

A History of Dark Matter

Gianfranco Bertone¹ and Dan Hooper^{2,3}

¹*GRAPPA, University of Amsterdam, Netherlands*

²*Center for Particle Astrophysics, Fermi National Accelerator Laboratory, USA and*

³*Department of Astronomy and Astrophysics, The University of Chicago, USA*

(Dated: May 26, 2016)

Although dark matter is a central element of modern cosmology, the history of how it became accepted as part of the dominant paradigm is often ignored or condensed into a brief anecdotal account focused around the work of a few pioneering scientists. The aim of this review is to provide the reader with a broader historical perspective on the observational discoveries and the theoretical arguments that led the scientific community to adopt dark matter as an essential part of the standard cosmological model.

CONTENTS

I. Preface	4
II. Prehistory	6
A. From Epicurus to Galileo	6
B. Dark Stars, Dark Planets, Dark Clouds	7
C. Dynamical Evidence	9
III. Galaxy Clusters	13
A. Zwicky and Smith	13
B. A Confusing Situation	15
IV. Galactic Rotation Curves	18
A. The Beginnings	18
B. The 1970s Revolution	21
C. Local Measurements	27
V. Dark Matter Particles	30
A. Neutrinos	31
B. Supersymmetry	36
C. Axions	40
D. The WIMP Paradigm	42
VI. Baryonic Dark Matter	44
A. Gravitational Microlensing	44
B. The Universe's Baryon Budget	46
C. Primordial Black Holes	48
VII. Modified Gravity	50
A. Toward a Realistic Theory of MOND	50
B. Observational Successes and Failures	52
VIII. Piecing the puzzle	55
A. Discrepancies At All Scales	55

	3
B. Cosmology	57
C. Numerical Simulations	58
IX. The Hunt for Dark Matter Particles	63
A. Scattering with Nuclei	63
B. Annihilation and Decay	67
A. Further Reading	71
References	72

I. PREFACE

Dark matter plays a central role in our understanding of modern cosmology. But despite its significance, many of the scientists active in this area of research know relatively little about its interesting history, and how it came to be accepted as the standard explanation for a wide variety of astrophysical observations. Most publications and presentations on this topic – whether at a technical or a popular level – either ignore the long history of this field or condense it into a brief anecdotal account, typically centered around the work on galaxy clusters by Fritz Zwicky in the 1930s and on galactic rotation curves by Vera Rubin in the 1970s. Only a small number of scientists, and an even smaller number of historians, have endeavoured to systematically analyze the development of the dark matter problem from an historical perspective, and it is surprisingly hard to find articles and books that do justice to the fascinating history of dark matter.

The aim of this article is to provide a review of the theoretical arguments and observations that led to the establishment of dark matter among the pillars of modern cosmology, as well as of the theories that have been proposed to explain its nature. Although we briefly discuss some early ideas and recent developments, the focus of this review is the 20th century, beginning with the first dynamical estimates of dark matter’s abundance in the Universe, and to its role in the current standard cosmological model, and the strategies that have been pursued to reveal its particle nature.

The first part of this article is largely based on the analysis of primary sources, mainly scanned versions of scientific articles and books published in the 19th and 20th centuries, freely accessible via NASA ADS and the Internet Archive Project. We study the emergence of the concept of dark matter in the late 19th century and identify a series of articles and other sources that describe the first dynamical estimates for its abundance in the known Universe (Chapter II). We then discuss the pioneering work of Zwicky within the context of the scientific developments of the early 20th century. And although his work clearly stands out in terms of methodology and significance, we find that his use of the term “dark matter” was in continuity with the contemporary scientific literature. We then go on to follow the subsequent development of the virial discrepancy that he discovered, with particular emphasis on the debate that took place around this issue in the 1960s (Chapter III).

The second part of this article focuses on more recent developments, which gave us the

opportunity to complement the analysis of the primary sources with extensive discussions with some of the pioneering scientists who contributed to the advancement of this field of research. We discuss the history of galactic rotation curves, from the early work in the 1920s and 1930s to the establishment of flat rotation curves in the 1970s, placing the famous work of Bosma and Rubin and collaborators in 1978 within the broader context of the theories and observations that were available at that time (Chapter IV). We then discuss the theories that have been put forward to explain the nature of dark matter, in terms of fundamental particles (Chapter V), astrophysical objects (Chapter VI), or manifestations of non-Newtonian gravity or dynamics (Chapter VII).

Finally, we discuss how the emergence of cosmology as a science in the 1960s and 1970s, the advent on numerical simulations in the 1980s, and the convergence between particle physics and cosmology, led most of the scientific community to accept the idea that dark matter was made of non-baryonic particles (Chapter VIII), and prompted the development of new ideas and techniques to search for dark matter candidates, many of which are still being pursued today (Chapter IX).

One of the main difficulties in reconstructing the history of dark matter is that the key developments took place in a continuously changing landscape of cosmology and particle physics, in which scientists were repeatedly forced to revise their theories and beliefs. The authors of this review are not professional historians, but scientists writing for other scientists. And although we have taken great care in reconstructing the contributions of individuals and groups of scientists, we have little doubt that that our work falls short of the standards of the historical profession. We nevertheless hope that this article will contribute to a better understanding and appreciation of the history of dark matter among our fellow astronomers and physicists, and that it will foster an interest among professional historians in this rich and fascinating field of research.

II. PREHISTORY

A. From Epicurus to Galileo

Throughout history, natural philosophers have speculated about the nature of matter, and even have contemplated the possibility that there may be forms of matter that are imperceptible – because they were either too far away, too dim, or intrinsically invisible. And although many of the earliest scientific inquiries were less than rigorous, and often inseparable from philosophy and theology, they reveal to us the longevity of our species’ desire to understand the world and its contents.

Although many early civilizations imagined their own cosmological systems, it was arguably the ancient Greeks who were the first to attempt the construction of such a model based on reason and experience. The atomists, most famously Leucippus and Democritus who lived in the 5th century BCE, were convinced that all matter was made of the same fundamental and indivisible building blocks, called atoms, and that these atoms were infinite in number, as was the infinite space that contained them. Epicurus (341 BCE – 270 BCE) further suggested in his “Letter to Herodotus” that an infinite number of other worlds existed as well, “some like this world, others unlike it”¹. Others speculated about unobservable matter that might be found within our own Universe. For example, the Pythagorean Philolaus conjectured the existence of the celestial body *Antichthon*, or counter-earth, which revolves on the opposite side of the “central fire” with respect to the Earth [187].

The cosmological model of Aristotle – which would dominate discourse throughout the Middle Ages – provided an elegant construction, in which the location of the Earth was fixed to the center of an immutable Universe. This model offered what seemed to many to be strong arguments against the existence of invisible or unknown forms of matter. Even the striking appearance of comets, which obviously had no place in Aristotle’s highly organized hierarchy of celestial spheres, was dismissed as an atmospheric phenomenon, a belief that continued to be held until Tycho Brahe measured the (absence of) parallax for a comet in 1577.

Although many offered challenges to the orthodoxy of Aristotelian cosmology, these attempts were not met without resistance. The statue of Giordano Bruno in Campo de’ Fiori

¹ Epicurus, Letter to Herodotus (c. 305 BCE), Extracted from Diogenes Laertius, Lives of Eminent Philosophers, trans. R. D. Hicks, vol. 2 (1925).

in downtown Rome serves as a reminder of the dangers that were inherent in such departures from the strict Aristotelian worldview embraced by the Catholic Church. It was at the location of that statue that Bruno was burned at the stake in 1600 by the Roman Inquisition, after being convicted on charges that included the holding of a heretical belief in the existence of infinite other worlds.

It was arguably Galileo – who himself had his share of trouble with the inquisition – who did the most to break the hold of Aristotelian cosmology. By pointing his telescope toward the sky, Galileo saw much that had been previously imperceptible. Among his many other discoveries, he learned that the faint glow of the Milky Way is produced by a myriad of individual stars, and that at least four satellites, invisible to the naked eye, are in orbit around Jupiter. Each of these observations encapsulate two lessons that remain relevant to dark matter today. First, the Universe may contain matter that cannot be perceived by ordinary means. And second, the introduction of new technology can reveal to us forms of matter that had previously been invisible.

B. Dark Stars, Dark Planets, Dark Clouds

The course of science, and of astronomy in particular, was transformed in 1687 when Isaac Newton published his treatise *Philosophiæ Naturalis Principia Mathematica*. Newton’s Laws of motion and Universal Gravitation provided scientists with new and formidable tools which, among many other things, enabled them to determine the gravitational mass of astronomical bodies by measuring their dynamical properties.

In 1783, John Michell, also famous for inventing the torsion balance for the measurement of the force of gravity, realized that if light is affected by the laws of gravity – as he reasoned it should, given the universal nature of gravity² – then there could potentially exist objects that are so massive that even light would not be able to escape their gravitational pull [213]. This proposal, also famously discussed a decade later by Pierre Simon Laplace, is often considered to be the first mention of what have become known as black holes. We mention it here, however, as an explicit example of a discussion of a class of invisible astrophysical objects, that populate the universe while residing beyond the reach of astronomical observations.

² This is already implicit in *Query 1* of Newton’s *Opticks*: “*Do not Bodies act upon Light at a distance, and by their action bend its Rays; and is not this action (ceteris paribus) strongest at the least distance?*”

The mathematician Friederich Bessel was perhaps the first to predict the existence of a specific undiscovered astronomical object, based only on its gravitational influence. In a letter published in 1844 [51], he argued that the observed proper motion of the stars Sirius and Procyon could only be explained by the presence of faint companion stars, influencing the observed stars through their gravitational pull:

If we were to regard Procyon and Sirius as double stars, their change of motion would not surprise us.

Bessel further argued in favor of the existence of many stars, possibly an infinite number of them, also anticipating the modern concept of the mass-to-light ratio:

But light is no real property of mass. The existence of numberless visible stars can prove nothing against the evidence of numberless invisible ones.

Only two years later, in 1846, the French astronomer Urbain Le Verrier and the English astronomer John Couch Adams, in order to explain some persistent anomalies in the motion of Uranus, proposed the existence of a new planet. Le Verrier's calculations were so precise that the German astronomer John Galle (assisted by Heinrich D'Arrest) identified the new planet at the Berlin observatory the same evening he received the letter from Le Verrier, within 1 degree of the predicted position.

Interestingly, it was Le Verrier himself who also later noticed the anomalous precession of the perihelion of Mercury, and proposed the existence of a perturbing planet to explain it. As it is well known, this "dark planet" – called *Vulcan* – was never observed, and the solution to this problem would have to await the advent of Einstein's theory of general relativity.

Beside dark stars and planets, astronomers in the 19th century also discussed dark matter in the form of dark clouds, or dark "nebulae". One of the earliest traces of this discussion can be found in a memoir written in 1877 by father Angelo Secchi, then Director of the Roman College Observatory, describing research on nebulae that had been carried out 20 years earlier [283]:

Among these studies there is the interesting probable discovery of dark masses scattered in space, whose existence was revealed thanks to the bright background on which they are projected. Until now they were classified as black cavities, but this explanation is highly improbable, especially after the discovery of the gaseous nature of the nebular masses.

Around the end of the 19th century, an interesting discussion began to take place within the astronomical community. As soon as astronomical photography was invented, scientists started to notice that stars were not distributed evenly on the sky. Dark regions were observed in dense stellar fields, and the question arose of whether they were dark because of a paucity of stars, or due to the presence of absorbing matter along the line-of-sight. The astronomer Arthur Ranyard, who was among the main proponents of the latter hypothesis, wrote in 1894 [252]:

The dark vacant areas or channels running north and south, in the neighborhood of [θ Ophiuchi] at the center seem to me to be undoubtedly dark structures, or absorbing masses in space, which cut out the light from the nebulous or stellar region behind them.

This debate went on for quite some time, and it sparked some interesting ideas. W. H. Wesley, who acted for 47 years as the assistant secretary of the Royal Astronomical Society, proposed a novel way to settle the question, involving a rudimentary simulation of the arrangement of stars in the Milky Way [329]:

It is better to solve the question experimentally. For this purpose [the author] repeated many times the experiment of sprinkling small splashes of Indian ink upon paper with a brush, revolving the paper between each sprinkling, so to avoid the chance of showing any artificial grouping in lines due to the direction in which the spots of ink were thrown from the hairs of the brush.

C. Dynamical Evidence

Lord Kelvin was among the first to attempt a dynamical estimate of the amount of dark matter in the Milky Way. His argument was simple yet powerful: if stars in the Milky Way can be described as a gas of particles, acting under the influence of gravity, then one can establish a relationship between the size of the system and the velocity dispersion of the stars [174]:

It is nevertheless probable that there may be as many as 10^9 stars [within a sphere of radius $3.09 \cdot 10^{16}$ kilometres] but many of them may be extinct and

dark, and nine-tenths of them though not all dark may be not bright enough to be seen by us at their actual distances. [...] Many of our stars, perhaps a great majority of them, may be dark bodies.

Kelvin also obtained an upper limit on the density of matter within such a volume, arguing that larger densities would be in conflict with the observed velocities of stars. Henri Poincaré was impressed by Lord Kelvin’s idea of applying the “theory of gases” to the stellar system of the Milky Way. In 1906 he explicitly mentioned “dark matter” (“matière obscure” in the original French), and argued that since the velocity dispersion predicted in Kelvin’s estimate is of the same order of magnitude as that observed, the amount of dark matter was likely to be less than or similar to that of visible matter [246] (for an English translation, see Ref. [247]. See also Ref. [248] for a more complete discussion):

There are the stars which we see because they shine; but might there not be obscure stars which circulate in the interstellar space and whose existence might long remain unknown? Very well then, that which Lord Kelvin’s method would give us would be the total number of stars including the dark ones; since his number is comparable to that which the telescope gives, then there is no dark matter, or at least not so much as there is of shining matter.

Along similar lines, in 1915, the Estonian astronomer Ernst Öpik built a model (published in Russian) of the motion of stars in the Galaxy, also concluding that the presence of large amounts of unseen matter was unlikely [103].

An important step forward in the understanding of the structure of the Milky Way was made by the Dutch astronomer Jacobus Kapteyn. In his most important publication, which appeared shortly before his death in 1922, Kapteyn attempted “a general theory of the distribution of masses, forces and velocities in the sidereal system” – that is, in the Milky Way.

Kapteyn was among the first to offer a quantitative model for the shape and size of the Galaxy, describing it as a flattened distribution of stars, rotating around an axis that points towards the Galactic Pole. He argued that the Sun was located close to the center of the Galaxy, and that the motion of stars could be described as that of a gas in a quiescent atmosphere. He then proceeded to establish a relationship between the motion of stars and their velocity dispersion, similar to what Öpik had done a few years earlier.

Kapteyn expressed the local density in terms of an effective stellar mass, by dividing the total gravitational mass by the number of observed stars – including faint ones, through an extrapolation of the luminosity function – and he explicitly addressed the possible existence of dark matter in the Galaxy:

We therefore have the means of estimating the mass of the dark matter in the universe. As matters stand at present, it appears at once that this mass cannot be excessive. If it were otherwise, the average mass as derived from binary stars would have been very much lower than what has been found for the effective mass.

In 1932, Kapteyn’s pupil, Jan Oort, published an analysis of the vertical kinematics of stars in the solar neighborhood [226]. In this work, Oort added to the list of estimates for the local dark matter density, including those by James Jeans (1922) [168] and by Bertil Lindblad (1926) [197]. In his analysis, Oort made a number of improvements on Kapteyn’s seminal work, relaxing for instance the assumption of the “isothermality” of the gas of stars.

Oort derived a most probable value for the total density of matter near the Sun of $0.092 M_{\odot}/\text{pc}^3$, corresponding to $6.3 \times 10^{-24} \text{ g/cm}^3$. He compared this number to the value obtained by Kapteyn, $0.099 M_{\odot}/\text{pc}^3$, and noticed that the agreement was “unexpectedly good”, given the differences in treatment and the data used. The numbers obtained by Jeans and Lindblad were each somewhat higher, $0.143 M_{\odot}/\text{pc}^3$ and $0.217 M_{\odot}/\text{pc}^3$, respectively.

In order to estimate the amount of dark matter, Oort then proceeded with an estimate for the contribution from stars to the local density, arguing that an extrapolation of the stellar mass function based on the observed stars seemed to be able to account for a substantial fraction of the inferred total density. It is interesting to recall the words used by Oort to illustrate the constraint on the amount of dark matter:

We may conclude that the total mass of nebulous or meteoric matter near the sun is less than $0.05 M_{\odot}\text{pc}^{-3}$, or $3 \cdot 10^{-24} \text{ g cm}^{-3}$; it is probably less than the total mass of visible stars, possibly much less.

We learn from this quote not only that the maximum allowed amount of dark matter was about half of the total local density, but also that astronomers thought that the dark matter was likely to consist of faint stars, that could be accounted for through a suitable extrapolation of the stellar mass function, along with “nebulous” and “meteoric” matter.

As we shall see in Chapter IV , the pioneering work of Kapteyn, Jeans, Lindblad, Öpik and Oort opened the path toward modern determinations of the local dark matter density, a subject that remains of importance today, especially for experiments that seek to detect dark matter particles through their scattering with nuclei.

III. GALAXY CLUSTERS

A. Zwicky and Smith

The Swiss-American astronomer Fritz Zwicky is arguably the most famous and widely cited pioneer in the field of dark matter. In 1933, he studied the redshifts of various galaxy clusters, as published by Edwin Hubble and Milton Humason in 1931 [162], and noticed a large scatter in the apparent velocities of eight galaxies within the Coma Cluster, with differences that exceeded 2000 km/s [346]. The fact that Coma exhibited a large velocity dispersion with respect to other clusters had already been noticed by Hubble and Humason, but Zwicky went a step further, applying the virial theorem to the cluster in order to estimate its mass.

This was not the first time that the virial theorem, borrowed from thermodynamics, was applied to astronomy; Poincare had done so more than 20 years earlier in his *Leçons sur les hypothèses cosmogoniques professées à la Sorbonne* [248]. But to the best of our knowledge, Zwicky was the first to use the virial theorem to determine the mass of a galaxy cluster.

Zwicky started by estimating the total mass of Coma to be the product of the number of observed galaxies, 800, and the average mass of a galaxy, which he took to be 10^9 solar masses, as suggested by Hubble. He then adopted an estimate for the physical size of the system, which he took to be around 10^6 light-years, in order to determine the potential energy of the system. From there, he calculated the average kinetic energy and finally a velocity dispersion. He found that 800 galaxies of 10^9 solar masses in a sphere of 10^6 light-years should exhibit a velocity dispersion of 80 km/s. In contrast, the observed average velocity dispersion along the line-of-sight was approximately 1000 km/s. From this comparison, he concluded:

If this would be confirmed, we would get the surprising result that dark matter is present in much greater amount than luminous matter.

This sentence is sometimes cited in the literature as the first usage of the phrase “dark matter”. It is not, as we have seen in the previous chapter, and it is not even the first time that Zwicky used it in a publication. He had, in fact, used the same phrase in a article published earlier the same year, pertaining to the sources of cosmic rays [347]:

According to the present estimates the average density of dark matter in our galaxy (ρ_g) and throughout the rest of the universe (ρ_u) are in the ratio $\rho_g/\rho_u > 100,000$.

Although he doesn't explicitly cite any article, it is obvious from this sentence that he was well aware of the work of Kapteyn, Oort and Jeans discussed in the previous chapter. His use of the term "dark matter" is, therefore, in continuity with the community of astronomers that had been studying the dynamics of stars in the local Milky Way.

In 1937, Zwicky published a new article – this time in English, in *the Astrophysical Journal* [348] – in which he refined and extended his analysis of the Coma Cluster. The purpose of this paper was to determine the mass of galaxies, and he proposed a variety of methods to attack this problem. In particular, he returned to the virial theorem approach that he had proposed in 1933, this time assuming that Coma contained 1000 galaxies within a radius of 2×10^6 light-years, and solving for the average galaxy's mass. From the observed velocity dispersion of 700 km/s, he obtained a conservative lower limit of $4.5 \times 10^{13} M_\odot$ on the mass of the cluster (to be conservative, he excluded a galaxy with a recession velocity of 5100 km/s as a possible outlier), corresponding to an average mass-per-galaxy of $4.5 \times 10^{10} M_\odot$. Assuming then an average absolute luminosity for cluster galaxies of 8.5×10^7 times that of the Sun, Zwicky showed that this led to a surprisingly high mass-to-light ratio of about 500.

Zwicky's work relied on Hubble's relationship between redshift and distance, and in the 1937 paper he used the results of Hubble and Humason [162], which pointed to a Hubble constant of $H_0 = 558$ km/s/Mpc, with an estimated uncertainty of 10-20%. If we rescale these results adopting the modern value of $H_0 = 67.27 \pm 0.66$ [245], we see that Zwicky over-estimated the mass-to-light ratio by a factor of $\sim 558/67.27 = 8.3$. Despite this substantial correction, Coma's velocity dispersion still implies a very high mass-to-light ratio and points to the existence of dark matter in some form.

What did Zwicky think that the dark matter in Coma and other galaxy clusters might be? An illuminating sentence in his 1937 paper provides a rather clear answer to this question:

[In order to derive the mass of galaxies from their luminosity] we must know how much dark matter is incorporated in nebulae in the form of cool and cold stars, macroscopic and microscopic solid bodies, and gases.

Meanwhile, another estimate for the mass of a cluster of galaxies had appeared in 1936,

this time from Sinclair Smith, who had studied the Virgo Cluster. Assuming that the outer galaxies were in circular motion around Virgo, Smith calculated a total mass for the cluster of $10^{14} M_{\odot}$. When divided by the number of observed galaxies, 500, he found an average mass per galaxy of $2 \times 10^{11} M_{\odot}$, which he pointed out was much higher than Hubble’s estimate of $10^9 M_{\odot}$.

Much like Zwicky, whose 1933 work he cites, Smith considers this high value for the mass-per-galaxy implied by his calculations to be a problem, in particular in light of its incompatibility with Hubble’s estimate. He also acknowledges, however, that both could be correct, and that:

the difference represents internebular material, either uniformly distributed or in the form of great clouds of low luminosity surrounding the [galaxies].

In his famous book *The Realm of Nebulae*, Hubble cites the work of Smith (and not that of Zwicky), and clearly states that he considers the discrepancy between the masses of galaxies inferred from the dynamics of clusters and those from the rotation of galaxies to be “real and important”. And although he argued that this problem might be solved, or at least diminished, by observing that the former were likely upper limits, while the latter lower limits, he acknowledged that this argument was not entirely satisfactory. A confusing situation had indeed arisen.

B. A Confusing Situation

There was no shortage of reasons for astronomers to be skeptical of the findings of Zwicky and Smith. The assumption that Virgo was a system in equilibrium, made by Smith, was questioned by Zwicky himself in his 1937 paper. In 1940, Erik Holmberg – who will appear again in this review as a pioneer of numerical simulations – described some of the concerns of the community regarding the work of Zwicky and Smith [158]:

It does not seem to be possible to accept the high velocities [in the Virgo and Coma cluster] as belonging to permanent cluster members, unless we suppose that a great amount of mass – the greater part of the total mass of the cluster – is contributed by dark material distributed among the cluster members – an unlikely assumption.

Objects	Distance (in kpc)	Luminosity (in sol. lum.)	Mass (in sol. mass)	Mass/Lum. f
Solar Neighborhood	—	—	—	4
Triangulum Nebula, M33	480	1.4×10^9	5×10^9	4
Large Magellanic Cloud	44	1.2×10^9	2×10^9	2
Andromeda Nebula	460	9×10^9	1.4×10^{11}	16
Globular Cluster, M92	11	1.7×10^5	$< 8 \times 10^5$	< 5
Elliptical Galaxy, NGC 3115	2100	9×10^8	9×10^{10}	100
Elliptical Galaxy, M32	460	1.1×10^8	2.5×10^{10}	200
Average S in Double Gal.	—	1.3×10^9	7×10^{10}	50
Average E in Double Gal.	—	8×10^8	2.6×10^{11}	300
Average in Coma Cluster	25000	5×10^8	4×10^{11}	800

FIG. 1. A snapshot of the dark matter problem in the 1950s: the distance, mass, luminosity, and mass-to-light ratio of several galaxies and clusters of galaxies, as compiled by M. Schwarzschild in 1954 [282].

