9 Gyr after infall, CDM galaxies

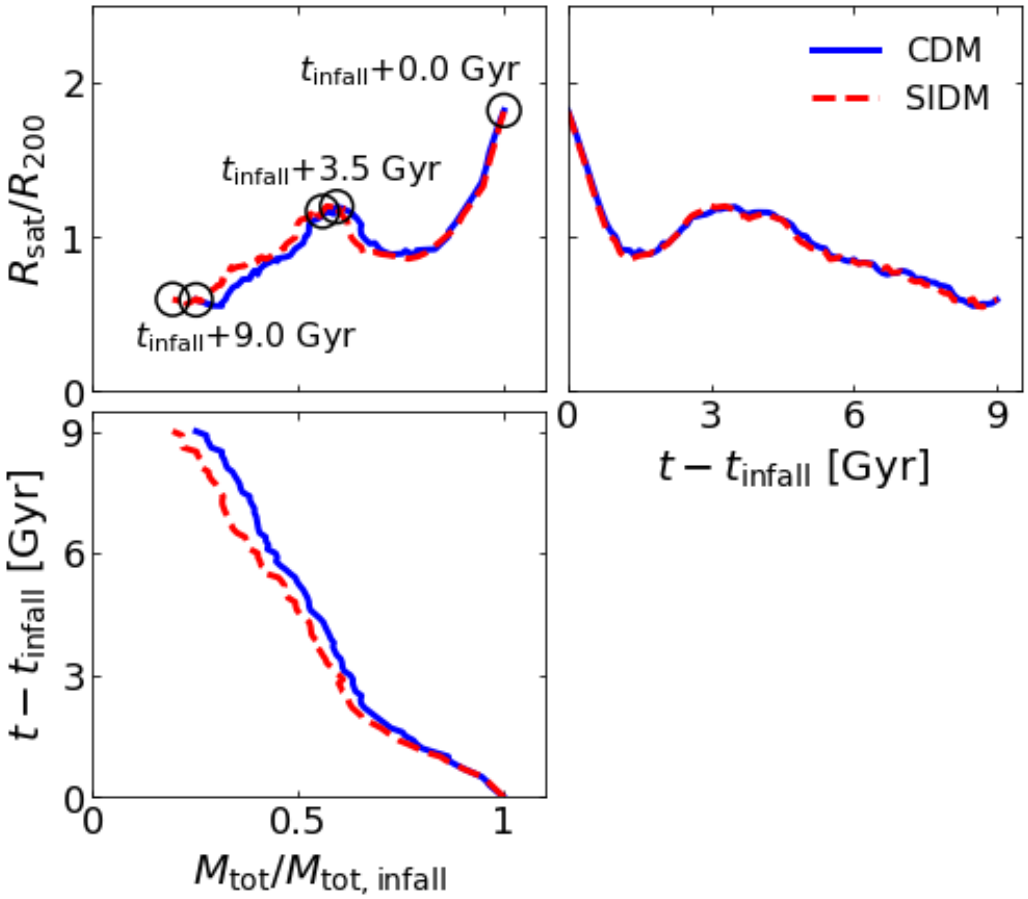


Figure 2.3: The median evolution since infall of cluster member galaxies with $M_\star \gtrsim 10^7\,\mathrm{M}_\odot$ at $z = 0$, in the CDM (solid blue) and SIDM (dashed red) versions of cluster CE-05. *Top left*: cluster-centric distance in units of $R_{200}$ *versus* galaxy mass in units of galaxy mass at infall. The labels indicate the time passed since infall, and the corresponding points on both tracks are encircled. *Top right*: cluster centric distance in units of $R_{200}$ as a function of time since infall. *Bottom right*: time since infall as a function of galaxy mass in units of the galaxy mass at infall. Note that a different number of galaxies contribute to the median at every point on the plot.

have lost $\sim 75$ per cent of their mass, while SIDM galaxies have lost $\sim 80$ per cent. However, we find no difference between the typical orbits of CDM and SIDM galaxies that survive to $z = 0$ (top right panel of figure 2.3; we shall later see very slight differences in the distribution of galaxies that do not survive).

Results are indistinguishable (but noisier) for galaxies with $M_\star \gtrsim 10^{10}\,\mathrm{M}_\odot$. Results are also very similar in CE-12, where CDM galaxies have lost 80 per cent of their mass after 9 Gyr, and SIDM galaxies have lost 90 per cent.

## 2.4 Observable differences between cluster galaxies in CDM and SIDM

We saw in section 2.3 that a galaxy made of SIDM has a higher rate of DM loss than an identical galaxy made of CDM. However, observations of the real Universe do not have the luxury of matched comparisons to a control sample or null test. In this section we investigate whether the increased rate of mass loss has observable effects on the population of galaxies in a cluster at $z = 0$.

### 2.4.1 Stellar-to-halo mass relation

At the mass scale of individual galaxies, the SHMR of field galaxies is indistinguishable between CDM and SIDM simulations (figures 2.4 and 2.5). This is expected because efficient gas cooling and star formation ensure that a baryon-dominated core retains a deep gravitational well (Robertson et al., 2019). Once a galaxy falls into a cluster, tidal forces preferentially remove DM, which is more diffuse than stars.

We first investigate the SHMR for matched pairs of galaxies with more than 10 star particles at $z = 0$ (figure 2.4). On average, SIDM cluster galaxies ended up with $\sim$0.12 dex (25 per cent) lower masses than their CDM counterparts. This effect increases to $\sim$0.2 dex (35 per cent) for the most massive cluster member galaxies.

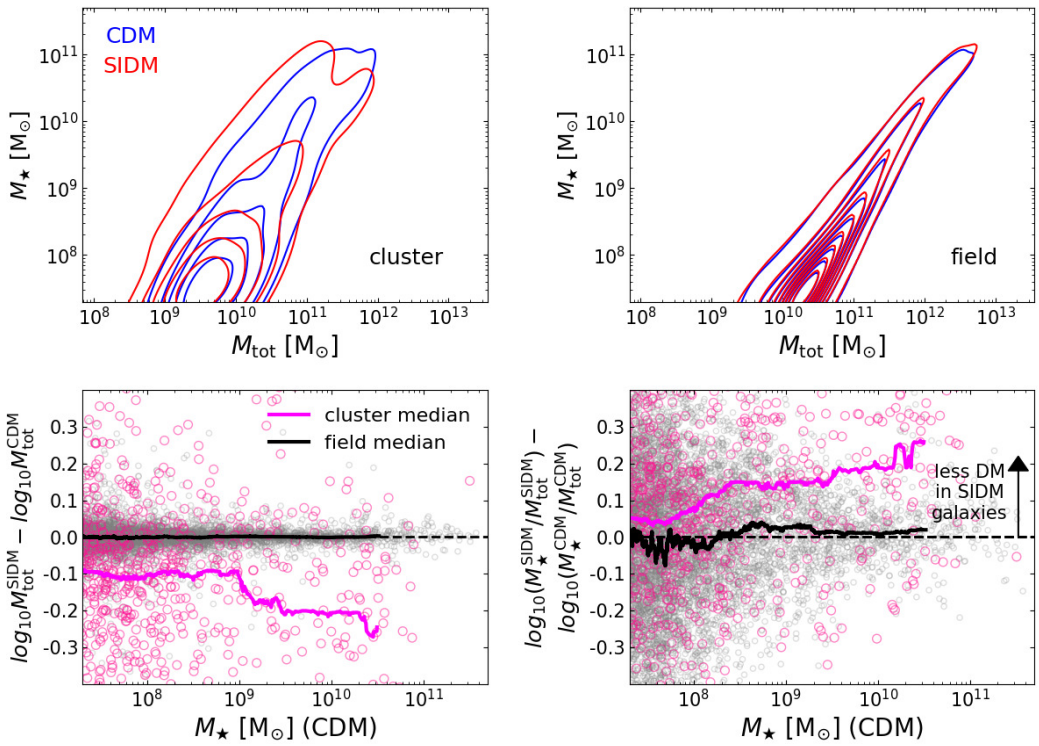


Figure 2.4: Stellar-to-halo mass relation for galaxy pairs with >10 star particles ($M_\star \gtrsim 5 \times 10^7\,\mathrm{M}_\odot$), matched between CDM and SIDM simulations. *Top panels*: number-density contours of the stellar mass *versus* total mass in cluster CE-12 (left) and the field (right). Both are smoothed with the same circular Gaussian kernel of width $\sigma = 0.35$ dex: the increased scatter inside a cluster is real. A version for all (including unmatched) galaxies looks qualitatively similar. *Bottom panels*: the difference in total mass (left) and stellar-to-halo mass (right) between the SIDM and CDM galaxy populations. Pink points show matched galaxy pairs in cluster CE-12, with the running median overlaid; grey points show pairs in the field. The effect of SIDM is greatest for more massive galaxies.

We then fit the Moster et al. (2013) relation, as described in section 2.2.3. We fit all galaxies, not just those matched between simulations (as would be done with observational data). Because this adds some almost-stripped galaxies, this raises the normalisation of the SHMR at low masses by a factor ∼1.5 for both CDM and SIDM, and moves the location of the turnover within its (considerable) statistical uncertainty. The best fits are shown in figure 2.5, and the best fitting parameters are listed in table 2.3.

We find that the SHMR of SIDM field galaxies is indistinguishable from that of CDM field galaxies, within the precision possible using our limited number of simu-

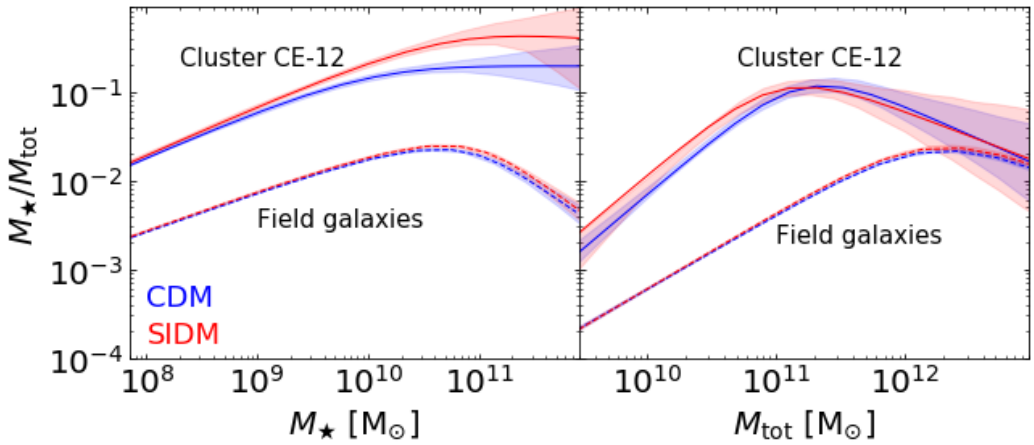


Figure 2.5: Fits to the SHMRs using equation 2.2. *Left*: the SHMR as a function of stellar mass. Fits to the cluster galaxies in CE-12 are shown as solid lines, and to field galaxies fits as dashed lines. Blue and red lines represent the CDM and SIDM versions of a given simulation respectively. Shaded regions represent the 68 per cent confidence regions, obtained from the $16^{\rm th}$ and $84^{\rm th}$ percentiles of the MCMC chain. *Right*: similar to the left panel, but now for the SHMR as a function of halo mass. The fits to the galaxies in CE-05 are similar but noisier, because that cluster has fewer member galaxies.

lated galaxies. This is expected since field galaxies are dominated by stars, and very inefficiently affected by SIDM interactions. The SHMR for SIDM cluster galaxies is also well fit using the functional form of Moster et al. (2013), but with different best-fit parameters to the CDM cluster galaxies.

The SHMRs for CDM and SIDM cluster galaxies are distinguishable at the high mass end, when binning by stellar mass. Fortunately, it is possible to measure this observationally. We find that cluster galaxies within $2R_{200}$ with stellar mass $10^{10-11}\,{\rm M}_\odot$ have $M_\star/M_{\rm tot}$ 8 times higher than field galaxies in a CDM universe, but 13 times higher in an SIDM universe, or in other words, the SHMR of the SIDM galaxies is $\log_{10}(13/8)\sim0.21$ dex above the SHMR of the CDM galaxies. For cluster galaxies within $R_{200}$, we find that these numbers increase to 10 and 20 (the best fitting parameters are included in table 2.3, but the fits are not shown on figure 2.5). There is considerable scatter, $\sigma_{\rm M}\sim0.4$ dex, in the SHMRs at these masses. To distinguish the SHMRs at $3\sigma$, the scatter needs to be less than $0.21/3 = 0.07$ dex. From $0.4/\sqrt{N} < 0.07$ we find that it would require noise-free measurements, e.g.

Table 2.3: The best fit parameters of the SHMR 2.2 for field galaxies and for cluster galaxies within $2R_{200}$ and $R_{200}$ of CE-12. The 68 per cent confidence intervals are the difference between the $16^{\text{th}}$ and $84^{\text{th}}$ percentiles of the marginalized 1D posteriors.

| Field galaxies | | | | |
|---|---|---|---|---|
| | Fit to $M_{\rm tot}(M_\star)$ | | Fit to $M_\star(M_{\rm tot})$ | |
| | CDM | SIDM | CDM | SIDM |
| $\log_{10}M_1$ | $12.09^{+0.06}_{-0.05}$ | $12.11^{+0.06}_{-0.05}$ | $12.22^{+0.05}_{-0.05}$ | $12.2^{+0.05}_{-0.05}$ |
| $N$ | $0.022^{+0.001}_{-0.001}$ | $0.024^{+0.001}_{-0.001}$ | $0.021^{+0.002}_{-0.001}$ | $0.023^{+0.002}_{-0.001}$ |
| $\beta$ | $0.81^{+0.02}_{-0.02}$ | $0.81^{+0.02}_{-0.02}$ | $0.84^{+0.02}_{-0.02}$ | $0.86^{+0.02}_{-0.02}$ |
| $\gamma$ | $0.46^{+0.04}_{-0.04}$ | $0.48^{+0.04}_{-0.04}$ | $0.6^{+0.07}_{-0.07}$ | $0.57^{+0.07}_{-0.07}$ |
| $\sigma_{\rm M}$ | $0.215^{+0.002}_{-0.002}$ | $0.214^{+0.002}_{-0.002}$ | $0.278^{+0.003}_{-0.003}$ | $0.281^{+0.003}_{-0.003}$ |

| Cluster galaxies ($R < 2R_{200}$) | | | | |
|---|---|---|---|---|
| | Fit to $M_{\rm tot}(M_\star)$ | | Fit to $M_\star(M_{\rm tot})$ | |
| | CDM | SIDM | CDM | SIDM |
| $\log_{10}M_1$ | $10.55^{+0.29}_{-0.24}$ | $10.84^{+0.47}_{-0.35}$ | $11.23^{+0.2}_{-0.24}$ | $11.02^{+0.25}_{-0.23}$ |
| $N$ | $0.1^{+0.02}_{-0.03}$ | $0.26^{+0.37}_{-0.11}$ | $0.11^{+0.02}_{-0.02}$ | $0.26^{+0.29}_{-0.03}$ |
| $\beta$ | $1.27^{+0.13}_{-0.17}$ | $1.27^{+0.13}_{-0.22}$ | $1.23^{+0.15}_{-0.18}$ | $1.27^{+0.15}_{-0.18}$ |
| $\gamma$ | $0.0^{+0.19}_{-0.19}$ | $0.08^{+0.42}_{-0.43}$ | $0.66^{+0.34}_{-0.39}$ | $0.08^{+0.37}_{-0.4}$ |
| $\sigma_{\rm M}$ | $0.36^{+0.01}_{-0.01}$ | $0.41^{+0.01}_{-0.01}$ | $0.62^{+0.02}_{-0.02}$ | $0.41^{+0.03}_{-0.03}$ |

| Cluster galaxies ($R < R_{200}$) | | | | |
|---|---|---|---|---|
| | Fit to $M_{\rm tot}(M_\star)$ | | Fit to $M_\star(M_{\rm tot})$ | |
| | CDM | SIDM | CDM | SIDM |
| $\log_{10}M_1$ | $10.53^{+0.29}_{-0.25}$ | $10.85^{+0.3}_{-0.2}$ | $11.25^{+0.29}_{-0.25}$ | $11.01^{+0.43}_{-0.43}$ |
| $N$ | $0.14^{+0.04}_{-0.04}$ | $0.5^{+0.34}_{-0.14}$ | $0.16^{+0.04}_{-0.05}$ | $0.5^{+0.92}_{-0.09}$ |
| $\beta$ | $1.28^{+0.14}_{-0.18}$ | $1.37^{+0.12}_{-0.17}$ | $1.19^{+0.2}_{-0.24}$ | $1.37^{+0.35}_{-0.55}$ |
| $\gamma$ | $0.06^{+0.21}_{-0.2}$ | $0.41^{+0.38}_{-0.26}$ | $0.6^{+0.44}_{-0.51}$ | $0.41^{+0.53}_{-0.63}$ |
| $\sigma_{\rm M}$ | $0.34^{+0.01}_{-0.01}$ | $0.39^{+0.01}_{-0.01}$ | $0.62^{+0.03}_{-0.03}$ | $0.39^{+0.04}_{-0.04}$ |

from galaxy-galaxy strong lensing, of ∼32 cluster galaxies to distinguish between these values at $3\sigma$, assuming that the SHMR for field galaxies is well known. It would be more challenging to measure other quantities like the slope of the SHMR at low masses, or the position of the turnover, because these vary by less than five per cent with different DM.

### 2.4.2 The stripping factor

Another measure used to express the mass lost from cluster galaxies is the 'stripping factor' (Niemiec et al., 2019)

$$\tau_{\rm strip}(M_\star) \equiv 1 - \frac{\widetilde{M}_{\rm tot,cluster}(M_\star)}{\widetilde{M}_{\rm tot,field}(M_\star)}, \tag{2.5}$$

where $\widetilde{M}_{\rm tot,cluster}(M_\star)$ and $\widetilde{M}_{\rm tot,field}(M_\star)$ are the median total mass of cluster and field galaxies in a bin of stellar mass $M_\star$. This definition is motivated by a model in which a galaxy's star formation is quenched as it enters a cluster. Since no new stars are formed, field galaxies of a given stellar mass act as the progenitors of cluster galaxies with the same stellar mass.

We split our sample of cluster (CE-12) and field galaxies into logarithmic bins of 1 dex in stellar mass ranging from $10^6$ to $10^{11}$ $\mathrm{M}_\odot$, and calculate the stripping factor in each bin; the result is shown in figure 2.6. The errors on the stripping factors are calculated using bootstrapping.

The difference between CDM and SIDM is not significant in this measure, although the largest hint of a difference again appears to be in galaxies with high stellar mass. The mean stripping factor of galaxies inside $2R_{200}$ at $z = 0$ is $0.86 \pm 0.03$ and $0.87 \pm 0.04$ for the CDM and SIDM version of cluster CE-12 respectively (blue solid and red dashed horizontal lines in figure 2.6), and there is little scatter about this value in the different stellar mass bins. For massive galaxies with $10^{10} < M_\star < 10^{11}\mathrm{M}_\odot$, the mean stripping factor for SIDM is $\mathcal{O}(10^{-2})$ higher than for CDM, but this is much smaller than statistical uncertainty. More stripping occurs in the inner parts of the cluster, and the stripping factors rises to $0.88 \pm 0.03$ and $0.90 \pm 0.05$ for galaxies inside $R_{200}$. Again there is little hope for observational discrimination.

Stripping factors are reduced in the lower mass cluster CE-05, to $0.83 \pm 0.04$ and $0.85 \pm 0.04$ for the CDM and SIDM versions of galaxies within $2R_{200}$ with again little scatter about these values. A more massive cluster seems to increase slightly both the stripping of mass and the effect of self-interactions.

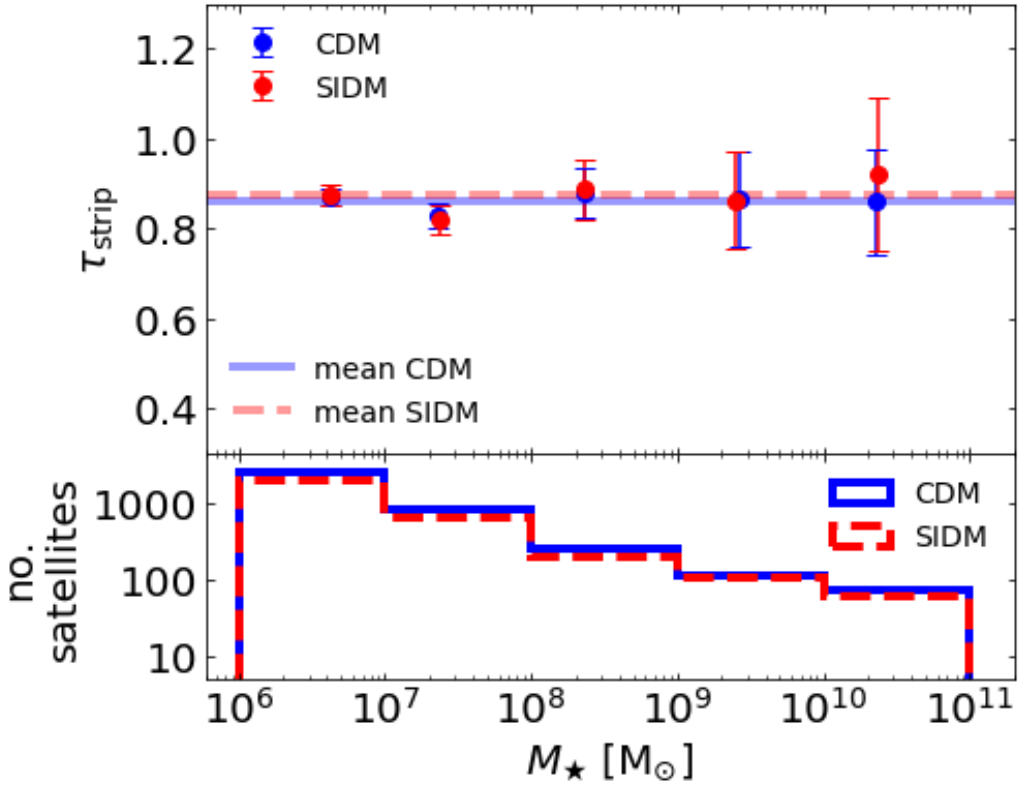


Figure 2.6: *Top*: $\tau_{\rm strip}$ (equation 2.5) as a function of mean stellar mass in five stellar bins. The results for the galaxies in the CDM and SIDM versions of CE-12 are plotted in blue and red respectively. The horizontal solid blue and dashed red line are the mean $\tau_{\rm strip}$ of the CDM and SIDM galaxies respectively. The mean stripping factor has a value of $0.86 \pm 0.03$ and $0.87 \pm 0.04$ for the CDM and SIDM satellites respectively. The results for cluster CE-05 are $0.83 \pm 0.04$ and $0.85 \pm 0.04$. *Bottom*: histogram of number of galaxies in the same five stellar bins as plotted in the top panel. Again blue represents CDM and redSIDM.

### 2.4.3 The number and radial distribution of cluster galaxies

There are ∼20 per cent fewer member galaxies in the SIDM version of a given cluster at $z = 0$ (table 2.1). Most of the discrepancy is in the central $\sim 100\,$kpc, which is also where the most disruption takes place of SIDM galaxies whose CDM counterparts survive (Figure 2.7). This is consistent with our earlier findings that SIDM barely changes the orbits of galaxies, but makes them more susceptible to disruption (section 2.3.2). Cluster outskirts contain similar numbers of galaxies, with the populations continually replenished by objects infalling from the field.

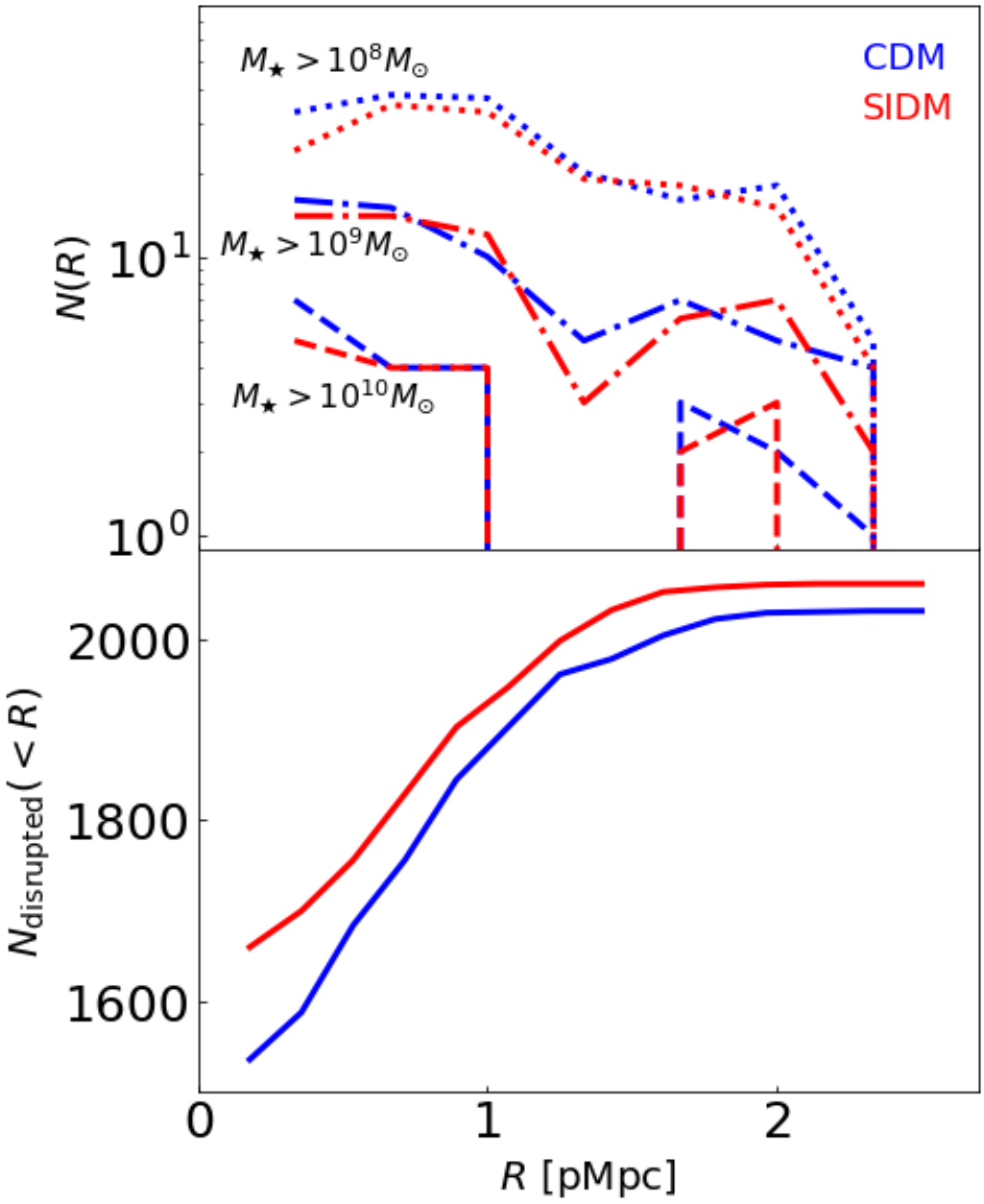


Figure 2.7: *Top*: The radial distribution of galaxies that survive until $z = 0$, in CDM and SIDM versions of cluster CE-05. The only useful difference is the slight reduction of SIDM galaxies inside the cluster core. *Bottom*: The last known location of galaxies that did not survive until $z = 0$. Cumulative number of galaxies inside a given radius, in the simulation snapshot immediately before they were disrupted.

