Initial Conditions for Cosmic Inflation, the History of the Dark Sector, and Dark-onium

by

Patrick John Fitzpatrick

B.A., University of California, Berkeley (2013)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

Author .....
Department of Physics
August 31, 2021

Certified by .....
Tracy R. Slatyer
Associate Professor of Physics
Thesis Supervisor

Certified by .....
David I. Kaiser
Professor of Physics
Germeshausen Professor of the History of Science
Thesis Supervisor

Accepted by .....
Deepto Chakrabarty
Associate Department Head of Physics

Initial Conditions for Cosmic Inflation, the History of the Dark Sector, and Dark-onium

by

Patrick John Fitzpatrick

Submitted to the Department of Physics
on August 31, 2021, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The question remains whether inflation is robust to inhomogeneous initial conditions. This thesis first describes the basics of cosmic inflation and the evolution of primordial inhomogeneities in the matter field and spacetime metric during inflation. Then a new approach for analyzing the onset of inflation amid backreaction from significant inhomogeneities is presented. This new approach incorporates certain nonlinear interactions among the coupled degrees of freedom by using the nonperturbative Hartree approximation. Results applying this approach to a single-field inflationary model find inflation to be robust for large-field models.

The particle nature of dark matter is still a mystery. This thesis very briefly summarizes what we know about dark matter and our current efforts to detect it. This thesis also provides the basic tools necessary to calculate the thermal history of the dark sector. Then the results for a full exploration of the thermal freezeout histories of a vector-portal dark matter model, in the region of parameter space in which the ratio of masses of the dark photon A' and dark matter χ is in the range $1 \lesssim m_{A'}/m_\chi \lesssim 2$, are presented. The temperatures of all species are carefully tracked, relaxing the assumption of previous studies that the dark and Standard Model sectors remain in thermal equilibrium throughout dark matter freezeout. A rich set of novel pathways which lead to the observed relic density of the dark matter is revealed. This thesis also examines the $1 \lesssim m_{A'}/m_\chi \lesssim 2$ regime of the vector-portal inelastic dark matter model, where the dark matter is made up of a Majorana ground state χ and excited state χ^* with a small mass splitting between them, carefully tracking the dark sector temperature throughout freezeout. The inelastic nature of the dark sector relaxes stringent cosmic microwave background and self-interaction constraints compared to symmetric dark matter models.

The spectrum of Weakly-Interacting-Massive-Particle (WIMP) dark matter generically possesses bound states when the WIMP mass becomes sufficiently large relative to the mass of electroweak gauge bosons. After a review of the treatment of bound

states in quantum electrodynamics, this thesis examines the formation and decay of bound states for dark matter inhabiting a more general nonabelian dark sector. The rate for SU(2) triplet dark matter (the wino) to bind into WIMPonium, and rates for the subsequent decays of these bound states, are computed. Results with applications beyond the wino case, e.g. for dark matter inhabiting a nonabelian dark sector, are also presented.

Thesis Supervisor: Tracy R. Slatyer
Title: Associate Professor of Physics

Thesis Supervisor: David I. Kaiser
Title: Professor of Physics
Germeshausen Professor of the History of Science

Acknowledgments

I only ever read the Acknowledgements sections of dissertations. The rest is too long to start and has no well defined audience. I expect a few people in particular might disagree with that last statement, after all in order for this thesis to be approved it had to be accepted by my thesis committee, Tracy Slatyer, David Kaiser, Jesse Thaler, and Mark Vogelsberger. That's true, but Tracy and Dave had already read many of the chapters of this dissertation many times over before I even began writing it, and Jesse had probably already read at least the abstracts of those chapters. Also, I was secretly expecting my thesis committee to more-or-less skim this dissertation and get just enough material to ask a couple difficult questions at my thesis defense. So it surprised me to receive two long lists of detailed comments on the first draft of this thesis from Dave and Tracy.

I want to thank my PhD advisors Tracy and Dave, for bringing me into the CTP and helping me grow from being a student at the start of grad school only just curious about switching from computational and observational astronomy to high energy theory, to a theoretical physicist capable of the work that appears in this dissertation. Tracy and Dave have both been incredibly generous with their time, which is limited in a way that I am only beginning to understand as life in physics unfortunately becomes increasingly professional. I learn so much from both of them every time I enter their offices and ask a question. Tracy helped me rehearse every research talk I've ever given, and sat in my Part III exam to keep me from getting too nervous. I've met few people who think about physics with the same energy, speed, creativity, and brightness that she does. She seems to be in a million places at one time, yet always present and always clear and brilliant. Sometimes when we're stuck on a research problem, I'll take an hour nap, wake up, open my computer and find an email from Dave with a ten page latex document carefully solving everything, and written well enough to appear in a textbook. That's barely an exaggeration. I don't know if Dave notices this but sometimes when we're working on a problem in his office and he's at the whiteboard working through a mess of equations, I get lost

looking at the titles on the surrounding walls of books. As academic work becomes increasingly specialized it's always been inspiring to look up and see Kant's *Critique of Pure Reason*, Kuhn's *The Structure of Scientific Revolutions*, and (many copies of) Misner, Thorne, and Wheeler's *Gravitation* all on the same shelf.

I want to thank all of my friends in my research group, Hongwan Liu, Lina Necib, Nick Rodd, Chih-Liang Liu, Wenzer Qin, Yitian Sun, and Gregory Ridgeway. All of you have been graduate students I've looked up to who set a standard for me to grow towards. We wouldn't have finished much of the work in this thesis if it wasn't for Hongwan. His ability to turn a mess of numerical results into an analytical explanation is amazing. And he's been a friend who every time I've needed help or advice throughout grad school has been available. I remember when Lina used to walk around the CTP in baggy cargo pants and huge DC shoes. Now she's one of the most important pioneers in our field, a scientist who has fully bridged the gap between theoretical particle physics and observational astronomy. She's always been a friend I could go to when I needed any kind of advice. When we all went to Berlin for the TeVPA conference I barely made it, having been robbed of all my money a week earlier in Spain. I arrived at 1 or 2 am and called everyone hoping someone was still awake. Nick woke up and let me stay in his hotel room. I showed up in the same clothes I'd been wearing for a few days eating a kebab that was apparently so smelly the hotel clerk pinched his nose and asked me to step outside while I waited (it was a nice hotel I guess). Then I asked Nick to borrow a few hundred dollars, and he didn't hesitate to say yes. Those days I was studying for Part III so I would always stay up a little later than Nick, too, studying with a dim light at a table right next to his bed. He never complained to me once, he made me feel welcome, and he helped me rehearse my talk for the conference. Now Nick is also one of the most important pioneers in our field. Chih-Liang and I suffered through the worst moments of Part III together, and along the way he was never too busy to help explain something to me. Whenever I would get disillusioned by the academic egos you can sometimes encounter in physics I would take inspiration from Chih-Liang, who is exceptionally talented and creative and doesn't seem to worry about whether you know it. Wenzer

is always relaxed and smiling, and I don't know how she manages that, doing so much diverse and interesting physics at the same time. The first time I met Yitian Sun was before he joined Tracy's group, I think we were working on a problem set for a holography class that I later dropped. We worked on the problem set all night until like 4am and he somehow didn't seem to get tired, constantly coming up with creative ways of solving the problems, which I mostly nodded my head pretending to understand. Greg is one of the most talented physicists I've ever met. The diversity of topics he has deep enough knowledge to come up with new ideas about continues to surprise me. It's an honor to have started and killed a physics project with him. And I learned how to be more social at physics conferences from Greg. That might seem trivial to a lot of people, but socializing in a large group of people who spend approximately all of their time either doing physics, sitting in meetings about physics, or answering emails about physics, and who may be reading your job applications one day, can sometimes be tricky. The more crude Greg's jokes get the more scared I am they're going to offend someone important, but everyone seems to be charmed by it. And then I make one joke about stealing bikes and everyone shuts up.

I also have many friends at and around MIT to thank. When we were working on stat mech problem sets together my good friend Luiz Gustavo Pimenta Martins would let me sleep on his couch so much I really should've been paying him rent. We would wake up and start doing stat mech again, using the pomodoro technique. We still need to jump in the Charles. Sarah Geller is the most inspiring person who loves physics more than anyone I know. Stella Schindler is a force of nature. Ever since she showed up to the CTP she's caused chaos, and the whole department is learning how to function better because of it. She gave me a lottery ticket as a going away present when I cleaned out my desk, I was one number away from winning some money from it, and she's always been there when I needed advice from a courageous person. Yu-Chien Huang always made the CTP a joyful place to be. She taught me QFT III and gave beautiful lectures about mathematical aspects of the Standard Model. Jasmine Brewer is the first person I met in the CTP. We learned QFT and studied for Part III together. She played drums in a physics rock and roll band and

is one of the most talented young physicists in her field. I wouldn't have made it past Part III if it wasn't for Nikhil Raghuram always being available to help me when I needed it. Twice he talked the same angry guy out of beating me up. Andrew Patrick Turner was the second person I met in the CTP, tied with Hongwan. I also wouldn't have made it past Part III if it wasn't for his help. I definitely wouldn't have made it past Part III if it wasn't for Gherardo Vita trying to get me to understand low-energy QCD and yelling at me at the chalkboard in Italian. Anthony Grebe is another good friend who is always around when I need some help or advice. He held the D-fence sign at my thesis defense. If not for him I would've missed a few holy days of obligation. Josué Lopez is one of the most inspiring people I've met at MIT. I keep vaguely promising him I'll give him one of my bikes but I haven't made the promise concrete yet. Michael DeMarco lets me into the beer socials without my ID sometimes. I think of Mason Biamonte as a físico maldito, although I don't know how he'd feel about that. Once when we were skating Mason told me he'd found a postdoc in the math department who was going to come skate with us. That was the first time I met Cris Negron. He did a kickflip frontside boardslide on the parking block ledge I'd built behind Walker Memorial first try. I miss skating with both of you. I miss skating with Chris, Serg, Stan, Adam, BDougs, and everyone at fountain. Sometimes I wonder if the reason Ian Moult is so good at physics is he has to justify the cost of having left a career in BMX, and whether that thought process is subconscious or not. Sometimes I wonder something similar about Brian Dolle. Pavel Chykov is as courageous as he is brilliant. He gives himself the freedom to let his curiosity take him wherever, and he seems not to have any sense of fear about getting stuck in the forest and never coming out. Somehow he always manages to come out with some beautiful idea. David Theurel is also brilliant and courageous. Every time I talk to David I remember the importance of thinking and researching. It may seem funny, but in the academy it can be easy to forget that, because the force of professional survival can sometimes suffocate curiosity, and almost always at least restrains it. David always seems to be committed to solving the problems about nature that are close to his heart, no matter what else. I first met Andrey Sadofyev drinking vodka and eating

pancakes at Pavel's house, and Nick was there, too. I had just arrived at MIT and I was complaining about how opaque I found Peskin and Schroeder. Andrey laughed and proclaimed that he had learned QFT from Landau and Lifshitz. He also said not to tell him my surname because he was grading my QFT problem sets. Andrey taught me about one-loop radiative corrections drawing loop diagrams with his finger in the dust on the surface of one of the picture frames that show I forgot which MIT professor receiving the nobel prize or something on a wall at one of the beer socials. After Andrey defended his thesis we got kicked out of the MIT sailing pavilion and then went to a bar. Then I woke up in the grad student lounge with cuts all over my legs and a swollen jaw, and was relieved to find my Leica M2 still hanging from my shoulder. I have a roll of film from that night that I am still waiting to develop.

I have many friends from my time before MIT to thank, who changed my path in a way that led me to be doing physics this long (for better or worse I don't know). Adam Becker, too much to say, so I'll just bring up a trivial early memory that for some reason stuck with me, shortly after we had first met each other, standing on Telegraph and Durant I think, I was eating a custard filled chocolate donut from Kingpin and we were arguing about something and you had me convinced you were a lapsed Thomist. Actually you still do. Pierre Christian, too much to say, so I'll just say that you're my favorite physicist. For being the most creative and the most delirious while somehow still productive physicist. Another trivial memory is making wacky mack at my place and drinking a bottle of sweet wine from a failed relationship and you not being able to walk Cinderella back home. Alwin Mao, you're actually tied with Pierre as my favorite physicist, for being the most creative and most delirious, although I have to be careful who I publicly confess that to. I first met Francisco Antonio in the physics reading room at Berkeley, he wasn't in my class but I think I randomly asked him a quantum mechanics question while he was staring at a circuit or something. A few months later and he's waking me up every day at 7am to go to Garber for daily Mass. Anthony Paredes helped me get my first research project with Genevieve Graves. We both became astrophysicists together. When I first arrived at Harvard for grad school (before leaving) he was living on the floor of my dorm. If it

wasn't for Philip Macias I would not have transferred to UC Berkeley, and I would not have studied physics and astronomy, and I would have missed out on all this. I met Phil when we were students at Cal State Fullerton in a physics class. I was studying mechanical engineering, but mostly I would just turn in my problem sets as soon as I could and go to the library with a bag of gummy worms and read. Even though I was reading courageous people like Nicanor Parra, Bolaño[1], Joyce, Rulfo, Augustine, Aquinas, Ginsberg...it was taking me a while to get the courage to admit that I didn't like where I was going. One day in physics class Phil told me he was going to apply to transfer to a UC to study physics and that I should too. The idea was planted and pretty soon it became an obsession, and knowing Phil was doing it too gave me the courage to do it and the confidence it was possible. The following year Phil was at UC Santa Barbara and I was at UC Berkeley. Three (or four?) years later Phil was the grad student who hosted my prospective grad student visit at UC Santa Cruz. He still hasn't beat me in a game of skate, despite the fact that I only have three consistent tricks.

I have too many important mentors to thank, so I'll just say a few things about the ones that come to mind. Genevieve Graves is the first person who had the patience to work with me on a scientific research project when I didn't have any experience, and I'd been rejected so many times I was about ready to give up. During our first meeting she taught me to read data files into Python and then eventually we wrote a paper together. It is still one of the papers I am most proud of. Genevieve taught me how to do scientific research, and is one of the most supportive people I've ever worked with. Many years later I still call her asking for advice. After I learned some research skills from Genevieve I started working with Imke de Pater. Imke gave me confidence as a researcher, and opened the doors for me. She even sent me to Hawaii to observe on Keck. I wrote a great paper with Imke, too, and when we were done with it I was amazed at how much I had grown by working with her. Imke is probably the most important person who guided me to pursue a PhD. After I graduated UC Berkeley I stayed for a year and worked with Imke de Pater

[1] "hasta las cabezas de los aristócratas nos pueden servir de armas"

and Daniel Kasen. Dan taught me how to start thinking about research problems like a theorist. Besides being one of my favorite astrophysicists he is a great example of humility and kindness. Enrico Ramirez-Ruiz has been another great mentor. An example of a great astrophysicist who is dedicated to his students. Someone I can always reach out to for advice. In fact I might wish I would have followed his advice on a few occasions when I didn't. Karl van Bibber is another great friend and mentor of mine, who has always been someone I can reach out to ever since I first arrived at Berkeley. I miss everyone at Garber house. When I was studying for Part III I would show up to Bob Jaffe's office every other day with a new question about the Standard Model, and he would never turn me away, usually spending an hour at a time carefully explaining something to me. The Part III exam that Jesse gave me before my Part III exam was way more difficult than my actual Part III exam. Peg Herlihy has spent countless hours helping me through stressful decisions and events in grad school with kindness and grace no matter how busy her day gets, and I'm not even a student at her university. When I left Harvard she took me to lunch. Most of the time I wait three years past the deadline to cash travel reimbursement checks and by then they're void, I'll find them when I'm shuffling through the stacks of paper on my desk looking for an old calculation, I take them to Scott Morley to get them reissued and he never scolds me about it or even asks why I waited three years, he just laughs and then a week later money appears in my bank account. He has the best style in the department. You hear stories about how you meet your favorite rock and roll idol and it turns out they aren't as cool as you hoped, like Van Morrison refuses to shake your hand or denies you an autograph or something. I'm thankful to know that the guy responsible for a paradigm shift in cosmology is also cool, humble and caring to his students. Thanks also, Alan, for enlightening discussions, answers to my questions, rides to seminars, and lots of pizza at group meetings. In the winter of 2019 I was in my office in the process of booking a plane ticket to DF when Greg appeared in my office to tell a crude joke. Then he mentioned he was booking his ticket to that winter school in Jerusalem he assumed I knew about, but I had missed the email. I looked up the winter school and a lot of the usual suspects in dark matter

physics were going to be there, so I apprehensively decided to make a compromise, I would fly to DF for only three days and then fly to Jerusalem. In Mexico it was the time of Guadalupe-Reyes, so it was a small Christmas miracle that I ever made it to Jerusalem. At the winter school I got to see Yotam Soreq again after a year or two had passed since his postdoc days in the CTP. I also got to meet a lot of other great dark matter physicists including Eric Kuflik and Yonit Hochberg. One morning I almost got arrested in the Church of the Holy Sepulchre. I woke up early before lectures started and went walking around the Old City looking for a Mass to attend. At an Armenian Catholic church a guy told me to come back in 30 minutes and there would be Mass, which was perfect timing to make it to lectures after. 30 minutes later I returned and he said everyone decided to start an hour later. I wouldn't make it to lectures on time if I waited another hour so after a little more searching I finally found a Divine Liturgy with the Greek Orthodox community in the Church of the Holy Sepulchre. I've only been to a few Divine Liturgies in my life but this was definitely the only one that ever had a bouncer. Just after the Great Litany this huge guy in intimidating black vestments kicked out an old couple who were speaking English really loudly and taking pictures with an oversized DSLR camera, then came up to me and asked if I was Greek Orthodox. When I said no he told me to leave. When I refused he started pushing me out. When I refused to leave still he told me he was going to get the police, which I thought was funny. He disappeared somewhere for a while and I didn't think much of it, then just before the Epistle out of nowhere an enormous force pulled me from behind and knocked me on my back, I look up and see two police officers dragging me across the floor from my shirt collar out of the chapel. Once we were outside the chapel one of the officers pulled out handcuffs, and that's when I gave up on any hope of making it to lectures that morning. But then luckily a bunch of tourists with blank expressions of curiosity started circling around me and the officers, and nearby a line of hundreds more tourists who were queuing up to see the tomb of Jesus were suddenly staring at us, which made the officers pause. They hesitated, looked around, and then just walked away. I got up, dusted myself off, and sat down on a bench just outside the chapel considering whether I would try to go

back in or not. I asked a monk who was sitting next to me what he thought about the whole thing and he just said something like 'sometimes people like to think they are God, let me show you another way'. He took me up some stairs to a balcony where I realized that the chapel I'd been dragged out of had no ceiling, I could celebrate the Divine Liturgy from the top. The bouncer noticed me up there and he started taking photos of me with his phone and showing them to what I would call altar boys, they were all laughing together. I thought about what the monk said and tried my best to ignore it. At the end of the Divine Liturgy I stood outside of the chapel to see the procession leave, when the monk passed by he gave me a wink and a nod, and when the bouncer passed by he shouldered me, then signaled for me to meet him outside, but then he never showed up. Anyway, about a week later I was visiting Technion, where I got to spend time with Yotam and Yael Shadmi. Yotam even arranged a hotel for me and let me use his office. Then we all joined up with everyone at I forgot which university for a joint particle physics meeting. From there I was returning to Jerusalem on the train with Eric, when suddenly the train stopped and started reversing back, everyone was saying someone had fallen into the tracks. We spent the next few hours on a series of buses which random walked back to Jerusalem. On one of those bus rides Eric asked me if it was true I'd gotten into a fistfight with a priest. I guess word spreads and changes quickly among Israeli physicists. It might have worried me to have everyone thinking I'd beaten up a priest, but Eric didn't seem to give any impression of negative judgement when he asked, on the contrary, it was almost like he was hoping for me to say yes and give him all the details. Just in case that rumor is still out there, this is my way of setting the record straight. The following week I had great discussions with Eric and Yonit who helped me realize some critical things about the work on freezeout mechanisms for vector-portal dark matter in this thesis. I am looking forward to seeing all of you soon.

Alexia Cassandra Macias Cadena, eres la persona mas guerrera que he conocido. Pequeños recuerdos, platicando tomando cafe en el sunburns de la alameda a las 5 de la mañana esperando el primer metrobus que nos lleve a la casa abandonada, tomando atole escuchando los saicos, nos gusta volar estaciones del tren.

To my mother Ofelia Fitzpatrick, especially for being with me during the long recovery of the last year and a half, and sharing some of the suffering. My injury will be the most outstanding memory I have when I think of my time as a PhD student, and the experience I'm learning the most from, que solo hay una forma de ser libre, ..., it will eventually be the most fruitful. I am leaving a record of it here so that the only option for the next months, years, whatever it takes, is to eventually look back at this paragraph with gratitude.

Contents

Chapter 1

Introduction

Cosmic inflation and dark matter are two mysteries of cosmology which unite the study of nature on length scales ranging from the sizes of elementary particles to the size of the observable universe, and over cosmic history from the earliest moments after the Big Bang to the present day. Both mysteries invite us to consider theories beyond the Standard Model (SM) of particle physics.

In this thesis I will discuss a few ways in which we push beyond the minimal cases of these theories. I will start by discussing the so-called "initial conditions problem" for cosmic inflation – the possible fine-tuning of the initial conditions required to allow for a period of cosmic inflation to occur in the very early universe. I will present an efficient numerical framework to investigate the robustness of cosmic inflation to inhomogeneous initial conditions in the very early universe, and apply this framework to a specific inflationary model.

I will then discuss ways we can push beyond the conventional weakly-interacting-massive-particle (WIMP) paradigm in which the late-time density of dark matter (DM) is set by the rate of two-body annihilations, and present a few alternative DM scenarios where other interactions control the final DM abundance. The alternative DM scenarios I will present naturally produce DM at light (sub-GeV) masses, where direct detection experiments are least sensitive. In particular, I will show that fully exploring the parameter space of a common benchmark DM model, the vector or kinetic-mixing portal DM model, while relaxing the assumption of efficient thermal

contact between the dark sector and the SM which had been made in previous studies, reveals a rich set of new mechanisms for obtaining the correct DM relic density, which lead to interesting cosmological histories and open new windows of parameter space for the DM where experiments have not yet explored.

Finally I will discuss the formation, properties, and phenomenological consequences of DM-DM bound states. In particular I will discuss the formation and decay of bound states of a specific WIMP called the wino, which is an $SU(2)_L$ triplet Majorana fermion with zero hypercharge.

The structure of this dissertation is as follows. In Chapter 2 I start by outlining the basics of Friedmann-Lemaître-Robertson-Walker (FLRW) spacetime and of cosmic inflation in Sections 2.2 and 2.3, respectively. In particular, I describe how inflation solves the flatness and horizon problems of standard cosmology. Then in Sections 2.6 and 2.7 I describe the basic formalism for describing the evolution of linearized quantum fluctuations in the matter field and spacetime metric during inflation, and give a basic description of how primordial quantum fluctuations evolve during inflation. In Section 2.8 I give only a very brief description of how the primordial density perturbations relate to the temperature anisotropies we observe on the CMB. At the end of Chapter 2 I introduce a possible remaining fine-tuning problem left by cosmic inflation, the initial-conditions problem. The overall goal of Chapter 2 is to give an introduction to the basic formalism of inflation and the treatment of primordial fluctuations during inflation to set up the foundation for understanding Chapter 3. The goal of Chapter 2 is also to illustrate the insufficiency of the linearized formalism for treating the evolution of primordial inhomogeneities to study the initial conditions problem, motivating a more sophisticated formalism like the one I present in Chapter 3.

In Chapter 3 I present a framework for numerically solving the equations of motion which describe the evolution of inhomogeneities in the early universe, in a way that is much more efficient than simulations involving full numerical general relativity. This approach captures the dominant nonlinear effect of inhomogeneities influencing the surrounding spacetime, and vice-versa. I present calculations which address to what

extent inflation can occur under inhomogeneous initial conditions in the very early universe for a specific inflationary model, and discuss application of this approach to a wider range of inflationary models. Chapter 3 contains our paper Ref. [1].

In Chapter 4 I start by giving a very brief introduction to ways in which we infer the existence of dark matter, the kinds of interactions we might expect it to have with SM particles, and the associated ways we might hope to detect DM. Then in Section 4.2 I begin describing the thermal history of the dark sector under the assumption that at some point in the early universe its interactions with the SM were strong enough to establish thermal and chemical equilibrium between the dark and SM sectors. I start this discussion by introducing the Boltzmann equations, the basic tool we need to describe the thermal evolution of the dark and SM sectors. Then in Section 4.3 I move on to describing DM production mechanisms, starting with the simplest and most-well known example of the thermal WIMP. In Section 4.3.2 I describe the motivation to move beyond the thermal WIMP, and in particular to explore the landscape of new models which naturally produce DM at light (sub-GeV) masses. In Sections 4.3.3 and 4.3.4 I review two specific DM production mechanisms alternative to the thermal WIMP, the strongly-interacting-massive-particle (SIMP) and elastically decoupling relic (ELDER) DM models. The overall goal of Chapter 4 is to provide the reader with the basic tools to calculate the thermal history of a dark sector given its structure and interactions, and to provide the groundwork for understanding Chapters 5 and 6.

In Chapter 5 I fully explore the thermal freezeout histories of a vector-portal DM model, in the region of parameter space in which the ratio of masses of the dark photon A' and the dark matter m_χ is in the range $1 \lesssim m_{A'}/m_\chi \lesssim 2$. In this region $2 \to 2$ and $3 \to 2$ annihilation processes within the dark sector, as well as processes that transfer energy between the dark sector and the SM, play important roles in controlling the thermal freezeout of the dark matter. I show the results of calculations which carefully track the temperatures of all species, relaxing the assumption of previous studies that the dark and SM sectors remain in thermal equilibrium throughout dark matter freezeout. Our calculations reveal a rich set of novel pathways which lead to the

observed relic density of DM, and we develop a simple analytic understanding of these different regimes. The viable parameter space in this model provides a target for future experiments searching for light (MeV-GeV) DM, and includes regions where the DM self-interaction cross section is large enough to affect the small-scale structure of galaxies. Chapter 5 contains our paper Ref. [2].

In Chapter 6 I return to the vector-portal DM model, but modify the baseline scenario presented in Chapter 5 to consider the case of *inelastic* DM, in which the DM is a pseudo-Dirac fermion which at low energies splits into two nearly-degenerate Majorana eigenstates. The inelastic nature of the DM relaxes stringent CMB and self-interaction constraints compared to the symmetric baseline model of Chapter 5. The unconstrained parameter space in this model serves as an ideal target for accelerator A' searches, and also provides a DM self-interaction cross section that is large enough to observably impact small-scale structure. Chapter 6 contains our paper Ref. [3].

In Chapter 7 I begin a discussion of DM-DM bound states, beginning with a basic discussion of how bound states are treated in quantum field theory. In Section 7.2.1 I quickly review how to describe bound states in quantum field theory by summarizing the well-known case of positronium decay to two photons. Then in Section 7.2.2 I summarize the formation of positronium from the radiative capture of an electron and positron. In Section 7.3 I present the nonabelian generalization of positronium and summarize a calculation of the rate for its formation, in order to introduce us to the basic tools necessary for describing the radiative formation of bound states in general dark sector models. The goal of Chapter 7 is to provide the groundwork for understanding Chapter 8, where we study the formation and properties of bound states in a particular dark sector model.

Finally, in Chapter 8 I present a calculation of the rate for $SU(2)$ triplet DM (the wino) to bind into WIMPonium, and study the subsequent decays of these bound states. In the specific case of the wino, we find that the rate for bound state formation is suppressed relative to direct annihilation, and so provides only a small correction to the overall annihilation rate. The soft photons radiated by the capture process and by bound state transitions could permit measurement of the DM's quantum numbers;

for wino-like DM, such photons are rare, but might be observable by a future ground-based gamma-ray telescope combining large effective area and a low energy threshold. Chapter 8 contains our paper Ref. [4].

Chapter 2

Primordial inhomogeneities during inflation

2.1 Introduction

Cosmic Inflation is extremely successful at explaining observations about our universe today, and has become a central part of our understanding of the very early universe. According to inflationary cosmology, a small fraction of a second after the Big Bang the spatial scale of the universe expanded exponentially quickly, growing from $\sim 10^{-26}$ m to ~ 1 m in a period of $\sim 10^{-35}$ s, before transitioning to a much slower rate of expansion for the rest of its history. A period of inflation successfully resolves observational puzzles which are left by the standard model of cosmology – most notably the horizon and flatness problems. The spectrum of primordial inhomogeneities predicted by inflation is also in remarkably close agreement with the pattern of temperature anisotropies we observe in the cosmic microwave background (CMB). In what follows I will briefly explain each of these successes of inflation, setting up the framework we need to describe inflationary dynamics and the evolution of primordial matter fluctuations during inflation. This is the subject of Chapter 2. Then in Chapter 3 I will scrutinize inflation by investigating a remaining fine-tuning problem that arises, the "initial conditions problem". By the end of Chapters 2 and 3 our goal is to have a good sense of how inflation works and just how robust inflation is to very

inhomogeneous initial conditions. For the discussion in this chapter I take a lot of inspiration from a handful of sources, especially: David Kaiser's "Informal Primers" on inflation [5–8], Daniel Baumann's TASI Lectures on Inflation [9], the nice review article "Inflationary Cosmology: Exploring the Universe from the Smallest to the Largest Scales" by Alan Guth and David Kaiser [10], and Viatcheslav Mukhanov's textbook "Physical Foundations of Cosmology" [11].

2.2 Basics of FLRW spacetime

On the largest observable scales the universe can be approximated as being homogeneous (space has the same properties at every point) and isotropic (space has the same properties in every direction). The CMB radiation is the clearest manifestation of homogeneity and isotropy in our universe. Astronomical observations also show us that space is expanding. We can therefore describe spacetime with the Friedmann-Lemaître-Robertson-Walker (FLRW) metric:

$$\begin{aligned} ds^2 &= g_{\mu\nu}(x)\, dx^\mu dx^\nu \\ &= -dt^2 + a^2(t)\left[\frac{dr^2}{(1-Kr^2)} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right], \end{aligned} \tag{2.1}$$

where t indicates cosmic time, the scale factor $a(t)$ describes the overall stretching of space over time, and the curvature constant K describes the overall curvature of space. Vanishing curvature constant, $K = 0$, indicates no spatial curvature (spatial sections at a given moment in time are Euclidean), $K = +1$ indicates positive spatial curvature (spatial sections curve around such that parallel lines in one region converge away from that region), and $K = -1$ indicates negative spatial curvature (spatial sections open out such that parallel lines in one region diverge from each other away from that region).

The homogeneity and isotropy of FLRW spacetime require that the energy-momentum tensor $T^{\mu\nu}$ take the form of a perfect fluid (a fluid which is incompressible and flows

without any anisotropic pressures or shears). $T^{\mu\nu}$ may then be written as:

$$T^{\mu\nu} = \left(\rho + \frac{p}{c^2}\right) u^\mu u^\nu + p g^{\mu\nu}, \tag{2.2}$$

where ρ is the energy density, p is the pressure, and u^μ is the velocity of the fluid, normalized such that $u_\mu u^\mu = -c^2$. The energy-momentum tensor in general is conserved:

$$D_\nu T^{\mu\nu} = 0. \tag{2.3}$$

The $\mu = t$ component of Equation 2.3, for example, describes how the energy density evolves over time:

$$D_\nu T^{t\nu} = \dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0 \tag{2.4}$$

The equation of state of matter in the universe is defined by the relation:

$$p = w\rho. \tag{2.5}$$

If we assume that w is constant, then Equations 2.5 and 2.4 allow us to relate the evolution of the energy density to the evolution of the scale factor according to the equation of state of matter:

$$\rho(t) = \rho_0 \left[a(t)\right]^{-3(1+w)}. \tag{2.6}$$

For example, nonrelativistic matter ($w \simeq 0$) has an equation of state $\rho(t) = \rho_0 \left[a(t)\right]^{-3}$, whereas relativistic matter ($w = 1/3$) has an equation of state $\rho(t) = \rho_0 \left[a(t)\right]^{-4}$.

The geometry of spacetime and the distribution of matter and energy are related through Einstein's field equations:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{1}{M_{\rm pl}^2} T_{\mu\nu}, \tag{2.7}$$

where the reduced Planck mass $M_{\rm pl} \equiv 1/\sqrt{8\pi G}$, with G Newton's gravitational con-

stant. The 00 component of Equation 2.7 gives the Friedmann equation:

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{1}{3M_{\rm pl}^2}\rho - \frac{K}{a^2}. \tag{2.8}$$

and subtracting Equation 2.8 from the ij components of Equation 2.7 yields an expression for the acceleration of the scale factor:

$$\frac{\ddot{a}}{a} = -\frac{1}{6M_{\rm pl}^2}\left(\rho + 3p\right). \tag{2.9}$$

Now let's make the discussion more concrete and consider the simplest inflationary model in which matter is described by a single scalar field ϕ with minimal coupling to gravity and canonical kinetic term, for which the action may be written:

$$S = \int d^4x\sqrt{-g}\left[\frac{M_{\rm pl}^2}{2}R - \frac{1}{2}g^{\mu\nu}\,\partial_\mu\phi\,\partial_\nu\phi - V(\phi)\right]. \tag{2.10}$$

Varying the action with respect to ϕ yields its equation of motion,

$$\Box\phi - V_{,\phi} = 0, \tag{2.11}$$

which together with Equation 2.7, with $T_{\mu\nu}$ given by:

$$\begin{aligned} T_{\mu\nu} &= \frac{-2}{\sqrt{-g}}\frac{\delta S^{(M)}}{\delta g^{\mu\nu}} \\ &= \partial_\mu\phi\,\partial_\nu\phi - g_{\mu\nu}\left[\frac{1}{2}g^{\alpha\beta}\,\partial_\alpha\phi\,\partial_\beta\phi + V(\phi)\right], \end{aligned} \tag{2.12}$$

constitute the coupled equations of motion for $\phi\,(t)$ and $a\,(t)$. As usual, the covariant d'Alembertian operator is given by

$$\Box\phi = \frac{1}{\sqrt{-g}}\partial_\mu\left[\sqrt{-g}\,g^{\mu\nu}\,\partial_\nu\phi\right]. \tag{2.13}$$

For an FLRW spacetime dominated by a single matter field ϕ, as in Equation

2.10, Equation 2.11, becomes:

$$\ddot{\phi} + 3H\dot{\phi} - V_{,\phi} = 0, \tag{2.14}$$

and the energy density and pressure are given by:

$$T_{00} \equiv \rho = \frac{1}{2}\dot{\phi}^2 + V(\phi) \tag{2.15}$$

and

$$T_{ij} \equiv g_{ij}p = g_{ij}\left[\frac{1}{2}\dot{\phi}^2 - V(\phi)\right]. \tag{2.16}$$

2.3 Basics of cosmic inflation

Now we have everything we need to consider the dynamics of an FLRW spacetime filled with a single scalar matter field. The Einstein's equations, Equation 2.7, describes how the scale factor $a(t)$ evolves over cosmic time, with the rate of acceleration proportional to the energy density and pressure, according to Equation 2.9. Before cosmic inflation became the leading paradigm for understanding the very early universe, in the standard model of cosmology, it was assumed that matter behaved "normally"; that is, matter would be assumed to have both positive mass-energy density and positive, or zero, pressure, such that $(\rho + 3p) > 0$. Under these "ordinary" circumstances the expansion of the universe would slow down according to Equation 2.9, with the energy density and pressure diluting according to Equation 2.6.

But as we learn by studying particle physics, matter can behave very differently from what we might ordinarily expect. Our most fundamental (and successful) theory today describes matter as a set of quantum fields permeating spacetime. For example, let's just consider a spacetime dominated by a single scalar matter field ϕ. This scalar field ϕ can get trapped in a false vacuum state, with its energy density ρ remaining nearly constant even as the universe expands. We see from Equation 2.4 that if ρ is approximately constant as the universe expands, this implies that matter during this period will have a negative pressure, with equation of state $p \simeq -\rho$, and

therefore $(\rho + 3p) < 0$. From Equation 2.9 we see that this will give rise to the accelerated expansion of the universe which characterizes inflation. During inflation the scale factor of the universe grows as $a \propto e^{Ht}$, blowing up by a factor of $\sim 10^{26}$ within 10^{-35} s. Let's summarize this more closely in the context of slowroll models of inflation.

For a wide range of models the $3H\dot{\phi}$ term in the equation of motion for ϕ, Equation 2.14, provides a frictional drag which damps out the evolution of the field ϕ, causing it to evolve slowly towards its equilibrium value, which can lead to the conditions described above driving inflation. This happens when the frictional damping effect causes ϕ to evolve slowly from a region of high potential energy $V(\phi)$ towards its equilibrium, while its kinetic energy remains much less than its potential energy, $\dot{\phi}^2/2 \ll V(\phi)$. This behavior is referred to as "slowroll" dynamics. The conditions which typically define "slowroll" evolution are:

$$\begin{aligned} \frac{1}{2}\dot{\phi}^2 &\ll V(\phi) \\ \left|\dot{H}\right| &\ll H^2 \\ \left|\ddot{\phi}\right| &\ll \left|3H\dot{\phi}\right|. \end{aligned} \tag{2.17}$$

And note that these conditions are not independent of one another.

So let's say at some point in the history of our universe, in some tiny region of space, the evolution of the matter field ϕ filling our universe enters a period of slowroll, in which the potential energy stored in the matter field dominated its kinetic energy:

$$\frac{1}{2}\dot{\phi}^2 \ll V(\phi), \tag{2.18}$$

then from Equations 2.15 and 2.16 we can see this implies an equation of state of matter:

$$p \simeq -\rho, \tag{2.19}$$

and from Equation 2.9 we can see that such an equation of state, with $w < -1/3$, drives a period of accelerated expansion of the universe, with $\ddot{a} > 0$. In particular,

for the case of $w \simeq -1$, from Equations 2.9, 2.15, and 2.16 we find:

$$\frac{\ddot{a}}{a} \simeq \frac{1}{3M_{\rm pl}^2} V(\phi) \,. \tag{2.20}$$

During this period the Friedmann equation, Equation 2.8, becomes:

$$H^2 \simeq \frac{1}{3M_{\rm pl}^2} V(\phi) \,, \tag{2.21}$$

and we see that during inflation $H^2 \simeq \ddot{a}/a$. From the definition of $H \equiv \dot{a}/a$ we have:

$$\dot{H} = \frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2 \tag{2.22}$$

which implies that for $H^2 \simeq \ddot{a}/a$, the Hubble parameter remains nearly constant, $\dot{H} \simeq 0$ during inflation. $\dot{H} \simeq 0$ implies that $a(t)$ grows exponentially with time:

$$a(t) = a_0 \exp\left[H(t - t_0)\right] . \tag{2.23}$$

The volume of space grows exponentially with time during the inflationary phase, so that even though inflation only needs to begin within a tiny region of space, that tiny region will quickly dominate the volume of space of surrounding regions which had not been inflating, and will characterize the spacetime dynamics.

It will be useful later to briefly describe how we typically estimate when slowroll dynamics will begin and come to an end, by introducing the slowroll parameters:

$$\epsilon \equiv -\frac{\dot{H}}{H^2} \tag{2.24}$$

and

$$\eta \equiv \epsilon - \frac{\ddot{\phi}}{H\dot{\phi}}, \tag{2.25}$$

When $\epsilon < 1$ (and $\eta < 1$) corresponds to when the system enters an inflationary phase of evolution, when $\epsilon \ll 1$ (and $\eta \ll 1$) corresponds to when the system is evolving deep along the slowroll attractor, and $\epsilon > 1$ (and $\eta > 1$) marks the end of accelerated

expansion and inflation. It can be shown that during slowroll:

$$\epsilon \simeq \frac{M_{\rm pl}^2}{2}\left(\frac{V_{,\phi}}{V}\right)^2 \tag{2.26}$$

and

$$\eta \simeq M_{\rm pl}^2\left(\frac{V_{,\phi\phi}}{V}\right), \tag{2.27}$$

which will be helpful later on.

It is also useful to introduce the number of "efolds" of inflation. Between the start of inflation, t_i, and the end of inflation $t_{\rm end}$ the scale factor grows by:

$$\frac{a\left(t_{\rm end}\right)}{a\left(t_i\right)} = e^{H(t_{\rm end}-t_i)} \equiv e^N \tag{2.28}$$

where

$$N \equiv \int_{t_i}^{t_{\rm end}} dt H\left(t\right) \simeq H\left(t_{\rm end}-t_i\right) \tag{2.29}$$

is the number of "efolds" of inflation. In order to solve the fine-tuning problems left by standard cosmology, the flatness and horizon problems, which I will describe below, we typically require inflation to provide at least 65 efolds of expansion:

$$N \gtrsim 65, \tag{2.30}$$

which corresponds to an increase in the universe's spatial volume by a factor of $a^3 = e^{3N} \gtrsim 10^{84}$.

2.4 Solving the flatness problem

Now we can see how inflation solves the flatness problem of standard cosmology, where the initial value of the Hubble constant H must be extremely fine tuned in order to produce a universe that is as spatially flat as the one we observe today. Let's start by defining the critical energy density and matter density parameter Ω. By rearranging

the Friedmann equation, Equation 2.8:

$$\frac{K}{a^2} = \frac{1}{3M_{\rm pl}^2}\rho - H^2, \tag{2.31}$$

we see that if the energy density takes the critical value ρ_{cr}:

$$\rho_{cr} \equiv 3M_{\rm pl}^2 H^2 \tag{2.32}$$

then the spatial curvature term K vanishes, $K = 0$; a universe that has energy density equal to the critical value is spatially flat, while a universe with energy density that exceeds or is less than the critical value will have positive or negative spatial curvature, respectively. Now let's introduce the dimensionless parameter Ω:

$$\Omega \equiv \frac{\rho}{\rho_{cr}} \tag{2.33}$$

in terms of which we can rewrite Equation 2.31 as:

$$\Omega = 1 + \frac{K}{a^2 H^2} \tag{2.34}$$

where the comoving Hubble radius, $(aH)^{-1}$, is the comoving size of the observable universe. We can already summarize the flatness problem from Equation 2.34, because in standard cosmology the comoving Hubble radius always grows with time, making a spatially flat universe with $\Omega = 1$ an unstable solution. To make this more clear let's write:

$$\frac{(\Omega - 1)}{\Omega} = \frac{3M_{\rm pl}^2 K}{a^2 \rho} \tag{2.35}$$

and substituting Equation 2.6 we find:

$$\frac{(\Omega - 1)}{\Omega} \propto a^{3(1+w)-2} \tag{2.36}$$

so that for matter that is either nonrelativistic ($w = 0$) or relativistic ($w = 1/3$),

spatial flatness is an unstable solution of Einstein's equations:

$$\frac{(\Omega - 1)}{\Omega} \propto \begin{cases} a(t) & \text{for nonrelativistic matter} \\ a^2(t) & \text{for relativistic matter.} \end{cases} \tag{2.37}$$

Current Planck measurements of the CMB find $\Omega = 1.0006 \pm 0.0019$ [12]. For Ω to be so close to 1 today, in standard cosmology, would require remarkable fine tuning at early times. For example, if Ω were 0.9 at $t = 1$ s after the big bang, it would be 10^{-14} today; and if Ω were 1.1 at $t = 1$ s after the big bang it would have grown so quickly that the universe would have recollapsed 45 s later [10].

Inflation offers a simple solution to the spatial flatness problem. Referring back to Equation 2.36, we see that if the matter filling the universe has an equation of state $w < -1/3$, then the quantity $(\Omega - 1)$ will decrease, driving Ω to 1. As we've discussed, the condition that $w < -1/3$ is the same as the condition leading to accelerated expansion of the universe, $\ddot{a} > 0$, which defines inflation.

2.5 Solving the horizon problem

We can similarly see how inflation solves the horizon problem of standard cosmology, in which the initial universe appears homogeneous over length scales which apparently encompass at least $\sim 10^{83}$ separate causally disconnected regions [13]. To see this let's introduce the comoving particle horizon τ, which is the causal horizon, or the maximum distance a light ray can travel between time $t = 0$ and a later time t:

$$\tau \equiv \int_0^t \frac{dt'}{a(t')} = \int_0^a \frac{da}{Ha^2} = \int_0^a d \ln a \left(\frac{1}{aH}\right). \tag{2.38}$$

The comoving horizon τ is the logarithmic integral of the comoving Hubble radius $(aH)^{-1}$. It is important to note a distinction between τ and $(aH)^{-1}$: if particles are separated by comoving distances greater than τ, they never could have been in contact with one another, and if they are separated by distances greater than $(aH)^{-1}$ they cannot be in contact with one another now.

In standard cosmology the comoving Hubble radius is always growing as the universe evolves, and hence we observe homogeneity across regions of space which are separated by more than a comoving Hubble radius, apparently implying that they never would have been in causal contact.

If a period of cosmic inflation occurs in the early universe, however, then it is possible for the comoving horizon τ to be much larger than the comoving Hubble radius now, $(aH)^{-1}$, so that particles which cannot communicate today could have been in causal contact at earlier times. This is because the same (equivalent) conditions which give a period of cosmic inflation, accelerated expansion of the universe driven by matter having an equation of state with $w < -1/3$, also imply a shrinking comoving Hubble radius during inflation:

$$\begin{aligned} &\frac{d^2a}{dt^2} > 0 \\ &\rho + 3p < 0 \\ &\frac{d}{dt}\left(\frac{1}{aH}\right) < 0. \end{aligned} \tag{2.39}$$

A shrinking comoving Hubble radius during inflation, which then begins to grow as we expect with Hubble expansion once inflation ends, implies that the large scales which are entering the present universe were inside the horizon, and in causal contact, before inflation. This way the uniformity of the CMB is no longer a mystery.

2.6 Linearized perturbations

Now let's turn to considering the primordial perturbations that we measure to high precision on the CMB. The temperature anisotropies that we see on the CMB are at the level of a part in 10^5, $(\Delta T/T \simeq 10^{-5})$, so to describe the late-time observations which match CMB observations is is sufficient to consider them as small perturbations around a classical background whose dynamics we described above. To do this, we typically expand the inflaton field and the spacetime metric around their classical

background values:

$$\begin{aligned} \phi(t,\mathbf{x}) &= \varphi(t) + \delta\phi(t,\mathbf{x}) \\ g_{\mu\nu}(t,\mathbf{x}) &= \bar{g}_{\mu\nu}(t) + \delta g_{\mu\nu}(t,\mathbf{x}) \end{aligned} \tag{2.40}$$

and expand the equations of motion to leading order in spatially varying quantities. Here's a brief description of how this is done.

2.6.1 Perturbing the spacetime metric

Let's specialize to the case of perturbations around a spatially flat ($K = 0$) FLRW background. Then the most general set of perturbations to the line element ds may be written:

$$\begin{aligned} ds^2 = &-(1+2A)\,dt^2 + 2a\,(\partial_i B - S_i)\,dt dx^i \\ &+ a^2\left[(1-2\psi)\,\delta_{ij} + 2\partial_i\partial_j E + 2\,(\partial_i F_j + \partial_j F_i) + \mathcal{H}_{ij}\right] dx^i dx^j. \end{aligned} \tag{2.41}$$

The vector perturbations $S^i(x)$ and $F^i(x)$ quickly redshift away as the FLRW spacetime expands, so we will ignore them from here on out; the tensor perturbations $\mathcal{H}_{ij}(x)$ represent gravitational waves, and decouple from the scalar perturbations, $A(x)$, $B(x)$, $\psi(x)$, and $E(x)$, at linear order. So let's focus on the scalar perturbations in what follows.

We should note that the scalar perturbations $A(x)$, $B(x)$, $\psi(x)$, and $E(x)$ actually contain redundant degrees of freedom, and we can completely characterize the scalar perturbations with just two functions. This subtlety means that there are multiple ways we can fix and eliminate two of the scalar metric functions to describe the same set of physical observables with the remaining two degrees of freedom. This is referred to as choosing a gauge. I won't go into detail here about the different gauge choices and how they amount to describing the same set of physical observables, but it's good to mention it.

Another approach to dealing with the redundant degrees of freedom in the set of scalar perturbations is to construct gauge-invariant combinations of them, and work in terms of those gauge-invariant quantities. Again I won't go into detail about this,

let's just move forward by working in terms of the gauge invariant Bardeen potentials:

$$\begin{aligned}\Phi &\equiv A - \frac{d}{dt}\left[a^2\left(\dot{E} - \frac{B}{a}\right)\right] \\ \Psi &\equiv \psi + a^2 H\left(\dot{E} - \frac{B}{a}\right)\end{aligned} \tag{2.42}$$

and we work in longitudinal gauge, in which $E = B = 0$. Now the perturbed line element takes the form:

$$ds^2 = -\left(1 + 2\Phi\right)dt^2 + a^2\left(1 - 2\Psi\right)\delta_{ij}dx^i dx^j. \tag{2.43}$$

2.6.2 Perturbing the Einstein's field equations

Let's first summarize perturbing the Einstein's field equations to linear order in spatially varying quantities:

$$\bar{G}_{\mu\nu} + \delta G_{\mu\nu} = \frac{1}{M_{\rm pl}^2}\left[\bar{T}_{\mu\nu} + \delta T_{\mu\nu}\right]. \tag{2.44}$$

Expanding the Einstein tensor to first order in Φ and Ψ, and separating its components into background and first-order quantities we obtain:

$$\begin{aligned}\bar{G}_{00} &= 3H^2 \\ \bar{G}_{0i} &= 0 \\ \bar{G}_{ij} &= -a^2\delta_{ij}\left(2\dot{H} + 3H^2\right)\end{aligned} \tag{2.45}$$

and

$$\begin{aligned}G^0_0 &\equiv g^{0\alpha}G_{0\alpha} = -3H^2 + 6H^2\Phi + 6H\dot{\Psi} - \frac{2}{a^2}\nabla^2\Psi \\ G^0_i &\equiv g^{0\alpha}G_{i\alpha} = -\left(1 - 2\Phi\right)2\partial_i\left(\dot{\Psi} + H\Phi\right) \\ G^i_j &\equiv g^{i\alpha}G_{j\alpha} = -\delta^i_j\left(2\dot{H} + 3H^2\right) \\ &\quad + 2\delta^i_j\left[\ddot{\Psi} + 3H\dot{\Psi} + H\dot{\Phi} + \left(2\dot{H} + 3H^2\right)\Phi + \frac{1}{2a^2}\nabla^2\left(\Phi - \Psi\right)\right] \\ &\quad - \frac{1}{a^2}\partial^i\partial_j\left(\Phi - \Psi\right)\end{aligned} \tag{2.46}$$

Now we turn to perturbing the energy-momentum tensor $T_{\mu\nu}$. The perturbed energy momentum tensor can generally be written as:

$$\begin{aligned} T_0^0 &= -\left(\rho + \delta\rho\right) \\ T_i^0 &= \partial_i \delta q \\ T_j^i &= \delta_j^i \left(p + \delta p\right) + \Pi_j^i, \end{aligned} \tag{2.47}$$

where $\delta\rho$ is the density perturbation, δq is the momentum flow, δp is the isotropic pressure perturbation, and Π^{ij} is the anisotropic pressure. We may write Π_{ij} in terms of a projection operator:

$$\Pi_{ij} = \left[\partial_i \partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right]\delta\Pi \tag{2.48}$$

which captures the fact that $\Pi_i^i = 0$, Π^{ij} has no time-like components, is symmetrical in its indices, and only has components on the off-diagonal.

Now we can write out the perturbed Einstein's equations component-by-component in terms of the perturbed fluid quantities, as well as Φ, and Ψ. The 00 component, $G_0^0 = M_{\rm pl}^{-2} T_0^0$, yields:

$$\begin{aligned} 3H^2 &= \frac{1}{M_{\rm pl}^2}\rho \\ 3H\left(\dot{\Psi} + H\Phi\right) - \frac{1}{a^2}\nabla^2\Psi &= -\frac{1}{2M_{\rm pl}^2}\delta\rho; \end{aligned} \tag{2.49}$$

the $0i$ component yields:

$$\dot{\Psi} + H\Phi = -\frac{1}{2M_{\rm pl}^2}\delta q; \tag{2.50}$$

the ij components yield:

$$\begin{aligned} \left(2\dot{H} + 3H^2\right) &= -\frac{p}{M_{\rm pl}^2} \\ \ddot{\Psi} + 3H\dot{\Psi} + H\dot{\Phi} + \left(2\dot{H} + 3H^2\right)\Phi &= \frac{1}{2M_{\rm pl}^2}\left[\delta p - \frac{2}{3}\nabla^2\delta\Pi\right] \\ \frac{1}{a^2}\partial^i\partial_j\left(\Phi - \Psi\right) &= \frac{1}{M_{\rm pl}^2}\partial^i\partial_j\delta\Pi. \end{aligned} \tag{2.51}$$

We see from Equation 2.51 that Φ and Ψ will be different if the system includes

anisotropic pressure $\delta\Pi \neq 0$, and the converse is true, if there is no anisotropic pressure $\delta\Pi = 0$ then $\Phi = \Psi$.