Holmberg argued instead that these galaxies were probably “temporary” members of the cluster, i.e. galaxies on hyperbolic orbits that had fallen into the gravitational potential of the cluster, but were not bound to it. In 1954, Martin Schwarzschild [282] – son of the famous Karl Schwarzschild who had made important contributions to general relativity – attempted to get rid of “interlopers”, and inferred a smaller radial velocity dispersion of 630 km/s. By adopting an updated Hubble parameter, and an average luminosity-per-galaxy of $5 \times 10^8 L_{\odot}$, he obtained the “bewildering high” mass-to-light ratio of 800. The distance, mass, luminosity, and mass-to-light ratio of the galaxies and clusters of galaxies compiled by Schwarzschild are shown in Fig. 1.

By the late 1950s, a number of other articles had been published on the mass-to-light ratios of galaxy clusters. Victor Ambartsumian rejected the possibility that dark matter existed in clusters and argued instead that they are unstable and rapidly expanding systems, to which the virial theorem cannot be applied. It was soon realised, however, (e.g. Burbidge and Burbidge [67] and Limber [196]) that this interpretation was in tension with the estimated age of the galaxies (requiring clusters that were younger than the galaxies they contained), and with that of the Universe (the clusters should have evaporated long ago). In August of 1961, a conference on the instability of systems of galaxies was held in Santa Barbara, and included as participants some of the most important astrophysicists active in that field of research. Jerzy Neyman, Thornton Page and Elizabeth Scott summarized the

discussions that took place around the mass discrepancy in galaxy clusters as follows:

Several possible explanations of this mass discrepancy were discussed at the Conference [..]. Many of those present consider that it might be real and due to invisible inter-galactic material in the clusters, totalling 90 to 99% of their mass. If these possibilities are excluded, however, the discrepancy in mass indicates positive total energy and instability of the system involved.

The overall situation was that of a community that was struggling to find a unified solution to a variety of problems. The dark matter hypothesis was not commonly accepted, nor was it disregarded. Instead, there was a consensus that more information would be needed in order to understand these systems.

In addition to the question of whether the dynamics of galaxy clusters required the presence of dark matter, astronomers around this time began to be increasingly willing to contemplate what this dark matter might be made of. Herbert Rood [263] (later confirmed by Simon White [332]) studied the relaxation process of galaxy clusters and argued that the mass responsible for their high mass-to-light ratios must to be found within the intergalactic space, and not in the galaxies themselves. Arno Penzias searched for free hydrogen in the Pegasus I cluster and set an upper limit of a tenth of its virial mass [241]. Neville Woolf suggested in 1967 that the gas could be ionised, and used radio, visible and X-ray observations to set limits on it [341]. Turnrose and Rood discussed the problems of this hypothesis in Ref. [314], and in 1971 Meekins et al. [210] obtained observational evidence for X-ray emission that limited the amount of hot intracluster gas to be less than 2% of that required for gravitational binding.

With gas ruled out as an explanation for the “missing mass” in galaxy clusters, scientists began to explore more or less exotic possibilities, including massive collapsed objects [317], HI snowballs [238], and M8 dwarf stars [307]. As we will discuss in Chapter V, these possibilities – and others like them – were eventually ruled out by measurements of the primordial light element abundances, which instead favor a non-baryonic nature for the dark matter.

IV. GALACTIC ROTATION CURVES

A. The Beginnings

The rotation curves of galaxies – i.e. the circular velocity profile of the stars and gas in a galaxy, as a function of their distance from the galactic center – played a particularly important role in the discovery of dark matter. Under some reasonable simplifying assumptions, it is possible to infer the mass distribution of galaxies from their rotation curves. Historically, it was the observation of approximately “flat” rotation curves at very large galactocentric distances that did the most to convince the scientific community that large amounts of dark matter is present in the outer regions of galaxies.

In 1914, ten years before Hubble convincingly demonstrated that Andromeda (M31) was a galaxy and located outside of the Milky Way, Max Wolf [340] and Vesto Slipher [297] noticed that the spectral lines from this system were inclined when the slit of the spectrogram was aligned with the galaxy’s major axis and straight when it was aligned with the minor axis, allowing them to conclude that Andromeda rotates. Based on 79 hours of observation in 1917 with the Mount Wilson Observatory’s 60-inch reflector, Francis Pease measured the rotation of the central region of Andromeda out to an angular radius of 2.5 arcminutes, finding that it rotates with an approximately constant angular velocity. Several authors used Andromeda’s observed rotational velocity to calculate its mass and discuss its mass-to-light ratio in comparison with the measured value for the solar neighborhood (see Chapter II), finding values that were in reasonable agreement, e.g. Hubble (1926) [163], Oort (1932) [226].

In a paper published in 1930 [203], Knut Lundmark made estimates for the mass-to-light ratios of five galaxies based on a comparison of their absolute luminosity – as estimated using novae as distance indicators – and their mass as inferred from spectroscopic observations. These mass-to-light ratios varied, quite unrealistically, from 100 for M81 to 6 for M33 – much larger than those found for the solar neighborhood (Lundmark also made early estimates for the mass of the Milky Way [202]). This demonstrates that astronomers at the time were open to the possibility that large amounts of dark matter might be present in astrophysical systems, in the form of “extinguished stars, dark clouds, meteors, comets, and so on”, as Lundmark writes in 1930³. Holmberg argued in 1937 that the large spread in mass-to-light

³ Translation from the German by Lars Bergström

ratios found by Lundmark was a consequence of the absorption of light “*produced by the dark matter*”, and that once this was correctly taken into account, all of the galaxies studied by Lundmark, including the Milky Way, would have mass-to-light ratios between 6 and 7 [157].

Fritz Zwicky, in his famous 1937 article on galaxy clusters, discussed the possibility of using the rotation curves of galaxies to infer their mass distribution, concluding that:

It is not possible to derive the masses of [galaxies] from observed rotations, without the use of additional information.

Beside the lack of information on the ellipticity of orbits, one of Zwicky’s main concerns was the possible internal “viscosity” resulting from the mutual interactions of stars. Only four years later, Chandrasekhar would demonstrate in his classic paper, “The Time of Relaxation of Stellar Systems”, that these interactions are completely negligible, allowing one to reliably describe galaxies as systems of non-interacting stars.

Meanwhile, in his 1939 PhD dissertation, Horace Babcock presented the rotation curve of M31 out to 100 arcminutes (i.e. about 20 kpc) away from its center [31]. Interestingly, he found very high values for the circular velocity at large radii – so high, in fact, that they are at odds with modern measurements. Approximating M31 as a sphere surrounded by a flattened ellipsoid, he calculated the mass distribution of the galaxy, recognizing that the observed rising rotation curve at large radii implied the existence of large amounts of mass in the outer parts of the galaxy. But when interpreting this result, he conservatively argued that:

the calculated ratio of mass to luminosity in proceeding outward from the nucleus suggests that absorption plays a very important role in the outer portion of the spiral, or, perhaps, that new dynamical considerations are required, which will permit of a smaller mass in the outer parts.

More than a decade later, observations made by Nicholas Mayall in 1951 at Mount Wilson [282] were used by Martin Schwarzschild to further study the dynamics of M31. In doing so, Schwarzschild showed that a model with a constant mass-to-light ratio was able to explain the rotational velocities measured by Mayall out to 115 arcminutes.

The German invasion of Poland in 1939 marked the official start of World War II. Hostilities brought death and destruction, but also unexpected benefits for science, as after the war

ended military radars began to be used for radio astronomical observations. The Netherlands was particularly active in this field, under the push of the visionary astronomer Jan Oort, who was not only a great scientist, but also a great organizer. A chain of so-called Würzburg antennas – 7.5 meter parabolic radars used at 54 cm wavelengths for aircraft tracking – had been left behind in the Netherlands by occupying German forces at the end of the war, and since the reflective surface and tracking precision were also suitable for shorter wavelengths, and in particular for the 21 cm line predicted by Oort’s student Hendrik van de Hulst, one was mounted in Kootwijk for the purpose of radio astronomy [318].

When Harold Ewen and Edward Purcell, from Harvard, detected the 21 cm line in 1951, van de Hulst was visiting Harvard, and so was F. J. Kerr from the Radiophysics Laboratory in Sydney. The Dutch and Australian groups were soon able to confirm the detection: the reports of the American and Dutch groups appeared in the same issue of *Nature*, together with a confirming telegram from the Australian group. This success provided an important boost to the young field of radio astronomy, and had a dramatic impact on the history of astrophysics and cosmology.

Back in the Netherlands, the construction of a new 25 meter radio telescope was completed in Dwingeloo, in 1955. Only two years later, van de Hulst, Jean Jacques Raimond, and Hugo van Woerden published the first radio rotation curve of M31, extending observations to 2 degrees away from its center [316]. Although the data seemed at first to be at odds with the rotation curve calculated by Schwarzschild, Maartin Schmidt argued in a paper accompanying the publication of van de Hulst et al. that a constant mass-to-light ratio provided a satisfactory explanation of the data, although also noting that “nothing as yet can be stated about the ratio in the innermost and outermost parts” of M31 [280].

In 1959, Franz Kahn and Lodewijk Woltjer proposed an ingenious method to determine the combined mass of M31 and the Milky Way. Since 21 cm observations of M31 indicated that it was approaching the Milky Way at a speed of 125 km/s, they derived a lower bound on the reduced mass of the M31-Milky Way system, assuming that the two galaxies are part of a bound system and that the orbital period is smaller than the age of the Universe. That lower bound was, however, six times larger than the currently accepted value of the reduced mass of the system [171]. The authors argued at the time that this provided evidence for intergalactic material in the form of gas stabilising the local group. In retrospect, this simple argument is one of the earliest clear indications of dark matter halos around galaxies.

In his detailed historical account [277], Robert Sanders argues that despite these developments there was no sense of crisis in the field of astrophysics at the end of the 1950s, or at least that there was no consensus that the observed rotation curves were in conflict with the current understanding of galaxies. A decade later, things began to dramatically change.

B. The 1970s Revolution

In the 1960s, Kent Ford developed an image tube spectrograph that Vera Rubin and he used to perform spectroscopic observations of the Andromeda Galaxy. The observations of the M31 rotation curve Rubin and Ford published in 1970 [267] represented a step forward in terms of quality. Their optical data extended out to 110 arcminutes away from the galaxy's center, and were compatible with the radio measurements obtained previously by Morton Roberts in 1966 [257].

It was also in 1970 that the first explicit statements began to appear arguing that additional mass was needed in the outer parts of some galaxies, based on comparisons of the rotation curves predicted from photometry and those measured from 21 cm observations. In the appendix of his seminal 1970 paper [126], Ken Freeman compared the radius at which the rotation curve was predicted to peak, under the assumption of an exponential disk with a scale length fit to photometric observations, to the observed 21 cm rotation curve. This combination of theoretical modelling and radio observations extending beyond the optical disk allowed Freeman to reach a striking conclusion. He found that for M33 (based on data summarised in Ref. [62]) and NGC 300 (based on data from Ref. [287]), the observed rotation curves peaked at larger radii than predicted, and – prompted by discussions with Roberts⁴ – concluded that:

if [the data] are correct, then there must be in these galaxies additional matter which is undetected, either optically or at 21 cm. Its mass must be at least as large as the mass of the detected galaxy, and its distribution must be quite different from the exponential distribution which holds for the optical galaxy.

This is perhaps the first convincing (or at least convinced) claim of a mass discrepancy in galaxies. D. Rogstad and G. Shostak performed a similar analysis in 1972 [262], by analyzing

⁴ K. Freeman, private communication.

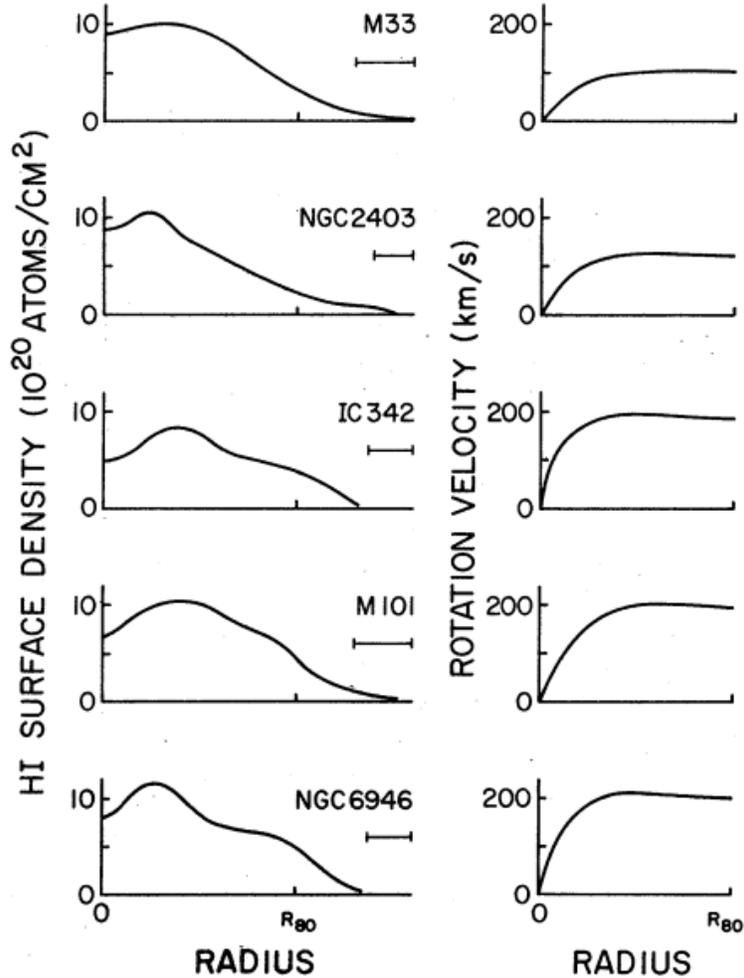


FIG. 2. Flat rotation curves began to emerge clearly from 21 cm observations in the early 1970s. Here we show the hydrogen surface density profile (left) and the rotation curves (right) of five galaxies as obtained by Rogstad and Shostak in 1972 [262]. The bars under the galaxy names indicate the average radial beam diameter, i.e. the effective spatial resolution. R_{80} is the radius containing 80% of the observed HI.

the rotation curves of five galaxies – M33, NGC 2403, IC 342, M101 and NGC 6946 – they had themselves obtained using the radio telescope at the Owens Valley Radio Observatory. They found that these rotation curves remained flat out to the largest radii observed (see Fig. 2) and, following the method of Freeman, they derived mass-to-light ratios as high as 20 at large radii. As explicitly said in their paper, they:

confirm[ed] the requirement of low-luminosity material in the outer regions

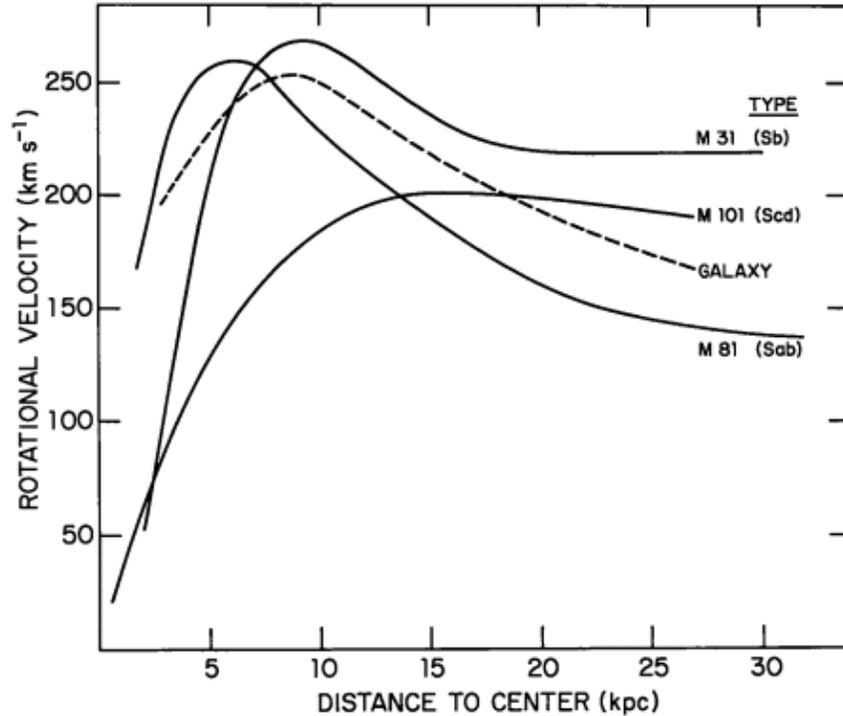


FIG. 3. The rotation curves for the galaxies M31, M101, and M81 (solid lines) obtained by Roberts and Rots in 1973. The rotation curve of the Milky Way Galaxy was included by the authors for comparison. From Ref. [260].

of these galaxies.

Morton Roberts was among the first to recognize the implications of the observed flatness of galactic rotation curves. Together with R. Whitehurst, he published in 1972 a rotation curve of M31 that extended to 120 arcminutes from its center [335]. In 1973, together with Arnold Rots, he extended the analysis to M81 and M101, and argued that these spiral galaxies each exhibited flat rotation curves in their outer parts [260] (see Fig. 3). The authors' interpretation of these data was unambiguous:

The three galaxies rotation curves decline slowly, if at all, at large radii, implying a significant mass density at these large distances. It is unreasonable to expect the last measured point to refer to the 'edge' of the galaxy, and we must conclude that spiral galaxies must be larger than indicated by the usual photometric measurements [...]. The present data also require that the mass to luminosity ratio vary with radius increasing in distance from the center.

In the Proceedings of the IAU Symposium No. 69, held in Besançon, France in September of 1974, Roberts reviewed the status of galactic rotation curves [259], highlighting the importance of radio observations, which extended well beyond the optical radius of the galaxies. When discussing the implications of the high mass-to-light ratios implied by these observations, he argued that the excess mass might take the form of intermediate and late dwarf M stars. He further tried to reassure his colleagues by arguing that the required radius-dependent luminosity function need not be alarming, since there was evidence of a dependence on the height above the Galactic Plane that exhibited a similar trend.

As we will discuss in Chapter VIII, two influential papers in 1974 brought together the observed mass discrepancies observed in clusters and in galaxies [109, 229]. Both of these papers clearly state in their first paragraph that the mass of galaxies had been until then underestimated by about a factor of ten. In support of this, Jerry Ostriker, Jim Peebles, and Amos Yahil [229] cited Roberts and Rots [260] and Rogstad and Shostak [262] for the observed flat rotation curves. For the same purpose, Jaan Einasto, Ants Kaasik and Enn Saar [109] cited a review written in 1975 by Roberts for a book edited by A. and M. Sandage together with J. Cristian [272]. In a separate paper that appeared in the same year, focusing on the “morphological evidence” of missing mass around galaxies, Einasto and collaborators cited the 1973 paper of Roberts and Rots [260]. Interestingly, Einasto and collaborators excluded the possibility that this missing mass was in the form of stars, and argued that the most likely explanation was the presence of large amounts of gas in the outer parts of galaxies, which they referred to as “coronas” [110].

By 1974, the flat rotation curves obtained by radio astronomers had done much to establish the existence of large amounts of mass in the outer parts of galaxies – at least to the eyes of the influential authors of Refs. [109, 229]. Portions of the astronomical community, however, were still not convinced of this conclusion [266]. In the late 1970s, this evidence was strengthened and corroborated by a series of new studies. In 1977, Nathan Krumm and Edwin Salpeter [191] observed six spiral galaxies with the Arecibo Observatory, and showed that they each exhibited a flat rotation curve out to radii larger than their optical extent, but these data turned out to be unreliable due to beam-smearing (see the discussion at the end of Ref. [270]).

In 1978, Albert Bosma published the results of his PhD thesis [60], including the radio observation of the velocity fields and corresponding rotation curves of 25 galaxies. This

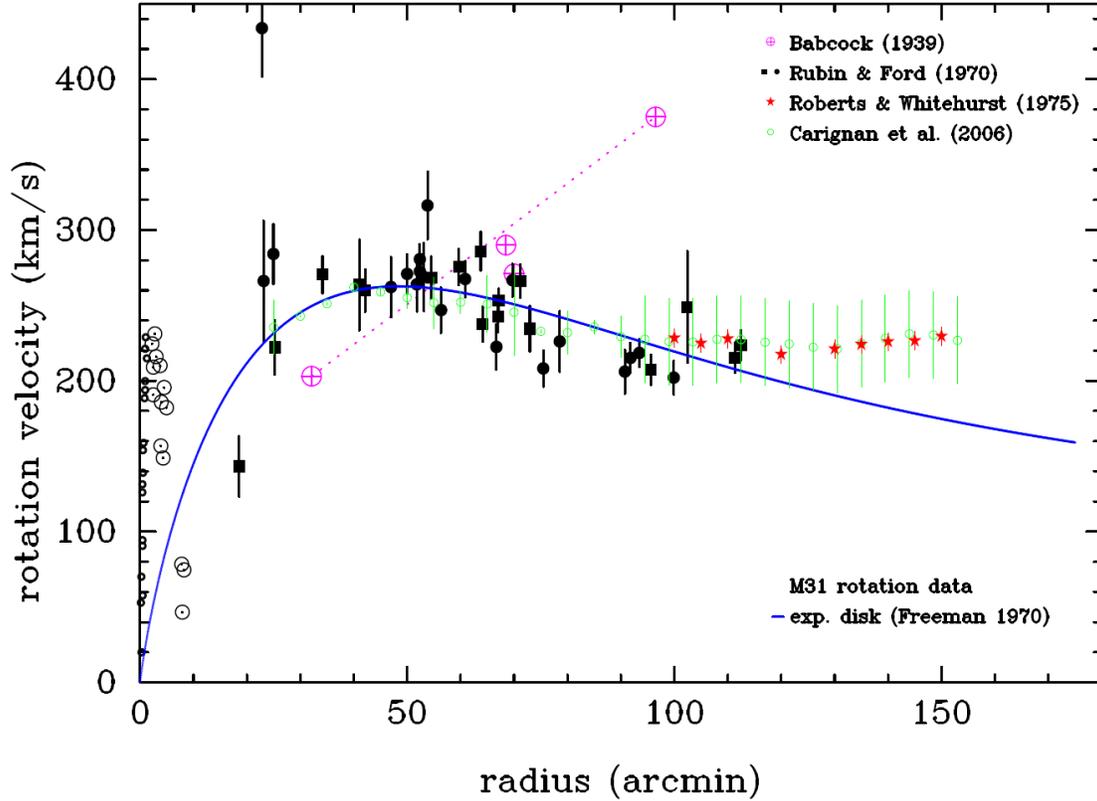


FIG. 4. Rotation curve data for M31. The purple points are emission line data in the outer parts from Babcock 1939 [31]. The black points are from Rubin and Ford 1970 [267] (squares for the SW data, filled circles for the NE data, and open circles for the data in the inner parts – the presence of non-circular motions in the inner parts makes the modelling of those data uncertain). The red points are the 21-cm HI line data from Roberts and Whitehurst 1975 [261]. The green points are 21-cm HI line data from Carignan et al. [73]. The black solid line corresponds to the rotation curve of an exponential disc with a scalelength according to the value given in Freeman 1970 [126], suitably scaled in velocity. 21-cm data demonstrate clearly the mass discrepancy in the outer parts. Figure courtesy of Albert Bosma.

work convincingly proved that most of these objects had flat rotation curves out to the largest observed radius, which again exceeded the optical size of the galaxies, therefore demonstrating that their mass continued to grow beyond the region occupied by the stars and gas (see also Fig. 5).