It would be difficult to distinguish between CDM and SIDM using cluster richness, given the intrinsic scatter in the mass-richness relation (Simet et al., 2017; Murata et al., 2019; Hilton et al., 2021). It is probably also difficult to distinguish between CDM and SIDM using the radial distribution of cluster galaxies. We find that 33 per cent and 36 per cent of galaxies reside inside $0.5R_{200}$ in the CDM version of clusters CE-05 and CE-12, compared to 30 per cent and 26 per cent in the SIDM versions. More simulations are needed to determine the population mean and intrinsic scatter, but the difference is likely to be washed out by projection effects (of outlying members in front of/behind the cluster core, and field galaxies onto cluster outskirts).

## 2.5 Discussion and conclusions

We studied the effects of self-interactions on the mass stripping of galaxies as they fall into galaxy clusters by comparing cosmological simulations with and without DM self-interactions. When a galaxy falls into a cluster, DM interactions accelerate the rate of mass stripping. Over 33 per cent of galaxies in an SIDM cluster can be entirely disrupted by the present time, compared to 20 per cent in a CDM cluster. Unfortunately, the disrupted galaxies (which are the most different between CDM and SIDM) are no longer observable. The orbits of surviving galaxies are essentially unchanged, and disrupted galaxies are continually replaced by new ones falling into the cluster. When comparing matched galaxies between the CDM and SIDM versions of a given cluster (section 2.3), we find significant differences in mass loss. However, when we only look at the population of galaxies remaining in the cluster at $z = 0$ (section 2.4), we find considerably smaller differences. SIDM galaxies are more susceptible to disruption, so there is a large group of disrupted SIDM galaxies which does not contribute to the signal at $z = 0$.

Potentially observable ways to discriminate between CDM and SIDM include the (high mass normalisation of the) stellar-to-halo mass relation of galaxies in clusters,

compared to galaxies in the field, or the stripping factor, both of which describe the mass of the DM in a galaxy of fixed stellar mass. We found a 25 per cent increase in the ratio of stellar-to-total mass of SIDM galaxies with stellar mass $M_\star > 5 \times 10^7\,\mathrm{M}_\odot$. The absolute normalisation of the relation is likely to be needed to discriminate SIDM from CDM, but this depends to some extent on the subgrid physics of the simulations. However, as in the field the relation is nearly indistinguishable for a CDM and SIDM universe, one could use the difference between the field and cluster relations at a given stellar mass to try and discriminate between the two models. From the left panel of figure 2.5, we find that, at approximately the stellar mass of the MW, $10^{10.5}\,\mathrm{M}_\odot$, the ratio $M_\star/M_{\rm tot}$ is 8 and 13 times higher in the cluster compared to the field for the CDM and SIDM versions of CE-12 respectively.

Previous, DM-only simulations (Bhattacharyya et al., 2022) predicted larger differences between SIDM and CDM, probably because of the way stars were assigned to galaxies after the simulation using a semi-analytic model. In DM-only SIDM simulations subhaloes form cores more easily than when baryons are included, making them more easily disrupted. In contrast, our simulations co-evolved a population of baryons and SIDM. In the full hydrodynamical simulation a large number of cluster galaxies fail to form cores or have their cores re-contracted by baryons, and so they are more durable.

We simulated a velocity-independent SIDM cross-section. As galaxies all orbit at the roughly the velocity dispersion of the cluster, they would experience the same effective cross-section even if a velocity dependence was introduced. However, the scattering rate of DM-DM interactions in the galaxy itself would be different for subhaloes of different masses. As internal scattering can change the structure of galaxy halos, tidal stripping could act differently at different masses. To test the effects on the SHMR, simulations would need to be run with a velocity dependence.

In the future, it would be informative to simulate more SIDM clusters (with and without velocity-dependence). While the C-EAGLE suite comprises 30 simulated

CDM clusters, only two have been re-run with SIDM. It is also important to note that a cross-section of $\sigma/m = 1$ cm$^2$g$^{-1}$ has arguably already been ruled out at the $\mathcal{O}(1000\,\mathrm{km\,s^{-1}})$ collision velocities between particles typical in clusters. Performing the same tests with simulations for a lower cross-section would presumably produce smaller differences and would require even higher signal-to-noise observations. Future surveys, such as the Euclid telescope (Laureijs et al., 2011), Rubin Legacy Survey of Space and Time (LSST; LSST Science Collaboration et al. 2009), SuperBIT (Romualdez et al., 2018), and JWST Cosmos-Webb survey (C. Casey & J. Kartaltepe, pers. comm. 2021) will provide data with higher signal-to-noise than ever before, potentially making such tests possible.

We will discuss the SuperBIT telescope in detail in chapter 4.

Chapter 3

# Merging clusters as a test-bed for self-interacting dark matter

*"Any view of things that is not strange, is false."*

— Neil Gaiman, *The Sandman*

## 3.1 Introduction

In a ΛSIDM universe, merging galaxy clusters could potentially act as 'DM colliders'. These mergers are defined by three major components: galaxies, which act as collisionless test particles, gas, which is dissociated from the galaxies through ram pressure stripping, and DM. If the DM is collisionless, it should remain coincident with the cluster galaxies. However, if the DM is able to interact, it can be offset from the galaxies due to drag from the DM self-interactions, with the trajectory of the DM set by the fundamental forces acting on it. Observationally, the stars are visible in a smoothed map of their optical emission, while the diffuse gas between galaxies is visible in X-ray emission. The DM can be located via weak gravitational lensing (see section 1.3.1).

The Bullet cluster (1E0657-558) is the most well-known observational example of two colliding clusters (post-collision). The 'Bullet' refers to the smaller cluster, presently moving away from the main cluster. Early attempts to use the Bullet Cluster to constrain the collisional nature of DM (Markevitch et al., 2004) found

that $\sigma/m < 5\,\mathrm{cm}^2/\mathrm{g}$ from limits on the offset between the DM and stars. This constraint, derived from analytical toy models, was improved by Randall et al. (2008) who ran N-body simulations of Bullet Cluster-like systems with SIDM. Combined with tighter constraints on any DM–galaxy separation (Bradač et al., 2008), they found $\sigma/m < 1.25\,\mathrm{cm}^2/\mathrm{g}$. Robertson et al. (2017a) relaxed this constraint to $\sigma/m < 2\,\mathrm{cm}^2/\mathrm{g}$ by using the first fully hydrodynamical simulations of the Bullet cluster.

Investigating an ensemble of mergers could possibly drive these upper limits down even further. The average DM particle velocity increases with halo mass, and so such detections could also characterise the velocity-dependence of the interaction and possibly constrain the mass of the particle acting as the mediator. By statistically combining observational measurements of major and minor mergers, Harvey et al. (2015) set a limit to the cross-section of $\sigma/m < 0.47\,\mathrm{cm}^2/\mathrm{g}$. However, by reweighting the offset measurements, to account for the fact that some orientations with respect to our line-of-sight should have more statistical power than others, Wittman et al. (2018) relaxed this upper limit to $\sigma/m < 2\,\mathrm{cm}^2/\mathrm{g}$.

In this work we focus on cosmological hydrodynamical simulations of galaxy clusters. We aim to compare the offsets between the DM and galaxies in cluster scale haloes of simulations run with CDM and SIDM physics — and to investigate whether the differences in offset would allow for an observational test to constrain the SIDM cross-section.

The rest of this chapter is organised as follows: in section 3.2, we present the simulation suite used in this work; in section 3.3 we discuss our method for measuring the positions of different mass components within a merging galaxy cluster, before applying this method to our simulations in section 3.4. We present our conclusions and discuss our results in section 3.5. We will also briefly discuss the next steps for this project in this final section.

## 3.2 Data

### 3.2.1 The BAHAMAS simulations

We use the BAryons and HAloes of MAssive Systems (BAHAMAS) cosmological simulations (McCarthy et al., 2017). The BAHAMAS project consists of a suite of simulations designed to test the impact of baryonic physics on the interpretation of large-scale structure tests of cosmology. The majority of the simulations are of periodic boxes, 400 $h^{-1}$ Mpc on a side, with $2 \times 1024^3$ particles. The BAHAMAS simulations have been run with differing cosmologies, however, we only use the WMAP 9-year cosmology* simulations (Hinshaw et al., 2013).

BAHAMAS was run with a modified version of the GADGET-3 code that includes radiative cooling, star formation, chemical evolution, and stellar and AGN feedback. For the simulations run with WMAP 9-year cosmology, the DM particle mass is $5.5 \times 10^9\,\mathrm{M}_\odot$, the initial gas particle mass is $1.1 \times 10^9\,\mathrm{M}_\odot$, and the gravitational softening length is fixed to 5.7 kpc in physical coordinates below $z = 3$ and fixed in comoving coordinates at higher redshifts.

The BAHAMAS boxes have been re-simulated from identical initial conditions in a ΛSIDM universe (see Robertson et al. 2018 for more details) with three different cross-sections. Our implementations of SIDM assume isotropic, velocity-independent interaction cross sections of $\sigma/m = [0.1, 0.3, 1]\,\mathrm{cm}^2\,\mathrm{g}^{-1}$. Similar to the project described in chapter 2, $\sigma/m = 1\,\mathrm{cm}^2\,\mathrm{g}^{-1}$ was chosen as the largest cross-section as it is around the current allowed upper limit, maximizing the observable consequences. From now on we shall refer to the SIDM runs with $\sigma/m = [0.1, 0.3, 1]\,\mathrm{cm}^2\,\mathrm{g}^{-1}$ as SIDM0.1, SIDM0.3, and SIDM1 respectively. The DM self-interactions were implemented using the same method as described in section 2.2.1 and in Robertson et al. (2017b): during each simulation timestep, $\Delta t$, DM particles

*$\Omega_\mathrm{m} = 0.2793$, $\Omega_\mathrm{b} = 0.0463$, $\Omega_\Lambda = 0.7207$, $\sigma_8 = 0.812$, $n_s = 0.972$ and $h = 0.700$.

scatter elastically off neighbours within radius $h_{\rm SI} = 5.7\,{\rm pkpc}$ with probability

$$P_{\rm scat} = \frac{(\sigma/m)\ m_{\rm DM}\, v\, \Delta t}{\frac{4}{3}\pi h_{\rm SI}^3}\ , \tag{3.1}$$

where $v$ is the particles' relative velocity and $m_{\rm DM}$ the DM particle mass.

### 3.2.2 Merging cluster sample

For our sample of major cluster mergers, we consider only those clusters that have massive substructures in their vicinity. In particular, we look for any subhaloes within a sphere with a radius of 4 pMpc centred on the centre of potential of a given cluster and with a total mass larger than five per cent of the total cluster mass. We only consider subhaloes around the 300 most massive clusters in each run of the simulations. This method yields around 100 clusters in each simulation with massive substructures, and a total of approximately 135 substructures across all chosen clusters. Table 3.1 provides a summary of the clusters and substructures included in our calculations.

Table 3.1: The number of clusters with a given number of subhaloes with masses larger than 5 per cent of the cluster's total mass within a sphere with a radius of 4 pMpc centred on the centre of potential. We also show the number of clusters with a given number of substructures, and the total number of substructures across all chosen clusters.

| | number of clusters with | | | | | |
|---|---|---|---|---|---|---|
| Simulation | $N_{\rm sub} \geq 1$ | $N_{\rm sub} = 1$ | $N_{\rm sub} = 2$ | $N_{\rm sub} = 3$ | $N_{\rm sub} = 4$ | $N_{\rm sub,tot}$ |
| CDM | 107 | 82 | 20 | 4 | 1 | 138 |
| SIDM0.1 | 103 | 79 | 19 | 2 | 3 | 135 |
| SIDM0.3 | 102 | 76 | 23 | 2 | 1 | 132 |
| SIDM1 | 105 | 83 | 17 | 3 | 2 | 134 |

To see the effects of DM self-interactions on both scales, we will show the results for the main cluster haloes and substructures separately.

### 3.2.3 Calculating offsets

Let us consider a triangle with vertices at the locations of the peaks of the stellar matter, gas, and DM, defining the side connecting the stellar and gas peaks as

the 'base'. We then define the 'intersection point' as the point on the base which intersects the line segment through the DM vertex that is perpendicular to the base. This perpendicular can lie inside or outside the triangle depending on the orientation of the three peaks.

From here on out we will refer to the offset from the stars to the gas as $\delta_{SG}$, and to the offsets from the stars and DM to the intersection point as $\delta_{SI}$ and $\delta_{DI}$ respectively. Figure 3.1 shows a possible configuration of the stellar, gas, and DM peaks, as well as the various offsets discussed above.

The offsets $\delta_{SI}$ and $\delta_{DI}$ can be calculated by taking the dot and cross products of the vector connecting the stars to the gas $\mathbf{r}_{SG}$ with the vector connecting the stars to the DM $\mathbf{r}_{SD}$ respectively, i.e.

$$\delta_{SI} = \pm|\mathbf{r}_{SI}| = \frac{\mathbf{r}_{SG} \cdot \mathbf{r}_{SD}}{|\mathbf{r}_{SG}|}, \tag{3.2}$$

and

$$\delta_{DI} = \pm|\mathbf{r}_{DI}| = \pm\frac{|\mathbf{r}_{SG} \times \mathbf{r}_{SD}|}{|\mathbf{r}_{SG}|}. \tag{3.3}$$

The signs of $\delta_{SI}$ and $\delta_{DI}$ depend on the orientation of the three distributions. By definition $\delta_{SG}$ is positive, and reflects the direction of motion of a merger as the gas is offset from the collisionless stars due to ram pressure. However, in the case that the centre of the stars lies in *between* the centres of the DM and the gas, $\delta_{SI}$ is negative*. The sign of $\delta_{SI}$ is determined by the dot product in equation 3.2. $\delta_{DI}$ being positive or negative does not have a physical meaning, but simply reflects that we have chosen a direction in which $\delta_{DI}$ is positive. In two dimensions, positive $\delta_{DI}$ means the centre of the DM distribution lies above the line connecting the centres of the stars and gas. In three dimensions, positive $\delta_{DI}$ means the centre lies above

*$\beta_{\parallel}$ can also be greater than one in case the DM lags behind the gas and both are offset from the stars in the same direction, i.e. the DM is affected even stronger than the gas.

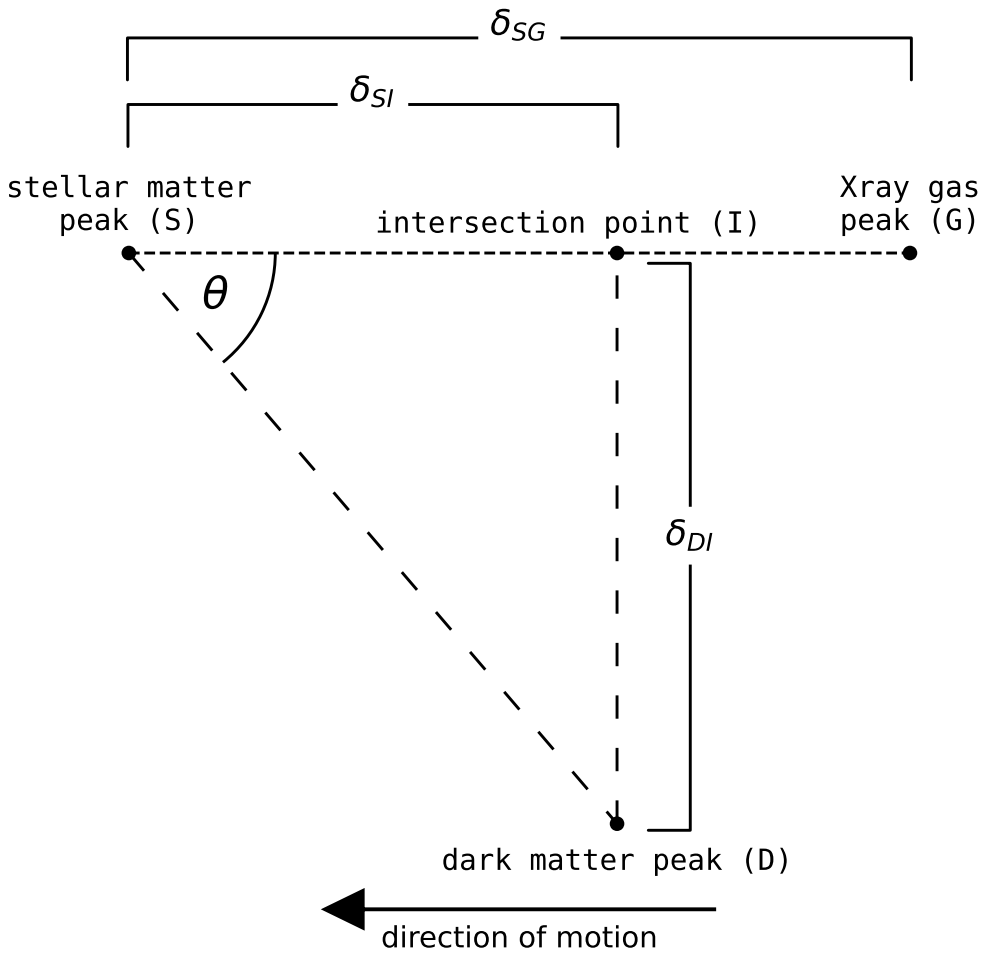


Figure 3.1: A possible configuration of the centres of the distributions during and after a cluster merger. The offset from the stars to the gas is given by $\delta_{SG}$, and to the DM in a parallel and perpendicular direction by $\delta_{SI}$ and $\delta_{DI}$ respectively. The direction of motion is defined by the vector connecting the gas to the stars, as the gas lags behind the stars due to ram pressure.

the plane defined by the triangle connecting the centre of the stars, the centre of the gas, and the origin of our coordinate system.

Similarly to Harvey et al. (2015), we measure the offset of the DM as a fractional lag given by

$$\beta_{\parallel} = \frac{\delta_{SI}}{\delta_{SG}}. \tag{3.4}$$

Using this dimensionless ratio as our measure of the DM offset has two main advant-

ages. First, in two dimensions it removes dependence on the angle of the collision with respect to the line-of-sight. Second, it represents a physical quantity that the analytic (approximate) model of SIDM dynamics in Harvey et al. (2015) suggests should be identical for all merger configurations, at all times during the merger, so measurements from different systems can be averaged*. We will discuss this analytical model in more detail in section 3.4. As a control test, we also measure the perpendicular lag,

$$\beta_{\perp} = \frac{\delta_{DI}}{\delta_{SG}}, \tag{3.5}$$

which should be consistent with zero on average, if the Universe does not have a handedness (and in the absence of systematics). As $\delta_{SG}$ is always positive, and $\delta_{SI}$ and $\delta_{DI}$ can be negative, both $\beta_{\parallel}$ and $\beta_{\perp}$ can be negative as well.

## 3.3 Measuring positions

In order to measure the offsets between different components, we first need a definition of position for each of the components. The methods to find the positions of the gas, galaxies and DM all differ when observing merging clusters. In turn, these methods differ from the methods used to find the positions in simulations of galaxy clusters. In our case, we can use the particle distributions directly to find the positions or centres of each component, a method only accessible for simulated clusters.

### 3.3.1 Shrinking-spheres

We use the *shrinking-spheres* method to determine the centre of a given set of particles (see, e.g., Power et al., 2003). A sphere is drawn centred on the centre of potential as determined by SUBFIND of the halo or subhalo in question, with the

*Figure 6 of Robertson et al. (2017a) suggests that there is actually a gradual increase in $\beta_{\parallel}$ immediately after a merger, but that it then moves towards a steady state.

initial radius chosen to be a third of the equivalent of $R_{200}$ for a given subhalo*. The radius is then shrunk by a factor $f = 0.9$ and the new centre is defined as the centre of mass of all the particles within the current sphere. The radius is shrunk again, and the process iterates until the number of particles within the sphere is equal or less than $N_{\rm min} = 100$. The centre of mass position of all the particles within the final sphere gives the position of a given set of particles. The method is run separately for each (sub)halo and for each different mass component involved in the merger, i.e. the gas, stars, and DM.

This method will give us the centres of the gas, stars, and DM in three dimensions. Observationally, only projected values for these centres are available. We project our three-dimensional centres onto the $x$- and $y$-axes to find two-dimensional equivalents for the centres. We will then use these three- and two-dimensional centres to find the offsets as outlined in section 3.2.3 and calculate $\beta_{\parallel}$ and $\beta_{\perp}$. We will compare our results in the next section.

The shrinking-spheres method can meander away to an incorrect local peak, even if the initial locations of the DM, stars, and gas were identical. To account for this effect, we require that for a given (sub)halo to be included in our calculation of $\beta_{\parallel}$ and $\beta_{\perp}$, the offset between the stars and the gas $\delta_{SG}$ is less than 250 pkpc. We impose this cut in both the two-dimensional and three-dimensional case. Such cuts are also necessary observationally, as the detected peaks in e.g. the stars and gas need to be matched with each other. This is typically done by requiring the stars and gas to be close to each other on the sky, i.e. for the star-gas offset to be below a given threshold.

*I.e. $R_{\rm ini} = \frac{1}{3}\left(M_{\rm sub}/\frac{4}{3}\pi\,\Delta_{\rm c}\,\Omega_{\rm crit}\right)^{1/3}$, where $R_{\rm ini}$ is the initial shrinking-spheres radius, $M_{\rm sub}$ is the mass of the (sub)halo as determined by SUBFIND, $\Omega_{\rm crit}$ the critical density, and $\Delta_{\rm c}$ the overdensity constant, which in our case equals 200.

## 3.4 Results

Figures 3.2 and 3.3 show the distributions of $\beta_{\parallel}$ for the substructures and main haloes respectively. Using the Python module curve_fit, we fitted a one-dimensional Gaussian to each $\beta_{\parallel}$ distribution. The mean of the Gaussian reflects the average (fractional) offset of the DM in each simulation, and the width represents the spread in these offsets. Figure 3.4 shows the best fit values for the widths of the Gaussians fitted to the two- and three-dimensional distributions of $\beta_{\parallel}$ for both the main haloes and the substructures. We find that the widths of the distributions increase with cross-section for both the main haloes and the substructures, with the widths being larger for the main haloes.

This increase in width with cross-section could be the result of wobbles of the BCG (we briefly touched on this in section 1.3.5). It is predicted that during the collision of galaxy clusters with *cored* density profiles, the BCG will be initially offset from the centre of the halo. A constant central density leads to a gravitational potential that is quadratic in radius, and so the offset BCG traces out the motion of a harmonic oscillator long after the halo has relaxed (Harvey et al., 2017; Kim et al., 2017). As a result, when the BCG is observed at a later time, there is a possibility it will be offset. When averaging a large number of systems, the average offset reflects the random phases of these BCG oscillations. Cored density profiles can be induced by DM self-interactions, however, in the CDM paradigm, the central density profile is generally cuspy and hence the BCG will be bound tight to the centre of the DM halo. The increase in width with cross-section could be a reflection of these BCG offsets. The fact that the widths are smaller for the substructures makes sense as we expect subhaloes that are still actively falling into a larger halo to be less affected by long-term wobbles of the galaxies residing inside them.