Now let's consider the conservation of the energy-momentum tensor $\nabla_\nu T^{\mu\nu} = 0$:

$$\begin{aligned}\nabla_\nu T^{0\nu} = &\left[\dot{\rho} + 3H\left(\rho + p\right)\right]\left(1 - 2\Phi\right) \\ &+ \left[\delta\dot{\rho} + 3H\left(\delta\rho + \delta p\right) + \frac{1}{a^2}\nabla^2\delta q - 3\left(\rho + p\right)\dot{\Psi}\right] = 0.\end{aligned} \tag{2.52}$$

Separating Equation 2.52 into background and first-order quantities yields:

$$\begin{aligned}\dot{\rho} + 3H\left(\rho + p\right) &= 0 \\ \delta\dot{\rho} + 3H\left(\delta\rho + \delta p\right) &= -\frac{1}{a^2}\nabla^2\delta q + 3\left(\rho + p\right)\dot{\Psi}\end{aligned} \tag{2.53}$$

2.6.3 Application to a single-field model

Now let's apply the discussion above to the single-field model of Equation 2.10, and expand the expression for $T_{\mu\nu}$ in Equation 2.12 to linear order in spatially varying quantities Ψ, Φ, and $\delta\phi$, so that we can express Equation 2.53 in terms of these quantities. For example, expanding the term in square brackets in the definition of $T_{\mu\nu}$, Equation 2.12, up to terms linear in spatially varying quantities, yields:

$$\left[\frac{1}{2}g^{\alpha\beta}\,\partial_\alpha\phi\,\partial_\beta\phi + V(\phi)\right] = -\frac{1}{2}\left(1 - 2\Phi\right)\dot{\varphi}^2 - \dot{\varphi}\delta\dot{\phi} + V\left(\varphi\right) + V_{,\phi}\left(\varphi\right)\delta\phi. \tag{2.54}$$

Proceeding similarly, the 00 component of $T_{\mu\nu}$, separating background and first-order quantities, yields:

$$\begin{aligned}\bar{T}_{00} &= \frac{1}{2}\dot{\varphi}^2 + V\left(\varphi\right) \\ \delta T_{00} &= \dot{\varphi}\delta\dot{\phi} + 2\Phi V\left(\varphi\right) + V_{,\phi}\left(\varphi\right)\delta\phi;\end{aligned} \tag{2.55}$$

the ij components of $T_{\mu\nu}$ yield:

$$\begin{aligned}\bar{T}_{ij} &= a^2\delta_{ij}\left(\frac{1}{2}\dot{\varphi}^2 - V\right) \\ \delta T_{ij} &= a^2\delta_{ij}\left[\dot{\varphi}\delta\dot{\phi} - \left(\Phi + \Psi\right)\dot{\varphi}^2 + 2\Psi V - V_{,\phi}\delta\phi\right];\end{aligned} \tag{2.56}$$

and the $0i$ components of $T_{\mu\nu}$ yield:

$$\begin{aligned} \bar{T}_{0i} &= 0 \\ \delta T_{0i} &= \dot{\varphi}\partial_i \delta\phi. \end{aligned} \tag{2.57}$$

By comparing Equations 2.55 through 2.57 with Equation 2.47 (remembering to raise the index) we can identify the perturbed fluid quantities:

$$\begin{aligned} \rho &= \frac{1}{2}\dot{\varphi}^2 + V \\ \delta\rho &= \dot{\varphi}\delta\dot{\phi} - \dot{\varphi}^2\Phi + V_{,\phi}\delta\phi, \end{aligned} \tag{2.58}$$

$$\delta q = -\dot{\varphi}\delta\phi, \tag{2.59}$$

and

$$\begin{aligned} p &= \frac{1}{2}\dot{\varphi}^2 - V \\ \delta p &= \dot{\varphi}\delta\dot{\phi} - \dot{\varphi}^2\Phi - V_{,\phi}\delta\phi \\ \Pi_{ij} &= 0. \end{aligned} \tag{2.60}$$

We find that in single-field models with minimal coupling the anisotropic pressure vanishes, and therefore, according to Equation 2.51, the Bardeen potentials are equal: $\Phi = \Psi$. From now on we'll express them in terms of Ψ.

Having identified the perturbed fluid quantities in Equations 2.58 through 2.60, we can now write the perturbed Einstein's equations component-by-component in terms of Ψ, $\delta\phi$, and $\delta\dot{\phi}$. From Equation 2.49 we have:

$$\begin{aligned} H^2 &= \frac{1}{3M_{\rm pl}^2}\left[\frac{1}{2}\dot{\varphi}^2 + V\right] \\ 3H\left(\dot{\Psi} + H\Psi\right) - \frac{1}{a^2}\nabla^2\Psi &= -\frac{1}{2M_{\rm pl}^2}\left[\dot{\varphi}\delta\dot{\phi} - \dot{\varphi}^2\Psi + V_{,\phi}\delta\phi\right]; \end{aligned} \tag{2.61}$$

from Equation 2.50 we have:

$$\dot{\Psi} + H\Psi = \frac{1}{2M_{\rm pl}^2}\dot{\varphi}\delta\phi; \tag{2.62}$$

from Equation 2.51 we have:

$$\begin{aligned} \left(2\dot{H} + 3H^2\right) &= -\frac{1}{M_{\rm pl}^2}\left(\frac{1}{2}\dot{\varphi}^2 - V\right) \\ \ddot{\Psi} + 3H\dot{\Psi} + H\dot{\Psi} + \left(2\dot{H} + 3H^2\right)\Psi &= \frac{1}{2M_{\rm pl}^2}\left[\dot{\varphi}\delta\dot{\phi} - \dot{\varphi}^2\Psi - V_{,\phi}\delta\phi\right] \end{aligned} \tag{2.63}$$

2.6.4 Perturbing the equation of motion for ϕ

Having obtained the perturbed Einstein's field equations for a universe dominated by a single minimally coupled scalar field, Equations 2.61 through 2.63, let's consider the perturbed equations of motion for the inflaton field ϕ. The equation of motion for ϕ is given by Equation 2.11, with $\Box\phi$ given by Equation 2.13. Expanding the determinant $\sqrt{-g}$ to linear order in $\Psi = \Phi$, we have:

$$\sqrt{-g} = a^3\left(1 - 2\Psi\right), \tag{2.64}$$

which according to Equation 2.13, and expanding ϕ to linear order in $\delta\phi$, yields:

$$\Box\left(\varphi + \delta\phi\right) = -\left(\ddot{\varphi} + 3H\dot{\varphi}\right)\left(1 - 2\Psi\right) + 4\dot{\Psi}\dot{\varphi} - \left(\delta\ddot{\phi} + 3H\delta\dot{\phi} - \frac{1}{a^2}\nabla^2\delta\phi\right). \tag{2.65}$$

We also find:

$$V_{,\phi}\left(\varphi + \delta\phi\right) = V_{,\phi}\left(\varphi\right) + V_{,\phi\phi}\left(\varphi\right)\delta\phi, \tag{2.66}$$

so putting everything together, the equations of motion for the background and linear-order components, φ and $\delta\phi$, are:

$$\ddot{\varphi} + 3H\dot{\varphi} + V_{,\phi} = 0, \tag{2.67}$$

and

$$\delta\ddot{\phi} + 3H\delta\dot{\phi} - \frac{1}{a^2}\nabla^2\delta\phi + V_{,\phi\phi}\delta\phi = -2\Psi V_{,\phi} + 4\dot{\varphi}\dot{\Psi}. \tag{2.68}$$

Equations 2.67 and 2.68, along with Equations 2.61 through 2.63, are the full equations of motion for the system of dynamical variables at background order, $\varphi(t)$, $a(t)$, and linear order $\delta\phi(\mathbf{x}, t)$, $\Psi(\mathbf{x}, t)$, (of course, we only need two of the equations among Equations 2.61 through 2.63 to solve along with Equations 2.67 and 2.68, say those in Equation 2.61, to completely describe the system). By examining these equations of motion it is important to note that to linear order in spatially varying quantities the equations of motion for field fluctuations and metric perturbations $\delta\phi$ and Ψ decouple from the equations of motion for the background quantities $\varphi(t)$, $a(t)$. This will be an important motivation for taking up the work of Chapter 3.

2.7 Quantum fluctuations during inflation

2.7.1 Canonical quantization of scalar field fluctuations

Let's briefly review canonical quantization of scalar field fluctuations in a flat ($K = 0$) background FLRW spacetime, since it will be useful for introducing some concepts which we will use later on. We promote fluctuations in the inflaton field, $\delta\phi$, and perturbations to the spacetime metric Ψ, as well as their conjugate momenta, to quantum operators:

$$\Psi(x^\mu) \to \hat{\Psi}(x^\mu) \tag{2.69}$$

and

$$\phi(x^\mu) \to \hat{\phi}(x^\mu) = \varphi(t) + \delta\hat{\phi}(x^\mu) \tag{2.70}$$

with

$$\begin{aligned} \langle\hat{\phi}(x^\mu)\rangle &\equiv \langle 0|\hat{\phi}(x^\mu)|0\rangle = \varphi(t) \\ \langle\delta\hat{\phi}(x^\mu)\rangle &= 0. \end{aligned} \tag{2.71}$$

We expand $\delta\hat{\phi}$ in terms of its Fourier components:

$$\begin{aligned}\delta\hat{\phi}(x) &= \int \frac{d^3\mathbf{k}}{(2\pi)^2}\left(\delta\phi_k(t)\,\hat{a}_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}} + \delta\phi_k^*(t)\,\hat{a}_{\mathbf{k}}^\dagger e^{-i\mathbf{k}\cdot\mathbf{x}}\right) \\ &= \int \frac{d^3\mathbf{k}}{(2\pi)^2}\left(\delta\phi_k(t)\,\hat{a}_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{x}} + \delta\phi_k^*(t)\,\hat{a}_{-\mathbf{k}}^\dagger e^{i\mathbf{k}\cdot\mathbf{x}}\right)\end{aligned} \tag{2.72}$$

with similar expressions for $\hat{\Psi}$ and the conjugate momenta. In the second line of Equation 2.72 we take $\mathbf{k} \to -\mathbf{k}$, using the fact that both the integration measure and the time-dependent mode functions $\delta\phi_k(t)$ are invariant under this transformation.

We impose the equal-time commutation relations for free scalar fields:

$$\left[\delta\hat{\phi}(t,\mathbf{x}), \delta\hat{\Pi}(t,\mathbf{y})\right] = i\delta^{(3)}(\mathbf{x}-\mathbf{y}), \tag{2.73}$$

where the momentum canonically conjugate to $\delta\hat{\phi}$, $\delta\hat{\Pi}$, is given by:

$$\delta\hat{\Pi} \equiv \frac{\partial\mathcal{L}}{\partial\dot{\delta\hat{\phi}}} = a^3(t)\,\dot{\delta\hat{\phi}}. \tag{2.74}$$

The creation and annihilation operators $\hat{a}_{\mathbf{k}}$ and $\hat{a}_{\mathbf{k}}^\dagger$ satisfy the canonical commutation relation:

$$\left[\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{q}}^\dagger\right] = (2\pi)^3\,\delta^3(\mathbf{k}-\mathbf{q}). \tag{2.75}$$

After expanding $\delta\hat{\Pi}$ in a way similar to Equation 2.72, and using the commutation relation for $\hat{a}_{\mathbf{k}}$ and $\hat{a}_{\mathbf{k}}^\dagger$ in Equation 2.75, the equal-time commutation relation Equation 2.73 imposes the constraint on the mode functions:

$$\delta\phi_k(t)\,\delta\Pi_k^*(t) - \delta\phi_k^*(t)\,\delta\Pi_k(t) = i. \tag{2.76}$$

The condition in Equation 2.76 provides one of the boundary conditions on the solutions of Equation 2.68, but in order to fix the mode functions completely we need a second boundary condition, which comes from considering the energetics of the system we want to describe in its initial state. We usually achieve this by choosing a vacuum state for the fluctuations.

2.7.2 Bunch-Davies vacuum and short-wavelength fluctuation mode behavior

To choose a vacuum state for the fluctuation modes:

$$\hat{a}_{\mathbf{k}}|0\rangle = 0, \tag{2.77}$$

let's make a standard choice and place the initial cosmological perturbations we're describing in the Bunch-Davies vacuum. The Bunch-Davies state of cosmological perturbations specifies a physical vacuum for the inflationary spacetime by requiring that the value of the Hamiltonian in the short-wavelength limit of the initial state is minimized; that is, the Bunch-Davies vacuum places the cosmological perturbations of a comoving observer in the far past, when perturbation modes begin well within the Hubble radius ($k \gg aH$), in the Minkowski vacuum. Because we will refer to it again later, let's summarize the derivation of the Bunch-Davies initial conditions for fluctuation modes.

Let's begin with the equation of motion for $\delta\phi$, Equation 2.68, which we rewrite here in terms of its Fourier components:

$$\delta\ddot{\phi}_k + 3H\delta\dot{\phi}_k + \frac{k^2}{a^2}\delta\phi_k + V_{,\phi\phi}\delta\phi_k = -2V_{,\phi}\Psi_k + 4\dot{\varphi}\dot{\Psi}_k, \tag{2.78}$$

where we used the fact that $\nabla^2\delta\phi = -k^2\delta\phi$. Let's take the limit of short-wavelength modes (those which start deep within the Hubble radius with $k \gg aH$) in the slowroll limit during inflation. Let's remind ourselves of the slowroll conditions:

$$\begin{aligned} \frac{1}{2}\dot{\phi}^2 &\ll V \\ \left|\dot{H}\right| &\ll H^2 \\ \left|\ddot{\phi}\right| &\ll \left|3H\dot{\phi}\right| \\ \left|\dot{\Psi}\right| &\ll |H\Psi|, \end{aligned} \tag{2.79}$$

where we now include a slowroll condition on the evolution of Ψ. It can be shown

that during slowroll:

$$\frac{\partial^2 V}{\partial \phi^2} = \frac{1}{M_{\rm pl}^2} V \eta \tag{2.80}$$

where η is the slowroll parameter defined in Equation 2.25. Recall that during inflation $\eta \ll 1$, and therefore:

$$\frac{\partial^2 V}{\partial \phi^2} \ll \frac{1}{M_{\rm pl}^2} V. \tag{2.81}$$

Also recall that during inflation, applying the slowroll condition to the Friedmann equation, Equation 2.8, yields:

$$H^2 \simeq \frac{1}{3M_{\rm pl}^2} V. \tag{2.82}$$

Therefore, Equations 2.81 and 2.82 imply:

$$\frac{\partial^2 V}{\partial \phi^2} \ll H^2. \tag{2.83}$$

And because we're considering modes which begin deep within the Hubble radius we have:

$$\frac{k^2}{a^2} \gg H^2 \gg \frac{\partial^2 V}{\partial \phi^2}, \tag{2.84}$$

so we can drop the last term on the left-hand-side of Equation 2.78, reducing it to:

$$\delta\ddot{\phi}_k + 3H\delta\dot{\phi}_k + \frac{k^2}{a^2}\delta\phi_k \simeq -2V_{,\phi}\Psi_k + 4\dot{\varphi}\dot{\Psi}_k. \tag{2.85}$$

Now let's focus on simplifying the right-hand-side of Equation 2.85 in the slowroll approximation. From Equation 2.67 in the slowroll approximation:

$$3H\dot{\varphi} + V_{,\phi} \simeq 0, \tag{2.86}$$

we can rewrite the right-hand-side of Equation 2.85 as:

$$\dot{\varphi}\left(6H\Psi_k + 4\dot{\Psi}_k\right), \tag{2.87}$$

and using the slowroll approximation summarized in Equation 2.79, we see that the

second term above can be dropped, reducing the right-hand-side of Equation 2.85 to:

$$6\dot{\varphi}H\Psi_k. \tag{2.88}$$

We can also simplify Equation 2.62 in the slowroll approximation:

$$\Psi_k \simeq \frac{1}{2M_{\rm pl}^2}\frac{1}{H}\dot{\varphi}\delta\phi_k, \tag{2.89}$$

which we substitute into Equation 2.88, reducing Equation 2.85 in the slowroll approximation to:

$$\delta\ddot{\phi}_k + 3H\delta\dot{\phi}_k + \frac{k^2}{a^2}\delta\phi_k \simeq 3\frac{1}{M_{\rm pl}^2}\dot{\varphi}^2\delta\phi_k. \tag{2.90}$$

Now let's use the fact that in the slowroll approximation:

$$3\frac{1}{M_{\rm pl}^2}\dot{\varphi}^2 \ll \frac{1}{3M_{\rm pl}^2}V \simeq H^2 \ll \frac{k^2}{a^2}, \tag{2.91}$$

where the first inequality comes from the slowroll approximation, the next equality comes from Equation 2.82, and the last inequality holds because we are considering modes which start deep within the Hubble radius, to further reduce Equation 2.90 to:

$$\delta\ddot{\phi}_k + 3H\delta\dot{\phi}_k + \frac{k^2}{a^2}\delta\phi_k \simeq 0. \tag{2.92}$$

Moving forward I find it more convenient to work in conformal time, so let's switch to conformal time, where $\tau = dt/a\,(t)$. After a conformal transformation Equation 2.92 takes the form:

$$\delta\phi_k^{''} + 2H\,(\tau)\,\delta\phi_k^{'} + k^2\delta\phi_k \simeq 0, \tag{2.93}$$

where primes indicate derivatives with respect to conformal time τ. Now let's make the substitution:

$$\delta\phi_k \equiv \frac{u_k}{a}, \tag{2.94}$$

in terms of which Equation 2.93 becomes:

$$u_k^{''} + \left(k^2 - \frac{a^{''}}{a}\right) u_k \simeq 0, \tag{2.95}$$

and for modes which begin deep inside the Hubble radius, $k \gg aH$, we can drop the second term in parentheses above, giving:

$$u_k^{''} + k^2 u_k \simeq 0. \tag{2.96}$$

Just to remind ourselves of where we are, Equation 2.96 is the equation of motion for field fluctuations according to a comoving observer in the far past, when $\tau \to -\infty$ (or $|k\tau| \gg 1$, or $k \gg aH$). We recognize this as the equation of motion for a simple harmonic oscillator, with independent solutions $u_k \propto e^{\pm ik\tau}$, and therefore the resulting solution for $\delta\phi_k$ is:

$$\delta\phi_k \simeq \frac{C_k}{a} e^{\pm ik\tau} \tag{2.97}$$

where C_k is a constant of integration which we will fix by considering that the initial scalar field modes arise as vacuum quantum fluctuations. The oscillating solution in Equation 2.97 makes sense, because it says that in the far past when fluctuation modes are much smaller than the initial Hubble radius (and the curvature scale), $k \ll aH$, they behave as if they are in Minkowski space. We also see that as fluctuation modes are stretched by the Hubble expansion, their amplitude decays as a^{-1}. Therefore inflation tends to wash out any pre-existing classical inhomogeneities as they are stretched out to large scales.

To fix C_k we want to minimize the vacuum expectation value $\langle 0|H|0\rangle$, where the Hamiltonian of field fluctuations is given by:

$$\hat{H}(\tau) = \int d^3\mathbf{x} \left[\delta\hat{\mathbf{\Pi}}^{\mathbf{2}}(\tau, \mathbf{x}) + \left(\nabla\delta\hat{\phi}(\tau, \mathbf{x})\right)^{\mathbf{2}}\right], \tag{2.98}$$

where for short-wavelength modes deep inside the Hubble radius an additional term which accounts for the interaction of field fluctuations with the gravitational back-

ground vanishes. Computing $\delta\hat{\Pi}$ and $\nabla\delta\hat{\phi}$ from Equations 2.72 and 2.74 and plugging into Equation 2.98, and taking the vacuum expectation value we find:

$$\langle 0|\hat{H}|0\rangle = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \left(\left|\delta\phi_k^{'}(\tau)\right|^2 + k^2 \left|\delta\phi_k(\tau)\right|^2 \right) \langle 0|\hat{a}_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{\dagger}|0\rangle \tag{2.99}$$

which is equal to

$$\langle 0|\hat{H}|0\rangle = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \left(\left|\delta\phi_k^{'}(\tau)\right|^2 + k^2 \left|\delta\phi_k(\tau)\right|^2 \right) \delta^3(0) \tag{2.100}$$

Dividing out the uninteresting divergence we see that in order to minimize $\langle 0|\hat{H}|0\rangle$ we want to minimize the quantity in parentheses in Equation 2.100:

$$\left|\delta\phi_k^{'}(\tau)\right|^2 + k^2 \left|\delta\phi_k(\tau)\right|^2, \tag{2.101}$$

for each mode k separately, subject to the normalization condition Equation 2.76. We can do this by first parameterizing ϕ_k as:

$$\delta\phi_k \equiv r_k e^{i\alpha_k}, \tag{2.102}$$

then substituting into Equation 2.101 we obtain:

$$r_k^{'2} + r_k^2\alpha_k^{'2} + k^2 r_k^2 \tag{2.103}$$

and substituting into Equation 2.76 we obtain the condition:

$$\alpha_k^{'} = -\frac{1}{2r_k^2}. \tag{2.104}$$

Substituting Equation 2.104 into Equation 2.103 we obtain:

$$r_k^{'2} + \frac{1}{4r_k^2} + k^2 r_k^2 \tag{2.105}$$

and we see that $\langle 0|\hat{H}|0\rangle$ is minimized for:

$$\begin{aligned} r_k^{'} &= 0 \\ r_k &= \pm\frac{1}{\sqrt{2k}}, \end{aligned} \tag{2.106}$$

which we can plug into Equation 2.104 to obtain the additional condition:

$$\alpha_k = -k\tau. \tag{2.107}$$

With this we finally have our initial conditions for the mode functions of short-wavelength ($k \ll aH$) field fluctuations which begin in the Bunch-Davies vacuum:

$$\begin{aligned} \delta\phi_k\left(\tau \to -\infty\right) &= r_k e^{i\alpha_k} = \pm\frac{1}{\sqrt{2k}}e^{-ik\tau} \\ \delta\phi_k^{'}\left(\tau \to -\infty\right) &= \mp i\sqrt{\frac{k}{2}}e^{-ik\tau}. \end{aligned} \tag{2.108}$$

Comparing Equation 2.108 with Equation 2.97 we infer that $C_k \sim k^{-1/2}$, and the evolution of modes preserves the vacuum spectrum.

2.7.3 Long-wavelength fluctuation mode behavior

Having characterized the behavior of the short wavelength modes ($k \gg aH$), those which are within the Hubble radius, during inflation, let's turn to considering the behavior of the long-wavelength modes ($k < aH$) once they have been stretched outside the Hubble radius with the Hubble expansion. For this calculation we'll closely follow the discussion in Section 8.2.2 of [11]. Let's again start with the equations of motion of field fluctuations and metric perturbations, Equations 2.68 and 2.62. In the long-wavelength limit we can neglect the spatial derivative term $\nabla^2\delta\phi$ in Equation 2.68. In the slowroll approximation we can also neglect terms proportional to $\delta\ddot{\phi}$ and $\dot{\Psi}$. Then Equations 2.68 and 2.62 become:

$$3H\delta\dot{\phi} + V_{,\phi\phi}\delta\phi + 2V_{,\phi}\Psi \simeq 0 \tag{2.109}$$

and

$$H\Psi \simeq \frac{1}{2M_{\rm pl}^2}\dot{\varphi}\delta\phi. \tag{2.110}$$

Now let's rewrite these in terms of the new variable:

$$y \equiv \frac{\delta\phi}{V_{,\phi}} \tag{2.111}$$

in terms of which Equations 2.109 and 2.110 simplify to:

$$3H\delta\dot{y} + 2\Psi \simeq 0 \tag{2.112}$$

and

$$H\Psi \simeq \frac{1}{2M_{\rm pl}^2}\dot{V}y. \tag{2.113}$$

And because $H^2 \simeq \frac{1}{3M_{\rm pl}^2}V$ during inflation, Equations 2.112 and 2.113 imply:

$$\frac{d\,(yV)}{dt} \simeq 0 \tag{2.114}$$

which we can integrate to give the solution:

$$y = \frac{A}{V}, \tag{2.115}$$

where A is a constant of integration. The solution for the behavior of long-wavelength modes during inflation is then:

$$\delta\phi_k = A_k\frac{V_{,\phi}}{V} \tag{2.116}$$

and

$$\Psi_k = -\frac{1}{2}A_k\left(\frac{V_{,\phi}}{V}\right)^2, \tag{2.117}$$

and comparing with Equation 2.26 we see that we can write this in terms of the slowroll parameter ϵ:

$$\delta\phi_k \propto \sqrt{\epsilon} \tag{2.118}$$

and

$$\Psi_k \propto -\epsilon, \tag{2.119}$$

and since the slowroll parameter ϵ grows slowly towards the end of inflation during inflation, we see that after horizon crossing ($a > k/H$) fluctuation modes cease to oscillate, remaining nearly constant, growing slowly towards the end of inflation. Since they remain nearly constant, modes are said to "freeze out" after horizon crossing.

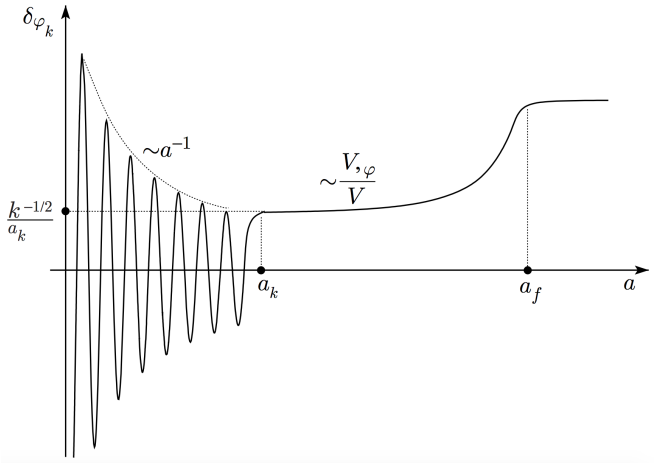

Figure 2-1: The behavior of a field fluctuation mode $\delta\phi_k$ that begins inside the Hubble radius. Fluctuation modes inside the horizon oscillate with decaying amplitude as they are stretched by the Hubble expansion, tending toward their minimum-energy state given by the Bunch-Davies vacuum, Equation 2.108. Once fluctuation modes cross outside the Hubble radius ($a > k/H$) they freeze out to nearly constant values, albeit with a modest growth towards the end of inflation given by Equation 2.118. Image taken from [11].

One gauge-invariant quantity we can construct from the field fluctuations $\delta\phi$ and metric perturbations Ψ, which directly relates to the temperature anisotropies we observe on the CMB, is the gauge-invariant curvature perturbation $\mathcal{R}$, which for a single-field model takes the simple form:

$$\mathcal{R} = \Psi + \frac{H}{\dot{\varphi}}\delta\phi. \tag{2.120}$$

Using the relation:

$$\dot{H} = -\frac{\dot{\varphi}^2}{2M_{\rm pl}^2}, \tag{2.121}$$

along with Equation 2.24, we can show that long-wavelength modes during inflation scale as:

$$\mathcal{R}_k \propto \epsilon^0 = \text{constant}. \tag{2.122}$$

The gauge-invariant curvature perturbation remains frozen on scales longer then the Hubble radius during inflation, with $\dot{\mathcal{R}}_k = 0$.

Now we have a complete picture of the behavior of field fluctuations (and corresponding metric perturbations) during inflation, which is summarized in Figure 2-1, which is borrowed from [11]. Fluctuation modes whose wavelength is smaller than the Hubble radius oscillate with decaying amplitude as they are stretched out by the Hubble expansion, as in Equation 2.97. Then any pre-existing inhomogeneities tend to get washed out by inflation as they are redshifted away and stretched out to large scales. As they evolve the sub-Hubble field fluctuations tend towards their minimum energy state which is given by the Bunch-Davies vacuum, as in Equation 2.108. New quantum fluctuations which appear on the shortest length scales as the universe expands are also generated in the Bunch-Davies vacuum, which corresponds to the quantum fluctuations generated by Heisenberg uncertainty in the Minkowski limit. Once fluctuation modes cross outside the Hubble radius ($a > k/H$) they stop oscillating and "freeze out", with the gauge-invariant curvature perturbation $\mathcal{R}_k$ remaining constant through the end of inflation.

2.8 Primordial fluctuations on the CMB

Having characterized the behavior of quantum field fluctuations $\delta\phi$ during inflation, we can give a brief picture of how we can connect the predictions of inflation to measurements of temperature anisotropies on the CMB. A detailed description of how we connect observations of the CMB and large-scale structure to information about the inflationary perturbation spectra is a big task and a bit outside the scope of this thesis, so I'll just give a general picture here, since it will help give us some context in which to understand the initial conditions problem for inflation, which I will focus on in the next chapter.

Let's first take account of the picture we've laid out so far: we've assumed that during inflation field fluctuations and metric perturbations are small enough that we can consider them as linearized quantum fluctuations that we sort of smear over a smooth classical background. This assumption is justified by the fact that the CMB temperature anisotropies we measure are tiny ($\Delta T/T \simeq 10^{-5}$, where the CMB temperature $T_0 = 2.725$ K). Therefore the linearized equations of motion we derive at the end of Sections 2.6.3 and 2.6.4 are a good description for the system of dynamical variables $\varphi(t)$, $a(t)$, $\delta\phi_k(t)$, and $\Psi_k(t)$. In Sections 2.7.2 and 2.6.4 we described how the quantized field fluctuations evolve during inflation: while inside the Hubble radius ($k > aH$) fluctuation modes oscillate with decaying amplitude, evolving toward the Bunch-Davies vacuum as they are stretched out by the Hubble expansion. This tends to wash out any initial inhomogeneous structure during inflation. With the Hubble expansion, as the comoving Hubble radius shrinks during inflation, eventually a fluctuation mode at a comoving length scale k crosses outside the Hubble radius ($k < aH$), and it ceases to oscillate, freezing out to a constant amplitude. Eventually inflation ends (once $\epsilon > 1$ from below) and the dynamical system begins to evolve as in standard cosmology, with the comoving Hubble radius increasing. Therefore, comoving scales which had crossed outside the horizon with the Hubble expansion during inflation will eventually re-enter the horizon at late times, after the end of inflation. And since fluctuation modes had frozen out after crossing outside the horizon, when they re-enter the horizon after inflation has ended they carry information about their primordial state during inflation at the moment they crossed out. Therefore, the anisotropies we measure in the CMB provide a snapshot of those primordial perturbations, at the moment they crossed outside the horizon during inflation, which correspond to the perturbations which re-entered the horizon during recombination.

This basic narrative is summarized in Figure 2-2, which I've borrowed from [9]. It is illustrated that once fluctuations modes cross out, while on super-horizon scales they no longer evolve. After the end of inflation they re-enter the horizon, and those which do so during recombination show up as temperature anisotropies on the CMB, which we measure today. This is represented by the quantity C_ℓ in Figure 2-2. I

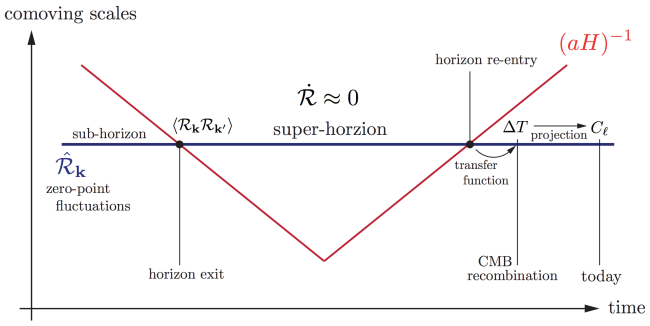

Figure 2-2: The evolution of perturbations before and after inflation. Image taken from [9].

won't go into detail about the CMB observables we analyze statistically, I'll just say that C_ℓ represents a Legendre moment in an angular decomposition of the two-point correlation function of temperature anisotropies, so we can think of it as a direct measure of ΔT. The basic idea of what happens is the fluctuations in the matter density (directly related to the $\delta\phi_k$ and Ψ_k whose evolution we've analyzed) interact with photons through their gravitational potentials. Since up to recombination, at a redshift $z \simeq 1000$, Compton scattering efficiently couples photons to the baryons in a thermalized photon-baryon fluid, fluctuations in the matter density which we've described correspond to fluctuations in the CMB temperature. At recombination, when neutral hydrogen is produced and the photons last scatter they carry the snapshot of these temperature fluctuations, which we observe on the CMB. Of course, relating the primordial fluctuations imprinted on the CMB to the temperature anisotropies we observe requires taking into account details of the time evolution, such as the effects of gravitational redshift on CMB photons, which I'm not going into here, but this is the basic picture.

When we compare the predictions of inflation in this picture we get a set of nearly scale-invariant perturbations which match the power spectrum of CMB temperature anisotropies over a wide range of wavelengths to remarkable accuracy. We typically parameterize the spectrum of primordial perturbations by the spectral index n_s, where a scale-invariant spectrum corresponds to $n_s = 1.0$. Inflationary models generally predict $n_s = 1$ to within $\sim 10\%$ [10], and the latest Planck measurements

find $n_s = 0.965 \pm 0.004$ [12]. The close agreement between the spectrum of primordial inhomogeneities and precision measurements of the CMB is one of the recent triumphs of cosmology, and one of the most compelling reasons for the central role inflation plays in how we describe the physics of the early universe.

2.9 A possible remaining fine-tuning problem

So far we've discussed how successful inflation has been in explaining several observable features of our universe today, including its spatial flatness, large-scale homogeneity, and the specific pattern of temperature anisotropies in the CMB. For additional reviews on inflation you can take a look at Refs. [14–22]. Now let's turn to scrutinizing inflation and investigating a possible remaining fine tuning problem which it leaves to question; that is, *if* inflation occurs it is extremely successful at alleviating the fine-tuning problems left by the standard model of cosmology, but what if inflation *itself* requires fine-tuned initial conditions to occur? For example, some studies have suggested that the onset of inflation may require a large degree of spatial homogeneity, with an initial smooth patch greater than a Hubble volume with radius $r > r_H = H^{-1}$, in order to begin. If this is true, if inflation fails to begin under sufficiently inhomogeneous initial conditions, then its naturalness is challenged: this is the "initial conditions problem". For reviews on this topic you can take a look at Refs. [23, 24].

The formalism we've discussed in this chapter is insufficient for investigating the initial conditions problem. We've treated the primordial inhomogeneities in the matter field and spacetime metric as quantum fluctuations which are small perturbations to a smooth classical FLRW background. We now see that this assumption is justified because *if* inflation occurs sub-Hubble fluctuation modes oscillate with decaying amplitude and tend toward their minimum-energy Bunch-Davies state, washing out any initial inhomogeneous structure until they cross outside the shirinking horizon and freeze out, and it is these fluctuation modes which re-enter the horizon during recombination, providing a snapshot of their primordial state during inflation which

can be seen in temperature anisotropies in the CMB, which we measure to be tiny ($\Delta T/T \simeq 10^{-5}$). By definition this formalism assumes that inflation has proceeded for a few efolds, when all of the information about the potentially inhomogeneous initial state which preceded the onset of inflation has been damped out.

The insufficiency of this formalism for describing the onset of inflation from its initial (pre-inflationary) state is evident in the absence of any coupling between the evolution of field fluctuations, $\delta\phi(\mathbf{x}, t)$, and metric perturbations, $\Psi(\mathbf{x}, t)$, with the classical background, $a(t)$, and $\varphi(t)$ (see Equations 2.67, 2.68, and 2.61 through 2.63). Of course, in order to investigate to what extent inflation can occur under inhomogeneous initial conditions we need to use a formalism that captures the nonlinear effects of backreaction between inhomogeneities and the surrounding spacetime. This will be the subject of Chapter 3.

Chapter 3

A bumpy start to a smooth ride: onset of inflation amid backreaction from inhomogeneities

This chapter is from [1], *by Jolyon K. Bloomfield, me, Kiriakos Hilbert, and David I. Kaiser.*

3.1 Introduction

The question remains whether the onset of inflation itself is fairly generic, or whether inflation requires fine-tuned initial conditions.

Many analyses have highlighted the effectiveness of an inflationary attractor: at least for single-field models, if one neglects spatial inhomogeneities, then one can show that large classes of models will flow into inflation, across broad ranges of initial conditions and couplings, even for regions of phase space in which the system is initially dominated by the field's kinetic (rather than potential) energy. The attractor behavior is especially effective in large-field models with sufficiently flat regions of the potential, for which $\epsilon_n \equiv (M_{\rm pl}\,\partial_\phi)^n \ln V(\phi) \ll 1$, where $M_{\rm pl} \equiv 1/\sqrt{8\pi G} = 2.43 \times 10^{18}$ GeV is the reduced Planck mass [15, 18, 25–32].

These analyses neglected perturbations in the spacetime metric, however, thereby

leaving open the question of whether the onset of inflation remains robust even in the presence of significant inhomogeneities. Pioneering efforts began as early as the mid-1980s to numerically simulate the onset of inflation amid such inhomogeneities [23, 24, 33–41]. Yet these early studies did not yield an unambiguous answer to the question of whether the onset of inflation required an initial patch of spacetime to be smooth across a characteristic length-scale as large as (or perhaps considerably larger than) the Hubble radius. (See also Refs. [42–44].)

Recently three groups have conducted studies of the onset of inflation that implement full $(3+1)$-dimensional numerical relativity; each found that large-field inflation (in single-field models) is strongly robust, even amid significant initial inhomogeneities [45–48]. (They also found that the onset of small-field inflation can be significantly more robust than previous semi-analytic treatments had suggested, although small-field models generically require more finely tuned initial conditions than large-field models do; see also Ref. [49].) In particular, each group found that large-field inflation would start even if the spacetime initially had no Hubble-sized smooth patches. Given the significant computational resources required to simulate the full Einstein field equations in $(3+1)$-dimensions, however, the latest studies have been restricted, to date, to a handful of specific forms of the potential $V(\phi)$ and to limited regions of parameter space [45–47].

In this chapter, we develop a complementary numerical approach. Although by design our approach cannot capture the full range of gravitational effects that could (in principle) affect the onset of inflation, it incorporates certain nonlinear interactions while remaining significantly more efficient than the full-bore numerical relativity studies. It therefore facilitates the study of a wide range of models, across broad regions of phase space and parameter space, while tracking the evolution of as many as $n \sim \mathcal{O}(10^2)$ coupled modes, spanning a wide range of length-scales. (It also offers a feasible means of generalizing to multifield models; see also Ref. [50].) For the case we have investigated so far — a single-field model that yields large-field inflation — our numerical results are consistent with the fully relativistic simulations of Refs. [45–47].

We work to linear order in metric perturbations, $\Psi(x^\mu)$, but incorporate nonlinear

interactions among the field fluctuations, $\delta\phi(x^\mu)$, by adopting the (nonperturbative) Hartree approximation. We therefore capture some of the nonlinear backreaction effects that are absent in studies that work only to linear order in fluctuations. Whereas an ordinary quantum loop expansion corresponds to a power series in the coupling constant λ and in $\hbar$, in the Hartree approximation we resum a particular class of Feynman diagrams (the so-called "cactus" or "daisy" diagrams) to all orders. Hence the Hartree approximation moves beyond perturbative series in λ, $\hbar$, or $\delta\phi$. (See, e.g., Refs. [51–55].)

By working only to linear order in $\Psi(x^\mu)$, our approach cannot capture extreme gravitational phenomena like the formation of trapped surfaces or the collapse of local regions to form black holes. But the fully relativistic simulations in Refs. [45–47] confirm that the formation of such black holes does not interrupt the overall flow of the system into inflation: regions outside of the collapse continue to expand, and the density of any such collapsed regions quickly dilutes. Moreover, as discussed in Ref. [46], no black hole that forms from pre-inflationary overdensities can grow so large as to encompass the entire Hubble sphere, so even the largest black holes that might form fail to disrupt the overall flow into inflation; see also Ref. [56]. Therefore our simplified treatment can complement the more computationally intensive studies.

Another feature of our approach is that we treat the origin of inhomogeneities as ultimately quantum-mechanical. One of the most significant achievements of inflationary cosmology is to provide a first-principles description of the primordial inhomogeneities that seed large-scale structure as arising from quantum-mechanical fluctuations of matter fields during inflation [14–22]. We aim to analyze the onset of inflation in a comparable way; after all, if there were a priori reasons to expect some pattern of classical inhomogeneities at arbitrarily early times, we wouldn't need a mechanism like inflation to account for large-scale structure. Given the high energies involved and the possibility of nonlinear interactions among the quantized matter fields at very early times, we consider scenarios in which the initial state of the quantum fluctuations departs significantly from the minimum-energy (Bunch-Davies) state. The larger quantum fluctuations seed significant spatial inhomogeneities and

affect the dynamics of the system.

Our goal in this chapter is to establish our formalism, introduce details of our numerical approach, and apply our system to a simple model; we defer detailed applications to a wider range of models, including both large-field and small-field models, single-field and multifield cases, to future work. In Section 3.2 we derive the coupled equations of motion for gravitational and matter degrees of freedom within the Hartree approximation. In Section 3.3 we discuss initial conditions for the field fluctuations and our UV regularization scheme. Section 3.4 presents our main numerical results, indicating that across broad regions of phase space, a simple model like $V(\phi) = \lambda\phi^4/4$ will flow into the inflationary attractor even for initial inhomogeneities as large as $|\Psi(x^\mu)| \lesssim 0.5$. In particular, we find a shift in the regions of phase space that support sufficient inflation (compared to the case in which we neglect all perturbations), but no reduction in the total volume of the space of initial conditions $(\varphi(t_0), \dot{\varphi}(t_0))$ that yields sufficient inflation, where φ is the vacuum expectation value of the quantized field that drives inflation. (We consider $N \geq 65$ efolds of inflation to be "sufficient" for addressing the usual shortcomings of standard hot big bang cosmology [15–17, 22, 57, 58].) Concluding remarks follow in Section 3.5. In Appendix A.1 we present more details of our discrete spectrum for the fluctuations. Appendix A.2 provides additional information about how we set initial conditions for the field fluctuations, while Appendix A.3 discusses the initialization of the metric perturbations. In Appendix A.4 we discuss various numerical convergence tests.

3.2 Equations of motion in the Hartree approximation

We work in $(3 + 1)$ spacetime dimensions and use units in which $c = \hbar = 1$. We consider scalar metric perturbations around a background Friedmann-Lemaître-Robertson-Walker (FLRW) line-element, and work in longitudinal gauge,

$$ds^2 = -(1 + 2\Phi)dt^2 + a^2(t)(1 - 2\Psi)h_{ij}(\mathbf{x})\, dx^i\, dx^j, \tag{3.1}$$

where $\Phi(x^\mu)$ and $\Psi(x^\mu)$ are scalar functions. As usual, the background metric on (comoving) spatial sections may be written

$$h_{ij}(\mathbf{x})\, dx^i\, dx^j = \frac{dr^2}{(1 - Kr^2)} + r^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2\right). \tag{3.2}$$

As we will see below, within the Hartree approximation the anisotropic pressure vanishes, so that $\Phi(x^\mu) = \Psi(x^\mu)$.

As is well known, there is no straightforward generalization of longitudinal gauge beyond linear order in $\Psi(x^\mu)$. In particular, at second order in metric perturbations, first-order scalar perturbations source tensor perturbations, and so on. Gauge subtleties therefore affect any perturbative calculation that aims to move beyond $\mathcal{O}(\Psi)$ while incorporating only scalar degrees of freedom. (For a recent discussion, see Appendix B of Ref. [59].) Hence we restrict our analysis to first order in spatially varying quantities.

The restriction to linear order in metric perturbations, treated in longitudinal gauge, can be reasonably well motivated. As demonstrated in Ref. [59], corrections from a fully relativistic treatment compared to a linear treatment in longitudinal gauge typically scale as $\mathcal{O}(\Psi^2)$, unlike the case for synchronous gauge, in which relativistic corrections to the linearized treatment can be as large as $\mathcal{O}(1)$. In fact, as found in Ref. [59], linearized scalar metric perturbations in longitudinal gauge tend to *exaggerate* gravitational effects on length-scales longer than the Hubble radius, $\ell > H^{-1}(t_0)$, compared to a fully relativistic treatment. Given the prior focus on whether inflation can start amid inhomogeneities with typical length-scales $\ell \leq H^{-1}(t_0)$ [23, 24, 33–44], we will be most interested in gravitational effects on sub-Hubble length-scales. In longitudinal gauge, meanwhile, large perturbations, $|\Psi(x^\mu)| \geq 0.5$, can lead to coordinate singularities. Hence the regime of interest is $|\Psi(x^\mu)| \lesssim \mathcal{O}(0.5)$.

We consider single-field models for which the action may be written

$$S = \int d^4x\sqrt{-g}\left[\frac{M_{\rm pl}^2}{2}R - \frac{1}{2}g^{\mu\nu}\,\partial_\mu\phi\,\partial_\nu\phi - V(\phi)\right]. \tag{3.3}$$

Varying the action with respect to ϕ and $g_{\mu\nu}$ yields the coupled equations of motion,

$$\Box\phi - V_{,\phi} = 0 \tag{3.4}$$

and

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{1}{M_{\rm pl}^2}T_{\mu\nu}, \tag{3.5}$$

with

$$T_{\mu\nu} = \partial_\mu\phi\,\partial_\nu\phi - g_{\mu\nu}\left[\frac{1}{2}g^{\alpha\beta}\,\partial_\alpha\phi\,\partial_\beta\phi + V(\phi)\right]. \tag{3.6}$$

As usual, the covariant d'Alembertian operator is given by

$$\Box\phi = \frac{1}{\sqrt{-g}}\partial_\mu\left[\sqrt{-g}\,g^{\mu\nu}\,\partial_\nu\phi\right]. \tag{3.7}$$

In addition to expanding the gravitational degrees of freedom to first order in $\Psi(x^\mu)$, we also consider fluctuations in the field ϕ. Upon quantizing the field, we have

$$\phi(x^\mu) \to \hat{\phi}(x^\mu) = \varphi(t) + \delta\hat{\phi}(x^\mu), \tag{3.8}$$

with

$$\langle\hat{\phi}(x^\mu)\rangle \equiv \langle 0|\hat{\phi}(x^\mu)|0\rangle = \varphi(t)\,,\ \ \langle\delta\hat{\phi}(x^\mu)\rangle = 0. \tag{3.9}$$

We expand the field fluctuations as

$$\delta\hat{\phi}(x^\mu) = \int d\tilde{\mu}(k)\,[\delta\phi_{k\ell m}(t)\,\hat{a}_{k\ell m}\,Z_{k\ell m}(\mathbf{x}) + H.c.]\,, \tag{3.10}$$

where "$H.c.$" denotes Hermitian conjugate, the measure $d\tilde{\mu}(k)$ is given by

$$\int d\tilde{\mu}(k) = \int_0^\infty dk \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}, \tag{3.11}$$

and $Z_{k\ell m}(\mathbf{x})$ is an eigenfunction of the comoving spatial Laplacian operator,

$$\nabla^2 Z_{k\ell m} \equiv \frac{1}{\sqrt{h}}\,\partial_i\left[\sqrt{h}\,h^{ij}\partial_j Z_{k\ell m}\right] = -k^2 Z_{k\ell m}. \tag{3.12}$$

The creation and annihilation operators obey the usual commutation relations

$$\begin{aligned}[\hat{a}_{k\ell m}, \hat{a}^\dagger_{k'\ell'm'}] &= \delta(k-k')\,\delta_{\ell\ell'}\,\delta_{mm'}\,, \\ [\hat{a}_{k\ell m}, \hat{a}_{k'\ell'm'}] &= [\hat{a}^\dagger_{k\ell m}, \hat{a}^\dagger_{k'\ell'm'}] = 0\end{aligned} \tag{3.13}$$

and satisfy

$$\hat{a}_{k\ell m}|0\rangle = 0\,,\ \langle 0|\hat{a}^\dagger_{k\ell m} = 0 \tag{3.14}$$

for all k, ℓ, and m. As we will see, upon expanding the field ϕ as in Eq. (3.8) and working to linear order in Ψ, Eqs. (3.4) and (3.5) will couple $\Psi(x^\mu)$ to terms linear in $\delta\hat{\phi}(x^\mu)$. To remain consistent, we therefore quantize $\Psi(x^\mu) \to \hat{\Psi}(x^\mu)$. Because we are working with single-field models, we only require a single set of operators $\hat{a}_{k\ell m}$ and $\hat{a}^\dagger_{k\ell m}$. We expand $\hat{\Psi}$ as

$$\hat{\Psi}(x^\mu) = \int d\tilde{\mu}(k)\,[\Psi_{k\ell m}(t)\,\hat{a}_{k\ell m}\,Z_{k\ell m}(\mathbf{x}) + H.c.]\,. \tag{3.15}$$

Given our interest in inflationary models, we focus on weakly coupled systems. For such systems, one expects the S matrix to be dominated by forward-scattering processes. Hard-scattering events, which impart large transverse momenta to the scattered particles, should be relatively rare, such that most of the time the particles will emerge from an interaction with (nearly) the same momentum as they had prior to the interaction. The dominant processes, in other words, involve particles propagating along a given trajectory, but moving with a modified, effective mass induced by the (self-)interactions. In such cases, we may approximate the effects of nonlinear self-interactions by calculating a dressed propagator: we select the dominant subset of Feynman diagrams at each order of perturbation theory that contribute to the effective mass, and sum all members of that subclass to all orders, while neglecting the other terms that would appear in a full expansion of the S matrix. (See, e.g., Section 4.7 of Ref. [60].) Within the Hartree approximation, we construct the dressed propagator by performing an infinite resummation of the so-called “cactus" or “daisy" diagrams [51–55].

The Hartree approximation becomes exact for an $O(\mathcal{N})$-symmetric model of $\mathcal{N}$ interacting scalar fields in the limit $\mathcal{N} \to \infty$. In that case, the amplitudes associated with the set of Feynman diagrams picked out for resummation remain parametrically larger than all other terms in the expansion of the S matrix, order by order [51–53]. Yet even in the case $\mathcal{N} = 1$, the diagrams that contribute to the Hartree approximation dominate among the contributions to the effective mass at a given order in perturbation theory, and hence Hartree is often used to incorporate nonperturbative effects even for $\mathcal{N} = 1$, as we will do here. (See, e.g., Refs. [54, 55, 61, 62].)

Operationally, this means that within the equations of motion, all terms that are higher order in the field fluctuations, of the form $(\delta\hat{\phi})^n$ for $n \geq 2$, are replaced by [53–55]

$$\begin{aligned}
(\delta\hat{\phi})^2 &\to \langle(\delta\hat{\phi})^2\rangle, \\
(\delta\hat{\phi})^3 &\to 3\langle(\delta\hat{\phi})^2\rangle\,\delta\hat{\phi}, \\
(\delta\hat{\phi})^4 &\to 3\langle(\delta\hat{\phi})^2\rangle^2, \\
(\delta\hat{\phi})^5 &\to 15\langle(\delta\hat{\phi})^2\rangle^2\,\delta\hat{\phi},
\end{aligned} \tag{3.16}$$

where the particular coefficients on the right-hand side arise from the combinatorics of the various Wick contractions. (One may continue in a similar manner for $(\delta\hat{\phi})^n$ with $n > 5$, though for the particular model of interest to us here, the higher-order terms will not be relevant.) Because the background spacetime around which we are perturbing is homogeneous and isotropic, the (dressed) two-point function $\langle(\delta\hat{\phi})^2\rangle$ is spatially homogeneous.

The Hartree approximation is nonperturbative, so we make no assumption about the relative magnitude of $\langle(\delta\hat{\phi})^2\rangle$ compared to φ^2. In particular, when expanding Eqs. (3.4) and (3.5), we retain terms of the form $\langle(\delta\hat{\phi})^2\rangle\hat{\Psi}$ as well as $\varphi^2\hat{\Psi}$. On the other hand, because we are working only to linear order in $\hat{\Psi}$, we do not include any terms of the form $\langle\hat{\Psi}\,\delta\hat{\phi}\rangle$. Within the Hartree approximation, such terms would arise from summing diagrams involving virtual $\hat{\Psi}$ quanta; yet even the bare propagator for such lines, $\Delta_F(x-y)$, is $\mathcal{O}(\hat{\Psi}^2)$, and hence remains beyond our approximation. To

linear order in metric perturbations, in other words, the perturbations $\hat{\Psi}(x^\mu)$ do not contribute to the Hartree corrections, though, as we will see, the evolution of $\hat{\Psi}(x^\mu)$ is sensitive to the nonlinear evolution of $\delta\hat{\phi}(x^\mu)$.

Lastly, we note that the Hartree approximation is spherically symmetric in k-space; it does not include any direct mode-mode coupling, which would arise from convolutions of the sort $\int d\tilde{\mu}(k')\, d\tilde{\mu}(q)\, \delta\phi_{k-k'-q}\, \delta\phi_{k-k'}\, \delta\phi_k$. (This is consistent with neglecting scattering events that would impart large transverse momenta.) By neglecting such rescattering effects, the Hartree approximation is less efficient at transferring power between modes of different length-scales than a fully nonlinear analysis would be. (See, e.g., the discussion in Ref. [63].) On the other hand, any such terms would require moving beyond linear order in the metric perturbations $\hat{\Psi}(x^\mu)$ — since they would be higher order in spatially varying quantities — and hence the Hartree approximation is especially well-suited for any study that is restricted to linear order in gravitational degrees of freedom.

Our procedure is to expand Eqs. (3.4) and (3.5) to linear order in $\hat{\Psi}$ and to arbitrarily higher order in $(\delta\hat{\phi})^n$; implement the Hartree approximation to replace higher-order terms $(\delta\hat{\phi})^n$ as in Eq. (3.16); and discard any remaining terms that are beyond linear order in spatially varying quantities. Expanding Eq. (3.4), we find a set of terms that are spatially homogeneous, and a set of terms that are linear in spatially varying quantities. Requiring each set to vanish yields the coupled equations of motion:

$$\ddot{\varphi} + 3H\dot{\varphi} + V^{(1)}(\varphi) + \frac{1}{2}V^{(3)}(\varphi)\langle(\delta\hat{\phi})^2\rangle = 0 \tag{3.17}$$

and

$$\begin{aligned} &\delta\ddot{\phi}_{k\ell m} + 3H\delta\dot{\phi}_{k\ell m} \\ &\quad + \left[\frac{k^2}{a^2} + V^{(2)}(\varphi) + \frac{1}{2}V^{(4)}(\varphi)\langle(\delta\hat{\phi})^2\rangle\right]\delta\phi_{k\ell m} \\ &\qquad = 2(\ddot{\varphi} + 3H\dot{\varphi})\Psi_{k\ell m} + 4\dot{\varphi}\dot{\Psi}_{k\ell m}, \end{aligned} \tag{3.18}$$

where overdots denote derivatives with respect to cosmic time t, $H(t) \equiv \dot{a}/a$, and we

use the notation

$$V^{(n)}(\varphi) \equiv \left(\frac{d^n V(\phi)}{d\phi^n} \right) \Bigg|_{\phi=\varphi}. \tag{3.19}$$

(Because we have in mind application to a model with $V(\phi) = \lambda\phi^4/4$ in this paper, no terms with $V^{(n)}(\varphi)$ appear for $n \geq 5$.) The term in Eq. (3.18) proportional to $V^{(4)}(\varphi)\,\langle(\delta\hat{\phi})^2\rangle$, which contributes to the effective mass for the modes $\delta\phi_{k\ell m}$, would not appear if we had performed a perturbative loop expansion. It appears in Eq. (3.18) because the Hartree approximation yields a self-consistent gap equation for the dressed propagator [54, 55].

From the $0i$ component of Eq. (3.5), we find

$$\partial_i \left(\dot{\hat{\Psi}} + H\hat{\Psi} \right) = \frac{1}{2M_{\rm pl}^2} \left(\dot{\varphi} + \delta\dot{\hat{\phi}} \right) \partial_i \delta\hat{\phi}. \tag{3.20}$$

The Hartree approximation stipulates that any terms quadratic in the field fluctuations $\delta\hat{\phi}$ should be replaced by the corresponding vacuum expectation value. In this case, the relevant term would be $\langle \delta\dot{\hat{\phi}}\, \partial_i \delta\hat{\phi} \rangle$, which will vanish: the resulting integrand is an odd function of k_i, integrated over symmetric limits. Hence this term vanishes within the Hartree approximation, and we find

$$\dot{\Psi}_{k\ell m} + H\Psi_{k\ell m} = \frac{1}{2M_{\rm pl}^2} \dot{\varphi}\, \delta\phi_{k\ell m}. \tag{3.21}$$

In a similar way, the anisotropic pressure that could arise from $T_{ij} = \partial_i \delta\hat{\phi}\, \partial_j \delta\hat{\phi}$ (for $i \neq j$) vanishes within the Hartree approximation, which is why the metric perturbations $\Phi(x^\mu)$ and $\Psi(x^\mu)$ in Eq. (3.1) remain equal to each other.