A few months later, Rubin, Ford and Norbert Thonnard published optical rotation curves

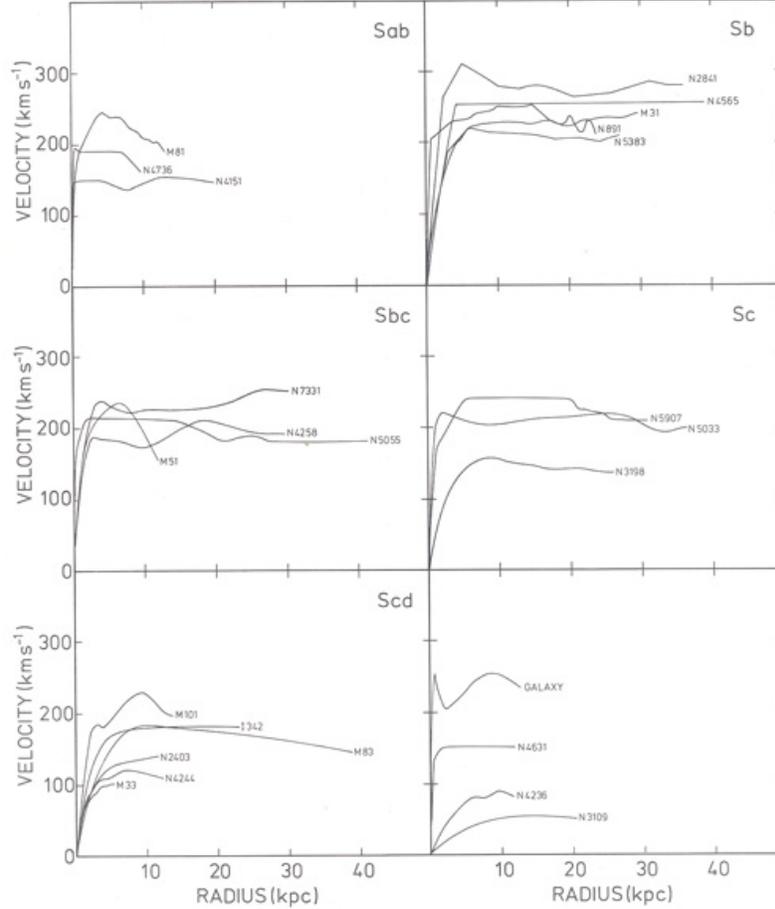


FIG. 5. The rotation curves of the 25 galaxies published by Albert Bosma in 1978 [60].

for ten high-luminosity spiral galaxies and found that they were flat out to the outermost measured radius [268]. This work has become one of the most well-known and widely cited in the literature, despite the fact that the optical measurements did not extend to radii as large as those probed by radio observations, thus leaving open the possibility that galaxies may not have dark matter halos, as pointed out, for example, by Agris J. Kalnajs in 1983 (see the discussion at the end of Ref. [150]) and by Stephen Kent in 1986 [175]. Rubin, Ford and Thonnard themselves acknowledged the credit that was due to the preceding analyses:

Roberts and his collaborators deserve credit for first calling attention to flat rotation curves. [...] These results take on added importance in conjunction with the suggestion of Einasto, Kaasik, and Saar (1974) and Ostriker, Peebles and Yahil (1974) that galaxies contain massive halos extending to large r .

A lucid and timely review of the status of galaxy masses and mass-to-light-ratios appeared in 1979, authored by Sandra Faber and John Gallagher [115]. We refer the reader to this excellent article for an overview of the various ideas that had been put forward in an effort to understand the complex and diverse observational data that was available at the time. The abstract of that article provides a clear indication of its contents:

The current status of the ‘missing mass’ problem is reviewed on the basis of standardized mass-to-light (M/L) ratios of galaxies. The stellar mass density in the immediate vicinity of the sun is examined, along with the mass of the Milky Way and the M/L ratios of spiral galaxies, E and S0 galaxies, and binary galaxies. The dynamics of small groups of galaxies is investigated, and mass derivations for cluster galaxies are discussed. It is concluded that the case for invisible mass in the universe is very strong and becoming stronger.

C. Local Measurements

We conclude this chapter with a brief overview of the efforts to determine the local dark matter density, i.e. the density of dark matter in the solar neighborhood. This quantity was historically important, as it provided the first – albeit rather weak – dynamical evidence for matter in the local Universe beyond visible stars. It is also important today, as the prospects for detecting dark matter particles in underground and astrophysical experiments strongly depend on this quantity.

As we have seen in the previous chapter, Kapteyn, Lindblad, Jeans and Oort had studied the dynamics of nearby stars, and compared the inferred gravitational mass with that of the visible stellar density. After decades of steady improvements (Oort 1932 [226], Hill (1960), Oort 1960 [227], Bahcall 1984 [32, 33]), Konrad Kuijken and Gerry Gilmore published a series of papers based on a refined method and a volume complete sample of K-dwarf data, to derive a much more precise value of the local density [192]. The advent of the Hipparcos, SDSS, and RAVE surveys has more recently triggered many new analyses. We refer the reader to the excellent review by Justin Read [253] for further details and references.

Alternatively, the local dark matter density can be constrained using measurements of the Milky Way’s rotation curve (e.g. Fich et al. 1989 [122], Merrifield 1992 [212], Dehnen

and Binney 1998 [89], Sofue et al. 2009 [298], Weber and de Boer 2010 [326], Catena and Ullio 2010 [76], Salucci et al. 2010 [271], Iocco et al. 2011 [166], Pato et al. 2015 [233]). Although rather precise determinations can be made using this approach, the results strongly depend on the assumptions one makes about the shape of the halo. Upcoming astronomical surveys – and in particular the Gaia satellite – are expected to lead to significant improvements in the reconstruction of the local density (Perryman et al. 2001 [243], Wilkinson et al. 2005 [338], Read 2014 [253], Silverwood et al. 2015 [292]).

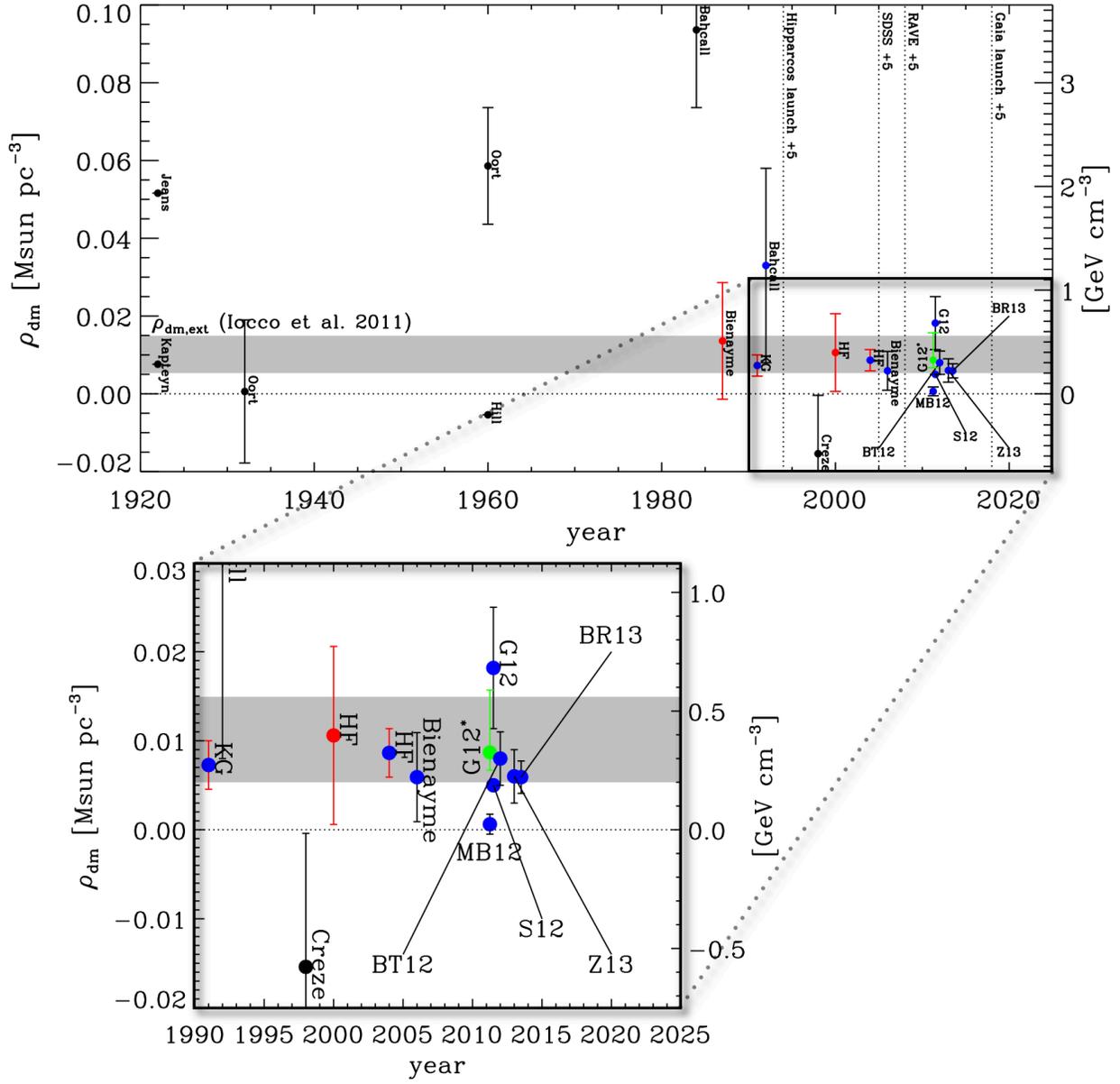


FIG. 6. Timeline of local dark matter density measurements. See Read (2014) for further details and references [253].

V. DARK MATTER PARTICLES

Over the past few decades, the very meaning of the phrase “dark matter” has evolved considerably. Today, this phrase is most frequently used as the name – a proper noun – of whatever particle species accounts of the bulk of our Universe’s matter density. When a modern paper discusses the distribution of dark matter, or the impact of dark matter on structure formation, or the prospects for detecting dark matter with a gamma-ray telescope, the reader does not have to ask themselves whether the authors might have in mind white dwarfs, neutron stars, or cold clouds of gas – they don’t. This is in stark contrast to the earlier usage of the phrase, in which the word “dark” was a mere adjective, and “dark matter” included all varieties of astrophysical material that happened to be too faint to be detected with available telescopes.

This linguistic transition reflects a larger change that has taken place over the past several decades within the astrophysics and particle physics communities. And although this transformation was driven and initiated by new scientific results and understanding, it also reflects a sociological change in the underlying scientific culture. Half a century ago, cosmology was something of a fringe-science, perceived by many astronomers and particle physicists alike to have little predictive power or testability. This can be easy to forget from our modern vantage point in the age of precision cosmology. Furthermore, prior to the last few decades, particle physicists did not often study or pursue research in astrophysics, and most astrophysicists learned and knew little about particle physics. As a result, these scientists did not frequently contribute to each other’s fields of research. When Fermilab founded its theoretical astrophysics group in 1983, for example, the decision to do so was seen by many as a radical departure from the lab’s particle physics mission. From the perspective of many particle physicists in the early 1980s, it was not obvious what astrophysics had to do with the questions being asked by particle physics. This view is shared by few today. As an illustration, we need only to note that the report of the US Particle Physics Project Prioritization Panel (P5) describes the “Cosmic Frontier”, along with the Energy and Intensity Frontiers, as co-equal areas of inquiry within the larger field of particle physics.

From our contemporary perspective, it can be easy to imagine that Fritz Zwicky, Vera Rubin, and the other early dark matter pioneers had halos of weakly interacting particles in mind when they discussed dark matter. In reality, however, they did not. But over time,

an increasing number of particle physicists became interested in cosmology, and eventually in the problem of dark matter. By the late 1980s, the hypothesis that the missing mass consists of one or more yet-unknown subatomic particle species had gained enough support to become established as the leading paradigm for dark matter. As alternatives were ruled out one-by-one (see Chapters VI and VII), this view came to be held almost universally among both particle physicists and astrophysicists, as well as among their new and now increasingly common hybrids – the particle-astrophysicists.

A. Neutrinos

When one considers the dark matter problem from the perspective of the standard model of particle physics, the three neutrinos clearly stand out. Unlike all other known particle species, the neutrinos are stable – or at least very long lived – and do not experience electromagnetic or strong interactions. These are essential characteristics for almost any viable dark matter candidate. And although we know today that dark matter in the form of standard model neutrinos would be unable to account for our Universe’s observed large scale structure, these particles provided an important template for the class of hypothetical species that would later be known as WIMPs – weakly interacting massive particles. In this way, standard model neutrinos served as an important gateway particle, leading astrophysicists and particle physicists alike to begin their experimentation with a variety of other, more viable, particle dark matter candidates. And although the first scientists to consider the role of neutrinos in cosmology did not have the dark matter problem in mind – many being unaware that there was any such problem to solve – their work helped to establish the foundations that the field of particle dark matter would later be built upon.

The earliest discussion of the role of neutrinos in cosmology appeared in a 1966 paper by S. S. Gershtein and Ya. B. Zeldovich [132]. To many scientists working in fields of cosmology and particle-astrophysics, it will be no surprise to see Zeldovich’s name attributed to this pioneering work. Yakov Borisovich Zeldovich was an utterly prolific and versatile physicist, making major contributions to the fields of material science, nuclear physics (including the Soviet weapons program), particle physics, relativity, astrophysics, and cosmology. In terms of research at the interface between particle physics and cosmology, it can sometimes seem like Zeldovich did almost everything first.

In the early 1960s, Zeldovich was one of only a handful of particle physicists who were also thinking about problems in cosmology. During this period, he made early contributions to black hole thermodynamics, recognized that accretion disks around black holes could power quasars, discussed the possibility of primordial black holes, and studied the problem of how the large scale structure of the Universe formed. He is probably most famous for his paper with Rashid Sunyaev, which predicted that the cosmic microwave background would be distorted by its inverse Compton scattering with high-energy electrons in galaxy clusters [306]. This so-called “Sunyaev-Zeldovich effect” was observed for the first time in 1983, and continues to be of considerable importance in modern cosmology. So sweeping were Zeldovich’s contributions to cosmology, that upon being introduced, Stephen Hawking is said to have expressed to him, “Before I met you, I believed you to be a collective author, like Bourbaki⁵.”

In their 1966 paper, Zeldovich and Gershtein considered the production of neutrinos under the conditions that existed shortly after the Big Bang. Making use of the knowledge of the newly discovered three degree cosmic microwave background [242], they predicted how many electron and muon neutrinos would have existed in thermal equilibrium in the early Universe, and at what temperature those particles would have ceased to efficiently self-annihilate, leading to a population of neutrinos that survived as a thermal relic⁶. Considering how the density of those neutrinos would impact the expansion history of the Universe, and comparing that to existing estimates of the Hubble constant and the age of the oldest observed stars, Zeldovich and Gershtein concluded that the masses of the electron and muon neutrinos must each be less than approximately 400 eV; if they had been heavier, the neutrinos would have unacceptably slowed, or even reversed, the rate of cosmic expansion. For the muon neutrino, this result represented an improvement of three orders of magnitude over the previously existing upper limits.

Looking back at this result from a modern perspective, we see the seeds of particle dark matter, and even WIMPs. In particular, Zeldovich and Gershtein showed that a neutrino species with a mass of a few tens of eV or greater would come to dominate the energy density of the Universe. But there was no mention in their paper of any missing mass that these neutrinos might be able to account for; they only required that the density of the relic

⁵ Nicolas Bourbaki was a collective pseudonym adopted by a group of 20th-century mathematicians.

⁶ As there existed no evidence for a third generation at the time, Gershtein and Zeldovich did not consider the tau neutrino.

neutrinos not be so high as to cause the expansion rate of the Universe to slow down faster than observed.

This is essentially the same perspective that was expressed years later, when papers on this topic began to appear in the West. The first of these papers appeared in 1972, in which Ram Cowsik and J. McClelland used an approach similar to Zeldovich and Gershtein’s to derive an upper limit of 8 eV on the mass of a single (Dirac) neutrino species [82] (see also Ref. [207]). If it had not been for this paper, one might be tempted to conclude that interest in this topic would have developed much sooner among American and Western European scientists if word of Zeldovich and Gershtein’s work had reached them earlier. But the paper by Cowsik and McClelland (who were both at the University of California, Berkeley, at the time) seems to disprove this counterfactual. Even after the appearance of this paper, there was no discernible rush to further explore the role of neutrinos (or other thermal relics) in the early Universe.

Eventually, however, interest in neutrino cosmology did begin to pick up. In 1976, A. S. Szalay and G. Marx published a paper that not only derived an upper limit on neutrino masses from cosmology, but also discussed the possibility that ~ 10 eV neutrinos might make up the “missing mass” in the Universe, and in galaxy clusters. Then, a few years later, a sequence of related papers appeared in rapid succession. In a paper received in April of 1977, Piet Hut presented a limit on the neutrino mass from cosmological considerations, ruling out masses in the range of 120 eV to 3 GeV [164]. In contrast to the authors of the preceding papers, Hut pointed out that quite heavy neutrinos ($m_\nu > 3$ GeV) would be produced in the Big Bang with an abundance that would not overclose the Universe. Only about a week later, Ben Lee and Steven Weinberg submitted a paper that included a very similar lower bound ($m_\nu > 2$ GeV) [194]. In the same month, a paper by K. Sato and H. Kobayashi [278] presented similar conclusions, and another by Duane Dicus, Edward “Rocky” Kolb and Vigdor Teplitz pointed out that such bounds could be evaded if neutrinos were unstable [91]. A month later, a new paper by Zeldovich (with M. I. Vysotskii and A. D. Dolgov) appeared, updating their own cosmological constraints on neutrino mass [322].

Despite the very interesting and important results of these papers, it is notable that most of them did not attempt to address, or even acknowledge, the possibility that neutrinos could account for the missing mass observed by astronomers on galactic and cluster scales. Exceptions to this include the 1976 paper of Szalay and Marx, and the 1977 paper of Lee

and Weinberg, whose final sentence reads as follows [194]:

Of course, if a stable heavy neutral lepton were discovered with a mass of order 1-15 GeV, the gravitational field of these heavy neutrinos would provide a plausible mechanism for closing the universe.

While this is still a long way from acknowledging the dynamical evidence for dark matter, it was an indication that physicists were beginning to realize that weakly interacting particles could be very abundant in our Universe, and may have had an observable impact on its evolution. The connection between particle physics and the missing mass problem did gradually become more appreciated over the years to come. In 1978, for example, a paper by James Gunn, Ben Lee, Ian Lerche, David Schramm, and Gary Steigman included the following statement in their abstract [144]:

... such a lepton is an excellent candidate for the material in galactic halos and for the mass required to bind the great clusters of galaxies.

By the end of the decade, a number of scientists – including Zeldovich and his Moscow group [99, 100, 344] – had begun to argue in favor of neutrinos as dark matter. Interest in this possibility grew considerably in and after 1980, when a group studying tritium beta decay reported that they had measured the mass of the electron anti-neutrino (and presumably also the electron neutrino) to be approximately 30 eV [201]. With a mass of this value, neutrinos would be expected to have played a very significant role in cosmology. And although this “discovery” was eventually refuted, it motivated many particle physicists to further investigate the cosmological implications of their research, and encouraged many astrophysicists to consider the possibility that the dark matter halos surrounding galaxies and galaxy clusters might not be made up of faint stars or other astrophysical objects, but instead might consist of a gas of non-baryonic particles.

By the middle of the 1980s, a new tool had come into use that would put neutrino dark matter to the test. This tool — numerical simulations — could be used to predict how large numbers of dark matter particles would evolve under the force of gravity in an expanding Universe, and thus was able to assess the cosmological role and impact of dark matter particles on the formation of large scale structure. Importantly, such tests could be used to discriminate between different dark matter candidates, at least in some cases.

The primary characteristic of a given particle dark matter candidate that can be probed by numerical simulations is whether it was relativistic (hot) or non-relativistic (cold) during the epoch of structure formation⁷. Standard model neutrinos, being very light thermal relics, are predicted to emerge from the early Universe with a highly relativistic velocity distribution, and thus represent an example of hot dark matter [240, 281]. Simulations have shown that hot dark matter particles would tend to collapse and form very large structures first, and only later go on to form smaller (i.e. galaxy-sized) halos through the fragmentation of larger halos. In contrast to this “top-down” sequence of structure formation, cold dark matter particles form structures through a “bottom-up” sequence, beginning with the smallest halos, which go on to form larger halos through a succession of mergers.

From these early simulations, it quickly became clear that hot and cold dark matter lead to very different patterns of large scale structure. By comparing the results of these simulations with those of galaxy surveys (in particular the CfA survey, which was the first extensive 3D survey of galaxies in the local Universe [85]), it was determined that standard model neutrinos – or any other examples of hot dark matter – could not account for most of the dark matter in the Universe [333]. In their 1983 paper, Simon White, Carlos Frenk and Marc Davis make the following statement about a neutrino-dominated Universe [333]:

We find [the coherence length] to be too large to be consistent with the observed clustering scale of galaxies... The conventional neutrino-dominated picture appears to be ruled out.

We will discuss numerical simulations, and their role in the history of dark matter, in greater detail in Sec. VIII C.

As it became accepted that standard model neutrinos could not make up most of the Universe’s dark matter⁸, it also became clear that there must exist at least one currently unknown particle species that makes up the missing mass. But although standard model neutrinos were far too light and hot to make up the dark matter, this new information did not preclude the possibility that other types of neutrino-like particles might make up this elusive substance (see, for example, Ref. [224]). In 1993, Scott Dodelson and Lawrence

⁷ The terms “hot” and “cold” dark matter were coined in 1983 by Joel Primack and Dick Bond (J. Primack, private communication).

⁸ A possible exception being the tau neutrino, whose mass would not be measured for another two decades, and thus could not at the time be ruled out as a cold dark matter candidate.

Widrow proposed a simple scenario in which an additional neutrino species, without the electroweak interactions experienced by standard model neutrinos, could be produced in the early Universe and realistically make up the dark matter [98]. Other than through gravity, the particles envisioned by Dodelson and Widrow interact only through a small degree of mixing with the standard model neutrinos. With such feeble interactions, such particles would have never been in thermal equilibrium in the early Universe, but instead would have been produced through the oscillations of the other neutrino species. Depending on their mass, such sterile neutrinos could be produced with a wide range of temperatures, and thus could constitute either a warm ($m_{\nu_s} \sim \text{keV}$) or a cold ($m_{\nu_s} \gg \text{keV}$) candidate for dark matter.

B. Supersymmetry

Among the particle species contained within the standard model, neutrinos are the only examples that are stable, electrically neutral, and not strongly interacting, and therefore are the only known particles that were viewed as potentially viable candidates for dark matter. Physicists’ imagination, however, would not remain confined to the standard model for long, but instead would turn to the contemplation of many speculative and yet undiscovered candidates for the dark matter of our Universe. In particular, beginning in the early 1970s, many physicists began to consider the possibility that nature may contain a space-time symmetry relating fermions to bosons, dubbed “supersymmetry” [133, 136, 186, 330]. Supersymmetry requires that for every fermion, a boson must exist with the same quantum numbers, and vice versa. Supersymmetry, therefore, predicts the existence of several new electrically neutral and non-strongly interacting particles, including the superpartners of the neutrinos, photon, Z boson, Higgs boson, and graviton. If any of these superpartners were stable, they could be cosmologically abundant, and may have played an important role in the history and evolution of our Universe.

The cosmological implications of supersymmetry began to be discussed as early as the late 1970s. In Piet Hut’s 1977 paper on the cosmological constraints on the masses of neutrinos (as described above), the discussion was not entirely limited to neutrinos, or even to weakly interacting particles. Even the abstract of that paper mentions another possibility [164]:

Similar, but much more severe, restrictions follow for particles that interact only

gravitationally. This seems of importance with respect to supersymmetric theories.