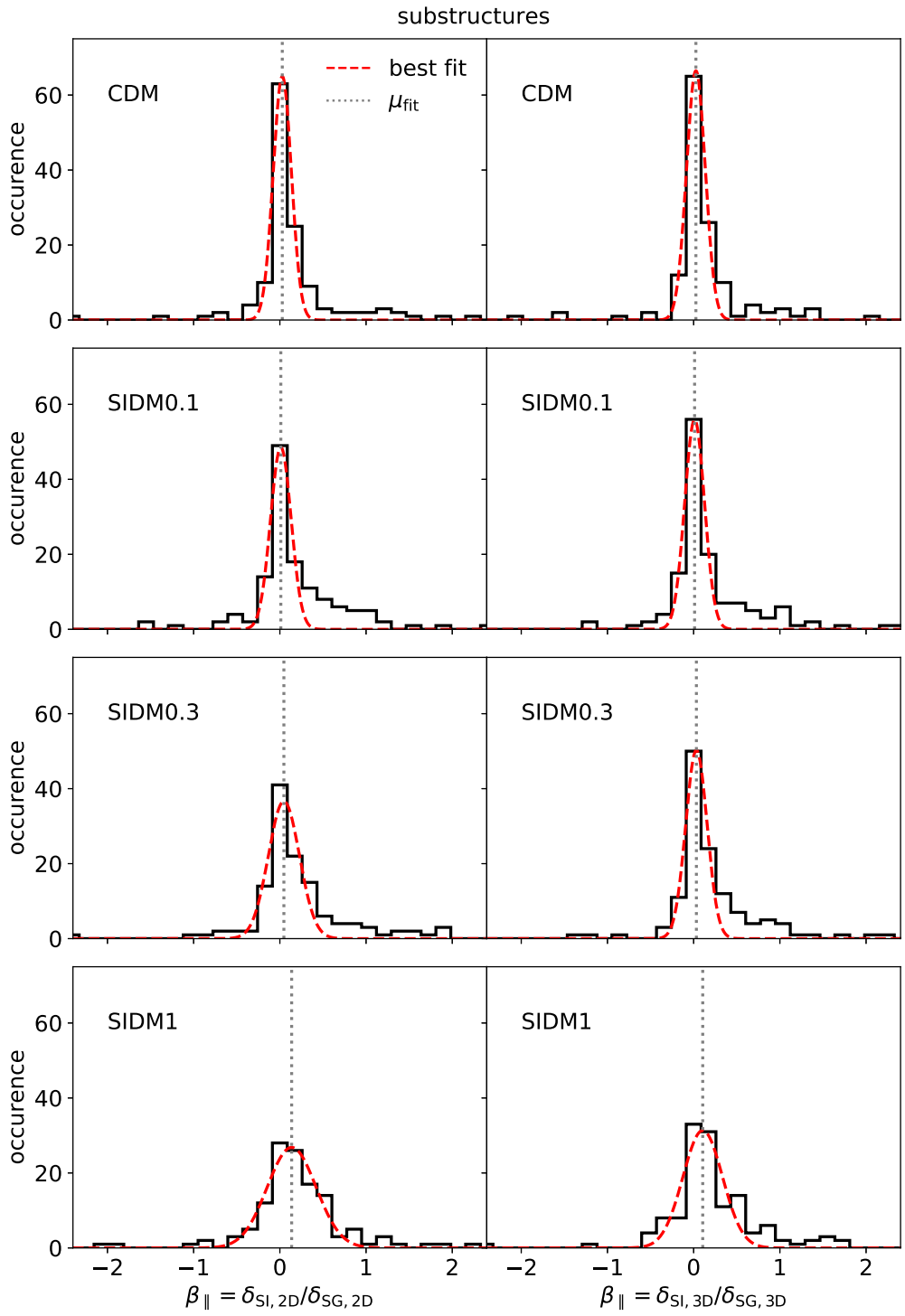


Figure 3.2: The distributions of $\beta_{\parallel}$ for the substructures of our sample of merging haloes defined in section 3.2.2. The red dashed lines are the best fit of a 1D Gaussian to the data. The grey dotted lines are the means of the Gaussians. From the top to the bottom panel, we show the distributions for the CDM, SIDM0.1, SIDM0.3, and SIDM1 simulations respectively. The left and right column shows the distributions of $\beta_{\parallel}$ calculated using the projected and three dimensional values for $\delta_{SI}$ and $\delta_{SG}$ respectively.

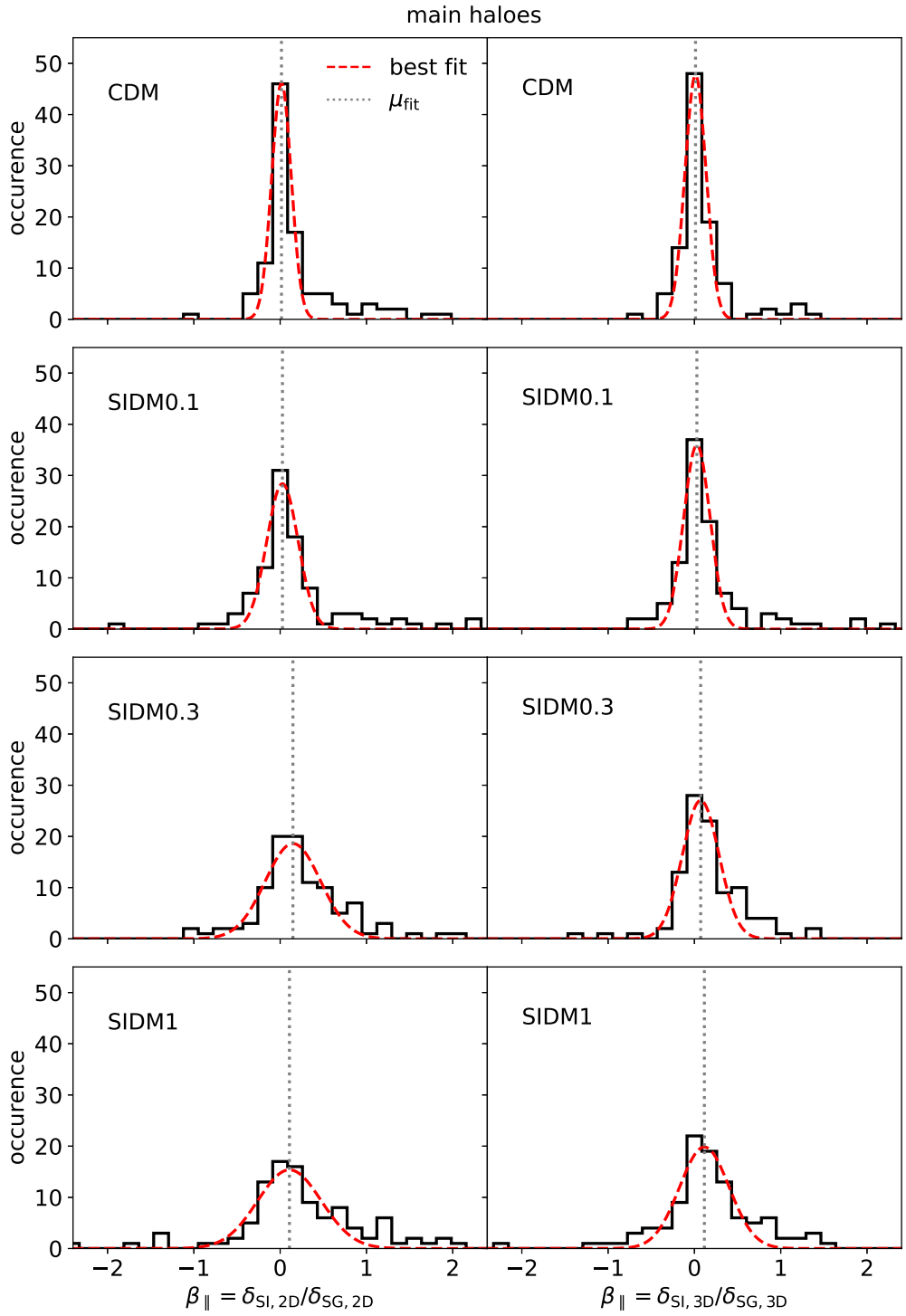


Figure 3.3: Similar to figure 3.2, but we now show the distributions for only the main haloes of the merging sample defined in section 3.2.2. Note that there are fewer main cluster haloes than there are substructures.

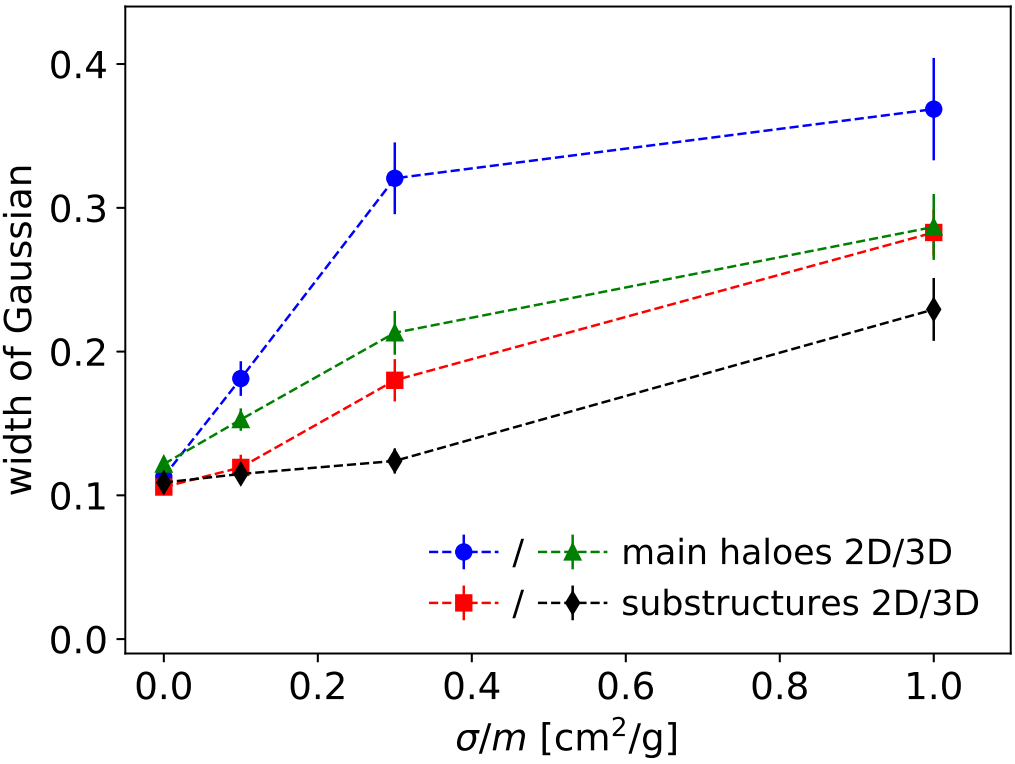


Figure 3.4: The width of the Gaussian fit to the distributions of $\beta_{\parallel}$ as a function of cross-section for the main haloes (blue circles for 2D, green triangles for 3D) and the substructures (red squares for 2D, black diamonds for 3D).

To compare $\beta_{\parallel}$ across the simulations, we calculate the median $\beta_{\parallel}$ of each simulation. We have chosen to use the median as opposed to the mean, as the median is much more robust to outliers. Figure 3.5 shows the median $\beta_{\parallel}$ as a function of DM cross-section. We show $\beta_{\parallel}$ calculated using the full three-dimensional information as well as $\beta_{\parallel}$ calculated using the projected values for $\delta_{SI}$ and $\delta_{SG}$. Note that we have offset the two-dimensional data for clarity. We find that using the projected values for the offsets only slightly increases the median $\beta_{\parallel}$ for each simulation. Using the Python module curve_fit, we also fitted the analytical model given by equation 33 from Harvey et al. (2014) to our results for $\beta_{\parallel}$:

$$\beta_{\parallel} = B\left(1 - e^{-(\sigma/m)/A}\right), \tag{3.6}$$

where $\sigma/m$ is the SIDM cross-section.

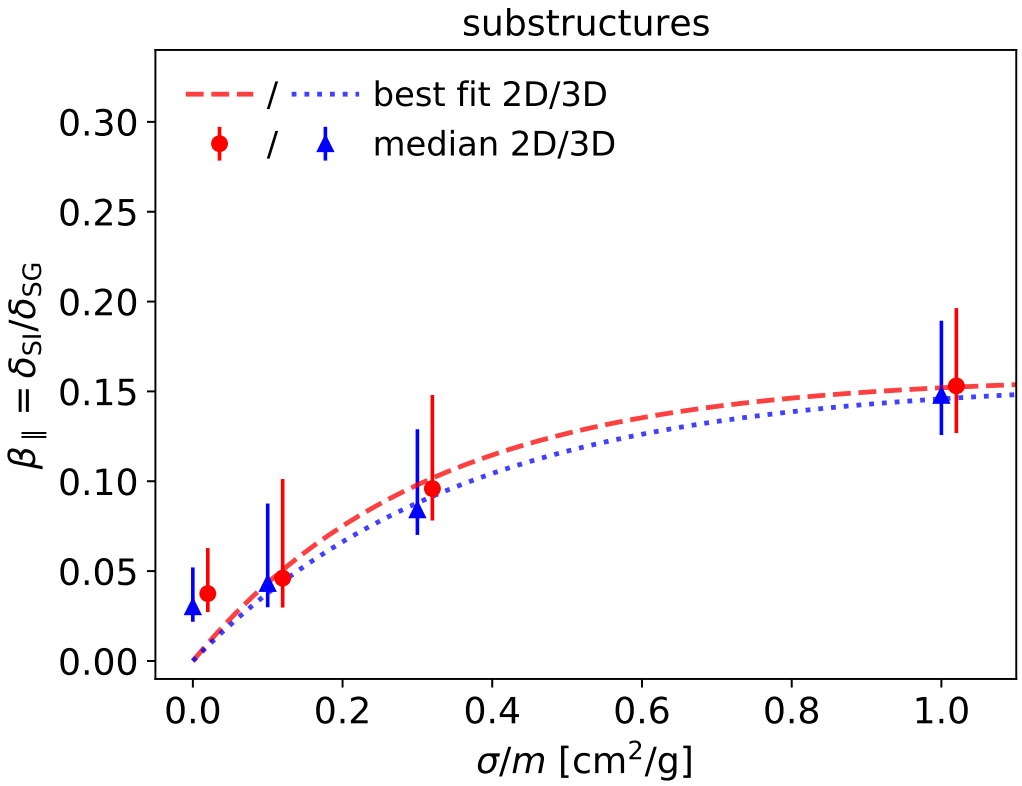


Figure 3.5: $\beta_\parallel$, defined by equation 3.4, as a function of the SIDM cross-section. We show both $\beta_\parallel$ calculated using the 3D information (red circles) as well as $\beta_\parallel$ calculated using the projected values for $\delta_{SI}$ and $\delta_{SG}$ (blue triangles). We have slightly offset the 2D data in the positive $x$-direction for clarity. The dashed red and dotted blue lines are the best fit of model 3.6 to the 2D and 3D data respectively.

The model was derived by interpolating between two well-understood extremes based on optical depth for the drag force acting on the self-interacting DM (see figure 2 of Harvey et al. 2014). For low interaction cross-sections, momentum exchange is slow and there is high preference to forward scattering. The resulting interactions are frequent with a small momentum transfer. In this optically thin regime, the effective drag force has a linear dependence on cross-section. For large cross-sections, the behaviour of the drag is assumed to be similar to that of the drag force acting on the gas: i.e. above a given cross-section threshold, the force depends only on the geometry of the DM substructure. In this optically thick regime, the drag is constant with cross-section. By considering all the forces* acting on the

*I.e. the force of the main cluster potential, the drag on the gas in a subhalo, the drag on the DM in the subhalo due to self-interactions, the force on the gas and galaxies due to the DM

gas, stars, and DM, expressions for $\delta_{SI}$ and $\delta_{SG}$, and thus $\beta_{\parallel}$, can be derived from the equations of motion, finally resulting in equation 3.6.

Essentially, coefficient $B$ in equation 3.6 reflects the relative behaviour of the DM and gas, and $A$ is the characteristic cross-section at which a halo of a given geometry becomes optically thick. From figure 3.5, it seems that the modelled $\beta_{\parallel}$ is still slightly increasing for cross-sections $\sigma/m > 1\,\mathrm{cm}^2\mathrm{g}^{-1}$, and has not plateaued quite yet.

Figure 3.6 shows our control test, the median $\beta_{\perp}$ as a function of cross-section. We show $\beta_{\perp}$ calculated using the full three-dimensional information as well as $\beta_{\perp}$ calculated using the projected values for $\delta_{DI}$ and $\delta_{SG}$. Note that we again have offset the two-dimensional data for clarity. The two-dimensional results are consistent with zero within $1\sigma$, the three-dimensional results within $2\sigma$.

### 3.4.1 Weighting of different mergers to maximise overall signal-to-noise

Wittman et al. (2018) pointed out that some merging systems would have more discriminating power than others. Some systems will have high values of $\delta_{SG}$, either because the motion occurs nearly in the plane of the sky, or the timing of our observations means that separation is maximised. In the presence of roughly constant measurement uncertainty on (DM, stellar, and gas) positions, these systems allow a larger dynamic range for $\delta_{SI}$, and therefore higher signal-to-noise in measurements of $\delta_{SI}$ and $\beta_{\parallel} = \delta_{SI}/\delta_{SG}$.

Assuming that the uncertainty on each offset measurement is approximately the same, i.e. $\sigma_{SG} = \sigma_{SI}$, standard propagation of errors gives for the error on $\beta_{\parallel}$

$$\sigma^2_{\beta_{\parallel}} \propto \frac{1}{\delta^2_{SG}} \left(1 + \frac{\delta^2_{SI}}{\delta^2_{SG}}\right). \tag{3.7}$$

subhalo potential, and the buoyancy on the gas and DM.

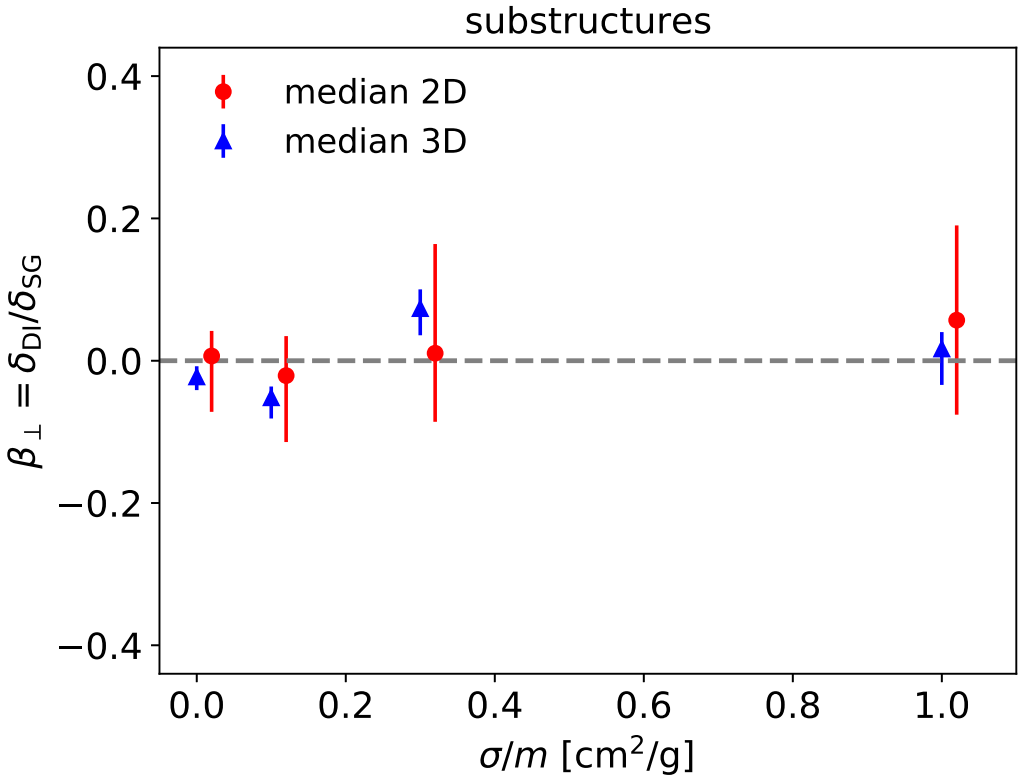


Figure 3.6: $\beta_\perp$, defined by equation 3.5, as a function of the SIDM cross-section. We show both $\beta_\perp$ calculated using the full 3D information (red circles) as well as $\beta_\perp$ calculated using the projected values for $\delta_{DI}$ and $\delta_{SG}$ (blue triangles). The results are consistent with zero within $1\sigma$. We have offset the 2D data in the positive $x$-direction for clarity.

With inverse-variance weighting, the weight $w_i$ of (sub)halo $i$ would be

$$w_i \propto \frac{\delta_{SG,i}^2}{1 + (\delta_{SI,i}/\delta_{SG,i})^2}. \tag{3.8}$$

In most cases $(\delta_{SI,i}/\delta_{SG,i})^2 < 1$, and so $w_i \propto \delta_{SG,i}^2$ is a good approximation (although we use the full expression). As such, with this weighting, substructures with large $\delta_{SG}$ dominantly contribute to the overall results.

In practice, we find that measurement uncertainty (in both simulations and observations) is not approximately constant. In addition to the statistical uncertainty on peak positions, individual merger systems are subject to large systematic uncertainties: for example misidentification of matched stellar and gas peaks that were

coincident at the start of infall, or misidentification of stellar centres (e.g. George et al., 2012) because of foreground galaxies or multiple BCGs). In this regime, the Wittman et al. (2018) prescription gives maximum weight to systems most likely to be systematically incorrect. Indeed, we find that the mean $\langle \beta_{\parallel} \rangle$ (not shown in this work) becomes highly unstable, dominated as it is by the biggest (and possibly incorrect) outliers, and is biased positive.

As a compromise, we here investigate the weighted median $\beta_{\parallel}$, where the weights are given by equation 3.8 (we ignore any proportionality). The weighted median is equal to the weighted $50^{\text{th}}$ percentile, where the weighted $100p^{\text{th}}$ percentile $(0 < p < 1)$ is calculated by sorting the data and finding the smallest set of data for which the weights sum to a fraction $p$ of the total weight.

Figure 3.7 shows the weighted median $\beta_{\parallel}$ calculated using the weighting scheme described above, as well as the best fits of model 3.6 to the weighted medians. Note that we have offset the two-dimensional data for clarity. Compared to the unweighted results, the weighting decreases $\beta_{\parallel}$ across the simulations, except for SIDM0.1 which shows a slight increase. The trend of increasing $\beta_{\parallel}$ with cross-section, however, is still present.

Imposing the cut of $\delta_{SG} < 250\,\text{pkpc}$ only removes one subhalo from the SIDM0.3 simulation. However, including the $\beta_{\parallel}$ of this single subhalo with projected $\delta_{SG} \approx 306\,\text{pkpc}$ and three-dimensional $\delta_{SG} \approx 321\,\text{pkpc}$, raises the weighted medians from 0.05 and 0.09 to 0.59 and 0.86 respectively. Similarly to using the (unweighted) mean of $\beta_{\parallel}$, the weighting scheme gives enormous weight to those (extreme) systems which only just enter the catalogue underneath the pair-matching cuts, and cause the results to become unstable. Including the same subhalo in the unweighted medians makes little difference in both the two- and three-dimensional case.

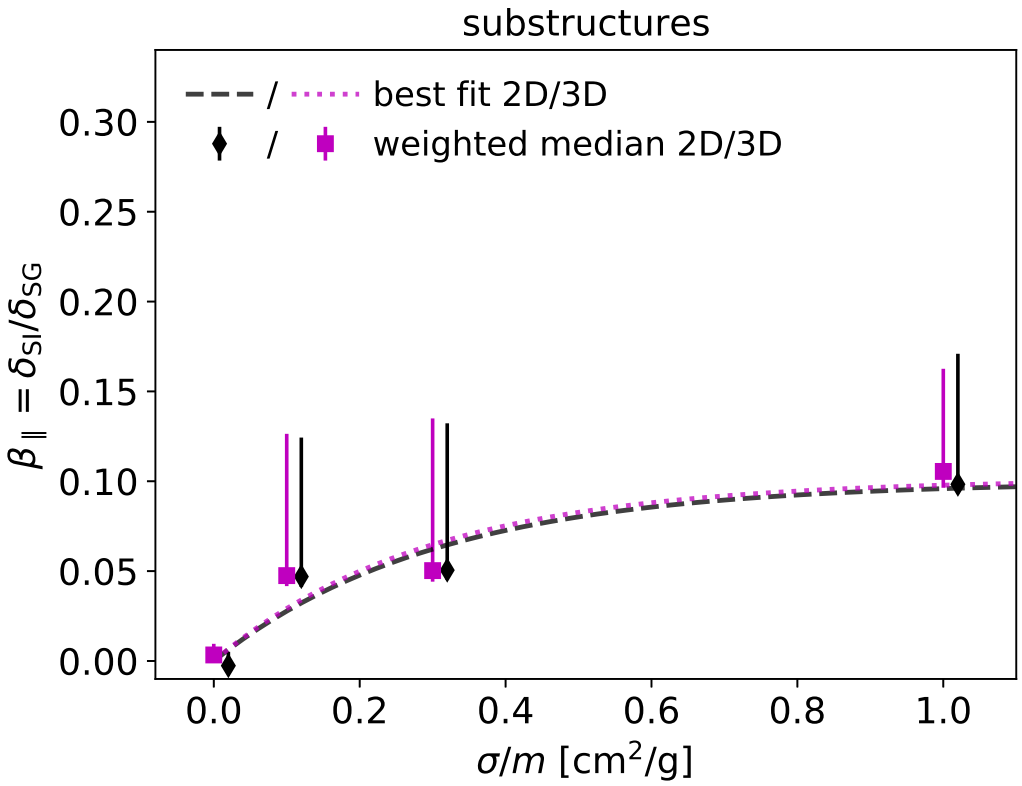


Figure 3.7: Weighted median (black diamonds for 2D, purple squares for 3D) $\beta_\parallel$ as a function of cross-section. The weighting was performed using the scheme outlined in Wittman et al. (2018), which gives greater weight to systems with large $\delta_{SG}$. The dashed black and dotted magenta lines are the best fit of model 3.6 to the 2D and 3D data respectively. We have offset the 2D data in the positive $x$-direction for clarity.

## 3.5 Conclusions and discussion

We have investigated the impact of DM self-interactions on the major mergers of simulated galaxy clusters. In particular, we measured the offset between the DM and the stars as a fractional offset $\beta_\parallel = \delta_{SI}/\delta_{SG}$. We find that $\beta_\parallel$ increases with cross-section, as predicted by the analytic model proposed by Harvey et al. (2014), see equation 3.6. Measurements of $\langle\beta\|\rangle$ could therefore be used as a potentially viable test of SIDM. As a control test, we also measured a perpendicular (fractional) offset $\beta_\perp = \delta_{DI}/\delta_{SG}$, which should on average be zero if the Universe has no handedness. We find that the median $\beta_\perp$ is indeed consistent with zero within

$1\sigma$ in the two-dimensional case, and within $2\sigma$ in the three-dimensional case.

However, there are some challenges with theoretical predictions about this measurement. In particular, we find that the median $\beta_{\|}$ is slightly positive for all values of interaction cross-section, $\sigma/m$ (figure 3.5). From our analytic model 3.6 we expect $\beta_{\|}$ to be consistent with zero for the CDM simulation with zero cross-section. Quirks in individual systems could produce a $\beta_{\|}$ larger than one or a negative $\beta_{\|}$, but we would expect these to average out to zero for an ensemble of haloes. Our positive measurement implies that it is more likely for the DM peak to be located between the stars and the gas than it is for the DM peak to be offset from the stars in the opposite direction of the gas. We speculate that this could be a result of the large gravitational field of the DM, which dominates the potential, pulling on the gas and thus offsetting both the gas and the DM in the same direction from the stars. Unless such an effect can be included in an analytic model, it will probably be necessary to interpret measurements from the real Universe by comparing to full simulations that include this effect.

We also investigated the effects of using the weighting scheme proposed by Wittman et al. (2018), which strongly favors systems with large $\delta_{SG}$, on our median $\beta_{\|}$. We found that our results become unstable when including this weighting. A single subhalo with large $\delta_{SG}$ that was excluded from our results, increases the weighted median $\beta_{\|}$ to nearly 10 times its value in both the two- and three-dimensional case. It seems this weighting scheme is most suitable when clean measurements are available, i.e. when random statistical errors dominate over systematic errors. In our case, the final centre found by the shrinking-spheres method is sometimes the centre of a nearby galaxy, e.g. because the initial radius was too large, resulting in an artificially high $\delta_{SG}$, giving the halo a large weight.