We next expand $T^0_{\ 0} = -\rho$ to find the various contributions to the energy density. We find three distinct contributions: $\bar{\rho}(t)$, which depends only on the (homogeneous) vacuum expectation value of the field, $\varphi(t)$; $\delta\hat{\rho}_{(1)}(x^\mu)$, which includes all terms that are linear in spatially varying quantities; and $\delta\rho_{(2)}(t)$, which includes all spatially

homogeneous terms that arise from the fluctuations:

$$\begin{aligned}
\bar{\rho}(t) &\equiv \frac{1}{2}\dot{\varphi}^2 + V(\varphi), \\
\delta\hat{\rho}_{(1)}(x^\mu) &\equiv \dot{\varphi}\,\delta\dot{\hat{\phi}} - \dot{\varphi}^2\hat{\Psi} - \langle(\delta\dot{\hat{\phi}})^2\rangle\,\hat{\Psi} + \frac{1}{a^2}\hat{\Psi}\langle(\nabla\delta\hat{\phi})^2\rangle \\
&\qquad + V^{(1)}(\varphi)\,\delta\hat{\phi} + \frac{1}{2}V^{(3)}(\varphi)\langle(\delta\hat{\phi})^2\rangle\,\delta\hat{\phi}, \\
\delta\rho_{(2)}(t) &\equiv \frac{1}{2}\langle(\delta\dot{\hat{\phi}})^2\rangle + \frac{1}{2a^2}\langle(\nabla\delta\hat{\phi})^2\rangle \\
&\qquad + \frac{1}{2}V^{(2)}(\varphi)\,\langle(\delta\hat{\phi})^2\rangle + \frac{1}{8}V^{(4)}(\varphi)\langle(\delta\hat{\phi})^2\rangle^2,
\end{aligned} \tag{3.22}$$

where

$$\langle(\nabla\delta\hat{\phi})^2\rangle \equiv h^{ij}\langle\partial_i\delta\hat{\phi}\,\partial_j\delta\hat{\phi}\rangle. \tag{3.23}$$

We see that $\langle\delta\hat{\rho}_{(1)}\rangle = 0$. Upon expanding $G^\mu{}_\nu$ to first order in Ψ and equating the terms from the 00 component of Eq. (3.5) that are spatially homogeneous, we find the effective Friedmann equation

$$H^2 = \frac{1}{3M_{\rm pl}^2}\left[\bar{\rho} + \delta\rho_{(2)}\right] - \frac{K}{a^2}. \tag{3.24}$$

Equating the terms in the 00 component that are linear in spatially varying quantities, we have

$$\begin{aligned}
&-6H\left(\dot{\Psi}_{k\ell m} + H\Psi_{k\ell m}\right) + \frac{2}{a^2}\left(3K - k^2\right)\Psi_{k\ell m} \\
&\quad = \frac{1}{M_{\rm pl}^2}\Bigg\{\dot{\varphi}\,\delta\dot{\phi}_{k\ell m} - \dot{\varphi}^2\Psi_{k\ell m} - \langle(\delta\hat{\phi})^2\rangle\Psi_{k\ell m} \\
&\qquad\quad + \frac{1}{a^2}\Psi_{k\ell m}\langle(\nabla\delta\hat{\phi})^2\rangle + V^{(1)}(\varphi)\,\delta\phi_{k\ell m} \\
&\qquad\qquad + \frac{1}{2}V^{(3)}(\varphi)\langle(\delta\hat{\phi})^2\rangle\,\delta\phi_{k\ell m}\Bigg\}.
\end{aligned} \tag{3.25}$$

From the ij component of Eq. (3.5), we are interested in extracting the spatially homogeneous contributions to the pressure, since these are relevant to the evolution of the background spacetime; in particular, we will use these terms to solve for $\dot{H}$. As usual we may parameterize the pressure as $T^i{}_i = 3p$, and hence, adopting notation as

above, we find

$$\begin{aligned}\bar{p}(t) &\equiv \frac{1}{2}\dot{\varphi}^2 - V(\varphi), \\ \delta p_{(2)}(t) &\equiv \frac{1}{2}\langle(\delta\dot{\hat{\phi}})^2\rangle - \frac{1}{6a^2}\langle(\nabla\delta\hat{\phi})^2\rangle \\ &\quad - \frac{1}{2}V^{(2)}(\varphi)\langle(\delta\hat{\phi})^2\rangle - \frac{1}{8}V^{(4)}(\varphi)\langle(\delta\hat{\phi})^2\rangle^2.\end{aligned} \tag{3.26}$$

(We will not need an explicit expression for $\delta\hat{p}_{(1)}$.) Equating the spatially homogeneous terms in the ij component of Eq. (3.5) yields

$$\left(2\dot{H} + 3H^2 + \frac{K}{a^2}\right) = -\frac{1}{M_{\rm pl}^2}\left[\bar{p} + \delta p_{(2)}\right]. \tag{3.27}$$

Combining Eqs. (3.24) and (3.27), we find

$$\begin{aligned}\dot{H} &= -\frac{1}{2M_{\rm pl}^2}\left[\bar{\rho} + \bar{p} + \delta\rho_{(2)} + \delta p_{(2)}\right] \\ &= -\frac{1}{2M_{\rm pl}^2}\left[\dot{\varphi}^2 + \langle(\delta\dot{\hat{\phi}})^2\rangle + \frac{1}{3a^2}\langle(\nabla\delta\hat{\phi})^2\rangle\right].\end{aligned} \tag{3.28}$$

From Eqs. (3.24) and (3.28), we see that the evolution of the background spacetime depends on the homogeneous field φ as well as on contributions from two-point functions of the fluctuations. A welcome feature of the Hartree approximation is that the Hartree-corrected energy-momentum tensor obeys the covariant conservation relation $\langle T^{\mu\nu}\rangle_{;\nu} = 0$, ensuring that these evolution equations remain mutually consistent with the equations of motion in Eqs. (3.17) and (3.18).

Finally, we may combine our expressions from the 00 and $0i$ components of Eq. (3.5) — in particular, Eqs. (3.21) and (3.25) — to derive a constraint equation relating the modes $\Psi_{k\ell m}$ to $\delta\phi_{k\ell m}$. Upon making algebraic substitutions from Eqs. (3.17) and (3.28), we find

$$\begin{aligned}&\left[\dot{H} + \frac{2}{3M_{\rm pl}^2a^2}\langle(\nabla\delta\hat{\phi})^2\rangle + \frac{1}{a^2}\left(k^2 - 3K\right)\right]\Psi_{k\ell m} \\ &\qquad = \frac{1}{2M_{\rm pl}^2}\left[\ddot{\varphi}\,\delta\phi_{k\ell m} - \dot{\varphi}\,\delta\dot{\phi}_{k\ell m}\right].\end{aligned} \tag{3.29}$$

In our numerical simulations, we use Eq. (3.29) only to set initial conditions for the modes $\Psi_{k\ell m}(t_0)$, based on the initial conditions for $\varphi(t_0)$, $\dot{\varphi}(t_0)$, $H(t_0)$, $\delta\phi_{k\ell m}(t_0)$, and $\delta\dot{\phi}_{k\ell m}(t_0)$; we then evolve the metric perturbations over time using Eq. (3.21). Although the source term in Eq. (3.21) is linear in $\delta\phi_{k\ell m}$, the evolution of $\varphi(t)$, $H(t)$, and $\delta\phi_{k\ell m}(t)$ each incorporates nonlinear backreaction effects from the Hartree corrections.

When working to linear order in $\hat{\Psi}$ and $\delta\hat{\phi}$, it is common to study the evolution of the gauge-invariant comoving curvature perturbation, $\hat{\mathcal{R}}(x^\mu)$, which (for single-field models) takes the form [16, 17]

$$\hat{\mathcal{R}} = \hat{\Psi} + \frac{H}{\dot{\varphi}}\,\delta\hat{\phi}. \tag{3.30}$$

(As is well-known, $\hat{\mathcal{R}}$ is proportional to the gauge-invariant Mukhanov-Sasaki variable, and is equivalent to the curvature perturbation on uniform-density hypersurfaces, $\hat{\zeta}$, in the limit $k \ll aH$ [16, 17].) Although $\hat{\mathcal{R}}$ only remains gauge-invariant for linear gauge transformations, it remains a useful quantity for considering the evolution of perturbations even when we incorporate the nonlinear, nonperturbative Hartree corrections, as we will see in Section 3.4.

3.3 Setting parameters

In our numerical simulations, we track the evolution of coupled modes within a finite (comoving) spatial volume; this restriction, in turn, means that for any Gaussian curvature K of the background spatial sections, we have a discrete spectrum of allowable wavenumbers. Then the integral over dk in the measure $d\tilde{\mu}(k)$ defined in Eq. (3.11) is replaced by a discrete sum, indexed by an integer $n \geq 1$:

$$\delta\hat{\phi}(x^\mu) = \sum_{n,\ell,m} \left[\delta\phi_{n\ell m}(t)\,\hat{a}_{n\ell m}\,Z_{n\ell m}(r,\theta,\phi) + H.c.\right], \tag{3.31}$$

and similarly for $\hat{\Psi}(x^\mu)$. In place of Eq. (3.13), the creation and annihilation operators now obey $[\hat{a}_{n\ell m}, \hat{a}^\dagger_{n'\ell'm'}] = \delta_{nn'}\delta_{\ell\ell'}\delta_{mm'}$, with $\hat{a}_{n\ell m}|0\rangle = \langle 0|\hat{a}^\dagger_{n\ell m} = 0$ for all $(n\ell m)$.

For the remainder of this paper we consider $K = 0$ and evolve the modes within a sphere of comoving radius R. (We defer the case of nonzero K to future work.) As described in Appendix A.1, for $K = 0$ the normalized spatial eigenfunctions $Z_{n\ell m}(r, \theta, \phi)$ take the form

$$Z_{n\ell m}(r, \theta, \phi) = N_{n\ell m}\, j_\ell(k_{n\ell} r)\, Y_{\ell m}(\theta, \phi). \tag{3.32}$$

Here $N_{n\ell m}$ is a normalization constant, $j_\ell(x)$ is the spherical Bessel function, and $Y_{\ell m}(\theta, \phi)$ is the usual spherical harmonic. We choose to use Dirichlet boundary conditions at $r = R$, which fixes $Z_{n\ell m}(R, \theta, \phi) = 0$ for all (θ, ϕ), which in turn restricts the allowable wavenumbers $k_{n\ell}$ to a discrete spectrum, namely

$$k_{n\ell}(R) \equiv \frac{x_{n\ell}}{R}, \tag{3.33}$$

where $x_{n\ell}$ is the nth zero of the Bessel function $j_\ell(x)$, that is, $j_\ell(x_{n\ell}) = 0$ for $n \geq 1$. (For $\ell = 0$, the k_{n0} take the simple form, $k_{n0} = n\pi/R$.)

Within the Hartree approximation, the evolution of $\varphi(t)$ and $H(t)$, as well as the modes $\Psi_{n\ell m}(t)$, depends on the initial conditions for the modes $\delta\phi_{n\ell m}(t_0)$ and $\delta\dot{\phi}_{n\ell m}(t_0)$. Because the Hartree approximation replaces higher-order interaction terms in the equation of motion for the fluctuations $\delta\hat{\phi}$ by an effective mass, we may follow many of the usual steps for quantizing a free scalar field in FLRW spacetime, and use this quantization procedure to parameterize initial conditions for $\delta\phi_{n\ell m}(t_0)$ and $\delta\dot{\phi}_{n\ell m}(t_0)$.

In the regime of interest, the field fluctuations are nearly massless around t_0. From Eq. (3.18), the effective mass is given by

$$m^2_{\text{eff}}(t) = V^{(2)}(\varphi) + \frac{1}{2}V^{(4)}(\varphi)\langle(\delta\hat{\phi})^2\rangle, \tag{3.34}$$

which is suppressed by the small coupling constant λ; hence we have $m^2_{\text{eff}}(t_0) \ll$

$H^2(t_0)$. Moreover, when we evolve the modes within a sphere of comoving radius R, we introduce an infrared cut-off given by $k_{\rm min} = \pi/R$, with $R \sim 1/[a(t_0)\, H(t_0)]$. Even for the longest-wavelength modes in our simulation, we therefore have $k_{n\ell}^2/a^2(t_0) \gtrsim m_{\rm eff}^2(t_0)$.

We do not assume that the system has attained the mininum energy state at the initial time t_0, and hence we consider initial conditions for $\delta\phi_{n\ell m}(t_0)$ and $\delta\dot{\phi}_{n\ell m}(t_0)$ that depart from the usual Bunch-Davies vacuum state. As described in Appendix A.2, we parameterize

$$\begin{aligned} \delta\phi_{n\ell m}(t_0) &= \frac{\alpha_{n\ell m}}{\sqrt{2k_{n\ell}}}, \\ \delta\dot{\phi}_{n\ell m}(t_0) &= \sqrt{\frac{k_{n\ell}}{2}} \left(-i\gamma_{n\ell m} + \delta_{n\ell m} - \frac{\alpha_{n\ell m} \bar{H}_0}{k_{n\ell}} \right), \end{aligned} \tag{3.35}$$

where $\alpha_{n\ell m}$, $\gamma_{n\ell m}$, and $\delta_{n\ell m}$ are each real-valued dimensionless constants and $\bar{H}_0$ is given by

$$\bar{H}_0^2 \equiv \frac{\bar{\rho}(t_0)}{3M_{\rm pl}^2}. \tag{3.36}$$

From Eq. (3.22) we note that $\bar{\rho}(t_0)$ is the energy density associated with the vacuum expectation value at the initial time, $\varphi(t_0)$. For the quantum fluctuations, the equal-time commutation relation at t_0 requires

$$\alpha_{n\ell m} = \frac{1}{\gamma_{n\ell m}} \tag{3.37}$$

for all $(n\ell m)$. The Bunch-Davies initial state corresponds to $\gamma_{n\ell m} = 1$ and $\delta_{n\ell m} = 0$ for all $(n\ell m)$. To depart from the Bunch-Davies initial state, we treat the coefficients $\gamma_{n\ell m}$ and $\delta_{n\ell m}$ as random variables for each mode, drawn from flat distributions within specific ranges. Once $\gamma_{n\ell m}$ is drawn for a given mode, $\alpha_{n\ell m}$ for that mode is fixed from Eq. (3.37).

In addition to selecting initial conditions for individual modes $\delta\phi_{n\ell m}(t_0)$ and $\delta\dot{\phi}_{n\ell m}(t_0)$, we also need to evaluate the various two-point functions that appear in the evolution equations for $\varphi(t)$, $\delta\phi_{n\ell m}(t)$, $H(t)$, and $\dot{H}(t)$; only then can we set initial conditions for the metric perturbation modes $\Psi_{n\ell m}(t_0)$ and evolve the coupled sys-

tem forward in time. As discussed in Appendix A.1, when we evaluate the two-point functions within a sphere of comoving radius R, only modes with $\ell = 0$ contribute to $\langle(\delta\hat{\phi})^2\rangle$ and $\langle(\delta\dot{\hat{\phi}})^2\rangle$, whereas only modes with $\ell = 1$ contribute to $\langle(\nabla\delta\hat{\phi})^2\rangle$, and we find

$$\begin{aligned}
\langle(\delta\hat{\phi})^2\rangle &= \frac{\pi}{2R^3}\sum_n n^2|\delta\phi_{n00}(t)|^2, \\
\langle(\delta\dot{\hat{\phi}})^2\rangle &= \frac{\pi}{2R^3}\sum_n n^2|\delta\dot{\phi}_{n00}(t)|^2, \\
\langle(\nabla\delta\hat{\phi})^2\rangle &= \frac{1}{6\pi R^3}\sum_n \sum_{m=-1,0,1} \frac{k_{n1}^2}{|j_2(k_{n1}R)|^2}|\delta\phi_{n1m}(t)|^2.
\end{aligned} \tag{3.38}$$

Although we have considered the $K = 0$ case here, the pattern of which ℓ modes contribute to which two-point functions holds for arbitrary K, since the (θ, ϕ) portion of the background metric in Eq. (3.2) does not depend on K.

Our expressions for the various two-point functions in Eq. (3.38) diverge in the limit $n \to \infty$; this is just the usual ultraviolet divergence that appears in the continuum limit for $k \to \infty$. Hence we must regularize all sums that appear in the various two-point functions. Since we expand the quantum fluctuations $\delta\hat{\phi}(x^\mu)$ as sums over modes of comoving wavenumber $k_{n\ell}$, we adopt a simple Gaussian regulator with a comoving UV scale κ:

$$F_{n\ell}(\kappa, R) = \exp\left[-\frac{k_{n\ell}^2(R)}{(2\kappa)^2}\right]. \tag{3.39}$$

We insert $F_{n\ell}(\kappa, R)$ within the sums when evaluating all two-point functions. For example,

$$\langle(\delta\hat{\phi})^2\rangle \to \langle(\delta\hat{\phi})^2\rangle_{\rm reg} = \frac{\pi}{2R^3}\sum_n n^2|\delta\phi_{n00}(t)|^2 F_{n0}(\kappa, R). \tag{3.40}$$

We use a single UV regulator scale κ, independent of ℓ. Once κ is fixed, the sums over n may be truncated at some finite number of modes, $n_{\rm max}$.

The regularized two-point functions depend on the UV regulator scale κ. In general, the two-point function for quantum fluctuations of a nearly massless scalar field in an FLRW background scales as $\langle(\delta\hat{\phi})^2\rangle \propto H^2$. We therefore parameterize $\kappa = b\bar{H}_0$, with b a real, dimensionless constant. To confirm the scaling of the regularized

two-point function with H, we use the fact that the random coefficients $\alpha_{n\ell m}$ for the mode functions $\delta\phi_{n\ell m}(t_0)$ vary independently of n, so we may take the term α_{n00}^2 out of the sum in Eq. (3.40) at t_0 and replace it by an average value. As detailed in Appendix A.2, this yields

$$\langle(\delta\hat{\phi}(t_0))^2\rangle_{\rm reg} \simeq (\alpha^2)_{\rm avg}\frac{b^2\bar{H}_0^2}{4\pi^2}. \tag{3.41}$$

(Although only modes with $\ell = m = 0$ contribute to the sum in Eq. (3.40), we draw the random variables $\gamma_{n\ell m} = 1/\alpha_{n\ell m}$ from the same distribution for all $(n\ell m)$, and hence the average value $(\alpha^2)_{\rm avg}$ holds for any n, ℓ, and m.) Following similar steps (see Appendix A.2) we find

$$\delta\rho_{(2)}(t_0) \simeq \frac{b^4\mathcal{C}}{4\pi^2}\bar{H}_0^4. \tag{3.42}$$

with

$$\mathcal{C} \equiv \left(\alpha^2 + \gamma^2 + \delta^2\right)_{\rm avg}. \tag{3.43}$$

The quantity $\mathcal{C}$ measures how strongly (on average) the amplitude of the initial state of the quantum fluctuations deviates from the minimum-energy (Bunch-Davies) state. Since $\gamma_{n\ell m} = \alpha_{n\ell m} = 1$ and $\delta_{n\ell m} = 0$ for all $(n\ell m)$ in the Bunch-Davies state, $\mathcal{C}_{\rm BD} = 2$. Throughout our analysis, we consider quantum fluctuations whose average initial energy density exceeds the Bunch-Davies value by an order of magnitude, with $\mathcal{C} \simeq 20$, by drawing the random coefficients for each mode from flat distributions within the ranges

$$\gamma_{n\ell m} \in \{0.09, 1\}\,, \;\; \delta_{n\ell m} \in \{-5, 5\}. \tag{3.44}$$

Once $\gamma_{n\ell m}$ is drawn for a given mode, $\alpha_{n\ell m}$ for that mode is fixed from Eq. (3.37). The ranges in Eq. (3.44) yield $(\alpha^2)_{\rm avg} = 11.11$, $(\gamma^2)_{\rm avg} = 0.37$, and $(\delta^2)_{\rm avg} = 8.33$.

The actual value of the initial Hubble scale $H_0 \equiv H(t_0)$ will be greater than $\bar{H}_0$, because H includes contributions from both $\bar{\rho}$ and $\delta\rho_{(2)}$, as in Eq. (3.24). We define

$f \equiv (H_0/\bar{H}_0) > 1$; using Eqs. (3.36) and (3.42) we find that on average

$$f_{\rm avg} = \left[1 + \frac{b^4 \mathcal{C}}{12\pi^2}\left(\frac{\bar{H}_0}{M_{\rm pl}}\right)^2\right]^{1/2} . \tag{3.45}$$

We aim to study initial conditions for the system such that $H_0 \gg H_{\rm infl}$, where $H_{\rm infl}$ is the Hubble scale of the slow-roll inflationary attractor. Yet we also need to keep H_0 low enough (compared to $M_{\rm pl}$) so that we can begin the simulations with significant power in fluctuations on sub-Hubble scales, $H_0 < k/a(t_0) < M_{\rm pl}$, while avoiding trans-Planckian ambiguities. Hence we focus on initial conditions such that $H_0 \sim 0.1\, M_{\rm pl}$. For perturbations that depart from the Bunch-Davies initial state, with $2 \leq \mathcal{C} \leq 20$, we find $\delta\rho_{(2)}(t_0) \lesssim \bar{\rho}(t_0)$ for $\bar{H}_0 = 0.1\, M_{\rm pl}$ if we fix $b = 5$. Eq. (3.45) then yields $f_{\rm avg} \leq 1.4$, corresponding to $H_0 \lesssim 0.14\, M_{\rm pl}$.

Because we are interested in effects of initial inhomogeneities on length-scales shorter than the initial Hubble radius, we fix the comoving radius $R = 1.5\pi\bar{H}_0^{-1} > 1.5\pi H_0^{-1}$. (We set $a(t_0) = 1$.) Then the longest comoving wavelength in the spectrum is $\lambda_{\rm max} = 2R > 3\pi H_0^{-1}$, corresponding to $k_{\rm min} = \pi/R$. This ensures that our longest wavelength modes begin on a superhorizon scale, but that most of our modes are initially subhorizon. Given the form of $F_{n\ell}(\kappa, R)$ in Eq. (3.39), meanwhile, we find strong suppression of the contribution to the various two-point functions by modes with comoving wavenumber $k_{n\ell} > k_{UV} = 3\kappa$, or $k_{UV} = 3b\bar{H}_0$. Setting $b = 5$, we find $k_{UV} \sim \mathcal{O}(10H_0) \sim M_{\rm pl}$. To ensure strong numerical convergence we fix $k_{\rm max} = 4\kappa$, which corresponds to $n_{\rm max} = 30$. Our simulations then involve $4n_{\rm max} = 120$ coupled modes: $n_{\rm max}$ each for $\ell = 0$ and for $\ell = 1$ with $m = -1, 0, 1$.

We implement our UV regularization via Eq. (3.39), but do not pursue formal renormalization. For one thing, we are interested in scenarios in which the system begins at high energies $H_0 \sim 0.1\, M_{\rm pl}$, and we aim to track effects of excited modes with wavenumbers up to $k_{UV} \sim M_{\rm pl}$; hence there are no arbitrarily large hierarchies between the physical energy scales of interest and the natural cut-off scale. More important, by studying initial states for the fluctuations that depart from the usual Bunch-Davies state, as in Eq. (3.35), any renormalization scheme would depend *both*

on the coupling constants in the Lagrangian *and* on the particular selection of initial parameters $\alpha_{n\ell m}, \gamma_{n\ell m}$, and $\delta_{n\ell m}$ for the modes. Hence any renormalization scheme would change, run by run, with the random draws for these parameters. (See, e.g., Refs. [64, 65].) We therefore defer questions about formal renormalization to future work.

3.4 Numerical results

In this section we first consider typical behavior of the coupled system for a particular set of initial conditions, before turning to a more systematic investigation across initial conditions. We study $V(\phi) = \lambda\phi^4/4$ with $\lambda = 10^{-10}$, and later compare results for $\lambda = 10^{-12}$.

When one ignores field fluctuations and metric perturbations, this model yields sufficient inflation, with $N_{\rm infl} \geq 65$ efolds, for $\varphi_0 \equiv \varphi(t_0) \sim 15 - 30\, M_{\rm pl}$, depending on the initial value of $\dot{\varphi}_0 \equiv \dot{\varphi}(t_0)$. For $\lambda = 10^{-10}$, this corresponds to a value of the Hubble parameter once the system has reached the slow-roll inflationary attractor (with $\dot{\varphi}^2 \ll V(\varphi)$) of $H_{\rm infl} \sim 10^{-3}\, M_{\rm pl}$; for $\lambda = 10^{-12}$, we have $H_{\rm infl} \sim 10^{-4}\, M_{\rm pl}$. We are therefore interested in the behavior of this system in the vicinity of $\varphi_0 \sim 15 - 30\, M_{\rm pl}$ when the nonlinear effects of fluctuations are incorporated into the dynamics.

For each value of λ, we varied $12\, M_{\rm pl} \leq \varphi_0 \leq 30\, M_{\rm pl}$ in steps of $\Delta\varphi_0 = 0.25\, M_{\rm pl}$, and $-0.25\, M_{\rm pl}^2 \leq \dot{\varphi}_0 \leq 0.25\, M_{\rm pl}^2$ in steps of $\Delta\dot{\varphi}_0 = 0.01\, M_{\rm pl}^2$, for a 73×51 search grid. For each grid point in $(\varphi_0, \dot{\varphi}_0)$, we computed 34 simulations: one with the Hartree corrections turned off (to neglect coupled fluctuations), one with the quantum fluctuations $(\delta\phi_{n\ell m}(t_0), \delta\dot{\phi}_{n\ell m}(t_0))$ initialized in the Bunch-Davies initial state with $\mathcal{C}_{\rm BD} = 2$, and 32 in which we initialized the quantum fluctuations with $\mathcal{C} \simeq 20$ by drawing random coefficients $\gamma_{n\ell m}$ and $\delta_{n\ell m}$ for each mode from the ranges in Eq. (3.44). This yielded a total of roughly 250,000 individual simulations. The simulations were performed on the Amazon Web Services EC2 service on a 16-core instance, and took a little under two days to complete. Our code was implemented in Python.

Our simulations are initialized in a number of steps. Starting from the values

$(\varphi_0, \dot{\varphi}_0)$, we construct $\bar{H}_0$ as in Eq. (3.36), from which R, κ, and $k_{\rm max}$ are computed. The spectrum of allowable wavenumbers $k_{n\ell}(R)$ with $\ell = 0$ and $\ell = 1$ is then constructed. Next, we construct the Bunch-Davies initial conditions for each mode $\delta\phi_{n\ell m}(t_0)$ and $\delta\dot{\phi}_{n\ell m}(t_0)$. For perturbed initial data, we draw random values for $\gamma_{k\ell m}$ and $\delta_{k\ell m}$ for each mode, and construct the initial mode data appropriately from Eq. (3.35). We then compute the appropriate two-point functions, initialize the modes $\Psi_{n\ell m}(t_0)$ from Eq. (3.29), and evaluate $\delta\rho_{(2)}(t_0)$ to construct the actual initial Hubble factor H_0.

Our simulations evolve the quantities $a(t)$, $\varphi(t)$, and $\dot{\varphi}(t)$. For each wavenumber $k_{n\ell}$ with $\ell = 0$ and $\ell = 1$, we also evolve two modes: one initialized with $(\delta\phi_{n\ell 0}(t_0), \delta\dot{\phi}_{n\ell 0}(t_0)) = (1, 0)$ and a second with $(0, 1)$ (along with a corresponding initialization for $\Psi_{n\ell 0}(t_0)$ for each mode). Given the time evolution of these modes and the initial conditions $\delta\phi_{n\ell m}(t_0)$, $\delta\dot{\phi}_{n\ell m}(t_0)$, and $\Psi_{k\ell m}(t_0)$, every mode can be reconstructed by exploiting the linearity of the equations of motion, since the nonlinear two-point functions are effectively functions of time that are independent of n, ℓ and m. Note that the modes that are evolved remain real. Doing this split makes little difference for the $\ell = 0$ modes, but reduces the computational cost threefold for the $\ell = 1$ modes. We evolve the system forward in time using a variable time-step RK45 algorithm. We declare that slow-roll begins once $\epsilon < 0.1$, and we terminate evolution thereafter at $\epsilon \geq 1$, where

$$\epsilon \equiv -\frac{\dot{H}}{H^2}. \tag{3.46}$$

3.4.1 Evolution of perturbations

In this subsection and the following, we present results from a representative sample of random initializations. We set $\lambda = 10^{-10}$ and consider the case $\varphi_0 = 25\, M_{\rm pl}$ and $\dot{\varphi}_0 = -0.25\, M_{\rm pl}^2$, for which the initial kinetic energy in the field φ exceeds the potential energy by a factor of 3200. We initialize the fluctuations $\delta\hat{\phi}(t_0, \mathbf{x})$ by parameterizing the mode functions as in Eq. (3.35) and drawing the random initial coefficients $\gamma_{n\ell m}$ and $\delta_{n\ell m}$ for each mode from the distributions in Eq. (3.44). Then $\mathcal{C} \simeq 20$ and the energy density in fluctuations $\delta\rho_{(2)}(t_0)$ begins about ten times greater than in the

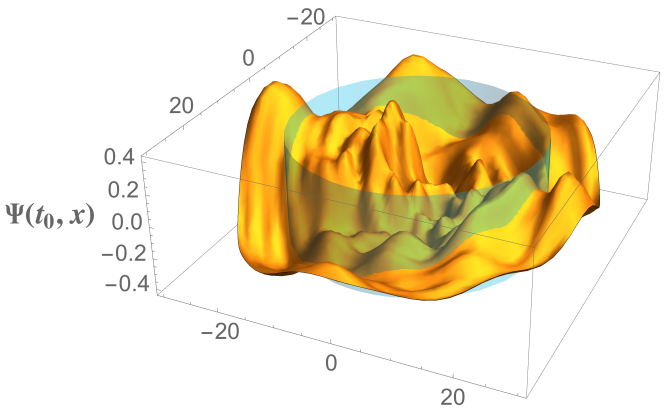

Figure 3-1: Typical initial surface of $\Psi(t_0, \mathbf{x})$ in the x-y plane (dependence on the polar angle θ in Eq. (3.2) is suppressed) for $\varphi_0 = 25\, M_{\rm pl}$ and $\dot{\varphi}_0 = -0.25\, M_{\rm pl}^2$. The field fluctuations $\delta\hat{\phi}(t_0, \mathbf{x})$ were initialized as in Eq. (3.35), with the random coefficients $\gamma_{n\ell m}$ and $\delta_{n\ell m}$ for each mode drawn from the ranges in Eq. (3.44), which in turn determined the modes $\Psi_{n\ell m}(t_0)$ from Eq. (3.29). Further details about the construction of the initial surface for $\Psi(t_0, \mathbf{x})$ are given in Appendix A.3. The blue cylinder has radius equal to $r_H(t_0) = \pi/H(t_0)$, in units of $M_{\rm pl}^{-1}$.

Bunch-Davies initial state. With these parameters, we find $\delta\rho_{(2)}(t_0) \simeq \bar{\rho}(t_0)$ and $H_0 = 0.14\, M_{\rm pl}$, two orders of magnitude greater than $H_{\rm infl}$.

For these initial conditions, the system begins with significant inhomogeneities on length-scales well within the initial Hubble radius. Fig. 3-1 shows $\Psi(t_0, \mathbf{x})$, constructed from modes $\Psi_{n\ell m}(t_0)$ whose amplitudes are set by Eq. (3.29). (Further details of how we construct $\Psi(t_0, \mathbf{x})$ are given in Appendix A.3.) The blue cylinder in Fig. 3-1 has a radius equal to $r_H(t_0) = \pi/H(t_0)$, such that modes with $k \geq a(t_0)H(t_0)$ have wavelengths that fit within the diameter $2r_H(t_0)$. For this choice of initial conditions, the metric perturbations $\Psi(t_0, \mathbf{x})$ begin with substantial structure on sub-Hubble length-scales, with spatial inhomogeneities as large as $|\Psi(t_0, \mathbf{x})| \simeq 0.4$.

Figures 3-2 and 3-3 show $|\mathrm{Re}(\delta\phi_{n\ell m})|$ and $|\mathrm{Re}(\Psi_{n\ell m})|$ versus $N \equiv \int H dt = \ln a$ for early times, for the $\ell = 0$ and $\ell = 1$ modes with $k_{n\ell} \geq a_0 H_0$. As expected, the modes oscillate with decaying amplitude while inside the Hubble radius, and their amplitudes freeze after Hubble crossing. At later times, after the physical wavelengths of the modes have redshifted to be exponentially larger than the Hubble radius, the amplitudes show a modest secular growth, rising as $|\delta\phi_{n\ell m}| \propto \sqrt{\epsilon}$ and $|\Psi_{n\ell m}| \propto \epsilon$,

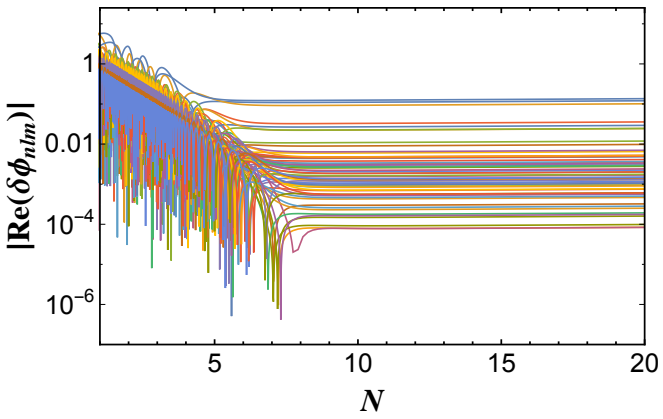

Figure 3-2: $|\mathrm{Re}(\delta\phi_{n\ell m})|$ versus $N = \ln a$ for the $\ell = 0, 1$ modes in our simulation with $k_{n\ell} \geq a_0 H_0$.

where ϵ is given in Eq. (3.46). This modest late-time growth matches the well-known behavior of perturbations deep in the infrared during the slow-roll regime, as treated in linear perturbation theory. (See, e.g., Section 8.2 of Ref. [15], as well as Ref. [66].) Nonetheless, the curvature perturbation, $\hat{\mathcal{R}}$, defined in Eq. (3.30), remains conserved once modes cross outside the Hubble radius. In Fig. 3-4 we plot the dimensionless power spectrum for the curvature perturbation,

$$\mathcal{P}_{\mathcal{R}}(k_{n\ell}) \equiv \frac{k_{n\ell}^3}{2\pi^2} |\mathcal{R}_{n\ell m}|^2 \tag{3.47}$$

for $\ell = 0$ and $\ell = 1$ modes that begin with $k_{n\ell} \geq a_0 H_0$. Consistent with the analytic results in Ref. [67], we find that the curvature perturbation remains conserved on super-Hubble length-scales, even when we incorporate nonlinear self-interactions. As shown in Fig. 3-4, substantial structure on sub-Hubble length-scales at early times damps out before modes cross outside the Hubble radius, producing a smooth patch on horizon scales, and remains exponentially suppressed for the duration of the simulation.

The energy density in fluctuations $\delta\rho_{(2)}$ begins with $\delta\rho_{(2)}(t_0) \simeq \bar{\rho}(t_0)$ and then begins to decay, as shown in Fig. 3-5. Because of the weak coupling λ, the effective mass for the fluctuations satisfies $m_{\mathrm{eff}}(t) \ll H(t)$ at early times, where m_{eff} is given in Eq. (3.34). While most modes are inside the Hubble radius, with $k/a > H \gg m_{\mathrm{eff}}$,

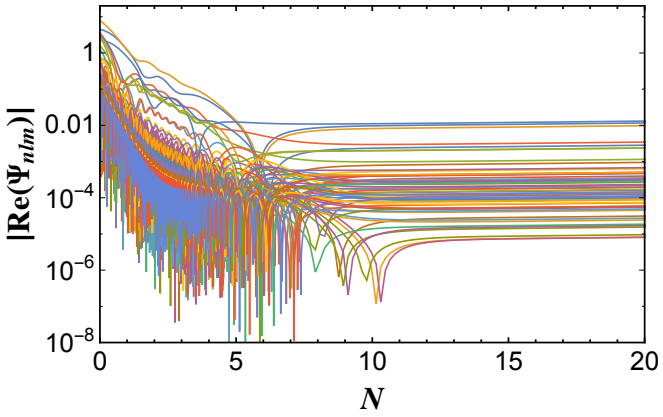

Figure 3-3: $|\mathrm{Re}(\Psi_{n\ell m})|$ versus N for the $\ell = 0, 1$ modes in our simulation with $k_{n\ell} \geq a_0 H_0$.

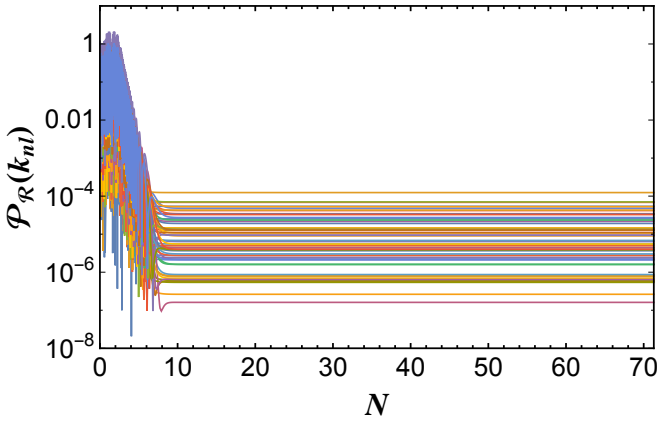

Figure 3-4: The dimensionless power spectrum of the curvature perturbation $\mathcal{P}_{\mathcal{R}}(k_{n\ell})$ versus N for the $\ell = 0, 1$ modes in our simulation with $k_{n\ell} \geq a_0 H_0$.

their energy density therefore evolves like a gas of (nearly) massless particles, with an equation of state like radiation, $\delta\rho_{(2)}(t) \propto a^{-4}(t)$. At later times, after the modes have crossed outside the Hubble radius and their amplitudes have frozen, $\delta\rho_{(2)}(t)$ becomes constant.

Next we consider the impact of these large initial inhomogeneities on the evolution of the Hubble parameter $H(t)$, shown in Fig. 3-6. The figure shows $H(t)$ for the same initial values $\varphi_0 = 25\, M_{\rm pl}$ and $\dot{\varphi}_0 = -0.25\, M_{\rm pl}^2$, when we neglect fluctuations (blue); when we initialize the fluctuations in the Bunch-Davies state, with $\mathcal{C}_{\rm BD} = 2$ (yellow); and for a particular simulation in which we initialized the system with large initial fluctuations, $\mathcal{C} = 20$ (pink). The energy density associated with φ, $\bar{\rho}(t)$, is dominated

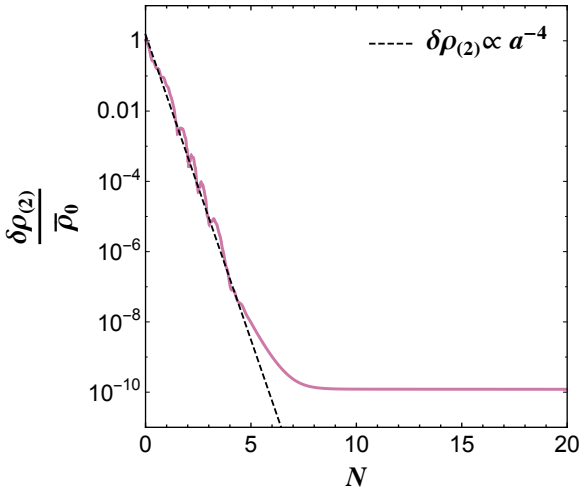

Figure 3-5: The energy density in fluctuations $\delta\rho_{(2)}(t)$ (normalized by the initial value $\bar{\rho}(t_0)$) versus N, with $\lambda = 10^{-10}$. At early times, while most modes are still inside the Hubble radius, $\delta\rho_{(2)}(t)$ decays like radiation.

at early times by the kinetic energy of φ, and hence it decays as $\bar{\rho}(t) \propto a^{-6}(t)$. When we neglect fluctuations, we therefore find $H(t) \propto a^{-3}(t)$ at early times. On the other hand, for large initial fluctuations with $\mathcal{C} = 20$ and hence $\delta\rho_{(2)}(t_0) \simeq \bar{\rho}(t_0)$, we find $H(t) \propto [\bar{\rho}(t) + \delta\rho_{(2)}(t)]^{1/2} \propto a^{-2}(t)$ at early times, while most fluctuations remain sub-Hubble and $\delta\rho_{(2)}(t)$ decays like radiation. For fluctuations that begin in the Bunch-Davies initial state, with $\mathcal{C}_{\rm BD} = 2$ and $\delta\rho_{(2)}(t_0) \sim 0.1\,\bar{\rho}(t_0)$, we find an evolution for $H(t)$ intermediate between these two cases. (The authors of Ref. [45] likewise found the volume-averaged quantities $\rho_{\rm avg}(t) \propto [a_{\rm avg}(t)]^{-4}$ and $H_{\rm avg}(t) \propto [a_{\rm avg}(t)]^{-2}$ at early times in their numerical simulations of large-field models with significant initial inhomogeneities. See also Ref. [32].)

The system begins to inflate, with $\ddot{a} > 0$, once $\epsilon < 1$. For the set of initial conditions we consider here, inflation begins by $N \sim 2$, and the system enters a phase of slow-roll inflation ($\epsilon < 0.1$) by $N \sim 3$. During slow-roll, $\bar{\rho}$ is dominated by $V(\varphi)$, while $\delta\rho_{(2)}$ continues to redshift like radiation until most of the modes have crossed outside the Hubble radius, by $N \sim 7$ (as shown in Fig. 3-5). Hence after slow-roll inflation begins, the system evolves with $\bar{\rho} \simeq V(\varphi) \gg \delta\rho_{(2)}$ and $H(t)$ settles onto a nearly constant value at $H_{\rm infl} \sim 10^{-3}\,M_{\rm pl}$.

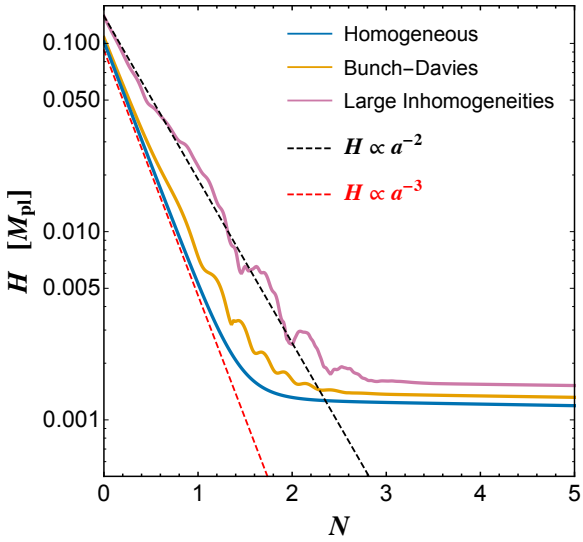

Figure 3-6: The evolution of the Hubble parameter $H(t)$ versus N for early times, with $\lambda = 10^{-10}$. In the absence of fluctuations (blue), $H(t) \propto a^{-3}(t)$. For large initial fluctuations (pink), $H(t) \propto a^{-2}(t)$ at early times. For fluctuations that begin in the Bunch-Davies initial state (yellow), the evolution of $H(t)$ falls between the other two cases.

As shown in Fig. 3-7, for the case with large initial fluctuations, the system reaches the slow-roll inflationary attractor while most of the power in fluctuations remains inside the Hubble radius. In the presented simulation, when slow-roll inflation begins (with $\epsilon \leq 0.1$), all of the modes that had begun inside the Hubble radius still remain inside the Hubble radius. Modes with comoving wavenumber up to the UV regulator scale $\kappa = 5\bar{H}_0 = M_{\rm pl}/2$ remain inside the Hubble radius for another 3 efolds after slow-roll begins, and the shortest-wavelength mode in the simulation, with $k_{\rm max} = 4\kappa = 2M_{\rm pl}$, crosses outside the Hubble radius more than 4 efolds after the system reaches the slow-roll attractor. Hence the early-time dynamics, during which the system enters a phase of slow-roll inflation, occurs with substantial inhomogeneity on sub-Hubble length scales. For this set of initial conditions, in other words, inflation is robust even amid large initial inhomogeneities and with initial conditions for $\varphi(t)$ dominated by kinetic energy.

Furthermore, as shown in Fig. 3-8, we find that for this set of initial conditions inflation actually persists considerably longer when we include large initial inhomogeneities ($N_{\rm infl} \simeq 69$ efolds of inflation) than when we ignore inhomogeneities

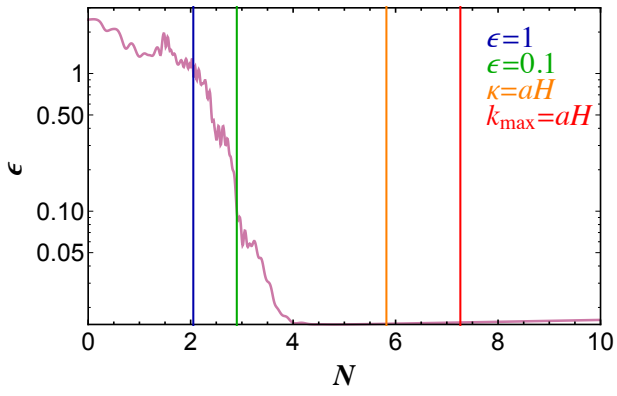

Figure 3-7: The slow-roll parameter ϵ versus N for early times with $\lambda = 10^{-10}$, for the system with large initial fluctuations. Inflation begins at $N \sim 2$ when $\epsilon < 1$, and slow-roll begins by $N \sim 3$ with $\epsilon < 0.1$. Note that modes with comoving wavenumber up to the UV regulator scale κ remain within the Hubble radius until $N \sim 6$, and the shortest-wavelength mode in the spectrum, with $k_{\rm max} = 4\kappa$, crosses outside the Hubble radius at $N \sim 7$.

($N_{\rm infl} \simeq 54$ efolds of inflation). (The authors of Ref. [50] found similar examples in their study of the onset of inflation in multifield models, when fluctuations $\delta\phi(x^\mu)$ and $\delta\psi(x^\mu)$ of the two fields were included.) Remarkably, significant initial inhomogeneities *extended* the duration of inflation in this case, enabling this set of initial conditions $(\varphi_0, \dot\varphi_0)$ to yield sufficient inflation, with $N_{\rm infl} > 65$.

3.4.2 Trajectories in phase space

We can understand the nontrivial effects of large initial inhomogeneities and their nonlinear backreaction on the evolution of $\varphi(t)$ and $H(t)$ by examining the evolution of the system through the phase space $(\varphi(t), \dot\varphi(t))$. As discussed in Refs. [30, 32], for single-field models and vanishing spatial curvature K, the variables $\varphi(t)$ and $\dot\varphi(t)$ define an effective phase space for the evolution of the spatially homogeneous system. Obviously $(\varphi(t), \dot\varphi(t))$ no longer serves as a proper phase space for the full dynamical system when we incorporate the coupled degrees of freedom $\delta\hat\phi(x^\mu)$ and $\hat\Psi(x^\mu)$, but studying the behavior of the system within $(\varphi(t), \dot\varphi(t))$ facilitates comparison with the case in which we neglect fluctuations.

In Fig. 3-9, we plot the evolution of the system $(\varphi(t), \dot\varphi(t))$ when we fix $\varphi_0 = 25\, M_{\rm pl}$ and select $\dot\varphi_0 = \pm 0.25\, M_{\rm pl}^2$. Dashed lines show the evolution of the system for $\epsilon > 0.1$,

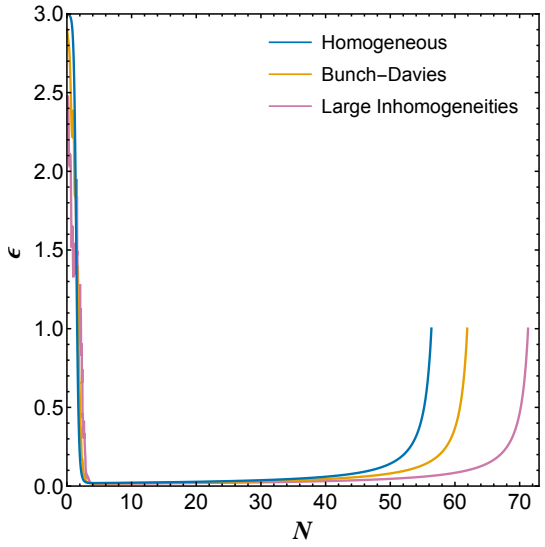

Figure 3-8: The evolution of the slow-roll parameter ϵ versus N with $\lambda = 10^{-10}$ when we neglect fluctuations (blue); when we initialize fluctuations in the Bunch-Davies state (yellow); and for a particular simulation that began with large initial fluctuations, with $\mathcal{C} = 20$ (pink).

before the system reaches the slow-roll inflationary attractor, and solid lines show the evolution once the system has entered slow-roll. We consider cases in which we neglect fluctuations (blue); in which we initialize the fluctuations in the Bunch-Davies state, $\mathcal{C}_{\rm BD} = 2$ (yellow); and in which we initialize the system with large fluctuations, $\mathcal{C} = 20$ (pink). As Fig. 3-9 makes clear, the value of the field when the system reaches the slow-roll attractor, $\varphi(t_{\rm sr})$, depends on the magnitude of the initial inhomogeneities. In particular, the field φ traverses a *shorter* distance before arriving at the slow-roll attractor when we incorporate fluctuations, compared to when we neglect fluctuations: either less far "up the hill" toward higher values of the potential for an initial field velocity $\dot{\varphi}_0 > 0$, or less far "down the hill" for $\dot{\varphi}_0 < 0$. This effect becomes more pronounced as the size of initial inhomogeneities increases.

We can make sense of this result analytically, using the scaling relations for $H(t)$ identified in the previous subsection. For these initial conditions and the coupling $\lambda = 10^{-10}$, the system begins with the kinetic energy in φ greatly exceeding the potential energy, and hence $\bar{\rho}(t) \propto a^{-6}(t)$ before the system enters slow-roll. This is

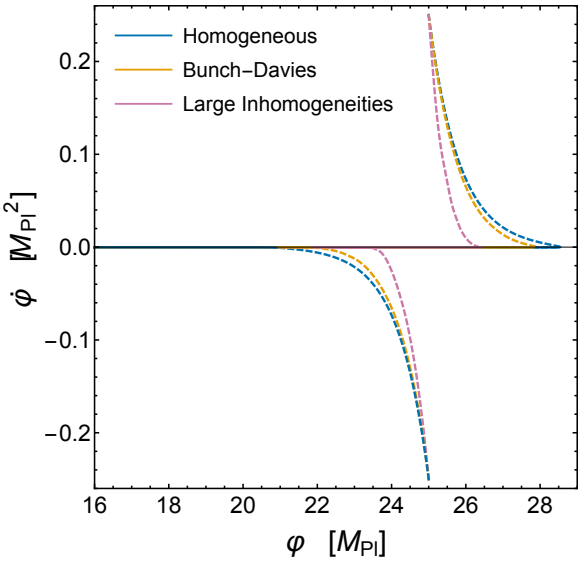

Figure 3-9: The evolution of the system $(\varphi, \dot{\varphi})$ when we fix $\varphi_0 = 25\, M_{\rm pl}$ and select $\dot{\varphi}_0 = \pm 0.25\, M_{\rm pl}^2$, with $\lambda = 10^{-10}$. Dashed lines indicate evolution of the system for $\epsilon > 0.1$. The line along $\dot{\varphi} = 0$ corresponds to the slow-roll inflationary attractor.

equivalent to

$$\dot{\varphi}(N) \simeq \frac{\dot{\varphi}_0}{a^3(N)} = \dot{\varphi}_0\, e^{-3N} \tag{3.48}$$

at early times. In that limit, Eq. (3.17) reduces to $\ddot{\varphi} + 3H\dot{\varphi} \simeq 0$, which we may integrate as

$$\varphi(t) \simeq \varphi_0 + \dot{\varphi}_0 \int_{a_0}^{a(t)} \frac{da}{a^4 H}. \tag{3.49}$$

When we neglect fluctuations, $H(t) \propto a^{-3}(t)$ at early times, and we find

$$\varphi_h(N) \simeq \varphi_0 + \frac{\dot{\varphi}_0}{H_0} N, \tag{3.50}$$

where the subscript "h" indicates evolution of the homogeneous system. On the other hand, when we include large initial fluctuations with $\delta\rho_{(2)}(t_0) \simeq \bar{\rho}(t_0)$, then $H(t) \propto a^{-2}(t)$ at early times, which yields

$$\varphi_q(N) \simeq \varphi_0 + \frac{\dot{\varphi}_0}{H_0} \left(1 - e^{-N}\right), \tag{3.51}$$

where the subscript "q" indicates the evolution of $\varphi(t)$ when we incorporate effects

from the coupled quantum fluctuations. Clearly the field φ will traverse a greater distance during early times when the fluctuations are neglected, as in Eq. (3.50), than when their effects are included, as in Eq. (3.51). The solutions for $\dot{\varphi}(N)$ in Eq. (3.48) and for $\varphi(N)$ in Eqs. (3.50) or (3.51) closely match the trajectories shown in Fig. 3-9 for the relevant cases, even though the curves in Fig. 3-9 come from our full numerical simulations.

We may use Eqs. (3.50)-(3.51) to estimate the values $\varphi_h(N_{\rm sr})$ and $\varphi_q(N_{\rm sr})$ at the time $N_{\rm sr}$ when the system reaches the slow-roll attractor. We estimate $N_{\rm sr}$ by setting $\dot{\varphi}^2(N_{\rm sr})/2 = V(\varphi(N_{\rm sr}))$. For $\lambda = 10^{-10}$ and $(\varphi_0, \dot{\varphi}_0) = (25\, M_{\rm pl}, -0.25\, M_{\rm pl}^2)$, we have $\bar{H}_0 = [\dot{\varphi}_0^2/(6M_{\rm pl}^2)]^{1/2}$ for the homogeneous case and $H_0 = \sqrt{2}\,\bar{H}_0$ for the case with $\delta\rho_{(2)}(t_0) \simeq \bar{\rho}(t_0)$. These yield $\varphi_h(N_{\rm sr}) = 21.4\, M_{\rm pl}$ and $\varphi_q(N_{\rm sr}) = 23.7\, M_{\rm pl}$, again closely matching the numerical results shown in Fig. 3-9.