The paper goes on to close with the first cosmological bounds on the mass of the supersymmetric partner of the graviton, the spin 3/2 gravitino:

Assuming the standard big bang model to be relevant in the context of supergravity theories, one can make the following remark. If there exist light massive spin 3/2 particles interacting only gravitationally, having four spin degrees of freedom, their mass must be less than 15 eV if they are their own antiparticles, otherwise their mass is less than 1.5 eV. Also, they may exist with masses very much larger than 1 TeV.

Although such bounds would be revised in the decades to follow, in particular being shown to depend on the temperature to which the Universe was reheated following inflation, this result is essentially the basis of what is known today as the “cosmological gravitino problem”.

In their 1982 paper, Heinz Pagels and Joel Primack also considered the cosmological implications of gravitinos [232]. But unlike Hut’s paper, or the other preceding papers that had discussed neutrinos as a cosmological relic, Pagels and Primack were clearly aware of the dark matter problem, and explicitly proposed that gravitinos could provide the solution by making up the missing mass [232]:

Gravitinos could also provide the dark matter required in galactic halos and small clusters of galaxies.

In many ways, Pagel and Primack’s letter reads like a modern paper on supersymmetric dark matter, motivating supersymmetry by its various theoretical successes and attractive features, and going on to discuss not only the missing mass in galaxies and clusters, but also the role that dark matter could play in the formation of large scale structure. At the time of Pagel and Primack’s submission, however, supersymmetry itself had not yet taken its modern form, and no truly realistic supersymmetric models had been proposed (although many important steps had been made in this direction [118–121]). This changed in December of 1981, when a paper by Savas Dimopoulos and Howard Georgi described a model that would become known as the minimal supersymmetric standard model (MSSM) [93].

The advent of the MSSM opened the door to considering superpartners other than the gravitino as cosmological relics. In particular, in the MSSM, the superpartners of the photon, the Z , and two neutral scalar Higgs bosons mix to form four particles that would become known as neutralinos. Over the past three and a half decades, neutralinos have been the single most studied candidate for dark matter, having been discussed in many thousands of scientific publications. In order to be the dark matter, however, something must stabilize the lightest neutralino, preventing these particles from decaying shortly after being created.

In supersymmetric extensions of the standard model, there exist interactions that violate the conservation of baryon and lepton number. Unless the relevant couplings are highly suppressed, such interactions are expected to cause the proton to decay on unacceptably short timescales, on the order of a year or less. It was recognized early in supersymmetry's development, however, that the proton's lifetime could be made to safely exceed observational limits if an additional – and well-motivated – symmetry known as R -parity [117, 118, 121, 269] is imposed. The R -parity of a given particle is defined as follows:

$$P_R = (-1)^{2s+3B+L}, \quad (1)$$

where s is the spin of the particle, and B and L are the particle's baryon number and lepton number, respectively. Under this definition, all of the standard model particles have positive R -parity, $P_R = +1$, while all of their superpartners have $P_R = -1$. As a consequence, this parity ensures that superpartners can only be created or destroyed in pairs. A heavy superpartner can decay into a lighter superpartner, along with any number of standard model particles, but the lightest of the superpartners cannot decay. Thus if the lightest superpartner of the MSSM is either a neutralino or a sneutrino (the superpartner of a standard model neutrino), R -parity will stabilize it, allowing it to be a potentially viable dark matter candidate. As far as we are aware, it was Pagels and Primack who were the first to invoke R -parity in order to stabilize a dark matter candidate [232].

Papers discussing the cosmological implications of stable neutralinos began to appear in 1983⁹. In the first two of these papers, Steven Weinberg [328] and Haim Goldberg [134] independently discussed the case of a photino – a neutralino whose composition is dominated by the superpartner of the photon – and derived a lower bound of 1.8 GeV on its mass by

⁹ Unstable but long-lived photinos had been considered earlier, in 1981, by Nicola Cabibbo, Glennys Farrar and Luciano Maiani [71]

requiring that the density of such particles does not overclose the Universe. A few months later, a longer paper by John Ellis, John Hagelin, Dimitri Nanopoulos, Keith Olive and Mark Srednicki considered a wider range of neutralinos as cosmological relics [113]. In Goldberg’s paper, there is no mention of the phrase dark matter or of any missing mass problem, and Ellis et al. took a largely similar approach, simply requiring that the cosmological abundance of neutralinos not be so large as to overly slow or reverse the Universe’s expansion rate. Ellis et al., however, did mention the possibility that neutralinos could make up the dark matter, although only in a single sentence [113]:

A more restrictive constraint follows from the plausible assumption that a non-relativistic [supersymmetric] fermion would participate in galaxy formation, in which case the limits on “dark matter” in galaxies allow one to deduce that $\rho_\chi \leq 2 \times 10^{-30} (\Omega h^2) \text{ gm/cm}^3$.

and in a passing footnote of Ref. [114]:

This bound comes from the overall density of the universe and is very conservative. One can argue that massive neutral fermions probably condense into galaxies in which case a more stringent limit coming from missing galactic matter could be applied.

Although far from a full embrace of a particle physics solution to the dark matter problem, these sentences (along with those expressed by Pagels and Primack [232], and by Jim Peebles within the context of massive neutrinos [240]) reflected the emergence of a new perspective¹⁰. Throughout the decades to follow, a countless number of particle physicists would motivate their proposals for physics beyond the standard model by showing that their theories could account for the Universe’s dark matter. Despite any other attractive features that a given theory might possess, if it cannot provide a dark matter candidate, it would come to be viewed as incomplete.

That supersymmetric particles, and the lightest neutralino in particular, have received so much attention as dark matter candidates is due, in large part, to the fact that the motivation for supersymmetry does not primarily rely on the dark matter problem. Particle

¹⁰ Early evidence for this transition can be found in the conferences that took place over this period of time, including the “Study Week on Cosmology and Fundamental Physics”, that was held at the Vatican in September and October of 1981.

physicists have been drawn to supersymmetry over the past four decades for its ability to solve the electroweak hierarchy problem, and to enable gauge coupling unification [94, 165, 206], combined with its unique nature as both a spacetime symmetry and an internal symmetry [148]. If in some other universe, astrophysicists had measured the cosmological density of matter to be consistent with the observed density of stars, gas, and other baryons, particle physicists in that universe may have been just as interested in supersymmetry as they are in ours. In this respect, supersymmetry's ability to provide a viable dark matter candidate is seen by many particle physicists as something of a bonus, rather than as the primary motivation to study such theories.

Supersymmetry, however, is not the only particle physics framework that is both strongly motivated in its own right, and able to provide a viable candidate for the dark matter of our Universe. In the next section, we will turn our attention to perhaps the second most studied candidate for dark matter, the axion.

C. Axions

By all measures, quantum chromodynamics (QCD) has been an incredibly successful theory, and describes the strong force and the quarks and gluons which experience it with remarkable precision. That being said, QCD does suffer from one troubling issue, known as the strong-CP problem. This problem comes down to the fact that the QCD Lagrangian contains the following term:

$$\mathcal{L}_{\text{QCD}} \supset \bar{\Theta} \frac{g^2}{32\pi^2} G^{a\mu\nu} \tilde{G}_{a\mu\nu}, \quad (2)$$

where $G^{a\mu\nu}$ is the gluon field strength tensor and $\bar{\Theta}$ is a quantity closely related to the phase of the QCD vacuum. If $\bar{\Theta}$ were of order unity, as would naively be expected, this term would introduce large charge-parity (CP) violating effects, causing the electric dipole moment of the neutron to be $\sim 10^{10}$ times larger than experimental upper bounds permit. Therefore, to be consistent with observations, the quantity $\bar{\Theta}$ must be smaller than $\sim 10^{-10}$. While this could be nothing more than a highly unlikely coincidence, it has been interpreted by many as an indication that some new physics comes in to explain why $\bar{\Theta}$ is so small. This is the essence of the strong-CP problem.

What is perhaps the most promising solution to this problem was proposed in 1977 by

Roberto Peccei and Helen Quinn [234, 235]. They showed that by introducing a new global $U(1)$ symmetry that is spontaneously broken, the quantity $\bar{\Theta}$ can be dynamically driven toward zero, naturally explaining the small observed value. Later in the same year, Frank Wilczek [336] and Steven Weinberg [327] each independently pointed out that such a broken global symmetry also implies the existence of a Nambu-Goldstone boson, called the axion. The axion acquires a small mass as a result of the $U(1)$ symmetry's chiral anomaly, on the order of $m_a \sim \lambda_{\text{QCD}}^2/f_{\text{PQ}}$, where f_{PQ} is the scale at which the symmetry is broken.

In its original conception, f_{PQ} was taken to be near the weak scale, leading to an MeV-scale axion mass. This scenario was quickly ruled out, however, by a combination of laboratory and astrophysical constraints. In particular, in contradiction with observation, axions heavier than ~ 10 keV are predicted to induce sizable rates for a number of exotic meson decays, such as $K^+ \rightarrow \pi^+ + a$ and $J/\psi \rightarrow \gamma + a$. Similarly, axions heavier than ~ 1 eV would lead to the very rapid cooling of red giant stars, again in contradiction with observations. Some years later, after the occurrence and observation of Supernova 1987A, even stronger constraints were placed on the axion mass, $m_a \lesssim 10^{-3}$ eV.

In order to evade these constraints, axions must be much lighter, and much more feebly interacting [96, 179, 286], than had been originally envisioned by Wilczek and Weinberg. Such light and “invisible” axions, however, can have very interesting consequences for cosmology. Being stable over cosmological timescales, any such axions produced in the early Universe will survive and, if sufficiently plentiful, could constitute the dark matter.

A number of mechanisms have been considered for the production of axions in the early Universe. As with other particle species, axions can be produced thermally [176, 313]. For axions light enough to avoid the above mentioned constraints, however, the thermal relic abundance is predicted to be very small, and would only be able to account for a small fraction of the dark matter density. There is, however, another production mechanism, related to the misalignment of the Peccei-Quinn field, that is likely to be more important in the mass range of interest [7, 95, 249]. Although the quantity $\bar{\Theta}$ is dynamically driven to zero by the mechanism proposed by Peccei and Quinn, its initial value was likely to be some much larger value, presumably determined through some random process. As the temperature of the Universe dropped below $T \sim \lambda_{\text{QCD}}$, and the value of $\bar{\Theta}$ was driven toward zero, the energy that had been stored in the Peccei-Quinn field gets transferred into the production of a non-thermal axion population. For typical initial conditions, this process

of misalignment production is predicted to generate a density of axions that is comparable to the dark matter density for masses on the order of $m_a \sim 10^{-5}$ eV. Alternatively, it was pointed out that as a consequence of $\bar{\Theta}$ taking on different initial values in different locations throughout space, a network of topological defects (axionic strings and domain walls) may be expected to form. The subsequent decay of these defects is predicted to generate a quantity of axions that is comparable to that resulting from misalignment production [86]. Inflation will erase this network of topological defects, however, unless it occurs prior to the breaking of the Peccei-Quinn symmetry.

In light of these considerations, axions with masses in the range of $m_a \sim 10^{-6} - 10^{-4}$ eV, and generated largely via misalignment production, have become one of the most popular and well-studied candidates for dark matter. Alternatively, it was also pointed out that if inflation occurs after the breaking of the Peccei-Quinn symmetry, then there may also be a viable anthropic scenario in which the axion mass could be much lighter [200, 308, 337]. In this scenario, the initial value of $\bar{\Theta}$ is of order unity in most regions, leading to very high axion densities and to the rapid contraction of space. In a small fraction of the overall cosmic volume, however, the initial value of $\bar{\Theta}$ will be much lower, leading to far less axion production. If we speculate that life is only able to emerge in those regions in which the Universe is allowed to expand for millions or billions of years or more, we should expect to find ourselves in a region with a density of axions that is similar to the observed density of dark matter, even if the axion is much lighter than non-anthropic estimates would lead us to expect.

D. The WIMP Paradigm

By the end of the 1980s, the conclusion that most of the mass in the Universe consists of cold and non-baryonic particles had become widely accepted, among many astrophysicists and particle physicists alike. And while alternatives continued to be discussed (see the following two chapters), cold dark matter in the form of some unknown species of elementary particle had become the leading paradigm. In addition to massive neutrinos (sterile or otherwise), supersymmetric particles (neutralinos, gravitinos, sneutrinos, axinos) and axions were each widely discussed as prospective dark matter candidates. And as the evidence in favor of non-baryonic dark matter became increasingly compelling, an ever greater number of particle

physicists began to openly speculate about the nature of this invisible substance. The result of this was a long and diverse list of exotic possibilities, ranging from topological defects produced through spontaneous symmetry breaking in the early Universe (monopoles, cosmic strings) [177], to macroscopic configurations of quark matter (centimeter-scale “nuggets”, with nuclear-scale densities) [339], and even “pyrgons” (Kaluza-Klein excitations) that could appear within the context of models with extra spatial dimensions [185].

While this proliferation of dark matter candidates was taking place, however, a commonality among many of the proposed particles was becoming increasingly appreciated. In order for a particle species to freeze-out of thermal equilibrium in the early Universe to become a cold relic, it must not be too light (roughly heavier than $\sim 1\text{-}100$ keV). Furthermore, for the predicted thermal relic abundance of such a species to match the observed dark matter density, the dark matter particles must self-annihilate with a cross section on the order of $\sigma v \sim 10^{-26}$ cm³/s (where v is the relative velocity between the annihilating particles). This number is strikingly similar to the cross section that arises from the weak force. For example, a stable neutrino with a mass of several GeV, annihilating through the exchange of a Z -boson, would freeze-out with a relic abundance that is roughly equal to the measured density of dark matter. Furthermore, such conclusions are not limited to neutrinos, but apply to a broad range of electroweak-scale dark matter candidates – including any number of stable particles with MeV-TeV masses and interactions that are mediated by the exchange of electroweak-scale particles. This observation, combined with theoretical arguments in favor of the existence of new physics at or around the electroweak scale, have elevated weakly interacting massive particles (WIMPs) [305] to the leading class of candidates for dark matter¹¹. WIMPs have been the subject of thousands of theoretical studies, leading to the refinement of many calculations, including that of the dark matter’s thermal relic abundance [137, 143, 300]. Furthermore, WIMPs (and to a somewhat lesser degree, axions) have motivated an expansive experimental program that continues to this day. With the advent of the Large Hadron Collider at CERN, and ever more sensitive astrophysical experiments, many believe that the moment of truth has come for WIMPs: either we will discover them soon, or we will begin to witness the decline of the WIMP paradigm [46].

¹¹ Although the term WIMP, as coined by Gary Steigman and Michael Turner in 1984, was originally intended to include all particle dark matter candidates, including axions, gravitinos, etc., the definition of this term has since evolved to more often denote only those particles that interact through the weak force.

VI. BARYONIC DARK MATTER

As the evidence in favor of dark matter in galaxies and galaxy clusters accumulated, more and more astronomers began to contemplate what might make up this faint material. To many astronomers and astrophysicists, the most obvious possibility was that this missing mass might consist of compact objects that were much less luminous than – but otherwise qualitatively similar to – ordinary stars. Possibilities for such objects included planets, brown dwarfs, red dwarfs, white dwarfs, neutron stars, and black holes. Kim Griest would later coin the term “MACHOs” – short for massive astrophysical compact halo objects – to denote this class of dark matter candidates, in response to the leading alternative of weakly interacting massive particles, “WIMPs”.

Although there is a consensus today that MACHOs do not constitute a large fraction of the dark matter, opinions differ as to which lines of evidence played the most important role in reaching that conclusion (for an example of some of the very early arguments that had been made against MACHOs as dark matter, see Ref. [152]). That being said, two lines of investigation would ultimately prove to be particularly important in resolving this question: searches for MACHOs using gravitational microlensing surveys, and determinations of the cosmic baryon density based on measurements of the primordial light element abundances and of the cosmic microwave background.

A. Gravitational Microlensing

The possibility that light could be deflected by gravity has a long history, extending back as far as Newton. In 1915, Einstein made the correct prediction for this phenomena using the framework of general relativity (which predicts twice the degree of deflection as Newtonian gravity). An early test of general relativity was famously conducted during the solar eclipse of 1919, which provided an opportunity to measure the bending of light around the Sun. Although the measurements obtained by Arthur Eddington favored the relativistic prediction, other simultaneous observations appeared to agree with the Newtonian expectation. Despite this apparent ambiguity, Eddington’s results were seen as persuasive by many astronomers, and served to elevate the status of Einstein’s theory.

In 1924, the Russian physicist Orest Chwolson returned to the topic of gravitational

lensing, pointing out that a massive body could deflect the light from a more distant source in such a way that would lead to the appearance of multiple images, or of a ring [80]. In 1936, Einstein himself published a paper on this topic [111], but concluded that due to the very precise alignment required, “there is no great chance of observing this phenomenon”.

The modern theory of gravitational lensing was developed in the 1960s, with contributions from Yu Klimov [180–182], Sidney Liebes [195], and Sjur Refsdal [255, 256], followed by the first observation of a lensed quasar by Dennis Walsh, Robert Carswell and Ray Weyman in 1979 [324]. In the same year, Kyongae Chang and Sjur Refsdal showed that individual stars could also act as lenses, leading to potentially observable variations over timescales of months [78]. In 1986, Bohdan Paczynski proposed that this phenomena of gravitational microlensing could be used to search for compact objects in the “dark halo” of the Milky Way [230], followed in 1987 by more detailed predictions for the probability and light curves of such events, described in the Ph.D. thesis of Robert Nemiroff [223]¹².

The strategy proposed by these authors was to simultaneously monitor large numbers of stars in a nearby galaxy (such as in the Large Magellanic Cloud), in an effort to detect variations in their brightness. If the halo consisted entirely of MACHOs, approximately one out of 2 million stars should be magnified at a given time, a ratio known as the microlensing optical depth. Furthermore, as the duration of a microlensing event is predicted to be $t \sim 130 \text{ days} \times (M/M_\odot)^{0.5}$, such a program would be best suited to detect objects with masses in the range of $\sim 10^{-7} M_\odot$ to $\sim 10^2 M_\odot$, corresponding to variations over timescales of hours to a year. These factors motivated the approaches taken by the MACHO, EROS (Experience pour la Recherche d’Objets Sombres), and OGLE (Optical Gravitational Lensing Experiment) Collaborations, who each set out to conduct large microlensing surveys in order to test the hypothesis that the Milky Way’s dark halo consisted of MACHOs.

Although the first claim of a microlensing event was reported in 1989, by Mike Irwin and collaborators [167], the implications of microlensing surveys for dark matter only began to take shape a few years later with the first results of the MACHO Collaboration. The MACHO Collaboration was a group of mostly American astronomers making use of the 1.27-meter telescope at the Mount Stromlo Observatory in Australia to simultaneously monitor millions of stars in the Large Magenellic Cloud. In October of 1993, they reported the

¹² The possibility that objects in the Milky Way’s dark halo could be detected through gravitational lensing was also discussed earlier, in a chapter of the 1981 Ph.D. thesis of Maria Petrou. On the advice of her supervisor, Petrou did not otherwise attempt to publish this work [315].

detection of their first microlensing event, consistent with a 0.03 to 0.5 M_{\odot} MACHO [20]. In the same month, the EROS Collaboration reported the detection of two such events, favoring a similar range of masses [30]. At the time, the rate of these events appeared to be consistent with that anticipated from a halo that was dominated by MACHOs. Kim Griest (a member of the MACHO Collaboration) recalled in 2000:

After the discovery of MACHOs in 1993, some thought that the dark matter puzzle had been solved.

But alas, it was not to be.

Over a period of 5.7 years, the MACHO Collaboration measured the light curves of 40 million individual stars, identifying between 14 and 17 candidate microlensing events. This was well above their expected background rate, and led them to conclude that between 8% and 50% of the Milky Way's halo mass consisted of compact objects, most of which had masses in the range of 0.15 to 0.9 M_{\odot} [21]. After collecting data for 6.7 years, however, the EROS Collaboration had identified only one microlensing candidate event, allowing them to place an upper limit of 8% on the halo mass fraction in MACHOs [193, 309]. Compact objects, at least within the mass range probed by microlensing surveys, do not appear to dominate the missing mass in the Milky Way's halo.

B. The Universe's Baryon Budget

Throughout much of the mid-twentieth century, the origin of the various nuclear species remained a subject of considerable mystery and speculation. As early as 1920, Arthur Eddington and others argued that the fusion of hydrogen into helium nuclei could be capable of providing the primary source of energy in stars, and suggested that it might also be possible to generate heavier elements in stellar interiors [104, 105]. In 1939, Hans Bethe expanded significantly upon this idea, describing the processes of the proton-proton chain and the carbon-nitrogen-oxygen cycle that are now understood to dominate the energy production in main sequence stars [52]. Fred Hoyle, in papers in 1946 and 1954, calculated that nuclei as heavy as iron could be synthesized in massive stars [160], and that even heavier nuclear species could be produced by supernovae [161].

An alternative to stellar nucleosynthesis was proposed in 1946 by George Gamow [131], and followed up upon two years later in a paper by Gamow and Hermann Alpher [22]. The

author list of this later paper also famously included Hans Bethe (who reportedly did not contribute to the research) in order to facilitate the pun that enabled it to become known as the “alpha-beta-gamma” paper. In this pair of papers, it was proposed that all nuclear species (both light and heavy) may have been produced in the early Universe through the process of neutron capture. While of historic significance, there were considerable technical problems with the calculations presented in these early papers, some of which were pointed out by Enrico Fermi, Chushiro Hayashi, and Anthony Turkevich in the years to follow. Among other flaws, Alpher and Gamow did not correctly account for Coulomb barriers in estimating the rates for nuclear fusion. Perhaps more importantly, they did not appreciate that the lack of stable nuclei with atomic numbers in the range of 5-8 would effectively prevent any significant nucleosynthesis from occurring beyond ${}^4\text{He}$. After accounting for these issues, Alpher, along with Robert Herman and James Follin, correctly predicted the abundance of helium produced in the early Universe, and reported in 1953 that the heavier elements could not be accounted for by this mechanism [23]. For these and other reasons, stellar nucleosynthesis remained the predominant theory throughout the 1950s and into the 1960s. That being said, by the late 1950s, it was becoming increasingly clear that stellar nucleosynthesis could not generate enough helium to accommodate the observed abundance, as summarized in the classic 1957 review paper by Margaret Burbidge, Geoffrey Burbidge, William Fowler, and Fred Hoyle [66].

The discovery of the cosmic microwave background in 1965 led to increased interest in Big Bang nucleosynthesis, and made it possible to further refine the predictions for the light element abundances. In particular, the temperature of this newly detected background favored a primordial helium fraction in the range of 26-28% [236, 323], consistent with observations. In 1973, a paper by Hubert Reeves, Jean Audouze, William Fowler and David Schramm focused on the production of deuterium in the early Universe [254]. As deuterium had been detected in the interstellar medium, but is not generated in stars, these authors argued that Big Bang nucleosynthesis offered the most plausible origin for the observed deuterium. In the same paper, the authors also used the measured light element abundances to derive an upper limit on the cosmological baryon density that was about one tenth of the critical density, $\Omega_b \lesssim 0.1 \Omega_{\text{crit}}$.