### 3.5.1 Future work

In the future, we wish to extend this work by applying observational techniques to the simulated data and compare our results to those presented here. Depending on the results, we hope to make projections for the upcoming telescope SuperBIT, see sections 4 and 5, which will be launched early 2023.

For the DM, we would use weak gravitational lensing (see section 1.3.1) to find the centre of the distribution. Gravitational lensing depends on the total mass of the object acting as the lens, i.e. we would be using the centre of the total mass as a proxy for the centre of the DM. We intent to follow a method similar to the one outlined in section 3.3 of Robertson et al. (2017a). In short, we would obtain surface density maps of our clusters, and then convert these surface density maps to convergence maps. From the Fourier transform of the convergence, we can derive the Fourier transforms of the shear components. Taking the inverse Fourier transform would then give us the shear components in real space. With a given mass model for our clusters, e.g. an elliptical NFW, we can then fit the shear field to get the centre of the total mass. In terms of the gas, we would obtain surface brightness maps. We could then use peak-finding software such as Source-Extractor (SExtractor; Bertin & Arnouts, 1996) to obtain the centres. The best method to identify the centre of luminous material is to use the BCG (see, e.g., George et al., 2012), which could also be found with SExtractor.

Chapter 4

# The Super-Pressure Balloon-borne Imaging Telescope

*"Seeing, contrary to popular wisdom, isn't believing. It's where belief stops, because it isn't needed any more."*

— Terry Pratchett, *Pyramids*

## 4.1 Introduction

Observing the night sky from outside of the atmosphere eliminates the blurring of images caused by the turbulence of the air, improving the resolution limit of the telescope, particularly important for weak gravitational lensing methods, which rely on accurately measuring the shapes of galaxies. It also allows for observation of the part of the electromagnetic spectrum that is blocked by the atmosphere. While space telescopes provide data of the highest quality, they are much more expensive than ground-based telescopes, and are difficult to maintain. A cheaper and practical alternative are *balloon-borne* telescopes, sub-orbital astronomical telescopes that are suspended on stratospheric balloons, allowing them to be lifted above a large fraction of the Earth's atmosphere.

Balloon-borne telescopes have been used to study the night sky from as early as the 1950s. Launched in 1957, Stratoscope I was the first ever unmanned balloon-borne telescope flown for astronomical research, and was used to study the turbulence

and granulation in the Sun's photosphere (Schwarzschild & Schwarzschild, 1959). Since then many more balloon-borne telescopes have been flown*. Arguably, one of the most of well-known balloon-borne telescopes is the Balloon Observations Of Millimetric Extragalactic Radiation ANd Geophysics experiment (BOOMERang; de Bernardis et al., 1999). It was the first experiment to make large, high-fidelity images of the CMB temperature anisotropies, and is best known for the discovery that the geometry of the Universe is close to flat (de Bernardis et al., 2000) .

The Super-Pressure Balloon-borne Imaging Telescope (SuperBIT) is an upcoming balloon-borne telescope that will be launched in early 2023 for its first fully operational science flight to measure gravitational lensing around ∼200 clusters of galaxies. With a budget of <US$10 million (£7.2 million), SuperBIT costs 100–1000 times less than a space telescope. In this chapter, we will briefly describe the instrument's astronomical background, its hardware and design, and SuperBIT's various engineering test flights that have taken place over the years in preparation for its future science flights.

## 4.2 Astronomical background

During its long duration flights, SuperBIT will float at an altitude of ∼40 km above sea-level, already above 99.2% of the atmosphere (Firanj Sremac & Salehi, 2018), ideal for astronomy in the optical and near-ultraviolet (NUV) bands due to significantly reduced atmospheric interference when compared with ground-based systems. This is illustrated in figure 4.1, which shows the atmospheric transmission as a function of wavelength at various altitudes above sea-level. Particularly for wavelengths below 400 nm, there is significantly reduced atmospheric absorption at balloon float altitude compared to sea-level as well as the ground-based observatory Mauna Kea, which is at an altitude of approximately ∼4.2 km.

*See the StratoCat website https://stratocat.com.ar/ for an extensive overview.

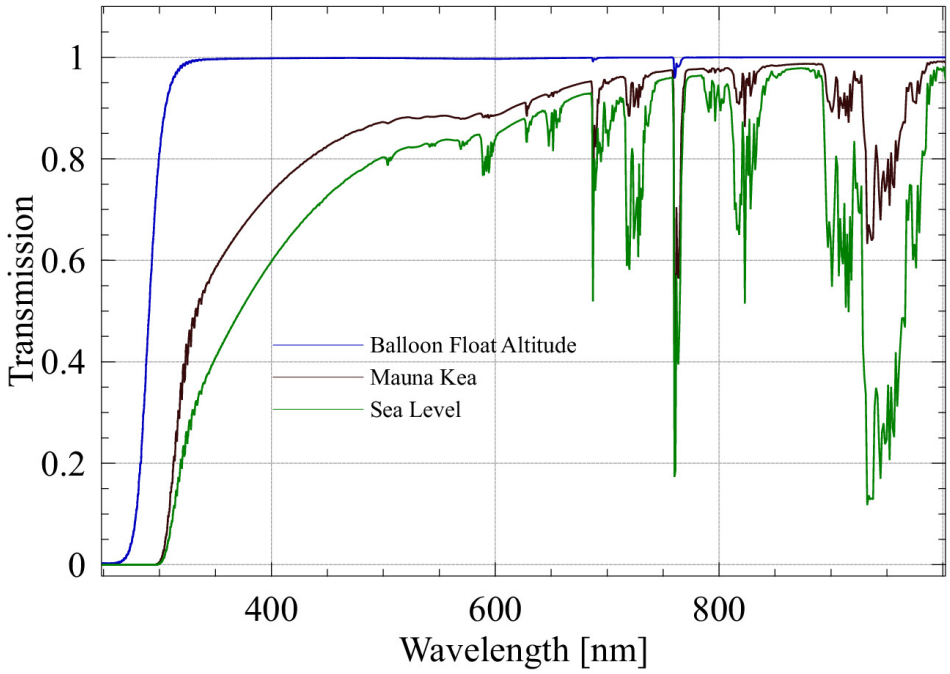


Figure 4.1: Atmospheric transmission as calculated by the MODTRAN4 software (Berk et al., 1999). At the float altitude of SuperBIT, ∼40 km, there is significantly reduced atmospheric absorption compared to sea level and the best land-based telescopes, particularly for wavelengths below 400 nm.

Combining diffraction limited angular resolution of $< 0.3$ arcseconds, extreme stability, space-like backgrounds, and long integrations, the SuperBIT platform will complement current and upcoming surveys like the James Webb Space Telescope (JWST; Gardner et al. 2006), the Euclid telescope (Laureijs et al., 2011), and the Nancy Grace Roman Space Telescope (NGRST; Akeson et al. 2019)*.

In addition, while JWST, Euclid, and NGRST will surpass the Hubble Space Telecope (HST)'s capabilities at red and near-infrared (NIR), after HST's demise there will be effectively no space-based capabilities in the blue and UV. The ground-based Rubin Legacy Survey of Space and Time (LSST; LSST Science Collaboration et al. 2009)†, will observe in the optical and will explore a large volume of the Universe (18,000 deg$^2$ to 27.5 r-band magnitude). However, being a ground-based survey, it will be limited by atmospheric seeing. The above mentioned space-based missions

*This telescope was formerly known as the Wide-Field Infrared Survey Telescope, or WFIRST.

†Originally, the observatory was named the Large Synoptic Survey Telescope, i.e. LSST, and has now been renamed to the Vera C. Rubin Observatory. At present 'LSST' refers to the astronomical survey carried out by the observatory.

will be either shallower or cover a smaller fraction of the sky than LSST. Euclid will have an overlap with LSST of ∼6000 deg$^2$, but will be ∼2 magnitudes shallower. On the other hand while NGRST will be as deep as LSST, its imaging will cover only 2300 deg$^2$. Operating at wavelengths of 300 to 900 nm with a field-of-view of 25 by 17 arcminutes, roughly 36 times larger than the HST's Advanced Camera for Surveys/Wide Field Camera 3, SuperBIT will restore the capabilities in the optical.

Within this wavelength range, the projected resolution and depth of SuperBIT imaging is sufficient to measure the (weak) gravitationally lensed shapes of distant ($z \approx 1$) galaxies behind foreground ($z \approx 0.3$) clusters of galaxies (Massey et al., 2007). SuperBIT's primary science goal then is to use strong and weak gravitational lensing to map out the distribution of DM in galaxy clusters and throughout the large scale structure of the Universe (see section 1.3.1 for a brief description of gravitational lensing). In addition, SuperBIT's wide field-of-view allows for an entire galaxy cluster to be imaged in one pointing, including its connection to the surrounding large-scale structure. Imaging is available in six selectable bands from 300 to 830 nm*, allowing for photo-metric redshift calibration. As such, cluster member galaxies can be identified via their 4000 Å break or the 3700 Å Balmer break in cluster dwarf galaxies for which this is suppressed.

However, its ability to obtain wide-field, high-resolution imaging, makes the instrument suitable for other proposed experiments related to, e.g., solar planet spectroscopy and exoplanet studies(for a comprehensive list see Romualdez et al., 2018). A brief description of the mechanical architecture of the SuperBIT instrument is provided in the following section.

*Specifically, these are the $U$ (330-430 nm), $B$ (370-570 nm), $G$ (520-700 nm), $R30$ (640-800 nm), and $S$ (530-830 nm) bands.

## 4.3 Mechanical architecture

Figure 4.2 shows the SuperBIT gondola as of its 2019 engineering flight, described in section 4.4. The gondola consists of three gimballed frames with the inner most frame containing the scientific payload, consisting of a 0.5 m NIR-to-NUV telescope with a field-of-view of 17 by 25 arcminutes, scientific charge-coupled device (CCD) readout electronics, and accompanying back-end stabilisation optics.

SuperBIT will maintain operations via the National Aeronautics and Space Administration (NASA)'s super pressure balloon (SPB) system. The volume of SPBs remains relatively constant with changes both in the ambient pressure outside the balloon and in the temperature of the lifting gas inside the balloon, allowing the balloon to keep a stable altitude for long periods of time. As such, one of the benefits of this SPB system over conventional zero-pressure balloon systems is that stratospheric operations can be supported through diurnal cycles. The SuperBIT launch vehicle consists of a SPB helium balloon with a volume of 1 million cubic m. The balloon is tethered to a 80–100 m long flight train, which constrains the parachute and is attached to the scientific payload/gondola through a pivot (figure 4.6).

While there are clear advantages to stratospheric balloon launch platforms, there are several unique challenges to balloon-borne telescopes as well. Perhaps the most challenging is correcting for the various pendulations the gondola is subject to in order to achieve diffraction limited sub-arcsecond resolution. Due to the stratospheric wind shears of the ballooning environment, the balloon and flight train induce gravity-driven compound pendulations. In addition, one needs to correct for the bulk sky rotation for long exposures (300-600 s) over the science payload field-of-view.

The three gimballed frames, shown in detail in figure 4.3, correct for these pendulations to provide sub-arcsecond stabilisation. Gimbal roll and pitch control is

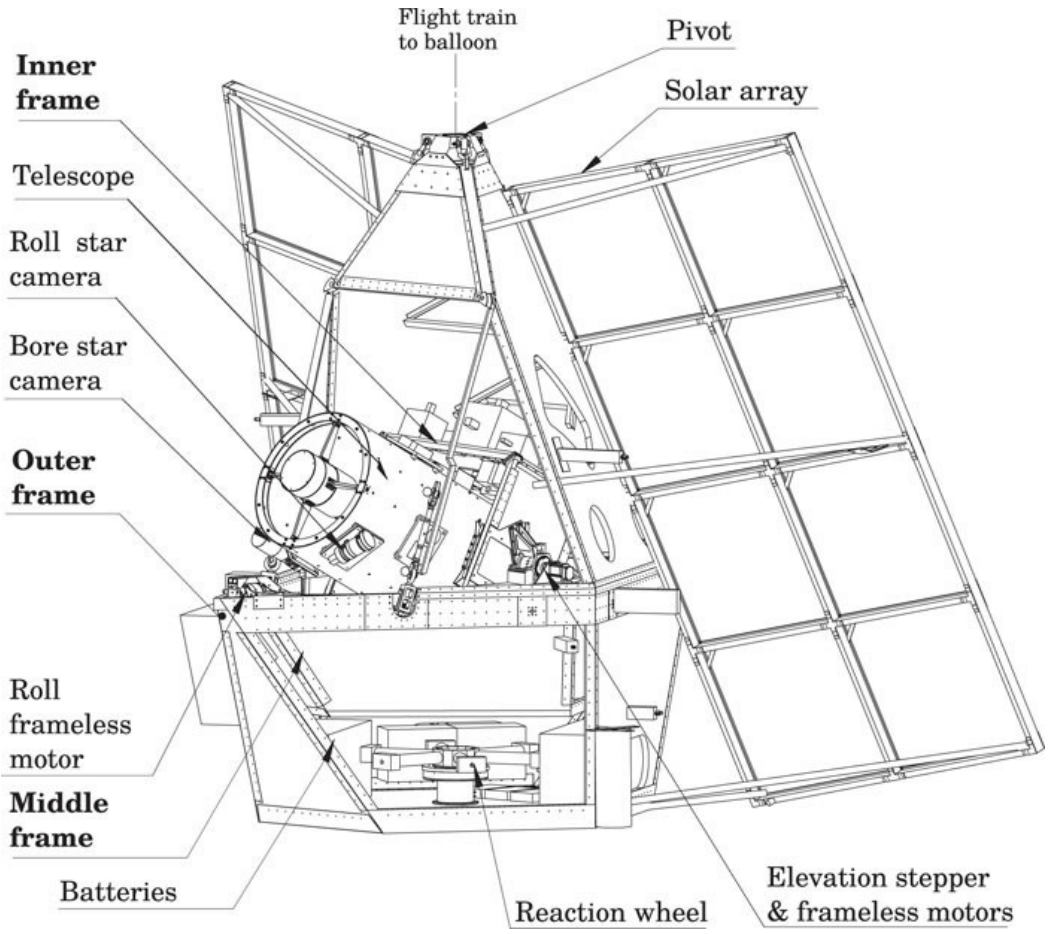


Figure 4.2: Figure 1 of Romualdez et al. (2020). The SuperBIT 2019 gondola. The main structure is comprised of three independently rotating gimbals. The telescope is connected to the balloon through the pivot. During the day, SuperBIT recharges its batteries using the solar panels.

facilitated per axis by motors, while a high-inertia reaction wheel facilitates yaw control and pendulation stability, with excess momentum dumped through the flight train to the balloon via the pivot connection.

Mounted to the inner frame are two wide-angle (2-3 deg) star tracking cameras — one pointing along the direction of the telescope's line-of-sight ('bore') and the other orthogonal to it ('roll') — that provide absolute sky-fixed pointing references at 1-50 Hz, while 1 kHz rate gyroscopes provide inertial stabilisation feedback. Based on feedback from the gyroscope sensors and the star tracking cameras, each frame corrects for motions along one of the Euler angles. Figure 4.3 shows the possible rotations of the three frames.

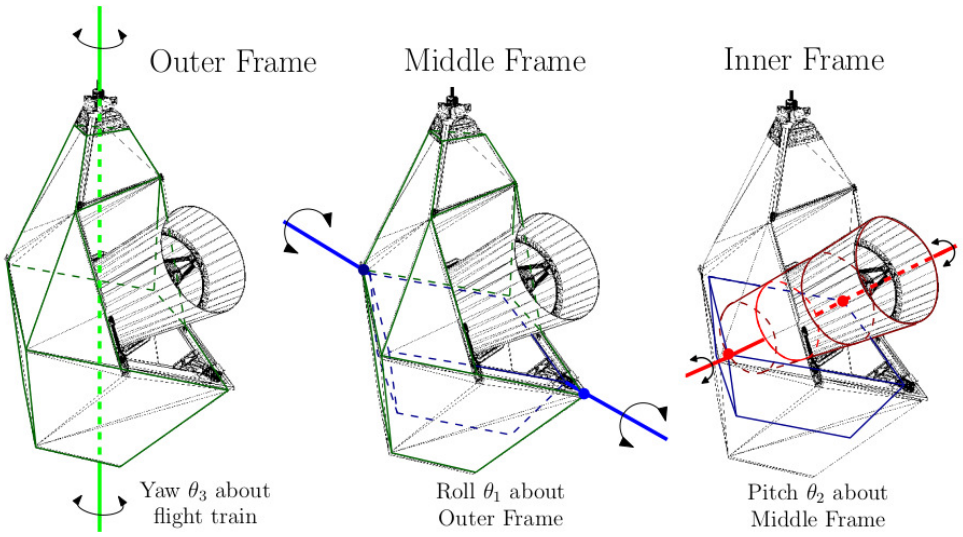


Figure 4.3: Figure 3 of Romualdez et al. (2016a). The three axes along which SuperBIT can rotate. From left to right: the outer, middle and inner frame (yaw, roll, and pitch).

While suspended, the SuperBIT gondola can rotate along the full 360 deg yaw range (left panel figure 4.3). However, due to the interference of the three frames, the roll of the middle frame is constrained to ±6 deg (middle panel figure 4.3). The pitch of the inner frame is restricted at 20 deg on the lower end due to the horizon and 55 deg on the upper end due to the obstruction caused by the helium balloon at full expansion (right panel figure 4.3).

When SuperBIT is fully assembled, it has a height of approximately 3 m from the base to the pivot and a weight between 800-1000 kg. The solar arrays provide the SuperBIT gondola with 1600 W solar power generation, which is stored in the 432 Ah power storage systems.

## 4.4 Engineering test flights

In this section, we give a brief overview of the four SuperBIT engineering test flights that have taken place from 2015 to 2019. Table 4.1 at the end of this section provides a summary of SuperBIT's performance over these four flights.

### 4.4.1 2015 BIT Timmins flight

The predecessor of SuperBIT, the Balloon-borne Imaging Testbed (BIT), had its inaugural engineering flight over the night of September 18, 2015 (for a detailed description, see Romualdez et al., 2016a). It was launched from the Timmins Stratospheric Balloon Base in Ontario, Canada. Facilities were provided by the Canadian Space Agency (CSA) and launch hardware was provided by the Centre National d'études Spatiales (CNES). BIT floated at 36 km altitude for 6.5 hours, and descended via a parachute in Northern Quebec. Afterwards, it was safely recovered with minimal damage.

The main goal of this particular flight was to test the pointing and stabilisation systems, and to model the vibrational modes in which the telescope could be unstable, and therefore needed to be controlled (Li et al., 2016). BIT's pointing instrumentation successfully stabilised the telescope to within 0.68 arc-seconds ($1\sigma$) for integration periods as long as 1.4 hours, and had target acquisition to within $< 0.1$ deg. However, images of large star fields taken with the science camera over 10-20 minutes long integration periods to assess the beam quality of the telescope *post*-launch showed that either telescope alignment pre-flight was insufficient or that alignment had suffered from shocks during launch.

This first iteration of the hardware used off-the-shelf optics and detectors (Clark et al., 2014). Despite the need for improvements to telescope optical alignment and image stabilisation hardware, the flight successfully demonstrated the ability to achieve sub-arcsecond pointing and image stability from a balloon-borne platform, which at that time had not been demonstrated at that level of precision, duration, and repeatability (Danielson et al., 1964). BIT was essentially a proof of concept for the SuperBIT instrument.

### 4.4.2 2016 SuperBIT Palestine flight

As a follow-up to the BIT Timmins flight, SuperBIT was launched in 2016 for its first engineering flight from NASA's Columbia Scientific Balloon Facility (CSBF) in Palestine, Texas, USA. The flight took place over a single night from June 30 to July 1. SuperBIT reached a float altitude of 33.5 km altitude, and landed West of Pecos, Texas, after a total flight time of 10.5 hours*.

Refurbishments were made to the BIT instrument to increase the overall bandwidth of the image stabilisation stage while improving the ability of the BIT system to accurately acquire targets of interest as SuperBIT will be driven by science objectives as opposed to engineering demonstrations (for a detailed description of the flight, see Romualdez et al., 2018). In addition, in order to better assess how the telescope's performance is affected due to changes in physical stress and temperature, a more rigorous approach to telescope alignment pre-flight was developed as well (see Redmond et al., 2018).

The BIT instrument was flown with a zero-pressure (variable-volume) balloon, however, the ultimate aim for SuperBIT is for it to be flown on a SPB system during science flights. Therefore, the main goal of this particular engineering flight was to demonstrate the ability to operate and calibrate the pointing systems during the flight using communications hardware and protocols similar to what would be used for an actual SPB flight, e.g. line-of-sight and over-the-horizon telemetry and commanding links at various bandwidths. Additional aims for this flight included: developing a more accurate and robust target acquisition stage, reconfirmation of telescope pointing stability at sub-arcsecond level, and improved image stability with redesigned, higher bandwidth tip-tilt hardware to correct for optical aberrations. Improved techniques for aligning the SuperBIT telescope pre-flight were used to increase the resistance of the optical alignment to any mechanical shock from launch.

*The full flight details and trajectory can be found at https://stratocat.com.ar/fichas-e/2016/PAL-20160701.html.

Figure 4.4: A picture of the Eagle nebula taken during the SuperBIT 2016 engineering flight. The image is a composition of 1-3 minute integrations in several observing bands ranging from NIR to NUV with a total observing time of 17 minutes.

During the flight, a number of targets were successfully acquired to within subarcsecond pointing accuracy. Figure 4.4 shows one of the 'glamour shots' taken during the 2016 flight. It is a composite image of the Eagle nebula, a young open cluster of stars (Hillenbrand et al., 1993).

As with the 2015 BIT flight, the observed beam quality suggested that alignment was effected by launch shocks and that possibly pre-flight alignment of the telescope optics was inadequate. These results emphasised the need for the ability to remotely realign the telescope after launch. This capability has since been implemented. However, the positive results from this flight highlighted the potential for the SuperBIT instrument to generate high resolution images during a prospective SPB flight.

### 4.4.3 2018 SuperBIT Palestine flight

SuperBIT's third overnight engineering flight took place over the night of 5 to 6 June of 2018. The telescope was again launched from Palestine with CSBF. This was the final engineering flight with the original 2016 BIT telescope and CCD. SuperBIT reached a float altitude of about ∼29 km with a total flight time of 21.2 hours. The telescope landed in an unpopulated zone 37 nautical miles (∼42.6 miles) south-east of San Angelo, Texas*.

The motivations for another test flight with CSBF included demonstrating enhanced image stability with upgraded hardware, improved flight operations with both redundant and more robust communications, and a refurbishment of the current model telescope to allow for in-flight real-time alignment and beam point-spread-function (PSF) correction. State-of-the-art fibre-optic rate gyroscopes were implemented alongside a high-speed, highly sensitive focal plane camera in order to increase the responsivity and bandwidth of the image stabilisation stage.

Science targets were chosen to assess the viability of the data analysis pipeline developed for SuperBIT. These and other similar engineering and science goals for the 2018 flight mainly served as probes to better inform methodologies for the future SPB flights of SuperBIT. An example of the imaging capability obtained during the 2018 test flight is shown in figure 4.5, which is an image of the spiral galaxy NGC 7331.

### 4.4.4 2019 SuperBIT Timmins flight

SuperBIT's final engineering flight took place over the course of two nights from 17 to 19 September 2019 (for a detailed description, see Romualdez et al., 2020). As with the BIT engineering flight, the telescope was launched from the CNES launch base in Timmins. This was the first flight that utilised high quality telescope optics

*The full flight details and trajectory can be found at https://stratocat.com.ar/fichas-e/2018/PAL-20180606.html.

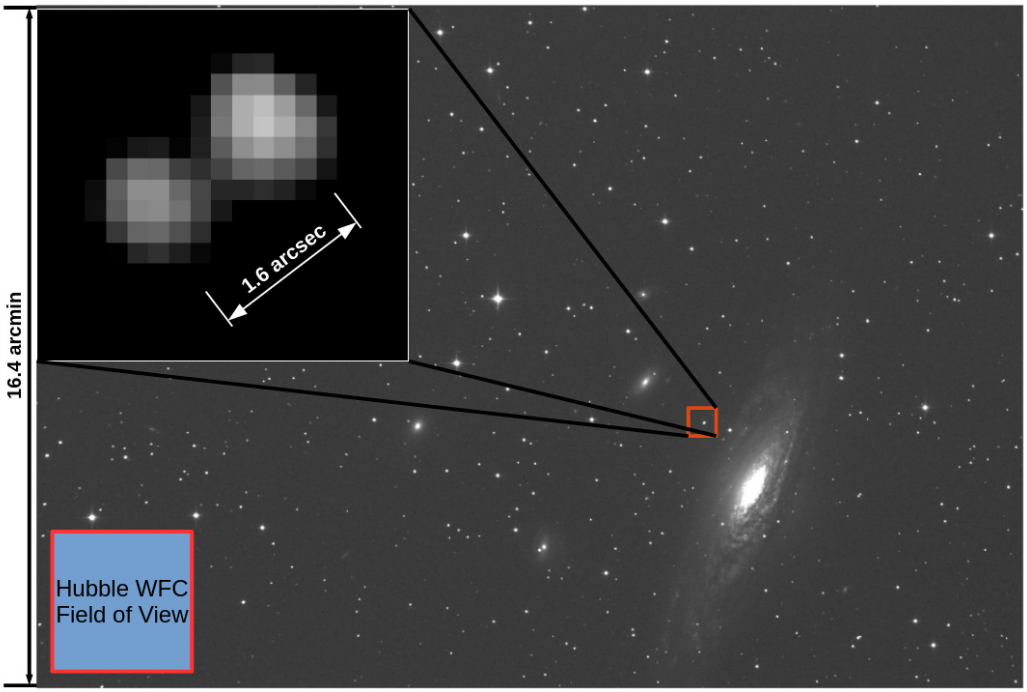


Figure 4.5: Figure 1 of Jones et al. (2019). A picture of the spiral galaxy NGC 7331 taken during the SuperBIT 2018 engineering flight. The image is a single 5 minute exposure at 500 nm. The field-of-view of HST's Wide Field Camera 3 is shown for scale. The inset shows that stratospheric diffraction-limited imaging could be used for the de-blending of future ground-based survey data.

for science imaging. Compared to the 2016 and 2018 flights, both the telescope and the CCD were upgraded in 2019. This flight demonstrated extraordinary pointing stability, with variation of less than one thirty-six thousandth of a degree for over an hour, enabling the telescope to obtain images of similar quality to HST. During the flight, measurements of the sky background level at different altitudes and wavelengths were also made in order to plan for upcoming science flights (Gill et al., 2020). Figure 4.6 shows SuperBIT just before the 2019 launch.