After the system reaches the slow-roll attractor, $\bar{\rho}(t) \sim$ constant while $\delta\rho_{(2)}(t)$ continues to redshift like radiation until most of the modes have crossed outside the Hubble radius, so the dynamics become dominated by $\bar{\rho} \gg \delta\rho_{(2)}$. In that regime, we may use the usual slow-roll approximation to estimate the duration of inflation,

$$\begin{aligned} N_{\rm infl} &\simeq -\frac{1}{M_{\rm pl}^2}\int_{\varphi_{\rm sr}}^{\varphi_{\rm end}} d\varphi \left(\frac{V(\varphi)}{V^{(1)}(\varphi)}\right) \\ &= \frac{1}{8M_{\rm pl}^2}\left(\varphi_{\rm sr}^2 - \varphi_{\rm end}^2\right), \end{aligned} \tag{3.52}$$

where $\varphi_{\rm sr} = \varphi(N_{\rm sr})$, and $\varphi_{\rm end} = \varphi(N_{\rm end})$ is determined by the condition $\epsilon(N_{\rm end}) = 1$. Again using the usual slow-roll estimate for late times, $\epsilon \simeq (M_{\rm pl}^2/2)(V^{(1)}(\varphi)/V(\varphi))^2$, we find $\varphi_{\rm end} = \sqrt{8}\, M_{\rm pl}$. Given our estimates of $\varphi_q(N_{\rm sr})$ and $\varphi_h(N_{\rm sr})$, we then find $N_{\rm infl} = 69.2$ efolds of inflation when we incorporate large initial quantum fluctuations, and $N_{\rm infl} = 56.2$ efolds when we neglect fluctuations — a close match to the behavior shown for the full numerical results in Fig. 3-8.

We can thus understand the most significant effect of the coupled fluctuations on the evolution of the system. Large fluctuations raise the initial value of the Hubble parameter compared to the case with no fluctuations, $H_0 > \bar{H}_0$, thereby increasing the initial Hubble drag on the field $\varphi(t)$. Even more significant, backreaction from

the fluctuations changes the scaling of $H(t)$ with $a(t)$ at early times, slowing the rate at which $H(t)$ falls, which further increases the effect of Hubble drag on the evolution of $\varphi(t)$. The backreaction dampens φ's motion as the system evolves toward the slow-roll inflationary attractor, such that $|\varphi_q(N_{\rm sr}) - \varphi_0| < |\varphi_h(N_{\rm sr}) - \varphi_0|$. Once the system reaches the attractor, the duration of inflation will be governed by the value $\varphi_q(N_{\rm sr})$. For an initial velocity "up the hill," with $\dot{\varphi}_0 > 0$, $\varphi_q(N_{\rm sr}) < \varphi_h(N_{\rm sr})$, and the system will spend less time evolving along the inflationary attractor than in the absence of fluctuations. For an initial velocity "down the hill," with $\dot{\varphi}_0 < 0$, $\varphi_q(N_{\rm sr}) > \varphi_h(N_{\rm sr})$, and the system will spend more time evolving along the inflationary attractor than in the absence of fluctuations.

3.4.3 Phase space of initial conditions

We turn now to discuss the effects of the coupled fluctuations on the evolution of the system across the phase space of initial conditions $(\varphi_0, \dot{\varphi}_0)$ for $\lambda = 10^{-10}$, as we vary $12\, M_{\rm pl} \leq \varphi_0 \leq 30\, M_{\rm pl}$ and $-0.25\, M_{\rm pl}^2 \leq \dot{\varphi}_0 \leq 0.25\, M_{\rm pl}^2$. To investigate the phase space behavior for the perturbed initial conditions, we construct averages from the 32 samples evolved at each point in $(\varphi_0, \dot{\varphi}_0)$. Figure 3-10 shows the average value of $\delta\rho_{(2)}(t_0)$ at each grid point compared to $\delta\rho_{(2)}(t_0)$ for Bunch-Davies initial conditions, confirming that for the ranges of coefficients in Eq. (3.44), we find initial energy densities about ten times greater than for the Bunch-Davies state. In Fig. 3-11, we plot the average of the initial value $\Psi_{\rm rms}(t_0) \equiv [\langle \hat{\Psi}^2(t_0) \rangle]^{1/2}$ at each grid point, confirming that for large quantum fluctuations, initialized such that $\mathcal{C} \simeq 20$, the system begins with $|\Psi(t_0, \mathbf{x})| \lesssim 0.5$.

Across $(\varphi_0, \dot{\varphi}_0)$, when the system begins with large initial inhomogeneities, the system reaches the slow-roll inflationary attractor ($\epsilon \leq 0.1$) while significant power remains in fluctuations on sub-Hubble scales. Figure 3-12 shows the average value of the ratio $N_{\rm sr}/N_\kappa$ at each grid point for fluctuations that begin with $\mathcal{C} \simeq 20$, where N_κ is the time when the mode with comoving wavenumber equal to the UV regulator scale κ crosses outside the Hubble radius, $\kappa = aH$. In all simulations, $N_{\rm sr}/N_\kappa \leq 0.55 \pm 0.06$. The ratio drops to zero at $\dot{\varphi}_0 = 0$, because at those locations in phase space the system

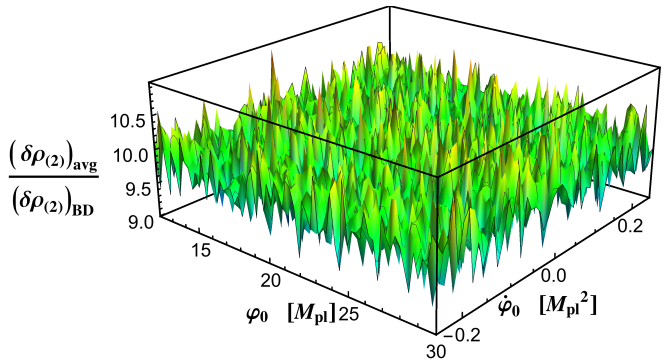

Figure 3-10: The average value of $\delta\rho_{(2)}(t_0)$ at each grid point in $(\varphi_0, \dot{\varphi}_0)$, when quantum fluctuations are initialized with random coefficients drawn from the ranges in Eq. (3.44), compared to the value of $\delta\rho_{(2)}(t_0)$ when fluctuations are initialized in the Bunch-Davies state, with $\lambda = 10^{-10}$.

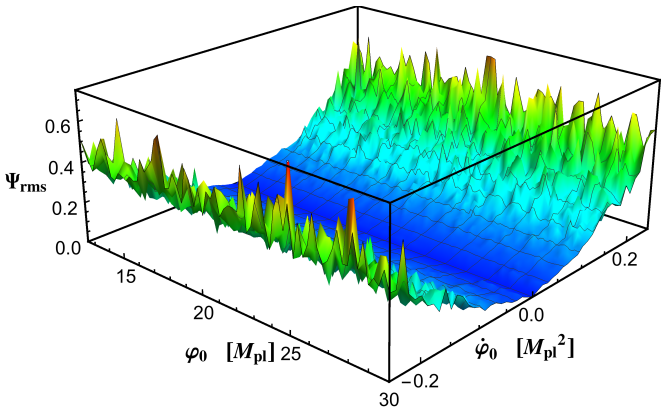

Figure 3-11: The average value of $\Psi_{\rm rms}(t_0)$ at each grid point in $(\varphi_0, \dot{\varphi}_0)$ when quantum fluctuations are initialized with random coefficients drawn from the ranges in Eq. (3.44), with $\lambda = 10^{-10}$. Note that $\Psi_{\rm rms}(t_0)$ is roughly proportional to $\dot{\varphi}_0^2$.

begins on the slow-roll attractor, and hence $N_{\rm sr} = 0$.

Results for the duration of inflation $N_{\rm infl}$ across these cases are shown in Fig. 3-13. Consider first the case in which the effects of the coupled fluctuations are neglected. For a given value φ_0, large initial velocities $\dot{\varphi}_0 > 0$ prolong the duration of inflation compared to the case with $\dot{\varphi}_0 = 0$, whereas large initial velocities $\dot{\varphi}_0 < 0$ decrease the duration of inflation. When initial inhomogeneities are included they backreact on $H(t)$, increasing the effect of Hubble drag on $\varphi(t)$, thereby affecting the shape of the contours of constant $N_{\rm infl}$ within $(\varphi_0, \dot{\varphi}_0)$. In particular, the effects of large

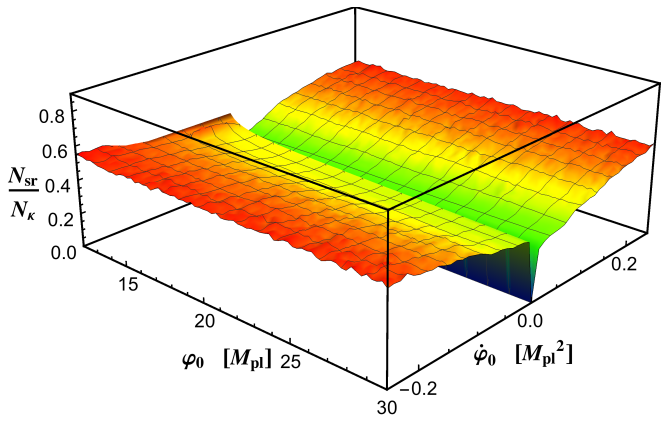

Figure 3-12: Average of the ratio of the time $N_{\rm sr}$ when the system first reaches the slow-roll inflationary attractor with $\epsilon \leq 0.1$, to the time N_κ when the mode with comoving wavenumber equal to the UV regulator scale crosses outside the Hubble radius, $\kappa = aH$, for the case of large initial fluctuations ($\mathcal{C} \simeq 20$), with $\lambda = 10^{-10}$. For all simulations across $(\varphi_0, \dot{\varphi}_0)$, we find $N_{sr}/N_\kappa \leq 0.55 \pm 0.06$.

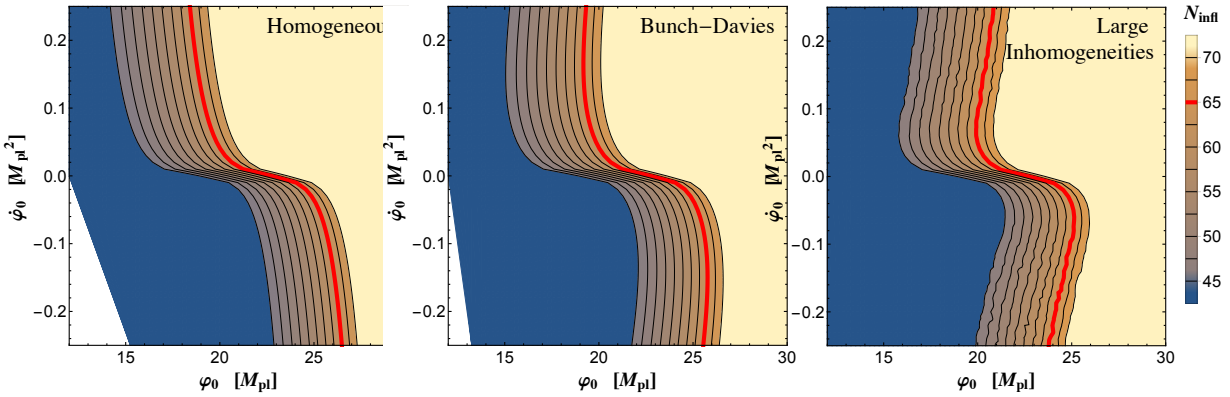

Figure 3-13: Contours of constant $N_{\rm infl}$ in $(\varphi_0, \dot{\varphi}_0)$ for $\lambda = 10^{-10}$ when the fluctuations are neglected (*left*); when the fluctuations are initialized in the Bunch-Davies state, $\mathcal{C}_{\rm BD} = 2$ (*middle*); and when the fluctuations are initialized with random coefficients $\gamma_{n\ell m}$ and $\delta_{n\ell m}$ for each mode drawn from the ranges in Eq. (3.44), which yields $\mathcal{C} \simeq 20$ (*right*). For the case of large initial fluctuations (*right*), the contours of constant $N_{\rm infl}$ were evaluated by averaging 32 simulations per grid point. In each plot, regions of dark blue indicate $N_{\rm infl} < 45$ and regions of light yellow indicate $N_{\rm infl} \geq 70$. In white regions in the lower left, the system never entered slow-roll. The critical lines that yield $N_{\rm infl} = 65$ efolds of inflation are indicated in red.

$|\dot{\varphi}_0|$ are more quickly damped than when the fluctuations are neglected, so that the field φ spends less time evolving along the slow-roll inflationary attractor than the corresponding case without fluctuations for $\dot{\varphi}_0 > 0$, and more time along the attractor for $\dot{\varphi}_0 < 0$.

Figure 3-14 shows the critical line in $(\varphi_0, \dot{\varphi}_0)$, to the right of which the system yields $N_{\rm infl} > 65$ efolds of inflation, for the homogeneous system (when we neglect fluctuations), for fluctuations in the Bunch-Davies initial state with $\mathcal{C}_{\rm BD} = 2$, and for larger initial fluctuations with $\mathcal{C} \simeq 20$. For the cases with large initial fluctuations, we plot the critical line based on averaging across 32 simulations per grid point, as well as $\pm 2\sigma$ contours.

Consistent with the analysis in the previous subsection, the effects of large initial fluctuations are symmetric for $\pm|\dot{\varphi}_0|$, and most significant for large $|\dot{\varphi}_0|$. As $|\dot{\varphi}_0|$ increases, the initial value $\bar{H}_0$ increases; the greater initial energy scale $\bar{H}_0$, in turn, yields more initial energy density in fluctuations, $\delta\rho_{(2)}(t_0)$, which raises $H_0 > \bar{H}_0$ and affects the scaling of $H(t)$ with $a(t)$ at early times. The larger fluctuations also seed larger initial inhomogeneities, $\Psi(t_0, \mathbf{x})$.

Although the effects of the nonlinear dynamics of the fluctuations are most pronounced at large $|\dot{\varphi}_0|$, the volume of the (projected) phase space of initial conditions $(\varphi_0, \dot{\varphi}_0)$ that yields sufficient inflation is *conserved*. The same volume of the $(\varphi_0, \dot{\varphi}_0)$ phase space that yields sufficient inflation which is lost in the region with $\dot{\varphi}_0 > 0$, compared to the homogeneous case, is gained in the region with $\dot{\varphi}_0 < 0$. This provides a useful quantitative measure of the robustness of single-field inflation to large initial inhomogeneities.

3.4.4 Varying the coupling constant

The main impact of lowering the coupling constant from $\lambda = 10^{-10}$ to $\lambda = 10^{-12}$ is to increase the ratio $H_0/H_{\rm infl}$ from $\mathcal{O}(10^2)$ to $\mathcal{O}(10^3)$. That prolongs the time during which the enhanced Hubble drag from the coupled fluctuations affects the evolution of $\varphi(t)$, compared to the case in which the fluctuations are neglected. We divided the phase space of initial conditions $(\varphi_0, \dot{\varphi}_0)$ into the same grid as for the $\lambda = 10^{-10}$ case,

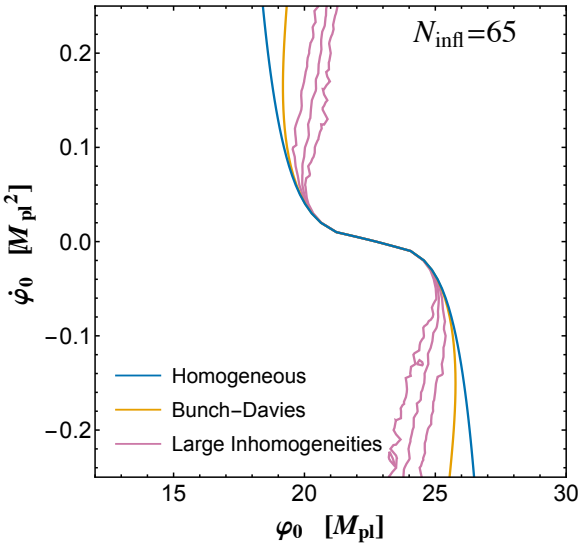

Figure 3-14: The critical line in $(\varphi_0, \dot{\varphi}_0)$ that yields sufficient inflation, with $N_{\rm infl} \geq 65$, for $\lambda = 10^{-10}$, for the cases of homogeneous evolution with no fluctuations (blue); fluctuations in the Bunch-Davies initial state (yellow); and large initial fluctuations (pink). (Points to the right of the critical lines achieve sufficient inflation.) For the latter, we show the critical line based on averaging across 32 simulations per grid point, as well as $\pm 2\sigma$ contours. The jitter in the pink curves arises from stochastic fluctuations, and it is expected that the curves would become smooth as we increase both the resolution of our sampling and the number of simulations at each point in phase space.

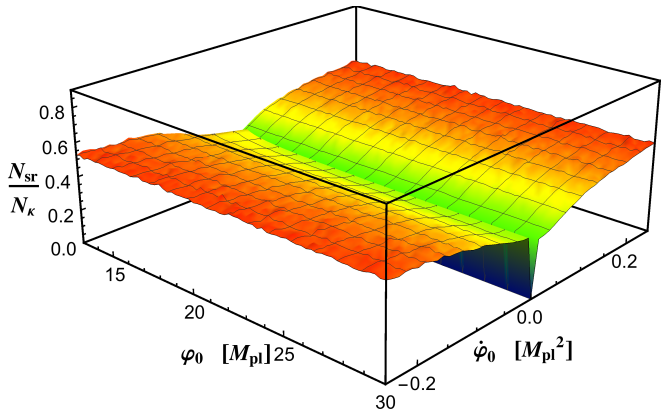

Figure 3-15: Average of the ratio of the time $N_{\rm sr}$ when the system first reaches the slow-roll inflationary attractor to the time N_κ when $\kappa = aH$ for the case of large initial fluctuations ($\mathcal{C} \simeq 20$), with $\lambda = 10^{-12}$. Across all simulations we find $N_{\rm sr}/N_\kappa \leq 0.53 \pm 0.04$.

and considered cases in which we neglected fluctuations, began with fluctuations in the Bunch-Davies initial state, and began with larger initial fluctuations, with the random coefficients for each mode $\delta\phi_{n\ell m}(t_0)$ and $\delta\dot{\phi}_{n\ell m}(t_0)$ drawn from the ranges in Eq. (3.44). Because the contributions to $\bar{\rho}(t_0)$ and $\delta\rho_{(2)}(t_0)$ that are proportional to λ remain subdominant for $\lambda = 10^{-12}$ as for $\lambda = 10^{-10}$, the initial values for $(\delta\rho_{(2)}(t_0))_{\rm avg}/(\delta\rho_{(2)}(t_0))_{\rm BD}$ and for $\Psi_{\rm rms}(t_0)$ are little changed from the results shown in Figs. 3-10 and 3-11. Likewise, we again find that throughout $(\varphi_0, \dot{\varphi}_0)$, $N_{\rm sr}/N_\kappa \leq 0.53 \pm 0.04$, as shown in Fig. 3-15.

Figure 3-16 shows contours of constant $N_{\rm infl}$ in $(\varphi_0, \dot{\varphi}_0)$ with $\lambda = 10^{-12}$ for the three cases of interest: no fluctuations, Bunch-Davies intial state ($\mathcal{C}_{\rm BD} = 2$), and larger initial fluctuations (with $\mathcal{C} \simeq 20$). In Fig. 3-17 we plot the critical line in $(\varphi_0, \dot{\varphi}_0)$ that yields $N_{\rm infl} = 65$ efolds of inflation for each of the three cases. For the cases with large initial fluctuations, we plot the critical line based on averaging across 32 simulations per grid point, as well as $\pm 2\sigma$ contours. The results are comparable to the case with $\lambda = 10^{-10}$. Although the effects of the coupled fluctuations are more dramatic with the smaller coupling — because the system takes more time to reach the inflationary attractor, and hence the enhanced Hubble drag in the presence of coupled fluctuations acts longer on the evolution of $\varphi(t)$ — the effects across the projected phase space $(\varphi_0, \dot{\varphi}_0)$ are once again symmetrical for $\pm|\dot{\varphi}_0|$, indicating that

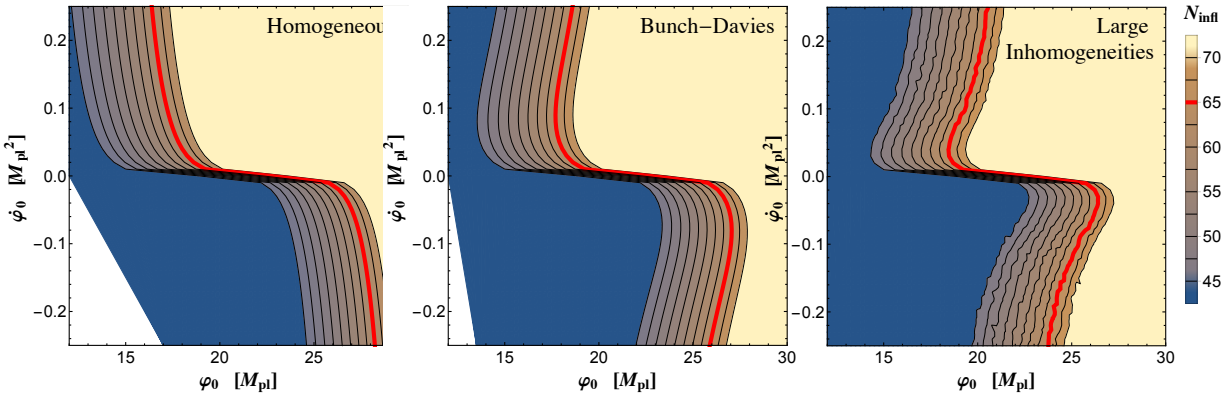

Figure 3-16: Contours of constant $N_{\rm infl}$ in $(\varphi_0, \dot{\varphi}_0)$ for $\lambda = 10^{-12}$ when the fluctuations are neglected (*left*); when the fluctuations are initialized in the Bunch-Davies state with $\mathcal{C}_{\rm BD} = 2$ (*middle*); and when the fluctuations are initialized with random coefficients $\gamma_{n\ell m}$ and $\delta_{n\ell m}$ for each mode drawn from the ranges in Eq. (3.44), which yields $\mathcal{C} \simeq 20$ (*right*). For the case of large initial fluctuations (*right*), the contours of constant $N_{\rm infl}$ were evaluated by averaging 32 simulations per grid point. In each plot, regions of dark blue indicate $N_{\rm infl} < 45$ and regions of light yellow indicate $N_{\rm infl} \geq 70$. In white regions in the lower left, the system never entered slow-roll. The critical lines that yield $N_{\rm infl} = 65$ efolds of inflation are indicated in red.

the total volume of the $(\varphi_0, \dot{\varphi}_0)$ phase space that yields sufficient inflation is conserved.

3.5 Conclusions

In this chapter we have analyzed the onset of inflation for a simple single-field model, $V(\phi) = \lambda\phi^4/4$, when the system begins with significant inhomogeneities on length-scales shorter than the initial Hubble radius. We incorporate certain nonlinear interactions among the coupled degrees of freedom by using the nonperturbative Hartree approximation, which resums an infinite set of Feynman diagrams involving the self-interacting quantum fluctuations $\delta\hat{\phi}(x^\mu)$ to construct a dressed propagator. By initializing the system in an excited state, with the energy density of fluctuations approximately ten times greater than in the minimum-energy Bunch-Davies state, our simulations begin with significant spatial inhomogeneities, parameterized by the scalar metric perturbation $|\Psi(t_0, \mathbf{x})| \lesssim 0.5$.

The energy density in fluctuations, $\delta\rho_{(2)}(t)$, backreacts on the evolution of the Hubble parameter $H(t)$, affecting the scaling of $H(t)$ with $a(t)$. This backreaction,

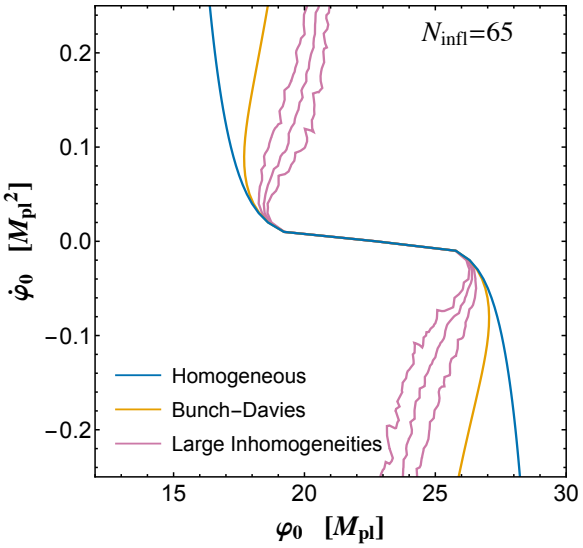

Figure 3-17: The critical line in $(\varphi_0, \dot{\varphi}_0)$ that yields $N_{\rm infl} \geq 65$ efolds of inflation for $\lambda = 10^{-12}$, for the cases of homogeneous evolution (blue); fluctuations in the Bunch-Davies initial state (yellow); and large initial fluctuations (pink). (Points to the right of the critical lines achieve sufficient inflation.) For the latter, we show the critical line based on averaging across 32 simulations per grid point, as well as $\pm 2\sigma$ contours.

in turn, leads to increased Hubble drag on the evolution of the vacuum expectation value of the field, $\varphi(t)$, affecting how quickly $\varphi(t)$ arrives at the slow-roll inflationary attractor, compared to the case in which one neglects fluctuations.

The impact of inhomogeneities on the evolution of the system is largest for initial conditions that yield the greatest initial value of $H(t_0)$, since the initial energy density in fluctuations scales as $H^4(t_0)$. Compared to those regions of $(\varphi_0, \dot{\varphi}_0)$ that yield $N_{\rm infl} \geq 65$ efolds of inflation when one neglects inhomogeneities, we find some regions that *fail* to yield sufficient inflation when we incorporate inhomogeneities, and an equal volume of regions that *succeed* in producing $N_{\rm infl} \geq 65$ but which had failed to do so in the absence of inhomogeneities. (See also Ref. [50].) In other words, the total volume of the space $(\varphi_0, \dot{\varphi}_0)$ that yields sufficient inflation is conserved when we incorporate nonlinear backreaction from inhomogeneities, compared to the case in which we neglect inhomogeneities.

Although we have analyzed the system numerically for a specific form of $V(\phi)$, the arguments about the robustness of inflation for such large-field models do not depend on our choice of $V(\phi)$. All that enters into our semi-analytic argument is

that the system can begin with large initial quantum fluctuations, such that $\delta\rho_{(2)}(t_0)$ is comparable to (or greater than) the initial energy density associated with the vacuum expectation value, $\bar{\rho}(t_0)$. For weakly coupled models — as required for large-field inflation, in order to produce a spectrum of primordial density perturbations consistent with observations — we generically expect that $\delta\rho_{(2)}(t)$ will evolve at early times with an equation of state like that of a gas of (nearly) massless particles, scaling as $\delta\rho_{(2)}(t) \propto a^{-4}(t)$ while most of the power in fluctuations is on sub-Hubble scales. This behavior for the fluctuations contrasts with the scaling of $\bar{\rho}(t)$ at early times, when the energy density associated with $\varphi(t)$ is dominated by kinetic energy, such that $\bar{\rho}(t) \propto a^{-6}(t)$. The backreaction of $\delta\rho_{(2)}(t)$ on $H(t)$ ensures that $\varphi(t)$ will traverse less distance en route to the slow-roll inflationary attractor than in the absence of inhomogeneities, thereby accounting for the differences we observe in the duration of inflation, $N_{\rm infl}$.

In our numerical analysis we initialize fluctuations at t_0 across the range of wavenumbers from $k_{\rm min}/a(t_0) \lesssim 2H_0/3$ up to $k_{\rm max}/a(t_0) = 30\,k_{\rm min}/a(t_0) \sim M_{\rm pl}$. Including any modes with $k > k_{\rm max}$ in the spectrum would lead to trans-Planckian ambiguities. One could nonetheless imagine initializing additional modes at later times — to simulate the notion that modes which had begun with $k/a(t_0) \gg M_{\rm pl}$ at t_0 later redshifted to $k/a(t) \leq M_{\rm pl}$ — but we do not expect such additional, trans-Planckian modes to change the qualitative behavior of the system. In our current framework, the system consistently arrives at the slow-roll inflationary attractor while most of the initial power in sub-Hubble fluctuations remains within the Hubble radius. The energy density associated with any modes that might be initialized at later times $t > t_0$ would be less than $\delta\rho_{(2)}(t_0)$, since $H(t) < H(t_0)$ and $\delta\rho_{(2)}(t) \propto H^4(t)$. Moreover, after a few efolds of slow-roll inflation, we expect that any newly initialized fluctuations should begin in the Bunch-Davies state, rather than in the more-energetic initial states that we consider here.

For next steps, we aim to generalize our formalism to include the evolution of systems with nonzero spatial curvature K, to consider small-field as well as large-field models, and to extend the formalism to multifield models (akin to Ref. [50],

but incorporating the coupled metric perturbations). Across each of these cases, we believe the approach we have developed here can complement the computationally intensive numerical-relativity approaches of Refs. [45–47].

Chapter 4

History of the dark sector

4.1 Introduction

Astrophysical observations spanning galactic to cosmological length scales have established the existence of dark matter. These observations include the initial observation of "missing" gravitating non-luminous matter in galaxy clusters ([68]), the observation of flat galactic rotation curves which imply the existence of non-luminous matter extending to much larger radii than visible matter seen in the galactic disk ([69]), the observation of a displacement between the baryonic component (X-ray emitting gas) and the main collisionless dark matter component (seen by weak gravitational lensing) in the Bullet Cluster system of two colliding galaxy clusters ([70]), observations of large-scale structure (e.g. [71, 72]), and the shape of the CMB anisotropy power spectrum [12]. The CMB anisotropy power spectrum (especially the third acoustic peak in the power spectrum) is particularly sensitive to the total amount of dark matter in the universe, as is illustrated in Figure 4-1 (taken from [73]), and the latest Planck measurements find [12]:

$$\Omega_c h^2 = 0.120 \pm 0.001, \tag{4.1}$$

where Ω_c is the dark matter abundance and $h = 0.674 \pm 0.0005$ is the dimensionless scaling factor for the Hubble expansion rate. This means that about 85% of all the

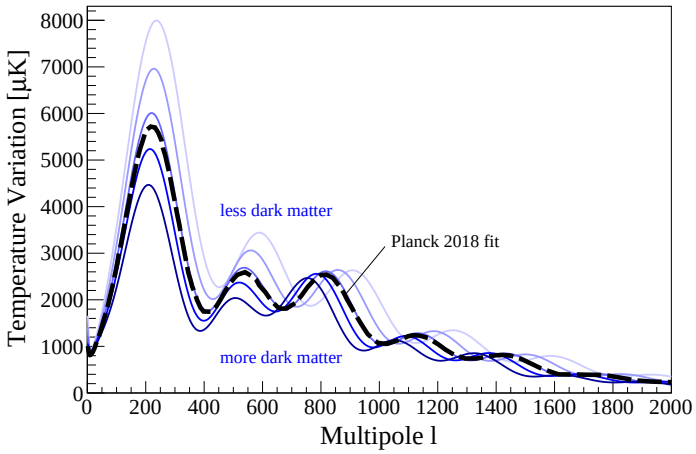

Figure 4-1: The temperature power spectrum of the CMB for a dark matter density contribution varying between $0.11 \leq \Omega_c \leq 0.43$ (blue lines) with all other cosmological parameters held fixed. The dashed black lines show the best fit to the 2018 Planck data [12]. Image is taken from [73].

matter in the universe is dark matter, measured at an uncertainty of 2%.

That is about all we can confidently say about dark matter. Although all of our experimental evidence is consistent with dark matter being a new particle not found in the Standard Model (SM), the basic properties of this particle and its interactions are completely unknown. Most of our theoretical efforts to understand dark matter therefore involve extending the SM in a way that provides one or more candidate particles which would constitute the dark matter, evading current experimental bounds and motivating future experiments looking for this new physics. This program, of course, assumes there to be interactions between dark matter and the SM. We might expect this assumption to be true since the abundance of dark matter is within an order of magnitude of baryonic matter, and because dark matter candidates which interact with the SM are provided by separately motivated theories, such as Supersymmetry, which is proposed to solve the hierarchy problem and predicts new physics at the weak scale, and the QCD axion, which is proposed to solve the strong CP problem.

Experimental searches for dark matter are mainly divided into three strategies, corresponding to three classes of interactions which can take place between DM and the SM:

- we can produce DM, χ, in interactions of the form SM SM $\to \chi\chi$. Therefore we look for DM production through proton-proton collisions at the Large Hadron Collider (LHC), $pp \to \chi\chi$. Of course, what we would actually hope to measure at the LHC from this process would be additional observable final states X, (e.g. $pp \to \chi\chi + X$) with missing energy carried away by the dark matter (see e.g. [74]);

- DM can scatter with SM particles in interactions of the form χ SM $\to \chi$ SM. An industry of direct detection experiments exploit this scattering process to detect DM through recoil of the SM particle. Direct detection experiments have set strong limits on the rate at which this scattering process can occur (see e.g. [75]);

- DM can annihilate through interactions of the form $\chi\chi \to$ SM SM, or it can decay through processes of the form $\chi \to$ SM SM. Strategies to detect dark matter through these interactions fall under the class of astrophysical and cosmological indirect detection experiments, including gamma-ray searches (see e.g. [76–82]), neutrino searches (see e.g. [83–85]), cosmic ray experiments (see e.g. [86–88]), and cosmological experiments mapping the CMB (see e.g. [12]).

If DM couples to the SM through one or more of the interactions summarized above, and if these interactions proceed at sufficient rates in the early universe they can establish thermal and chemical equilibrium between the dark and SM sectors. As the universe expands and cools the rates for the processes which thermally and chemically couple the dark and SM sectors will become inefficient, until eventually these processes become too inefficient to maintain thermal and chemical equilibrium, when they are said to "freeze out", and thereafter the comoving number density of DM remains constant, at the late-time relic abundance that we observe today. The many models whose relic abundance is determined by such a scenario are called "thermal relics", and they have the attractive feature that the late-time relic abundance is entirely determined by the cross section for the process which controls the thermal and chemical coupling of the DM and SM sectors (variations of this statement constitute

Chapters 5 and 6 of this thesis). Hence, knowing the late-time abundance of DM gives us direct insight into the microscopic properties of the thermal relic models which might have produced the DM. To make this more concrete we need to solve the Boltzmann equations. In this chapter I take inspiration from Mariangela Lisanti's "Lectures on Dark Matter Physics" [89], Edward Kolb and Michael Turner's textbook "The Early Universe" [90], Tracy Slatyer's "TASI Lectures on Indirect Detection of Dark Matter" [91], and the papers on SIMP DM (e.g. [92]) and ELDER DM ([93, 94]).

4.2 Calculating the thermal history of the dark sector

We calculate the thermal history and relic abundance of DM in a given model by solving the Boltzmann equations, which I'll now describe.

4.2.1 First moment of the Boltzmann equation: number density evolution

The probability of finding a DM particle, χ, in a phase space volume $d^3\mathbf{x}d^3\mathbf{p}$ of some astrophysical environment (say a galactic halo) is given by $f\left(t;\mathbf{x},\mathbf{p}\right)d^3\mathbf{x}d^3\mathbf{p}$, where $f\left(t;\mathbf{x},\mathbf{p}\right)$ is the phase space density of the DM particle. Conservation of probability gives the relation:

$$\int f\left(t;\mathbf{x},\mathbf{p}\right)d^3\mathbf{x}d^3\mathbf{p}=1. \tag{4.2}$$

We can obtain the DM number and energy density from the phase space density:

$$n_\chi = g_\chi \int \frac{d^3\mathbf{p}}{\left(2\pi\right)^3} f_\chi, \tag{4.3}$$

and

$$\rho_\chi = g_\chi \int \frac{d^3\mathbf{p}}{\left(2\pi\right)^3} E f_\chi, \tag{4.4}$$

where g_χ is the number of degrees of freedom in χ, and $E = \sqrt{|\mathbf{p}|^2 + m_\chi^2}$.

The Boltzmann equation says:

$$L[f] = C[f], \tag{4.5}$$

where L is the Liouville operator and C is the collision operator. The Liouville operator describes the time and phase-space evolution of the phase space density, including gravitational effects and effects of the Hubble expansion, and the collision operator describes the effects of interactions of DM with itself and other particles. The general form of the Liouville operator is:

$$L[f] = p^\alpha \frac{\partial f}{\partial x^\alpha} - \Gamma^\alpha_{\beta\gamma} p^\beta p^\gamma \frac{\partial f}{\partial p^\alpha}, \tag{4.6}$$

where $\Gamma^\alpha_{\beta\gamma}$ is the affine connection, which captures gravitational effects. As we described in Chapter 2, for the FLRW spacetime the phase space density is spatially homogeneous and isotropic, and has the form: $f(t; \mathbf{p})$. For the FLRW spacetime the Liouville operator has the form:

$$L[f] = E\frac{\partial f}{\partial t} - H|\mathbf{p}|^2 \frac{\partial f}{\partial E}. \tag{4.7}$$

Taking the first moment of the Boltzmann equation, and applying the definition of n_χ in Equation 4.3, we obtain:

$$g_\chi \int L[f] \frac{d^3\mathbf{p}}{(2\pi)^3} = \frac{1}{a^3}\frac{d}{dt}\left(na^3\right) = \frac{dn}{dt} + 3Hn = g_\chi \int C[f] \frac{1}{E}\frac{d^3\mathbf{p}}{(2\pi)^3}. \tag{4.8}$$

We see from Equation 4.8 that in the absence of interactions with the DM, captured by the collision operator $C[f]$, then the number of DM particles in a comoving volume would remain constant, $d(na^3)/dt = 0$, as we expect. Let's now consider a general interaction between a DM particle, χ, and other particles which may be either DM or SM particles, $(\chi + a + ... \leftrightarrow i + j + ...)$. The collision term in Equation 4.8 for this

process is given by:

$$\begin{aligned} g_\chi \int C\left[f\right] \frac{1}{E} \frac{d^3\mathbf{p}}{\left(2\pi\right)^3} = &- \sum_{\text{spins}} \int d\Pi_\chi d\Pi_a ... d\Pi_i d\Pi_j ... \\ &\times \left(2\pi\right)^4 \delta^4 \left(p_\chi + p_a + ... - p_i - p_j - ...\right) \\ &\times \left[\left| \mathcal{M}_{\chi+a+...\rightarrow i+j+...} \right|^2 f_\chi f_a ... f_i f_j ... \left(1 \pm f_i\right) \left(1 \pm f_j\right) ... \right. \\ &\left. - \left| \mathcal{M}_{i+j+...\rightarrow \chi+a+...} \right|^2 f_i f_j ... f_\chi f_a ... \left(1 \pm f_\chi\right) \left(1 \pm f_a\right) ... \right], \end{aligned} \tag{4.9}$$

where $\mathcal{M}_{\chi+a+...\rightarrow i+j+...}$ is the matrix element for the reaction $\chi + a + ... \rightarrow i + j + ...$, and in factors of the form $(1 \pm f)$ the (+) sign applies to bosons and the (-) sign applies to fermions. The Lorentz-invariant phase space integration factors $d\Pi$ are given by:

$$d\Pi_i = \frac{d^3\mathbf{p}_i}{\left(2\pi\right)^3 2E_i}. \tag{4.10}$$

We can make a few assumptions in order to bring Equation 4.9 into a more manageable form. The assumption of CP (or T) invariance means the forward and backward processes give the same matrix amplitudes:

$$\left| \mathcal{M}_{\chi+a+...\rightarrow i+j+...} \right|^2 = \left| \mathcal{M}_{i+j+...\rightarrow \chi+a+...} \right|^2. \tag{4.11}$$

We also assume that rather than Fermi-Dirac statistics for fermions and Bose-Einstein statistics for bosons, we can use Maxwell-Boltzmann statistics for all species. This assumption is accurate when kinetic equilibrium is maintained among each species, and when the temperature of each species satisfies $T_i \ll E_i - \mu_i$. Under this assumption, the effects of Bose condensation or Fermi degeneracy can be ignored, so $(1 \pm f) \simeq 1$, and the phase space distributions take the form $f_i\left(E_i\right) \simeq \exp\left[-\left(E_i - \mu_i\right)/T\right]$. In this case the Boltzmann equation takes the simplified form:

$$\begin{aligned} \frac{dn_\chi}{dt} + 3Hn_\chi = &- \sum_{\text{spins}} \int d\Pi_\chi d\Pi_a ... d\Pi_i d\Pi_j ... \left(2\pi\right)^4 \left| \mathcal{M}_{\chi+a+...\rightarrow i+j+...} \right|^2 \\ &\delta^4 \left(p_\chi + p_a + ... - p_i - p_j - ...\right) \left[f_\chi f_a ... - f_i f_j ...\right]. \end{aligned} \tag{4.12}$$

Now, in order to write Equation 4.12 in a more simple form let's specialize to the

case of two particles in the initial state, $\chi + a$, this will allow us to write Equation 4.12 in terms of the familiar definition of the (initial-state-averaged and final-state-summed) cross section:

$$\begin{aligned} d\sigma_{\chi+a\to i+j+...} = \frac{1}{4E_\chi E_a}\frac{1}{|\mathbf{v}_\chi - \mathbf{v_a}|}\left(2\pi\right)^4\delta^4\left(p_\chi + p_a - p_i - p_j - ...\right) \\ \times \prod_\ell \frac{d^3\mathbf{p}_\ell}{\left(2\pi\right)^3 2E_\ell}\langle|\mathcal{M}|^2\rangle, \end{aligned} \tag{4.13}$$

which we can generalize to a Lorentz-invariant form:

$$\begin{aligned} d\sigma_{\chi+a\to i+j+...} = \frac{1}{4\sqrt{\left(p_\chi\cdot p_a\right)^2 - m_\chi^2 m_a^2}}\left(2\pi\right)^4\delta^4\left(p_\chi + p_a - p_i - p_j - ...\right) \\ \times \prod_\ell \frac{d^3\mathbf{p}_\ell}{\left(2\pi\right)^3 2E_\ell}\langle|\mathcal{M}|^2\rangle, \end{aligned} \tag{4.14}$$

where ℓ runs over all final states, and $\langle|\mathcal{M}|^2\rangle$ is the initial-state-averaged and final-state-summed matrix element:

$$\langle|\mathcal{M}|^2\rangle \equiv \frac{1}{g_\chi g_a}\sum_{\text{spins}}|\mathcal{M}|^2\,. \tag{4.15}$$

Plugging Equation 4.14 back into Equation 4.12 we have:

$$\begin{aligned} \frac{dn_\chi}{dt} + 3Hn_\chi = -C_1\frac{1}{S_{id}}g_\chi g_a\int d\Pi_\chi d\Pi_a 4\sqrt{\left(p_\chi\cdot p_a\right)^2 - m_\chi^2 m_a^2} \\ \times d\sigma_{\chi+a\to i+j+...} f_\chi f_a \\ - \left(i+j+...\to\chi+a\right), \end{aligned} \tag{4.16}$$

where the term in parentheses represents the equivalent expression for the reverse process $(i + j + ... \to \chi + a)$. Also note that in Equation 4.16 we have included a factor of $1/S_{id}$, where S_{id} contains a factor of $k!$ for every k indistinguishable particles in either the initial or final states, and the factor C_1 contains a factor of l for every l number of particles of species χ that are either eliminated (or added in the case of negative C_1) by the $(\chi + a \to i + j + ...)$ process. The equivalent term for the reverse

$(i + j + ... \to \chi + a)$ process contains similar factors.

We can rewrite Equation 4.16 in terms of the relativistic Møller velocity:

$$\begin{aligned}\frac{dn_\chi}{dt} + 3Hn_\chi = &- C_1 \frac{1}{S_{id}} g_\chi g_a \int \frac{d^3\mathbf{p}_\chi}{(2\pi)^3} \frac{d^3\mathbf{p}_a}{(2\pi)^3} \left(v_{\text{Møll}}\right)_{\chi a} d\sigma_{\chi+a\to i+j+...} f_\chi f_a \\ &- (i + j + ... \to \chi + a)\,,\end{aligned} \tag{4.17}$$

where the relativistic Møller velocity is defined as:

$$\left(v_{\text{Møll}}\right)_{\chi a} = \frac{\sqrt{\left(p_\chi \cdot p_a\right)^2 - \left(m_\chi m_a\right)^2}}{E_\chi E_a}. \tag{4.18}$$

From now on I'll drop all the subscripts on $\left(v_{\text{Møll}}\right)_{\chi a}$ and write it as v, unless it's not clear in context. Finally, we can write Equation 4.17 as

$$\frac{dn_\chi}{dt} + 3Hn_\chi = -C_1 n_\chi n_a \langle\sigma v\rangle_{\chi+a\to i+j+...} + C_2 n_i n_j ... \langle\sigma v^{n-1}\rangle_{i+j+...\to\chi+a}, \tag{4.19}$$

where n is the number of particles $i, j, ...$ in the initial state of the $i + j + ... \to a + \chi$ process. In Equation 4.19 we have defined the thermally averaged quantity:

$$n_\chi n_a \langle\sigma v\rangle_{\chi+a\to i+j+...} = \frac{1}{S_{id}} \sum_{\text{spins}} \int \prod_r \frac{d^3\mathbf{p}_r}{(2\pi)^3 2E_r} (2\pi)^4 \delta^4\left(\sum p\right) f_\chi f_a \left|\mathcal{M}\right|^2, \tag{4.20}$$

where r runs over all initial and final states.

It is helpful to define the thermally averaged cross section $\langle\sigma v\rangle$ according to the discussion above. Assuming Maxwell-Boltzmann statistics for all particle species, we can write the phase-space densities as:

$$f_i = \frac{n_i}{n_i^{(0)}} e^{-E_i/T}, \tag{4.21}$$

where $n_i^{(0)}$ represents the equilibrium distribution for particle i, given by:

$$n_i^{(0)}(T) = \frac{g_i m_i^2 T}{2\pi^2} K_2\left(m_i/T\right), \tag{4.22}$$

where K_2 is the modified Bessel function of the second kind. Because we are usually describing dark matter in the galactic halo moving at nonrelativistic velocities, it is useful to have the expression for $n_i^{(0)}$ in the nonrelativistic limit:

$$n_i^{(0)}(T) = g_i \left(\frac{m_i T}{2\pi}\right)^{3/2} \exp\left(-m_i/T\right). \tag{4.23}$$

Making the replacement $f_i = n_i/n_i^{(0)} \exp\left(-E_i/T\right)$ in Equation 4.20, we obtain:

$$\langle \sigma v \rangle_{\chi+a\to i+j+\ldots} = \frac{1}{S_{id}} \sum_{\text{spins}} \int \prod_r \frac{d^3\mathbf{p}_r}{(2\pi)^3 2E_r} (2\pi)^4 \delta^4\left(\sum p\right) \frac{1}{n_\chi^{(0)}} \frac{1}{n_a^{(0)}} |\mathcal{M}|^2, \tag{4.24}$$

and plugging in the full expression for $n_i^{(0)}$ in Equation 4.22, we obtain:

$$\begin{aligned} \langle \sigma v \rangle_{\chi+a\to i+j+\ldots} =& \frac{2\pi^2}{g_\chi m_\chi^2 T K_2\left(m_\chi/T\right)} \frac{2\pi^2}{g_a m_a^2 T K_2\left(m_a/T\right)} \frac{1}{S_{id}} g_\chi g_a \\ & \times \int \prod_r \frac{d^3\mathbf{p}_r}{(2\pi)^3 2E_r} (2\pi)^4 \delta^4\left(\sum p\right) e^{-E_\chi/T} e^{-E_a/T} \langle |\mathcal{M}|^2 \rangle. \end{aligned} \tag{4.25}$$

Using the definition of the cross section in Equation 4.13 we can rewrite Equation 4.25 as:

$$\begin{aligned} \langle \sigma v \rangle_{\chi+a\to i+j+\ldots} =& \frac{2\pi^2}{m_\chi^2 T K_2\left(m_a/T\right)} \frac{2\pi^2}{m_a^2 T K_2\left(m_a/T\right)} \\ & \times \frac{1}{S_{id}} \int \frac{d^3\mathbf{p}_\chi}{(2\pi)^3} \frac{d^3\mathbf{p}_a}{(2\pi)^3} e^{-E_\chi/T} e^{-E_a/T} \sigma v \end{aligned} \tag{4.26}$$

4.2.2 Second moment of the Boltzmann equation: energy density evolution

In Section 4.2.1 above we took the first moment of the Boltzmann equation, as in Equation 4.8, and we obtained the Boltzmann equation for the evolution of the number density n_χ, Equation 4.19. Now let's take the second moment of the Boltzmann equation:

$$g_\chi \int EL[f] \frac{d^3\mathbf{p}}{(2\pi)^3} = g_\chi \int C[f] E \frac{1}{E} \frac{d^3\mathbf{p}}{(2\pi)^3}, \tag{4.27}$$

and proceeding as in Section 4.2.1, skipping all the similar details, we obtain the Boltzmann equation for the energy density evolution:

$$\frac{\partial \rho_\chi}{\partial t}+3H\left(\rho_\chi+P_\chi\right)=-\langle\sigma v^{n-1}\delta E\rangle_{\chi+a+...\to i+j+...}n_\chi n_a...+\left(\begin{array}{c}\text{all other processes}\\ \text{which add or deplete}\\ \text{energy from the } \chi \text{ particles}\end{array}\right), \tag{4.28}$$

where the thermal averaged per-χ-particle energy transfer rate $\langle\sigma v^{n-1}\delta E\rangle_{\chi+a+...\to i+j+...}n_\chi n_a...$ is given by:

$$\begin{aligned}\langle\sigma v^{n-1}\delta E\rangle_{\chi+a+...\to i+j+...}n_\chi n_a...=\frac{1}{S_{id}}\sum_{\text{spins}}\int\prod_r\frac{d^3\mathbf{p}_r}{\left(2\pi\right)^3 2E_r}\\ \times\left(2\pi\right)^4\delta^4\left(p_\chi+p_a+...-p_i-p_j-...\right)\\ \times f_\chi f_a...\delta E\left|\mathcal{M}\right|^2,\end{aligned} \tag{4.29}$$

where r runs over all initial (χ,a,...) and final (i,j,...) states, and δE is the amount of energy depleted from the χ particles in a single $\chi+a+...\to i+j+...$ process. The energy density ρ is given by Equation 4.4. Note that Equation 4.28 is written for the case of a single species of dark sector particle. In the case of multiple dark sector particle species which are in thermal equilibrium with each other Equation 4.28 will contain equivalent terms for all dark sector species. We will see an example of this in Chapters 5 and 6. The term $\langle\sigma v^{n-1}\delta E\rangle_{\chi+a+...\to i+j+...}n_\chi n_a...$ represents any process which takes place among the χ and other particles in the initial state (χ, a, ...) which changes the energy in the dark sector χs by δE. All such processes which change the dark sector energy are included as source terms in Equation 4.28, as expressed by the term in parentheses. For example, specializing to the case in which the only interaction which changes the energy content of the χs is an elastic scattering process $\chi+f\to\chi+f$, Equation 4.28 takes the form:

$$\frac{\partial \rho_\chi}{\partial t}+3H\left(\rho_\chi+P_\chi\right)=-\langle\sigma v\delta E\rangle_{\chi f\to\chi f}n_\chi n_f. \tag{4.30}$$

Again Equation 4.28 is easier to solve when we assume Maxwell-Boltzmann distributions for all species, which allows us to relate the energy density and pressure to their equilibrium values via the number density n_χ, according to the relations:

$$\begin{aligned} \rho_\chi &= \left(\frac{n_\chi}{n_\chi^{(0)}}\right) \rho_\chi^{(0)} \\ P_\chi &= \left(\frac{n_\chi}{n_\chi^{(0)}}\right) P_\chi^{(0)}, \end{aligned} \tag{4.31}$$

where the equilibrium distributions $\rho_\chi^{(0)}$ and $P_\chi^{(0)}$ are given by:

$$\begin{aligned} \rho_\chi^{(0)} &= \frac{g_\chi m_\chi^2 T}{2\pi^2} \left(m_\chi K_1\left(m_\chi/T\right) + 3T K_2\left(m_\chi/T\right)\right) \\ P_\chi^{(0)} &= \frac{g_\chi m_\chi^2 T^2}{2\pi^2} K_2\left(m_\chi/T\right). \end{aligned} \tag{4.32}$$

Equations 4.28 and 4.19 give a complete set of coupled Boltzmann equations which we can use to solve for the evolution of n_χ and the dark sector temperature T', and therefore solve for the thermal history of the dark sector before thermal freezeout of the dark matter.

In practice when solving the Boltzmann equations we need to have calculated the thermal averaged cross sections appearing as source terms in Equation 4.19 and the thermal averaged energy transfer rates appearing as source terms in Equation 4.28. Since deriving these from scratch always goes the same up until we need to plug in the specific form of the matrix element, I'll summarize their derivations in Section 4.4 as a nice reference to go back to later. We also need to explicitly rewrite the Boltzmann equations in terms of a single independent variable. It's usually convenient to choose the independent variable to be the SM temperature T (or an equivalent variable, such as $x \equiv m_\chi/T$). Then we solve the Boltzmann equations for the evolution of the DM number density $n_\chi\left(T\right)$ and dark sector temperature $T'\left(T\right)$ as functions of the SM temperature T. Rewriting the left-hand-side of the Boltzmann equations explicitly as functions of T is also explained in Section 4.4.

Before going into those details required to numerically solve the Boltzmann equa-

tions in Section 4.4, based on what I've already discussed so far about calculating the thermal history of the dark sector we can understand a few important dark matter production mechanisms.

4.3 Dark matter production

Now that we understand the overall structure of the Boltzmann equations in Section 4.2, we can analytically calculate the thermal history of dark matter using some parametric estimates and see how its relic abundance is produced. Let's use this approach to summarize a few important dark matter production mechanisms.

4.3.1 Thermal WIMP

Let's start with the simplest thermal relic, which has also received the most attention, the thermal weakly-interacting-massive-particle (WIMP). That is, let's consider a dark matter particle χ with its corresponding antiparticle $\bar{\chi}$ which interacts with the SM and freezes out through $2 \to 2$ annihilations of DM to SM particles ($\chi\bar{\chi} \to$ SM+SM). What I mean by "freeze out" will be defined more precisely in this section. Note that another definition used to describe WIMPs is a DM particle which interacts with the SM via the SM weak interaction; I'm not using this definition here, as I'm not specifying the portal by which the DM interacts with the SM to produce the $2 \to 2$ annihilations.