Constraints on the cosmological baryon density became increasingly stringent over the decades to follow. Of particular importance were the first high-precision measurements of

the primordial deuterium abundance, which were carried out in the late 1990s by Scott Burles, David Tytler, and others [69, 70, 225]. These measurements were used to determine the baryonic abundance with roughly 10% precision, $\Omega_b h^2 = 0.020 \pm 0.002$ (95% CL) [68]; leaving little room for baryonic MACHOs [129]. At around the same time, measurements of the angular power spectrum of the cosmic microwave background were also becoming sensitive to this quantity. In particular, the ratio of the heights of the odd and even peaks in this power spectrum is primarily set by the baryonic density. Although limited measurements of the second peak were made by ground- and balloon-based experiments in the late 1990s, it was not until the satellite-based WMAP experiment that these determinations became competitive with (and superior to) those based on the measured light element abundances. WMAP ultimately achieved a measurement of $\Omega_b h^2 = 0.02264 \pm 0.00050$ (68% CL) [154], while the most recent analysis from the Planck Collaboration arrives at a constraint of $\Omega_b h^2 = 0.02225 \pm 0.00016$, corresponding to a fractional uncertainty of less than one percent [245]. When this is compared to the total matter density as inferred by these and other experiments, one is forced to the conclusion that less than 20% of the matter in the Universe is baryonic.

C. Primordial Black Holes

By the late 1990s, it had become clear that baryonic dark matter does not constitute a large fraction of the Universe's dark matter. Although these results seem to imply that the dark matter must consist of one or more new particle species, there remains a caveat to this conclusion: the dark matter might instead consist of black holes that formed before the epoch of Big Bang nucleosynthesis and with masses below the sensitivity range of microlensing surveys.

The possibility that black holes may have formed in the early Universe was discussed by Barnard Carr and Stephen Hawking as early as 1974 [74]. Such primordial black holes exhibit a characteristic mass that is on the order of the mass contained within the horizon at the time of formation, $M_{\text{horizon}} \sim 10^{15} \text{ kg} \times (10^7 \text{ GeV}/T)^2$, allowing for a very large range of possible masses. A lower limit on this mass range can be placed, however, from the lack of Hawking-radiated gamma-rays from a primordial black hole population [205, 231]. Combining gamma-ray constraints [178, 343] with the null results of microlensing surveys

yields an acceptable mass range of 10^{14} kg to 10^{23} kg for dark matter in the form of primordial black holes.

A major factor that has tempered the enthusiasm for primordial black hole dark matter pertains to the number of such objects that are expected to have formed in the early Universe. If one assumes an approximately scale-invariant spectrum of density fluctuations (normalized to that observed at large scales), the predicted formation rate is cosmologically negligible. To generate a relevant abundance of such black holes, one must postulate a large degree of non-gaussianity or other such features in the primordial power spectrum.

VII. MODIFIED GRAVITY

In February of 1982, Mordehai Milgrom submitted a trio papers to the *Astrophysical Journal* [214–216]. These papers, which Milgrom developed at the Weizmann Institute in Israel and while on sabbatical at Princeton’s Institute of Advanced Study, provided the foundation for what would become the leading alternative to dark matter. This proposal, known as Modified Newtonian Dynamics, or MOND, was a seemingly simple one, but with extremely far reaching consequences. At the heart of MOND is the recognition that if instead of obeying Newton’s second law, $F = ma$, the force due to gravity scaled as $F = ma^2/a_0$ in the limit of very low accelerations ($a \ll a_0 \sim 1.2 \times 10^{-10} \text{ m/s}^2$), then it would be possible to account for the observed motions of stars and gas within galaxies without postulating the presence of any dark or otherwise hidden matter. In Milgrom’s proposal, there was no dark matter. Instead, what astronomers had discovered was evidence of a new framework for gravity and dynamics, beyond that described by Newtonian physics or even by general relativity.

A. Toward a Realistic Theory of MOND

Milgrom’s initial proposal was not intended to represent a realistic theory, but rather was presented as the approximate weak-field limit of some unknown, but more complete framework. In its original form, it was not even clear whether MOND was merely a modification of the behavior of gravity, or was instead a more general correction to Newton’s second law, applicable to all forces. Within the context of either interpretation, however, it has proven challenging to embed MOND-like behavior within a realistic theoretical framework. First of all, in its original formalism, MOND does not conserve momentum, angular momentum, or energy. Furthermore, Milgrom did not initially propose any means by which MOND could be embedded within a theory consistent with general relativity. Before MOND could be considered a viable alternative to dark matter, a more realistic version of this theory would have to be developed. And while significant progress has been made toward this goal over the past three decades, this progress has often been accomplished at great expense in terms of economy and simplicity.

A first step in this direction was made in 1984 through the collaboration of Milgrom with

Jacob Bekenstein, and their proposal of the AQUAdratic Lagrangian theory (AQUAL) [38]. In AQUAL, Bekenstein and Milgrom began with a modification of the Lagrangian of Newtonian gravity, rather than with a modification of Newton's second law. As a result, this approach automatically preserves the conservation of momentum, angular momentum, and energy, and respects the weak equivalence principle. And while the predictions of AQUAL are identical to those of MOND only in special and highly symmetric cases, the differences between the predictions of these two theories are typically modest (at the $\sim 10\%$ level) [217].

Despite its advantages over the original version of MOND, AQUAL was still a modification of Newtonian dynamics, and is not compatible with the general theory of relativity. In order for any variation of MOND to be taken seriously, it would need to be able to account for the many varieties of relativistic phenomena that have been observed, including those of gravitational lensing and cosmological expansion. The first attempts to embed MOND into a relativistic framework involved theories with more than one metric. In relativistic AQUAL (RAQUAL) [38], for example, the dynamics of matter and radiation are dictated by a metric that is distinct from the standard spacetime metric that applies to the gravitational field. The difference between these two metrics is the result of the presence of an additional scalar field which, along with matter, contributes to the gravitational potential. In this respect, RAQUAL shares some of the features of much earlier scalar-tensor theories of gravity [63, 169]. To avoid causal problems resulting from the superluminal motion of the scalar field, however, RAQUAL had to be further modified [37], and these changes were to the detriment of the theory's consistency with precision solar system tests. Even more problematic was the fact that these early attempts at a relativistic theory of MOND failed to adequately describe the phenomena of gravitational lensing.

Although the deflection of light is predicted to occur in RAQUAL and other relativistic formulations of MOND, the magnitude of such lensing is generally expected to be proportional to the amount of (baryonic) mass that is present in the deflecting system. In contrast, the degree of lensing that is observed around galaxy clusters is much larger than can be accounted for by the mass of the baryons alone. In this respect, RAQUAL cannot address the dark matter problem on cluster scales.

Although other efforts to resolve this issue were attempted [40, 274], it was not until 2004 that Bekenstein proposed the first realistic solution to the problem of gravitational lensing in relativistic theories of MOND [39]. Since its proposal, Bekenstein's TeVeS theory

– short for Tensor-Vector-Scalar gravity – has become the leading theory of MOND, and has received a great deal of attention. Beyond those of general relativity, TeVeS contains two additional fields, three free parameters, and one free function. On the one hand, this freedom makes TeVeS somewhat limited in its predictive power. On the other, it provides TeVeS with enough flexibility to potentially be consistent with gravitational lensing observations and other cosmological considerations, such as those pertaining to structure formation and the cosmic microwave background.

B. Observational Successes and Failures

Early in its history, it was appreciated that MOND was capable of explaining the observed dynamics of many spiral and elliptical galaxies. MOND also, however, made predictions for the behavior of low surface brightness galaxies, whose dynamics had not yet been well measured. The fact that such systems were later found to be compatible with MOND [75, 208] served to bolster interest in the theory. Today, MOND appears to be compatible with the observed rotations curves of hundreds of spiral galaxies [36, 218, 219, 273].

In addition to galactic rotation curves, MOND also provides an explanation for the empirical Tully-Fisher formula [311], which relates the intrinsic luminosities and rotational velocities of spiral galaxies, $L \propto V_{\text{rot}}^\alpha$, where $\alpha \approx 4$. If one assumes a common mass-to-light ratio for all galaxies, MOND predicts precisely this relationship, with a value of $\alpha = 4$, which is consistent with observations [209].

On the scale of galaxy clusters, MOND has not been nearly as successful. While MOND does reduce the need for additional mass in clusters, significant quantities of dark matter are still required. If the three known species of neutrinos were as heavy as $m_\nu \sim 1\text{-}2$ eV (near the upper limits from beta decay experiments), it has been suggested that they might be able to account for this discrepancy, essentially acting as dark matter in clusters [26, 275, 276]. Massive neutrinos can also help to reduce to some degree the discrepancy between measurements of the cosmic microwave background and the predictions of TeVeS [295].

In recent years, the debate over MOND has been focused on the use of gravitational lensing to measure the mass profiles of galaxy clusters. The idea that lensing could be used to determine the mass of a galaxy or a galaxy cluster was first proposed by Fritz Zwicky in his famous paper of 1937 (see Chapter III A). It was more than 40 years later that the first

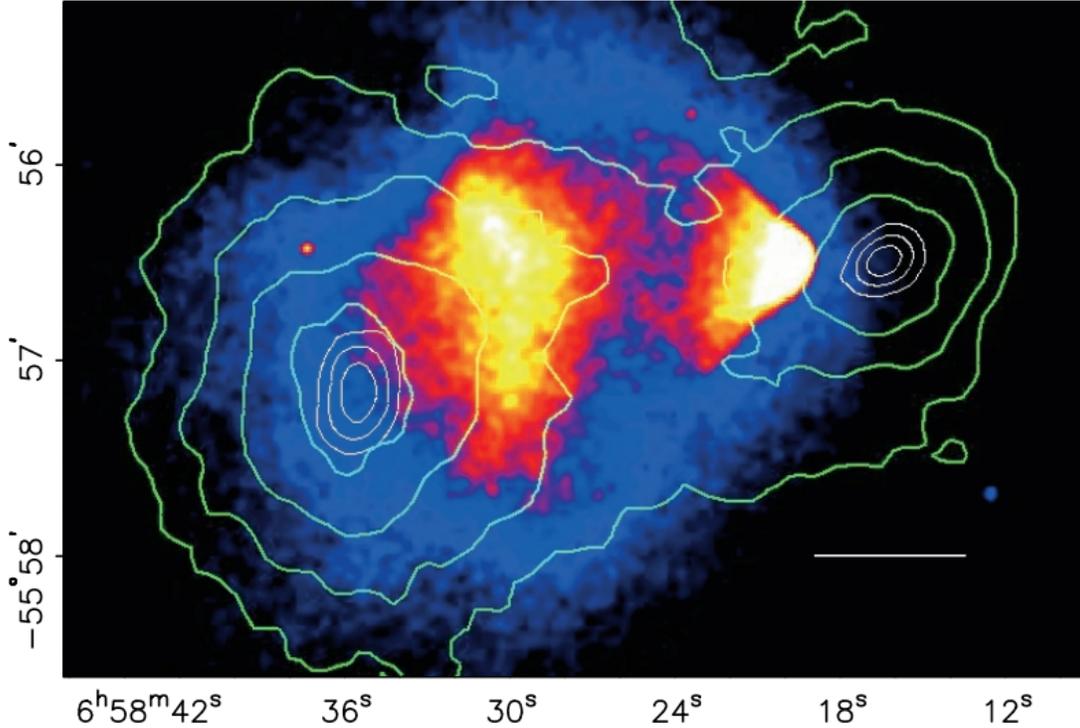


FIG. 7. The bullet cluster. The colored map represents the X-ray image of this system of merging clusters, obtained in a 500 second exposure with Chandra. The white bar is shown for scale, and represents a distance of 200 kpc at the location of the cluster. The green contours denote the reconstructed lensing signal, proportional to the projected mass in the system. From Ref. [81].

gravitational lens was observed [324] – two mirror images of a quasar – and another decade after that before the first observations were made of lensing by a galaxy cluster [204, 299]. Today, gravitational lensing is frequently used to study the properties of clusters (see e.g. Refs. [155, 211] for recent reviews).

In 2006, a group of astronomers including Douglas Clowe transformed the debate between dark matter and MOND with the publication of an article entitled, “A direct empirical proof of the existence of dark matter”. In this paper, the authors described the observations of a pair of merging clusters collectively known as the “bullet cluster” (and also known as 1E0657-558) [81]. As a result of the clusters’ recent collision, the distribution of stars and galaxies is spatially separated from the hot X-ray emitting gas (which constitutes the majority of the baryonic mass in this system). A comparison of the weak lensing and X-ray maps of the bullet cluster clearly reveals that the mass in this system does not trace the distribution

of baryons (see Fig. 7). Another source of gravitational potential, such as that provided by dark matter, must instead dominate the mass of this system.

Following these observations of the bullet cluster (and of other similar systems), many researchers expected that this would effectively bring the MOND hypothesis to an end. In the years since, however, anything but has taken place. Since the introduction of TeVeS, MOND has continued to attract a great deal of attention, despite its failure to address the dynamics of galaxy clusters, and in particular the bullet cluster. In addition to massive neutrinos, some authors have considered the possibility that TeVeS's vector field might source the gravitational potential of the bullet cluster, itself acting much like dark matter on cluster scales. Similarly, the failure of TeVeS to predict the observed ratio of the second and third peaks of the cosmic microwave background's angular power spectrum might be plausibly averted if some of TeVeS's additional degrees-of-freedom behaved much like cold dark matter during the early history of the Universe. And although this possibility goes somewhat against the original spirit of MOND, it is hard to rule out at this time. Taken together, the bullet cluster and other increasingly precise cosmological measurements have been difficult to reconcile with all proposed versions of MOND, and it remains unclear whether TeVeS, in some form, might be compatible with these observations [97, 293, 295]. For reviews, See Refs. [116, 294].

VIII. PIECING THE PUZZLE

A. Discrepancies At All Scales

When Fritz Zwicky proposed in 1933 that dark matter might be responsible for the high velocity dispersion of galaxies in the Coma Cluster (see Chapter III A), he was familiar with the concept of dark matter, and with earlier attempts to dynamically measure the density of dark matter in the Galaxy. Over the decades that followed, however, the presence of dark matter in clusters and in galaxies were discussed largely independently of each other. It wasn't until the 1960s that mass discrepancies on multiple scales began again to be considered within a common context.

In his pioneering paper of 1963, Arrigo Finzi cited Zwicky's 1933 work on galaxy clusters, the 1957 observation of M31's rotation curve from van de Hulst et al., as well as more recent determinations of the mass of the Milky Way, and argued in favor of a common interpretation for these phenomena [125]. He then went on to consider various possible forms of what we would today call "baryonic" dark matter, ruling them out one-by-one. He even went as far as to suggest that these phenomena might be explained by modifying Newton's gravitational force law, so that it scaled as $r^{-3/2}$ at large distances.

Despite the highly original and prescient nature of Finzi's work, it was largely ignored by the scientific community [277], attracting only 50 citations over the past 50 years. Although it is impossible to unambiguously identify the precise reasons for this, the very bold nature of Finzi's conclusions may have been difficult for many of his colleagues to accept, or even seriously consider. In any case, this work had little impact, and it would be another decade before other scientists began to pursue similar lines of inquiry.

As mentioned in Chapter IV, two independent groups published groundbreaking papers in 1974, each presenting a strong case for the existence of large amounts of mass in the outer parts of galaxies. The first of these papers, by the Estonian astronomers Jaan Einasto, Ants Kaasik and Enn Saar, was submitted on April 10 and was entitled "Dynamic evidence on massive coronas of galaxies" [109]. These authors began with a discussion of galactic rotation curves, citing the work of Roberts that would be eventually published in Ref. [258], and presented an analysis of rotation curve data that included estimates for the contributions from stars for five galaxies of different mass. They argued that the discrepancy between

the total mass and the stellar mass implied the existence of a “corona”, consisting of a “previously unrecognised, massive population”. They then used 105 pairs of galaxies to estimate the total mass and dimensions of their galactic coronas, concluding that the total mass of galaxies exceeded that in stars by an order of magnitude. Finally, the authors argued that these new mass estimates could also explain the mass discrepancy that had been observed in clusters. Similar arguments had also been sketched earlier by the Einasto, including at the 1972 IAU meeting in Athens [107].

On May 28 – about six weeks after Einasto et al. – Jerry Ostriker, Jim Peebles and Amos Yahil submitted a paper of similar content and scope, entitled “The size and mass of galaxies, and the mass of the universe” [229]. This paper did not present any new observations, but instead compiled existing estimates for the masses of (mostly giant spiral) galaxies. They begin with galactic rotation curves, citing the papers of Roberts and Rots [260] and Rogstad and Shostak [262] as evidence for their flatness in the outer parts of galaxies. The authors then went on to build a case for the existence of large amounts of dark matter in the outer parts of galaxies, based on mass estimates from galaxy pairs, the dynamics of dwarf galaxies, and the so-called timing argument for the Local Group. And although the observations presented in this paper were not new, and were subject to large uncertainties, the authors appear to have been confident in their conclusions, stating that the trend of increasing mass with increasing radius is “almost certainly real”, and arguing that this trend was in line with the “virial discrepancy” that had been observed in clusters and groups of galaxies [124, 264, 265]. The first sentences of this paper’s body summarizes well the sentiment of the authors:

“There are reasons, increasing in number and quality, to believe that the masses of ordinary galaxies may have been underestimated by a factor of 10 or more. Since the mean density of the Universe is computed by multiplying the observed number density of galaxies by the typical mass per galaxy, the mean density of the Universe would have been underestimated by the same factor.”

In 1979, Sandra Faber and John Gallagher published an influential review, “Masses and mass-to-light ratios of galaxies” [115], which played an important role in crystallizing the opinion among cosmologists and astronomers that dark matter was indeed abundant in the Universe. Interestingly, they chose not to use the terms “corona” or “halo”, as suggested

by the two above mentioned papers, but instead adopted the phrase “massive envelope” to describe the distribution of dark matter in astrophysical systems¹³.

B. Cosmology

As astronomers continued to gather information on the masses of galaxies, and on other observables of cosmological relevance, cosmologists began to increasingly reflect upon the implications of those findings for the formation of structure and the evolution of the Universe. In 1974, the same year as the two key papers described above [109, 229], Richard Gott, James Gunn, David Schramm and Beatrice Tinsley published a paper that provides us with an illuminating snapshot of the status of cosmology at that time [141]. The conclusions of this paper, entitled “An unbound universe”, appear within the original abstract:

“A variety of arguments strongly suggest that the density of the universe is no more than a tenth of the value required for closure. Loopholes in this reasoning may exist, but if so, they are primordial and invisible, or perhaps just black.”

In this paper, the authors argued that the body of astronomical data indicated that there was simply not enough matter in the Universe – even accounting for the large mass-to-light ratios observed among galaxies – to equal or exceed the critical density of the Universe. Among other caveats to this conclusion, they considered possible contributions from low-mass neutrinos, as had been suggested by Cowsik and McClelland, but ultimately ruled out this possibility as well.

In the early 1980s, the introduction of the theory of inflation profoundly changed the thinking of the cosmological community, and allowed one for the first time to make specific predictions for the total cosmological density and for the spectrum of density perturbations [34, 145, 147, 151, 198, 302]. This began a decade long struggle to reconcile models of structure formation with what had by then become the “theoretical imperative” of a flat Universe [84]. This struggle was exacerbated by estimates of the cosmological matter

¹³ In the discussions that took place as part of our research for this historical review, we encountered a considerable range of opinions regarding the relative importance of galactic rotation curves in establishing the existence of dark matter. This supports a picture in which different groups of scientists found quite different lines of evidence to be compelling during this period of time. Despite these disagreements regarding the strengths and weaknesses of the various observations and arguments, a consensus nonetheless began to emerge in favor of dark matter’s existence.

density arising from galaxy clusters which pointed toward a total abundance of matter – including dark matter, by then accepted by most cosmologists – that was clearly insufficient to close the Universe [334]. The resolution to this problem had to await the discovery of the accelerating expansion rate of the Universe, and the contribution to the total energy density arising from a cosmological constant or “dark energy”.

Meanwhile, Jim Peebles had pointed out that the absence of fluctuations in the cosmic microwave background at a level of $\sim 10^{-4}$ was incompatible with a Universe that was composed of only baryonic matter, and argued that this problem would be relieved if the Universe was instead dominated by massive, weakly interacting particles, whose density fluctuations could begin to grow prior to decoupling [239] (see also, Ref. [79]). This and other papers that will be discussed in the next section received enormous attention from the scientific community, and rapidly led to the establishment of cold dark matter as the leading paradigm to describe the structure and evolution of the Universe at all scales.

C. Numerical Simulations

Much of our current understanding of the structure and evolution of dark matter halos in the Universe is based on the results of computer simulations. Such explorations have a longer history than one might expect. Working in the 1940s, the ingenious Swedish scientist Erik Holmberg exploited the fact that light follows the same inverse square law as the gravitational force, and performed the first simulation of the interaction between two galaxies on an analog computer that consisted of 74 light-bulbs, photo-cells and galvanometers. He then calculated the amount of light received by each cell, and manually moved the light bulbs in the direction that received the most light.

Holmberg published his paper in November of 1941, shortly before the United States entered World War II. In the following years, the work of many research institutes ground to a halt, but science meanwhile continued to make progress, thanks in large part to the enormous resources made available to military research, especially at the Los Alamos National Laboratory, at which computers and advanced numerical techniques were developed within the context of the Manhattan Project. The first application of such computers to gravitational systems was arguably performed by John Pasta and Stanislaw Ulam in 1953. Their numerical experiments were performed on the Los Alamos computer, which by then

had been applied to a variety of other problems, including early attempts to decode DNA sequences and the first chess-playing program. A number of other pioneering studies exploring the evolution of a system of gravitationally interacting massive particles appeared throughout the 1960s and 1970s, gradually increasing the number of simulated particles from $N \sim 100$ in the early works to $N \sim 1000$ [3, 237, 320, 321, 331].

By the early 1970s, it had become possible to numerically simulate the dynamics of galaxies. Simulations carried out by Richard Miller, Kevin Prendergast and William Quirk [220] as well as by Frank Hohl [156] each found rotationally supported galaxies consisting of a stellar disk to be unstable, in contradiction with observations. Instead of reaching an equilibrium configuration, such systems were found to change rapidly, forming bars and evolving toward a more elliptical and pressure supported configuration. The solution to this problem was proposed in 1973 by Jerry Ostriker and Jim Peebles, who recognized that a rotationally supported stellar disk could be stable if embedded within a massive spherical halo [228].

The first attempt to numerically solve the formation and evolution of cosmological structures in an expanding universe was presented in a famous paper published in 1974 by William Press and Paul Schechter [250]. This was followed by a number of developments in the late 1970s and early 1980s that significantly advanced the power of such endeavors (see, for example, Refs. [4, 77, 106, 142, 183, 184, 312]). First, a combination of improvements in processor speed and in numerical techniques made it possible for the first time to simultaneously simulate millions of particles. Second, the newly proposed theory of inflation [146, 199] offered a physical means by which initial density perturbations could be generated, providing the initial conditions for cosmological simulations. And third, the results of the first large 3D survey of galaxies (the CfA redshift survey) were published in 1982, providing a distribution that could be directly compared with the output of simulations.

In some ways, the results of cosmological simulations do not depend much on what the dark matter consists of. In particular, they are largely insensitive to the electroweak or other non-gravitational interactions that may (or may not) be experienced by dark matter particles – for the purposes of structure formation, such particles are effectively “collisionless”. What does impact the results of such simulations, however, is the initial velocity distribution of the dark matter particles [58, 59, 99]. Importantly, this provides cosmologists with a way to discriminate between different classes of dark matter candidates. Standard model neutrinos, for example, decoupled from thermal equilibrium in the early Universe at a temperature that

is much greater than their mass, and thus remained highly relativistic throughout the epoch of structure formation. In contrast, supersymmetric neutralinos are predicted to freeze-out of thermal equilibrium at a temperature below their mass, and are thus non-relativistic throughout cosmic history. Axions generated through misalignment production are also predicted to be produced with non-relativistic velocities.