Table 4.1 shows the progression of SuperBIT's performance over the four engineering test flights from 2015 to 2019 described above. Currently, SuperBIT is undergoing preparations for its first fully operational science flight, scheduled to launch from Wanaka, New Zealand in early 2023. In addition, a successor to SuperBIT is currently being developed (Romualdez, 2018).

Figure 4.6: Figure 3 of Romualdez et al. (2020). SuperBIT just before the 2019 engineering launch. Top panel: the telescope is secured by a launch support vehicle beneath the tow balloon. Bottom panel: the telescope seen from a distance with both the tow (right) and primary (left) balloon. The smaller tow balloon provides neutral buoyancy for launch and is secured during inflation of the primary balloon.

## 4.5 Further applications

The development of SuperBIT has produced contributions in the fields of balloon-borne engineering and techniques for suborbital operations in general. Specifically, SuperBIT has established standards and general design methodologies for balloon-borne payloads in the areas of suborbital mechanical modelling and design (Li, 2015), altitude dynamics and control (Romualdez et al., 2020), and thermal modelling and mitigation (Redmond et al., 2018; Redmond, 2018). In addition, from the

Table 4.1: Adapted from table II of Romualdez et al. (2020). Summary of the absolute pointing and image stabilisation performance for the four SuperBIT test flights over five years. We show the stability over the course of five minutes, representative of science camera integration periods, as well as the stability over an extended period of 30 minutes.

| | | | Best achieved sky-fixed stability ($1\sigma$) [arcseconds] | | | |
|---|---|---|---|---|---|---|
| Year | Launch site | Provider | Telescope stabilisation | | Image stabilisation | |
| | | | @5 min. | @30 min. | @5 min. | @30 min. |
| 2015 | Timmins | CNES-CSA | 0.5 | 1.5 | 0.085 | 0.5 |
| 2016 | Palestine | CSBF-NASA | 0.5 | 1.1 | 0.070 | 0.2 |
| 2018 | Palestine | CSBF-NASA | 0.4 | 0.8 | 0.065 | 0.090 |
| 2019 | Timmins | CNES-CSA | 0.3 | 0.5 | 0.046 | 0.048 |

development SuperBIT various other endeavours and hardware have come about with applications outside of the SuperBIT instrument and the field of astronomy.

### 4.5.1 StarSpec Technologies

StarSpec Technologies is a company that aims to lower the barrier to entry to space as well as near-space environments by reducing the cost and overall development time*. Given the success of SuperBIT, the company was formed by the core team instrumental in developing the telescope. StarSpec Technologies is currently contracted for a number of instrumentation projects such as NASA's EXoplanet Climate Infrared TElescope (EXCITE; Pascale et al., 2021) project, which aims to characterise exoplanet atmospheres from the stratosphere.

### 4.5.2 The SuperBIT data recovery system

To keep up with SuperBIT's relatively high data acquisition rate, as well of its successor (estimated to obtain about 20 times *more* data), a system for physical 'downloading' was developed, launched, and tested during SuperBIT's 2019 engineering test flight. This system was named the SuperBIT Data Recovery

*For more information see https://www.starspectechnologies.com/.

System (DRS) (Sirks et al., 2020). Currently, the DRS's main purpose is *data* retrieval, however, we imagine it can be adapted to retrieve small physical samples as well. The DRS is described in detail in the next chapter, as well as the software I developed which predicts the descent trajectories and landing sites of the DRSs.

Chapter 5

# Download by Parachute: Retrieval of Assets from High Altitude Balloons

*"Any sufficiently advanced technology is indistinguishable from magic."*

— Arthur C. Clarke, *Profiles of the Future, 1962*

## 5.1 Introduction

High altitude balloon (HAB) missions are increasing in number, duration, and expense. Some acquire enough data that transmitting it to the ground would be impossible due to limited band-width or cost; others acquire physical samples that must be returned to the ground for full analysis. Mid-flight retrieval could improve a mission's efficiency, by using early results to optimise later data acquisition. Retrieval at any time mitigates the critical risk of total loss if the main hardware were damaged upon landing or lost, e.g. at sea.

Examples of small balloons include the ∼2000 radiosondes launched every day for weather forecasting, as well as instruments flown by amateur groups for scientific or educational purposes. Less than 20% of the ∼US$200 radiosondes launched in the USA are recovered, which prohibits upgrades to ∼US$1000 ozonesondes (Flores

et al., 2013), or increases in the number of weather stations, whose sparsity in the Southern hemisphere particularly limits forecasting precision (Karoly & Vincent, 1998).

An example of a large scale HAB mission is the Super-Pressure Balloon-borne Imaging Telescope (SuperBIT) (Romualdez et al., 2016b, 2020). SuperBIT is an astronomical telescope that rises above 99% of the Earth's turbulent atmosphere to achieve stabilised (Li et al., 2016; Redmond et al., 2018) high-resolution imaging at visible and near-UV wavelengths, with a field of view 36 times larger than the HST's Advanced Camera for Surveys/Wide Field Camera. SuperBIT is currently scheduled for a 50–100 day long duration flight, during which it will obtain ∼50 GB of uncompressed science data per day; a successor is already being designed that will obtain 20 times more (Romualdez, 2018).

Line-of-sight radio communications can achieve 100 Mbps but, on a long duration flight, global satellite communication systems are limited to 1 Mbps (10.5 GB per day), which is not exclusively used for image transfer, and cost up to US$0.50 per MB.*

We have developed the SuperBIT Data Recovery System (DRS) to recover assets from any balloon, any time it is over land. In default configuration, each DRS capsule includes 5 TB of storage, accessible over Wi-Fi™ Ethernet. These are attached to a HAB platform before launch, and ascend as usual. Following a remote command, they descend via parachute, transmitting their location via Iridium message – and continuing to transmit as well as beep audibly after landing. I have calibrated and tested software to predict the descent trajectory and landing site. This software helps to optimise the moment of release, so the DRS lands safely but accessibly, and assists retrieval on the ground. We successfully used two DRS capsules during SuperBIT's science commissioning test flight, and intend to use several more during its long duration mission. We also welcome interest from other HAB mission teams for whom the technology would be useful.

*See https://www.mailasail.com/Communication/Iridium-Pilot-Airtime

The rest of this chapter is organized as follows. Section 5.2 details safety and other requirements. Section 5.3 describes the DRS hardware and its release mechanism. Section 5.4 describes the algorithm we use to predict its landing site. Section 5.5 describes an end-to-end test of the DRS during the 2019 SuperBIT commissioning flight. We draw conclusions, and outline plans for future improvements in section 5.6.

## 5.2 Requirements

This section summarises the main safety requirements for a DRS to be allowed to be jettisonned from a balloon launched by the CSA and CNES from the Timmins Stratospheric Balloon Base in Ontario, Canada in September 2019. The requirements were set in conjunction with the International Civil Aviation Organization's Convention on International Civil Aviation Rules of the Air (Annex 2), but note that requirements may differ at other launch sites or for other agencies.

Relevant safety requirements include (but are not limited to)

(R1) *Electrical safety*: To prevent risk of fire, the gondola and/or DRS must be equipped with a fuse. All cables must be rated for a current greater than the fuse, and must also be insulated, protected, and secured. Electrical connectors must be designed so that there is no ambiguity in their connection. Static charges must be drained away.

(R2) *Mechanical safety*: The DRS capsule must not detach from the HAB platform unless commanded. In particular, the release mechanism must be sufficiently robust to withstand shocks during launch and descent (in case it is not released). The maximum vertical and horizontal acceleration for a 750 kg payload on a 14 million cubic foot zero-pressure balloon are $6.4g$ (vertical) and $1.3g$ (horizontal), which occur during parachute deployment.* We add

*According to CNES internal document BSO-MU-0-4793-CN-VA.

these in quadrature, with a safety factor of $\times 2$, and adopt a requirement on the DRS to withstand accelerations up to $13g$.

(R3) *Control of Fault Propagation*: Two or more active steps must be taken by an operator to initiate the release of a DRS capsule. In the event of power failure, there must be no change in the state of any safety barrier, and systems must switch to safe mode. It must not be possible for an electrical circuit to be activated as a result of an action on any other circuit, or through the effect of external events.

(R4) *Descent safety*: As the DRS reaches ground level, it must have vertical speed

$$|v_z| < \left(5 + \frac{3.4\,\mathrm{kg}}{m}\right)\mathrm{m\,s^{-1}}\,, \tag{5.1}$$

where $m$ is its mass.* This safety criterion applies to any package with total mass $< 2\,\mathrm{kg}$ and areal density $< 13\,\mathrm{g\,cm^{-2}}$, defined as the mass of the package divided by the area of its smallest surface.

To be useful, the DRS must also meet several practical requirements

(R5) *Easy to find*: The DRS must be easy to find after landing, visibly and audibly.

(R6) *Labelling*: In case the DRS is found by a person not associated with the HAB mission, it must be labelled with a safety warning about the electrical hazards, and contact details for more information or where to return the capsule.

(R7) *Predictable*: It must be possible to predict the descent trajectory and landing site of the DRS within 5 km (requirement) or 1 km (goal), in order to make go/no-go decisions about release. More accurate performance will open more potential landing sites that avoid e.g. towns and lakes, and cluster near remote roads to aid recovery. This code must run in $< 30\,\mathrm{s}$, so that accurate decisions

*Equation provided by CSA.

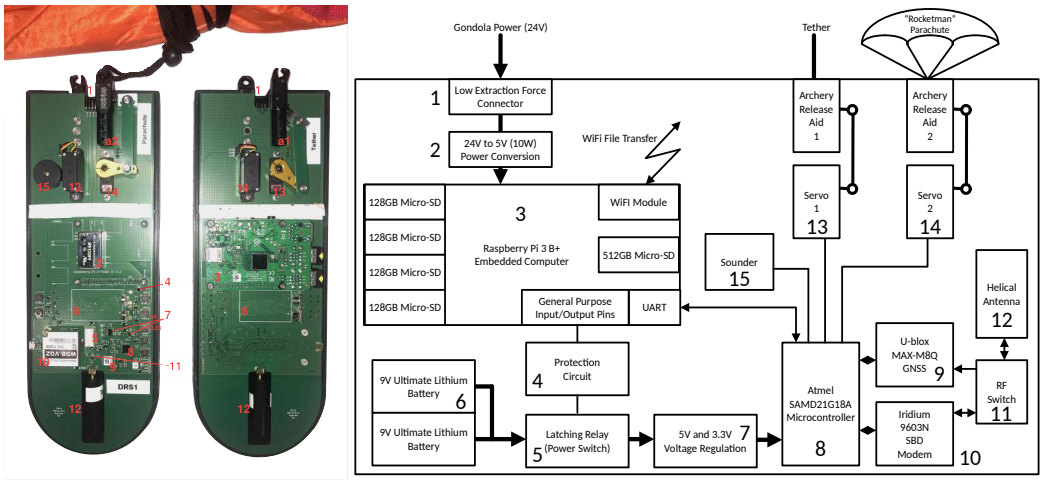


Figure 5.1: Left: Front side of the PCB. Middle: Rear side of the PCB. The red numbers refer to the numbers on the block diagram. Right: Block diagram of the PCB. The **a1** and **a2** indicate the archery release mechanisms 1 and 2 respectively.

can be made about the timing of release from even a fast-moving HAB. A slower but more accurate prediction may also be useful to assist recovery, in the event of communication loss.

## 5.3 Hardware

The DRS hardware design and operations software are open source.* All components are integrated onto on a custom 300 mm×100 mm printed circuit board (PCB). Throughout this section, numbers in square brackets refer to component labels in figure 5.1.

The main function of the DRS is to carry large quantities of science data to the ground and allow its recovery. It is, in effect, 'remote storage with benefits' for the main data acquisition computer (IFC). Data could be transferred into that remote storage either over a wired interface, such as USB or Ethernet, or wirelessly. In the case of SuperBIT, the IFC and the DRS are physically separated, making USB an unwise choice as, e.g., USB2.0 has a maximum cable length of 5 m. We originally selected wireless rather than wired Ethernet in order to avoid having to use an

*Available, with full operating instructions, from https://github.com/PaulZC/Data_Recovery_System.

8-way connector, although our experience with low extraction force connectors since then has suggested that an Ethernet interface would work well, and we have incorporated either option into the latest design (see Section 5.7.1).

The IFC manages many tasks, such as command forwarding, telemetry downlink, and science camera housekeeping. It is essential that the file transfer into the DRS does not take resources from those operations since the IFC is the gateway to the rest of the SuperBIT payload. Using a Raspberry Pi single board computer in the DRS allowed us to implement a wireless or Ethernet interface in a straightforward way. It also simplified the mirroring of files from the IFC into DRS storage, by having essentially a Unix computer at both ends of the transfer.

### 5.3.1 Enclosure

The PCB is protected by a 3D-printed ABS-like cover, which is manufactured in two identical halves and sealed around the lower two thirds to limit water ingress (with a moisture barrier vent to allow pressure equalisation).* This is enclosed inside a softer outer shell, made from moulded expanding polyurethane (PU) foam. Nylon paracord of diameter 2.4 mm is embedded into the foam, so it can be tied over the top of the cover to secure it; and a nylon sheet lining the mould forms a smooth outer surface on which warnings and contact details can be written in permanent marker (R6). The entire DRS, including parachute and batteries, weighs 1029 grams and has areal density $5.8\,\mathrm{g\,cm^{-2}}$.

### 5.3.2 Power

The DRS capsule will be powered down during most of the HAB mission. This prevents accidental or erroneous release. When the DRS is required, remotely switched (and fused) 12–48 V DC power is supplied from the gondola, via a low

*The material is similar to ABS, but is a bit easier to work with and does not suffer from the same delamination problems. See https://e3d-online.com/spoolworks-edge.

extraction force connector [1] with three pins arranged symmetrically and with redundancy on ground (R1). A medical-grade, switch-mode DC-DC converter [2] regulates power to 5 V. The embedded Raspberry Pi computer [3] automatically boots up, enables its Wi-Fi$^{TM}$ network, and connects to the main gondola flight computer. In its current configuration, the DRS uses a power cable with only 3 pins, to minimise the force required to disconnect. Further tests have shown that a connector with 8 pins (arranged in an asymmetric configuration to meet requirement (R1)) will also work, so future versions of the DRS may use wired Ethernet with Power-over-Ethernet.

Immediately before descent, a latching power relay [5] is switched, and two Energizer Ultimate Lithium 9V (PP3) batteries [6] supply similarly regulated [7] power to a tracking subsystem [8–15]. These batteries will henceforth remain powered, and are the only components of the jettisoned DRS that could be considered potentially hazardous (R1). However, they are compliant with safety test criteria T1–T8 defined in Section 38.3 of UN, Committee of Experts on the Transport of Dangerous Goods (2019), which include transportation safety and altitude simulation. Indeed, we have used these batteries without incident in $>$ 30 HAB flights (Clark et al., 2019).

### 5.3.3 Raspberry Pi

The Raspberry Pi provides the front-end user interface for the DRS, accessible during the mission via `ssh` from the main gondola flight computer. For SuperBIT, it is also the heart of the 'recoverable assets', hosting up to 5 TB of solid-state data storage (1 TB micro SD card that includes the operating system, plus $4 \times 1$ TB micro SD cards, through 480 MB s$^{-1}$ USB2.0). Data can be copied to this at any time before release, using gondola power. As a useful backup in case of faults e.g. due to cosmic rays in the space-like environment, data is constantly uploaded instead of all at once right before release.

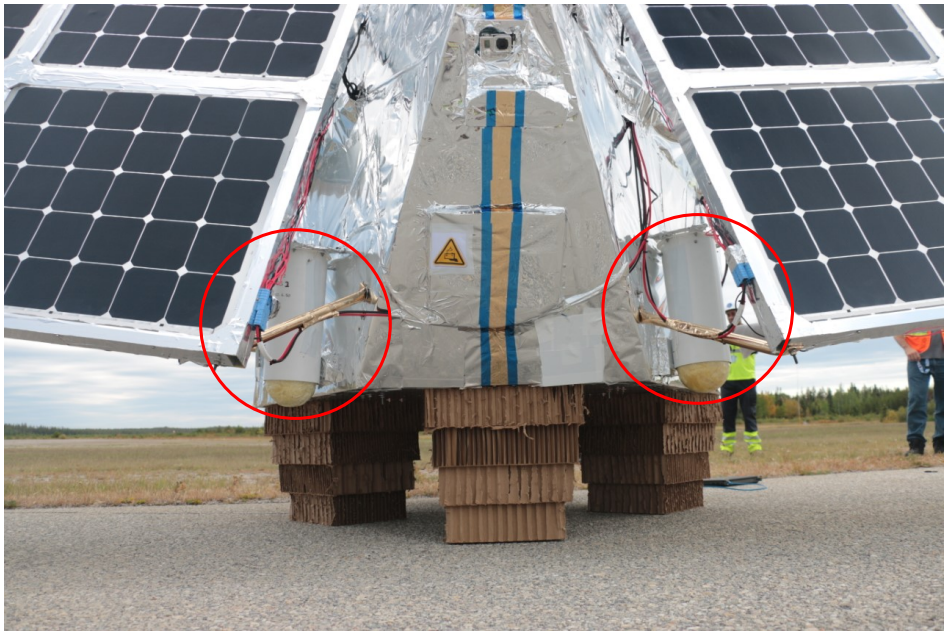

Figure 5.2: Two DRS capsules (highlighted by red circles), mounted on the back of the SuperBIT telescope just before launch on September 17, 2019. The white launch tubes stay attached to the telescope when the capsules are dropped. The PU foam surrounding the circuit boards can be seen protruding from the bottom of the tubes. The cardboard crush pads underneath SuperBIT are intended to soften impact upon landing.

We have also considered using the Raspberry Pis to pre-process and analyse science data during flight, but found they overheated when used for long durations in vacuum and inside the PU foam enclosure: implementing this would require thermal redesign.

### 5.3.4 Release mechanism

Each DRS capsule is packaged inside a plastic drainpipe (diameter 150 mm, length 350 mm), to limit swinging and to constrain the parachute before release (figure 5.2). These 'launch tubes' remain attached to the gondola after the DRS is released.

Inside each tube is a short power cable and a loop of 2.4 mm diameter nylon paracord. As with our balloon tracking payload (Clark et al., 2019), the DRS grips the loop using a sprung release-aid mechanism developed for archery, and oper-

ated here via a servo [13] stripped, cleaned and re-lubricated with 'space-grease' (Castrol's Braycote 601 EF). The strength of the release mechanism was tested against requirement (R2) by holding the PCB upside-down and hanging 13 kg of lead bricks from the nylon cord. The release mechanism held, and no damage to the PCB or nylon cord was observed.

### 5.3.5 Two-step instructions for release

Two further actions are required to release the DRS (R3), once the Raspberry Pi is powered up. First, the ground team must `ssh` into the Raspberry Pi and run a 'Power On' python script, which configures its GPIO pins to switch on the latching relay [5]. A discrete logic protection circuit [4] requires three of the GPIO pins to be in the correct state before the relay is triggered. The pins and states have been selected to prevent the relay from being accidentally triggered as the Pi goes through its boot process. Once the relay is triggered, the DRS's internal batteries power the microcontroller [8], which goes through its own start-up procedure and starts to monitor its serial (UART) port for a 'Go' command. The Global Navigation Satellite System (global navigation satellite system (GNSS)) receiver [9] is also powered up and starts to establish a fix. The GNSS NMEA messages are sent through the serial port of the microcontroller and logged by the Raspberry Pi. This can be monitored and, if required, the drop can be delayed until it is confirmed that the GNSS has established a fix.

Second, the ground team must use `ssh` to run another python script that sends a 'Go' command to the microcontroller via its serial (UART) port, then immediately shuts down the Raspberry Pi. 30 seconds later (time for the Pi to shut down gracefully), the microcontroller enables 5 V power to the servo via a P-channel FET then generates the correct Pulse Width Modulated (PWM) signal to move the servo to the open position. As the DRS is released, the low extraction force connector pulls apart, disconnecting power to the Raspberry Pi, which will remain inactive until recovery. If the 'Go' script is accidentally run before the first 'Power

On' script, the script will have no effect as the microcontroller will be unpowered and the 'Go' command ignored. If either of the microcontroller actions fail, e.g. due to its code crashing, the release will not open.

### 5.3.6 Parachute

The parachute is initially folded on top of the DRS, inside the plastic launch tube (figure 5.2). It unfolds when the capsule slides out of the white tube. We use a 4 foot (1.22 m) Rocketman parachute, which is expected to slow the descent of our 1029 g payload to terminal velocity $< 4\,\mathrm{m\,s^{-1}}$ at ground level, easily meeting requirement (R4).* It is coloured bright orange, to aid recovery on the ground (R5).†

### 5.3.7 Tracking and recovery

During descent and after landing, communication is maintained with the DRS via Iridium 9603N satellite modem [10]. The microcontroller alternately switches [11] between monitoring its location vis GNSS then transmitting this information via Mobile Originated Iridium SBD messages. A large, helical antenna [12] is shared for these tasks, saving weight while achieving superior performance than a patch antenna, especially after landing horizontally on ground, in trees or on water. A small Radio Frequency (RF) switch is used to connect the antenna to either the GNSS or the Iridium modem. The switch shields the GNSS during Iridium transmit bursts. This subsystem is a modified version of Clark et al. (2019)'s HAB tracking toolkit.

*See https://the-rocketman.com/recovery-html/.

†Optionally, a second servo [14] and archery release can be used to release the parachute once it has been confirmed to have reached the ground. This option could prevent the DRS from being dragged by the parachute, or allow it to fall to the ground if the parachute has become caught in a tree. However, it introduces a risk of the parachute being released prematurely, through human error. To militate against this risk, the second release can only be opened by sending a Mobile Terminated (MT) SBD message containing a time code. The microcontroller will only respond if the time code matches GNSS time to within an appropriate interval; it will ignore (and delete) all other messages, so old queued or erroneous MT messages have no effect.

A sounder [15] begins beeping after the 'Go' command is received. Thus a recovery crew can head to GNSS coordinates (in a worst case, transmitted immediately before landing), then look for a bright orange parachute and listen for beeps (R5). The sounder can be disabled (or re-enabled), and the frequency with which the DRS reports its location can be adjusted, via Iridium MT message to the DRS. Depending on this frequency, the batteries have an expected operating lifetime of 2–6 weeks. Electrical hazard warnings and contact information written on the nylon surface in permanent marker are easily visible after this time, even in wet conditions (R6).

## 5.4 Software to predict descent trajectories

The key remaining requirement (R7) is software to quickly and accurately predict the landing site of the DRS. I have adapted open source python code, originally written to simulate the trajectories of tropospheric sounding balloons*. Such trajectories included an ascent phase on a weather balloon and a descent phase of the payload on a parachute. We are principally interested in the descent phase, and I have improved and calibrated its accuracy. The code remains open source.†

### 5.4.1 Data

#### Weather models

I use Global Forecast System (GFS) weather models produced by the National Centers for Environmental Prediction (NCEP). They are generated every six hours, at 00:00, 6:00, 12:00, and 18:00 GMT, then become publicly available ∼3.5 hours later (for current weather conditions) to 5 hours later (for a forecast up to 16 days into the future).‡

*I.e. https://github.com/pnuu/pyBalloon by Panu Lahtinen, currently at the Finnish Meteorological Institute.

†See https://github.com/EllenSirks/pyBalloon.

‡See https://www.ncdc.noaa.gov/data-access/model-data.

The forecasts include air density, temperature, wind speeds, and geopotential heights in voxels across the globe, with a horizontal resolution of 0.5 degrees, and at 34 air pressure levels, ranging from 1000 mb (low altitude) to 0.4 mb (high altitude)*. The geopotential heights represent the height above sea level of a given pressure level†. This is an estimated height based on temperature and pressure data. At relatively low altitudes, the geopotential height is approximately equal to the geometric height. E.g. at the SuperBIT flight altitude, ∼30 km, the difference is less than 150 m. The models have a vertical resolution between ∼200 m near ground level to 5 km at stratospheric altitudes (∼50 km).

Conditions are forecast with a time resolution of 3 hours. The difference between the production time of forecasts and the trajectory time has a large effect on our accuracy, and so I introduce variable $t_{\text{future}}$, the number of hours a forecast is predicting into the future. For example, for conditions at 16:00, the forecast nearest in time is produced at 12:00 with $t_{\text{future}} = 3$ hours. An ensemble of weather forecasts, generated from slightly perturbed initial conditions, are also available for 9 days (after which their files are deleted, and the main model is moved to archival storage). I have experimented using the ensemble forecasts to estimate uncertainty – but find their variance to be smaller than other sources of uncertainty in our calculations, and cannot access them for historic flights, so do not exploit them.

I require a look-up table of atmospheric conditions at higher resolution than the GFS forecasts. I shall therefore interpolate all variables in vertical columns using a cubic B-spline, in latitude and longitude using bilinear interpolation, then linearly in time. Compared to this scheme, nearest neighbour interpolation degrades the accuracy of our landing site predictions by 28% (4% from spatial interpolation and 23% from temporal interpolation).

*GFS models are calculated at air pressure levels: 0.4, 1, 2, 3, 5, 7, 10, 15, 20, 30, 40, 50, 70, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 925, 950, 975, and 1000 mb.

†See 'height' at https://w1.weather.gov/glossary.

### Altitude of the ground

For locations with a latitude between -60 and 60 degrees, I use tables of ground altitude as a function of latitude and longitude with a resolution of 1 arcsecond or approximately 30 m at the equator.* For any other latitudes I use tables with slightly lower resolution of 3 arcseconds as the high resolution data are not available for these regions.† At a given location, I assign the altitude of the closest grid point in the tables as the elevation.

### Test flights

I have access to the trajectories of 30 flights in which a real payload ascended by weather balloon then descended via parachute (Clark et al., 2019). These took place between 2018 and 2019, in Switzerland (20), Greenland (4), and Morocco (6), and are listed in table 5.1. During each flight, the longitude, latitude, and altitude of the payload was recorded by GNSS in ∼5 minute intervals.