To describe the thermal history of this scenario we start with the Boltzmann equation:

$$\frac{dn_\chi}{dt} + 3Hn_\chi = -n_\chi^2 \langle \sigma v \rangle_{\chi\bar{\chi}\to \mathrm{SM+SM}} + n_{\mathrm{SM}}^2 \langle \sigma v \rangle_{\mathrm{SM+SM}\to\chi\bar{\chi}}. \tag{4.33}$$

Note that there is a corresponding equation for $\bar{\chi}$, but in the absence of an asymmetry we can assume $n_\chi = n_{\bar{\chi}}$ and the evolution is identical. Up until freezeout the DM χ and its SM annihilation products are in chemical equilibrium, so we can relate the

rates of the forward and reverse processes in Equation 4.33 using detailed balance:

$$n_\chi^2 \langle \sigma v \rangle_{\chi\bar{\chi}\to \mathrm{SM+SM}} \simeq n_{\mathrm{SM}}^2 \langle \sigma v \rangle_{\mathrm{SM+SM}\to\chi\bar{\chi}}, \tag{4.34}$$

up to order $\mathcal{O}(n_\chi H)$. We can use this to rewrite Equation 4.33 as:

$$\frac{dn_\chi}{dt} + 3Hn_\chi \simeq -\left(n_\chi^2 - \left(n_\chi^{(0)}\right)^2\right) \langle \sigma v \rangle_{\chi\bar{\chi}\to \mathrm{SM+SM}}. \tag{4.35}$$

In order to fully solve the thermal history of the WIMP in this scenario we need to solve Equation 4.35. Let's make some parametric estimates to get a nice analytic picture of the thermal history. Again, when at early times the annihilation rate $n_\chi \langle \sigma v \rangle_{\chi\bar{\chi}\to \mathrm{SM+SM}}$ is large the DM is in equilibrium with the SM, and the rates of the forward ($\chi\bar{\chi} \to \mathrm{SM} + \mathrm{SM}$) and reverse ($\mathrm{SM} + \mathrm{SM} \to \chi\bar{\chi}$) processes cancel down to $\mathcal{O}(H)$. Therefore at early times n_χ is driven towards its equilibrium value, $n_\chi \simeq n_\chi^{(0)}$, given by Equation 4.23 (we take the non-relativistic limit for DM since typical velocities in the galactic halo are $\sim 10^{-3}$). As the universe expands and cools, the rate of annihilations, $n_\chi^{(0)} \langle \sigma v \rangle_{\chi\bar{\chi}\to \mathrm{SM+SM}}$ falls exponentially as $\sim \exp(-m_\chi/T)$. This proceeds up until the point at which the rate of annihilations is no longer sufficient to maintain chemical equilibrium between the DM and the SM. When this happens the DM freezes out to a constant comoving number density $d(na^3)/dt = 0$. The relevant scale against which to compare the rate of $2 \to 2$ annihilations to determine when freezeout takes place is the Hubble rate H; freezeout of the DM occurs when the time needed for a pair of χ and $\bar{\chi}$ particles to find each other and annihilate becomes longer than the timescale of Hubble expansion:

$$n_\chi(T_f) \langle \sigma v \rangle_{\chi\bar{\chi}\to \mathrm{SM+SM}} \sim H(T_f). \tag{4.36}$$

Let's make the approximation of instantaneous freezeout, in which we assume that up until the moment of freezeout the DM follows its equilibrium distribution $n_\chi^{(0)}$, and immediately after freezeout it evolves with constant comoving number density, $n_\chi(T < T_f) \propto a^{-3}$. Under this assumption we can write the freezeout condition in

Equation 4.36 as:

$$n_\chi^{(0)}\left(T_f\right)\left\langle\sigma v\right\rangle_{\chi\bar{\chi}\to\mathrm{SM+SM}} \sim H\left(T_f\right), \tag{4.37}$$

with $n_\chi^{(0)}$ given by Equation 4.23 (we assume that DM freezes out while nonrelativistic since the case of DM that freezes while relativistic results in changes to structure formation inconsistent with observations). We also assume that freezeout occurs during the radiation-dominated epoch, since a scenario where freezeout occurs after matter-radiation equality also results in changes to structure formation inconsistent with observations, and therefore, from the Friedmann equation:

$$H \sim \frac{1}{M_{\mathrm{pl}}}\sqrt{\rho} \sim \frac{1}{M_{\mathrm{pl}}}T^2. \tag{4.38}$$

Then we can write Equation 4.37 as:

$$\begin{aligned} n_\chi^{(0)}\left(T_f\right)\left\langle\sigma v\right\rangle_{\chi\bar{\chi}\to\mathrm{SM+SM}} &\sim H\left(x=1\right)x_f^{-2} \\ &\sim \frac{1}{M_{\mathrm{pl}}}T_f^2 \end{aligned} \tag{4.39}$$

where I introduced the quantity $x \equiv m_\chi/T$ (and $x_f \equiv m_\chi/T_f$). Now, ignoring changes in the number of degrees of freedom coupled to the thermal bath, the number density of photons and DM both redshift as a^{-3} with the Hubble expansion after freezeout:

$$\frac{m_\chi n_\chi}{n_\gamma}\left(x > x_f\right) \simeq \frac{m_\chi n_\chi^{(0)}}{n_\gamma}\left(x_f\right) \sim \frac{1}{M_{\mathrm{pl}}}\frac{x_f}{\left\langle\sigma v\right\rangle_{\chi\bar{\chi}\to\mathrm{SM+SM}}}. \tag{4.40}$$

Let's now fix ρ_χ/n_γ to its present day value. We know the DM mass density is about 5 times the baryon mass density, $\rho_{\mathrm{DM}} \sim 5\times\rho_b$, and that the baryon-to-photon ratio is $\sim 5\times10^{-10}$, so $\rho_{\mathrm{DM}} \sim 5\times n_\gamma \times 5\times10^{-10}$ GeV, and since $M_{\mathrm{pl}} \sim 10^{19}$ GeV, we obtain:

$$\frac{\rho_\chi}{n_\gamma} \sim 3\times10^{-9}\ \mathrm{GeV} \sim \frac{x_f}{\left\langle\sigma v\right\rangle_{\chi\bar{\chi}\to\mathrm{SM+SM}}}10^{-19}\ \mathrm{GeV}^{-1}, \tag{4.41}$$

which rearranged, and plugging in the present-day value of $n_\gamma = 410.7(3)\ \mathrm{cm}^{-3}$, gives:

$$\left\langle\sigma v\right\rangle_{\chi\bar{\chi}\to\mathrm{SM+SM}} \sim x_f \times 10^{-10.5}\ \mathrm{GeV}^{-2} \tag{4.42}$$

Now, since $n_\chi^{(0)}$ falls exponentially for $x > 1$ let's first assume $x_f \sim \mathcal{O}(1)$. If we assume that the SM particles in the $2 \to 2$ annihilations are much lighter than the DM χ, then parametrically the thermal averaged cross section scales as $\langle \sigma v \rangle_{\chi\bar{\chi}\to \mathrm{SM+SM}} \sim \alpha_D^2/m_\chi^2$, where α_D is the dark sector coupling, and we get:

$$\frac{\alpha_D^2}{m_\chi^2} \sim 10^{-10.5}\ \mathrm{GeV}^{-2}. \tag{4.43}$$

In this rough approximation we can already see the weak scale emerge. If we take $\alpha_D \sim 0.01$ comparable to the SM electroweak coupling, then we find $m_\chi \sim 10^3$ GeV, at the weak scale.

We can now get a better approximation for x_f. Starting with Equation 4.39 and rearranging, and dropping the $\log x_f$ compared to x_f, we have:

$$\begin{aligned} x_f &\sim \log\left(\frac{m_\chi^3 \langle \sigma v \rangle_{\chi\bar{\chi}\to \mathrm{SM+SM}}}{H\,(x=1)}\right) \\ &\sim \log\left(\frac{m_\chi^3 \alpha_D^2}{m_\chi^2 H\,(x=1)}\right), \end{aligned} \tag{4.44}$$

and again plugging in the values of $\alpha_D \sim 0.01$, $m_\chi \sim 10^3$ GeV, we get $x_f \sim 25$. If we plug this back into Equation 4.41 we find $\langle \sigma v \rangle_{\chi\bar{\chi}\to \mathrm{SM+SM}} \sim 10^{-26}$ cm^3/s for DM with a mass $m_\chi \sim$ TeV.

Because I think it's useful to have written here I'll repeat the same calculation above taking a different approach which will make it more clear how the DM mass scales with known quantities in the WIMP freezeout scenario. Let's start again with the freezeout condition Equation 4.36:

$$n_\chi\,(T_f)\,\langle \sigma v \rangle_{\chi\bar{\chi}\to \mathrm{SM+SM}} \sim H\,(T_f) \sim \frac{T_f^2}{M_{\mathrm{pl}}}. \tag{4.45}$$

After freezeout the DM number density dilutes with the Hubble expansion as $n_\chi \propto a^{-3} \propto T^3$. And at the period of matter-radiation equality the DM mass density and

radiation energy density are equal:

$$m_\chi n_\chi\left(T_{\rm MRE}\right) \sim m_\chi n_\chi\left(T_f\right) \frac{T_{\rm MRE}^3}{T_f^3} \sim T_{\rm MRE}^4, \tag{4.46}$$

which we can rearrange to write $n_\chi\left(T_f\right)$ as:

$$n_\chi\left(T_f\right) \sim \frac{T_f^3 T_{\rm MRE}}{m_\chi}, \tag{4.47}$$

which we can plug into Equation 4.45 to obtain the freezeout condition:

$$\langle\sigma v\rangle_{\chi\bar{\chi}\to{\rm SM+SM}} \sim \frac{m_\chi}{M_{\rm pl} T_f T_{\rm MRE}}. \tag{4.48}$$

With $x_f \sim 25$ Equation 4.48 becomes:

$$\langle\sigma v\rangle_{\chi\bar{\chi}\to{\rm SM+SM}} \sim \frac{25}{M_{\rm pl} T_{\rm MRE}} \sim 10^{-9}\ {\rm GeV}^{-2} \sim 10^{-26}{\rm cm}^3/{\rm s}. \tag{4.49}$$

If the thermally averaged cross section scales as $\langle\sigma v\rangle_{\chi\bar{\chi}\to{\rm SM+SM}} \sim \alpha_D^2/m_\chi^2$, the freeze-out condition becomes:

$$m_\chi \sim \alpha_D \sqrt{\frac{M_{\rm pl} T_{\rm MRE}}{25}}. \tag{4.50}$$

If we take $\alpha_D \sim 0.01$ at the SM electroweak scale, then we again find $m_\chi \sim$TeV.

A more exact calculation is made by solving Equation 4.35, which yields a thermal relic cross section for the WIMP of $\langle\sigma v\rangle_{\chi\bar{\chi}\to{\rm SM+SM}} \simeq 2 - 3 \times 10^{-26}\ {\rm cm}^3/{\rm s}$, almost independent of the DM mass (see e.g. Ref. [95]). The natural emergence of the weak scale in the freezeout of the thermal WIMP, where new physics had been expected, is called the "WIMP miracle".

4.3.2 Beyond the Thermal WIMP: light DM

With increasingly strong experimental constraints being placed on the thermal WIMP scenario, we are motivated to consider alternative scenarios where other interactions control the final DM abundance. There has been considerable recent interest in ex-

ploring thermal relic scenarios that naturally produce DM at light (sub-GeV) masses (see for example Ref. [96]), as existing direct detection constraints are much less sensitive to sub-GeV mass DM (e.g. Refs. [97–100]). Existing beam dump experiments are sensitive to sub-GeV DM but leave much of the parameter space unconstrained [101–110]. New accelerator and direct detection experiments are projected to soon explore the parameter space of light DM with unprecedented sensitivity (see Ref. [96] and references therein), so it's an important time to understand the landscape of models which naturally populate this sub-GeV region. In the following two sections I'll describe two previously studied mechanisms for producing light (sub-GeV) DM: the SIMP and ELDER scenarios. Thinking about these two mechanisms will help us introduce some concepts which will help us introduce the material of Chapters 5 and 6.

4.3.3 Strongly-Interacting-Massive-Particle (SIMP) DM

One identified mechanism for thermally producing sub-GeV DM is the strongly-interacting-massive-particle (SIMP) scenario in which strong $3 \to 2$ self-annihilations among DM particles control the thermal relic abundance [92]. Let's take a brief look at how this scenario works.

In addition to the annihilation process ($\chi\chi \to \text{SM}+\text{SM}$) process which determines the DM relic abundance in the thermal WIMP scenario, the DM particle χ can interact with the SM via elastic scattering ($\chi + \text{SM} \to \chi + \text{SM}$), and with itself via self-annihilations ($\chi\chi \leftrightarrow \chi\chi ... \chi$, where three or more DM particles are represented on the right-hand-side of the arrow). Let's imagine a scenario in which the interactions which control the thermal history of the DM are elastic scattering and $3 \to 2$ self-annihilations ($\chi\chi\chi \leftrightarrow \chi\chi$). Then the relevant Boltzmann equations are:

$$\frac{dn_\chi}{dt} + 3Hn_\chi \simeq -\langle\sigma v^2\rangle_{3\to2}\left(n_\chi^3 - n_\chi^2 n_\chi^{(0)}\right) \tag{4.51}$$

and

$$\frac{d\rho_\chi}{dt} + 3H\left(\rho_\chi + P_\chi\right) \simeq \langle\sigma v \delta E\rangle_{\text{el}} n_\chi n_\gamma^{(0)}, \tag{4.52}$$

where we've assumed that the dominant coupling between the DM and SM in the elastic scattering process is to the photon. For now let's assume that the elastic scattering process is efficient enough throughout freezeout to keep the dark and SM sectors thermally coupled, then we can focus just on the Boltzmann Equation 4.51 (in the next section we'll see what interesting things can happen when we break this assumption).

We can make some parametric estimates like in Section 4.3.1 to see what energy scale of the DM gives the observed relic abundance in this scenario. In this case the $\chi\chi\chi \leftrightarrow \chi\chi$ process controls thermal freezeout, so the freezeout condition is:

$$\langle \sigma v^2 \rangle_{3\to 2} \left(n_\chi^{(0)} \left(T_f \right) \right)^2 \sim H \sim \frac{T_f^2}{M_{\rm pl}}, \tag{4.53}$$

where again we have assumed instantaneous freezeout which occurs during the radiation-dominated era.

After freezeout the number density of DM dilutes with the Hubble expansion as $n_\chi \left(T < T_f \right) \propto a^{-3} \propto T^3$. And at the period of matter-radiation equality the DM mass density and radiation energy density are equal. Then we can write the relation:

$$m_\chi n_\chi \left(T_{\rm MRE} \right) \sim m_\chi n_\chi^{(0)} \left(T_f \right) \frac{T_{\rm MRE}^3}{T_f^3} \sim T_{\rm MRE}^4, \tag{4.54}$$

in order to write $n_\chi^{(0)} \left(T_f \right)$ as:

$$n_\chi^{(0)} \left(T_f \right) \sim \frac{T_f^3 T_{\rm MRE}}{m_\chi}, \tag{4.55}$$

which we can plug into Equation 4.53 to obtain the freezeout condition:

$$\langle \sigma v^2 \rangle_{3\to 2} \sim \frac{1}{M_{\rm pl}} \frac{m_\chi^2}{T_f^4 T_{\rm MRE}^2}. \tag{4.56}$$

As in the case of the thermal WIMP freezeout occurs at $x_f \sim 20$, so Equation 4.56 becomes:

$$\langle \sigma v^2 \rangle_{3\to 2} \sim \frac{1}{M_{\rm pl}} \frac{10^5 m_\chi^{-2}}{T_{\rm MRE}^2}. \tag{4.57}$$

By dimensional analysis we expect $\langle \sigma v^2 \rangle_{3\to 2}$ to scale as:

$$\langle \sigma v^2 \rangle_{3\to 2} \sim \frac{\alpha_{\text{eff}}^3}{m_\chi^5}, \tag{4.58}$$

where α_{eff} is the dark sector coupling. Plugging this into Equation 4.57 and rearranging we the freezeout condition:

$$m_\chi \sim 10^{-5/3} \alpha_{\text{eff}} \left(M_{\text{pl}} T_{\text{MRE}}^2 \right)^{1/3}, \tag{4.59}$$

and plugging in the values $M_{\text{pl}} \simeq 10^{19}$ GeV and $T_{\text{MRE}} \simeq 1$ eV, we find:

$$m_\chi \sim 0.05 \alpha_{\text{eff}} \text{ GeV}. \tag{4.60}$$

The SIMP scenario naturally leads to strongly-coupled DM ($\alpha_{\text{eff}} \sim 1$) with mass similar to the QCD scale ($m_\chi \sim 10 - 100$ MeV).

4.3.4 Elastically Decoupling Relic (ELDER) DM

Now let's imagine a scenario in which the DM has the same dominant interactions as in the SIMP case: elastic scattering ($\chi + \text{SM} \to \chi + \text{SM}$) and $3 \to 2$ self-annihilations ($\chi\chi\chi \to \chi\chi$). Again the relevant Boltzmann equations are Equations 4.51 and 4.52. However this time let's see what happens when the elastic scattering process which maintains thermal equilibrium between the dark and SM sectors is no longer efficient enough to keep the two sectors thermally coupled throughout freezeout; that is, whereas in the SIMP scenario $3 \to 2$ self-annihilations decouple before elastic scattering, in this case the elastic scattering process decouples before $3 \to 2$ self-annihilations. Before elastic decoupling the system evolves as in the SIMP case: $3 \to 2$ self-annihilations maintain chemical equilibrium in the dark sector while elastic scattering maintains thermal equilibrium between the dark and SM sectors; the dark and SM sectors evolve with equal temperatures while the number density of DM depletes through $3 \to 2$ self-annihilations and redshifts with the Hubble expansion, evolving according to Equation 4.23 (in the nonrelativistic limit $x > 1$). This be-

havior proceeds up until elastic decoupling occurs, when the rate of elastic scattering is no longer efficient enough to maintain thermal equilibrium between the dark and SM sectors. After elastic decoupling, the dark and SM sectors evolve with separate temperatures T' and T, respectively. Elastic decoupling is defined by the condition:

$$m_\chi R_{\rm el} < \dot{K}_\chi \tag{4.61}$$

where $R_{\rm el}$ defines the rate of elastic scattering, given by:

$$R_{\rm el} \equiv \langle \sigma v \delta E \rangle_{\chi+{\rm SM}\to\chi+{\rm SM}} n_\gamma\left(T\right) \tag{4.62}$$

and $\dot{K}_\chi$ is the per-χ-particle rate at which the $3 \to 2$ self-annihilations release kinetic energy into the dark sector, given by:

$$\dot{K}_\chi \equiv m_\chi \frac{\dot{n}_\chi\left(T'\right)}{n_\chi}. \tag{4.63}$$

It is important to understand how the temperature of the dark sector T' evolves after kinetic decoupling ($T' < T'_d$), so let's focus on the Boltzmann equation describing the energy density of the dark sector, Equation 4.52. After kinetic decoupling the rate of elastic scattering is slower than the Hubble expansion (more specifically, $m_\chi R_{\rm el} < \dot{K}_\chi$, and one can show that this is nearly equivalent to the condition that $R_{\rm el} < H$), and therefore it's a reasonable approximation to drop the source terms on the right-hand-side of Equation 4.52:

$$\frac{d\rho_\chi\left(T'\right)}{dt} + 3H\left(\rho_\chi\left(T_\prime\right) + P_\chi\left(T_\prime\right)\right) \simeq 0. \tag{4.64}$$

As we will see more explicitly below, Equation 4.64 says that the comoving entropy density of the dark sector is separately conserved after kinetic decoupling, since the dark and SM sectors can no longer efficiently transfer energy/entropy between each other.

Let's also restrict the discussion to the case in which elastic decoupling occurs while

the DM is nonrelativistic ($x_d > 1$), where $P_\chi \simeq nT \ll \rho_\chi$, and in the nonrelativistic limit $\rho_\chi \simeq m_\chi n_\chi$, so we can approximate Equation 4.64 as:

$$\frac{dm_\chi n_\chi\left(T'\right)}{dt} + 3Hm_\chi n_\chi\left(T'\right) \simeq 0. \tag{4.65}$$

Because kinetic decoupling occurs while the $3 \to 2$ self-annihilations are still active, they maintain chemical equilibrium with zero chemical potential in the dark sector, and the number density of DM follows its equilibrium distribution, $n_\chi\left(T'\right) \simeq n_\chi^{(0)}\left(T'\right)$, which in the nonrelativistic limit is given by Equation 4.23. The conservation of total entropy (DM+SM) can also be used to derive the relation:

$$\begin{aligned} \frac{\partial}{\partial t} &= \frac{xH\left(T\right)}{1 + \frac{1}{3}\frac{T}{g_{\text{eff,s}}}\frac{\partial g_{\text{eff,s}}}{\partial T}}\frac{\partial}{\partial x} \\ &\simeq xH\left(T\right)\frac{\partial}{\partial x} \end{aligned} \tag{4.66}$$

where we approximate by neglecting effects of the number of relativistic degrees of freedom $g_{\text{eff,s}}$. It's also useful to have the expression for the time derivative of $n_\chi^{(0)}\left(T'\right)$ in the nonrelativistic limit:

$$\frac{\partial}{\partial x}n_\chi^{(0)}\left(T'\right) = n_\chi^{(0)}\left(T'\right)\left(\frac{3}{2}\frac{1}{T'} + \frac{m_\chi}{\left(T'\right)^2}\right)\frac{\partial T'}{\partial x}. \tag{4.67}$$

We can then directly plug in Equations 4.23, 4.66, and 4.67 into Equation 4.65, dropping the first term in parentheses in Equation 4.67 in the nonrelativistic ($x_d > 1$) limit, to obtain the following relation between T' and T after kinetic decoupling x_d:

$$\frac{\partial T'}{\partial T} \simeq \frac{3\left(T'\right)^2}{m_\chi T}. \tag{4.68}$$

It's helpful for our intuition to note that Equation 4.68 says that the comoving entropy density is conserved in the dark sector after kinetic decoupling. We can also derive Equation 4.68 directly from the conservation of comoving entropy density of the dark sector:

$$\frac{\partial}{\partial t}\left(s_D a^3\left(T\right)\right) = 0 \tag{4.69}$$

where the dark sector entropy density s_D is given by:

$$s_D = \frac{m_\chi n_\chi (T')}{T'}. \tag{4.70}$$

Plugging in Equations 4.23, 4.66, and 4.67 into Equation 4.69, again dropping the first term in parentheses in Equation 4.67 in the nonrelativistic ($x_d > 1$) limit, we obtain Equation 4.68.

Now we can integrate Equation 4.68 to obtain the relation between T' and T after kinetic decoupling while the $3 \to 2$ process is still efficient. Doing so, we find the relation:

$$T' \simeq \frac{T_d}{1 - \frac{3T_D}{m_\chi} \ln (T/T_d)}, \tag{4.71}$$

that is, while the SM temperature redshifts as $T \propto a^{-1}$, the DM temperature evolves much more slowly, scaling logarithmically with T as $T' \propto 1/\log a$. This nearly constant scaling of T' with a is due to a process called "cannibalization" [111]: after kinetic decoupling the dark and SM sectors can no longer efficiently exchange entropy, and the comoving entropy density of the dark sector is conserved. At the same time the active $3 \to 2$ self-annihilations convert mass to kinetic energy in the dark sector, heating it up relative to the SM bath such that T' remains nearly constant, with only subleading (logarithmic) dependence on a, as in Equation 4.71. Since T' remains nearly constant during the period of cannibalization, likewise the comoving number density of DM remains nearly constant during this period. Cannibalization of the dark sector proceeds after kinetic decoupling up until the freezeout of the $3 \to 2$ self-annihilations, which signals the freezeout of the DM. However, since the comoving number density of DM remains nearly constant during cannibalization, the DM relic abundance is mostly set by the elastic decoupling process; that is, the relic density of DM in this scenario is controlled by the strength of the elastic scattering process, with only subleading (logarithmic) dependence on the strength of the $3 \to 2$ self-annihilations. The ELDER scenario also leads to light DM, with m_χ ranging from a few to ~ 100 MeV, with $\mathcal{O}(1)$ strength of self-interactions in the dark sector [93, 94].

It is important to note that in starting with the thermal SIMP scenario [92] of Section 4.3.3 and relaxing the assumption of thermal equilibrium between the dark and SM sectors during the freezeout of DM in the ELDER scenario [93, 94], a novel freezeout history is revealed in which the relic abundance of DM has drastically different dependence on the interaction processes and model parameters. This will be a useful lesson to keep in mind when we get into Chapters 5 and 6.

4.4 Numerically solving the Boltzmann equations

To calculate the thermal history of dark matter more accurately than the parametric estimates we made above, we have to numerically solve the coupled Boltzmann Equations 4.19 and 4.28. This section provides the tools we need to numerically solve the Boltzmann equations. First in Section 4.4.1 I show how to rewrite the Boltzann Equations 4.19 and 4.28 explicitly in terms of the independent variable Standard Model temperature T. Then in Sections 4.4.2, 4.4.3, and 4.4.4 I show how to calculate the thermally averaged cross sections for a $3 \to 2$ process and a $2 \to 2$ process (both in the threshold approximation), and the thermally averaged energy transfer rate $\langle \sigma v \delta E \rangle$ for an elastic scattering process ($\chi f \to \chi f$), respectively.

4.4.1 Rewriting the Boltzmann equations explicitly as functions of a single independent variable Standard Model temperature T

In order to numerically solve the coupled set of Boltzmann Equations 4.19 and 4.28 we need to first write these as explicit functions of the independent variable T, (or an equivalent variable, such as $x \equiv m_\chi/T$). We need to prove the relation:

$$\frac{\partial}{\partial t} = \frac{xH}{1 + \frac{1}{3}\frac{T}{g_{\text{eff},s}}\frac{\partial g_{\text{eff},s}}{\partial T}} \frac{\partial}{\partial x}, \tag{4.72}$$

where $g_{\text{eff},s}$ are the effective entropy degrees of freedom, and all quantities on the right-hand-side of Equation 4.72 are functions of T.

In order to derive Equation 4.72 we can first rewrite:

$$\frac{\partial}{\partial t} = \frac{\partial a}{\partial t}\frac{\partial}{\partial a} = Ha\frac{\partial}{\partial a} = Ha\frac{\partial T}{\partial a}\frac{\partial}{\partial T}. \tag{4.73}$$

Now, to find an expression for $\partial T/\partial a$, use the relation describing the conservation of total entropy of the system:

$$\begin{aligned} sa^3 &= \text{constant} \\ g_{\text{eff},s}\left(T\right)T^3a^3 &= \text{constant} \\ g_{\text{eff},s}^{1/3}\left(T\right)Ta &= \text{constant}, \end{aligned} \tag{4.74}$$

where s is the total entropy density $s = \frac{2\pi^2}{45}g_{\text{eff},s}T^3$. Take the time derivative of Equation 4.74:

$$\begin{aligned} \frac{\partial}{\partial t}\left(g_{\text{eff},s}^{1/3}\left(T\right)Ta\right) &= 0 \\ Ha\frac{\partial T}{\partial a}\frac{\partial}{\partial T}\left(g_{\text{eff},s}^{1/3}\left(T\right)Ta\right) &= 0 \end{aligned} \tag{4.75}$$

which can be rearranged to obtain:

$$\frac{\partial a}{\partial T} = -\frac{a}{T}\left(1 + \frac{1}{3}\frac{T}{g_{\text{eff},s}}\frac{\partial g_{\text{eff},s}}{\partial T} + 1\right) \tag{4.76}$$

which together with Equation 4.73 yields Equation 4.72, which can be further rewritten as:

$$\frac{\partial}{\partial t} = \frac{xH}{1 - \frac{1}{3}\frac{x}{g_{\text{eff},s}(T)}\frac{\partial g_{\text{eff},s}(T)}{\partial x}}\frac{\partial}{\partial x} \tag{4.77}$$

This relation is clearly a function of x (or T).

Rewriting the Boltzmann Equation for n in terms of the independent variable T

With Equation 4.77 we can rewrite the left-hand-side of the Boltzmann Equation 4.19 as:

$$\frac{\partial n_{\chi}\left(T'\right)}{\partial t}+3H\left(T\right)n_{\chi}\left(T'\right)=3H\left(T\right)n_{\chi}\left(T'\right)+\frac{xH}{1-\frac{1}{3}\frac{x}{g_{\mathrm{eff},s}(T)}\frac{\partial g_{\mathrm{eff},s}(T)}{\partial x}}\frac{\partial n_{\chi}\left(T'\right)}{\partial x}. \quad (4.78)$$

Because n_{χ} dilutes exponentially with the Hubble expansion, it is usually more numerically tractable to solve Equation 4.78 in terms of the dependent variable $y\left(T\right)\equiv n_{\chi}\left(T'\right)/n_{\chi}^{(0)}\left(T\right)$, with the effects of the Hubble expansion cancelling out of the n_{χ} and $n_{\chi}^{(0)}$ factors in $y\left(T\right)$ (often the similar variable $Y\equiv n_{\chi}/s$ is used). Rewriting Equation 4.78 in terms of $y\left(T\right)$, we get:

$$\frac{\partial n_{\chi}\left(T'\right)}{\partial t}+3H\left(T\right)n_{\chi}\left(T'\right)=3H\left(T\right)n_{\chi}^{(0)}\left(T\right)y\left(T\right)+\frac{xH}{1-\frac{1}{3}\frac{x}{g_{\mathrm{eff},s}(T)}\frac{\partial g_{\mathrm{eff},s}(T)}{\partial x}}\frac{\partial\left(n_{\chi}^{(0)}\left(T\right)y\left(T\right)\right)}{\partial x}. \quad (4.79)$$

At this point I would like to clear up any possible confusion about whether we can treat $y\left(T\right)\equiv n_{\chi}\left(T'\right)/n_{\chi}^{(0)}\left(T\right)$ as only a function of T, or if we need to handle it explicitly as a function of $y\left(T,T'\left(T\right)\right)$ when taking the derivative in Equation 4.79; that is, can we treat the partial derivative in Equation 4.79 as a total derivative d/dx? To show that we can, let's expand the derivative term in Equation 4.79 as a total derivative with respect to x (or T), and show that it is equal to the derivative term in Equation 4.78; that is let's show the following relation is true:

$$\left(\frac{dy\left(T\right)}{dT}n_{\chi,(0)}\left(T\right)+y\left(T\right)\frac{dn_{\chi,(0)}\left(T\right)}{dT}\right)=\frac{dn_{\chi}\left(T'\right)}{dT'}\frac{dT'}{dT}. \quad (4.80)$$

Expanding the left-hand-side of that relation, we get:

$$\begin{aligned}&\left(\frac{dn_{\chi}\left(T'\right)}{dT'}\frac{dT'}{dT}\frac{1}{n_{\chi,(0)}\left(T\right)}-\frac{n_{\chi}\left(T'\right)}{n_{\chi,(0)}^{2}\left(T\right)}\frac{dn_{\chi,(0)}\left(T\right)}{dT}\right)n_{\chi,(0)}\left(T\right)\\&+\frac{n_{\chi}\left(T'\right)}{n_{\chi,(0}\left(T\right)}\frac{dn_{\chi,(0)}\left(T\right)}{dT}=\frac{dn_{\chi}\left(T'\right)}{dT'}\frac{dT'}{dT},\end{aligned} \quad (4.81)$$

the last two terms of Equation 4.81 cancel and we explicitly see that the relation in Equation 4.80 holds. So we see that we can treat $y(T)$ explicitly as a function of only the dependent variable T when solving Equation 4.79.

Let's make the T (or x) dependence in H as it appears in Equation 4.79 more explicit. Because $H \propto \sqrt{g_{\text{eff}}(T)}T^2$, where g_{eff} are the effective energy degrees of freedom, we can write $H(T)$ as:

$$\begin{aligned} H(T) &= H(T = m_\chi)\sqrt{\frac{g_{\text{eff}}(T)}{g_{\text{eff}}(T = m_\chi)}}\frac{T^2}{m_\chi^2} \\ &= H(T = m_\chi)\sqrt{\frac{g_{\text{eff}}(T)}{g_{\text{eff}}(T = m_\chi)}}\frac{1}{x^2}, \end{aligned} \tag{4.82}$$

in terms of which we can finally write Equation 4.79 as:

$$\begin{aligned} &\frac{\partial n_\chi(T')}{\partial t} + 3H(T)\, n_\chi(T') = \\ &H(T = m_\chi)\sqrt{\frac{g_{\text{eff}}(T)}{g_{\text{eff}}(T = m_\chi)}}\frac{1}{x^2}\left[3n_\chi^{(0)}(T)\, y(x) + \left(1 - \frac{1}{3}x\frac{1}{g_{\text{eff},s}(T)}\frac{\partial g_{\text{eff},s}(T)}{\partial x}\right)^{-1} x\frac{\partial}{\partial x}\left(n_\chi^{(0)}(T)\, y(x)\right)\right] \end{aligned} \tag{4.83}$$

Equation 4.83 can be used to rewrite the Boltzmann Equation 4.19 explicitly in terms of the independent variable x (or T).

Rewriting the Boltzmann equation for ρ in terms of the independent variable T

Rewriting the Boltzmann Equation 4.28 for the energy density evolution goes the same as the discussion above. Let's start with the Boltzmann Equation 4.30, for the specific case in which the only interaction which changes the energy content of DM χ is the elastic scattering $\chi + f \to \chi + f$, and rewrite that equation using Equations

4.31 and 4.72:

$$\begin{aligned} & xH\left(T\right)\left(1+\frac{1}{3}\frac{T}{g_{\mathrm{eff},s}\left(T\right)}\frac{\partial g_{\mathrm{eff},s}\left(T\right)}{\partial T}\right)^{-1}\frac{\partial}{\partial x}\left(\frac{n_{\chi}\left(T'\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)\right) \\ & +3H\left(T\right)\left(\frac{n_{\chi}\left(T'\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)+\frac{n_{\chi}\left(T'\right)}{n_{\chi,(0)}\left(T'\right)}P_{\chi,(0)}\left(T'\right)\right) \\ & =\left\langle\sigma v\delta E\right\rangle_{\chi f\rightarrow\chi f}n_{\chi}\left(T'\right)n_{f}\left(T\right). \end{aligned} \tag{4.84}$$

Again we want to substitute in $y\equiv n_{\chi}\left(T'\right)/n_{\chi,(0)}\left(T\right)$, since it is more numerically tractable to factor out the effects of the Hubble expansion in y, so we rewrite Equation 4.84 as:

$$\begin{aligned} & xH\left(T\right)\left(1+\frac{1}{3}\frac{T}{g_{\mathrm{eff},s}\left(T\right)}\frac{\partial g_{\mathrm{eff},s}\left(T\right)}{\partial T}\right)^{-1}\frac{\partial}{\partial x}\left(\frac{n_{\chi}\left(T'\right)}{n_{\chi,(0)}\left(T\right)}\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)\right) \\ & +3H\left(T\right)\left(\frac{n_{\chi}\left(T'\right)}{n_{\chi,(0)}\left(T\right)}\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)+\frac{n_{\chi}\left(T'\right)}{n_{\chi,(0)}\left(T\right)}\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}P_{\chi,(0)}\left(T'\right)\right) \\ & =\left\langle\sigma v\delta E\right\rangle_{\chi f\rightarrow\chi f}\frac{n_{\chi}\left(T'\right)}{n_{\chi,(0)}\left(T\right)}n_{\chi,(0)}\left(T\right)n_{f}\left(T\right) \end{aligned} \tag{4.85}$$

which can be rewritten as:

$$\begin{aligned} & xH\left(T\right)\left(1+\frac{1}{3}\frac{T}{g_{\mathrm{eff},s}\left(T\right)}\frac{\partial g_{\mathrm{eff},s}\left(T\right)}{\partial T}\right)^{-1}\frac{\partial}{\partial x}\left(y\left(T\right)\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)\right) \\ & +3H\left(T\right)\left(y\left(T\right)\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)+y\left(T\right)\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}P_{\chi,(0)}\left(T'\right)\right) \\ & =\left\langle\sigma v\delta E\right\rangle_{\chi f\rightarrow\chi f}y\left(T\right)n_{\chi,(0)}\left(T\right)n_{f}\left(T\right). \end{aligned} \tag{4.86}$$

Again we note that we can treat $y\left(T\right)$ as a function only of the variable T when taking the derivative term in Equation 4.86. Now expand the derivative term in the first line of Equation 4.86:

$$\begin{aligned} & xH\left(T\right)\left(1+\frac{1}{3}\frac{T}{g_{\mathrm{eff},s}\left(T\right)}\frac{\partial g_{\mathrm{eff},s}\left(T\right)}{\partial T}\right)^{-1}\left(\frac{\partial y\left(T\right)}{\partial x}\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)+y\left(T\right)\frac{\partial}{\partial x}\left(\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)\right)\right) \\ & +3H\left(T\right)\left(y\left(T\right)\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)+y\left(T\right)\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}P_{\chi,(0)}\left(T'\right)\right) \\ & =\left\langle\sigma v\delta E\right\rangle_{\chi f\rightarrow\chi f}y\left(T\right)n_{\chi,(0)}\left(T\right)n_{f}\left(T\right). \end{aligned} \tag{4.87}$$

Now multiply all of Equation 4.87 by $\frac{n_{\chi,(0)}(T')}{n_{\chi,(0)}(T)}\frac{1}{\rho_{\chi,(0)}(T')}$:

$$\begin{aligned}
&xH\left(T\right)\left(1+\frac{1}{3}\frac{T}{g_{\mathrm{eff},s}\left(T\right)}\frac{\partial g_{\mathrm{eff},s}\left(T\right)}{\partial T}\right)^{-1}\left(\frac{\partial y\left(T\right)}{\partial x}+\frac{n_{\chi,(0)}\left(T'\right)}{n_{\chi,(0)}\left(T\right)}\frac{1}{\rho_{\chi,(0)}\left(T'\right)}y\left(T\right)\frac{\partial}{\partial x}\left(\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)\right)\right)\\
&+3H\left(T\right)\left(y\left(T\right)+y\left(T\right)\frac{P_{\chi,(0)}\left(T'\right)}{\rho_{\chi,(0)}\left(T'\right)}\right)\\
&=\left\langle\sigma v\delta E\right\rangle_{\chi f\rightarrow\chi f}y\left(T\right)\frac{n_{\chi,(0)}\left(T'\right)n_{f}\left(T\right)}{\rho_{\chi,(0)}\left(T'\right)}.
\end{aligned}\tag{4.88}$$

Now multiply all of Equation 4.88 by $\frac{1}{xH(T)}\left(1+\frac{1}{3}\frac{T}{g_{\mathrm{eff},s}(T)}\frac{\partial g_{\mathrm{eff},s}(T)}{\partial T}\right)$:

$$\begin{aligned}
&\frac{\partial y\left(T\right)}{\partial x}+\frac{n_{\chi,(0)}\left(T'\right)}{n_{\chi,(0)}\left(T\right)}\frac{1}{\rho_{\chi,(0)}\left(T'\right)}y\left(T\right)\frac{\partial}{\partial x}\left(\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)\right)\\
&+\frac{3}{x}\left(1+\frac{1}{3}\frac{T}{g_{\mathrm{eff},s}\left(T\right)}\frac{\partial g_{\mathrm{eff},s}\left(T\right)}{\partial T}\right)\left(y\left(T\right)+y\left(T\right)\frac{P_{\chi,(0)}\left(T'\right)}{\rho_{\chi,(0)}\left(T'\right)}\right)\\
&=\frac{1}{xH\left(T\right)}\left(1+\frac{1}{3}\frac{T}{g_{\mathrm{eff},s}\left(T\right)}\frac{\partial g_{\mathrm{eff},s}\left(T\right)}{\partial T}\right)\left\langle\sigma v\delta E\right\rangle_{\chi f\rightarrow\chi f}y\left(T\right)\frac{n_{\chi,(0)}\left(T'\right)n_{f}\left(T\right)}{\rho_{\chi,(0)}\left(T'\right)}.
\end{aligned}\tag{4.89}$$

Now multiply all of Equation 4.89 by $1/y\left(T\right)$:

$$\begin{aligned}
&\frac{1}{y\left(T\right)}\frac{\partial y\left(T\right)}{\partial x}+\frac{n_{\chi,(0)}\left(T'\right)}{n_{\chi,(0)}\left(T\right)}\frac{1}{\rho_{\chi,(0)}\left(T'\right)}\frac{\partial}{\partial x}\left(\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)\right)\\
&+\frac{3}{x}\left(1+\frac{1}{3}\frac{T}{g_{\mathrm{eff},s}\left(T\right)}\frac{\partial g_{\mathrm{eff},s}\left(T\right)}{\partial T}\right)\left(1+\frac{P_{\chi,(0)}\left(T'\right)}{\rho_{\chi,(0)}\left(T'\right)}\right)\\
&=\frac{1}{xH\left(T\right)}\left(1+\frac{1}{3}\frac{T}{g_{\mathrm{eff},s}\left(T\right)}\frac{\partial g_{\mathrm{eff},s}\left(T\right)}{\partial T}\right)\left\langle\sigma v\delta E\right\rangle_{\chi f\rightarrow\chi f}\frac{n_{\chi,(0)}\left(T'\right)n_{f}\left(T\right)}{\rho_{\chi,(0)}\left(T'\right)}.
\end{aligned}\tag{4.90}$$

Now make the replacement $H\left(T\right)=H\left(T=m_{\chi}\right)\sqrt{\frac{g_{\mathrm{eff}}(T)}{g_{\mathrm{eff}}(T=m_{\chi})}}\frac{1}{x^{2}}$, and also make the replacement $\frac{T}{g_{\mathrm{eff}}(T)}\frac{\partial g_{\mathrm{eff}}(T)}{\partial T}=-\frac{1}{g_{\mathrm{eff}}(T)}x\frac{\partial g_{\mathrm{eff}}(T)}{\partial x}$:

$$\begin{aligned}
&\frac{1}{y\left(T\right)}\frac{\partial y\left(T\right)}{\partial x}+\frac{n_{\chi,(0)}\left(T'\right)}{n_{\chi,(0)}\left(T\right)}\frac{1}{\rho_{\chi,(0)}\left(T'\right)}\frac{\partial}{\partial x}\left(\frac{n_{\chi,(0)}\left(T\right)}{n_{\chi,(0)}\left(T'\right)}\rho_{\chi,(0)}\left(T'\right)\right)\\
&+\frac{3}{x}\left(1-\frac{1}{3}\frac{x}{g_{\mathrm{eff},s}\left(T\right)}\frac{\partial g_{\mathrm{eff},s}\left(T\right)}{\partial x}\right)\left(1+\frac{P_{\chi,(0)}\left(T'\right)}{\rho_{\chi,(0)}\left(T'\right)}\right)\\
&=\frac{1}{H\left(m_{\chi}\right)}\sqrt{\frac{g_{\mathrm{eff}}\left(m_{\chi}\right)}{g_{\mathrm{eff}}\left(T\right)}}x\left(1-\frac{1}{3}\frac{x}{g_{\mathrm{eff},s}\left(T\right)}\frac{\partial g_{\mathrm{eff},s}\left(T\right)}{\partial x}\right)\left\langle\sigma v\delta E\right\rangle_{\chi f\rightarrow\chi f}\frac{n_{\chi,(0)}\left(T'\right)n_{f}\left(T\right)}{\rho_{\chi,(0)}\left(T'\right)}.
\end{aligned}\tag{4.91}$$

4.4.2 $3 \to 2$ thermally averaged cross section in the threshold approximation

When describing the dynamics of DM particles it is often appropriate to make the the threshold (low momentum and low temperature relative to DM mass) approximation. Therefore it's useful to have general expressions for the thermally averaged cross section in this approximation.

Let's calculate the thermally averaged cross section for a general $3 \to 2$ process, $\langle \sigma v \rangle_{123 \to 45}$, in the threshold limit, starting with the definition Equation 4.20:

$$\begin{aligned} n_1 n_2 n_3 \langle \sigma v \rangle_{123 \to 45} = \frac{1}{S_{id}} \sum_{\text{spins}} \int \prod_{i=1}^{5} \frac{d^3 \mathbf{p}_i}{(2\pi)^3 2E_i} (2\pi)^4 \delta^4 (p_1 + p_2 + p_3 - p_4 - p_5) \\ \times f_1 f_2 f_3 |\mathcal{M}|^2 . \end{aligned} \tag{4.92}$$

Again, if we assume Maxwell-Boltzmann statistics for all particle species we can write the phase-space densities as:

$$f_i = \frac{n_i}{n_i^{(0)}} e^{-E_i/T}, \tag{4.93}$$

and we can rewrite Equation 4.92 as:

$$\begin{aligned} n_1 n_2 n_3 \langle \sigma v \rangle_{123 \to 45} = \frac{1}{S_{id}} \sum_{\text{spins}} \int \prod_{i=1}^{5} \frac{d^3 \mathbf{p}_i}{(2\pi)^3 2E_i} (2\pi)^4 \delta^4 (p_1 + p_2 + p_3 - p_4 - p_5) \\ \times \frac{n_1 n_2 n_3}{n_1^{(0)} n_2^{(0)} n_3^{(0)}} e^{-(E_1+E_2+E_3)/T} |\mathcal{M}|^2 . \end{aligned} \tag{4.94}$$

If we rewrite Equation 4.94 in terms of the initial- and final-state averaged matrix element, it becomes:

$$\begin{aligned} n_1 n_2 n_3 \langle \sigma v \rangle_{123 \to 45} = \frac{1}{S_{id}} \int \prod_{i=1}^{5} \frac{g_i d^3 \mathbf{p}_i}{(2\pi)^3 2E_i} (2\pi)^4 \delta^4 (p_1 + p_2 + p_3 - p_4 - p_5) \\ \times \frac{n_1 n_2 n_3}{n_1^{(0)} n_2^{(0)} n_3^{(0)}} e^{-(E_1+E_2+E_3)/T} \langle |\mathcal{M}|^2 \rangle. \end{aligned} \tag{4.95}$$

Now taking the low-momentum limit we can simplify terms of the form:

$$\begin{aligned}\int \frac{g_i d^3\mathbf{p}_i}{(2\pi)^3 3E_i} e^{-E_i/T} &\simeq \int \frac{g_i d^3\mathbf{p}_i}{(2\pi)^3} \frac{1}{2m_i}\left(1-\frac{|\mathbf{p}_i|^2}{2m_i^2}\right) e^{-\frac{|\mathbf{p}_i|^2}{2m_i T}} e^{-\frac{m_i}{T}} \\ &\simeq \frac{g_i}{(2\pi)^3}\frac{\pi^{3/2}T^2\left(2m_i-3T\right)}{\sqrt{2m_i T}} e^{-m_i/T},\end{aligned} \tag{4.96}$$

and terms of the form:

$$\begin{aligned}n_i^{(0)}(T) = \int \frac{g_i d^3\mathbf{p}_i}{(2\pi)^3} e^{-E_i/T} &\simeq \int \frac{g_i d^3\mathbf{p}_i}{(2\pi)^3} e^{-\frac{|\mathbf{p}_i|^2}{2m_i T}} e^{-\frac{m_i}{T}} \\ &\simeq \frac{2\sqrt{2}g_i}{(2\pi)^3}\pi^{3/2}\left(m_i T\right)^{3/2} e^{-\frac{m_i}{T}}.\end{aligned} \tag{4.97}$$

Now taking the low-temperature as well as low-momentum limit, we find:

$$\frac{1}{n_i^{(0)}}\int \frac{g_1 d^3\mathbf{p}_i}{(2\pi)^3 2E_i} e^{-E_i/T} \simeq \frac{1}{2m_i}, \tag{4.98}$$

therefore in this limit we can simplify Equation 4.95 to:

$$\langle\sigma v\rangle_{123\to 45} \simeq \frac{1}{S_{id}}\frac{1}{8m_1m_2m_3}\int\prod_{i=4}^{5}\frac{g_i d^3\mathbf{p}_i}{(2\pi)^3 2E_i}(2\pi)^4\delta^4\left(p_1+p_2+p_3-p_4-p_5\right)\langle|\mathcal{M}|^2\rangle, \tag{4.99}$$

which can now be easily integrated (assuming a constant $\langle|\mathcal{M}|^2\rangle$) to give:

$$\langle\sigma v\rangle_{123\to 45} \simeq \frac{g_4 g_5}{64\pi S_{id} m_1 m_2 m_3}\lambda^{1/2}\left(m_1+m_2+m_3, m_4, m_5\right)\langle|\mathcal{M}|^2\rangle, \tag{4.100}$$

where $\lambda(x,y,z)$ is defined as:

$$\lambda(x,y,z) = \left(1-\frac{(z+y)^2}{x^2}\right)\left(1-\frac{(z-y)^2}{x^2}\right). \tag{4.101}$$

4.4.3 $2 \to 2$ thermally averaged cross section in the threshold approximation

Starting with the definition of the thermally averaged cross section for a $2 \to 2$ process:

$$n_1 n_2 \langle \sigma v \rangle_{12\to 34} = \frac{1}{S_{id}} \sum_{\text{spins}} \int \prod_{i=1}^{4} \frac{d^3\mathbf{p}_i}{(2\pi)^3 2E_i} (2\pi)^4 \delta^4 (p_1 + p_2 - p_3 - p_4) \times f_1 f_2 |\mathcal{M}|^2 , \tag{4.102}$$

and taking the low-momentum low-temperature limit, the calculation goes the same as in Section 4.4.2 above. The result is:

$$\langle \sigma v \rangle_{12\to 34} \simeq \frac{g_3 g_4}{32\pi S_{id} m_1 m_2} \lambda^{1/2} (m_1 + m_2, m_3, m_4) \langle |\mathcal{M}|^2 \rangle. \tag{4.103}$$

4.4.4 Elastic scattering ($\chi f \to \chi f$) thermal averaged energy transfer rate $\langle \sigma v \delta E \rangle$

Let's calculate the thermal averaged energy transfer rate for the elastic scattering process $\chi f \to \chi f$, $\langle \sigma v \delta E \rangle_{\chi f \to \chi f}$.

Without choosing a specific coordinate system yet, the explicit forms of the Mandelstam variables s and t are:

$$\begin{aligned} s &= (p_{f_1} - p_{\chi_1})^2 \\ &= m_f^2 + m_\chi^2 + 2 (E_{f_1} E_{\chi_1} - \mathbf{p}_{f_1} \cdot \mathbf{p}_{\chi_1}) \end{aligned} \tag{4.104}$$

and

$$\begin{aligned} t &= (p_{\chi_2} - p_{\chi_1})^2 \\ &= 2m_\chi^2 - 2 (E_{\chi_2} E_{\chi_1} - \mathbf{p}_{\chi_2} \cdot \mathbf{p}_{\chi_1}) . \end{aligned} \tag{4.105}$$

In what follows we'll keep careful track of what happens with the angular dependence of $|\mathcal{M}|^2$ during each phase space integral, so bear with me through all the details.

We start with the definition of the quantity:

$$\langle \sigma v \delta E \rangle_{\chi f \to \chi f} n_\chi n_f = \frac{1}{S_{id}} \sum_{\text{spin}} \int \prod_{i=1}^{4} \frac{d^3\mathbf{p}_i}{(2\pi)^3 2E_i} (2\pi)^4 \delta^4 \left(\sum p \right) (E_{\chi_2} - E_{\chi_1}) \times f_{\chi_1} f_{f_1} |\mathcal{M}|^2 . \tag{4.106}$$

For simplicity I will ignore factors of $S_{id} = 1$ and $\sum_{\text{spins}}$ for now and rewrite them back in later. I will also suppress the $\chi f \to \chi f$ subscript where it is obvious. First make the substitution: $f_i = n_i e^{-E_i/T} / n_i^{(0)}$, which gives:

$$\langle \sigma v \delta E \rangle = \frac{1}{n_{\chi,(0)}} \frac{1}{n_{f,(0)}} \int \prod_{i=1}^{4} \frac{d^3\mathbf{p}_i}{(2\pi)^3 2E_i} (2\pi)^4 \delta^4 \left(\sum p \right) (E_{\chi_2} - E_{\chi_1}) \times e^{-E_{\chi_1}/T'} e^{-E_{f_1}/T} |\mathcal{M}|^2 . \tag{4.107}$$

Now, we play a trick to make the calculation easier, by noting that by the definition of the delta function:

$$\int d^4 p_0 \delta (p_0 - p_{\chi_1} - p_{f_1}) \delta^4 (p_0 - p_{\chi_2} - p_{f_2}) = \delta^4 (p_{\chi_1} + p_{f_1} - p_{\chi_2} - p_{f_2}) , \tag{4.108}$$

which we can use to rewrite Equation 4.107 as:

$$\langle \sigma v \delta E \rangle = \frac{1}{n_{\chi,(0)}} \frac{1}{n_{f,(0)}} \int \prod_{i=1}^{4} \frac{d^3\mathbf{p}_i}{(2\pi)^3 2E_i} (2\pi)^4 d^4 p_0 \delta (p_0 - p_{\chi_1} - p_{f_1}) \delta^4 (p_0 - p_{\chi_2} - p_{f_2}) (E_{\chi_2} - E_{\chi_1}) e^{-E_{\chi_1}/T'} e^{-E_{f_1}/T} |\mathcal{M}|^2 . \tag{4.109}$$

Now, eliminate the delta function $\delta^3 (\mathbf{p}_0 - \mathbf{p}_{\chi_2} - \mathbf{p}_{f_2})$ by integrating over $d^3\mathbf{p}_{f_2}$:

$$\langle \sigma v \delta E \rangle = \frac{1}{n_{\chi,(0)}} \frac{1}{n_{f,(0)}} \int \prod_{i=1}^{3} \frac{d^3\mathbf{p}_i}{(2\pi)^3 2E_i} \frac{1}{(2\pi)^3} \frac{1}{2E_{f_2}} (2\pi)^4 d^4 p_0 \delta^4 (p_0 - p_{\chi_1} - p_{f_1}) \times \delta (E_0 - E_{\chi_2} - E_{f_2}) (E_{\chi_2} - E_{\chi_1}) \times e^{-E_{\chi_1}/T'} e^{-E_{f_1}/T} |\mathcal{M}|^2 |_{\text{conditions}}, \tag{4.110}$$

where I used the notation $|_{\text{conditions}}$ to indicate the conditions which we continue to

impose as we integrate out each delta function. Right now {conditions} : $\{\mathbf{p}_0 = \mathbf{p}_{\chi_2} + \mathbf{p}_{f_2}\}$.