At the largest scales – those associated with galaxy clusters and superclusters – cosmological simulations predict a pattern of structure that is largely insensitive to the initial velocities of the dark matter. At smaller scales, however, density fluctuations can be washed out by the random thermal motion of individual dark matter particles. As a result, the growth of small scale structure is predicted to be suppressed if the dark matter is relativistic, or “hot” [240, 281]. Non-relativistic, or “cold” dark matter particles undergo a very different sequence of structure formation. The much shorter free-streaming length of such particles allow them to form very low mass halos; roughly in the range of $\sim 10^{-3}$ to $\sim 10^{-9} M_{\odot}$ for a typical neutralino, for example. These very small halos form very early in the Universe’s history, and then go on to merge with one another, gradually building up larger and larger dark matter structures. This bottom-up, or hierarchical, process of structure formation is in stark contrast to the top-down sequence predicted for hot dark matter.

Simulations of large scale structure are, of course, only useful if their results can be compared to the actual patterns of structure found in the Universe. This was made possible with the CfA survey, which was the first extensive 3D survey of galaxies in the local Universe [85]. Among other features, CfA revealed the first indications of the “cosmic web”, which described the distribution of matter on the largest scales. This survey also identified the presence of significant structure on sub-cluster scales, in conflict with the predictions of hot dark matter simulations [333].

In the wake of the failures of hot dark matter, it was quickly becoming appreciated that cold dark matter could do a much better job of accounting for the observed patterns of large scale structure. To quote the 1984 paper by George Blumenthal, Sandra Faber, Joel Primack, and Martin Rees [54]:

“We have shown that a universe with ~ 10 times as much cold dark matter as baryonic matter provides a remarkably good fit to the observed universe. This model predicts roughly the observed mass range of galaxies, the dissipational nature of galaxy collapse, and the observed Faber-Jackson and Tully-Fisher rela-

tions. It also gives dissipationless galactic halos and clusters. In addition, it may also provide natural explanations for galaxy-environment correlations and for the differences in angular momenta between ellipticals and spiral galaxies.”

The first simulations of cold dark matter were carried out by Marc Davis, George Efstathiou, Carlos Frenk, and Simon White, who published their results in 1985 [84]. The resemblance of their simulated distribution of dark matter halos to that of the galaxies in the CfA survey was clear, serving to further elevate the status of cold dark matter within the cosmological community.

By middle of the 1980s, the paradigm of cold dark matter was well on its way to becoming firmly established. And although scenarios involving mixed dark matter (containing significant quantities of both cold and hot dark matter) and warm dark matter (suppressing structure only on the scale of dwarf galaxies and below) would each continue to be discussed in the literature, the possibility that the dark matter was dominated by neutrinos or other relativistic particles was quickly abandoned.

A decade later, the predictions of cosmological simulations had shifted in focus from the distribution of cold dark matter halos to the shapes of those halos. In 1996, Julio Navarro, Carlos Frenk and Simon White published a remarkable result, based on an analysis of the halos generated in their high-resolution cold dark matter simulations [222]:

The spherically averaged density profiles of all our halos can be fit over two decades in radius by scaling a simple universal profile. The characteristic overdensity of a halo, or equivalently its concentration, correlates strongly with halo mass in a way which reflects the mass dependence of the epoch of halo formation.

The simple fitting formula derived by the authors became known as the Navarro-Frenk-White profile. This parametrization is still widely used today, and represents the primary benchmark for most dark matter detection studies, despite the fact that it is expected to be inaccurate in the innermost regions of galaxies, where baryons dominate the gravitational potential.

In more recent years, the frontier for cosmological simulations has focused on the implementation of baryonic physics, including the hydroynamical evolution of gas in astrophysical structures, stellar formation, and feedback from supernova explosions and black

holes. Current simulations are not yet able to resolve all relevant scales – which range between sub-parsec distances for stellar formation to Gpc scales for cosmological structures – but implement baryonic physics through the introduction of suitable “sub-grid” parameters which attempt to encode the collective behaviour of large amounts of gas and stars. Such parameters are generally tuned to match observable quantities, such as the galaxy mass function and the galaxy-central black hole mass relation, as in e.g. the recent suite of Eagle simulations [279].

IX. THE HUNT FOR DARK MATTER PARTICLES

As particle physicists became increasingly interested in the problem of the missing matter of the Universe, some began to turn their attention toward ways that individual particles of dark matter might be detected, either directly or indirectly. Although many of the leading techniques were first conceived of in the 1980s, dark matter searches have continued with vigor ever since, occupying the attentions of generations of experimental particle-astrophysicists.

A. Scattering with Nuclei

In 1984, an article by Andrzej Drukier and Leo Stodolsky at the Max Planck Institute in Munich appeared in *Physical Review D*, discussing techniques that might be used to detect neutrinos scattering elastically off nuclei [102]. Among other possibilities, the article proposed the use of a superconducting colloid detector, consisting of micron-scale superconducting grains maintained at a temperature just below their superconducting transition. Even a very small quantity of energy deposited by the recoil of an incident neutrino could cause a superconducting grain to flip into the normal state, collapsing the magnetic field and producing a potentially measurable electromagnetic signal. In January 1985, Mark Goodman and Ed Witten submitted a paper to the same journal, arguing that this technology could also be used to detect some types of dark matter particles [140]¹⁴. Although Drukier and Stodolsky’s original detector concept was never employed at a scale sensitive to dark matter, the broader notion of experiments capable of detecting ~ 1 -100 keV nuclear recoils provided a path through which it appeared possible to test the WIMP hypothesis.

In their original paper, Goodman and Witten considered three classes of dark matter candidates: 1) those that undergo coherent scattering with nuclei (also known as spin-independent scattering), 2) those that scatter with nuclei through spin-dependent couplings, and 3) those with strong interactions. The first two of these three categories provide the basis for how most direct dark matter detection results have since been presented. If mediated by unsuppressed couplings to the Z boson (an important early benchmark), coherent scattering was predicted to lead to large scattering rates, typically hundreds or thousands of

¹⁴ A similar paper by Ira Wasserman [325] was submitted shortly after Goodman and Witten’s.

events per day per kilogram of target material. With such high rates, the prospects for detecting dark matter in the form of a heavy neutrino or sneutrino appeared very encouraging. Dark matter candidates that scatter with nuclei only through spin-dependent couplings, in contrast, were generally predicted to yield significantly lower rates, and would require larger and more sensitive detectors to test. Even as early as in this first paper, Goodman and Witten pointed out that such experiments would have difficulty detecting dark matter particles lighter than $\sim 1\text{-}2$ GeV, due to the modest quantity of momentum that would be transferred in the collisions.

The first experiment to place constraints on the scattering cross section of dark matter with nuclei was carried out in 1986 at the Homestake Mine in South Dakota by a collaboration of scientists at the Pacific Northwest National Laboratory, the University of South Carolina, Boston University, and Harvard [18]. Using a low-background germanium ionization detector (originally designed to search for neutrinoless double beta decay), they accumulated an exposure of 33 kg-days, yielding a limit that significantly constrained dark matter candidates with unsuppressed spin-independent scattering cross sections with nuclei (such as heavy neutrinos or sneutrinos) [18]. Shortly thereafter, similar results were obtained by an independent collaboration of scientists from the Universities of California at Santa Barbara and Berkeley [72].

Despite the importance of these first dark matter scattering limits, the reach of such detectors quickly became limited by their backgrounds, making it difficult to achieve significant improvements in sensitivity. One possible solution to this problem, first suggested by Andrzej Drukier, Katherine Freese, and David Spergel [101], was to search for an annual variation in the rate of dark matter induced events in such an experiment, as was predicted to result from the combination of the Earth's motion around the Sun and the Sun's motion through the dark matter halo. Such a technique could, in principle, be used to identify a signal of dark matter scattering over a large rate of otherwise indistinguishable background events. The most well known group to employ this technique was the DAMA/NaI Collaboration (and later DAMA/LIBRA). The original DAMA/NaI experiment consisted of nine 9.70 kg scintillating thallium-doped sodium iodide crystals, located in Italy's deep underground Gran Sasso Laboratory. In 1998, they published their first results, reporting the observation of an annually modulating rate consistent with dark matter scattering [43]. Over the past nearly two decades, DAMA's signal has persisted and become increasingly statistically

significant as more data was collected [42], including with the more recent DAMA/LIBRA detector [44, 45]. At this point in time, it seems hard to reconcile dark matter interpretations of the DAMA/LIBRA signal with the null results of other direct detection experiments. On the other hand, no convincing alternative explanation for this signal has been so far identified.

During the period of time that DAMA/NaI was being developed and collecting its first data, experimental techniques were being pursued that could discriminate dark matter-like nuclear recoil events from various backgrounds. These efforts ultimately lead to the technologies employed by the CDMS (Cryogenic Dark Matter Search), EDELWEISS (Experience pour DEtecter Les Wimps En Site Souterrain), and CRESST (Cryogenic Rare Event Search with Superconducting Thermometers) Collaborations. These experiments each made use of two-channel detectors, capable of measuring both ionization and heat (CDMS, EDELWEISS) or scintillation and heat (CRESST), the ratio of which could be used to discriminate nuclear recoil events from electron recoils generated by gamma and beta backgrounds. All three of these experiments employed crystalline target materials, maintained at cryogenic temperatures, consisting of germanium and silicon, germanium, and calcium tungstate, respectively. Throughout most the first decade of the 21st century, the CDMS and EDELWEISS experiments lead the field of direct detection, providing the most stringent constraints and improving in sensitivity by more than two orders of magnitude over that period of time (see Fig. 8).

In order to continue to increase the sensitivity of direct dark matter experiments, it was necessary for experiments to employ ever larger targets, gradually transitioning from the kilograms of detector material used by EDELWEISS and CDMS (9.3 kg in the case of SuperCDMS) to the ton-scale and beyond. Cryogenic solid state detectors, however, have proven to be costly to scale up into ton-scale experiments. In the late 1990s, Pio Picchi, Hanguo Wang and David Cline pioneered an alternative technique that exploited liquid noble targets (most notably liquid xenon). Like solid state detectors, such experiments discriminate nuclear recoils from electron recoils by measuring two quantities of deposited energy; in this case scintillation and ionization. Between 2010 and 2015, the XENON100 and LUX experiments (each of which utilize a liquid xenon target) have improved upon the limits placed by CDMS by approximately two orders of magnitude. It is generally anticipated that future experiments employing liquid xenon targets (XENON1T, LZ, XENON-NT) will

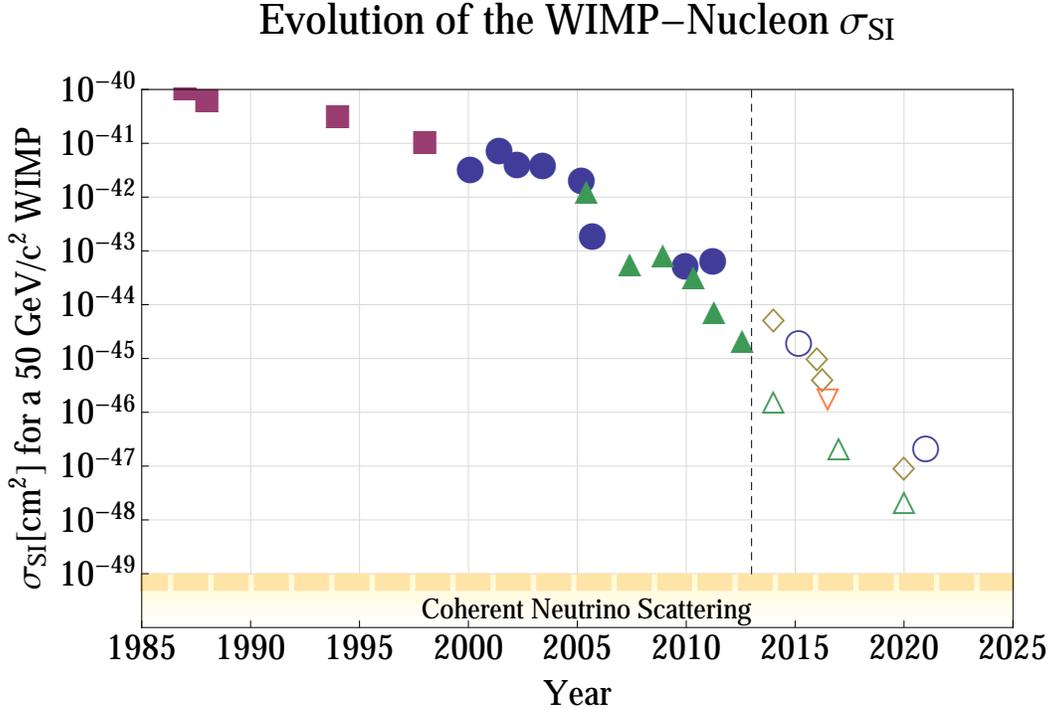


FIG. 8. The past and projected evolution of the spin-independent WIMP-nucleon cross section limits for a 50 GeV dark matter particle. The shapes correspond to limits obtained using different detectors technologies: cryogenic solid state detectors (blue circles), crystal detectors (purple squares), liquid argon detectors (brown diamonds), liquid xenon detectors (green triangles), and threshold detectors (orange inverted triangle). Taken from Ref. [83].

continue along this trajectory for years to come.

As CDMS, EDELWEISS, XENON100, LUX and other direct detection experiments have increased in sensitivity over the past decades, they have tested and ruled out an impressive range of particle dark matter models. And although results from the CoGeNT [1, 2], CRESST [24], and CDMS [14] experiments were briefly interpreted as possible dark matter signals, they now appear to be the consequences of poorly understood backgrounds [25, 173] and/or statistical fluctuations. While many viable WIMP models remain beyond the current reach of this experimental program, a sizable fraction of the otherwise most attractive candidates have been excluded. Of particular note is the fact that these experiments now strongly constrain dark matter particles that scatter coherently with nuclei through Higgs exchange, representing an important theoretical benchmark.

B. Annihilation and Decay

In the 1978 Valentine’s Day issue of *Physical Review Letters*, there appeared two articles that discussed – for the first time – the possibility that the annihilations of pairs of dark matter particles might produce an observable flux of gamma rays. And although each of these papers (by Jim Gunn, Ben Lee,¹⁵ Ian Lerche, David Schramm and Gary Steigman [144], and by Floyd Stecker [303]) focused on dark matter in the form of a heavy stable lepton (i.e. a heavy neutrino), similar calculations would later be applied to a wide range of dark matter candidates. On that day, many hopeless romantics became destined to a lifetime of searching for signals of dark matter in the gamma-ray sky.

At the time, the most detailed measurement of the astrophysical gamma-ray background was that made using data from the Small Astronomy Satellite (SAS) 2 [123]. Although the intensity of 35-100 MeV gamma rays measured by this telescope ($\sim 6 \times 10^{-5} \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$) was several orders of magnitude higher than that predicted from annihilating dark matter particles smoothly distributed throughout the Universe, it was recognized that inhomogeneities in the dark matter distribution could increase this prediction considerably. In particular, annihilations taking place within high-density dark matter halos, such as that of the Milky Way, could plausibly produce a flux of gamma rays that was not much fainter than that observed at high galactic latitudes, and with a distinctive gradient on the sky [144, 303]. Focusing on GeV-scale dark matter particles, Gunn et al. went as far as to state that such a signal “may be discoverable in future γ -ray observations”.

Several years later, in 1984, Joe Silk and Mark Srednicki built upon this strategy, considering not only gamma rays as signals of annihilating dark matter particles, but also cosmic-ray antiprotons and positrons [291] (see also, Refs. [112, 172, 304]). They argued that the observed flux of ~ 0.6 - 1.2 GeV antiprotons [64] provided the greatest sensitivity to annihilating dark matter, and noted that ~ 10 GeV WIMPs would be predicted to produce a quantity of cosmic-ray antiprotons that was comparable to the observed flux.

In 1985, Lawrence Krauss, Katherine Freese, David Spergel and William Press published a paper suggesting that neutrinos might be detected from dark matter annihilating in the core of the Sun [189] (see also, Ref. [251]). Shortly thereafter, Silk, Olive, and Srednicki pointed out that not only could elastic scattering cause dark matter particles to become

¹⁵ In regards to Ben Lee, who died in a traffic accident in 1977, this article was published posthumously.

gravitationally bound to and captured within the Sun, but that the number of WIMPs captured over the age of the Solar System could be sufficiently high to attain equilibrium between the processes of capture and annihilation [290]. Observations over the subsequent few years by the proton decay experiments IMB, FREJUS, and Kamioka capitalized on this strategy, strongly constraining some classes of dark matter candidates, most notably including light electron or muon sneutrinos. Similar approaches using dark matter capture by the Earth were also proposed around the same time [127, 190].

In the decades that followed, measurements of astrophysical gamma ray, antimatter, and neutrino fluxes improved dramatically. In parallel, the scientific community’s understanding of the astrophysical sources and propagation of such particles also matured considerably. Information from successive gamma-ray satellite missions, including COS-B [153], EGRET (Energetic Gamma Ray Experiment Telescope) [301], and the Fermi Gamma-Ray Space Telescope, gradually lead to the conclusion that most of the observed gamma-ray emission could be attributed to known gamma-ray source classes (such as active galactic nuclei), although it remains possible that a non-negligible component of the high-latitude background could originate from dark matter [9].

Motivated by their high densities of dark matter and low levels of baryonic activity, dwarf spheroidal galaxies – satellites of the Milky Way – have in recent years become a prime target of gamma-ray telescopes searching for evidence of dark matter annihilations. Fermi’s study of dwarf galaxies has provided the strongest limits on the dark matter annihilation cross section to date, strongly constraining WIMPs lighter than ~ 100 GeV or so in mass [10]. Ground based gamma-ray telescopes have also used observations of dwarf galaxies to constrain the annihilations of heavier dark matter candidates. Although complicated by imperfectly understood backgrounds, gamma-ray observations of the Milky Way’s Galactic Center are also highly sensitive to annihilating WIMPs. A significant excess of GeV-scale gamma-rays has been identified from this region, consistent with arising from the annihilations of ~ 50 GeV particles [87, 139]. An active debate is currently taking place regarding the interpretation of these observations. Alternative targets for indirect searches have also been proposed, including Galactic dark matter subhalos not associated with dwarf galaxies [92, 244], and density “spikes” of dark matter around black holes [49, 138, 345].

Over approximately the same period of time, great progress has also been made in the measurement of the cosmic-ray antiproton spectrum, including successive advances by the

CAPRICE [56, 57], BESS [8, 27], AMS [16], and PAMELA [13] experiments. When these measurements are combined with our current understanding of cosmic-ray production and propagation, they appear to indicate that the observed cosmic ray antiproton spectrum originates largely from conventional secondary production (cosmic-ray interactions with gas), although a significant contribution from dark matter remains a possibility. These measurements generally yield constraints on annihilating dark matter that are not much less stringent than those derived from gamma-ray observations.

Compared to antiprotons, measurements of the cosmic-ray positron spectrum have been more difficult to interpret. Building upon earlier measurements [55, 135, 221], the balloon-borne HEAT experiment observed in 1994, 1995, and 2000 indications of an excess of cosmic-ray positrons at energies above ~ 10 GeV, relative to the rate predicted from standard secondary production [35]. This was later confirmed, and measured in much greater detail, by a series of space-based experiments: AMS [17], PAMELA [12], and AMS-02 [15]. Although this positron excess received much attention as a possible signal of annihilating dark matter, this possibility is now strongly constrained by a variety of arguments (e.g. Ref. [47, 130, 296]), and plausible astrophysical explanations have also been proposed (e.g. Ref. [159]).

As large volume neutrino telescopes began to be deployed, such experiments became increasingly sensitive to dark matter annihilating in the interiors of the Sun and Earth. The AMANDA detector at the South Pole [19], along with Super-Kamiokande in Japan [90], each significantly improved upon previous limits, to be followed most notably by IceCube [5] and ANTARES [11]. Constraints from neutrino telescopes are currently competitive with those derived from direct detection experiments for the case of WIMPs with spin-dependent interactions with nuclei.

Many of the strategies employed to search for annihilating dark matter have also been used to constrain the rate at which dark matter particles might decay. In addition to constraints on gravitinos and other potentially unstable particles, such searches are particularly interesting within the context of sterile neutrino dark matter. Sterile neutrinos with masses in the range of ~ 1 -100 keV are predicted to decay (into an active neutrino and a photon) at a rate that could generate a potentially observable X-ray line [6]. In fact, considering the standard case of Dodelson-Widrow production (as discussed in Chapter V), the combination of constraints from X-ray observations and measurements of the Lyman- α forest [284, 319]

disfavor sterile neutrino dark matter over this entire mass range. Models with enhanced production in the early Universe [285] can evade such constraints, however, and continue to receive considerable interest. In particular, reports of a 3.55 keV line observed from a collection of galaxy clusters [61, 65] have recently received a great deal of attention within the context of a decaying sterile neutrino.

C. Axion Experiments

For some time, there has been an active experimental program searching for dark matter axions, most notably in the form of the Axion Dark Matter eXperiment (ADMX). The idea behind this effort is to make use of the photon-photon-axion coupling, generically present in axion models, to convert dark matter axions in a strong and static magnetic field into a signal of nearly monochromatic microwave photons. This possibility was first suggested by Pierre Sikivie in 1983 [288], and was later expanded upon by Sikivie [289], along with Lawrence Krauss, John Moody, Frank Wilczek and Donald Morris [188]. As the signal in such an experiment is maximized for a specific cavity frequency (corresponding to a specific axion mass), it is necessary that the resonant frequency of the cavity be tunable, making it possible to scan over a range of axion masses.

The first laboratory constraints on dark matter axions were presented in the late 1980s, by a number of groups [88, 149, 342]. While the frequency range covered by these experiments was well suited to axion masses favored by dark matter abundance considerations (covering approximately $m_a \simeq 4.5 - 16.3 \mu\text{eV}$), their sensitivity was orders of magnitude below that required to test realistic axion models. In 2003, however, the ADMX Collaboration reported results that constrained realistic axion dark matter models, although only for a relatively narrow range of masses, $1.9 - 3.3 \mu\text{eV}$ [29]. With anticipated upgrades [28], ADMX is expected to be sensitive to a much larger range of axion masses and couplings, significantly constraining the axion dark matter parameter space in the coming years.

Acknowledgements. We are grateful to the many pioneering physicists and astronomers who helped us to reconstruct the methods, ideas and circumstances that led to the establishment of the dark matter paradigm. In particular, we would to thank Andrea Biviano, Lars Bergström, Albert Bosma, Shantanu Desai, Jaco de Swart, Jaan Einasto, John Ellis, Pierre Fayet, Ken Freeman, Katherine Freese, Steve Kent, Rocky Kolb, Keith Olive, Jerry Ostriker, Sergio Palomares-Ruiz, Jim Peebles, Joel Primack, Morton Roberts, Bernard Sadoulet, Bob Sanders, Gary Steigman, Alar Toomre, Juri Toomre, Scott Tremaine, Virginia Trimble, Michael Turner and Simon White. We also acknowledge many useful discussions with Jeroen van Dongen and Jaco de Swart. This work would not have been possible without the vast collection of articles and books, dating back to the 19th century, made freely available by NASA ADS and the Internet Archive Project. GB acknowledges support from the European Research Council through the ERC starting grant *WIMPs Kairos*. DH is supported by the US Department of Energy.

Appendix A: Further Reading

- Robert H. Sanders. *The Dark Matter Problem: A Historical Perspective*. Cambridge University Press, 2010 [277].
- Virginia Trimble. *History of Dark Matter in Galaxies*, 2013 [310].
- Jaan Einasto. *Dark matter and Cosmic Web Story*. World Scientific, 2014 [108].
- H. S. Kragh. *Conceptions of Cosmos: From Myths to the Accelerating Universe*. Oxford University Press, 2006.
- Andrea Biviano. *From Messier to Abell: 200 Years of Science with Galaxy Clusters*, 2000 [53].
- Gianfranco Bertone, Dan Hooper and Joseph Silk. *Particle dark matter: Evidence, candidates and constraints*, 2004 [48].
- Lars Bergström. *Nonbaryonic dark matter: Observational evidence and detection methods*, 2000 [41].