I exploit these trajectories to calibrate our software and test its accuracy; however, they were not originally intended for this purpose. For example, the payload mass was ∼1.6 kg (and always $< 2$ kg for legal reasons) but not accurately recorded on each occasion. The parachute was a 7 foot (2.13 m) Rocketman parachute, of the same design as our DRS but larger. Furthermore, the time at which the balloon burst was not recorded (even though it was detected via the on-board accelerometer). At the highest point recorded by GNSS, the payload could be either ascending or descending. It was only guaranteed to be descending at the time and location recorded *after* the highest point. I therefore use this as the initial condition for descent trajectories.

*See http://srtm.csi.cgiar.org/srtmdata/.
†See http://viewfinderpanoramas.org/.

Table 5.1: Descent trajectories of real payloads, logged via GNSS. I use the top 30 descents to calibrate and test the software. During each of these flights, a payload ascended via weather balloon, then descended via parachute. The exact cutdown point was not recorded, so I list the time, date and location of the first GNSS location after its highest: the first moment at which the payload is guaranteed to be descending. The landing site is the mean position of the GNSS locations recorded with the same altitude. The range $d$ is the straight-line horizontal distance between the release point and landing site. $\langle t_{\text{future}} \rangle$ is the mean $t_{\text{future}}$ of all the forecasts used at each altitude step in a given predicted trajectory. The bottom two rows show the position and time of release of the DRS capsules launched with SuperBIT for an end-to-end demonstration. The flights marked by an asterisk had limited data logs or extreme trajectories, and were excluded from our statistical analysis.

| Date | Time | Location | Release point | | | Landing site | | Time of | Horizontal | $\langle t_{\text{future}} \rangle$ | Predicted landing site | | Error in |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Lat. | Lon. | Altitude | Lat. | Lon. | flight | range $d$ | | Lat. | Lon. | prediction |
| [GMT] | | | [degrees] | [degrees] | [m] | [degrees] | [degrees] | [minutes] | [km] | [hours] | [degrees] | [degrees] | $\Delta r$ [km] |
| 2018-03-04* | 10:11 | Switzerland | 46.72 | 6.56 | 26091 | 46.98 | 7.12 | 65.0 | 51.7 | 4.5 | 46.91 | 7.0 | 12.08 |
| 2018-04-06 | 08:54 | Switzerland | 46.62 | 7.0 | 28988 | 46.63 | 7.3 | 54.0 | 22.8 | 9.2 | 46.62 | 7.3 | 1.38 |
| 2018-04-06 | 18:36 | Switzerland | 46.59 | 6.89 | 23178 | 46.59 | 7.09 | 54.0 | 15.1 | 6.9 | 46.59 | 7.07 | 2.11 |
| 2018-05-06 | 07:47 | Switzerland | 46.63 | 7.05 | 25895 | 46.57 | 6.79 | 55.0 | 20.8 | 8.1 | 46.6 | 6.8 | 3.53 |
| 2018-05-11 | 10:18 | Switzerland | 46.61 | 6.88 | 28932 | 46.65 | 6.98 | 52.2 | 8.4 | 4.6 | 46.64 | 6.96 | 1.69 |
| 2018-05-19 | 08:35 | Switzerland | 46.52 | 6.74 | 28426 | 46.47 | 6.89 | 54.0 | 12.1 | 8.9 | 46.48 | 6.88 | 1.18 |
| 2018-06-16 | 11:11 | Switzerland | 46.68 | 6.77 | 21803 | 46.67 | 6.95 | 67.1 | 13.4 | 5.5 | 46.67 | 6.94 | 0.98 |
| 2018-06-30 | 19:23 | Switzerland | 46.69 | 6.67 | 25882 | 46.6 | 6.88 | 48.3 | 19.2 | 7.7 | 46.61 | 6.86 | 2.23 |
| 2018-08-03* | 19:06 | Greenland | 67.75 | -48.52 | 28557 | 67.97 | -48.31 | 48.0 | 26.1 | 7.4 | 67.95 | -48.28 | 2.41 |
| 2018-08-06 | 07:24 | Greenland | 69.31 | -50.81 | 22134 | 69.34 | -50.56 | 51.0 | 10.3 | 7.7 | 69.33 | -50.57 | 1.04 |
| 2018-08-11 | 11:00 | Greenland | 69.74 | -51.2 | 8765 | 69.87 | -51.16 | 42.2 | 15.1 | 5.2 | 69.86 | -51.16 | 1.14 |
| 2018-08-13 | 17:00 | Greenland | 69.43 | -51.52 | 21457 | 69.56 | -51.63 | 42.3 | 15.1 | 5.3 | 69.57 | -51.64 | 1.54 |
| 2018-10-28 | 11:30 | Morocco | 30.78 | -5.61 | 25892 | 30.78 | -4.85 | 45.0 | 72.8 | 5.8 | 30.79 | -4.83 | 1.92 |
| 2018-11-17 | 10:24 | Switzerland | 46.36 | 7.54 | 24621 | 46.09 | 6.98 | 44.0 | 51.8 | 4.7 | 46.12 | 6.99 | 2.65 |
| 2018-12-15 | 10:11 | Switzerland | 45.85 | 7.9 | 26435 | 45.39 | 8.36 | 51.0 | 62.6 | 4.5 | 45.41 | 8.37 | 2.59 |
| 2019-01-01 | 13:41 | Switzerland | 45.78 | 8.18 | 25863 | 45.39 | 8.43 | 52.0 | 47.1 | 8.0 | 45.4 | 8.43 | 0.85 |
| 2019-02-02 | 13:30 | Switzerland | 46.58 | 6.67 | 20706 | 46.79 | 6.79 | 45.0 | 24.6 | 7.8 | 46.74 | 6.76 | 5.74 |
| 2019-02-09 | 15:30 | Switzerland | 45.94 | 8.16 | 14030 | 45.85 | 8.69 | 33.0 | 42.4 | 3.8 | 45.85 | 8.71 | 1.77 |
| 2019-03-02 | 12:24 | Switzerland | 46.01 | 6.83 | 25950 | 45.98 | 7.0 | 39.0 | 13.7 | 6.6 | 45.96 | 7.03 | 3.46 |
| 2019-03-11 | 17:53 | Switzerland | 45.11 | 7.31 | 27358 | 44.42 | 7.79 | 48.0 | 85.8 | 6.2 | 44.37 | 7.83 | 6.28 |
| 2019-04-19 | 18:00 | Switzerland | 31.27 | -6.86 | 23546 | 31.15 | -6.27 | 44.5 | 56.8 | 6.3 | 31.15 | -6.26 | 1.75 |
| 2019-04-21 | 18:18 | Switzerland | 31.25 | -6.96 | 24493 | 30.9 | -6.61 | 42.0 | 50.3 | 6.6 | 30.93 | -6.6 | 2.88 |
| 2019-04-22 | 13:30 | Switzerland | 30.78 | -7.42 | 25594 | 30.57 | -7.14 | 48.0 | 35.0 | 7.8 | 30.61 | -7.15 | 4.28 |
| 2019-04-24 | 14:30 | Switzerland | 31.26 | -6.91 | 23507 | 31.17 | -6.57 | 39.0 | 33.5 | 8.8 | 31.18 | -6.55 | 2.61 |
| 2019-04-25 | 13:41 | Switzerland | 31.25 | -9.35 | 24881 | 31.3 | -9.11 | 45.5 | 22.8 | 8.0 | 31.31 | -9.11 | 1.47 |
| 2019-05-31* | 12:11 | Morocco | 46.06 | 6.02 | 23838 | 45.95 | 5.99 | 30.0 | 12.4 | 6.5 | 45.88 | 5.97 | 7.81 |
| 2019-07-26 | 14:30 | Switzerland | 32.03 | -6.5 | 28321 | 31.94 | -6.45 | 42.2 | 10.6 | 8.8 | 31.93 | -6.48 | 2.2 |
| 2019-07-28 | 19:53 | Switzerland | 31.36 | -9.31 | 27829 | 31.43 | -9.23 | 51.0 | 11.3 | 8.2 | 31.42 | -9.22 | 1.78 |
| 2019-07-30* | 08:11 | Switzerland | 29.52 | -9.94 | 23879 | 29.56 | -9.82 | 30.0 | 12.2 | 8.5 | 29.59 | -9.8 | 4.06 |
| 2019-08-01 | 07:00 | Switzerland | 31.27 | -4.13 | 25439 | 31.47 | -3.97 | 48.0 | 26.0 | 7.3 | 31.45 | -4.0 | 3.41 |
| 2019-09-18 | 02:58 | Canada | 47.76 | -80.72 | 28400 | 47.49 | -80.56 | 38.6 | 32.4 | 9.2 | 47.5 | -80.56 | 0.54 |
| 2019-09-18 | 17:52 | Canada | 47.49 | -82.4 | 29848 | 47.42 | -82.17 | 35.4 | 18.6 | 6.1 | 47.42 | -82.19 | 1.12 |

### 5.4.2 Method: dynamical modelling

#### Initial Conditions

The user inputs the starting location, $\mathbf{r}^{\text{release}}$=(longitude, latitude) and altitude $z$, as well as the date and time of release (this defaults to now). If desired, a 'drift time' can be specified, during which the DRS travels horizontally with the HAB platform before release. The code automatically determines and downloads the most appropriate GFS weather data for these inputs.

Upon release, I assume that the DRS instantly reaches terminal velocity. Balancing gravitational acceleration $g$ acting downwards and drag force acting upwards, this is

$$v_z^{\text{predicted}} = -\lambda \left( \frac{m}{A\, C_d} \right)^{\frac{1}{2}} \left( \frac{2g}{\rho} \right)^{\frac{1}{2}}, \tag{5.2}$$

where $m$ is the mass of the payload, $\rho$ is the density of air, $A$ is the area of the parachute, and $C_d$ is its coefficient of drag. I initially adopt the manufacturer's design specifications for $A$ and $C_d$ (see section 5.3.6), but calibrate these via free parameter $\lambda$ (see section 5.3.6). Both $g$ and $\rho$ depend on altitude; I calculate $g(z)$ assuming the Earth is a perfect sphere with a radially symmetric distribution of mass and interpolate $\rho$ from the GFS weather model.

#### Iterated descent trajectory

We split the descent into altitude steps of height $\Delta z$ (I set a requirement on this in section 5.4.1). For each altitude step, I calculate the time $\Delta t$ to descend from top to bottom, assuming that the parachute moves vertically with the terminal velocity evaluated at the midpoint of the altitude step, directly below its starting position. The main strength of this 'leapfrog' method of updating the velocity is that it better conserves the energy of the dynamical system and therefore does not allow the system to drift substantially over time. By using this method, I

better approximate the true velocity versus altitude curve than if instead I used the velocity at the beginning of the altitude step.

I neglect updraughts and downdraughts in the GFS model, finding these negligible to the terminal velocity and having no measurable effect on the accuracy of our predicted landing sites.

During each altitude step, I assume that the parachute and payload travel horizontally with North-South ('$u$') and East-West ('$v$') wind speeds, again evaluated directly below the starting position, at the midpoint of the altitude step. I update the latitude and longitude of the DRS using the haversine formula, then iterate to the next altitude step.

#### Termination criterion

The code iterates the position of the DRS until it reaches sea level (altitude $z = 0$). This is generally below ground. I do not test for this during descent, because calls to evaluate ground level are relatively slow, and fast horizontal speeds near the ground necessitate a new call at each step.* I instead work backwards from $z = 0$, checking whether each point in the predicted trajectory was above or below ground. Once we find a pair of coordinates straddling ground level, I interpolate linearly between them to predict the latitude and longitude of the landing site, $\mathbf{r}$.

#### Convergence test

The choice of altitude step size $\Delta z$ represents a tradeoff between precision and run-time. Run-time is important for real-time predictions of the landing site, to optimise the moment of release from a fast-moving HAB (requirement R7). To predict the landing site $\mathbf{r}_1$ with the greatest possible precision (but slowly), I use altitude step size $\Delta z = 1\,\mathrm{m}$ to calculate trajectories from all the initial conditions in

*Checking that the DRS is above ground at each time step adds 1 s to runtime if $\Delta z = 100\,\mathrm{m}$, or 20 s for $\Delta z = 1\,\mathrm{m}$.

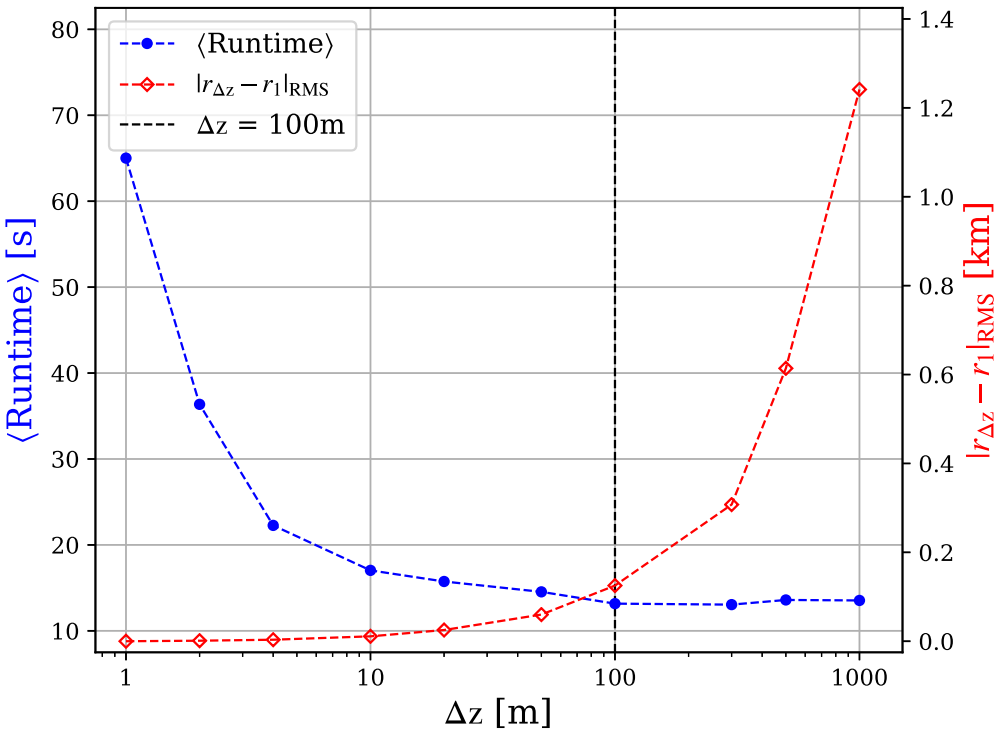


Figure 5.3: Code convergence test, and tradeoff between precision versus speed. *R*ed: the root mean square horizontal error in predicted landing site as a function of altitude step size $\Delta z$, compared to the most accurate prediction using $\Delta z = 1\,\mathrm{m}$. *B*lue: mean wallclock runtime per trajectory calculation, on a 1.7GHz laptop. In both cases, trajectories are calculated from, and averaged over all 30 initial conditions in table 5.1. The vertical dashed black line indicates our choice of nominal altitude step $\Delta z = 100\,\mathrm{m}$ that is used for all further analysis in this paper.

table 5.1, as a representative sample of possible release locations. I then recompute the trajectories with different step sizes, and record predicted landing sites $\mathbf{r}_{\Delta z}$. The mean error $\langle \mathbf{r}_{\Delta z} - \mathbf{r}_{1\mathrm{m}} \rangle$, and the wall-clock runtime on a laptop with a 1.7 GHz CPU are shown in figure 5.3. Note that during calculation of the trajectories, I did not check for ground elevation.

Predictions for the landing site converge successfully if the altitude step size fully samples the (maximum 200 m) vertical resolution of the GFS models. A practical compromise is $\Delta z = 100\,\mathrm{m}$. In a runtime of 13 seconds, this achieves a mean landing site precision of 125 m: an error that is subdominant to other sources of uncertainty. All further analysis will be performed with this step size.

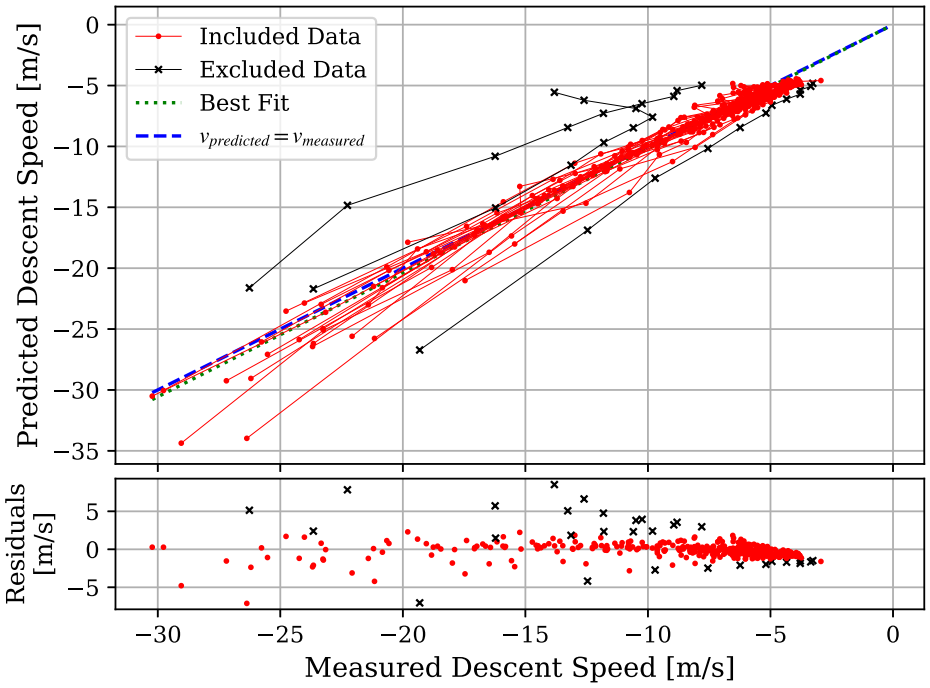


Figure 5.4: Calibration of our parachute descent model, by comparing (only) the vertical speed predicted for and recorded during the 30 test flights. *Top panel*: the descent speed for each flight (red and black lines for included and excluded flights respectively); trajectories start on the left and end on the right, with data points recorded every ∼5 minutes. If our trajectory calculation were perfect, the predicted and actual descent speeds would be equal (blue dashed line). The best-fit linear perturbation from this is consistent with the speeds having been overestimated by $(3.7 \pm 0.4)\%$ (green dotted, which is constrained to pass through the origin). *Bottom panel*: residuals of the best fit to the data in the top panel.

### 5.4.3 Trajectory calibration and validation

#### Vertical Descent Speeds

I compare the vertical component of the predicted descent speeds to the altitude difference between successive GNSS measurements, for 29 of the 30 test flights (figure 5.4).* The predicted and measured speeds would be equal, if the design specification of the parachute's drag coefficient and area were correct, and the pay-

*The GNSS failed to record during most of the 2018-08-03 flight in Greenland (most likely due to cold), so I exclude this flight from figure 5.4 and all subsequent analysis.

load masses were recorded correctly. To refine our knowledge of these parameters, I fit the free parameter $\lambda$ from equation (5.2) across all flights, as

$$v_z^{\text{predicted}} = \lambda v_z^{\text{measured}}. \tag{5.3}$$

The best-fit value is $\lambda_{\text{bf}} = 1.019 \pm 0.006$. There is a marginal evidence that the predicted speeds are approximately correct at high speed (high altitude), but 10–20% too low at low speed (low altitude). This might be due to additional drag in the higher density air – but without further evidence to support and quantify this hypothesis, I shall consider it useful margin in safety requirement (R4), and empirically incorporate it into our uncertainty in the predicted landing sites.

In our test data, the payload mass and parachute diameter were not precisely recorded. To test whether these varied between flights, I refit $\lambda$ for each individual flight. Three flights in particular (2018-03-04 in Switzerland, 2019-05-31 and 2019-07-30 in Morocco) have large ($> 4\,\text{km}$) errors in their predicting landing sites (see table 5.1) and also have the most anomalous values of $\lambda_{\text{bf}}$. They are so different from $\lambda_{\text{bf}} = 1$ that either $m < 1\,\text{kg}$ (unlikely for practical reasons), $m > 2\,\text{kg}$ (impossible for legal reasons), or (most likely) a different parachute was used. I exclude these three flights from further quantitative analysis. All other 26 test flights have descent rates consistent with a mean value of $\langle 1/\lambda_{\text{bf}} \rangle = 0.967 \pm 0.005$. Individual values of $\lambda_{\text{bf}}$ vary by $< 20\%$; if I use these values to recompute the trajectory, the mean error in landing site (compared to the truth) changes negligibly from 2.40 km to 2.37 km. I thus conclude that both the parachutes and payload masses were likely constant for these flights. Nonetheless, because $\lambda_{\text{bf}}$ is always consistent with 1, yet the true payload mass remains uncertain, I henceforth adopt $\lambda = 1$ for all further calculations. If the payload masses did vary between flights, this approach will lead to a slight increase in our estimate of uncertainty. However, it should avoid biasing the calculation of future trajectories with different payload masses.

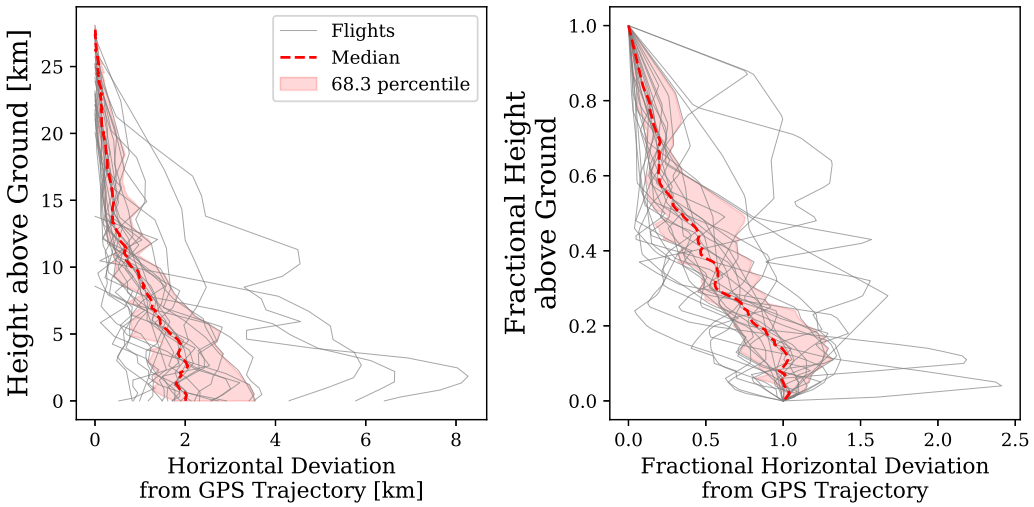


Figure 5.5: Accuracy of trajectories predicted for the descent of 26 parachutes, compared to the true trajectories recorded by GNSS. Trajectories begin at the top, and end at the bottom. *Left panel*: absolute horizontal deviation of each true trajectory from the prediction, at heights above ground level whenever the GNSS location was recorded, every ∼5 minutes. Each descent begins from a slightly different altitude. The red line indicates the median of the 26 flights, and the red area indicates the 68.3% region. *Right panel*: as before, but with the vertical and horizontal distance covered by each trajectory normalised to start or end at the same fractional altitude or horizontal deviation.

### Horizontal position

The most important aspect of a predicted trajectory is its horizontal accuracy, which culminates in the distance of its predicted landing site from the true landing site, $\Delta\mathbf{r} = (\mathbf{r}_{\text{predicted}} - \mathbf{r}_{\text{true}})$. I find that the predicted trajectories are most accurate at high altitude, which is traversed quickly, and near the ground, where the weather forecast is higher resolution and perhaps more accurate (figure 5.5).

Most of the deviation from the predicted trajectory builds while the parachute descends through the jet stream, where horizontal speeds are also greatest. Thus, the accuracy of our predictions is probably more limited by the accuracy of weather forecasts than the accuracy of our time-stepping algorithm.

I model uncertainty in the predicted landing site as

$$\sigma_{\parallel}^2 = (q\sigma_{\perp})^2 \equiv \sigma_0^2 + h d_{\text{predicted}}^2 + k\langle t_{\text{future}}\rangle^2, \tag{5.4}$$

Table 5.2: Best-fit parameters for model (5.4) of the uncertainty in predicted landing sites, after predicting all the descents in table 5.1. The two sets of parameters represent predictions made using only those weather forecasts available before release, or also those spanning the time of release and available shortly after.

| Weather forecast models | $\sigma_0$ [km] | $h$ [$10^{-4}$] | $k$ [$10^{-3}\,\mathrm{km}^2/\mathrm{hour}^2$] | $q$ |
|---|---|---|---|---|
| Available at launch | $1.77 \pm 0.14$ | $3.1 \pm 1.5$ | $3.6 \pm 0.9$ | $1.14 \pm 0.06$ |
| Available with hindsight | $1.63 \pm 0.13$ | $6.4 \pm 1.6$ | $3.3 \pm 1.1$ | $1.20 \pm 0.07$ |

where $d_{\mathrm{predicted}}$ is the horizontal distance between the release point and predicted landing site, $\langle t_{\mathrm{future}} \rangle$ the average $t_{\mathrm{future}}$ of the forecasts used at each altitude step in a predicted trajectory — and $\sigma_\parallel$, $\sigma_\perp$, $q$, $\sigma_0$, $h$ and $k$ are free parameters. In particular, $\sigma_\parallel$ ($\sigma_\perp$) is our model uncertainty in (perpendicular to) the mean direction of predicted travel, and $q$ is the axis ratio between them.

I fit the free parameters using Python code EMCEE (Foreman-Mackey et al., 2013) to maximise log-likelihood

$$\ln \mathcal{L} \equiv -\frac{1}{2} \sum_{i=1}^{26} \left[ (\Delta r_{\parallel,i} - \sigma_{\parallel,i})^2 + (\Delta r_{\perp,i} - \sigma_{\perp,i})^2 \right], \tag{5.5}$$

where $\Delta r_{\parallel,i}$ ($\Delta r_{\perp,i}$) is the component of $\Delta \mathbf{r}$ in (perpendicular to) the direction of $d_{\mathrm{predicted}}$, for each descent in table 5.1. I compute two sets of predicted trajectories. The first set is relevant to assess the safety and optimum timing of a live release, and uses only those weather forecasts that would be available at release (or earlier, to constrain $k$). The second set is the most accurate that could be made to aid recovery, if communications were lost with DRS capsules immediately after release. These interpolate between weather forecasts available before and after launch, and also use $\Delta z = 1\,\mathrm{m}$, for a slower but slightly more accurate calculation.