Next, we eliminate the delta function $\delta^3(\mathbf{p}_0 - \mathbf{p}_{\chi_1} - \mathbf{p}_{f_1})$ by integrating over $d^3\mathbf{p}_{f_1}$:

$$\begin{aligned}\langle \sigma v \delta E \rangle = \frac{1}{n_{\chi,(0)}} \frac{1}{n_{f,(0)}} \int \frac{d^3\mathbf{p}_{\chi_1}}{(2\pi)^3 2E_{\chi_1}} \frac{d^3\mathbf{p}_{\chi_2}}{(2\pi)^3 2E_{\chi_2}} \frac{1}{(2\pi)^3} \frac{1}{2E_{f_1}} \frac{1}{(2\pi)^3} \frac{1}{2E_{f_2}} \\ \times (2\pi)^4 d^4p_0 \delta(E_0 - E_{\chi_1} - E_{f_1}) \delta(E_0 - E_{\chi_2} - E_{f_2}) \\ \times (E_{\chi_2} - E_{\chi_1}) e^{-E_{\chi_1}/T'} e^{-E_{f_1}/T} |\mathcal{M}|^2 |_{\text{conditions}},\end{aligned} \tag{4.111}$$

where now {conditions} : $\{\mathbf{p}_0 = \mathbf{p}_{\chi_2} + \mathbf{p}_{f_2}; \mathbf{p}_0 = \mathbf{p}_{\chi_1} + \mathbf{p}_{f_1}\}$.

Now that I've imposed two conditions on the momenta in Equation 4.111, I'll more specifically define a coordinate system and keep track of the important angles which appear as we integrate the rest of the phase space integral. I'll define θ_1 and θ_2 as the angles of the two incoming particles with respect to $\mathbf{p}_0$:

$$\begin{aligned}\theta_1 &= \cos^{-1}(\hat{\mathbf{p}}_{\chi_1} \cdot \hat{\mathbf{p}}_0) \\ \theta_2 &= \cos^{-1}(\hat{\mathbf{p}}_{f_1} \cdot \hat{\mathbf{p}}_0).\end{aligned} \tag{4.112}$$

These definitions, combined with the conditions of the phase space integrals imply:

$$\hat{\mathbf{p}}_{\chi_1} \cdot \hat{\mathbf{p}}_{f_1} = \cos(\theta_1 + \theta_2) \tag{4.113}$$

which fixes the form of s in terms of the angles θ_1 and θ_2.

I'll also define θ_3 and θ_4 as the angles of the two outgoing particles with respect to $\mathbf{p}_0$:

$$\begin{aligned}\theta_3 &= \cos^{-1}(\hat{\mathbf{p}}_{\chi_2} \cdot \hat{\mathbf{p}}_0) \\ \theta_4 &= \cos^{-1}(\hat{\mathbf{p}}_{f_2} \cdot \hat{\mathbf{p}}_0).\end{aligned} \tag{4.114}$$

These definitions, combined with conditions of the phase space integral imply:

$$\hat{\mathbf{p}}_{\chi_2} \cdot \hat{\mathbf{p}}_{\chi_1} = \cos(\theta_1 - \theta_3). \tag{4.115}$$

With our coordinate system defined, let's focus back on the phase space integral.

I'll rewrite the integration measure in Equation 4.111 as:

$$\frac{|\mathbf{p}_{\chi_1}|^2 \, d\,|\mathbf{p}_{\chi_1}| \, d\Omega_1}{(2\pi)^3 \, 2E_{\chi_1}} \frac{|\mathbf{p}_{\chi_2}|^2 \, d\,|\mathbf{p}_{\chi_2}| \, d\Omega_3}{(2\pi)^3 \, 2E_{\chi_2}} \tag{4.116}$$

and the term $d\Omega_1 d\Omega_3 = d(\cos\theta_1)\, d(\cos\theta_3)\, d\phi_1 d\phi_3$, where I'm still working in a coordinate system in which all angles are defined relative to the momentum vector $\mathbf{p}_0$.

I want to integrate out the angles θ_1 and θ_3 in Equation 4.111 by rewriting the delta functions as explicit functions of those angles. First working with θ_1, I can rewrite the delta function:

$$\delta(E_0 - E_{\chi_1} - E_{f_1}) \equiv \delta(x(\cos\theta_1)), \tag{4.117}$$

and I can rewrite E_{f_1} to bring out the explicit θ_1 dependence:

$$\begin{aligned} x(\cos\theta_1) &= E_0 - E_{\chi_1} - \sqrt{|\mathbf{p}_{f_1}|^2 + m_f^2} \\ &= E_0 - E_{\chi_1} - \sqrt{|(\mathbf{p}_0 - \mathbf{p}_{\chi_1})|^2 + m_f^2} \\ &= E_0 - E_{\chi_1} - \sqrt{|\mathbf{p}_0|^2 - 2\,|\mathbf{p}_0|\,|\mathbf{p}_{\chi_1}|\cos\theta_1 + |\mathbf{p}_{\chi_1}|^2 + m_f^2}. \end{aligned} \tag{4.118}$$

I can rewrite Equation 4.117 using the well-known delta function identity:

$$\delta(x(\cos\theta_1)) = \sum_i \frac{\delta((\cos\theta_1) - (\cos\theta_1)_i)}{|x'((\cos\theta_1)_i)|} \tag{4.119}$$

where $(\cos\theta_1)_i$ are the zeros of the function $x(\cos\theta_1)$ and $\sum_i$ is a sum over the zeros. There is only one zero of $x(\cos\theta_1)$ in our case, it is that which implies $E_0 = E_{\chi_1} + E_{f_1}$. Therefore the zero $(\cos\theta_1)_i$ is given by:

$$E_0 - E_{\chi_1} = \sqrt{|\mathbf{p}_0|^2 - |\mathbf{p}_0|\,|\mathbf{p}_{\chi_1}|\cos\theta_1 + |\mathbf{p}_{\chi_1}|^2 + m_f^2} \tag{4.120}$$

which gives:

$$\cos\theta_1 = \frac{-(E_0 - E_{\chi_1})^2 + |\mathbf{p}_0|^2 + |\mathbf{p}_{\chi_1}|^2 + m_f^2}{|\mathbf{p}_0|\,|\mathbf{p}_{\chi_1}|}. \tag{4.121}$$

And $x'(\cos\theta_1)$ is given by:

$$\frac{\partial x(\cos\theta_1)}{\partial(\cos\theta_1)} = \frac{|\mathbf{p}_0||\mathbf{p}_{\chi_1}|}{E_{f_1}}. \tag{4.122}$$

We can therefore rewrite the phase space integral Equation 4.111 as:

$$\begin{aligned}\langle\sigma v\delta E\rangle = \frac{1}{n_{\chi,(0)}}\frac{1}{n_{f,(0)}}\int &\frac{|\mathbf{p}_{\chi_1}|^2 d|\mathbf{p}_{\chi_1}| d\phi_1 d(\cos\theta_1)}{(2\pi)^3 2E_{\chi_1}}\frac{|\mathbf{p}_{\chi_2}|^2 d|\mathbf{p}_{\chi_2}| d\phi_3 d(\cos\theta_3)}{(2\pi)^3 2E_{\chi_2}}\\ &\times\frac{1}{(2\pi)^3}\frac{1}{2E_{f_1}}\frac{1}{(2\pi)^3}\frac{1}{2E_{f_2}}(2\pi)^4 d^4p_0\\ &\times\frac{\delta((\cos\theta_1)-(\cos\theta_1)_i)}{\frac{|\mathbf{p}_0||\mathbf{p}_{\chi_1}|}{E_{f_1}}}\delta(E_0-E_{\chi_2}-E_{f_2})\\ &\times(E_{\chi_2}-E_{\chi_1})e^{-E_{\chi_1}/T'}e^{-E_{f_1}/T}|\mathcal{M}|^2|_{\text{conditions}},\end{aligned} \tag{4.123}$$

and integrating over $d\cos\theta_1$ we obtain:

$$\begin{aligned}\langle\sigma v\delta E\rangle = \frac{1}{n_{\chi,(0)}}\frac{1}{n_{f,(0)}}\int &\frac{|\mathbf{p}_{\chi_1}|^2 d|\mathbf{p}_{\chi_1}| d\phi_1}{(2\pi)^3 2E_{\chi_1}}\frac{|\mathbf{p}_{\chi_2}|^2 d|\mathbf{p}_{\chi_2}| d\phi_3 d(\cos\theta_3)}{(2\pi)^3 2E_{\chi_2}}\\ &\times\frac{1}{(2\pi)^3}\frac{1}{2E_{f_1}}\frac{1}{(2\pi)^3}\frac{1}{2E_{f_2}}(2\pi)^4 d^4p_0\\ &\times\frac{E_{f_1}}{|\mathbf{p}_0||\mathbf{p}_{\chi_1}|}\delta(E_0-E_{\chi_2}-E_{f_2})\\ &\times(E_{\chi_2}-E_{\chi_1})e^{-E_{\chi_1}/T'}e^{-E_{f_1}/T}|\mathcal{M}|^2|_{\text{conditions}},\end{aligned} \tag{4.124}$$

where now the set of conditions is:

$$\begin{aligned}\{\text{conditions}\} : \{&\mathbf{p}_0 = \mathbf{p}_{\chi_2} + \mathbf{p}_{f_2}\\ &\mathbf{p}_0 = \mathbf{p}_{\chi_1} + \mathbf{p}_{f_1}\\ &\cos\theta_1 = \frac{-(E_0-E_{\chi_1})^2 + |\mathbf{p}_0|^2 + |\mathbf{p}_{\chi_1}|^2 + m_f^2}{|\mathbf{p}_0||\mathbf{p}_{\chi_1}|}\}\end{aligned} \tag{4.125}$$

Now I'll go through the same process integrating over $d(\cos\theta_3)$. First rewrite the delta function $\delta(E_0 - E_{\chi_2} - E_{f_2})$ as:

$$\delta(y(\cos\theta_3)) = \frac{\delta((\cos\theta_3)-(\cos\theta_3)_i)}{|y'((\cos\theta_3)_i)|} \tag{4.126}$$

where $(\cos\theta_3)_i$ is the zero of the function $y(\cos\theta_3)$:

$$\begin{aligned} y(\cos\theta_3) &= E_0 - \sqrt{|\mathbf{p}_{\chi_2}|^2 + m_\chi^2} - \sqrt{|\mathbf{p}_{f_2}|^2 + m_f^2} \\ &= E_0 - \sqrt{|\mathbf{p}_{\chi_2}|^2 + m_\chi^2} - \sqrt{|(\mathbf{p}_0 - \mathbf{p}_{\chi_2})|^2 + m_f^2} \\ &= E_0 - \sqrt{|\mathbf{p}_{\chi_2}|^2 + m_\chi^2} - \sqrt{|\mathbf{p}_0|^2 + |\mathbf{p}_{\chi_2}|^2 - 2|\mathbf{p}_0||\mathbf{p}_{\chi_2}|\cos\theta_3 + m_f^2} \end{aligned} \tag{4.127}$$

and $y'(\cos\theta_3)$ is given by:

$$\frac{\partial y(\cos\theta_3)}{\partial(\cos\theta_3)} = \frac{|\mathbf{p}_0||\mathbf{p}_{\chi_2}|}{E_{f_2}} \tag{4.128}$$

so we can write Equation 4.124 as:

$$\begin{aligned} \langle\sigma v \delta E\rangle = \frac{1}{n_{\chi,(0)}}\frac{1}{n_{f,(0)}} \int & \frac{|\mathbf{p}_{\chi_1}|^2 d|\mathbf{p}_{\chi_1}| d\phi_1}{(2\pi)^3 2E_{\chi_1}} \frac{|\mathbf{p}_{\chi_2}|^2 d|\mathbf{p}_{\chi_2}| d\phi_3 d(\cos\theta_3)}{(2\pi)^3 2E_{\chi_2}} \\ & \times \frac{1}{(2\pi)^3}\frac{1}{2E_{f_1}}\frac{1}{(2\pi)^3}\frac{1}{2E_{f_2}}(2\pi)^4 d^4p_0 \\ & \times \frac{E_{f_1}}{|\mathbf{p}_0||\mathbf{p}_{\chi_1}|} \frac{\delta((\cos\theta_3) - (\cos\theta_3)_i)}{\frac{|\mathbf{p}_0||\mathbf{p}_{\chi_2}|}{E_{f_2}}} \\ & \times (E_{\chi_2} - E_{\chi_1}) e^{-E_{\chi_1}/T'} e^{-E_{f_1}/T} |\mathcal{M}|^2 |_{\text{conditions}}, \end{aligned} \tag{4.129}$$

and integrating over $d(\cos\theta_3)$ we obtain:

$$\begin{aligned} \langle\sigma v \delta E\rangle = \frac{1}{n_{\chi,(0)}}\frac{1}{n_{f,(0)}} \int & \frac{|\mathbf{p}_{\chi_1}|^2 d|\mathbf{p}_{\chi_1}| d\phi_1}{(2\pi)^3 2E_{\chi_1}} \frac{|\mathbf{p}_{\chi_2}|^2 d|\mathbf{p}_{\chi_2}| d\phi_3}{(2\pi)^3 2E_{\chi_2}} \\ & \times \frac{1}{(2\pi)^3}\frac{1}{2E_{f_1}}\frac{1}{(2\pi)^3}\frac{1}{2E_{f_2}}(2\pi)^4 d^4p_0 \\ & \times \frac{E_{f_1}}{|\mathbf{p}_0||\mathbf{p}_{\chi_1}|} \frac{E_{f_2}}{|\mathbf{p}_0||\mathbf{p}_{\chi_2}|} \\ & \times (E_{\chi_2} - E_{\chi_1}) e^{-E_{\chi_1}/T'} e^{-E_{f_1}/T} |\mathcal{M}|^2 |_{\text{conditions}}, \end{aligned} \tag{4.130}$$

where now conditions also imposes the energy constraint $E_0 = E_{\chi_2} + E_{f_2}$ through the

condition on $(\cos\theta_3)$:

$$\begin{aligned}\{\text{conditions}\} : \{&\mathbf{p}_0 = \mathbf{p}_{\chi_2} + \mathbf{p}_{f_2} \\ &\mathbf{p}_0 = \mathbf{p}_{\chi_1} + \mathbf{p}_{f_1} \\ &\cos\theta_1 = \frac{-\left(E_0 - E_{\chi_1}\right)^2 + \left|\mathbf{p}_0\right|^2 + \left|\mathbf{p}_{\chi_1}\right|^2 + m_f^2}{2\left|\mathbf{p}_0\right|\left|\mathbf{p}_{\chi_1}\right|} \\ &\cos\theta_3 = \frac{-\left(E_0 - E_{\chi_2}\right)^2 + \left|\mathbf{p}_0\right|^2 + \left|\mathbf{p}_{\chi_2}\right|^2 + m_f^2}{2\left|\mathbf{p}_0\right|\left|\mathbf{p}_{\chi_2}\right|}\}.\end{aligned} \tag{4.131}$$

Now, just to keep track of where we are, we have two angles left to fix in order for all variables to be dynamically fixed: θ_2 and θ_4. We have left to integrate $d\left|\mathbf{p}_{\chi_1}\right|$, $d\left|\mathbf{p}_{\chi_2}\right|$, $d\phi_1$, $d\phi_3$, d^4p_0. We should be free to rotate the two azimuthal angles ϕ_1 and ϕ_3 freely without changing any of the other dynamical quantities, so those integrals should just give us a factor of $(2\pi)^2$. Also, we make the substitution $E_{\chi_1}dE_{\chi_1} = \left|\mathbf{p}_{\chi_1}\right| d\left|\mathbf{p}_{\chi_1}\right|$ and $E_{\chi_2}dE_{\chi_2} = \left|\mathbf{p}_{\chi_2}\right| d\left|\mathbf{p}_{\chi_2}\right|$. Then we can write Equation 4.130 as:

$$\begin{aligned}\langle\sigma v\delta E\rangle = \frac{1}{n_{\chi,(0)}}\frac{1}{n_{f,(0)}}\frac{1}{(8\pi)^2}\int\frac{d^4p_0}{(2\pi)^4}dE_{\chi_1}dE_{\chi_2}\frac{(E_{\chi_2}-E_{\chi_1})}{\left|\mathbf{p}_0\right|^2} \\ \times e^{-E_{\chi_1}/T'}e^{-E_{f_1}/T}\left|\mathcal{M}\right|^2|_{\text{conditions}}.\end{aligned} \tag{4.132}$$

Now, s is given by:

$$\begin{aligned}s &= \left(p_{\chi_1} + p_{f_1}\right)^2 \\ &= \left(E_{\chi_1} + E_{f_1}\right)^2 - \left(\mathbf{p}_{\chi_1} + \mathbf{p}_{f_1}\right)\cdot\left(\mathbf{p}_{\chi_1} + \mathbf{p}_{f_1}\right) \\ &= E_0^2 - \left|\mathbf{p}_0\right|^2.\end{aligned} \tag{4.133}$$

We can write d^4p_0 as:

$$\begin{aligned}d^4p_0 &= dE_0 d^3\mathbf{p}_0 \\ &= dE_0\left|\mathbf{p}_0\right|^2 d\left|\mathbf{p}_0\right| d\Omega_0.\end{aligned} \tag{4.134}$$

And from Equation 4.133 we can rewrite $d\left|\mathbf{p}_0\right|$ in terms of ds:

$$2\left|\mathbf{p}_0\right| d\left|\mathbf{p}_0\right| = -ds \tag{4.135}$$

where E_0 is held fixed in relating $d\left|\mathbf{p}_0\right|$ to ds.

Making the above substitutions for d^4p_0, as well as making the substitution $\left|\mathbf{p}_0\right| = \sqrt{E_0^2 - s}$, we can write Equation 4.132 as:

$$\begin{aligned}\langle \sigma v \delta E \rangle = \frac{1}{n_{\chi,(0)}} \frac{1}{n_{f,(0)}} \frac{1}{512\pi^5} \int ds dE_0 dE_{\chi_1} dE_{\chi_2} \frac{(E_{\chi_2} - E_{\chi_1})}{\sqrt{E_0^2 - s}} \\ \times e^{-E_{\chi_1}/T'} e^{-E_{f_1}/T} \left|\mathcal{M}\right|^2 \big|_{\text{conditions}}.\end{aligned} \tag{4.136}$$

And then rewriting $E_{f_1} = E_0 - E_{\chi_1}$ we can rewrite Equation 4.136 as:

$$\begin{aligned}\langle \sigma v \delta E \rangle = \frac{1}{n_{\chi,(0)}} \frac{1}{n_{f,(0)}} \frac{1}{512\pi^5} \int ds dE_0 dE_{\chi_1} dE_{\chi_2} \frac{(E_{\chi_2} - E_{\chi_1})}{\sqrt{E_0^2 - s}} \\ \times e^{-E_{\chi_1}(1/T' - 1/T)} e^{-E_0/T} \left|\mathcal{M}\right|^2 \big|_{\text{conditions}},\end{aligned} \tag{4.137}$$

with the conditions in Equation 4.137 given by Equation 4.131.

Now, let's keep track of all the kinematics again. In general the matrix element $\left|\mathcal{M}\right|^2$ in Equation 4.137 will be a function of s and t: $\left|\mathcal{M}\right|^2 = \left|\mathcal{M}\right|^2(s,t)$. The expression for s is given by Equation 4.104:

$$\begin{aligned} s &= (p_{f_1} - p_{\chi_1})^2 \\ &= m_f^2 + m_\chi^2 + 2\left(E_{f_1}E_{\chi_1} - \mathbf{p}_{f_1} \cdot \mathbf{p}_{\chi_1}\right) \\ &= m_f^2 + m_\chi^2 + 2\left(\left(E_0 - E_{\chi_1}\right)E_{\chi_1} - \left(\mathbf{p}_0 - \mathbf{p}_{\chi_1}\right) \cdot \mathbf{p}_{\chi_1}\right) \\ &= m_f^2 + m_\chi^2 + 2\left(E_0 E_{\chi_1} - m_\chi^2 - \mathbf{p}_0 \cdot \mathbf{p}_{\chi_1}\right) \\ &= m_f^2 - m_\chi^2 + 2E_0E_{\chi_1} - 2\left|\mathbf{p}_0\right|\left|\mathbf{p}_{\chi_1}\right|\cos\theta_1, \end{aligned} \tag{4.138}$$

so we see the variable s, along with E_0 and E_{χ_1}, fixes the angle θ_1.

We can use Equation 4.138, along with Equation 4.133, to write the conditions

relating the angles θ_1 and θ_3 to dynamical variables appearing in Equation 4.137 as:

$$\begin{aligned}\{\text{conditions}\} : \{&\mathbf{p}_0 = \mathbf{p}_{\chi_2} + \mathbf{p}_{f_2} \\ &\mathbf{p}_0 = \mathbf{p}_{\chi_1} + \mathbf{p}_{f_1} \\ &\cos\theta_1 = \frac{2E_0E_{\chi_1} - s + m_f^2 - m_\chi^2}{2\sqrt{E_0^2 - s}\sqrt{E_{\chi_1}^2 - m_\chi^2}} \\ &\cos\theta_3 = \frac{2E_0E_{\chi_2} - s + m_f^2 - m_\chi^2}{2\sqrt{E_0^2 - s}\sqrt{E_{\chi_2}^2 - m_\chi^2}}\}.\end{aligned} \tag{4.139}$$

And we can write t from Equation 4.105 as:

$$\begin{aligned}t &= (p_{\chi_2} - p_{\chi_1})^2 \\ &= 2m_\chi^2 - 2\left(E_{\chi_2}E_{\chi_1} - \mathbf{p}_{\chi_2}\cdot\mathbf{p}_{\chi_1}\right) \\ &= 2m_\chi^2 - 2\left(E_{\chi_2}E_{\chi_1} - |\mathbf{p}_{\chi_2}|\,|\mathbf{p}_{\chi_1}|\cos\left(\theta_1 - \theta_3\right)\right) \\ &= 2m_\chi^2 - 2\left(E_{\chi_2}E_{\chi_1} - \sqrt{E_{\chi_2}^2 - m_\chi^2}\sqrt{E_{\chi_1}^2 - m_\chi^2}\cos\left(\theta_1 - \theta_3\right)\right).\end{aligned} \tag{4.140}$$

Now, with t given by Equation 4.140, and θ_1 and θ_3 given by Equation 4.139, we have everything we need to evaluate Equation 4.137 to obtain $\langle\sigma v\delta E\rangle$ in terms of T and T' for a given $|\mathcal{M}|^2(s,t)$.

We perform the remaining integral in Equation 4.137 numerically using the following limits of integration:

$$\begin{aligned}&E_{\chi_2}^{\min,\max} = E_{\chi_1}^{\min,\max} = \frac{E_0\left(m_\chi^2 + s\right)}{2s} \pm \frac{|\mathbf{p}_0|}{2}\lambda^{1/2}\left(\sqrt{s}, m_\chi, 0\right) \\ &\sqrt{s} \le E_0 < \infty \\ &m_\chi^2 \le s < \infty\end{aligned} \tag{4.141}$$

where λ is defined as:

$$\lambda(x,y,z) = \left(1 - \frac{(z+y)^2}{x^2}\right)\left(1 - \frac{(z-y)^2}{x^2}\right). \tag{4.142}$$

Chapter 5

New Pathways to the Relic Abundance of Vector-Portal Dark Matter

This chapter is from [2], *by me, Hongwan Liu, Tracy R. Slatyer, and Yu-Dai Tsai.*

5.1 Introduction

In Chapter 4 we saw that there has been a lot of recent effort to move beyond the thermal WIMP scenario and understand alternative DM production mechanisms which naturally produce light (sub-GeV) DM. In particular, we took a brief look at the strongly-interacting-massive-particle (SIMP) scenario, in which strong $3 \to 2$ self-annihilations among DM particles control the thermal relic abundance, leading to strongly-coupled DM ($\alpha_D \sim 1$) with mass similar to the QCD scale ($m_\chi \sim 10$–$100\,\mathrm{MeV}$) [92]. As we mentioned, in the SIMP scenario the DM and SM sectors remain in thermal equilibrium throughout freezeout via elastic scattering between DM and SM particles, and we also saw that an alternative thermal production mechanism for light DM arises when this condition is relaxed; in the elastically decoupling relic (ELDER) scenario the DM and SM sectors thermally decouple through the elastic DM-SM scattering while strong $3 \to 2$ self-annihilations are still active [93, 94]. In

the ELDER scenario, although thermal freezeout proceeds through the $3 \to 2$ DM self-annihilations, the DM relic abundance is nevertheless determined by the decoupling of DM-SM elastic scattering. This is achieved through a dark sector process called "cannibalization" [111], which occurs immediately after elastic decoupling and proceeds until $3 \to 2$ freezeout. During cannibalization, while the DM and SM sectors are thermally secluded, $3 \to 2$ DM self-annihilations convert mass to kinetic energy and heat the dark sector. As a result, the dark sector temperature evolves slowly (logarithmically as a function of SM temperature) during cannibalization, and likewise, the DM abundance evolves slowly. This leads to a DM relic abundance that is primarily determined by its value at kinetic decoupling. The ELDER scenario also naturally leads to MeV-GeV mass DM.

Distinctive thermal production mechanisms for light DM have also been realized in the well-studied vector-portal DM model of a Dirac fermion DM particle χ charged under a hidden U(1) gauge symmetry with dark gauge boson A', which is coupled to the SM photon through kinetic mixing. In the region of parameter space in which the dark photon is more massive than the DM ($r \equiv m_{A'}/m_\chi > 1$), the kinematically suppressed $2 \to 2$ annihilations of DM to heavier A's ($\chi\bar{\chi} \to A'A'$) can control the relic abundance. In this "forbidden DM" (FDM) mechanism [112, 113] the exponential suppression of the $2 \to 2$ process setting the relic abundance of DM naturally gives rise to DM exponentially lighter than the weak scale. The FDM mechanism was shown to be a viable mechanism for producing sub-GeV DM.

More recently, Ref. [114] showed that in the region of parameter space of the dark photon model in which $1.5 \lesssim r \lesssim 2$, the kinematic suppression of the $\chi\bar{\chi} \to A'A'$ annihilation process is compensated for by a kinematically allowed $3 \to 2$ ($\chi\chi\bar{\chi} \to \chi A'$) annihilation channel, which can then play a dominant role in setting the thermal relic abundance of DM. This "not-forbidden dark matter" (NFDM) scenario is analogous to the thermal SIMP scenario in that $3 \to 2$ processes can determine the DM relic abundance, realized in the simple and well-studied vector-portal DM model. The NFDM scenario was also demonstrated to be a viable mechanism for naturally producing sub-GeV DM. In both the FDM and NFDM scenarios, the DM and SM sectors

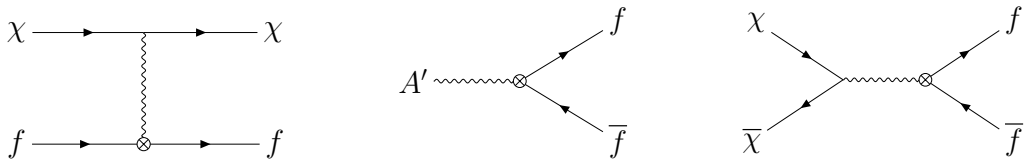

Figure 5-1: Tree-level interactions between dark sector and SM particles, including *(left)* $\chi f \to \chi f$ elastic scattering, *(center)* $A' \to f\overline{f}$ decay, and *(right)* $\chi\overline{\chi} \to f\overline{f}$ annihilation into SM fermions.

were assumed to remain thermally coupled throughout the freezeout of DM.

In this chapter, we extend both of these frameworks to consider cases in which the DM and SM sectors are allowed to kinetically decouple during thermal freezeout of the DM. We fully explore the $1 \lesssim r \lesssim 2$ region of parameter space of the dark photon model, in which the kinematically suppressed $2 \to 2$ ($\chi\bar{\chi} \to A'A'$) channel and the kinematically allowed $3 \to 2$ ($\chi\chi\bar{\chi} \to \chi A'$) channel play important roles in controlling thermal freezeout, and relax the condition that kinetic equilibrium is maintained between the two sectors throughout the freezeout process. We find a rich set of novel cosmological histories leading to a range of different mechanisms for obtaining the correct DM relic density. Among these, we identify a general class of mechanisms in which the DM relic abundance is determined by processes controlling the kinetic decoupling of the DM and SM sectors (which we call the KINetically DEcoupling Relic – KINDER). This KINDER scenario in the dark photon model generalizes the ELDER scenario to cases in which multiple processes control the thermal coupling between dark and SM sectors, and in which a $3 \to 2$ annihilation process among multiple dark sector species supports heating of the dark sector.

The outline of this chapter is as follows. In Section 5.2, we describe the dark photon model in the $1 \lesssim r \lesssim 2$ region we consider, including the primary interactions controlling chemical equilibrium in the dark sector, and those between the dark sector and SM particles. In Section 5.3 we discuss general features of dark sector freezeout in our model, including the relevant interaction processes, the Boltzmann equations, which describe the thermodynamic evolution of the system, and the freezeout conditions of relevant processes. In this section, we also classify three thermodynamic phases (A, B, and C), which generally describe the various stages of the thermal

histories realized in our model.

In Sections 5.4 and 5.5, we characterize the thermal freezeout histories for $1.5 \lesssim r \lesssim 2$ and $1 \lesssim r \lesssim 1.5$ respectively. In each case, we identify a rich set of freezeout histories and analytically determine the parameter space regions where they occur. These different histories are naturally classified into specific regions in the ϵ–α_D plane, where ϵ describes the mixing between the dark photon and the SM photon, and α_D is the dark sector coupling. In Section 5.4 we study the $1.5 \lesssim r \lesssim 2$ region of our model, where the $2 \leftrightarrow 2$ process freezes out before the $3 \leftrightarrow 2$ process; the possible histories can be classified into the WIMP, NFDM and KINDER regimes. In Section 5.5, we examine the $1 \lesssim r \lesssim 1.5$ region of our model, where the $3 \to 2$ process freezes out prior to the $2 \to 2$ process, and find four distinct regimes in addition to the WIMP regime (Regimes I – IV). In Section 5.6 we discuss the relevant experimental and cosmological constraints; finally, in Section 5.7 we summarize our conclusions.

Throughout this chapter, we make use of Planck 2018 cosmological parameters [115], using the TT,TE,EE+lowE+lensing results; we take the DM abundance to be the central value of $\Omega_\chi h^2 = 0.12$, with $h = 0.6736$. All quantities are expressed in natural units, with $\hbar = c = k_B = 1$. Finally, we use many different symbols for approximations in this chapter, and have attempted to keep them consistent with the following definitions: *(i)* we use "$\simeq$" when the approximation is a physical limit, e.g., a nonrelativistic limit; *(ii)* we use "$\approx$" for statements that are true within an order of magnitude, but which we will take to be an equality for the purpose of analytic results; *(iii)* finally, we use "$\sim$" for statements that are true within an order of magnitude, but we do not use the fact either analytically or numerically.

5.2 Model

In the mass basis, the Lagrangian of the dark photon model we consider is:

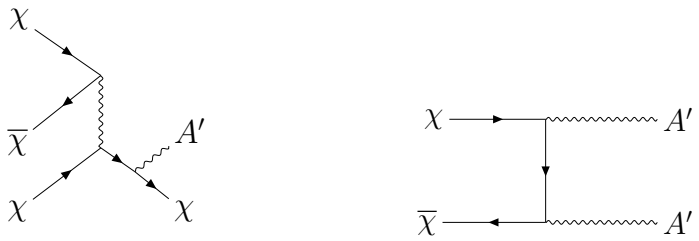

Figure 5-2: The dark-sector-only *(left)* $\chi\overline{\chi}\chi \leftrightarrow A'\chi$ $(3 \leftrightarrow 2)$ process and *(right)* the $\chi\overline{\chi} \leftrightarrow A'A'$ $(2 \leftrightarrow 2)$ process.

$$\mathcal{L} \supset -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} + \frac{1}{2}m_{A'}^2 A'_\mu A'^\mu \\ \overline{\chi}(i\not{D} - m_\chi)\chi + \epsilon A'_\mu J^\mu_{\rm EM} \tag{5.1}$$

where the gauge coupling is $\alpha_D = g_D^2/4\pi$, and $\not{D} \equiv \not{\partial} - ig_D\not{A}'$. The dark photon A' kinetically mixes with the SM photon, giving rise to a small coupling between the dark photon and the SM electromagnetic current $J^\mu_{\rm EM}$, set by the kinetic mixing parameter ϵ. The value of ϵ can naturally range from as small as 10^{-13} up to 10^{-1} [116]. The hidden U(1) symmetry can be spontaneously broken through a Higgs-like mechanism with the dark Higgs taken to be heavy enough to be excluded from this low-energy effective description, since we will always be considering energies $\lesssim m_\chi, m_{A'}$. The kinetic mixing generates the tree-level interactions between dark and SM particles shown in Fig. 5-1: $\chi f \to \chi f$, $\chi\bar{\chi} \to f\bar{f}$, $A' \to f\bar{f}$.

We are primarily interested in scenarios in which the dominant DM-number-changing interactions are the $\chi\chi\bar{\chi} \leftrightarrow \chi A'$ $(3 \leftrightarrow 2)$ process and the kinematically suppressed $\chi\bar{\chi} \leftrightarrow A'A'$ $(2 \leftrightarrow 2)$ process, shown in Fig. 5-2. This restricts us to the region of parameter space in which $1 \lesssim r \lesssim 2$.

At lower values of r, where $r < 1$, the dominant process controlling thermal freeze-out is $\chi\bar{\chi} \to A'A'$ (which is then kinematically allowed), and the A' decays promptly to SM particles. This regime is strongly ruled out by cosmic microwave background (CMB) constraints on the annihilation cross section of DM into SM particles for $m_\chi \lesssim 10\,\text{GeV}$ [117].

At higher values of r, where $r > 2$, the s-channel annihilation of $\chi\bar{\chi} \to f\bar{f}$ via an off-shell A' dominates the DM-number-changing interactions: the $\chi\overline{\chi} \to A'A'$ process is very kinematically suppressed and the $\chi\chi\bar{\chi} \to A'\chi$ process reduces to a scattering process among the χs as the final-state A' promptly decays back to $\chi\bar{\chi}$. Dark sector freezeout proceeds via the classic WIMP freezeout scenario, which also runs into stringent CMB constraints on the s-wave annihilation of Dirac fermion dark matter below 10 GeV [115].

In the intermediate ($1 \lesssim r \lesssim 2$) region of interest to us, which of the $2 \to 2$ or $3 \to 2$ processes dominates during thermal freeze-out depends on the ratio r. The $2 \to 2$ process receives a kinematic suppression from χ particles annihilating into heavier A' particles (with an exponential factor of the form $e^{-2(r-1)m_\chi/T}$), while the $3 \to 2$ receives a Boltzmann suppression from an extra factor of χ number density in the initial state (with an exponential factor of the form $e^{-m_\chi/T}$). In the lower half of the range in r we consider ($1 < r \lesssim 1.5$), the Boltzmann suppression is more severe for the $3 \to 2$ process, and therefore the $2 \to 2$ process dominates during thermal freezeout. In this regime, and for the case in which the DM and SM sectors remain thermally coupled throughout thermal freezeout, the $2 \to 2$ process determines the relic abundance; this is the FDM scenario described in the introduction.

In the upper half of the range in r we consider ($1.5 \lesssim r < 2$), in contrast, the large kinematic suppression of the $2 \to 2$ process renders it subdominant to the $3 \to 2$ process at freezeout. In this regime, and for the case in which the DM and SM sectors remain thermally coupled throughout thermal freezeout, the $3 \to 2$ process determines the thermal relic abundance; this is the NFDM scenario described in the introduction.

5.3 Dark Sector Freezeout

Before we detail all of the different regimes in which the dark sector can evolve to obtain the final DM relic density, we will begin by discussing some general features of the dark sector freezeout in this model. By "dark sector freezeout", we mean the

cosmological evolution from the initial state, when both sectors are in thermal equilibrium, to the point where the DM has attained its final comoving relic abundance.

5.3.1 Thermodynamic Variables

Throughout freezeout, for the parameter space we consider, χ and A' remain in thermal equilibrium with each other through χ–A' scattering. The dark sector can therefore be described by a single dark sector temperature T'. The general expressions for the number densities of the particles in the nonrelativistic limit are:

$$\begin{aligned} n_\chi(T') &\equiv \frac{2g_\chi m_\chi^2 T'}{2\pi^2} K_2\left(\frac{m_\chi}{T'}\right) \\ &\simeq 2g_\chi \left(\frac{m_\chi T'}{2\pi}\right)^{3/2} e^{-m_\chi/T'} e^{\mu_\chi(T')/T'} , \end{aligned} \tag{5.2}$$

$$\begin{aligned} n_{A'}(T') &\equiv \frac{g_{A'} m_{A'}^2 T'}{2\pi^2} K_2\left(\frac{m_{A'}}{T'}\right) \\ &\simeq g_{A'} \left(\frac{r m_\chi T'}{2\pi}\right)^{3/2} e^{-r m_\chi/T'} e^{\mu_{A'}(T')/T'} , \end{aligned} \tag{5.3}$$

where $g_\chi = 2$ and $g_{A'} = 3$ are the numbers of degrees of freedom associated with each particle. The factor of two in Eq. (5.2) accounts for the fact that we are including both χ and $\overline{\chi}$ in the definition of n_χ. We have also included effective chemical potentials μ_χ and $\mu_{A'}$ which are in general nonzero; we denote the number densities of χ and A' with zero chemical potential as $n_{\chi,0}(T')$ and $n_{A',0}(T')$ respectively. We will also frequently use the inverse dimensionless temperatures, $x \equiv m_\chi/T$ and $x' \equiv m_\chi/T'$.

The energy densities and pressures of χ and A' are related to their number densities by:

$$\rho_\chi\left(T'\right) = \frac{n_\chi\left(T'\right)}{n_{\chi,0}\left(T'\right)} \rho_{\chi,0}\left(T'\right) \tag{5.4}$$

and

$$P_\chi\left(T'\right) = \frac{n_\chi\left(T'\right)}{n_{\chi,0}\left(T'\right)} P_{\chi,0}\left(T'\right) , \tag{5.5}$$

where the Maxwell-Boltzmann distributions with zero chemical potential for $\rho_{\chi,0}(T')$ and $P_{\chi,0}(T')$ are given by

$$\rho_{\chi,0}(T') \equiv \frac{2g_\chi m_\chi^2 T'}{2\pi^2}\left(m_\chi K_1\left(\frac{m_\chi}{T'}\right) + 3T' K_2\left(\frac{m_\chi}{T'}\right)\right),$$
$$P_{\chi,0}(T') \equiv \frac{2g_\chi m_\chi^2 T'^2}{2\pi^2} K_2\left(\frac{m_\chi}{T'}\right) = n_{\chi,0}(T')T'. \tag{5.6}$$

Similar relations hold for A'. The entropy of the dark sector is conserved when no heat is transferred between the dark sector and the SM through processes that involve both dark sector and SM particles. The entropy density of the dark sector is

$$s_D \equiv \frac{(\rho_\chi + \rho_{A'}) + (P_\chi + P_{A'}) - \mu_\chi n_\chi - \mu_{A'} n_{A'}}{T'}. \tag{5.7}$$

Entropy conservation of the dark sector in the limit where heat transfer processes are inefficient is a useful fact that we will use extensively in obtaining an analytic understanding of our results. When the dark sector entropy is conserved, $d(s_D a^3)/dt = 0$, where a is the expansion scale factor.

Since we will be discussing the time evolution of n_χ and $n_{A'}$ frequently in the context of analytic estimates, we will derive here several expressions related to $\dot{n}_\chi$ and $\dot{n}_{A'}$ that will be useful throughout the chapter. First, taking the time derivative of n_χ gives

$$\dot{n}_\chi \simeq -\left[\frac{3}{2T'} + \frac{m_\chi}{T'^2} + \frac{d}{dT'}\left(\frac{\mu_\chi}{T'}\right)\right]\frac{dT'}{dT} HT n_\chi, \tag{5.8}$$

where we have used $dT/dt \simeq -HT$. We will often make the approximation that $m_\chi/T' \gg 1$ during freezeout, and so the term $3/(2T')$ can often be neglected, unless $\mu_\chi \sim m_\chi$. Similarly,

$$\dot{n}_{A'} \simeq -\left[\frac{3}{2T'} + \frac{r m_\chi}{T'^2} + \frac{d}{dT'}\left(\frac{\mu_{A'}}{T'}\right)\right]\frac{dT'}{dT} HT n_{A'}. \tag{5.9}$$

We will often be interested in comparing the final number density of the dark matter after the dark sector completely freezes out to the number density required to

achieve the relic abundance of dark matter today. Defining $Y_\chi \equiv n_\chi / s_{\rm SM}$, where $s_{\rm SM}$ is the entropy density of the SM sector after the dark sector has completely decoupled, the correct relic abundance is obtained when [118]

$$Y_\chi \equiv \frac{n_\chi}{s_{\rm SM}} = 4.32 \times 10^{-10} \left(\frac{\rm GeV}{m_\chi} \right) . \tag{5.10}$$

5.3.2 Relevant Processes

In the conventional WIMP regime, DM freezes out through the process $\chi\overline{\chi} \to f\overline{f}$, where f is a SM fermion. Once the mixing parameter $\epsilon \lesssim 10^{-5}$–$10^{-4}$, however, $\chi\overline{\chi} \to f\overline{f}$ freezes out while other dark sector processes are still active, and these processes play a significant role in the freezeout of the dark sector [119, 120].

Outside the WIMP regime, there are four main processes that play important roles during the freezeout of the dark sector when $1 \lesssim r \lesssim 2$:

1. *The* $2 \leftrightarrow 2$ *dark sector process,* $\chi\overline{\chi} \leftrightarrow A'A'$. This process was shown in Ref. [119] to be responsible for the freezeout of the dark sector for $1 \lesssim r \lesssim 1.5$, under the assumption that the dark sector was in full thermal equilibrium with the SM. As we described in the introduction, this process is kinematically forbidden for $r > 1$ for stationary χ particles, leading to a velocity-averaged annihilation cross section that is exponentially suppressed as a function of the dark sector temperature T'. Explicitly, the annihilation cross section is given by [119]

$$\begin{aligned} \langle \sigma v \rangle_{\chi\overline{\chi} \to A'A'} &= \frac{n_{A',0}^2}{n_{\chi,0}^2} \langle \sigma v \rangle_{A'A' \to \chi\overline{\chi}} \\ &= \frac{9}{16} r^3 e^{2(1-r) m_\chi / T'} \langle \sigma v \rangle_{A'A' \to \chi\overline{\chi}} . \end{aligned} \tag{5.11}$$

   We provide the expression for $\langle \sigma v \rangle_{A'A' \to \chi\overline{\chi}}$ in App. B.1; to make our analytic estimates more convenient, however, we parametrize this annihilation cross section as follows:

$$\langle \sigma v \rangle_{A'A' \to \chi\overline{\chi}} \equiv \frac{\alpha_D^2}{m_\chi^2} g(r) \tag{5.12}$$

r	1.2	1.3	1.4	1.5	1.6	1.7	1.8
$f(r)$	9.47	14.1	23.7	45.9	105.7	312.9	1427
$g(r)$	4.44	5.49	5.90	5.94	5.77	5.50	5.19

Table 5.1: List of $f(r)$ and $g(r)$ values, as defined in Eqs. (5.12) and (5.13), evaluated at typical r-values of interest in this chapter.

where $g(r)$ is a function of r that captures the nontrivial r-dependence. Typical values of $g(r)$ are shown in Table 5.1.

As r increases, the rate of the forward process becomes exponentially more suppressed as the mass difference between χ and A' increases. Note that in the forward direction, $\chi\overline{\chi} \to A'A'$ removes kinetic energy from the dark sector; the rate of the forward reaction also becomes exponentially suppressed as T' decreases, since less kinetic energy is available to χ particles for conversion into the rest mass of A' particles.

2. *The* $3 \leftrightarrow 2$ *dark sector process,* $\chi\overline{\chi}\chi \leftrightarrow A'\chi$. For $1.5 \lesssim r \lesssim 2$, the freezeout of the dark sector is mainly controlled by this process, as examined in Ref. [120], once again under the assumption of a dark sector in thermal equilibrium with the SM. The forward process is a $3 \to 2$ process, with velocity-averaged annihilation cross section given by

$$\langle \sigma v^2 \rangle_{\chi\overline{\chi}\chi \to A'\chi} \equiv \langle \sigma v^2 \rangle \equiv \frac{\alpha_D^3}{m_\chi^5} f(r)\,, \tag{5.13}$$

where $f(r)$ encodes the nontrivial r-dependence of the cross section; once again, the full expression for $\langle \sigma v^2 \rangle_{\chi\overline{\chi}\chi \to \chi A'}$ is given in App. B.1. For ease of notation, we will drop the subscript on the thermally averaged cross section from here on, unless it is needed to avoid ambiguity. Typical values for $f(r)$ across the range of r considered in this chapter are shown in Table 5.1. Note that the forward reaction converts rest mass to kinetic energy, and heats the dark sector, similar to other $3 \to 2$ processes found in cannibal dark matter models [93, 94, 121–123].

3. $A' \leftrightarrow f\overline{f}$. The dark photon kinetically mixes with the SM photon, and can decay into a pair of SM fermions. This process is an important number-changing process for A' particles, and is one of two important processes responsible for transferring energy between the two sectors. The decay width Γ of A' is given in full in App. B.1.

4. $\chi f \leftrightarrow \chi f$. This elastic scattering process, and all possible processes related by conjugation, allows χ to directly transfer energy to or from the SM. This process as well as $A' \leftrightarrow f\overline{f}$ together determine how efficiently energy gets transferred between the two sectors. Once both $\chi f \leftrightarrow \chi f$ and $A' \leftrightarrow f\overline{f}$ become sufficiently inefficient, the dark sector and the SM can lose thermal contact and kinetically decouple, falling out of thermal equilibrium.

There are additional $3 \leftrightarrow 2$ dark-sector-only processes that we do not consider, such as $\chi\bar{\chi}A' \rightarrow A'A'$ and $A'A'A' \rightarrow \chi\bar{\chi}$. Since we are only considering $m_{A'} > m_\chi$, these $3 \rightarrow 2$ processes have rates that are parametrically suppressed by at least one power of $\exp(-(r-1)m_\chi/T')$ compared to $\chi\overline{\chi}\chi \rightarrow A'\chi$, and α_D times at least one power of $\exp(-m_{A'}/T')$ compared to $\chi\overline{\chi} \rightarrow A'A'$. These slower processes are therefore relatively unimportant compared to the much faster $3 \leftrightarrow 2$ and $2 \leftrightarrow 2$ processes shown here.

We also neglect the processes $A'f \leftrightarrow \gamma f$ and $A'\gamma \leftrightarrow f\overline{f}$: these processes are suppressed by an additional factor of the electromagnetic fine structure constant $\alpha_{\rm EM}$ relative to $A' \rightarrow f\overline{f}$, and are also Boltzmann suppressed by $n_{A'} \ll n_\chi$ relative to $\chi f \rightarrow \chi f$. Consequently, they never control when thermal decoupling between the two sectors occurs. They also do not play any important role based on the analytic understanding that we will develop below; they may only appear as terms proportional to $n_{A'} - n_{A',0}(T)$ in the Boltzmann equations, and can therefore be treated as small corrections to energy transfer rate arising from decays.

5.3.3 Boltzmann Equations

The evolution of the system is governed by the coupled Boltzmann equations for the number densities of χ and A', n_χ and $n_{A'}$, respectively, along with their energy densities ρ_χ, $\rho_{A'}$ and pressures P_χ, $P_{A'}$:

$$\begin{aligned}\frac{dn_\chi}{dt} + 3Hn_\chi = & -\frac{1}{4}\langle\sigma v^2\rangle\left[n_\chi^3 - \frac{n_{\chi,0}(T')^2}{n_{A',0}(T')}n_\chi n_{A'}\right] \\ & + \langle\sigma v\rangle_{A'A'\to\bar{\chi}\chi}\left[n_{A'}^2 - \frac{n_{A',0}(T')^2}{n_{\chi,0}(T')^2}n_\chi^2\right] \\ & - \frac{1}{2}\langle\sigma v\rangle_{\chi\bar{\chi}\to f\bar{f}}\left[n_\chi^2 - n_{\chi,0}^2(T)\right] ,\end{aligned} \tag{5.14}$$

$$\begin{aligned}\frac{dn_{A'}}{dt} + 3Hn_{A'} = & \frac{1}{8}\langle\sigma v^2\rangle\left[n_\chi^3 - \frac{n_{\chi,0}(T')^2}{n_{A',0}(T')}n_\chi n_{A'}\right] \\ & - \langle\sigma v\rangle_{A'A'\to\bar{\chi}\chi}\left[n_{A'}^2 - \frac{n_{A',0}(T')^2}{n_{\chi,0}(T')^2}n_\chi^2\right] \\ & - \Gamma\left[n_{A'} - n_{A',0}(T)\right] ,\end{aligned} \tag{5.15}$$

and

$$\begin{aligned}&\frac{d(\rho_\chi + \rho_{A'})}{dt} + 3H(\rho_\chi + \rho_{A'} + P_\chi + P_{A'}) \\ &\quad = -\langle\sigma v\delta E\rangle_{\chi f\to\chi f}n_\chi n_f - m_{A'}\Gamma\left[n_{A'} - n_{A',0}(T)\right] \\ &\qquad - \frac{1}{2}m_\chi\langle\sigma v\rangle_{\chi\overline{\chi}\to f\overline{f}}\left[n_\chi^2 - n_{\chi,0}^2(T)\right] ,\end{aligned} \tag{5.16}$$

where n_f is the number density of charged SM particles, which for simplicity we assume to consist only of electrons and positrons. This assumption is justified because we are considering sub-GeV dark matter and dark photons, so thermal equilibrium between the SM and dark sector typically holds down to temperatures of $T \lesssim 100\,\mathrm{MeV}$, at which point all other SM particles have annihilated or decayed away. Note that all dark sector (SM) variables are evaluated at the dark sector temperature T' (SM temperature T) unless otherwise stated. The prefactors for each term account for our convention of including both χ and $\overline{\chi}$ in n_χ, and for initial state symmetry factors.

Our convention, as well as the derivation of the dark sector annihilation cross sections, can be found in Ref. [120]. We take the limit of nonrelativistic χ and A' for the $A' \to f\bar{f}$ and $\chi\bar{\chi} \to f\bar{f}$ energy transfer rates. Details on the energy transfer rate for elastic scattering $\chi f \to \chi f$ can be found in App. B.2; in particular, we highlight the fact that we have calculated $\langle \sigma v \delta E \rangle_{\chi f \to \chi f}$ analytically without assuming that f is relativistic, which is to our knowledge a new result. This result is important when $m_\chi \sim \mathcal{O}(\text{MeV})$.

Eqs. (5.14)–(5.16) contain three unknowns: n_χ, $n_{A'}$ and T', and can be solved numerically for the coupled evolution of these variables as a function of the SM temperature T. The numerical solution of these equations is used for all of the results throughout the chapter.

We will also rely significantly on analytic approximations to gain some intuition for these results. To this end, it is useful to write the energy density Boltzmann equation Eq. (5.16) in the nonrelativistic limit. Expanding the energy densities to first order in $1/x'$, which is a small parameter once $T, T' \ll m_\chi$, we find

$$\begin{aligned} \rho_\chi &\simeq m_\chi n_\chi \left(1 + \frac{3}{2x'}\right) , \\ \dot{\rho}_\chi &\simeq m_\chi \dot{n}_\chi \left(1 + \frac{3}{2x'}\right) - \frac{3x}{2x'^2}\frac{dx'}{dx} H m_\chi n_\chi , \end{aligned} \tag{5.17}$$

and similarly for $\rho_{A'}$. We have also made the approximation $dx/dt \simeq Hx$. With these expansions, we obtain

$$\begin{aligned} (\dot{n}_\chi + r\dot{n}_{A'})\left(1 + \frac{3}{2x'}\right) + 3H(n_\chi + r n_{A'})\left(1 + \frac{5}{2x'} - \frac{x}{2x'^2}\frac{dx'}{dx}\right) \\ \simeq -\langle \sigma v \delta E \rangle \frac{n_\chi n_f}{m_\chi} - r\Gamma\left[n_{A'} - n_{A',0}(T)\right] . \end{aligned} \tag{5.18}$$

We have neglected DM annihilation into SM fermions in this analytic estimate for simplicity, since this process is typically not important in the regions of parameter space we will be interested in. We also find numerically that $(x/2x'^2)dx'/dx \ll \mathcal{O}(1)$ in all scenarios, and thus can be neglected in Eq. (5.18). The simplified Boltzmann

energy density equation to leading order then reads

$$\begin{aligned}\dot{n}_\chi + r\dot{n}_{A'} + 3H(n_\chi + rn_{A'}) \simeq &- \langle\sigma v \delta E\rangle \frac{n_\chi n_f}{m_\chi} \\ &- r\Gamma\left[n_{A'} - n_{A'0}(T)\right] .\end{aligned} \tag{5.19}$$

Comparing this with the sum of the χ and A' number density Boltzmann equations, Eqs. (5.14) and (5.15), which is given by

$$\begin{aligned}\dot{n}_\chi + \dot{n}_{A'} + 3H(n_\chi + n_{A'}) = &- \frac{1}{8}\langle\sigma v^2\rangle \left[n_\chi^3 - \frac{n_{\chi,0}^2}{n_{A',0}} n_\chi n_{A'}\right] \\ &- \Gamma[n_{A'} - n_{A',0}(T)] ,\end{aligned} \tag{5.20}$$

we finally obtain the following compact expression for the χ number density evolution:

$$\begin{aligned}\dot{n}_\chi + 3Hn_\chi \simeq &- \frac{n_\chi n_f}{(1-r)m_\chi}\langle\sigma v \delta E\rangle \\ &+ \frac{r}{8(1-r)}\langle\sigma v^2\rangle \left[n_\chi^3 - \frac{n_{\chi,0}^2}{n_{A',0}} n_\chi n_{A'}\right] .\end{aligned} \tag{5.21}$$

Comparing this expression with the number density Boltzmann equation for χ, we find

$$\begin{aligned}\langle\sigma v\rangle_{A'A'\to\chi\overline{\chi}} \left[n_{A'}^2 - \frac{n_{A',0}^2}{n_{\chi,0}^2} n_\chi^2\right] \simeq &\frac{2-r}{8(1-r)}\langle\sigma v^2\rangle \left[n_\chi^3 - \frac{n_{\chi,0}^2}{n_{A',0}} n_\chi n_{A'}\right] \\ &- \frac{n_\chi n_f}{(1-r)m_\chi}\langle\sigma v \delta E\rangle .\end{aligned} \tag{5.22}$$

With this relation, we can also reformulate the number density Boltzmann equation for A' as

$$\begin{aligned}\dot{n}_{A'} + 3Hn_{A'} \simeq &- \frac{1}{8(1-r)}\langle\sigma v^2\rangle \left[n_\chi^3 - \frac{n_{\chi,0}^2}{n_{A',0}} n_\chi n_{A'}\right] \\ &+ \frac{n_\chi n_f}{(1-r)m_\chi}\langle\sigma v \delta E\rangle - \Gamma\left[n_{A'} - n_{A',0}(T)\right] .\end{aligned} \tag{5.23}$$

These equations show that in the nonrelativistic limit, the Boltzmann equations establish certain relations between the rates of the various processes, determined ultimately by number and energy conservation. These equations will prove to be extremely useful for gaining analytic understanding of our numerical results.