- Gerard Jungman, Marc Kamionkowski and Kim Griest. *Supersymmetric Dark Matter*, 1996 [170].
- Edmund Bertschinger. *Simulations of Structure Formation in the Universe*, 1998 [50].
- Carlos Frenk and Simon White. *Dark matter and cosmic structure*, 2012 [128].
- Jeremiah P. Ostriker and Simon Mitton. *Heart of Darkness: Unraveling the Mysteries of the Invisible Universe*. Princeton University Press, 2013.

-
- [1] Aalseth, C., P. Barbeau, J. Colaresi, J. Collar, J. Diaz Leon, *et al.* (2011), Phys.Rev.Lett. **107**, 141301, arXiv:1106.0650 [astro-ph.CO].
 - [2] Aalseth, C., *et al.* (CoGeNT) (2011), Phys.Rev.Lett. **106**, 131301, arXiv:1002.4703 [astro-ph.CO].
 - [3] Aarseth, S. J. (1963), MNRAS **126**, 223.
 - [4] Aarseth, S. J., E. L. Turner, and J. R. Gott, III (1979), Astrophys. J. **228**, 664.
 - [5] Aartsen, M., *et al.* (IceCube) (2013), Phys.Rev.Lett. **110** (13), 131302, arXiv:1212.4097 [astro-ph.HE].
 - [6] Abazajian, K., G. M. Fuller, and W. H. Tucker (2001), Astrophys.J. **562**, 593, arXiv:astro-ph/0106002 [astro-ph].
 - [7] Abbott, L., and P. Sikivie (1983), Phys.Lett. **B120**, 133.
 - [8] Abe, K., H. Fuke, S. Haino, T. Hams, A. Itazaki, *et al.* (2008), Phys.Lett. **B670**, 103, arXiv:0805.1754 [astro-ph].
 - [9] Ackermann, M., *et al.* (Fermi-LAT) (2015), arXiv:1501.05464 [astro-ph.CO].
 - [10] Ackermann, M., *et al.* (Fermi-LAT) (2015), Phys. Rev. Lett. **115** (23), 231301, arXiv:1503.02641 [astro-ph.HE].
 - [11] Adrian-Martinez, S., *et al.* (ANTARES) (2013), JCAP **1311**, 032, arXiv:1302.6516 [astro-ph.HE].
 - [12] Adriani, O., *et al.* (PAMELA) (2009), Nature **458**, 607, arXiv:0810.4995 [astro-ph].
 - [13] Adriani, O., *et al.* (PAMELA) (2010), Phys.Rev.Lett. **105**, 121101, arXiv:1007.0821 [astro-ph.HE].

- [14] Agnese, R., *et al.* (CDMS) (2013), *Phys.Rev.Lett.* **111** (25), 251301, arXiv:1304.4279 [hep-ex].
- [15] Aguilar, M., G. Alberti, B. Alpat, A. Alvino, G. Ambrosi, K. Andeen, H. Anderhub, L. Ar-ruda, P. Azzarello, A. Bachlechner, and *et al.* (2013), *Physical Review Letters* **110** (14), 141102.
- [16] Aguilar, M., *et al.* (AMS) (2002), *Phys.Rept.* **366**, 331.
- [17] Aguilar, M., *et al.* (AMS 01) (2007), *Phys.Lett.* **B646**, 145, arXiv:astro-ph/0703154 [ASTRO-PH].
- [18] Ahlen, S., F. Avignone, R. Brodzinski, A. Drukier, G. Gelmini, *et al.* (1987), *Phys.Lett.* **B195**, 603.
- [19] Ahrens, J., *et al.* (AMANDA) (2002), *Phys.Rev.* **D66**, 032006, arXiv:astro-ph/0202370 [astro-ph].
- [20] Alcock, C., C. W. Akerlof, R. A. Allsman, T. S. Axelrod, D. P. Bennett, S. Chan, K. H. Cook, K. C. Freeman, K. Griest, S. L. Marshall, H.-S. Park, S. Perlmutter, B. A. Peterson, M. R. Pratt, P. J. Quinn, A. W. Rodgers, C. W. Stubbs, and W. Sutherland (1993), *Nature* (London) **365**, 621, astro-ph/9309052.
- [21] Alcock, C., R. A. Allsman, D. R. Alves, T. S. Axelrod, A. C. Becker, D. P. Bennett, K. H. Cook, N. Dalal, A. J. Drake, K. C. Freeman, M. Geha, K. Griest, M. J. Lehner, S. L. Marshall, D. Minniti, C. A. Nelson, B. A. Peterson, P. Popowski, M. R. Pratt, P. J. Quinn, C. W. Stubbs, W. Sutherland, A. B. Tomaney, T. Vandehei, and D. Welch (2000), *Astrophys. J.* **542**, 281, astro-ph/0001272.
- [22] Alpher, R. A., H. Bethe, and G. Gamow (1948), *Physical Review* **73**, 803.
- [23] Alpher, R. A., J. W. Follin, and R. C. Herman (1953), *Physical Review* **92**, 1347.
- [24] Angloher, G., M. Bauer, I. Bavykina, A. Bento, C. Bucci, *et al.* (2012), *Eur.Phys.J.* **C72**, 1971, arXiv:1109.0702 [astro-ph.CO].
- [25] Angloher, G., *et al.* (CRESST-II) (2014), *Eur.Phys.J.* **C74** (12), 3184, arXiv:1407.3146 [astro-ph.CO].
- [26] Angus, G. W., H. Shan, H. Zhao, and B. Famaey (2007), *Astrophys.J.* **654**, L13, arXiv:astro-ph/0609125 [astro-ph].
- [27] Asaoka, Y., Y. Shikaze, K. Abe, K. Anraku, M. Fujikawa, *et al.* (2002), *Phys.Rev.Lett.* **88**, 051101, arXiv:astro-ph/0109007 [astro-ph].

- [28] Asztalos, S., *et al.* (ADMX) (2010), Phys.Rev.Lett. **104**, 041301, arXiv:0910.5914 [astro-ph.CO].
- [29] Asztalos, S. J., *et al.* (ADMX) (2004), Phys.Rev. **D69**, 011101, arXiv:astro-ph/0310042 [astro-ph].
- [30] Aubourg, E., P. Bareyre, S. Bréhin, M. Gros, M. Lachièze-Rey, B. Laurent, E. Lesquoy, C. Magneville, A. Milsztajn, L. Moscoso, F. Queinnec, J. Rich, M. Spiro, L. Vigroux, S. Zylberajch, R. Ansari, F. Cavalier, M. Moniez, J.-P. Beaulieu, R. Ferlet, P. Grison, A. Vidal-Madjar, J. Guibert, O. Moreau, F. Tajahmady, E. Maurice, L. Prévôt, and C. Gry (1993), Nature (London) **365**, 623.
- [31] Babcock, H. W. (1939), Lick Observatory Bulletin **19**, 41.
- [32] Bahcall, J. N. (1984), Astrophys. J. **287**, 926.
- [33] Bahcall, J. N. (1984), Astrophys. J. **276**, 169.
- [34] Bardeen, J. M., P. J. Steinhardt, and M. S. Turner (1983), Phys. Rev. D **28**, 679.
- [35] Barwick, S., *et al.* (HEAT) (1997), Astrophys.J. **482**, L191, arXiv:astro-ph/9703192 [astro-ph].
- [36] Begeman, K., A. Broeils, and R. Sanders (1991), Mon.Not.Roy.Astron.Soc. **249**, 523.
- [37] Bekenstein, J. (1988), Phys.Lett. **B202**, 497.
- [38] Bekenstein, J., and M. Milgrom (1984), Astrophys.J. **286**, 7.
- [39] Bekenstein, J. D. (2004), Phys.Rev. **D70**, 083509, arXiv:astro-ph/0403694 [astro-ph].
- [40] Bekenstein, J. D., and R. H. Sanders (1994), Astrophys.J. **429**, 480, arXiv:astro-ph/9311062 [astro-ph].
- [41] Bergstrm, L. (2000), Rept. Prog. Phys. **63**, 793, arXiv:hep-ph/0002126 [hep-ph].
- [42] Bernabei, R., P. Belli, F. Cappella, R. Cerulli, F. Montecchia, *et al.* (2003), Riv.Nuovo Cim. **26N1**, 1, arXiv:astro-ph/0307403 [astro-ph].
- [43] Bernabei, R., P. Belli, F. Montecchia, W. Di Nicolantonio, A. Incicchitti, *et al.* (1998), Phys.Lett. **B424**, 195.
- [44] Bernabei, R., *et al.* (DAMA) (2008), Eur.Phys.J. **C56**, 333, arXiv:0804.2741 [astro-ph].
- [45] Bernabei, R., *et al.* (DAMA, LIBRA) (2010), Eur.Phys.J. **C67**, 39, arXiv:1002.1028 [astro-ph.GA].
- [46] Bertone, G. (2010), Nature **468**, 389, arXiv:1011.3532 [astro-ph.CO].
- [47] Bertone, G., M. Cirelli, A. Strumia, and M. Taoso (2009), JCAP **0903**, 009, arXiv:0811.3744

- [astro-ph].
- [48] Bertone, G., D. Hooper, and J. Silk (2005), Phys. Rept. **405**, 279, arXiv:hep-ph/0404175 [hep-ph].
- [49] Bertone, G., A. R. Zentner, and J. Silk (2005), Phys. Rev. **D72**, 103517, arXiv:astro-ph/0509565 [astro-ph].
- [50] Bertschinger, E. (1998), ARA&A **36**, 599.
- [51] Bessel, F. W. (1844), MNRAS **6**, 136.
- [52] Bethe, H. A. (1939), Physical Review **55**, 434.
- [53] Biviano, A. (2000), in *Constructing the Universe with Clusters of Galaxies*, p. 1, astro-ph/0010409.
- [54] Blumenthal, G. R., S. M. Faber, J. R. Primack, and M. J. Rees (1984), Nature (London) **311**, 517.
- [55] Boezio, M., P. Carlson, T. Francke, N. Weber, M. Suffert, *et al.* (2000), Astrophys.J. **532**, 653.
- [56] Boezio, M., *et al.* (WIZARD) (1997), Astrophys.J. **487**, 415.
- [57] Boezio, M., *et al.* (WIZARD/CAPRICE) (2001), Astrophys.J. **561**, 787, arXiv:astro-ph/0103513 [astro-ph].
- [58] Bond, J. R., G. Efstathiou, and J. Silk (1980), Physical Review Letters **45**, 1980.
- [59] Bond, J. R., and A. S. Szalay (1983), Astrophys. J. **274**, 443.
- [60] Bosma, A. (1978), *The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types*, Ph.D. thesis (PhD Thesis, Groningen Univ., (1978)).
- [61] Boyarsky, A., O. Ruchayskiy, D. Iakubovskiy, and J. Franse (2014), Phys.Rev.Lett. **113**, 251301, arXiv:1402.4119 [astro-ph.CO].
- [62] Brandt, F. C. (1965), MNRAS **129**, 309.
- [63] Brans, C., and R. H. Dicke (1961), Phys.Rev. **124**, 925.
- [64] Buffington, A., S. Schindler, and C. Pennypacker (1981), Astrophys.J. **248**, 1179.
- [65] Bulbul, E., M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein, *et al.* (2014), Astrophys.J. **789**, 13, arXiv:1402.2301 [astro-ph.CO].
- [66] Burbidge, E. M., G. R. Burbidge, W. A. Fowler, and F. Hoyle (1957), Reviews of Modern Physics **29**, 547.
- [67] Burbidge, G. R., and E. M. Burbidge (1959), Astrophys. J. **130**, 629.

- [68] Burles, S., K. M. Nollett, and M. S. Turner (2001), *ApJ* **552**, L1, astro-ph/0010171.
- [69] Burles, S., and D. Tytler (1998), *Astrophys. J.* **499**, 699, astro-ph/9712108.
- [70] Burles, S., and D. Tytler (1998), *Astrophys. J.* **507**, 732, astro-ph/9712109.
- [71] Cabibbo, N., G. R. Farrar, and L. Maiani (1981), *Phys. Lett.* **B105**, 155.
- [72] Caldwell, D. O., R. Eisberg, D. Grumm, M. S. Witherell, B. Sadoulet, *et al.* (1988), *Phys.Rev.Lett.* **61**, 510.
- [73] Carignan, C., L. Chemin, W. K. Huchtmeier, and F. J. Lockman (2006), *ApJ* **641**, L109, astro-ph/0603143.
- [74] Carr, B. J., and S. W. Hawking (1974), *MNRAS* **168**, 399.
- [75] Casertano, S., and J. van Gorkom (1991), *Astron.J.* **101**, 1231.
- [76] Catena, R., and P. Ullio (2010), *J. Cosmology Astropart. Phys.* **8**, 004, arXiv:0907.0018.
- [77] Centrella, J., and A. L. Melott (1983), *Nature (London)* **305**, 196.
- [78] Chang, K., and S. Refsdal (1979), *Nature (London)* **282**, 561.
- [79] Chernin, A. D. (1981), *Soviet Astronomy* **25**, 14.
- [80] Chwolson, O. (1924), *Astronomische Nachrichten* **221**, 329.
- [81] Clowe, D., M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky (2006), *ApJ* **648**, L109, astro-ph/0608407.
- [82] Cowsik, R., and J. McClelland (1972), *Phys.Rev.Lett.* **29**, 669.
- [83] Cushman, P., C. Galbiati, D. McKinsey, H. Robertson, T. Tait, *et al.* (2013), arXiv:1310.8327 [hep-ex].
- [84] Davis, M., G. Efstathiou, C. S. Frenk, and S. D. M. White (1985), *Astrophys. J.* **292**, 371.
- [85] Davis, M., J. Huchra, D. W. Latham, and J. Tonry (1982), *Astrophys. J.* **253**, 423.
- [86] Davis, R. L. (1986), *Phys.Lett.* **B180**, 225.
- [87] Daylan, T., D. P. Finkbeiner, D. Hooper, T. Linden, S. K. N. Portillo, *et al.* (2014), arXiv:1402.6703 [astro-ph.HE].
- [88] De Panfilis, S., A. Melissinos, B. Moskowitz, J. Rogers, Y. Semertzidis, *et al.* (1987), *Phys.Rev.Lett.* **59**, 839.
- [89] Dehnen, W., and J. Binney (1998), *MNRAS* **294**, 429, astro-ph/9612059.
- [90] Desai, S., *et al.* (Super-Kamiokande) (2004), *Phys.Rev.* **D70**, 083523, arXiv:hep-ex/0404025 [hep-ex].
- [91] Dicus, D. A., E. W. Kolb, and V. L. Teplitz (1977), *Phys.Rev.Lett.* **39**, 168.

- [92] Diemand, J., M. Kuhlen, and P. Madau (2007), *Astrophys. J.* **657**, 262, arXiv:astro-ph/0611370 [astro-ph].
- [93] Dimopoulos, S., and H. Georgi (1981), *Nucl.Phys.* **B193**, 150.
- [94] Dimopoulos, S., S. Raby, and F. Wilczek (1981), *Phys.Rev.* **D24**, 1681.
- [95] Dine, M., and W. Fischler (1983), *Phys.Lett.* **B120**, 137.
- [96] Dine, M., W. Fischler, and M. Srednicki (1981), *Phys. Lett.* **B104**, 199.
- [97] Dodelson, S., and M. Liguori (2006), *Phys.Rev.Lett.* **97**, 231301, arXiv:astro-ph/0608602 [astro-ph].
- [98] Dodelson, S., and L. M. Widrow (1994), *Phys.Rev.Lett.* **72**, 17, arXiv:hep-ph/9303287 [hep-ph].
- [99] Doroshkevich, A. G., Y. B. Zeldovich, R. A. Sunyaev, and M. Khlopov (1980), *Sov. Astron. Lett.* **6**, 252, [Pisma Astron. Zh.6,457(1980)].
- [100] Doroshkevich, A. G., Y. B. Zeldovich, R. A. Sunyaev, and M. Khlopov (1980), *Sov. Astron. Lett.* **6**, 257, [Pisma Astron. Zh.6,465(1980)].
- [101] Drukier, A., K. Freese, and D. Spergel (1986), *Phys.Rev.* **D33**, 3495.
- [102] Drukier, A., and L. Stodolsky (1984), *Phys.Rev.* **D30**, 2295.
- [103] E., O. (1915), *Bull. de la Soc. Astr. de Russie* **21**, 150.
- [104] Eddington, A. S. (1920), *The Observatory* **43**, 341.
- [105] Eddington, A. S. (1920), *Nature (London)* **106**, 14.
- [106] Efstathiou, G., and J. W. Eastwood (1981), *MNRAS* **194**, 503.
- [107] Einasto, J. (1972), *Tartu Astrofüüsika Observatoorium Teated* **40**, 3.
- [108] Einasto, J. (2014), *Dark Matter and Cosmic Web Story, Edited by Jaan Einasto. Published by World Scientific Publishing Co. Pte. Ltd., 2013. ISBN #9789814551052* (World Scientific Publishing Co).
- [109] Einasto, J., A. Kaasik, and E. Saar (1974), *Nature (London)* **250**, 309.
- [110] Einasto, J., E. Saar, A. Kaasik, and A. D. Chernin (1974), *Nature (London)* **252**, 111.
- [111] Einstein, A. (1936), *Science* **84**, 506.
- [112] Ellis, J., R. A. Flores, K. Freese, S. Ritz, D. Seckel, and J. Silk (1988), *Physics Letters B* **214**, 403.
- [113] Ellis, J. R., J. Hagelin, D. V. Nanopoulos, K. A. Olive, and M. Srednicki (1984), *Nucl.Phys.* **B238**, 453.

- [114] Ellis, J. R., J. S. Hagelin, D. V. Nanopoulos, and M. Srednicki (1983), Phys. Lett. **B127**, 233.
- [115] Faber, S. M., and J. S. Gallagher (1979), ARA&A **17**, 135.
- [116] Famaey, B., and S. McGaugh (2012), Living Rev.Rel. **15**, 10, arXiv:1112.3960 [astro-ph.CO].
- [117] Farrar, G. R., and P. Fayet (1978), Phys.Lett. **B76**, 575.
- [118] Fayet, P. (1975), Nucl.Phys. **B90**, 104.
- [119] Fayet, P. (1976), Phys. Lett. **B64**, 159.
- [120] Fayet, P. (1977), Phys. Lett. **B70**, 461.
- [121] Fayet, P. (1977), Phys. Lett. **B69**, 489.
- [122] Fich, M., L. Blitz, and A. A. Stark (1989), Astrophys. J. **342**, 272.
- [123] Fichtel, C. E., G. A. Simpson, and D. J. Thompson (1978), Astrophys.J. **222**, 833.
- [124] Field, G. B., and W. C. Saslaw (1971), Astrophys. J. **170**, 199.
- [125] Finzi, A. (1963), MNRAS **127**, 21.
- [126] Freeman, K. C. (1970), Astrophys. J. **160**, 811.
- [127] Freese, K. (1986), Phys.Lett. **B167**, 295.
- [128] Frenk, C. S., and S. D. M. White (2012), Annalen Phys. **524**, 507, arXiv:1210.0544 [astro-ph.CO].
- [129] Fukugita, M., C. J. Hogan, and P. J. E. Peebles (1998), Astrophys. J. **503**, 518, arXiv:astro-ph/9712020 [astro-ph].
- [130] Galli, S., F. Iocco, G. Bertone, and A. Melchiorri (2009), Phys. Rev. **D80**, 023505, arXiv:0905.0003 [astro-ph.CO].
- [131] Gamow, G. (1946), Physical Review **70**, 572.
- [132] Gershtein, S., and Y. a. Zeldovich (1966), JETP Lett. **4**, 120.
- [133] Gervais, J.-L., and B. Sakita (1971), Nucl.Phys. **B34**, 632.
- [134] Goldberg, H. (1983), Phys.Rev.Lett. **50**, 1419.
- [135] Golden, R., *et al.* (WIZARDS) (1996), Astrophys.J. **457**, L103.
- [136] Golfand, Y., and E. Likhtman (1971), JETP Lett. **13**, 323.
- [137] Gondolo, P., and G. Gelmini (1991), Nucl. Phys. **B360**, 145.
- [138] Gondolo, P., and J. Silk (1999), Phys. Rev. Lett. **83**, 1719, arXiv:astro-ph/9906391 [astro-ph].
- [139] Goodenough, L., and D. Hooper (2009), arXiv:0910.2998 [hep-ph].

- [140] Goodman, M. W., and E. Witten (1985), Phys.Rev. **D31**, 3059.
- [141] Gott, J. R., III, J. E. Gunn, D. N. Schramm, and B. M. Tinsley (1974), Astrophys. J. **194**, 543.
- [142] Gott, J. R., III, E. L. Turner, and S. J. Aarseth (1979), Astrophys. J. **234**, 13.
- [143] Griest, K., and D. Seckel (1991), Phys. Rev. **D43**, 3191.
- [144] Gunn, J., B. Lee, I. Lerche, D. Schramm, and G. Steigman (1978), Astrophys.J. **223**, 1015.
- [145] Guth, A. H. (1981), Phys. Rev. D **23**, 347.
- [146] Guth, A. H. (1981), Phys.Rev. **D23**, 347.
- [147] Guth, A. H., and S.-Y. Pi (1982), Physical Review Letters **49**, 1110.
- [148] Haag, R., J. T. Lopuszanski, and M. Sohnius (1975), Nucl.Phys. **B88**, 257.
- [149] Haggmann, C., P. Sikivie, N. Sullivan, and D. Tanner (1990), Phys.Rev. **D42**, 1297.
- [150] Haud, U., and J. Einasto (1983), in *Internal Kinematics and Dynamics of Galaxies*, IAU Symposium, Vol. 100, edited by E. Athanassoula, pp. 81–86.
- [151] Hawking, S. W. (1982), Physics Letters B **115**, 295.
- [152] Hegyi, D. J., and K. A. Olive (1983), Physics Letters B **126**, 28.
- [153] Hermsen, W., L. Blitz, and J. Bloemen (1994), Adv.Space Res. **3**, 83.
- [154] Hinshaw, G., D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R.olta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright (2013), ApJS **208**, 19, arXiv:1212.5226.
- [155] Hoekstra, H., M. Bartelmann, H. Dahle, H. Israel, M. Limousin, and M. Meneghetti (2013), Space Sci. Rev. **177**, 75, arXiv:1303.3274 [astro-ph.CO].
- [156] Hohl, F. (1971), Astrophys. J. **168**, 343.
- [157] Holmberg, E. (1937), Annals of the Observatory of Lund **6**, 3.
- [158] Holmberg, E. (1940), Astrophys. J. **92**, 200.
- [159] Hooper, D., P. Blasi, and P. D. Serpico (2009), JCAP **0901**, 025, arXiv:0810.1527 [astro-ph].
- [160] Hoyle, F. (1946), MNRAS **106**, 343.
- [161] Hoyle, F. (1954), ApJS **1**, 121.
- [162] Hubble, E., and M. L. Humason (1931), Astrophys. J. **74**, 43.
- [163] Hubble, E. P. (1926), Astrophys. J. **64**, 10.1086/143018.
- [164] Hut, P. (1977), Phys.Lett. **B69**, 85.