In both cases, the uncertainty is slightly greater in the direction of travel ($q > 1$); I convert the best-fit parameters into error ellipses on the predicted landing sites.

## 5.5 End-to-end system test

I shall now describe an end-to-end test of the DRS hardware and software performed during the 2019 science commissioning flight of the SuperBIT telescope. In general, DRS capsules could be released at any time during a HAB mission, with only a few minute's notice. For convenient retrieval, we planned to release one DRS shortly after reaching ceiling (so that it would land near the launch base) and the second shortly before termination (so that it would land near the main gondola). To save cost, the DRS capsules were configured for this test with only 1 TB of storage ($1 \times 512$ GB plus $4 \times 128$ GB) instead of the maximum 5 TB.

### 5.5.1 Launch and release

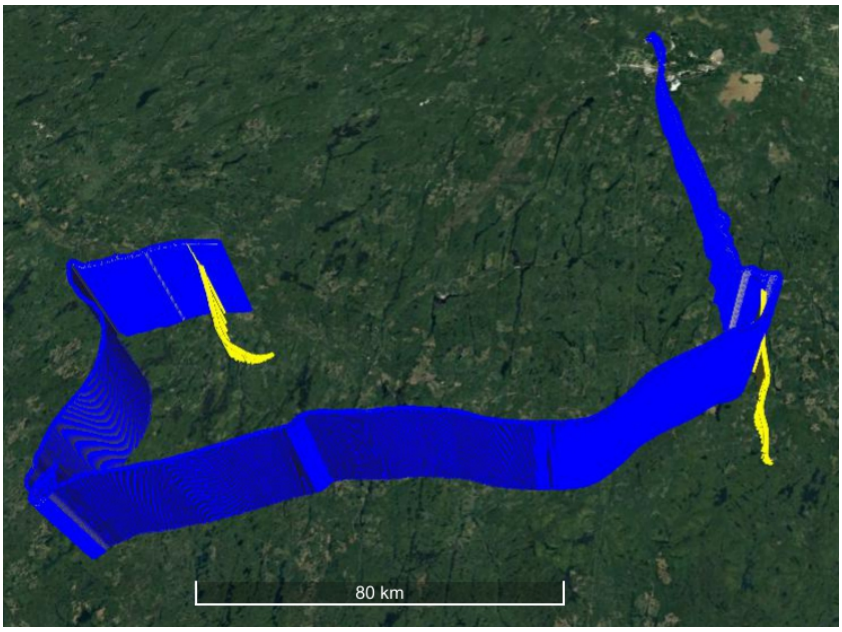


Figure 5.6: The flight path of the two DRS capsules, while they were attached to SuperBIT (blue) and while descending independently by parachute (yellow). The trajectory starts near the top right corner of the figure, and continues clockwise. It does not include SuperBIT's descent because the main gondola powers down before termination.

The SuperBIT telescope was launched from the CNES Stratospheric Launch Base in Timmins, Ontario on 2019-09-17 at 20:34 GMT-4, carrying two DRS capsules

(figure 5.2). During ascent, we obtained science calibration data from the telescope, and copied it to the DRS capsules. Shortly after ascent through ∼28 km altitude, I used my trajectory prediction software to target an area of forest without lakes or population, yet still near enough to the launch facility for convenient retrieval. We waited until the DRS would land near remote but usable roads identified in satellite imagery, then released the first DRS capsule with predicted $1\sigma$ uncertainties on the landing site of 2.0 km and 1.7 km in the directions parallel and perpendicular to the direction of travel respectively.

The SuperBIT mission continued, performing telescope calibration and alignment – followed by 3.5 hours acquiring science data that was copied to the second DRS. We planned to release the second DRS shortly before mission termination, so that it would land near the SuperBIT gondola, convenient for retrieval. In the event, the mission was terminated early because SuperBIT's balloon had a leak. We still released the DRS shortly before termination but, because of time constraints, did not have opportunity to run our prediction software in advance. This was acceptable from a safety perspective because the main gondola was predicted (by proprietary CNES software) to land well away from population, and had a similar value of $m/AC_d$ as the DRS. We released the DRS, and afterwards ran the prediction software for the moment of release, using weather forecasts that would have been available in advance. Predicted $1\sigma$ uncertainties on the landing site were 1.9 km and 1.6 km in the directions parallel and perpendicular to the direction of travel respectively.

Figure 5.6 shows the full trajectory of SuperBIT, recorded by its own GNSS receiver, and the trajectories of both DRS capsules. Coordinates of the DRS release points are included in table 5.1.

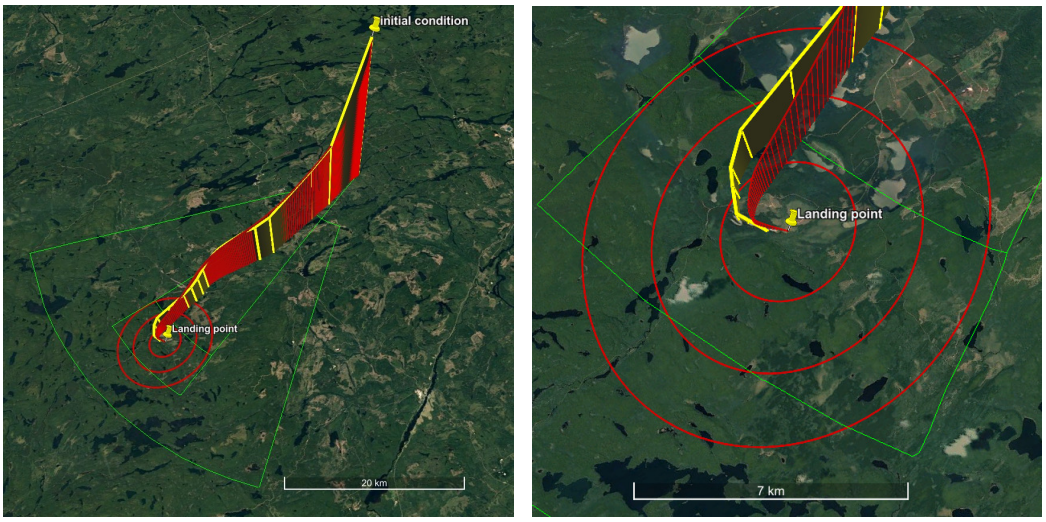


Figure 5.7: The predicted trajectory of the first DRS capsule, using GFS weather forecast data available at launch (red), and its actual trajectory recorded by GNSS (yellow). The yellow pin labeled 'initial condition' on the top right marks its release location. The yellow pin labeled 'Landing point' marks its predicted landing location, surrounded by red ellipses indicating 1, 2, and 3$\sigma$ uncertainty. Narrow and wide green cones show the 1 and 3$\sigma$ predictions from CNES software. The right panel is a zoom of the left.

### 5.5.2 Descent and landing

Both DRS capsules began logging GNSS coordinates before release, and continued transmitting them via Iridium, every ∼2 minutes (17 and 20 times) during descents lasting 35 and 39 minutes. We had increased the frequency of these transmission for better localisation in case of lost contact, because of high winds at ground level that week. Indeed, western Canada is covered by dense forest (Massey et al., 2018), so GNSS lock from the forest floor was not guaranteed.

Both capsules maintained Iridium link after landing, and continued reporting GNSS coordinates with standard deviation in latitude and longitude of 7 m from the first DRS, and 10 m from the second. We waited to receive a few dozen GNSS readings, to average away this noise, then commanded the capsules via Iridium MT message to conserve battery life and report back only every 2 hours. Both capsules had landed safely, on dry land.

The predicted trajectories were more accurate than expected (figures 5.7 and 5.8).

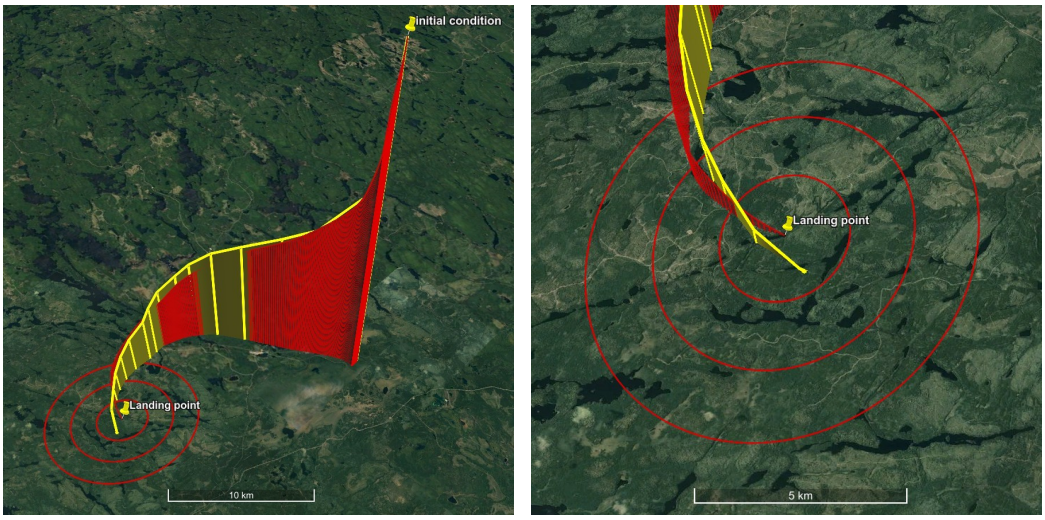


Figure 5.8: As figure 5.7, but showing the predicted (red) and GNSS (yellow) descent trajectory of the second DRS capsule. The prediction from the CNES software was used before dropping the capsule, but is no longer available for inclusion in this figure.

Predicted landing sites were within 300 m and 600 m of the true locations, which would have been adequate for successful recovery even without GNSS measurements. I obtained live predictions using an older version of the software than that available on github.* The current version is more accurate in general but – for these particular initial conditions – predicts landing sites within 600 m and 1100 m of the true locations, consistent with the expected uncertainty. Our live runs were noisier, and their particularly high accuracy was good luck.

### 5.5.3 Recovery

To aid recovery, the capsules are equipped with a sounder, and the parachutes are bright orange. A recovery crew went to the GNSS coordinates of both landing sites, and found both DRS capsules within a few minutes each. They had both fallen to the forest floor (figure 5.9), so no further action was necessary.

Upon return to the launch facility, the cases were opened to remove batteries and deactivate the sounders (they could have been deactivated remotely, but were in the

*For example, the 'leapfrog' method of updating the position and velocity discussed in section 5.4.2 was not implemented in the older version of the code.

Figure 5.9: Photos of the two capsules on the ground taken by the CNES recovery team Sébastian Lafrance and Francis Martin. The capsules are indicated by red circles. The parachutes can be clearly seen in bright orange.

back of an effectively soundproof truck). A few pine needles had entered the upper chamber of one DRS, but the inner chamber of both DRS capsules was clean. The Raspberry Pis were plugged into external power, and the data successfully retrieved.

## 5.6 Conclusions

Retrieving assets from a High Altitude Balloon (HAB) platform can mitigate the risk of total loss if the platform is damaged or lost upon landing. Mid-flight retrieval can also increase a mission's efficiency, if its initial performance is assessed, and subsequent operation improved. One solution to retrieve physical samples, or digital data acquired at too high a rate for transmission to the ground, is to jettison a small capsule that descends via parachute.

We have developed, and successfully tested the SuperBIT Data Recovery System (DRS) to 'download' up to 5 TB of data via parachute. We released two DRS capsules from $\sim$ 30 km altitude during a commissioning flight of the SuperBIT telescope in September 2019. SuperBIT is an astronomical telescope that operates

in the stratosphere for up to 100 days at a time. Both capsules landed safely, a few hundred metres from their predicted landing sites, and were easily recovered.

Hardware worked as envisaged. Several times during flight, the main gondola logged in to the DRS capsules via 2.4 GHz Wi-Fi™, and copied data onto them. At two different times, we issued a two-stage 'release' command to one DRS, via `ssh`. The capsules dropped 30 seconds later, and their parachutes opened. During and after descent, they measured their location via GNSS and transmitted it back to the ground station via Iridium message.

Software to predict the descent trajectory also worked well. After travelling a horizontal distance of 31 and 19 km from their release points, the DRS capsules landed within 300 m and 600 m of their expected landing sites. Calibrated on 30 parachute descents from the stratosphere, our software can predict landing sites all over the world with $1\sigma$ uncertainty of $\sim$1.5 km. This uncertainty accumulates most rapidly while the capsules descend through the jet stream. Our software thus appears limited mainly by the accuracy of (GFS) weather models at this altitude. Nonetheless, it satisfies safety requirements to permit immediate release — and it can also be used to predict the best time to release a capsule so that it can be conveniently recovered. This takes the form of a landing strip on the ground, roughly underneath the future path that the software predicts for the HAB platform.

During this test with SuperBIT, we used the DRS capsules as a means to retrieve digital data. However, we envisage that they could be used to retrieve a variety of assets, including hardware or physical samples. We welcome interest from other HAB teams for whom the system may be useful.

## 5.7 Updates for future flights

For SuperBIT's upcoming first fully operational science flight in 2023, we have decided to implement two major hardware updates. These are: (1) switching from data transfer via Wi-Fi$^{TM}$ to wired Ethernet, and (2) a thermal redesign to prevent the Raspberry Pis from overheating during, e.g. data transfer.

### 5.7.1 Wired Ethernet

We have opted to switch from data transfer via Wi-Fi$^{TM}$ to wired Ethernet, for faster data transfer, and to avoid any potential for radio frequency (RF) electromagnetic interference. We have flown 2.4 GHz Wi-Fi$^{TM}$ networks on both NASA and CSA/CNES balloons without any problems, but testing for that interference has frequently slowed payload integration, and has even delayed launch on one occasion. Additionally, this would extend the possible applications for the DRS to CMB experiments (e.g. SPIDER; Filippini et al., 2010) which are extremely sensitive to RF and would be unable to tolerate an onboard Wi-Fi$^{TM}$ network.

We have opted for Ethernet with Power-over-Ethernet, such that we could remove the original power cable and minimise the number of cables attached to the PCB. As we have switched to data transfer over wired Ethernet we have also switched to a 9-way low force extraction connector*.

### 5.7.2 Thermal redesign

The operating temperature for a Raspberry Pi is between 0° C and 85° C. Specifically, the CPU or System on Chip (SoC, the integrated circuit that does the Raspberry Pi's processing) is qualified from -40° C to 85° C; the USB and Ethernet controller of the Raspberry Pi is specified by the manufacturers as being qualified

*In particular, we now use D02PB906MSTH/D02PB906FSTAH, see www.smithsinterconnect.com/products/connectors/circular/d-series/.

from 0° C to 70° C. Effectively the maximum operating temperatures of a Raspberry Pi's key components are 70° C and 85° C. During the 2019 test flight, the Raspberry Pis reached high temperatures, especially considering we only transferred a small amount of data to test the system. To prevent the Raspberry Pis from overheating on future flights, when we will transfer much larger quantities of data, we have added aluminium heatsinks to the CPU as well as a smaller copper heatsink directly onto the RAM.

Figures 5.10 and 5.11 show the aluminium heatsink from the side and top respectively for one of our updated DRSs. At float, these heatsinks will be exposed to space to act as a radiator. The small copper heatsink, not attached to a Raspberry Pi, is shown in figure 5.12 with a pen for scale. The thermal redesign could possibly allow for the Raspberry Pis to be used for pre-processing and analysing science data during flight. This, however, has not yet been implemented or tested.

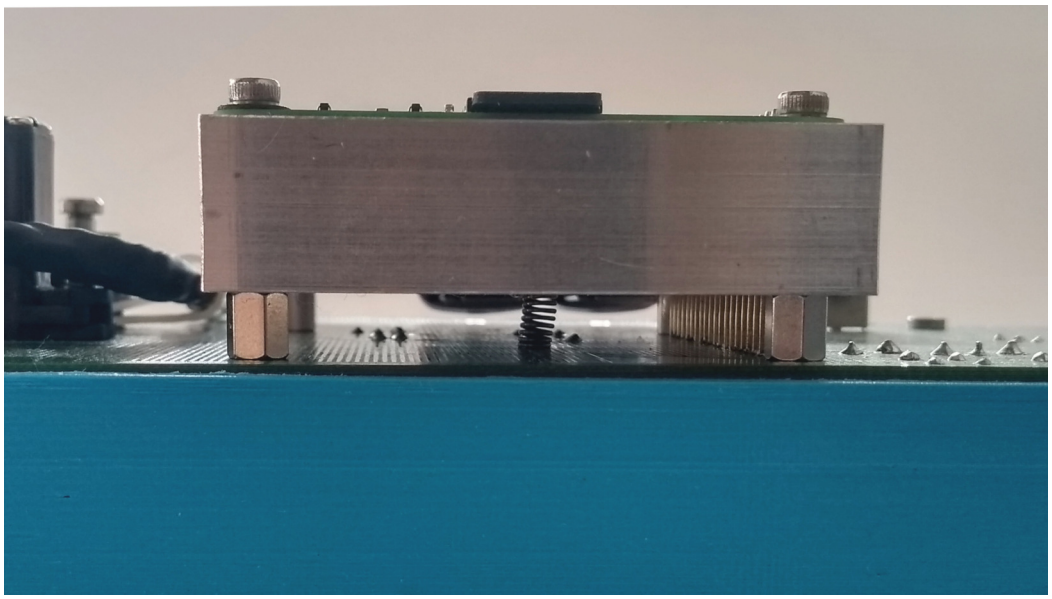

Figure 5.10: A Raspberry Pi of one of the DRSs as seen from the side. The Raspberry Pi is obscured by the aluminium heatsink. A spring can be seen underneath the Pi, pushing the heatsink against the main CPU but allowing a range of motion during thermal expansion.

Figure 5.13 shows images taken of a Raspberry Pi with an infrared camera. The left image shows the side of the Raspberry Pi with the CPU, and was taken while

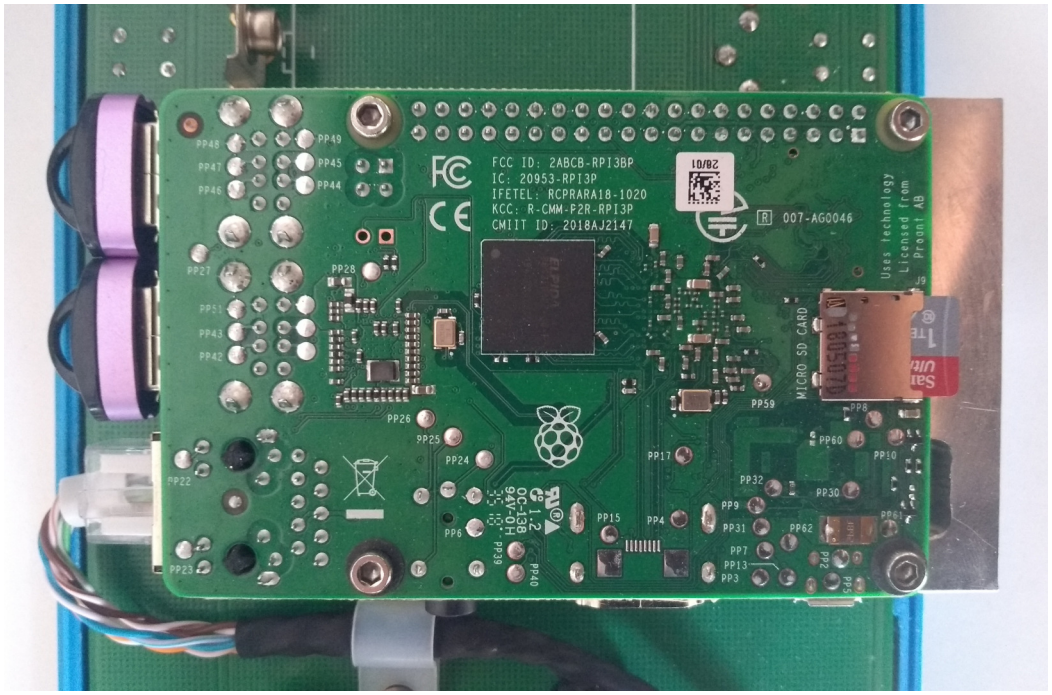


Figure 5.11: A Raspberry Pi of one of the DRSs as seen from the top. The aluminium heatsink can be seen sticking out on the right side of the image. The Ethernet cable, which doubles as the power cable, is plugged into the port on the bottom left below the four (only two are visible) purple SD card readers. The copper heatsink is not shown here.

Figure 5.12: A copper heatsink that will be added directly to the processor chip of the Raspberry Pis. The pen is shown for scale.

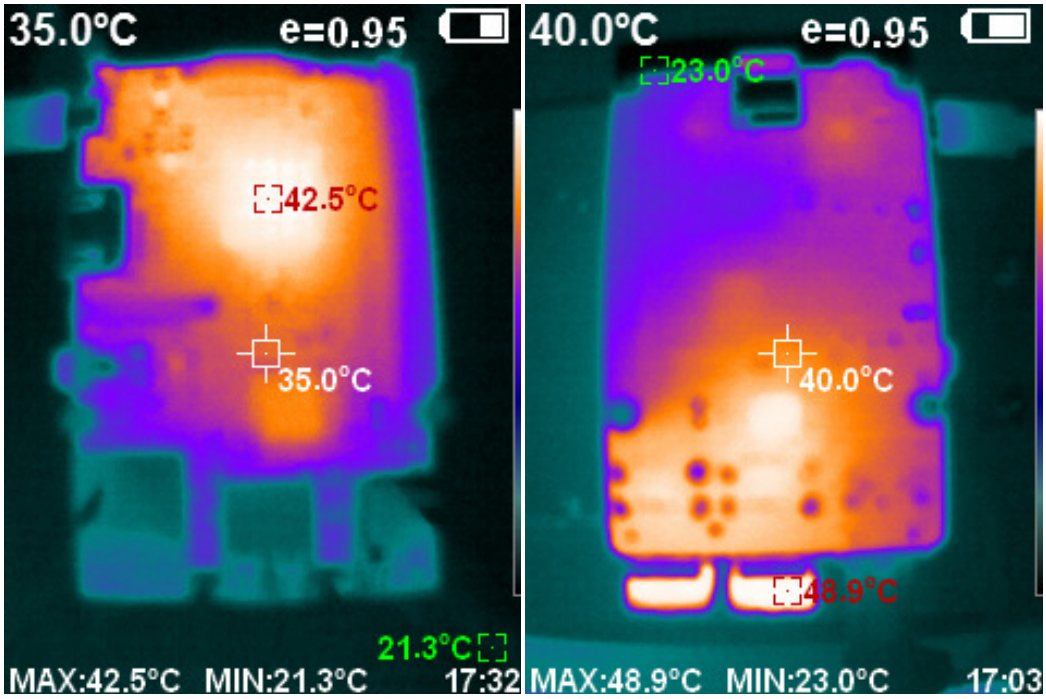


Figure 5.13: Images of a Raspberry Pi taken with an infrared camera. Each image shows one side of the Raspberry Pi. Left: This image was taken while a Python script printing some text 4000 times to the command line was running. No USB card readers were attached to the Raspberry Pi at this time. Right: This image was taken while a 5 GB file was being transferred to the storage of the DRS. The USB readers can be seen at the bottom left, the heatsink and the SD card containing the operating system of the Raspberry Pi at the top. The $e = 0.95$ at the top of each image represents the thermal emissivity.

a simple Python script printing some text 4000 times to the command line was running. The infrared camera showed a temperature of 42.5° C for the CPU, while the Raspberry Pi itself reported a CPU temperature of 49.4° C. The image on the right shows the side of the Raspberry Pi with the SD card containing the operating system. The image was taken while a file with a size of 5 GB was being transferred to the storage system. During transfer, the USB card readers heated up to a temperature of almost 50.0° C. We hope to mitigate the heating of the USB readers/storage system by throttling the file transfer. The heatsink became warmer than its surroundings, reaching a temperature of 23.0° C, showing that it is able to transfer heat away from the CPU. That fact that both of these images were taken during basic operations confirms the need for thermal mitigation.

### 5.7.3 Casing

Due to the addition of the heatsink and the Ethernet cable, the plastic case that contains the DRS had to be slightly modified. I added a gap to the side as the heatsink sticks too far out of the Raspberry Pi, as well as a notch in the wall of the casing that separates the Raspberry Pi from the servo and archery release system to accommodate the Ethernet/power cable. Figure 5.14 shows the side of the case with the gap for the heatsink.

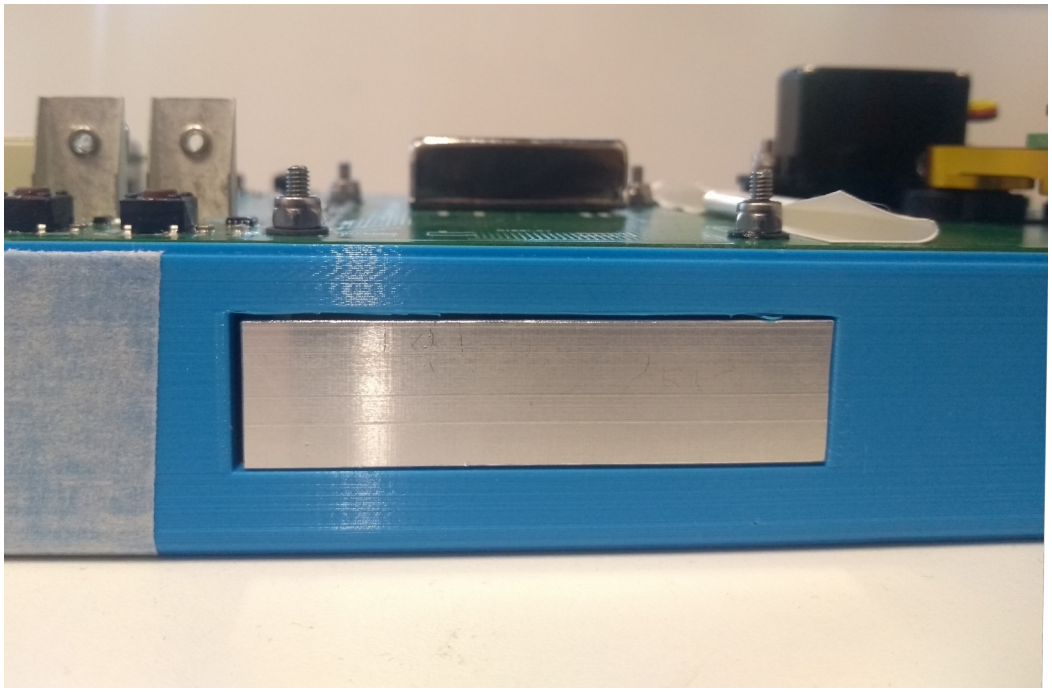

Figure 5.14: The casing as seen from the side. The aluminium heatsink fits snugly into the hole in the side. Upon installation, the heatsink will be painted white. At float, it will be exposed to space to act as a radiator.