5.3.4 Fast Reactions and Freezeout

To gain an understanding of the freezeout behavior of our dark sector, it is useful to understand when processes are occurring at rates fast enough to influence the freezeout process, and when they cease to be important. For temperatures $T \gtrsim m_\chi$, the rates of all of the process are generally fast, i.e. the rates of all processes in one direction are all much larger than the Hubble rate. For example, the $3 \leftrightarrow 2$ process is considered fast when

$$\frac{1}{4} n_\chi \langle \sigma v^2 \rangle \gg H(T)\,. \tag{5.24}$$

While a process is fast, the corresponding terms in square brackets in Eqs. (5.14) and (5.15) will generically be small, e.g. for the $3 \leftrightarrow 2$ process,

$$n_\chi^3 \approx \frac{n_{\chi,0}^2}{n_{A',0}} n_\chi n_{A'} \quad (\text{fast } 3 \leftrightarrow 2) \tag{5.25}$$

such that

$$\frac{1}{4} \langle \sigma v^2 \rangle \left[n_\chi^3 - \frac{n_{\chi,0}^2}{n_{A',0}} n_{A'} n_\chi \right] \approx H n_\chi \quad (\text{fast } 3 \leftrightarrow 2)\,; \tag{5.26}$$

otherwise, the $3 \leftrightarrow 2$ process can change the number densities of both χ and A' within a time much faster than the Hubble time, until Eq. (5.26) is satisfied.

Similarly, the $2 \leftrightarrow 2$ process is fast when

$$n_\chi \langle \sigma v \rangle_{\chi\overline{\chi} \to A'A'} \gg H(T)\,, \tag{5.27}$$

with

$$n_{A'}^2 \approx \frac{n_{A',0}^2}{n_{\chi,0}^2} n_\chi^2 \quad (\text{fast } 2 \leftrightarrow 2) \tag{5.28}$$

such that

$$\langle \sigma v \rangle_{\chi\overline{\chi}\to A'A'} \left[n_{A'}^2 - \frac{n_{A',0}^2}{n_{\chi,0}^2} n_\chi^2 \right] \approx H n_\chi \quad (\text{fast } 2 \leftrightarrow 2)\,. \tag{5.29}$$

Once $T \ll m_\chi$, the number densities of both χ and A' are Boltzmann suppressed and rapidly decrease. At some point, the forward rates of these processes become comparable to the Hubble rate, and the process freezes out. For the $3 \to 2$ process, this happens when

$$\frac{1}{4} n_\chi \langle \sigma v^2 \rangle \approx H(T) \quad (3 \leftrightarrow 2 \text{ freezeout}) \tag{5.30}$$

and for the $2 \leftrightarrow 2$ process,

$$n_\chi \langle \sigma v \rangle_{\chi\overline{\chi}\to A'A'} \approx H(T) \quad (2 \leftrightarrow 2 \text{ freezeout})\,. \tag{5.31}$$

Similar results hold for $\chi\overline{\chi} \to f\overline{f}$, just like in the conventional WIMP scenario.

The approximate relations found in Eqs. (5.25) and (5.28) when the $3 \leftrightarrow 2$ and $2 \leftrightarrow 2$ processes are fast can be rewritten in terms of the effective chemical potential μ_χ and $\mu_{A'}$ as

$$2\mu_\chi \approx \mu_{A'} \qquad (\text{fast } 3 \leftrightarrow 2)\,, \tag{5.32}$$

and

$$\mu_\chi \approx \mu_{A'} \qquad (\text{fast } 2 \leftrightarrow 2) \tag{5.33}$$

respectively. Note that when both processes are fast, these relations together enforce $\mu_\chi \approx \mu_{A'} \approx 0$.

For processes that are responsible for transferring heat between the SM and dark sector, the criterion for when these processes are "fast" depend on how much heat is generated/removed due to the $2 \leftrightarrow 2$ and $3 \leftrightarrow 2$ processes described above. Since the energy density of the dark sector for $T \ll m_\chi$ is dominated by the χ as $n_\chi \gg n_{A'}$,

the rate of change of dark sector energy density per dark sector particle is given approximately by $m_\chi \dot{n}_\chi / n_\chi$; processes are considered "fast" if they can transfer heat between the sectors at a comparable rate.

As discussed in Sec. 5.3.2, the two most important processes transferring energy between the two sectors are $A' \leftrightarrow f\overline{f}$ and $\chi f \leftrightarrow \chi f$. Let us first focus on the process $A' \leftrightarrow f\overline{f}$. In scenarios where both the $3 \leftrightarrow 2$ and $2 \leftrightarrow 2$ processes are fast, the number densities of the dark sector particles are given by $n_{\chi,0}(T')$ and $n_{A',0}(T')$. When $T \gg m_\chi$, $A' \leftrightarrow f\overline{f}$ is generally fast enough to maintain thermal equilibrium between the two sectors, so that $T' = T$. However, once $m_\chi > T$, $n_{A',0}(T)$ drops rapidly, and the number densities of the dark sector $n_{\chi,0}(T)$ and $n_{A',0}(T)$ evolve to a point where

$$\frac{n_{A',0}}{n_{\chi,0}} r m_\chi \Gamma \approx \frac{m_\chi \dot{n}_{\chi,0}}{n_{\chi,0}} \,. \tag{5.34}$$

After this point, the term on the left-hand side starts to become small relative to the right-hand side, and $A' \leftrightarrow f\overline{f}$ becomes ineffective at maintaining both sectors in thermal equilibrium. Similarly, the dark sector number densities can evolve to a point where

$$n_f \langle \sigma v \delta E \rangle_{\chi f \to \chi f} \approx \frac{m_\chi \dot{n}_{\chi,0}}{n_{\chi,0}} \frac{T - T'}{T} \,, \tag{5.35}$$

after which $\chi f \leftrightarrow \chi f$ is too slow to maintain thermal equilibrium. Once both Eq. (5.34) and (5.35) have been met, kinetic decoupling occurs, and the dark sector temperature T' starts to diverge from the SM temperature T. Keep in mind that $\langle \sigma v \delta E \rangle_{\chi f \to \chi f}$ is proportional to $(T' - T)/T$; the comparison made in Eq. (5.35) is therefore between the heat transfer rate when $|T' - T|/T \sim \mathcal{O}(1)$ and the energy lost due to $n_{\chi,0}$ decreasing.

We are now ready to understand the broad features of the thermodynamic evolution of the dark sector. There are three thermodynamic phases that the dark sector in our model may go through:

1. *Thermodynamic phase A: dark sector in thermal equilibrium with the SM.* Interactions between the dark sector and the SM allow the two sectors to exchange heat. If these interactions are sufficiently fast, the dark sector stays in thermal equilibrium with the SM with $T' = T$, and the number densities of χ and A' are simply given by $n_{\chi,0}(T)$ and $n_{A',0}(T)$;

2. *Thermodynamic phase B: $T' \neq T$ with zero chemical potential.* Once $A' \to f\overline{f}$ and $\chi f \to \chi f$ become too slow, the dark sector kinetically decouples, and develops a temperature different from T. The $2 \leftrightarrow 2$ and $3 \leftrightarrow 2$ dark sector processes can inject or remove heat from the dark sector. While both processes are fast, Eqs. (5.32) and (5.33) enforce $\mu_\chi \approx \mu_{A'} \approx 0$.

3. *Thermodynamic phase C: $T' \neq T$, with nonzero chemical potential.* If either the $3 \to 2$ or the $2 \to 2$ process freezes out after the SM-dark sector processes become slow, χ and A' develop a chemical potential $\mu_\chi(T')$ and $\mu_{A'}(T')$ respectively, according to either Eqs. (5.32) and (5.33).

In some parts of parameter space in the models we study, the dark sector goes through all three phases sequentially; in other parts of parameter space, a nonzero chemical potential develops once T' starts diverging from T, leading to a direct transition from phase A to C without spending any significant time in phase B.

Previous studies investigating this model [119, 120] have assumed that the dark sector only stays in thermodynamic phase A, with Ref. [119] making the further assumption that $n_{A'} = n_{A',0}(T)$ throughout in their thermally coupled model. However, we shall see that for values of ϵ as large as 10^{-5}, the dark sector does not stay in thermodynamic phase A throughout the process of freezeout, changing the dependence of the relic abundance on the model parameters drastically.

Throughout this chapter, we will mostly be interested in ϵ values that are small, of order 10^{-5} or smaller. However, if ϵ is too small, the dark sector and the SM sector need not have been in thermal contact at any point, calling into question the basic assumption we make that the two sectors start out in thermal equilibrium. To obtain an estimate for the minimum value of ϵ above which we are guaranteed to

have the dark sectors in thermal equilibrium at $T \sim m_\chi$, we follow Ref. [124], and set this minimum value of ϵ to be when the $f\overline{f} \to A'$ rate exceeds the Hubble rate at $T = m_{A'}$. When this condition is met, A' particles can be produced at a rate much faster than Hubble at $T \sim m_\chi$, allowing the whole dark sector to come into chemical equilibrium with the SM prior to the onset of the Boltzmann suppression from A' and χ going nonrelativistic. This condition can be written as [124]

$$\frac{\pi^2}{12\zeta(3)}\frac{m_{A'}}{T}\Gamma \sim \frac{T^2}{M_{\rm pl}}\,, \tag{5.36}$$

where ζ is the Riemann zeta function, and $M_{\rm pl}$ is the Planck mass. Using the expression for Γ in App. B.1 and setting $T = m_{A'}$, we obtain the following estimate for $\epsilon_{\rm eq}$, the minimum value of ϵ at which thermal equilibrium is guaranteed by $T \sim m_\chi$:

$$\epsilon_{\rm eq} \sim 7 \times 10^{-9} \left(\frac{m_{A'}}{\rm GeV}\right)^{1/2}\,. \tag{5.37}$$

In practice, experimental constraints will limit us to values of $\epsilon \gtrsim 10^{-8}$; we can therefore safely assume the dark sector to be thermally coupled to the SM at $T \sim m_\chi$ throughout this chapter.

5.4 $1.5 \lesssim r \lesssim 2$

We begin our discussion of the freezeout of the vector-portal dark matter model with $1.5 \lesssim r \lesssim 2$. For these values of r, the $2 \leftrightarrow 2$ process freezes out before the $3 \leftrightarrow 2$ process. Under the assumption that the dark sector stays in thermodynamic phase A with $T' = T$, this regime — which we call the "classic not-forbidden dark matter (NFDM)" regime — was studied in Ref. [120], and was found to be a viable model for sub-GeV dark matter with appreciable self-interaction rates and thus the potential to affect the small-scale structure of galaxies. Here, we explore $1.5 \lesssim r \lesssim 2$ including the temperature evolution of the dark sector.

5.4.1 "Classic Not-Forbidden" Regime

For sufficiently small values of ϵ with $r \gtrsim 1.5$, the $3 \leftrightarrow 2$ process eventually freezes out later than $\chi\overline{\chi} \to f\overline{f}$ — the process that controls conventional WIMP freezeout — and starts to become the main process that controls the final abundance of χ. This transition occurs when

$$n_\chi \langle \sigma v \rangle_{\chi\overline{\chi}\to f\overline{f}} \approx H \approx \frac{1}{4} n_\chi^2 \langle \sigma v^2 \rangle \,, \tag{5.38}$$

i.e. when both processes freeze out at roughly the same time. Using the analytic expressions for the quantities above, we obtain an estimate for $\epsilon_{\rm N/W}$, the value of ϵ that sets the boundary between the 'classic NFDM' regime and the WIMP regime:

$$\epsilon_{\rm N/W} \sim 2 \times 10^{-5} (4 - r^2) \left(\frac{g_*(x_f)}{10.75}\right)^{1/8} \times \left(\frac{\alpha_D}{1.0}\right)^{1/4} \left(\frac{20}{x_f}\right)^{1/2} \left(\frac{m_\chi}{\rm GeV}\right)^{1/4} \left(\frac{f(r)}{105.7}\right)^{1/4} , \tag{5.39}$$

where $x_f \equiv m_\chi / T_f$, and T_f is the temperature at which freezeout of either of these two processes occur. g_* is the effective number of relativistic degrees of freedom that enters into the Hubble parameter, $H(T) = 1.66 g_*^{1/2}(T) T^2 / M_{\rm pl}$. Further requiring that the final relic abundance of DM is equal to the observed one today gives a relation between α_D, ϵ and m_χ. In the WIMP regime, where freezeout is controlled by $\chi\overline{\chi} \to f\overline{f}$, the correct relic abundance is obtained when Eq. (5.10) is satisfied. This allows us to predict:

$$\epsilon_{\rm N/W} \sim 10^{-5} (4 - r^2) \left(\frac{\alpha_D}{1.0}\right)^{1/2} \left(\frac{g_{*,s}(x_f)}{10.75}\right)^{1/6} \times \left(\frac{g_*(x_f)}{10.75}\right)^{1/12} \left(\frac{20}{x_f}\right)^{5/6} \left(\frac{f(r)}{105.7}\right)^{1/3} \tag{5.40}$$

as the boundary between the conventional WIMP-like regime and the "classic NFDM" regime when the correct relic abundance is achieved.

For $\epsilon < \epsilon_{\rm N/W}$, the freezeout of the $3 \leftrightarrow 2$ process determines the abundance of

DM, and the parameters that generate the correct relic abundance become virtually independent of ϵ, provided that ϵ is large enough that the system remains in thermodynamic phase A (i.e. $T' = T$) throughout freezeout.

5.4.2 Kinetically Decoupling Relic (KINDER) Regime

As ϵ decreases further, processes that exchange energy between the dark and SM sectors become gradually less efficient; eventually, thermal equilibrium between the two sectors is lost even prior to $2 \leftrightarrow 2$ freezeout. This scenario, which we call the kinetic decoupling relic (KINDER) regime, is starkly different from the "classic NFDM" regime explained above. Notably, the abundance of DM after freezeout is governed primarily by when kinetic decoupling occurs, and therefore depends on both ϵ and α_D. With thermal equilibrium between the two sectors lost prior to the freezeout of dark sector processes, the dark sector now goes through the different thermodynamic phases described in Sec. 5.3.4.

General Features

In Fig. 5-3, we show the abundances of χ and A', as well as the dark sector temperature T' as a function of x for our benchmark parameter values in the KINDER regime: $m_\chi = 10\,\mathrm{MeV}$, $\alpha_D = 1$, $\epsilon = 4\times10^{-8}$, and $r = 1.8$. For ease of presentation, we plot the abundance as $m_\chi Y_\chi$ and $m_\chi Y_{A'}$, where Y_i is defined in Eq. (5.10). In Fig. 5-4, we show the number density and energy density rates for the relevant dark sector processes; explicitly, these are the terms for each process that appear on the right-hand side of Eqs. (5.14) and (5.15) divided by n_χ for number density rates, and the right-hand side of Eq. (5.16) divided by n_χ for energy density rates. At this parameter point (which is representative of the KINDER regime), the dark sector freezeout proceeds through the following stages:

1. *Kinetic decoupling, transition from thermodynamic phase A to B.* While either $\chi f \to \chi f$ and $A' \to f\overline{f}$ occur at rates larger than or comparable to the kinetic energy production rate of χ (i.e. the left-hand side of Eqs. (5.34) or (5.35) are

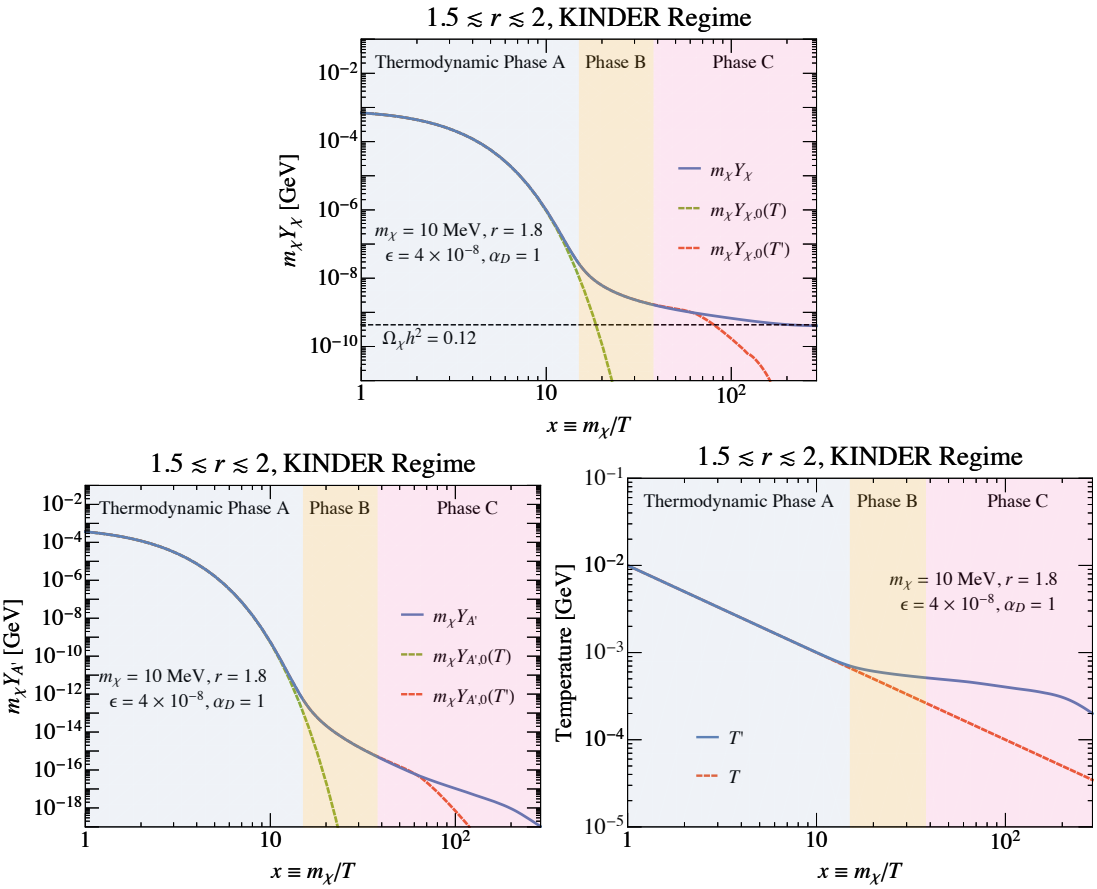

Figure 5-3: Dark sector evolution in the KINDER regime for $1.5 \lesssim r \lesssim 2$, with parameters $m_\chi = 10\,\text{MeV}$, $r = 1.8$, $\epsilon = 4 \times 10^{-8}$ and $\alpha_D = 1$. In all three plots, thermodynamic phases A, B and C as defined in Sec. 5.3.4 are shown in light blue, yellow and pink respectively. (*Top left*) χ abundance (given as $m_\chi Y_\chi$) as a function of x (blue line), with the zero chemical potential abundance at the SM temperature $m_\chi Y_{\chi,0}(T)$ (green dashed line) and the dark sector temperature $m_\chi Y_{\chi,0}(T')$ (red dashed line) shown for reference. The observed DM abundance is indicated by the horizontal black dashed line, as defined in Eq. (5.10). (*Top right*) A' abundance (given as $m_\chi Y_{A'}$) as a function of x (blue line), with $Y_{A',0}(T)$ (green dashed line) and $Y_{A',0}(T')$ (red dashed line) once again given for reference. (*Bottom*) The dark sector temperature T' (blue line), as a function of the SM temperature (red dashed line).

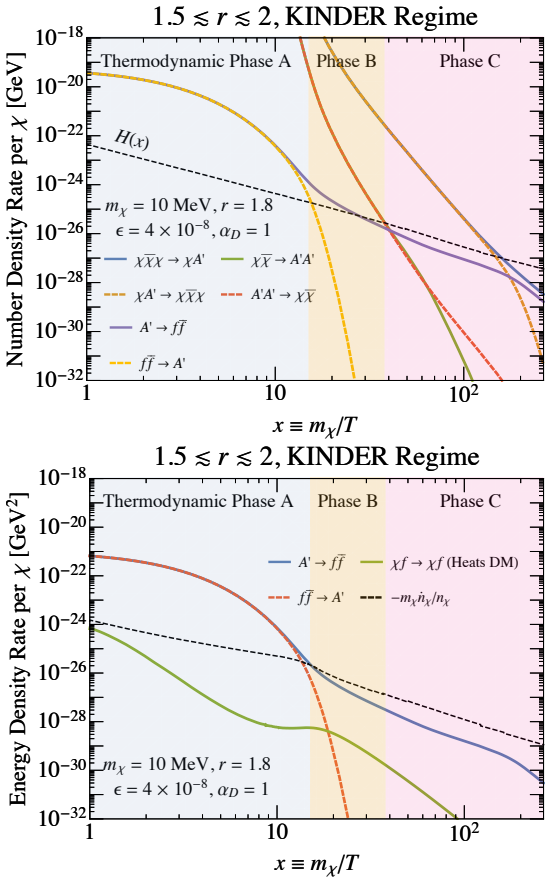

Figure 5-4: Rates of change in number density and energy density of the dark sector in the $1.5 \lesssim r \lesssim 2$ KINDER regime, with $m_\chi = 10\,\text{MeV}$, $r = 1.4$, $\epsilon = 4 \times 10^{-8}$ and $\alpha_D = 1$ In both plots, thermodynamic phases A, B and C as defined in Sec. 5.3.4 are shown in light blue, yellow and pink respectively. (*Left*) Number density rates for $\chi\overline{\chi}\chi \to \chi A'$ (blue line), $\chi A' \to \chi\overline{\chi}\chi$ (dark orange dashed line), $\chi\overline{\chi} \to A'A'$ (green line), $A'A' \to \chi\overline{\chi}$ (red dashed line) are shown, with solid lines indicating processes that net deplete χ's, and dashed lines indicating processes that net produce it instead. Also shown are the rates for $A' \to f\overline{f}$ (purple line) and $f\overline{f} \to A'$ (dashed yellow line). The Hubble parameter is shown in the black dashed line. (*Right*) Energy density rates for $A' \to f\overline{f}$ (blue line), $f\overline{f} \to A'$ (red dashed line) and $\chi f \to \chi f$ (green line), which has the net effect of heating the dark sector. The rate at which the energy density of dark matter is changing $-m_\chi \dot{n}_\chi / n_\chi$ (black dashed line) is shown for reference.

large compared to the RHS, kinetic equilibrium between the dark sector and SM particles is maintained at a common temperature $T = T'$. Once this is no longer true, i.e. after both $\chi f \to \chi f$ and $A' \to f\overline{f}$ become slow, kinetic decoupling occurs, and T' begins to diverge from T. For our benchmark parameter values, kinetic decoupling occurs when $A' \to f\overline{f}$ becomes slow, as shown in Fig. 5-4.

2. *Cannibalization in thermodynamic phase B.* After this point, both $2 \leftrightarrow 2$ and $3 \leftrightarrow 2$ processes remain fast, and the dark sector enters thermodynamic phase B, where $T' \neq T$ and $\mu_\chi \approx \mu_{A'} \approx 0$, since both processes are fast. The net effect of the dark sector processes is to convert mass to kinetic energy in the dark sector so as to deplete χ, and because this happens after the dark sector has kinetically decoupled from the SM, the dark sector heats up. This shares many similarities with dark matter models with a cannibal phase [93, 94, 121–123], but with two different species involved in the $3 \leftrightarrow 2$ process sustaining cannibalization instead of one. Like other cannibal DM models, the dark sector particles have zero chemical potential, and x' evolves in an approximately logarithmic manner with respect to x, with Y_χ evolving slowly. Unlike previous models, however, the entropy of the dark sector is not quite conserved, with $A' \to f\overline{f}$ decays remaining relatively efficient at depositing heat from the dark sector to the SM, but not fast enough to ensure equal temperatures; we will discuss this point in more detail below.

3. *Freezeout of* $2 \leftrightarrow 2$ *process, continued cannibalization.* After this point, the dark sector enters thermodynamic phase C with $2\mu_\chi \approx \mu_{A'}$, since the $3 \leftrightarrow 2$ process continues to be fast. Both χ and A' develop a nonzero chemical potential in thermodynamic phase C, and the logarithmic evolution of x' and slow evolution of Y_χ with respect to x continues until the $3 \leftrightarrow 2$ process freezes out. This is an extension of the conventional cannibal dark matter scenario that we will investigate in greater detail below.

4. *Freezeout of* $3 \leftrightarrow 2$ *process.* Finally, the $3 \to 2$ rate falls below the Hubble rate. With no other active number changing processes, the dark matter number

density n_χ evolves proportionally to a^{-3}.

Because the slow evolution of Y_χ takes place from the time of kinetic decoupling until the freezeout of the $3 \leftrightarrow 2$ process, the DM thermal relic density is governed mainly by the kinetic decoupling process. In this regime, the vector-portal DM model therefore shares many similarities with elastically decoupling (ELDER) dark matter [93], with the main differences being the existence of thermodynamic phase C mentioned in the last paragraph, and the fact that kinetic decoupling in vector-portal DM is frequently governed by $A' \leftrightarrow f\overline{f}$, instead of elastic scattering processes, i.e. $\chi f \leftrightarrow \chi f$. The dark sector entropy is also not fully conserved due to the existence of $A' \leftrightarrow f\overline{f}$.

Similarly to the boundary between the WIMP and "classic NFDM" regimes, we can estimate the value of ϵ at which we transition from the KINDER regime to the "classic NFDM" regime, by finding the value of ϵ for which kinetic decoupling and $3 \leftrightarrow 2$ freezeout occur at roughly the same time. We find that $A' \leftrightarrow f\overline{f}$ is often the process that governs kinetic decoupling, and so the boundary between these regimes occurs at the value of $\epsilon = \epsilon_{\mathrm{K/N}}$ where both Eqs. (5.30) and (5.34) are satisfied at the same SM temperature T. Analytically, we find

$$\begin{aligned}\epsilon_{\mathrm{K/N}} \sim 10^{-7} e^{9.9(r-1.6)} \left(\frac{\alpha_D}{1.0}\right)^{\frac{3(r-1)}{4}} \\ \times \left(\frac{1.6}{r}\right)^{9/4} \left(\frac{x_f}{20}\right)^{-\frac{r+3}{4}} \left(\frac{g_*(x_f)}{10.75}\right)^{-\frac{r-3}{8}} \\ \times \left(\frac{\mathrm{GeV}}{m_\chi}\right)^{\frac{r-3}{4}} \left(\frac{f(r)}{105.7}\right)^{\frac{r-1}{4}}\end{aligned} \tag{5.41}$$

as the boundary between the 'classic NFDM' regime and the KINDER regime, with x_f denoting the dimensionless inverse temperature at the freezeout of the $3 \leftrightarrow 2$ process. To obtain an expression analogous to Eq. (5.40) under the additional assumption that the correct relic abundance is obtained, i.e. that Eq. (5.10) is satisfied, we need to understand how the freezeout abundance of DM scales with the model parameters analytically in the KINDER regime. In the next few sections, we will review each thermodynamic phase of the KINDER regime, providing where possible an analytic

understanding of the KINDER freezeout process.

Kinetic Decoupling and Cannibalization

As we discussed in Sec. 5.3.4, kinetic decoupling occurs at the point when both Eqs. (5.34) and (5.35) have just been satisfied. We find that kinetic decoupling is usually controlled by $A' \leftrightarrow f\overline{f}$, i.e. the condition Eq. (5.34) is fulfilled after Eq. (5.35). Therefore, for the purpose of analytic estimates, we will assume that this is always true; our numerical results show that elastic scattering can become the process controlling kinetic decoupling at $m_\chi \sim \mathcal{O}(\text{GeV})$ and large α_D.

Let us first obtain an analytic estimate of x_d, the dimensionless inverse temperature at kinetic decoupling, to see how it depends on the parameters of our model. Using the expression in Eq. (5.8) with $T' = T$ and $\mu_\chi = 0$, Eq. (5.34) reads

$$x_d^2 e^{(1-r)x_d} \approx 2.2 \frac{g_*^{1/2}(x_d)}{r^{7/2}} \frac{m_\chi^2}{M_{\rm pl}\Gamma} \quad \text{(kinetic decoupling)}\,. \tag{5.42}$$

For our benchmark parameters in this regime, the value of x where this condition is met is shown in the right panel of Fig. 5-4 at the transition between thermodynamic phases A and B.

After kinetic decoupling the dark sector temperature T' deviates from the SM temperature T, as indicated in Fig. 5-3, while the $2 \to 2$ and $3 \to 2$ processes continue to proceed at rates larger than the Hubble expansion rate. The dark sector enters thermodynamic phase B, with both the $2 \leftrightarrow 2$ and $3 \leftrightarrow 2$ processes maintaining chemical equilibrium in the dark sector and forcing the chemical potentials to zero, as discussed in Sec. 5.3.4. During this phase, the dark sector is cannibalistic, undergoing a net conversion of mass to kinetic energy in the dark sector, which then causes the dark sector to heat up.

In the limit where no energy is transferred to the SM, the dark sector entropy

$s_D a^3$ is conserved. The dark sector entropy density can be approximated as

$$\begin{aligned} s_D &= \frac{\rho_\chi + \rho_{A'} + P_\chi + P_{A'} - \mu_\chi n_\chi - \mu_{A'} n_{A'}}{T'} \\ &\simeq \frac{m_\chi n_\chi - \mu_\chi n_\chi - \mu_{A'} n_{A'}}{T'}\,, \end{aligned} \tag{5.43}$$

where in the second line we can neglect $\rho_{A'}$ due to its relatively large Boltzmann suppression compared to ρ_χ, and we used the fact that $P_{A'} \ll P_\chi = n_\chi T' \ll m_\chi n_\chi$ for $x' \gg 1$. Conservation of entropy enforces $d\left(s_D a^3\right)/dt = 0$, with no processes active between the dark sector and the SM. In this limit, we have $\mu_\chi n_\chi \gg \mu_{A'} n_{A'}$ since μ_χ and $\mu_{A'}$ are of the same order, and $\mu_\chi \dot{n}_\chi + \mu_{A'} \dot{n}_{A'} \simeq 0$, since the fast dark sector processes are responsible for both setting the chemical potentials and the number density evolution of the dark sector particles. Making use of Eq. (5.43) and the expression of $\dot{n}_\chi$ in Eq. (5.8), entropy conservation in the dark sector implies the following relation between T' and T:

$$\frac{3}{T}\left(1 - \frac{\mu_\chi}{m_\chi}\right)\frac{dT}{dT'} \simeq \frac{1}{2T'} + \frac{m_\chi}{T'^2} + \left(1 - \frac{T'}{m_\chi}\right)\frac{d}{dT'}\left(\frac{\mu_\chi}{T'}\right) \qquad (s_D a^3 \text{ conserved}). \tag{5.44}$$

In thermodynamic phase B, we have $\mu_\chi \approx \mu_{A'} \approx 0$ and $m_\chi \gg T'$, giving:

$$\frac{m_\chi}{T'^2}\frac{dT'}{dT} \simeq \frac{3}{T} \qquad (s_D a^3 \text{ conserved}), \tag{5.45}$$

which we can integrate from $T'_d = T_d \equiv m_\chi / x_d$ up to some dark sector temperature T' to get

$$x' \simeq x_d + 3\log\left(\frac{x}{x_d}\right) \qquad (s_D a^3 \text{ conserved}). \tag{5.46}$$

We see that the dark sector temperature T' is approximately fixed by the temperature of kinetic decoupling T_d, with x' evolving slowly (logarithmically) with x thereafter.

If entropy were perfectly conserved, then the corresponding evolution in n_χ would be approximately

$$n_\chi \approx n_{\chi,0}(T_d) \times \frac{T^3}{T_d^3} \qquad (s_D a^3 \text{ conserved}). \tag{5.47}$$

which would indicate an approximately constant $n_\chi a^3$ and Y_χ in phase B.

While entropy conservation arguments are sufficient to get a crude approximation of the behavior of the dark sector in this phase, the true picture is significantly more complicated; for example, in Fig. 5-3, while Y_χ stops exponentially decreasing in phase B, it is clearly not constant. In Fig. 5-4, we see that the dark sector enters thermodynamic phase B after kinetic decoupling occurs at around $x_d \sim 15$ for our benchmark parameters. After kinetic decoupling, the $A' \leftrightarrow f\overline{f}$ is no longer fast enough to keep the dark sector and SM in thermal equilibrium. As a result, the dark sector begins to heat, as shown in the bottom panel of Fig. 5-3. With the increase in T', however, comes an increase in $n_{A'}$, which also increases the rate at which energy density is transferred by $A' \to f\overline{f}$ to the SM. As a result, the energy density transfer from the dark sector to the SM remains relatively large even after kinetic decoupling; this can be seen in Fig. 5-4, which shows that this rate stays close to the rate of change of the dark sector energy density per χ particle, given approximately by $m_\chi \dot{n}_\chi / n_\chi$. Dark sector entropy is thus not quite conserved.

A better analytic understanding for the dark sector evolution thermodynamic phase B can be obtained from the argument above: since T' always evolves in such a way as to keep $A' \to f\overline{f}$ relatively efficient at transferring energy from the dark sector to the SM, we find that

$$m_{A'} \Gamma \frac{n_{A'}}{n_\chi} \approx \frac{m_\chi \dot{n}_\chi}{n_\chi}\,. \tag{5.48}$$

Thermodynamic phase B is characterized by zero chemical potentials for both species, i.e. $n_\chi \approx n_{\chi,0}(T')$, and likewise for $n_{A'}$. Given the expression for $\dot{n}_\chi$ in Eq. (5.8), this

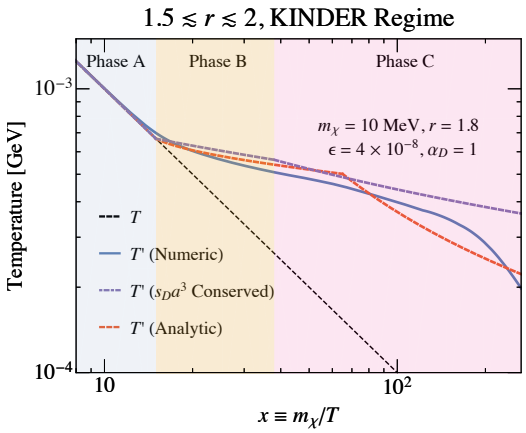

Figure 5-5: $1.5 \lesssim r \lesssim 2$, KINDER regime comparison between the improved analytic estimate of T' (red dashed line) using Eq. (5.50) in thermodynamic phase B and Eq. (5.62) in phase C and the full numeric calculation from the Boltzmann equations (blue line). We also show the predicted temperature assuming dark sector entropy conservation, using Eq. (5.46) in phase B and Eq. (5.62) in phase C is given by the purpled dashed line. The SM temperature T is shown for reference (black dashed line), with the thermodynamic phases A, B and C marked in light blue, yellow and pink.

approximation gives

$$\frac{3}{4}\Gamma r^{5/2} e^{(1-r)x'} \approx H\left(\frac{3x}{2x'} + x\right)\frac{dx'}{dx}\,. \tag{5.49}$$

Taking $3x/2x' \ll x$ and $x^2 H(x) \approx x_d^2 H(x_d)$, this differential equation is easily integrated to get

$$x' \approx x_d + \frac{1}{r-1}\log\left[1 + \frac{3}{8}\frac{\Gamma r^{5/2}(r-1)e^{(1-r)x_d}}{x_d^2 H(x_d)}(x^2 - x_d^2)\right]\,. \tag{5.50}$$

Fig. 5-5 shows the comparison between this analytic temperature evolution and the numerical evolution computed directly from the Boltzmann equations. We see that the analytic result assuming entropy conservation overestimates the temperature somewhat, since it neglects the transfer of energy to the SM, and our modified analytic estimate is in better agreement with the phase B numerical results.

As we indicated earlier, a very similar logarithmic evolution of x' in a kinetically decoupled dark sector with zero chemical potential has already been found in other

dark sector models [93, 94, 121–123]. However, as discussed above in Sec. 5.4.2, in the dark photon model parameter space we are studying, a second stage of cannibalization begins when the universe expands and cools to the point where the $2 \leftrightarrow 2$ process freezes out.

$2 \leftrightarrow 2$ Freezeout and Continued Cannibalization

The $2 \leftrightarrow 2$ process freezes out when the $\chi\overline{\chi} \to A'A'$ rate falls below the Hubble expansion rate, triggering a nonzero chemical potential in the dark sector; this is indicated on the left panel of Fig. 5-4 by the transition from phase B to C. We will label the temperatures of the SM and dark sector at which $2 \leftrightarrow 2$ freezeout occurs as T_2 and T_2' respectively, and correspondingly x_2 and x_2'.

To understand the behavior of the dark sector in this phase analytically, we rely on Eq. (5.21) and drop the contribution from elastic scattering, which is unimportant by the time the dark sector is in thermodynamic phase C. This gives

$$\dot{n}_\chi + 3Hn_\chi \simeq \frac{r}{8(1-r)}\langle\sigma v^2\rangle \left[n_\chi^3 - \frac{n_{\chi,0}^2}{n_{A',0}} n_\chi n_{A'}\right] \tag{5.51}$$

for the χ number density evolution, and

$$\dot{n}_{A'} + 3Hn_{A'} \simeq -\frac{1}{8(1-r)}\langle\sigma v^2\rangle \left[n_\chi^3 - \frac{n_{\chi,0}^2}{n_{A',0}} n_\chi n_{A'}\right] - \Gamma\left[n_{A'} - n_{A',0}(T)\right] . \tag{5.52}$$

for the A' number density.

In general, provided that the dark-sector number densities n_i ($i = \chi, A'$) are such that they would be in a steady state in the absence of the cosmic expansion, their time derivatives will be parametrically controlled by H and can be approximated as being of order Hn_i (the prefactor, of course, being important to the details of the solution). During the two cannibalization stages, when the comoving number density evolution is slow, we furthermore expect the prefactor to be an $\mathcal{O}(1)$ number.

Therefore, Eq. (5.51) shows that:

$$Hn_\chi \approx \frac{r}{8(r-1)}\langle\sigma v^2\rangle\left[n_\chi^3 - \frac{n_{\chi,0}^2}{n_{A',0}} n_\chi n_{A'}\right] . \tag{5.53}$$

The $3 \leftrightarrow 2$ term on the right-hand side of Eq. (5.53) also appears in the Boltzmann equation for A' shown in Eq. (5.15); however, since in general $n_{A'} \ll n_\chi$, we see that

$$\frac{1}{8(r-1)}\langle\sigma v^2\rangle\left[n_\chi^3 - \frac{n_{\chi 0}^2}{n_{A',0}} n_\chi n_{A'}\right] \gg Hn_{A'} . \tag{5.54}$$

In the parameter space of interest for obtaining the correct relic abundance in thermodynamic phase C, we generally have $\Gamma \gg H$ by the time $T \sim m_\chi$, as well as $n_{A'} \gg n_{A',0}(T)$, i.e.

$$\Gamma(n_{A'} - n_{A',0}) \approx \Gamma n_{A'} \gg Hn_{A'} . \tag{5.55}$$

As we argued above, we expect the right-hand side of Eq. (5.52) to be on the order of $Hn_{A'}$; since both terms on the right-hand side are large compared to $Hn_{A'}$, we expect these terms to be comparable in magnitude. Given that the $3 \leftrightarrow 2$ rate is on the order of Hn_χ as shown in Eq. (5.53), we therefore arrive at the following important approximate relation that is valid in phase C:

$$\frac{r}{8(r-1)}\langle\sigma v^2\rangle\left[n_\chi^3 - \frac{n_{\chi,0}^2}{n_{A',0}} n_\chi n_{A'}\right] \approx Hn_\chi \approx r\Gamma n_{A'} . \tag{5.56}$$

How well the last approximation in the equation above is satisfied determines the accuracy of our analytic results: in Fig. 5-4, we see that this approximation is satisfied up to a factor of 3 throughout phase C.

In thermodynamic phase C with a fast $3 \leftrightarrow 2$ process, recall from Eq. (5.32) that the chemical potentials of χ and A' are related by $\mu_{A'} \approx 2\mu_\chi$. We can therefore rewrite Eq. (5.56) as

$$\frac{4}{3} r^{-3/2} e^{(r-1)x'} e^{-\mu_\chi/T'} \approx \frac{r\Gamma}{H(T)} . \tag{5.57}$$

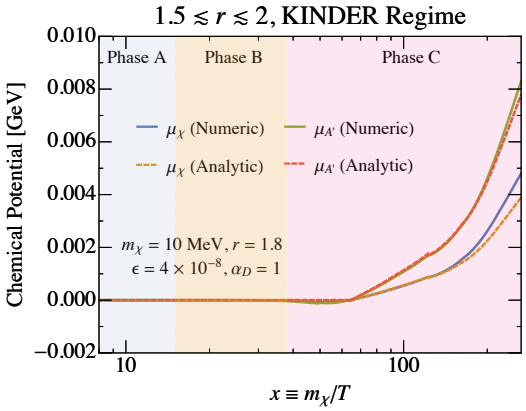

Figure 5-6: $1.5 \lesssim r \lesssim 2$, KINDER regime comparison between the analytic estimates of μ_χ (orange dashed line) and $\mu_{A'}$ (red dashed line) based on Eq. (5.66) and $\mu_{A'} \approx 2\mu_\chi$ with the numerical μ_χ (blue line) and $\mu_{A'}$ (green line) based on integrating the full Boltzmann equations.

At the point of $2 \to 2$ freezeout, with the dark sector and SM temperatures being T_2' and T_2 respectively, we still have $\mu_\chi(T_2') = 0$, and so we have

$$\frac{4}{3} r^{-3/2} e^{(r-1)x_2'} \approx \frac{r\Gamma}{H(T_2)}\,, \tag{5.58}$$

from which we finally obtain the following approximate relation for μ_χ as a function of T and T':

$$\frac{\mu_\chi}{T'} \approx (r-1)\, m_\chi \left[\frac{1}{T'} - \frac{1}{T_2'}\right] - \log\left[\frac{H(T_2)}{H(T)}\right]\,. \tag{5.59}$$

To obtain a full, analytic understanding of the dark sector evolution, we now need to determine T' as a function of T after $2 \to 2$ freezeout. We can once again obtain a rough approximation by taking the dark sector entropy to be conserved, in which case Eq. (5.44) determines the evolution of T' as a function of T. In order to get analytic control of the temperature evolution, we can make the approximations $T' \ll m_\chi$ and $\mu_\chi \ll m_\chi$; the latter condition is true early in phase C since the chemical potential starts at zero. Using the expression for μ_χ/T' derived in Eq. (5.59), we find

$$\frac{dT}{dT'} \approx \frac{T}{2T'} + (2-r)\frac{m_\chi T}{T'^2} \qquad (s_D a^3 \text{ conserved}). \tag{5.60}$$

After $2 \to 2$ freezeout, for values of r that are not too close to 2, we typically have $2(2-r)x' \gg 1$, and so we may drop the first term in the equation above to find that

$$\frac{dT}{dT'} \approx (2-r)\frac{m_\chi T}{T'^2} \qquad (s_D a^3 \text{ conserved}). \tag{5.61}$$

We may integrate this approximate expression to obtain

$$x' \approx x'_2 + \frac{1}{2-r}\log\left(\frac{x}{x_2}\right) \qquad (s_D a^3 \text{ conserved}). \tag{5.62}$$

which shows that even during thermodynamic phase C with a nonzero chemical potential in the dark sector, the dark sector temperature T' still evolves logarithmically with the SM temperature T. After the freezeout of the $2 \leftrightarrow 2$ process, the $3 \leftrightarrow 2$ process alone is sufficient to maintain cannibalization of the dark sector, even though a nonzero dark chemical potential μ has developed. This second stage of cannibalization which occurs in the KINDER scenario is an extension of the conventional cannibalization scenario. It is a critical part of the thermal history of KINDER, because it ensures that after $2 \leftrightarrow 2$ freezeout and before $3 \leftrightarrow 2$ freezeout, the dark sector temperature T' and comoving number density ($n_\chi a^3$) continue to evolve slowly, as Fig. 5-3 shows, remaining mostly fixed by their values at kinetic decoupling. We will explore this slow evolution of n_χ in more detail in Sec. 5.4.2.

As before, entropy conservation is not strictly obeyed due to the fact that $A' \to f\overline{f}$ remains quite efficient at transferring energy from the dark sector to the SM; a more sophisticated analytic understanding can once again be attained by examining the Boltzmann equations closely. First, with elastic scattering being unimportant, Eq. (5.22) shows that there is an approximate relationship between the $2 \leftrightarrow 2$ and $3 \leftrightarrow 2$ rates that is applicable even after $2 \leftrightarrow 2$ freezeout:

$$\langle\sigma v\rangle_{A'A' \to \chi\overline{\chi}}\left[n_{A'}^2 - \frac{n_{A',0}^2}{n_{\chi,0}^2}n_\chi^2\right] \simeq \frac{2-r}{8(1-r)}\langle\sigma v^2\rangle\left[n_\chi^2 - \frac{n_{\chi,0}^2}{n_{A',0}}n_\chi n_{A'}\right]. \tag{5.63}$$

As we argued in Eq. (5.53), the $3 \leftrightarrow 2$ rate is comparable to Hn_χ, which leads us to

conclude that

$$\langle\sigma v\rangle_{A'A'\to\chi\overline{\chi}}\left[n_{A'}^2-\frac{n_{A',0}^2}{n_{\chi,0}^2}n_\chi^2\right]\approx\frac{2-r}{r}Hn_\chi\,. \tag{5.64}$$

This expression demonstrates that just after the point of $2\leftrightarrow 2$ freezeout, defined in Eq. (5.31), the $\chi\overline{\chi}\to A'A'$ and $A'A'\to\chi\overline{\chi}$ rates remain close to each other, until

$$n_\chi\langle\sigma v\rangle_{\chi\overline{\chi}\to A'A'}\approx\frac{2-r}{r}H\,. \tag{5.65}$$

The fact that these rates are close even after $2\leftrightarrow 2$ freezeout can be seen in Fig. 5-4, immediately after the transition between phases B and C.

Before the condition in Eq. (5.65) is satisfied, we must therefore have $\mu_\chi\approx\mu_{A'}$ as well, which together with the fast $3\leftrightarrow 2$ requirement that $\mu_{A'}\approx 2\mu_\chi$ maintains the chemical potential of the dark sector at approximately zero. Moreover, temperature evolution continues to obey the temperature evolution derived in phase B, shown in Eq. (5.50). Eventually, n_χ decreases to a point where Eq. (5.65) becomes satisfied at some SM temperature T_μ and corresponding $x_\mu\equiv m_\chi/T_\mu$.

Above x_μ, the previous argument used to obtain Eq. (5.59) can be used to obtain a similar expression:

$$\frac{\mu_\chi}{T'}\approx(r-1)m_\chi\left[\frac{1}{T'}-\frac{1}{T'_\mu}\right]-\log\left[\frac{H(T_\mu)}{H(T)}\right]\,, \tag{5.66}$$

and the condition shown in Eq. (5.56) reduces the χ number density evolution to the following compact form:

$$\dot{n}_\chi+3Hn_\chi\simeq-r\Gamma n_{A'}\,. \tag{5.67}$$

Using the expression for $\dot{n}_\chi$ found in Eq. (5.8) as well as the expression for the chemical potential derived in Eq. (5.59), we obtain

$$-\left[\frac{3}{2T'}+(2-r)\frac{m_\chi}{T'^2}\right]\frac{dT'}{dT}\simeq-\frac{1}{T}-\frac{r\Gamma}{H(T_\mu)T}\frac{3}{4}r^{3/2}e^{(1-r)m_\chi/T'_\mu}\,, \tag{5.68}$$

If we make the approximation that $3/2 \ll (2-r)m_\chi/T'$, we can integrate this expression to obtain

$$x' \approx x'_\mu + \frac{1+C}{2-r}\log\left(\frac{x}{x_\mu}\right) , \tag{5.69}$$

where

$$C \equiv \frac{r\Gamma}{H(T_\mu)}\frac{3}{4}r^{3/2}e^{(1-r)x'_\mu} . \tag{5.70}$$

Compared to the estimate for x' in phase C obtained using entropy conservation in Eq. (5.62), we see that this more sophisticated analytic treatment *(i)* correctly identifies the delay in the onset of a nonzero chemical potential, and *(ii)* introduces a correction to the temperature evolution encapsulated by the factor C ($C \simeq 3.4$ for our benchmark parameters). The result of our analytic estimate for the temperature is shown in Fig. 5-5, and shows reasonable agreement with the fully numerical solution, up till the complete freezeout of the dark sector at $x \sim 200$. The agreement between the analytic estimate and the numerical result deteriorates at larger x as the approximation $Hn_\chi \approx r\Gamma n_{A'}$ becomes poor (we should only expect them to be equal up to an $\mathcal{O}(1)$ factor). The result for our improved analytic estimate for the chemical potentials using Eq. (5.66) is shown in Fig. 5-6, and shows good agreement with the numerical results.

$3 \to 2$ Freezeout and Relic Abundance

Cannibalization of the dark sector continues until the universe expands and cools to the point at which $3 \leftrightarrow 2$ annihilations freeze out at temperature T_3 (and corresponding x_3). This marks the freezeout of DM χ, at $x_3 \sim 200$ for our benchmark KINDER parameter point, as demonstrated in Figs. 5-3 and 5-4. After freezeout, the comoving DM abundance Y_χ settles to its constant relic value, and the dark sector temperature begins to evolve as $T' \propto T^2$, as expected for a completely decoupled nonrelativistic fluid.

Given the analytic estimates derived in the previous sections, we can now obtain an analytic estimate for the number density of DM particles at $3 \leftrightarrow 2$ freezeout, given by the condition shown in Eq. (5.30). We use the assumption of dark sector entropy conservation for simplicity, although a similar conclusion can be reached by using the more accurate analytic results described previously.

The number density of DM at freezeout can be written given the chemical potential in Eq. (5.59), giving

$$n_\chi(T_3') \approx 4 \left(\frac{m_\chi^2}{2\pi x_3'}\right)^{3/2} \frac{x_2^2}{x_3^2} e^{(r-2)x_3'} e^{(1-r)x_2'} \,. \tag{5.71}$$

However, the approximate expression for the temperature evolution in thermodynamic phase C found in Eq. (5.62) allows us to rewrite this as

$$n_\chi(T_3') \approx 4 \left(\frac{m_\chi T_3'}{2\pi}\right)^{3/2} \frac{x_2^3}{x_3^3} e^{-x_2'} \,. \tag{5.72}$$

Finally, using the expression for the temperature evolution during thermodynamic phase B in Eq. (5.50), we can rewrite x_2' in terms of x_d, the dimensionless inverse temperature at which kinetic decoupling occurs, giving

$$n_\chi(T_3') \simeq 4 \left(\frac{m_\chi^2}{2\pi x_3'}\right)^{3/2} e^{-x_d} \frac{x_d^3}{x_3^3} \,. \tag{5.73}$$

This remarkable expression shows explicitly that the freezeout abundance is mostly controlled by kinetic decoupling, being exponentially sensitive to x_d, up to small power law corrections.

Since the temperature of the dark sector evolves logarithmically after kinetic decoupling, we can make the approximation $x_3' \approx x_d$ in Eq. (5.73). Substituting the resulting expression into Eq. (5.30), we obtain the following estimate for n_χ at freezeout:

$$n_\chi(T_3') \simeq 8.2 e^{x_d/2} \left(\frac{g_*^{1/2}(x_3)}{x_d M_{\rm pl} \langle \sigma v^2 \rangle}\right)^{3/4} \,. \tag{5.74}$$

We are now ready to obtain an analytic estimate for $\epsilon_{K/N}$ as shown in Eq. (5.41), when the "classic NFDM" regime transitions into the KINDER regime in the α_D–ϵ plane, but now with the requirement that the correct relic abundance is achieved by choosing m_χ appropriately at each point in this parameter space. At the regime boundary, kinetic decoupling and $3 \to 2$ freezeout occur at roughly the same time, i.e. $x_d \approx x_3$. Combining the requirement shown in Eq. (5.10) for the correct relic abundance of χ with Eq. (5.74) gives

$$\left(\frac{m_\chi}{\text{GeV}}\right) \sim 0.1 \left(\frac{20}{x_d}\right)^{4/3} \left(\frac{g_{*,s}(x_d)}{10.75}\right)^{2/3} \left(\frac{g_*(x_d)}{10.75}\right)^{-1/6} \left(\frac{\alpha_D}{1.0}\right) \left(\frac{f(r)}{105.7}\right)^{1/3} . \quad (5.75)$$

Note that typical values of x_d are $x_d \simeq 18.5$ for $m_\chi \simeq$ MeV and $x_d \simeq 23.4$ for $m_\chi \simeq$ GeV. Substituting this expression into Eq. (5.41) leads to

$$\begin{aligned} \epsilon_{\text{K/N}} \sim 4 \times 10^{-8} e^{10.5(r-1.6)} & \left(\frac{1.6}{r}\right)^{9/4} \\ & \times \left(\frac{g_{*,s}(x_d)}{10.75}\right)^{\frac{3-r}{6}} \left(\frac{g_*(x_d)}{10.75}\right)^{\frac{3-r}{12}} \\ & \times \left(\frac{x_d}{20}\right)^{\frac{r-21}{12}} \left(\frac{\alpha_D}{1.0}\right)^{r/2} \left(\frac{f(r)}{105.7}\right)^{r/6} . \end{aligned} \quad (5.76)$$

Summary of regimes and boundaries for $1.5 < r < 2$

Fig. 5-7 shows the ϵ–α_D parameter space of this model with $1.5 \lesssim r \lesssim 2$, with contours at fixed values of m_χ indicating the values of ϵ and α_D for each m_χ at which the observed relic abundance of $\Omega_\chi h^2 = 0.12$ is obtained. We show $r = 1.6$ and $r = 1.8$ as two examples for this range of r values. The three different regimes that we have discussed in this section — the WIMP, "classic NFDM" and KINDER regimes — are shown in this parameter space, with the boundaries between the regimes given by $\epsilon_{\text{N/W}}$ defined in Eq. (5.39) between the WIMP and "classic NFDM" regimes, and by $\epsilon_{\text{K/N}}$ defined in Eq. (5.76) between the "classic NFDM" and KINDER regimes.