- [165] Ibanez, L. E., and G. G. Ross (1981), Phys.Lett. **B105**, 439.
- [166] Iocco, F., M. Pato, G. Bertone, and P. Jetzer (2011), JCAP **1111**, 029, arXiv:1107.5810 [astro-ph.GA].
- [167] Irwin, M. J., R. L. Webster, P. C. Hewett, R. T. Corrigan, and R. I. Jedrzejewski (1989), AJ **98**, 1989.
- [168] Jeans, J. H. (1922), MNRAS **82**, 122.
- [169] Jordan, P. (1955), *Schwerkraft und Weltall*, Vieweg (Braunschweig)..
- [170] Jungman, G., M. Kamionkowski, and K. Griest (1996), Phys. Rept. **267**, 195, arXiv:hep-ph/9506380 [hep-ph].
- [171] Kahn, F. D., and L. Woltjer (1959), Astrophys. J. **130**, 705.
- [172] Kamionkowski, M., and M. S. Turner (1991), Phys. Rev. **D43**, 1774.
- [173] Kelso, C., J. Collar, *et al.* (2014), Talk at TeV Particle Astrophysics/Identification of Dark Matter Conference, June 23-28, Amsterdam..
- [174] Kelvin, B. (1904), Baltimore lectures on molecular dynamics and the wave theory of light, <https://archive.org/details/baltimorelecture00kelviala>.
- [175] Kent, S. M. (1986), AJ **91**, 1301.
- [176] Kephart, T. W., and T. J. Weiler (1987), Phys.Rev.Lett. **58**, 171.
- [177] Kibble, T. (1976), J.Phys. **A9**, 1387.
- [178] Kim, H. I., C. H. Lee, and J. H. MacGibbon (1999), Phys. Rev. D **59** (6), 063004, astro-ph/9901030.
- [179] Kim, J. E. (1979), Physical Review Letters **43**, 103.
- [180] Klimov, Y. G. (1963), AZh **40**, 874.
- [181] Klimov, Y. G. (1963), Soviet Physics Doklady **8**, 119.
- [182] Klimov, Y. G. (1963), Soviet Physics Doklady **8**, 431.
- [183] Klypin, A. A., and S. F. Shandarin (1983), MNRAS **204**, 891.
- [184] Klypin, A. A., and S. F. Shandarin (1983), MNRAS **204**, 891.
- [185] Kolb, E. W., and R. Slansky (1984), Phys.Lett. **B135**, 378.
- [186] Kolkov, D. V., and V. P. Akulov (1972), Pisma Zh. Eksp. Teor. Fiz. **16**, 621..
- [187] Kragh, H. (2006), *Conceptions of Cosmos:From Myths to the Accelerating Universe* (OUP Oxford).
- [188] Krauss, L., J. Moody, F. Wilczek, and D. E. Morris (1985), Phys.Rev.Lett. **55**, 1797.

- [189] Krauss, L. M., K. Freese, W. Press, and D. Spergel (1985), *Astrophys. J.* **299**, 1001.
- [190] Krauss, L. M., M. Srednicki, and F. Wilczek (1986), *Phys.Rev.* **D33**, 2079.
- [191] Krumm, N., and E. E. Salpeter (1977), *A&A* **56**, 465.
- [192] Kuijken, K., and G. Gilmore (1989), *MNRAS* **239**, 605.
- [193] Lasserre, T., C. Afonso, J. N. Albert, J. Andersen, R. Ansari, É. Aubourg, P. Bareyre, F. Bauer, J. P. Beaulieu, G. Blanc, A. Bouquet, S. Char, X. Charlot, F. Couchot, C. Coutures, F. Derue, R. Ferlet, J. F. Glicenstein, B. Goldman, A. Gould, D. Graff, M. Gros, J. Haissinski, J. C. Hamilton, D. Hardin, J. de Kat, A. Kim, É. Lesquoy, C. Loup, C. Magneville, B. Mansoux, J. B. Marquette, É. Maurice, A. Milsztajn, M. Moniez, N. Palanque-Delabrouille, O. Perdereau, L. Prévot, N. Regnault, J. Rich, M. Spiro, A. Vidal-Madjar, L. Vigroux, S. Zylberajch, and EROS Collaboration (2000), *A&A* **355**, L39, astro-ph/0002253.
- [194] Lee, B. W., and S. Weinberg (1977), *Phys.Rev.Lett.* **39**, 165.
- [195] Liebes, S. (1964), *Physical Review* **133**, 835.
- [196] Limber, D. N. (1962), in *Problems of Extra-Galactic Research*, IAU Symposium, Vol. 15, edited by G. C. McVittie, p. 239.
- [197] Lindblad, (1926), *Uppsala Meddelanden* **11**, 30.
- [198] Linde, A. D. (1982), *Physics Letters B* **108**, 389.
- [199] Linde, A. D. (1982), *Phys.Lett.* **B108**, 389.
- [200] Linde, A. D. (1991), *Phys.Lett.* **B259**, 38.
- [201] Lubimov, V. A., E. G. Novikov, V. Z. Nozik, E. F. Tretyakov, and V. S. Kosik (1980), *Phys.Lett.* **B94**, 266.
- [202] Lundmark, K. (1925), *MNRAS* **85**, 865.
- [203] Lundmark, K. (1930), *Meddelande frn Lunds astronomiska Observatorium*.
- [204] Lynds, R., and V. Petrosian (1986), p. 1014.
- [205] MacGibbon, J. H. (1987), *Nature* **329**, 308.
- [206] Marciano, W. J., and G. Senjanovic (1982), *Phys.Rev.* **D25**, 3092.
- [207] Marx, G., and A. Szalay (1972), *Talk at Neutrino '72*, pg. 123, 1972, Budapest, Hungary..
- [208] McGaugh, S. S., and W. de Blok (1998), *Astrophys.J.* **499**, 41, arXiv:astro-ph/9801123 [astro-ph].
- [209] McGaugh, S. S., J. M. Schombert, G. D. Bothun, and W. de Blok (2000), *Astrophys.J.* **533**, L99, arXiv:astro-ph/0003001 [astro-ph].

- [210] Meekins, J. F., G. Fritz, T. A. Chubb, and H. Friedman (1971), *Nature (London)* **231**, 107.
- [211] Mellier, Y. (2010), “Gravitational lensing and dark matter,” in *Particle Dark Matter : Observations, Models and Searches*, edited by G. Bertone (Cambridge University Press) p. 56.
- [212] Merrifield, M. R. (1992), *AJ* **103**, 1552.
- [213] Michell, J. (1784), *Philosophical Transactions of the Royal Society of London Series I* **74**, 35.
- [214] Milgrom, M. (1983), *Astrophys.J.* **270**, 365.
- [215] Milgrom, M. (1983), *Astrophys.J.* **270**, 371.
- [216] Milgrom, M. (1983), *Astrophys.J.* **270**, 384.
- [217] Milgrom, M. (1986), *Astrophys.J.* **302**, 617.
- [218] Milgrom, M., and R. H. Sanders (2005), *Mon.Not.Roy.Astron.Soc.* **357**, 45, arXiv:astro-ph/0406487 [astro-ph].
- [219] Milgrom, M., and R. H. Sanders (2007), *Astrophys.J.* **658**, L17, arXiv:astro-ph/0611494 [astro-ph].
- [220] Miller, R. H., K. H. Prendergast, and W. J. Quirk (1970), *Astrophys. J.* **161**, 903.
- [221] Mueller, D., and K. K. Tang (1987), *Astrophys.J.* **312**, 183.
- [222] Navarro, J. F., C. S. Frenk, and S. D. M. White (1996), *Astrophys. J.* **462**, 563, astro-ph/9508025.
- [223] Nemiroff, R. J. (1987), *Prediction and analysis of basic gravitational microlensing phenomena*, Ph.D. thesis (Pennsylvania Univ., Philadelphia.).
- [224] Olive, K. A., and M. S. Turner (1982), *Phys. Rev.* **D25**, 213.
- [225] O’Meara, J. M., D. Tytler, D. Kirkman, N. Suzuki, J. X. Prochaska, D. Lubin, and A. M. Wolfe (2001), *Astrophys. J.* **552**, 718, astro-ph/0011179.
- [226] Oort, J. H. (1932), *Bull. Astron. Inst. Netherlands* **6**, 249.
- [227] Oort, J. H. (1960), *Bull. Astron. Inst. Netherlands* **15**, 45.
- [228] Ostriker, J. P., and P. J. E. Peebles (1973), *Astrophys. J.* **186**, 467.
- [229] Ostriker, J. P., P. J. E. Peebles, and A. Yahil (1974), *ApJ* **193**, L1.
- [230] Paczynski, B. (1986), *Astrophys. J.* **304**, 1.
- [231] Page, D. N., and S. W. Hawking (1976), *Astrophys. J.* **206**, 1.
- [232] Pagels, H., and J. R. Primack (1982), *Phys.Rev.Lett.* **48**, 223.
- [233] Pato, M., F. Iocco, and G. Bertone (2015), *JCAP* **1512** (12), 001, arXiv:1504.06324 [astro-ph.GA].

- [234] Peccei, R., and H. R. Quinn (1977), *Phys.Rev.* **D16**, 1791.
- [235] Peccei, R., and H. R. Quinn (1977), *Phys.Rev.Lett.* **38**, 1440.
- [236] Peebles, P. J. E. (1966), *Astrophys. J.* **146**, 542.
- [237] Peebles, P. J. E. (1970), *AJ* **75**, 13.
- [238] Peebles, P. J. E. (1971), *Princeton Series in Physics, Princeton, N.J.: Princeton University Press, 1971.*
- [239] Peebles, P. J. E. (1982), *ApJ* **263**, L1.
- [240] Peebles, P. J. E. (1982), *Astrophys. J.* **258**, 415.
- [241] Penzias, A. A. (1961), *AJ* **66**, 293.
- [242] Penzias, A. A., and R. W. Wilson (1965), *Astrophys.J.* **142**, 419.
- [243] Perryman, M. A. C., K. S. de Boer, G. Gilmore, E. Høg, M. G. Lattanzi, L. Lindegren, X. Luri, F. Mignard, O. Pace, and P. T. de Zeeuw (2001), *A&A* **369**, 339, astro-ph/0101235.
- [244] Pieri, L., G. Bertone, and E. Branchini (2008), *Mon. Not. Roy. Astron. Soc.* **384**, 1627, arXiv:0706.2101 [astro-ph].
- [245] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and et al. (2015), ArXiv e-prints arXiv:1502.01589.
- [246] Poincare, H. (1906), *L' Astronomie* , 158.
- [247] Poincare, H. (1906), *Popular Astronomy* **14**, 475.
- [248] Poincare, H., and H. Vergne (1911), *Paris, A. Hermann et fils, 1911.*
- [249] Preskill, J., M. B. Wise, and F. Wilczek (1983), *Phys.Lett.* **B120**, 127.
- [250] Press, W. H., and P. Schechter (1974), *Astrophys. J.* **187**, 425.
- [251] Press, W. H., and D. N. Spergel (1985), *Astrophys.J.* **296**, 679.
- [252] Ranyard, A. (1894), *Knowledge* **17**, 253.
- [253] Read, J. I. (2014), *Journal of Physics G Nuclear Physics* **41** (6), 063101, arXiv:1404.1938.
- [254] Reeves, H., J. Audouze, W. A. Fowler, and D. N. Schramm (1973), *Astrophys. J.* **179**, 909.
- [255] Refsdal, S. (1964), *MNRAS* **128**, 307.
- [256] Refsdal, S. (1964), *MNRAS* **128**, 295.
- [257] Roberts, M. S. (1966), *Astrophys. J.* **144**, 639.
- [258] Roberts, M. S. (1975), "Radio Observations of Neutral Hydrogen in Galaxies," in *Galaxies and the Universe*, edited by A. Sandage, M. Sandage, and J. Kristian (the University of

- Chicago Press) p. 309.
- [259] Roberts, M. S. (1975), in *Dynamics of the Solar Systems*, IAU Symposium, Vol. 69, edited by A. Hayli, p. 331.
- [260] Roberts, M. S., and A. H. Rots (1973), *A&A* **26**, 483.
- [261] Roberts, M. S., and R. N. Whitehurst (1975), *Astrophys. J.* **201**, 327.
- [262] Rogstad, D. H., and G. S. Shostak (1972), *Astrophys. J.* **176**, 315.
- [263] Rood, H. J. (1965), *The Dynamics of the Coma Cluster of Galaxies.*, Ph.D. thesis (The University of Michigan).
- [264] Rood, H. J., V. C. A. Rothman, and B. E. Turnrose (1970), *Astrophys. J.* **162**, 411.
- [265] Rood, H. J., II (1974), *Astrophys. J.* **188**, 451.
- [266] Rubin, V. (2004), “A brief history of dark matter,” in *The Dark Universe: Matter, Energy and Gravity*, edited by M. Livio, pp. 1–13.
- [267] Rubin, V. C., and W. K. Ford, Jr. (1970), *Astrophys. J.* **159**, 379.
- [268] Rubin, V. C., N. Thonnard, and W. K. Ford, Jr. (1978), *ApJ* **225**, L107.
- [269] Salam, A., and J. Strathdee (1975), *Nucl.Phys.* **B87**, 85.
- [270] Salpeter, E. E. (1978), in *Structure and Properties of Nearby Galaxies*, IAU Symposium, Vol. 77, edited by E. M. Berkhuijsen and R. Wielebinski, pp. 23–26.
- [271] Salucci, P., F. Nesti, G. Gentile, and C. Frigerio Martins (2010), *A&A* **523**, A83, arXiv:1003.3101.
- [272] Sandage, A., M. Sandage, and J. Kristian (1975), *Galaxies and the Universe*.
- [273] Sanders, R. (1996), *Astrophys.J.* **473**, 117, arXiv:astro-ph/9606089 [astro-ph].
- [274] Sanders, R. (1997), *Astrophys.J.* **480**, 492, arXiv:astro-ph/9612099 [astro-ph].
- [275] Sanders, R. (2003), *Mon.Not.Roy.Astron.Soc.* **342**, 901, arXiv:astro-ph/0212293 [astro-ph].
- [276] Sanders, R. (2007), *Mon.Not.Roy.Astron.Soc.* **380**, 331, arXiv:astro-ph/0703590 [ASTRO-PH].
- [277] Sanders, R. H. (2010), *The Dark Matter Problem: A Historical Perspective by Robert H. Sanders. Cambridge University Press, 2010. ISBN: 9780521113014.*
- [278] Sato, K., and M. Kobayashi (1977), *Prog.Theor.Phys.* **58**, 1775.
- [279] Schaye, J., R. A. Crain, R. G. Bower, M. Furlong, M. Schaller, T. Theuns, C. Dalla Vecchia, C. S. Frenk, I. G. McCarthy, J. C. Helly, A. Jenkins, Y. M. Rosas-Guevara, S. D. M. White, M. Baes, C. M. Booth, P. Camps, J. F. Navarro, Y. Qu, A. Rahmati, T. Sawala, P. A.

- Thomas, and J. Trayford (2015), MNRAS **446**, 521, arXiv:1407.7040.
- [280] Schmidt, M. (1957), Bull. Astron. Inst. Netherlands **14**, 17.
- [281] Schramm, D. N., and G. Steigman (1981), Astrophys. J. **243**, 1.
- [282] Schwarzschild, M. (1954), AJ **59**, 273.
- [283] Secchi, A. (1877), *L'Astronomia in Roma nel pontificato di Pio IX: memoria* (Tipografia della Pace, Roma).
- [284] Seljak, U., A. Slosar, and P. McDonald (2006), JCAP **0610**, 014, arXiv:astro-ph/0604335 [astro-ph].
- [285] Shi, X.-D., and G. M. Fuller (1999), Phys.Rev.Lett. **82**, 2832, arXiv:astro-ph/9810076 [astro-ph].
- [286] Shifman, M. A., A. I. Vainshtein, and V. I. Zakharov (1980), Nucl. Phys. **B166**, 493.
- [287] Shobbrook, R. R., and B. J. Robinson (1967), Australian Journal of Physics **20**, 131.
- [288] Sikivie, P. (ADMX) (1983), Phys.Rev.Lett. **51**, 1415.
- [289] Sikivie, P. (ADMX) (1985), Phys.Rev. **D32**, 2988.
- [290] Silk, J., K. A. Olive, and M. Srednicki (1985), Phys.Rev.Lett. **55**, 257.
- [291] Silk, J., and M. Srednicki (1984), Phys.Rev.Lett. **53**, 624.
- [292] Silverwood, H., S. Sivertsson, P. Steger, J. I. Read, and G. Bertone (2015), arXiv:1507.08581 [astro-ph.GA].
- [293] Skordis, C. (2008), Phys.Rev. **D77**, 123502, arXiv:0801.1985 [astro-ph].
- [294] Skordis, C. (2009), Class.Quant.Grav. **26**, 143001, arXiv:0903.3602 [astro-ph.CO].
- [295] Skordis, C., D. Mota, P. Ferreira, and C. Boehm (2006), Phys.Rev.Lett. **96**, 011301, arXiv:astro-ph/0505519 [astro-ph].
- [296] Slatyer, T. R., N. Padmanabhan, and D. P. Finkbeiner (2009), Phys. Rev. **D80**, 043526, arXiv:0906.1197 [astro-ph.CO].
- [297] Slipher, V. M. (1914), Lowell Observatory Bulletin **2**, 66.
- [298] Sofue, Y., M. Honma, and T. Omodaka (2009), PASJ **61**, 227, arXiv:0811.0859.
- [299] Soucail, G., B. Fort, Y. Mellier, and J. P. Picat (1987), A&A **172**, L14.
- [300] Srednicki, M., R. Watkins, and K. A. Olive (1988), Nucl. Phys. **B310**, 693.
- [301] Sreekumar, P., *et al.* (EGRET) (1998), Astrophys.J. **494**, 523.
- [302] Starobinsky, A. A. (1982), Physics Letters B **117**, 175.
- [303] Stecker, F. (1978), Astrophys.J. **223**, 1032.

- [304] Stecker, F. W., S. Rudaz, and T. F. Walsh (1985), *Physical Review Letters* **55**, 2622.
- [305] Steigman, G., and M. S. Turner (1985), *Nucl.Phys.* **B253**, 375.
- [306] Sunyaev, R., and Y. Zeldovich (1970), *Astrophys.Space Sci.* **7**, 3.
- [307] Tarter, J., and J. Silk (1974), *QJRAS* **15**, 122.
- [308] Tegmark, M., A. Aguirre, M. Rees, and F. Wilczek (2006), *Phys.Rev.* **D73**, 023505.
- [309] Tisserand, P., L. Le Guillou, C. Afonso, J. N. Albert, J. Andersen, R. Ansari, É. Aubourg, P. Bareyre, J. P. Beaulieu, X. Charlot, C. Coutures, R. Ferlet, P. Fouqué, J. F. Glicenstein, B. Goldman, A. Gould, D. Graff, M. Gros, J. Haissinski, C. Hamadache, J. de Kat, T. Lasserre, É. Lesquoy, C. Loup, C. Magneville, J. B. Marquette, É. Maurice, A. Maury, A. Milsztajn, M. Moniez, N. Palanque-Delabrouille, O. Perdureau, Y. R. Rahal, J. Rich, M. Spiro, A. Vidal-Madjar, L. Vigroux, S. Zylberajch, and EROS-2 Collaboration (2007), *A&A* **469**, 387, astro-ph/0607207.
- [310] Trimble, V. (2013), “History of Dark Matter in Galaxies,” in *Planets, Stars and Stellar Systems. Volume 5: Galactic Structure and Stellar Populations*, edited by T. D. Oswalt and G. Gilmore, p. 1091.
- [311] Tully, R., and J. Fisher (1977), *Astron.Astrophys.* **54**, 661.
- [312] Turner, E. L., S. J. Aarseth, N. T. Blanchard, R. D. Mathieu, and J. R. Gott, III (1979), *Astrophys. J.* **228**, 684.
- [313] Turner, M. S. (1987), *Phys.Rev.Lett.* **59**, 2489.
- [314] Turnrose, B. E., and H. J. Rood (1970), *Astrophys. J.* **159**, 773.
- [315] Valls-Gabaud, D. (2006), *AIP Conf. Proc.* **861**, 1163, arXiv:1206.1165 [physics.hist-ph].
- [316] van de Hulst, H. C., E. Raimond, and H. van Woerden (1957), *Bull. Astron. Inst. Netherlands* **14**, 1.
- [317] van den Bergh, S. (1969), *Nature (London)* **224**, 891.
- [318] van Woerden, H., and R. G. Strom (2006), *Journal of Astronomical History and Heritage* **9**, 3.
- [319] Viel, M., J. Lesgourgues, M. G. Haehnelt, S. Matarrese, and A. Riotto (2005), *Phys.Rev.* **D71**, 063534.
- [320] von Hoerner, S. (1960), *ZAp* **50**, 184.
- [321] von Hoerner, S. (1963), *ZAp* **57**, 47.
- [322] Vysotsky, M., A. Dolgov, and Y. Zeldovich (1977), *JETP Lett.* **26**, 188.

- [323] Wagoner, R. V., W. A. Fowler, and F. Hoyle (1967), *Astrophys. J.* **148**, 3.
- [324] Walsh, D., R. F. Carswell, and R. J. Weymann (1979), *Nature (London)* **279**, 381.
- [325] Wasserman, I. (1986), *Phys.Rev.* **D33**, 2071.
- [326] Weber, M., and W. de Boer (2010), *A&A* **509**, A25, arXiv:0910.4272 [astro-ph.CO].
- [327] Weinberg, S. (1978), *Phys.Rev.Lett.* **40**, 223.
- [328] Weinberg, S. (1983), *Phys.Rev.Lett.* **50**, 387.
- [329] Wesley, W. H. (1894), *Journal of the British Astronomical Association* **5**, 45.
- [330] Wess, J., and B. Zumino (1974), *Nucl.Phys.* **B70**, 39.
- [331] White, S. D. M. (1976), *MNRAS* **177**, 717.
- [332] White, S. D. M. (1977), *MNRAS* **179**, 33.
- [333] White, S. D. M., C. S. Frenk, and M. Davis (1983), *ApJ* **274**, L1.
- [334] White, S. D. M., J. F. Navarro, A. E. Evrard, and C. S. Frenk (1993), *Nature (London)* **366**, 429.
- [335] Whitehurst, R. N., and M. S. Roberts (1972), *Astrophys. J.* **175**, 347.
- [336] Wilczek, F. (1978), *Phys.Rev.Lett.* **40**, 279.
- [337] Wilczek, F. (2004), In *Carr, Bernard (ed.): Universe or multiverse* 151-162. arXiv:hep-ph/0408167 [hep-ph].
- [338] Wilkinson, M. I., A. Vallenari, C. Turon, U. Munari, D. Katz, G. Bono, M. Cropper, A. Helmi, N. Robichon, F. Thévenin, S. Vidrih, T. Zwitter, F. Arenou, M.-O. Baylac, G. Bertelli, A. Bijaoui, F. Boschi, F. Castelli, F. Crifo, M. David, A. Gomboc, A. Gómez, M. Haywood, U. Jauregi, P. de Laverny, Y. Lebreton, P. Marrese, T. Marsh, S. Mignot, D. Morin, S. Pasetto, M. Perryman, A. Prša, A. Recio-Blanco, F. Royer, A. Sellier, A. Siviero, R. Sordo, C. Soubiran, L. Tomasella, and Y. Viala (2005), *MNRAS* **359**, 1306, astro-ph/0506083.
- [339] Witten, E. (1984), *Phys.Rev.* **D30**, 272.
- [340] Wolf, M. (1914), *Vierteljahres schr. Astron. Ges.* **49**, 162.
- [341] Woolf, N. J. (1967), *Astrophys. J.* **148**, 287.
- [342] Wuensch, W., S. De Panfilis-Wuensch, Y. Semertzidis, J. Rogers, A. Melissinos, *et al.* (1989), *Phys.Rev.* **D40**, 3153.
- [343] Yokoyama, J. (1998), *Phys. Rev. D* **58** (10), 107502, gr-qc/9804041.
- [344] Zeldovich, Ya. B., A. A. Klypin, M. Yu. Khlopov, and V. M. Chechetkin (1980), *Sov. J. Nucl. Phys.* **31**, 664, [Yad. Fiz.31,1286(1980)].

- [345] Zhao, H.-S., and J. Silk (2005), *Phys. Rev. Lett.* **95**, 011301, arXiv:astro-ph/0501625 [astro-ph].
- [346] Zwicky, F. (1933), *Helvetica Physica Acta* **6**, 110.
- [347] Zwicky, F. (1933), *Phys. Rev.* **43**, 147.
- [348] Zwicky, F. (1937), *Astrophys. J.* **86**, 217.