### 5.7.4 Current state

Currently, one updated DRS has been build and fully tested. This DRS was sent to Palestine, Texas to be integrated with SuperBIT. We are currently testing four more DRSs and hope to send them to Palestine by the end of September. One more DRS is being build such that six DRSs will be ready to fly with SuperBIT

during its science flight. We have calculated that the mission will gather up to 20 Tb of data, so five DRSs are required, including one spare in case a DRS lands in a location from which we cannot retrieve it*.

*E.g. a lake...

Chapter 6

# Summary and Conclusions

*"Going home without my burden,*
*Going home behind the curtain,*
*Going home without this costume that I wore."*

— Leonard Cohen, *Old Ideas*

The existence of a component of matter that does not interact with the electromagnetic force has been proposed as early as the 1930s. However, after decades of research and experiments, we still do not know exactly what this DM is. Presently, the presence of DM has only been inferred through its gravitational effects on its surroundings. In this thesis, we investigated some of the effects of a type of DM particle which can elastically scatter from other DM particles, known as SIDM. In particular, we investigated the effects of DM self-interactions on galaxy clusters and the galaxies that reside in these environments.

We began in chapter 2 with an investigation of the effects of DM self-interactions on the mass loss of galaxies accreted unto galaxy clusters. As galaxies fall into clusters, they are subject to violent interactions with their environment. In an ΛSIDM universe, interactions between the DM of the galaxy halo and the DM of the cluster halo can cause the DM to be scattered out, thus providing a channel for mass loss in addition to tidal stripping alone. This additional mass loss is referred to as *subhalo evaporation*.

We used hydrodynamical cosmological simulations run with CDM and SIDM physics to compare the mass loss of galaxies falling into clusters. The CDM and SIDM versions of our simulated clusters were started from identical initial conditions, and as such we could directly match galaxies between the simulations and compare their evolution. When comparing the phase-space properties of individual galaxies, we indeed saw evidence of tidal mass loss in both the CDM and SIDM galaxies, as well as additionally lost material due to the subhalo evaporation in the SIDM galaxy. We then considered all galaxies that ever fell into the clusters, including those that have since been disrupted, have merged, or have left the cluster. We found that by present time a larger fraction of the SIDM galaxies had disrupted compared to the galaxies in the CDM version of the same cluster and that the remaining SIDM galaxies had on average lost more DM as well. Over 33 per cent of galaxies in an SIDM cluster can be entirely disrupted by present time, compared to 20 per cent in a CDM cluster. When comparing matched galaxies between the CDM and SIDM versions of a given cluster, we find significant differences in mass loss. However, when we looked at the population of galaxies remaining in the cluster at $z = 0$, we find considerably smaller differences. This is most likely due to the fact that there is a large group of disrupted SIDM galaxies which does not contribute to the signal at $z = 0$.

We found that potentially observable ways to discriminate between CDM and SIDM include the high mass normalisation of the stellar-to-halo mass relation of galaxies in clusters, compared to galaxies in the field, which describes the mass of the DM in a galaxy of fixed stellar mass. The absolute normalisation of the relation would make it much easier to discriminate SIDM from CDM, but this depends to some extent on the subgrid physics of the simulations. However, as in the field the relation is nearly indistinguishable for a CDM and SIDM universe, one could use the difference between the field and cluster relations at a given stellar mass to try and discriminate between the two models. We find that, at approximately the stellar mass of the MW, the ratio $M_\star/M_{\rm tot}$ is 8 and 13 times higher in the cluster

compared to the field for the CDM and SIDM versions of the most massive of our simulated clusters respectively. While challenging, we estimate that with noise-free measurements of ∼32 cluster galaxies such a test could be performed.

In chapter 3, we investigated the effects of self-interactions on the major mergers between galaxy clusters. Galaxies are collisionless and essentially pass through a merger unaffected. The gas, however, gets dissociated from the galaxies due to ram pressure and lags behind the galaxies after the merger. If the SIDM cross-section is zero, the DM should remain incident with the galaxies. On the other hand if DM particles can indeed interact and scatter, an offset from the galaxies could be induced.

We used the shrinking-spheres method to determine the centres of the particle distributions of the DM, gas and stars of a number of simulated galaxy clusters and massive substructures residing in their vicinity. We then measured the offset of the DM as a dimensionless fractional lag, given by the offset between the DM and the stars divided by the offset between the stars and gas. Using a fractional lag removes dependence on the angle of the collision with respect to the line-of-sight, and it represents a physical quantity that the analytic models suggest should be identical for all merger configurations, so measurements from different systems can be averaged. As expected from analytical models, we find that the average offset of the DM increases with cross-section, and could potentially be used to discriminate between models of DM. However, from these analytical models, zero cross-section of the CDM simulations is expected to on average produce zero offset. We find that this is not the case, and that our median offset is slightly positive. We suggest this could be a result of the large gravitational potential of the DM pulling the gas with it and thus offsetting the DM and gas in the same direction from the stars. It may therefore be necessary to compare observations against full simulations that include these effects, rather than against simple, analytic models.

Finally, we looked at the effects of introducing weights to find our median offset. In particular, we used weights that give great importance to systems with a large

offset between the stars and the gas. The idea is that the fractional lag is highly uncertain when the stars-gas offset is small compared to the uncertainty in the stars-DM offset. Conversely, a large star-gas offset provides a stable baseline from which to measure the fractional lag. A single subhalo with large star-gas offset increases the weighted median offset to nearly 10 times its value. It seems this weighting scheme is most suitable when clean measurements are available, i.e. when random statistical errors dominate over systematic errors. In our case, the final centre found is sometimes the centre of a nearby halo, resulting in an artificially high star-gas offset, giving the halo a large weight.

We wish to extend this work by applying observational techniques to the same simulated data. In particular, we wish to use weak gravitational lensing methods to find the centre of the distribution. Gravitational lensing depends on the total mass of the object acting as the lens, i.e. we would be using the centre of the total mass as a proxy for the centre of the DM. In terms of the gas, we would obtain surface brightness maps, and then use peak-finding software to obtain the centres. For the galaxies we would take the BCG to be the centre, which could also be found with peak-finding software. Having found the centres of the DM, gas, and stars using observational techniques, we will compare our results to those presented in chapter 3.

In the second half of this thesis we focused on the balloon-borne astronomical telescope SuperBIT. We introduced the astronomical background and technical aspects of the telescope in chapter 4, and briefly described the test flights performed in preparation for its upcoming first fully operational science flight scheduled for the summer of 2023. The main science goals of SuperBIT are to map out the DM around galaxy clusters and the large scale structure of the Universe. Considering that the data is expected to be of similar quality to HST, SuperBIT could be used to compare our theoretical predictions from chapter 2 to observations. After we have applied observational techniques to the simulated data from chapter 3, we hope to make projections for the SuperBIT flight as well.

In chapter 5 we described the DRS, a toolkit for recovering data from any stratospheric balloon platform any time it is over land, developed to keep up with SuperBIT's high data rate. The DRS is currently capable of 'downloading' up to 5 TB of data via parachute. We also introduced software we developed for the prediction of the flight trajectories of the DRSs given the date, time, and location of release. During SuperBIT's 2019 test flight, we released two DRS capsules from ∼30 km altitude using the prediction software to target safe and easily accessible landing sites. The hardware worked as envisaged, and both capsules landed safely, a few hundred metres from their predicted landing sites, and were easily recovered.

Finally, we described some hardware updates that are currently being implemented to the DRS for the SuperBIT 2023 flight. We have opted to switch from data transfer over WiFi$^{TM}$ to wired Ethernet. To mitigate overheating during data transfer we have added a heatsink to the Raspberry Pi, which constitutes the front-end user interface for the DRS as well as the heart of the recoverable assets hosting the data storage. Six updated DRSs are currently being developed and tested.

Our hope was to bring the three projects together and use the SuperBIT hardware to measure the behaviour of DM and calibrate it against the cosmological simulations. Unfortunately SuperBIT's first science flight was delayed due to the COVID-19 pandemic, and it was not yet possible to measure the DM effects on real astronomical data. We hope to perform the suggested tests from chapters 2 and 3 on the data from SuperBIT's 2023 flight, and possibly constrain the SIDM cross-section with these tests.

In this thesis we have discussed various ways of utilising galaxy clusters to constrain the DM self-interacting cross-section. Yoo et al. (2022) studied the correlation between the spatial distribution of DM in clusters with various luminous components, such as satellites galaxies, the BCG, and the inter-cluster light (ICL). They developed a new methodology to quantify the similarity of two-dimensional spatial distributions. With this novel methodology, Yoo et al. (2022) found that

that the best luminous tracer for DM is the combination of the BCG and the ICL. Moreover, they found that galaxy clusters that were more relaxed showed tighter correlations, which could allow the method to be used as a dynamical stage indicator for clusters. As such, their method could possibly be used to constrain DM models such as the SIDM or CDM model, since these models predict different tidal interaction histories as we have shown in chapter 2. We hope to use the method proposed by Yoo et al. (2022) to study the C-EAGLE clusters run with CDM and SIDM physics, and compare the results.

# Bibliography


Akerib D. S., et al., 2013, Nuclear Instruments and Methods in Physics Research A, 704, 111

Akeson R., et al., 2019, arXiv e-prints, p. arXiv:1902.05569

Alam S., et al., 2021, Phys. Rev. D, 103, 083533

Alpher R. A., Bethe H., Gamow G., 1948, Phys. Rev., 73, 803

Aprile E., et al., 2017, Phys. Rev. Lett., 119, 181301

Aprile E., et al., 2020, JCAP, 2020, 031

Atwood W. B., et al., 2009, ApJ, 697, 1071

Bahé Y. M., et al., 2017, MNRAS, 470, 4186

Barnes D. J., et al., 2017, MNRAS, 471, 1088

Bartelmann M., 2010, Classical and Quantum Gravity, 27, 233001

Bechtol K., et al., 2015, ApJ, 807, 50

Bennett C. L., et al., 1996, ApJL, 464, L1

Berk A., et al., 1999, in Larar A. M., ed., Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 3756, Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research III. pp 348–353, doi:10.1117/12.366388

de Bernardis P., et al., 1999, NAR, 43, 289

de Bernardis P., et al., 2000, Nature, 404, 955

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Bhattacharyya S., Adhikari S., Banerjee A., More S., Kumar A., Nadler E. O., Chatterjee S., 2022, ApJ, 932, 30

Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton University Press, Princeton

de Blok W. J. G., Bosma A., McGaugh S., 2003, MNRAS, 340, 657

van den Bosch F. C., Aquino D., Yang X., Mo H. J., Pasquali A., McIntosh D. H., Weinmann S. M., Kang X., 2008, MNRAS, 387, 79

Bose S., et al., 2017, MNRAS, 464, 4520

Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, MNRAS, 415, L40

Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2012, MNRAS, 422, 1203

Bradač M., Allen S. W., Treu T., Ebeling H., Massey R., Morris R. G., von der Linden A., Applegate D., 2008, ApJ, 687, 959

Bringmann T., Huang X., Ibarra A., Vogl S., Weniger C., 2012, JCAP, 2012, 054

Brooks A. M., Kuhlen M., Zolotov A., Hooper D., 2013, ApJ, 765, 22

Brown T. M., et al., 2014, ApJ, 796, 91

Bullock J. S., Boylan-Kolchin M., 2017, ARA&A, 55, 343

Bullock J. S., Kravtsov A. V., Weinberg D. H., 2000, ApJ, 539, 517

Burkert A., 2000, ApJL, 534, L143

Clark P., et al., 2014, Journal of Astronomical Instrumentation, 3, 1440003

Clark P., et al., 2019, Journal of Instrumentation, 14, P04003

Clifton T., Ferreira P. G., Padilla A., Skordis C., 2012, Phys. Rep., 513, 1

Clowe D., Bradač M., Gonzalez A. H., Markevitch M., Randall S. W., Jones C., Zaritsky D., 2006, ApJL, 648, L109

Copeland E. J., Sami M., Tsujikawa S., 2006, Int. J. Mod. Phys. D, 15, 1753

Crain R. A., et al., 2015, MNRAS, 450, 1937

Dalcanton J. J., Stilp A. M., 2010, ApJ, 721, 547

Danielson R. E., Gaustad J. E., Schwarzschild M., Weaver H. F., Woolf N. J., 1964, AJ, 69, 344

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

Dawson W. A., et al., 2012, ApJ, 747, L42

Dekker A., Ando S., Correa C. A., Ng K. C. Y., 2021, arXiv e-prints, p. arXiv:2111.13137

Di Cintio A., Tremmel M., Governato F., Pontzen A., Zavala J., Bastidas Fry A., Brooks A., Vogelsberger M., 2017, MNRAS, 469, 2845

Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497

Dooley G. A., Peter A. H. G., Vogelsberger M., Zavala J., Frebel A., 2016, MNRAS, 461, 710

Drlica-Wagner A., et al., 2015, ApJ, 813, 109

Einstein A., de Sitter W., 1932, PNAS, 18, 213

Elbert O. D., Bullock J. S., Kaplinghat M., Garrison-Kimmel S., Graus A. S., Rocha M., 2018, ApJ, 853, 109

Fattahi A., Navarro J. F., Sawala T., Frenk C. S., Sales L. V., Oman K., Schaller M., Wang J., 2016, arXiv e-prints, p. arXiv:1607.06479

Faucher-Giguŕe C.-A., 2017, MNRAS, 473, 3717

Ferrero I., Abadi M. G., Navarro J. F., Sales L. V., Gurovich S., 2012, MNRAS, 425, 2817

Filippini J. P., et al., 2010, in Holland W. S., Zmuidzinas J., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7741, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy V. p. 77411N (*arXiv:1106.2158*), doi:10.1117/12.857720

Firanj Sremac A., Salehi A., 2018, Agricultural Meteorology and Climatology. Firenze University Press

Flores R. A., Primack J. R., 1994, ApJL, 427, L1

Flores F., Rondanelli R., Díaz M. A., Querel R. R., Mundnich K., Herrera L. A., Pola D. A., Carricajo T., 2013, Bulletin of the American Meteorological Society, 94, 187

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Frenk C. S., White S. D. M., 2012, Annalen der Physik, 524, 507

Friedmann A., 1922, Zeitschrift fur Physik, 10, 377

Gardner J. P., et al., 2006, Space Sci. Rev., 123, 485

Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., Kirby E. N., 2014, MNRAS, 444, 222

George M. R., et al., 2012, ApJ, 757, 2

Gill A., et al., 2020, AJ, 160, 266

Hafez I., 2010, PhD thesis, James Cook University

Hargis J. R., Willman B., Peter A. H. G., 2014, ApJL, 795, L13

Harvey D., et al., 2014, MNRAS, 441, 404

Harvey D., Massey R., Kitching T., Taylor A., Tittley E., 2015, Science, 347, 1462

Harvey D., Courbin F., Kneib J. P., McCarthy I. G., 2017, MNRAS, 472, 1972

Hillenbrand L. A., Massey P., Strom S. E., Merrill K. M., 1993, AJ, 106, 1906

Hilton M., et al., 2021, ApJS, 253, 3

Hinshaw G., et al., 2013, ApJS, 208, 19

Holder J., et al., 2008, in Aharonian F. A., Hofmann W., Rieger F., eds, American Institute of Physics Conference Series Vol. 1085, American Institute of Physics Conference Series. pp 657–660 (*arXiv:0810.0474*), doi:10.1063/1.3076760

Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581

Hoskin M. A., 1976, Journal for the History of Astronomy, 7, 169

Hubble E. P., 1925, Popular Astronomy, 33, 252

Hubble E. P., 1929, PNAS, 15, 168

Jiang L., Helly J. C., Cole S., Frenk C. S., 2014, MNRAS, 440, 2115

Jones W., et al., 2019, in Bulletin of the American Astronomical Society. p. 171

Jungman G., Kamionkowski M., Griest K., 1996, Phys. Rep., 267, 195

Kaplinghat M., Keeley R. E., Linden T., Yu H.-B., 2014, Phys. Rev. Lett., 113, 021302

Karoly D. J., Vincent D. G., 1998, Meteorological Monographs, 27, 1

Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201

Kim S. Y., Peter A. H. G., Wittman D., 2017, MNRAS, 469, 1414

Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, ApJ, 522, 82

Klypin A., Karachentsev I., Makarov D., Nasonova O., 2015, MNRAS, 454, 1798

LSST Science Collaboration et al., 2009, arXiv e-prints, p. arXiv:0912.0201

Laureijs R., et al., 2011, Euclid Definition Study Report (*arXiv:1110.3193*)

Lemaître G., 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Li L., 2015, PhD thesis, University of Toronto

Li L., Damaren C. J., Romualdez L. J., Netterfield C. B., Hartley J. W., Galloway M. N., Massey R. J., Clark P., 2016, Journal of Aerospace Engineering, 29, 04016051

Markevitch M., Gonzalez A. H., Clowe D., Vikhlinin A., Forman W., Jones C., Murray S., Tucker W., 2004, ApJ, 606, 819

Marsh D. J. E., 2016, Phys. Rep., 643, 1

Massey R., et al., 2007, MNRAS, 376, 13

Massey R., Kitching T., Richard J., 2010, Rep. Prog. Phys., 73, 086901

Massey R., Goetz S. J., Berner L. T., Mack M. C., Rogers B. M., 2018, in AGU Fall Meeting Abstracts. pp B14C–06

Mateo M. L., 1998, ARA&A, 36, 435

McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, MNRAS, 383, 593

McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, MNRAS, 465, 2936

Meneghetti M., Yoshida N., Bartelmann M., Moscardini L., Springel V., Tormen G., White S. D. M., 2001, MNRAS, 325, 435

Milgrom M., 1983, ApJ, 270, 365

Miralda-Escudé J., 2002, ApJ, 564, 60

Moore B., 1994, Nature, 370, 629

Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJL, 524, L19

Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121

Murata R., et al., 2019, PASJ, 71, 107

Navarro J. F., Eke V. R., Frenk C. S., 1996, MNRAS, 283, L72

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Newman A. B., Treu T., Ellis R. S., Sand D. J., Richard J., Marshall P. J., Capak P., Miyazaki S., 2009, ApJ, 706, 1078

Newman A. B., Treu T., Ellis R. S., Sand D. J., 2011, ApJ, 728, L39

Newman A. B., Treu T., Ellis R. S., Sand D. J., Nipoti C., Richard J., Jullo E., 2013a, ApJ, 765, 24

Newman A. B., Treu T., Ellis R. S., Sand D. J., 2013b, ApJ, 765, 25

Niemiec A., Jullo E., Giocoli C., Limousin M., Jauzac M., 2019, MNRAS, 487, 653

Oh S.-H., de Blok W. J. G., Brinks E., Walter F., Kennicutt Robert C. J., 2011, AJ, 141, 193

Oman K. A., et al., 2015, MNRAS, 452, 3650

Oman K. A., Bahé Y. M., Healy J., Hess K. M., Hudson M. J., Verheijen M. A. W., 2021, MNRAS, 501, 5073

Oñorbe J., Boylan-Kolchin M., Bullock J. S., Hopkins P. F., Kereš D., Faucher-Giguère C.-A., Quataert E., Murray N., 2015, MNRAS, 454, 2092

Papastergis E., Giovanelli R., Haynes M. P., Shankar F., 2015, A&A, 574, A113

Pascale E., Butler N., Nagler P., Netterfield C. B., Tucker G., The Excite Collaboration 2021, in European Planetary Science Congress. pp EPSC2021–98, doi:10.5194/epsc2021-98

Peñarrubia J., Pontzen A., Walker M. G., Koposov S. E., 2012, ApJL, 759, L42

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Peter A. H. G., Rocha M., Bullock J. S., Kaplinghat M., 2013, MNRAS, 430, 105

Pineda J. C. B., Hayward C. C., Springel V., Mendes de Oliveira C., 2016, MNRAS, 466, 63

Planck Collaboration et al., 2014, A&A, 571, A16

Planck Collaboration et al., 2016, A&A, 594, A13

Planck Collaboration et al., 2020, A&A, 641, A6

Pontzen A., Governato F., 2012, MNRAS, 421, 3464

Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, MNRAS, 338, 14

Randall S. W., Markevitch M., Clowe D., Gonzalez A. H., Bradač M., 2008, ApJ, 679, 1173

Read J. I., 2014, Journal of Physics G Nuclear Physics, 41, 063101

Redmond S., 2018, PhD thesis, University of Toronto

Redmond S., et al., 2018, in Marshall H. K., Spyromilio J., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 10700,

Ground-based and Airborne Telescopes VII. p. 107005R (*arXiv:1807.09869*), doi:10.1117/12.2312339

Riess A. G., et al., 1998, AJ, 116, 1009

Robertson A., 2017, PhD thesis, Durham University

Robertson A., Massey R., Eke V., 2017a, MNRAS, 465, 569

Robertson A., Massey R., Eke V., 2017b, MNRAS, 467, 4719

Robertson A., et al., 2018, MNRAS, 476, L20

Robertson A., Harvey D., Massey R., Eke V., McCarthy I. G., Jauzac M., Li B., Schaye J., 2019, MNRAS, 488, 3646

Rocha M., Peter A. H. G., Bullock J. S., Kaplinghat M., Garrison-Kimmel S., Oñorbe J., Moustakas L. A., 2013, MNRAS, 430, 81–104

Rodríguez-Torres S. A., et al., 2016, MNRAS, 460, 1173

Romualdez L. J., 2018, PhD thesis, University of Toronto (Canada)

Romualdez L. J., Clark P., Damaren C. J., Galloway M. N., Hartley J. W., Li L., Massey R. J., Netterfield C. B., 2016a, arXiv e-prints, p. arXiv:1603.01161

Romualdez L. J., et al., 2016b, arXiv e-prints, p. arXiv:1608.02502

Romualdez L. J., et al., 2018, in Evans C. J., Simard L., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 10702, Ground-based and Airborne Instrumentation for Astronomy VII. p. 107020R (*arXiv:1807.02887*), doi:10.1117/12.2307754

Romualdez L. J., et al., 2020, Review of Scientific Instruments, 91, 034501

Rubin V. C., Ford W. Kent J., 1970, ApJ, 159, 379

Sand D. J., Treu T., Ellis R. S., 2002, ApJ, 574, L129

Sand D. J., Treu T., Smith G. P., Ellis R. S., 2004, ApJ, 604, 88

Sawala T., et al., 2016, MNRAS, 457, 1931

Schaye J., et al., 2015, MNRAS, 446, 521

Schwarzschild M., Schwarzschild B., 1959, Scientific American, 200, 52

Shen S., Madau P., Conroy C., Governato F., Mayer L., 2014, ApJ, 792, 99

Simet M., McClintock T., Mandelbaum R., Rozo E., Rykoff E., Sheldon E., Wechsler R. H., 2017, MNRAS, 466, 3103

Sirks E. L., et al., 2020, Journal of Instrumentation, 15, P05014

Slipher V. M., 1913, Lowell Observatory Bulletin, 1, 56

Slipher V. M., 1915, Popular Astronomy, 23, 21

Slipher V. M., 1917, Proceedings of the American Philosophical Society, 56, 403

Smoot G. F., et al., 1992, ApJL, 396, L1

Sobacchi E., Mesinger A., 2013, MNRAS, 432, 3340

Somerville R. S., 2002, ApJL, 572, L23

Spergel D. N., Steinhardt P. J., 2000, Phys. Rev. Lett., 84, 3760

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Springel V., Frenk C. S., White S. D. M., 2006, Nature, 440, 1137

Springel V., et al., 2008, MNRAS, 391, 1685

The LZ Collaboration et al., 2019, arXiv e-prints, p. arXiv:1910.09124

Trujillo-Gomez S., Klypin A., Colín P., Ceverino D., Arraki K. S., Primack J., 2014, MNRAS, 446, 1140

Tulin S., Yu H.-B., 2018, Phys. Rep., 730, 1

UN, Committee of Experts on the Transport of Dangerous Goods 2019, Recommendations on the Transport of Dangerous Goods.: Manual of tests and criteria. UN, New York, http://digitallibrary.un.org/record/3846833

Vargya D., Sanderson R., Sameie O., Boylan-Kolchin M., Hopkins P. F., Wetzel A., Graus A., 2021, arXiv e-prints, p. arXiv:2104.14069

Vega-Ferrero J., Dana J. M., Diego J. M., Yepes G., Cui W., Meneghetti M., 2021, MNRAS, 500, 247

Wang J., Frenk C. S., Navarro J. F., Gao L., Sawala T., 2012, MNRAS, 424, 2715

Wheeler C., Phillips J. I., Cooper M. C., Boylan-Kolchin M., Bullock J. S., 2014, MNRAS, 442, 1396

Wittman D., Golovich N., Dawson W. A., 2018, ApJ, 869, 104

Wong K. C., et al., 2020, MNRAS, 498, 1420

Wright T., 2014, An Original Theory or New Hypothesis of the Universe, Founded upon the Laws of Nature: And Solving by Mathematical Principles the General Phænomena of the Visible Creation, and Particularly the Via Lactea. Cambridge Library Collection - Astronomy, Cambridge University Press, doi:10.1017/CBO9781107447370

Yoo J., et al., 2022, ApJS, 261, 28

Yoshida N., Springel V., White S. D. M., Tormen G., 2000, ApJ, 544, L87

Zavala J., Vogelsberger M., Walker M. G., 2013, MNRAS Lett., 431, L20

Zwicky F., 1933, Helvetica Physica Acta, 6, 110

## Colophon

This thesis is based on a template developed by Matthew Townson and Andrew Reeves. It was typeset with LaTeX $2_\varepsilon$. It was created using the *memoir* package, maintained by Lars Madsen, with the *madsen* chapter style. The font used is Latin Modern, derived from fonts designed by Donald E. Kunith.