For large ϵ values above $\epsilon_{\text{N/W}}$, the contours follow a constant value of $\epsilon^2 \alpha_D$, the parameter combination that appears in the expression for $\langle \sigma v \rangle_{\chi\overline{\chi} \to f\overline{f}}$; this corresponds

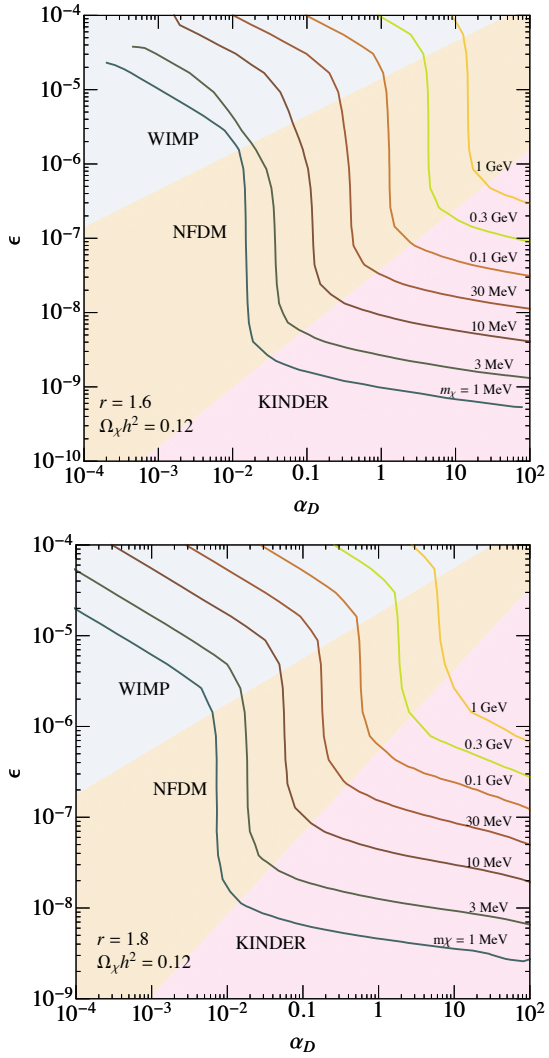

Figure 5-7: Contours of the observed relic abundance ($\Omega_\chi h^2 = 0.12$) in the α_D–ϵ plane for $1.5 \lesssim r \lesssim 2$, for *(left)* $r = 1.6$ and *(right)* $r = 1.8$, for various values of m_χ. The KINDER regime (pink), "classic NFDM" regime (orange) and WIMP regime (blue) are indicated, with the boundaries obtained using Eq. (5.39) for the WIMP/"classic NFDM" boundary and Eq. (5.41) for the "classic NFDM"/KINDER boundary.

to the WIMP regime.

Below $\epsilon_{\mathrm{N/W}}$, the freezeout of the dark sector transitions into the 'classic NFDM' regime, with the dark sector remaining in thermal contact up till the point of freeze-out, and with the abundance controlled solely by when the $3 \leftrightarrow 2$ process freezes out. Consequently — as previously discussed in Sec. 5.4.1 — the correct relic abundance does not depend on ϵ and is only determined by the value of α_D, leading to vertical contours.

For yet smaller values of ϵ, we eventually encounter the NFDM-KINDER boundary $\epsilon_{\mathrm{K/N}}$. Within the KINDER regime, the dark matter abundance is determined by the kinetic decoupling process; over much of the parameter space this process is controlled by $A' \leftrightarrow f\overline{f}$, which *only* depends on ϵ, leading to roughly horizontal contours of approximately constant ϵ. At larger values of m_χ, the elastic scattering process (which depends on α_D) becomes more important, and starts to play a bigger role in determining when kinetic decoupling occurs.

5.5 $1 \lesssim r \lesssim 1.5$

We will now focus on the behavior of the dark sector when $1 \lesssim r \lesssim 1.5$. For these values of r, the $2 \leftrightarrow 2$ process freezes out after the $3 \leftrightarrow 2$ process, leading to qualitatively different behavior in the dark sector. Solving the full Boltzmann equations given in Eqs. (5.14)–(5.16) reveals a rich and complicated picture, with both the freezeout of DM and the temperature of the dark sector showing drastically different behavior depending on the parameter values.

For $\epsilon \gtrsim 10^{-4}$, the dark sector is once again in the WIMP regime, and freezes out via $\chi\overline{\chi} \to f\overline{f}$. For smaller values of ϵ, we find four different regimes when $1 \lesssim r \lesssim 1.5$:

1. *Regime I: the "classic forbidden" scenario.* ϵ is large enough that $A' \leftrightarrow f\overline{f}$ is fast, so that $n_{A'} \simeq n_{A',0}(T)$; furthermore, $\chi f \to \chi f$ elastic scattering is sufficiently fast to ensure that the dark sector temperature is nearly equal to the SM temperature throughout the freezeout. The dark sector stays in thermodynamic phase A until the $2 \leftrightarrow 2$ process freezes out, and no dark sector

number-changing processes remain. This regime is precisely the limit studied in Ref. [119].

2. *Regime II: $n_{A'} = n_{A',0}(T)$, slight cooling.* At slightly smaller values of ϵ, the process $A' \leftrightarrow f\overline{f}$ is still fast enough to maintain $n_{A'} \simeq n_{A',0}(T)$. However, this condition is insufficient to keep the dark sector in thermal contact with the SM, which cools due to the net conversion of kinetic energy in χ particles into rest mass of the heavier A' particles through $\chi\overline{\chi} \to A'A'$. In regime II, ϵ is large enough for the elastic scattering process, $\chi f \leftrightarrow \chi f$, to transfer some heat from the SM to the dark sector, slowing the cooling.

3. *Regime III: $n_{A'} = n_{A',0}(T)$, rapid cooling.* Going to still smaller values of ϵ, the $A' \leftrightarrow f\overline{f}$ process is still fast enough to lock the number density of A' to $n_{A',0}(T)$, but $\chi f \to \chi f$ is too inefficient to transfer any heat from the SM to the dark sector at any point after $3 \leftrightarrow 2$ freezeout. In this limit, the rate of cooling is independent of ϵ, and the dark sector cools in a manner that only depends on α_D.

4. *Regime IV: KINDER.* For the smallest values of ϵ that we consider, kinetic decoupling of the dark sector from the SM occurs while both the $3 \to 2$ and $2 \to 2$ processes have rates that are much faster than Hubble. This shares many of the features of the KINDER regime discussed for $1.5 \lesssim r \lesssim 2$: the dark sector first enters thermodynamic phase B with zero chemical potential and logarithmic evolution of T' with respect to T, and then transitions to thermodynamic phase C after $3 \leftrightarrow 2$ freezeout.

We will first discuss the broad features of how the dark sector temperature evolves for $1 \lesssim r \lesssim 1.5$, before examining each of these regimes in turn, focusing on getting some analytic intuition for them. All of our results are once again obtained by solving the Boltzmann equations, Eqs. (5.14)– (5.16), numerically.

5.5.1 Dark Sector Temperature Evolution

In Regime I, the "classic forbidden" DM regime, the temperature evolution of the dark sector is trivially given by $T' = T$. For the other regimes, the $3 \leftrightarrow 2$ freezeout divides the dark sector temperature evolution into two important phases.

Temperature Evolution Before $3 \leftrightarrow 2$ Freezeout

In Regimes II and III, while both the $2 \leftrightarrow 2$ and $3 \leftrightarrow 2$ processes are fast, the simultaneous conditions imposed on the chemical potentials shown in Eqs. (5.32) and (5.33) are satisfied only if $\mu_\chi \approx \mu_{A'} \approx 0$. At the same time, ϵ is large enough such that $n_{A'} = n_{A',0}(T)$; therefore, we must have $n_\chi = n_{\chi,0}(T)$ as well, i.e. $T' = T$. Prior to $3 \leftrightarrow 2$ freezeout, Regimes II and III thus stay in thermodynamic phase A.

For the KINDER-like Regime IV during this phase, the temperature evolution is identical to the KINDER regime with $1.5 \lesssim r \lesssim 2$, with $T' = T$ prior to kinetic decoupling, and the dark sector entering thermodynamic phase B once decoupling occurs. While in thermodynamic phase B, the dark sector particles have zero chemical potential, and the temperature evolves as in Eq. (5.50).

Temperature Evolution After $3 \leftrightarrow 2$ Freezeout

Once the $3 \leftrightarrow 2$ process freezes out, the only process which depletes χ particles is $\chi\overline{\chi} \to A'A'$. This process converts lighter χ particles into heavier A' particles, removing kinetic energy from the dark sector, resulting in a cooling of the dark sector. The $2 \leftrightarrow 2$ process enforces $\mu_\chi \approx \mu_{A'}$, which start to take on nonzero values.

As we derived in Sec. 5.3.3, the Boltzmann equations enforce certain relations between the rates of the $3 \leftrightarrow 2$ process, the $2 \leftrightarrow 2$ process, $A' \leftrightarrow f\overline{f}$ and elastic scattering in the nonrelativistic limit. As shown in Eq. (5.21), we can approximately express the number density evolution of χ particles purely in terms of the elastic scattering rate and the $3 \leftrightarrow 2$ rate. In Regime III, the number density evolution between $3 \leftrightarrow 2$ and $2 \leftrightarrow 2$ freezeout is dominated solely by the $3 \leftrightarrow 2$ rate, with the elastic scattering term being negligible. Since the $3 \to 2$ rate has dropped below the

Hubble rate in this phase, Regime III is characterized by $n_\chi a^3$ being approximately constant, with the dark sector temperature being dependent only on the $3 \to 2$ rate. In Regime II, the number density evolution is instead dominated by the elastic scattering rate before $2 \leftrightarrow 2$ freezeout, leading to more rapid evolution of n_χ, and less deviation of T' from the SM temperature. In the limit of large elastic scattering, $n_\chi \to n_{\chi,0}(T)$ with $T' \to T$, which is the condition found in Regime I.

To understand the behavior of Regimes II and III more quantitatively, we can expand $\dot{n}_\chi$ in Eq. (5.21) using Eq. (5.8) to obtain

$$\left[\frac{3}{2T'} + \frac{m_\chi}{T'^2} + \frac{d}{dT'}\left(\frac{\mu_\chi}{T'}\right)\right]\frac{dT'}{dT} \simeq \frac{3}{T} - \frac{r}{8(1-r)HT}\langle\sigma v^2\rangle\left[n_\chi^2 - \frac{n_{\chi,0}^2}{n_{A',0}}n_{A'}\right] + \frac{n_f}{(1-r)m_\chi HT}\langle\sigma v \delta E\rangle_{\chi f \to \chi f}\,. \quad (5.77)$$

In Regimes II and III, approximations for μ_χ/T' after $3 \to 2$ freezeout can be found. In these regimes, the value of ϵ is large enough such that

$$n_{A'} \approx n_{A',0}(T) = 3\left(\frac{r m_\chi T}{2\pi}\right)^{3/2} e^{-r m_\chi/T}\,. \quad (5.78)$$

We emphasize, however, that the dark sector temperature T' is *not* equal to T; rather, the chemical potential $\mu_{A'}$ evolves in such a way as to maintain the relation above. The $\chi\overline{\chi} \to A'A'$ process removes kinetic energy from the dark sector, and the exact evolution of T' depends on the efficiency of the heat exchange processes between the dark sector and the SM. Writing out the full expression for $n_{A'}$ in Eq. (5.3) and making use of the fact that while the $2 \to 2$ process is the only process that is fast, Eq. (5.33) must hold i.e. $\mu_\chi \approx \mu_{A'}$, we find that the chemical potential must satisfy the following relation:

$$e^{\mu_\chi/T'} \approx e^{\mu_{A'}/T'} \approx \left(\frac{x'}{x}\right)^{3/2} e^{-r(x-x')}\,. \quad (5.79)$$

Furthermore, the ratio of n_χ and $n_{A'}$ is completely specified by x' since the chemical

potentials cancel out:

$$\frac{n_\chi}{n_{A'}} \approx \frac{4}{3} r^{-3/2} e^{(r-1)x'} \,. \tag{5.80}$$

Eqs. (5.79) and (5.80) show that given T' as a function of T, we will be able to obtain n_χ and $n_{A'}$ as a function of the SM temperature in Regimes II and III. Eq. (5.79) provides an expression for μ_χ/T', which combined with Eq. (5.77) gives an expression for T' as a function of T after the freezeout of the $3 \leftrightarrow 2$ process, with $n_{A'}(T') \approx n_{A',0}(T)$:

$$\begin{aligned}(1-r)\frac{m_\chi}{T'^2}\frac{dT'}{dT} \approx -\frac{r m_\chi}{T^2} + \frac{3}{2T} \\ - \frac{r}{8(1-r)HT}\langle\sigma v^2\rangle \left[n_\chi^2 - \frac{n_{\chi,0}^2}{n_{A',0}} n_{A'}\right] \\ + \frac{n_f}{(1-r)m_\chi HT}\langle\sigma v \delta E\rangle_{\chi f \to \chi f} \,.\end{aligned} \tag{5.81}$$

If we make the further approximation that $m_\chi \gg T, T'$, this equation takes a particularly simple form,

$$\begin{aligned}\frac{dT'}{dT} \approx \frac{r}{r-1}\frac{T'^2}{T^2} - \frac{3T'^2}{2(r-1)m_\chi T} \\ - \frac{rT'^2}{8(r-1)^2 H m_\chi T}\langle\sigma v^2\rangle \left[n_\chi^2 - \frac{n_{\chi,0}^2}{n_{A',0}} n_{A'}\right] \\ + \frac{n_f T'^2}{(r-1)^2 HT m_\chi^2}\langle\sigma v \delta E\rangle_{\chi f \to \chi f} \,.\end{aligned} \tag{5.82}$$

We note that the second term on the right-hand side is typically smaller than the term before it since $T \ll m_\chi$, but has been included to improve the accuracy of this

analytic result. In terms of x' and x, we have

$$\frac{dx'}{dx} \approx \frac{r}{r-1} - \frac{3}{2(r-1)x} - \frac{r}{8(r-1)^2 Hx}\langle\sigma v^2\rangle \left[n_\chi^2 - \frac{n_{\chi,0}^2}{n_{A',0}} n_{A'}\right] + \frac{n_f}{(r-1)^2 m_\chi Hx}\langle\sigma v \delta E\rangle_{\chi f\to\chi f}\,. \quad (5.83)$$

The relative importance of each term on the right-hand side of Eq. (5.82), which governs the temperature evolution after $3 \leftrightarrow 2$ freezeout, separates Regimes I–III. Since the $3 \leftrightarrow 2$ term is typically less than $\mathcal{O}(1)$, the different regimes are distinguished by how large the elastic scattering term is compared to $r/(r-1)$. In Regime I, throughout the period between $3 \leftrightarrow 2$ freezeout and $2 \leftrightarrow 2$ freezeout, we have

$$n_f\langle\sigma v\delta E\rangle_{\chi f\to\chi f} \gg r(r-1)\frac{Hm_\chi^2}{T^2}(T'-T) \quad \text{(Regime I)}\,, \quad (5.84)$$

keeping in mind that $n_f\langle\sigma v\delta E\rangle_{\chi f\to\chi f} \propto (T'-T)$ (see Eq. (B.16) for an expression for $\langle\sigma v\delta E\rangle_{\chi f\to\chi f}$). The fast elastic scattering enforces $T' \simeq T$, the assumption of the "classic forbidden" regime. In Regime II, we have instead

$$n_f\langle\sigma v\delta E\rangle_{\chi f\to\chi f} \sim r(r-1)\frac{Hm_\chi^2}{T^2}(T'-T) \quad \text{(Regime II)} \quad (5.85)$$

at some point between the two dark sector freezeout events. In this regime, since $n_f\langle\sigma v\delta E\rangle_{\chi f\to\chi f} \propto (T'-T)$, the dark sector begins to cool immediately after $3 \to 2$ freezeout, but once T' starts differing significantly from T, the elastic scattering term becomes large enough to slow the cooling process.

Finally, in Regime III, between the $3 \leftrightarrow 2$ and $2 \leftrightarrow 2$ freezeout events, we always have

$$n_f\langle\sigma v\delta E\rangle_{\chi f\to\chi f} \ll r(r-1)\frac{Hm_\chi^2}{T^2}(T'-T) \text{ (Regime III)}\,, \quad (5.86)$$

This is the limit where the elastic scattering process is too inefficient to transfer heat

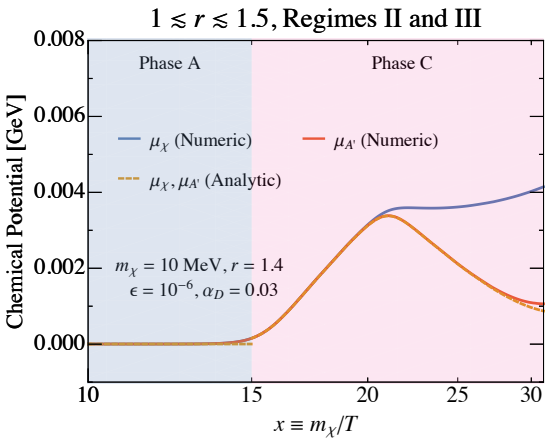

Figure 5-8: Chemical potential in the dark sector in Regimes II and III. Here, we choose a set of benchmark parameters ($m_\chi = 10\,\text{MeV}$, $r = 1.4$, $\epsilon = 10^{-6}$, $\alpha_D = 0.03$) within Regime II, but a similar result is obtained in Regime III as well. The numerically computed chemical potentials of χ (blue line) and A' (red line) are shown, together with the analytic result from Eq. (5.79) (orange dashed line).

between the two sectors, and therefore the dark sector cooling is rapid and becomes independent of ϵ.

5.5.2 Regime Boundaries and Characteristics

We will now describe some general characteristics of each regime, providing where we can an analytic description of the dark sector freezeout process. We also explain how to numerically estimate the value of ϵ on the α_D–ϵ plane at which the boundary between the regimes is located.

Regime I

For $\epsilon \gtrsim 10^{-4}$, freezeout of the dark sector is controlled by $\chi\overline{\chi} \leftrightarrow f\overline{f}$, corresponding to the conventional WIMP regime. For values of ϵ smaller than this, we enter regime I, the "classic forbidden" regime, with $T' \approx T$ until the final freezeout of the dark sector. This regime was studied in Ref. [119], where they showed that the dark sector freezeout is determined entirely by when the $2 \leftrightarrow 2$ freezeout occurs, a purely dark sector process which is independent of ϵ.

The "classic forbidden"-WIMP boundary occurs when the $2 \leftrightarrow 2$ dark sector pro-

cess freezes out and approximately the same time as $\chi\overline{\chi} \to f\overline{f}$, i.e.

$$n_\chi \langle \sigma v \rangle_{\chi\overline{\chi}\to A'A'} \approx H \approx n_\chi \langle \sigma v \rangle_{\chi\overline{\chi}\to f\overline{f}} \quad \text{(WIMP/I)}\,. \tag{5.87}$$

If we further require the freezeout to produce the observed relic abundance and fulfil Eq. (5.10), we obtain the following analytic estimate for $\epsilon_{\text{WIMP/I}}$, the value of ϵ at the WIMP/Regime I boundary, and specializing to $r = 1.4$ for illustration:

$$\epsilon_{\text{WIMP/I}} \sim 4 \times 10^{-4} \left(\frac{\alpha_D}{1.0}\right)^{3/14} \left(\frac{g_*(x_f)}{10.75}\right)^{1/14} \left(\frac{20}{x_f}\right)^{3/7} \left(\frac{g_{*,s}(x_f)}{10.75}\right)^{1/7}, \tag{5.88}$$

where $x_f \sim 20$ gives the temperature of freezeout of both the $2 \leftrightarrow 2$ and the $\chi\overline{\chi} \leftrightarrow f\overline{f}$ processes.

At the low-ϵ end of Regime I, the elastic scattering energy transfer rate becomes gradually small enough such that Eq. (5.84) is no longer satisfied at all points between $3 \leftrightarrow 2$ freezeout and $2 \leftrightarrow 2$ freezeout, and the dark sector transitions into Regime II. The boundary between Regimes I and II is therefore marked by when the elastic scattering condition for Regime II, Eq. (5.85), becomes fulfilled just as $2 \to 2$ freezeout occurs, i.e.

$$\begin{aligned} n_f \langle \sigma v \delta E \rangle_{\chi f\to\chi f} &\approx r(r-1)\frac{H(T_2)m_\chi^2}{T_2^2}(T_2' - T_2)\,, \\ n_\chi \langle \sigma v \rangle_{\chi\overline{\chi}\to A'A'} &\approx H(T_2) \qquad\qquad \text{(I/II)}\,, \end{aligned} \tag{5.89}$$

where T_2 and T_2' are the SM and dark sector temperatures at $2 \to 2$ freezeout. Note that both $\langle \sigma v \delta E \rangle_{\chi f\to\chi f}$ and $\langle \sigma v \rangle_{\chi\overline{\chi}\to A'A'}$ depend on T and T'. Together with Eq. (5.10) for the relic abundance, we can obtain a numerical estimate for $\epsilon_{\text{I/II}}$, the value of ϵ as a function of α_D at the boundary between Regimes I and II.

Regime II

Regime II is characterized by Eq. (5.85) between $3 \leftrightarrow 2$ and $2 \leftrightarrow 2$ freezeout, which ensures that $T' < T$ due to the $\chi\overline{\chi} \to A'A'$ process, but with some heat being

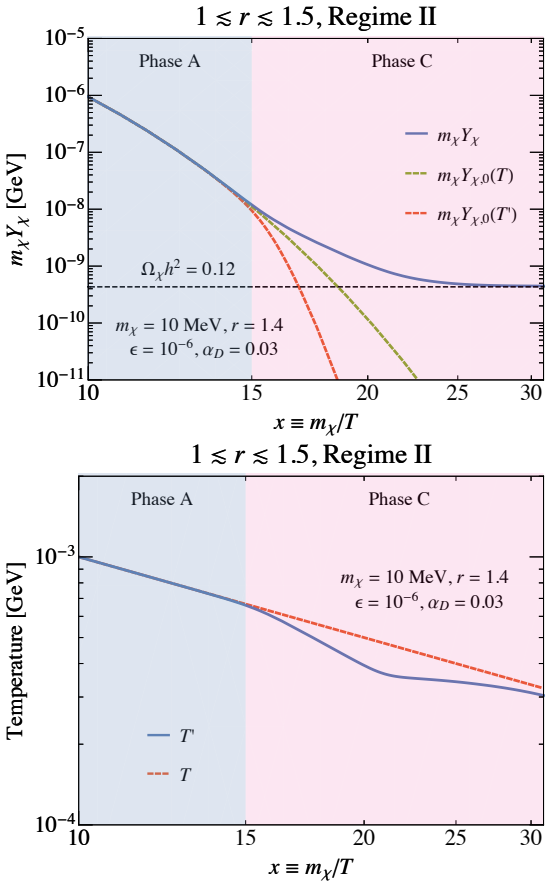

Figure 5-9: Dark sector evolution in Regime II for $1 \lesssim r \lesssim 1.5$, with parameters $m_\chi = 10\,\mathrm{MeV}$, $r = 1.4$, $\epsilon = 10^{-6}$ and $\alpha_D = 0.03$. In both plots, thermodynamic phases A and C as defined in Sec. 5.3.4 are shown in light blue and pink respectively. (*Left*) χ abundance (given as $m_\chi Y_\chi$) as a function of x (blue line), with the zero chemical potential abundance at the SM temperature $m_\chi Y_{\chi,0}(T)$ (green dashed line) and the dark sector temperature $m_\chi Y_{\chi,0}(T')$ (red dashed line) shown for reference. The observed DM abundance is indicated by the horizontal black dashed line, as defined in Eq. (5.10). (*Right*) The dark sector temperature T' (blue line), as a function of the SM temperature (red dashed line). The A' abundance evolves trivially as $n_{A'} = n_{A',0}(T)$ in this regime.

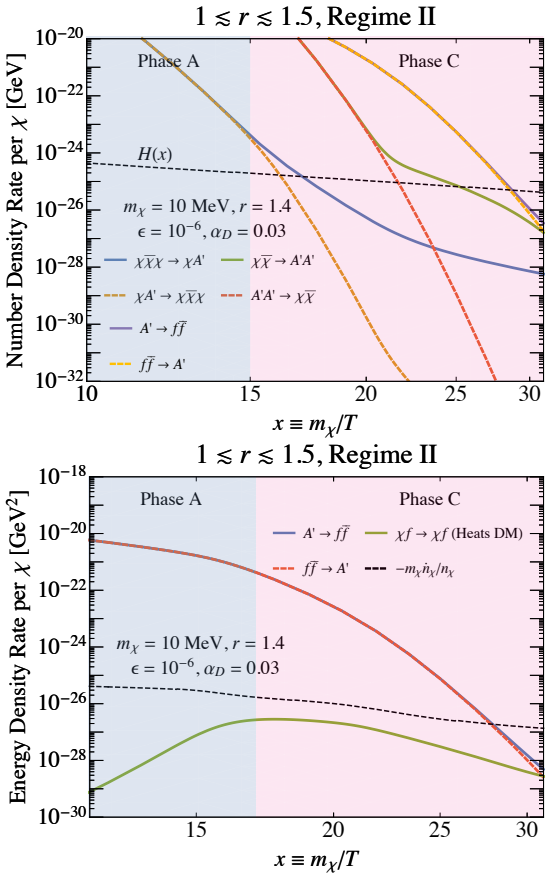

Figure 5-10: Rates of change in number density and energy density of the dark sector in Regime II for $1 \lesssim r \lesssim 1.5$, with parameters $m_\chi = 10\,\text{MeV}$, $r = 1.4$, $\epsilon = 10^{-6}$ and $\alpha_D = 0.03$. In both plots, thermodynamic phases A and C as defined in Sec. 5.3.4 are shown in light blue and pink respectively. (*Left*) χ Number density rates for $\chi\overline{\chi}\chi \to \chi A'$ (blue line), $\chi A' \to \chi\overline{\chi}\chi$ (dark orange dashed line), $\chi\overline{\chi} \to A'A'$ (green line), $A'A' \to \chi\overline{\chi}$ (red dashed line), $A' \to f\overline{f}$ (purple line) and $f\overline{f} \to A'$ (yellow dashed line) are shown. The Hubble rate is shown as a black dashed line. (*Right*) Energy density rates for $A' \to f\overline{f}$ (blue line), $f\overline{f} \to A'$ (red dashed line) and $\chi f \leftrightarrow \chi f$ (green line), which has the net effect of heating the dark sector. The rate at which the energy density of DM is changing $-m_\chi \dot{n}_\chi/n_\chi$ (black dashed line) is shown for reference.

transferred from the SM to the dark sector to impede the cooling of the dark sector due to $\chi f \to \chi f$. At the same time, the decay rate Γ is large enough such that $n_{A'} \approx n_{A',0}(T)$ throughout the freezeout of the dark sector. This condition immediately determines the chemical potentials μ_χ, given analytically by the expression Eq. (5.79), as well as $\mu_{A'} \approx \mu_\chi$. In Fig. 5-8, we show this analytic result in comparison with the numeric calculation of the chemical potential, for our Regime II benchmark point of $m_\chi = 10\,\mathrm{MeV}$, $r = 1.4$, $\epsilon = 10^{-6}$ and $\alpha_D = 0.03$. Note that the agreement deteriorates rapidly once $2 \leftrightarrow 2$ freezeout occurs at $x \sim 21$, after which the DM particle has completely frozen out, and the assumption that $\mu_\chi \approx \mu_{A'}$ breaks. A similar result is obtained in Regime III as well, where $n_{A'} \approx n_{A',0}(T)$ also holds.

Fig. 5-9 shows the evolution of the χ number density and T' at the same benchmark parameters. $n_{A'}$ evolves trivially as $n_{A',0}(T)$, and therefore need not be separately plotted. Since a chemical potential develops immediately after the dark sector kinetically decouples from the SM at the point of $3 \leftrightarrow 2$ freezeout at $x \sim 15$, the dark sector passes from thermodynamic phase A to C directly. The characteristic cooling of the dark sector is apparent in the right panel of Fig. 5-9, and is governed by Eq. (5.83). In this regime, this differential equation does not appear to be analytically integrable; we show only the numerical result, obtained directly from the full Boltzmann equations.

In Fig. 5-10, we show the number density and energy density rates of all relevant dark sector processes. The transition between phases A and C occurs at roughly $x \sim 15$, when the backward and forward $3 \leftrightarrow 2$ rates cease to be approximately equal. This occurs when the $3 \to 2$ rate is still much larger than the Hubble rate, due to the relation between the rates of the $3 \leftrightarrow 2$ and $2 \leftrightarrow 2$ processes enforced by Eq. (5.22), where the $2 \leftrightarrow 2$ rate being of order Hn_χ allows the $3 \leftrightarrow 2$ total rate to be much larger than the Hubble rate. Once the dark sector transitions into phase C, we see that the elastic scattering energy density rate per χ particle becomes just a factor of a few smaller than $-m_\chi \dot{n}_\chi / n_\chi$, meeting the Regime II criterion laid out in Eq. (5.85). This shows that a significant amount of heat is transferred from the SM to the dark sector, slowing the cooling rate compared to what happens in Regime III, which we will discuss next.

Within Regime II, as ϵ decreases still further, Eq. (5.85) is met increasingly earlier, leading to a colder dark sector due to the diminishing ability of $\chi f \to \chi f$ to heat the dark sector. Eventually, the condition Eq. (5.85) is only met at the point of $3 \to 2$ freezeout, and no significant amount of heat is transferred to the dark sector after that. This marks the boundary between Regime II and III, i.e.

$$\begin{aligned} n_f \langle \sigma v \delta E \rangle_{\chi f \to \chi f} &\approx r(r-1) \frac{H(T_3) m_\chi^2}{T_3^2} (T_3' - T_3) \,, \\ \frac{1}{4} n_\chi^2 \langle \sigma v^2 \rangle &\approx H(T_3) \qquad\qquad \text{(II/III)} \,, \end{aligned} \tag{5.90}$$

where T_3 and T_3' are the SM and dark sector temperatures at $3 \leftrightarrow 2$ freezeout respectively. An analytic estimate for $\epsilon_{\text{II/III}}$, the value of ϵ when these two conditions are satisfied, is

$$\begin{aligned} \epsilon_{\text{II/III}} \sim 8 \times 10^{-7} \left(\frac{1.0}{\alpha_D}\right) \left(\frac{g_*(x_f)}{10.75}\right)^{1/4} \\ \times \left(\frac{r}{1.4}\right)^{5/2} \left(\frac{r-1}{0.4}\right)^{1/2} \left(\frac{m_\chi}{\text{GeV}}\right)^{1/2} \left(\frac{x_f}{20}\right)^3 . \end{aligned} \tag{5.91}$$

Once again, combining the boundary conditions shown above with the observed relic abundance in Eq. (5.10) allows us to eliminate m_χ and x_f from the expression above numerically. This numerical expression for $\epsilon_{\text{II/III}}$ forms the boundary between Regimes II and III.

Regime III

Fig. 5-11 shows the evolution of the χ-abundance and the dark sector temperature in Regime III, for our benchmark parameters in this regime, $m_\chi = 10\,\text{MeV}$, $r = 1.4$, $\epsilon = 3 \times 10^{-8}$ and $\alpha_D = 0.13$. In Fig. 5-12, we show the number density and energy density rates per χ particle through the dark sector freezeout. In this regime, the dark sector temperature once again cools rapidly after $3 \to 2$ freezeout and enters thermodynamic phase C; unlike Regime II, however, elastic scattering plays no significant role in influencing this evolution between $3 \leftrightarrow 2$ freezeout and $2 \leftrightarrow 2$

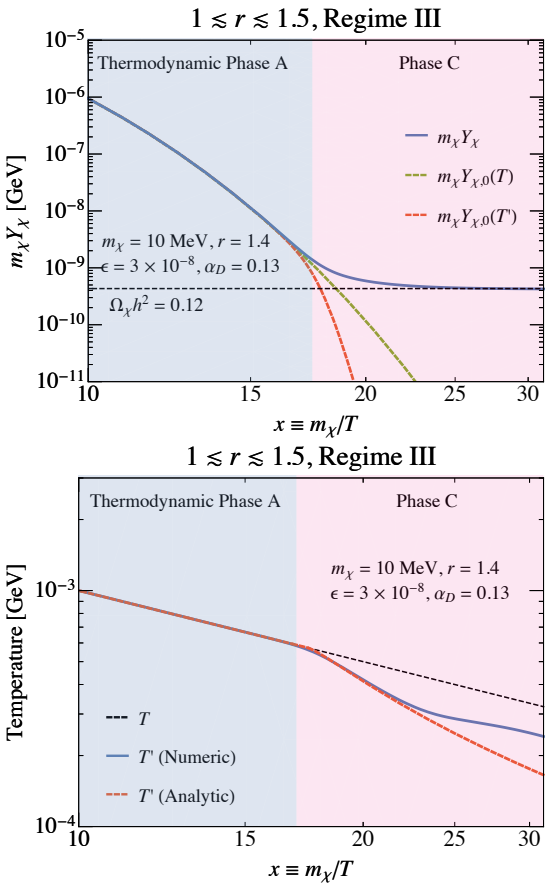

Figure 5-11: Dark sector evolution in Regime III for $1 \lesssim r \lesssim 1.5$, with parameters $m_\chi = 10\,\text{MeV}$, $r = 1.4$, $\epsilon = 3 \times 10^{-8}$ and $\alpha_D = 0.13$. In both plots, thermodynamic phases A and C as defined in Sec. 5.3.4 are shown in light blue and pink respectively. (*Left*) χ abundance (given as $m_\chi Y_\chi$) as a function of x (blue line), with the zero chemical potential abundance at the SM temperature $m_\chi Y_{\chi,0}(T)$ (green dashed line) and the dark sector temperature $m_\chi Y_{\chi,0}(T')$ (red dashed line) shown for reference. The observed DM abundance ($\Omega_\chi h^2 = 0.12$) is indicated by the horizontal black dashed line, as defined in Eq. (5.10). (*Right*) The dark sector temperature T' (blue line), as a function of the SM temperature (black dashed line). An analytic estimate for T', given in Eq. (5.93), is shown by the red dashed line. The A' abundance evolves trivially as $n_{A'} = n_{A',0}(T)$ in this regime.

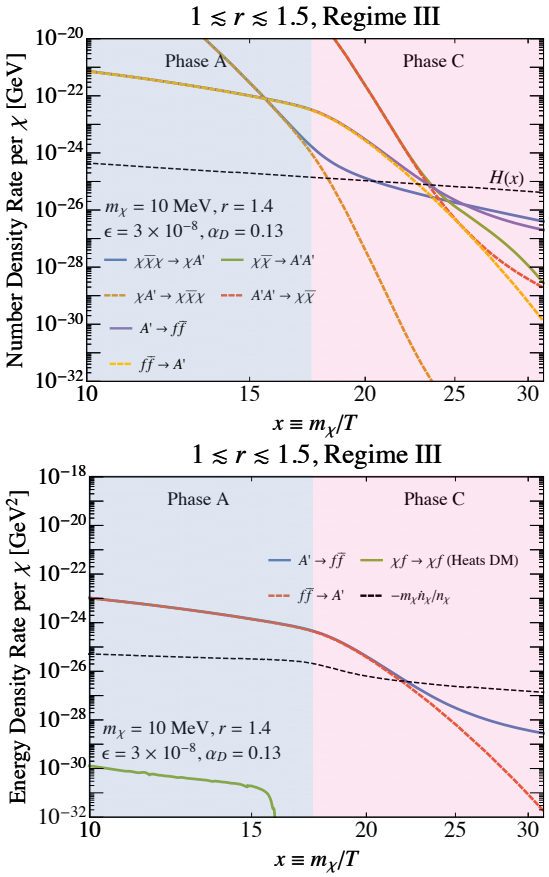

Figure 5-12: Rates of change in number density and energy density of the dark sector in Regime III for $1 \lesssim r \lesssim 1.5$, with parameters $m_\chi = 10\,\text{MeV}$, $r = 1.4$, $\epsilon = 3 \times 10^{-8}$ and $\alpha_D = 0.13$. In both plots, thermodynamic phases A and C as defined in Sec. 5.3.4 are shown in light blue and pink respectively. (*Left*) χ Number density rates for $\chi\overline{\chi}\chi \to \chi A'$ (blue line), $\chi A' \to \chi\overline{\chi}\chi$ (dark orange dashed line), $\chi\overline{\chi} \to A'A'$ (green line), $A'A' \to \chi\overline{\chi}$ (red dashed line), $A' \to f\overline{f}$ (purple line) and $f\overline{f} \to A'$ (yellow dashed line) are shown. The Hubble rate is shown as a black dashed line. (*Right*) Energy density rates for $A' \to f\overline{f}$ (blue line), $f\overline{f} \to A'$ (red dashed line) and $\chi f \to \chi f$ (green line), which has the net effect of heating the dark sector. The rate at which the energy density of DM is changing $-m_\chi \dot{n}_\chi/n_\chi$ (black dashed line) is shown for reference.

freezeout, as can be seen in the right panel of Fig. 5-12. The dark sector temperature evolution after $3 \leftrightarrow 2$ freezeout can be obtained by setting $\langle \sigma v \delta E \rangle \to 0$ in Eq. (5.83) and neglecting the $2 \to 3$ rate (which is much smaller than the forward rate after $3 \to 2$ freezeout), i.e.

$$\frac{dx'}{dx} \approx \frac{r}{r-1} - \frac{3}{2(r-1)x} - \frac{r}{8(r-1)^2 H x} \langle \sigma v^2 \rangle n_\chi^2 \,. \tag{5.92}$$

Given the approximation for the chemical potential μ_χ in Eq. (5.79), this differential equation can be integrated exactly, starting from $x_3 = x_3'$, to give

$$\begin{aligned} x' \approx x_3 + \frac{r}{r-1}(x - x_3) - \frac{3}{2(r-1)} \log\left(\frac{x}{x_3}\right) \\ - \frac{1}{2(r-1)} \log\left[1 + \frac{r m_\chi^6 \langle \sigma v^2 \rangle e^{-2x_3}}{8\pi^3 (r-1) H(T_3) x_3^3} \left(1 - \frac{x_3^4}{x^4}\right)\right] . \end{aligned} \tag{5.93}$$

In the right panel of Fig. 5-11, we show this analytic result in comparison with the numeric result obtained from the full Boltzmann equation, and find excellent agreement between them, up to $2 \leftrightarrow 2$ freezeout at $x \sim 23$.

Throughout Regime III, $n_{A'} \approx n_{A',0}(T)$ due to the highly efficient $A' \leftrightarrow f\overline{f}$ process; as ϵ decreases, however, $A' \leftrightarrow f\overline{f}$ becomes less and less rapid, and eventually this process becomes too inefficient to keep the dark sector in thermal equilibrium at the point of $3 \leftrightarrow 2$ freezeout. Below this point, kinetic decoupling between the two sectors occurs before either of the dark sector processes freezes out, leading to the KINDER-like Regime IV. We can estimate the boundary between Regimes III and IV by requiring the $3 \to 2$ freezeout and kinetic decoupling to occur at the same time, i.e.

$$\begin{aligned} \frac{r m_\chi \Gamma}{n_{\chi,0}(T_3)} n_{A',0}(T_3) &\approx \frac{m_\chi \dot{n}_{\chi,0}(T_3)}{n_{\chi,0}(T_3)} \,, \\ \frac{1}{4} n_\chi^2 \langle \sigma v^2 \rangle &\approx H(T_3) \qquad\qquad \text{(III/IV)} \,. \end{aligned} \tag{5.94}$$

These conditions are however identical to the conditions used for estimating the boundary between the KINDER and the NFDM regime for $1.5 \lesssim r \lesssim 2$ in Eq. (5.41).

This equation can be restated as

$$\epsilon_{\rm III/IV} \sim 10^{-8} e^{9.6(r-1.4)} \left(\frac{\alpha_D}{1.0}\right)^{\frac{3(r-1)}{4}} \times \left(\frac{1.4}{r}\right)^{9/4} \left(\frac{x_f}{20}\right)^{-\frac{r+3}{4}} \left(\frac{g_*(x_f)}{10.75}\right)^{-\frac{r-3}{8}} \times \left(\frac{\rm GeV}{m_\chi}\right)^{\frac{r-3}{4}} \left(\frac{f(r)}{23.7}\right)^{\frac{r-1}{4}} , \quad (5.95)$$

where $\epsilon_{\rm III/IV}$ is the value of ϵ between Regimes III and IV as a function of various model parameters. Finally, we may once again combine Eq. (5.95) with the condition for the observed relic abundance in Eq. (5.10) to numerically derive the boundary between these regimes.

Regime IV

Fig. 5-13 shows the evolution of the χ-abundance and the dark sector temperature in Regime IV, for our benchmark parameters in this regime, $m_\chi = 10\,\text{MeV}$, $r = 1.4$, $\epsilon = 2 \times 10^{-9}$ and $\alpha_D = 0.6$. In Fig. 5-14, we show the number density and energy density rates per χ particle throughout dark sector freezeout. In Regime IV, kinetic decoupling occurs before either of the $2 \leftrightarrow 2$ or $3 \leftrightarrow 2$ processes become slow. This regime is similar to the KINDER regime with $1.5 \lesssim r \lesssim 2$, exhibiting heating in the dark sector, with the key difference being that the $3 \to 2$ process is now slower than the $2 \to 2$ process. In thermodynamic phase A and B, the physics in this regime is identical to that of the KINDER regime with $1.5 \lesssim r \lesssim 2$, as discussed in Sec. 5.4.2. Kinetic decoupling occurs first at a temperature given approximately by Eq. (5.42), after which the dark sector enters phase B. An approximation for the evolution of T' can be obtained by assuming dark sector entropy conservation, leading to

$$x' \simeq x_d + 3\log\left(\frac{x}{x_d}\right) , \quad (5.96)$$

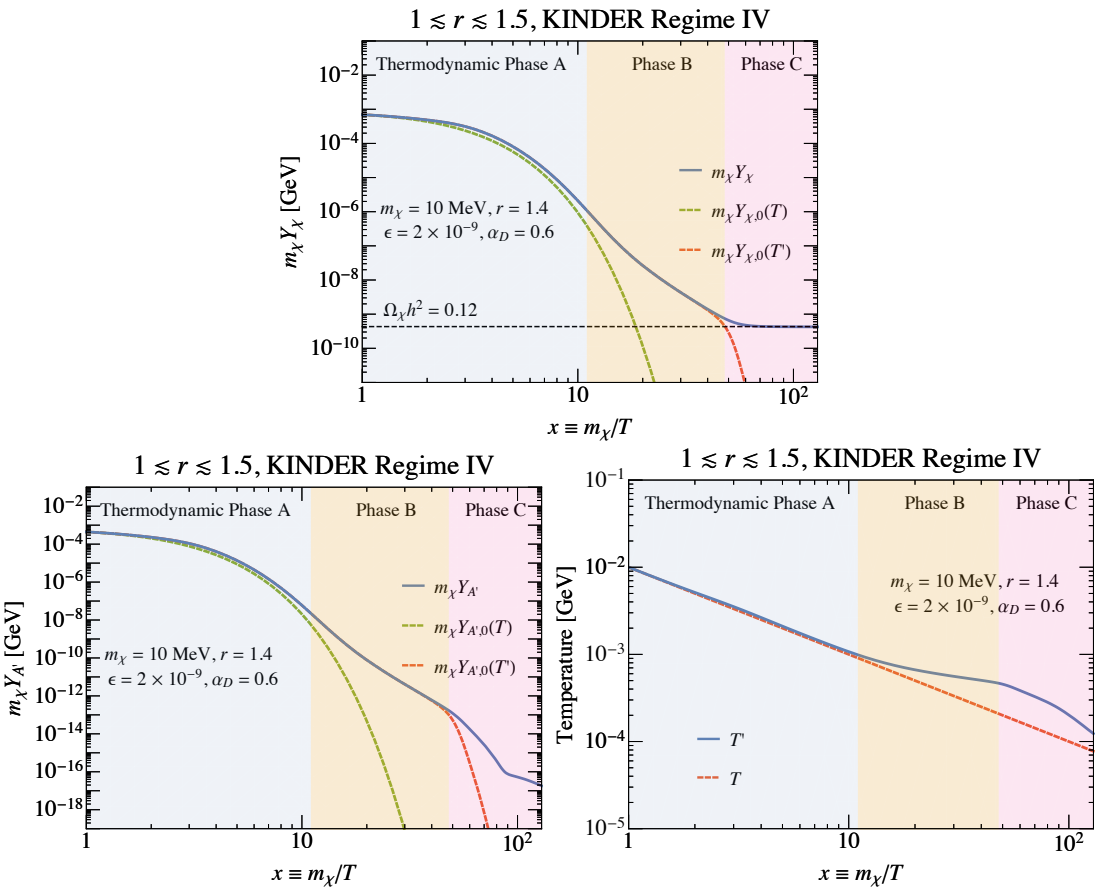

Figure 5-13: Dark sector evolution in regime IV for $1 \lesssim r \lesssim 1.5$, with parameters $m_\chi = 10\,\text{MeV}$, $r = 1.4$, $\epsilon = 2 \times 10^{-9}$ and $\alpha_D = 0.6$. In all three plots, thermodynamic phases A, B and C as defined in Sec. 5.3.4 are shown in light blue, yellow and pink respectively. (*Top left*) χ abundance (given as $m_\chi Y_\chi$) as a function of x (blue line), with the zero chemical potential abundance at the SM temperature $m_\chi Y_{\chi,0}(T)$ (green dashed line) and the dark sector temperature $m_\chi Y_{\chi,0}(T')$ (red dashed line) shown for reference. The observed DM abundance is indicated by the horizontal black dashed line, as defined in Eq. (5.10). (*Top right*) A' abundance (given as $m_\chi Y_{A'}$) as a function of x (blue line), with $Y_{A',0}(T)$ (green dashed line) and $Y_{A',0}(T')$ (red dashed line) once again given for reference. (*Bottom*) The dark sector temperature T' (blue line), as a function of the SM temperature (red dashed line).

while a more detailed examination of the Boltzmann equations leads to the improved approximation in Eq. (5.50), i.e.

$$x' \approx x_d + \frac{1}{r-1} \log \left[1 + \frac{3}{8} \frac{\Gamma r^{5/2} (r-1) e^{(1-r)x_d}}{x_d^2 H(x_d)} (x^2 - x_d^2) \right] . \tag{5.97}$$

Once the $3 \leftrightarrow 2$ process freezes out, the dark sector enters thermodynamic phase

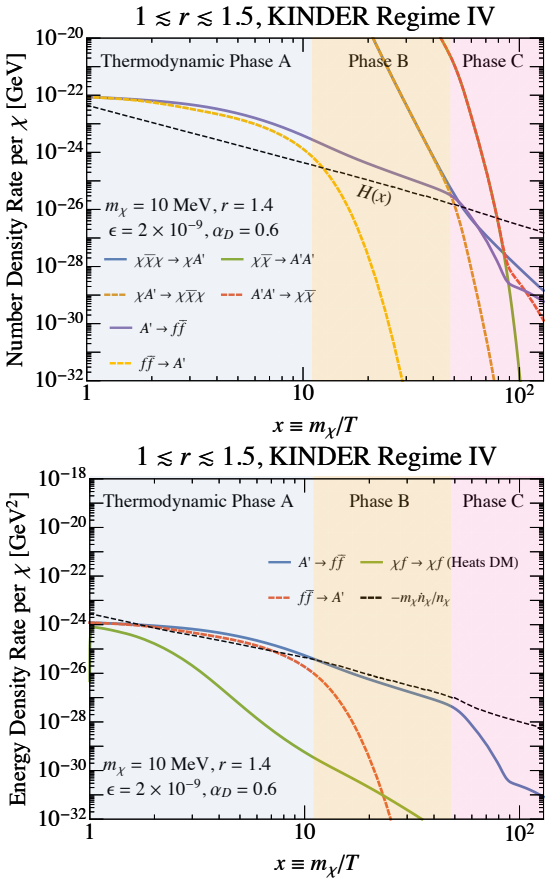

Figure 5-14: Rates of change in number density and energy density per χ particle of the dark sector in regime IV for $1 \lesssim r \lesssim 1.5$; the model parameters are $m_\chi = 10\,\text{MeV}$, $r = 1.4$, $\epsilon = 2 \times 10^{-9}$ and $\alpha_D = 0.6$. In both plots, thermodynamic phases A, B and C as defined in Sec. 5.3.4 are shown in light blue, yellow and pink respectively. (*Left*) Number density rates for $\chi\overline{\chi}\chi \to \chi A'$ (blue line), $\chi A' \to \chi\overline{\chi}\chi$ (dark orange dashed line), $\chi\overline{\chi} \to A'A'$ (green line), $A'A' \to \chi\overline{\chi}$ (red dashed line), $A' \to f\overline{f}$ (purple line) and $f\overline{f} \to A'$ (yellow dashed line) are shown. Also shown are the rates for $A' \to f\overline{f}$ (purple line) and $f\overline{f} \to A'$ (dashed yellow line). The Hubble parameter is shown in the black dashed line. (*Right*) Energy density rates for $A' \to f\overline{f}$ (blue line), $f\overline{f} \to A'$ (red dashed line) and $\chi f \leftrightarrow \chi f$ (green line), which has the net effect of heating the dark sector. The rate at which the energy density of dark matter is changing $-m_\chi \dot{n}_\chi / n_\chi$ (black dashed line) is shown for reference.

C. As before, the χ number density evolution is given by Eq. (5.21), i.e.

$$\dot{n}_\chi + 3Hn_\chi \simeq \frac{r}{8(1-r)}\langle\sigma v^2\rangle\left[n_\chi^3 - \frac{n_{\chi,0}^2}{n_{A',0}} n_\chi n_{A'}\right]. \tag{5.98}$$

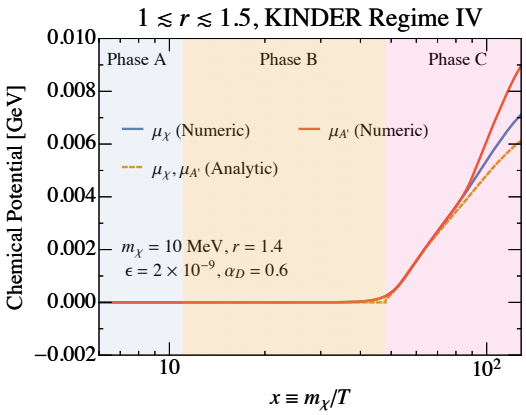

Figure 5-15: $1 \lesssim r \lesssim 1.5$, KINDER Regime IV (benchmark values $m_\chi = 10\,\text{MeV}$, $r = 1.4$, $\epsilon = 2 \times 10^{-9}$ and $\alpha_D = 0.6$) comparison between the analytic estimate of $\mu_\chi \approx \mu_{A'}$ (orange dashed line) given in Eq. (5.101), and the numerical computation of μ_χ (blue line) and $\mu_{A'}$ (red line) based on integrating the full Boltzmann equations.

For $1 \lesssim r \lesssim 1.5$, however, the $3 \to 2$ process is slow in phase C, meaning that

$$\dot{n}_\chi + 3Hn_\chi \simeq 0\,, \tag{5.99}$$

i.e. $n_\chi \propto a^{-3}$ in phase C, with χ frozen out.

More accurately, Eqs. (5.54) and (5.55) are still true in this regime, since $n_{A'} \ll n_\chi$; we therefore still have the following approximate relation after $3 \leftrightarrow 2$ freezeout occurs:

$$\frac{r}{8(r-1)}\langle\sigma v^2\rangle n_\chi^3 \approx r\Gamma n_{A'}\,, \tag{5.100}$$

where we have neglected the $2 \to 3$ rate since the $3 \leftrightarrow 2$ freezeout has occurred. This approximate relation gives us an expression for $\mu_\chi \simeq \mu_{A'}$:

$$\frac{\mu_\chi}{T'} \approx \frac{3-r}{2}x' + \frac{3}{2}\log x' + \frac{1}{2}\log\left[\frac{8(r-1)\Gamma}{\langle\sigma v^2\rangle}\frac{3(2\pi)^3 r^{3/2}}{64 m_\chi^6}\right]. \tag{5.101}$$

A comparison between this analytic approximation and the numerical result in phase C is shown in Fig. 5-15, demonstrating good agreement up till the point of $2 \leftrightarrow 2$ freezeout.

We can substitute our analytic expression for μ_χ/T' into Eq. (5.98) using the

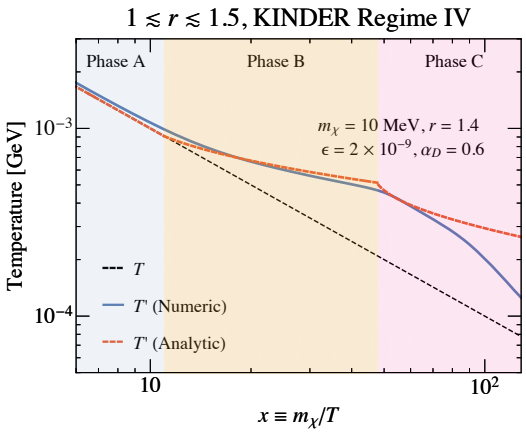

Figure 5-16: $1 \lesssim r \lesssim 1.5$, KINDER Regime IV (benchmark values $m_\chi = 10\,\text{MeV}$, $r = 1.4$, $\epsilon = 2 \times 10^{-9}$ and $\alpha_D = 0.6$) comparison between the analytic estimate of T' (red dashed line) given in Eq. (5.103), and the numerical computation of T' (blue line) based on integrating the full Boltzmann equations. The SM temperature is shown for reference (black dashed line).

expression for $\dot{n}_\chi$ in Eq. (5.8), giving

$$\frac{1-r}{2}\frac{m_\chi}{T'^2}\frac{dT'}{dT} \approx -\frac{3}{T} - \frac{r\Gamma}{H(T_3)}\frac{T_3^2}{T^3}\frac{3}{4}r^{3/2}e^{(1-r)m_\chi/T'}\,, \tag{5.102}$$

where T_3 is the temperature at $3 \leftrightarrow 2$ freezeout. This expression can be integrated exactly to give

$$x' \approx x'_3 + \frac{6}{r-1}\log\left(\frac{x}{x_3}\right) + \frac{1}{r-1}\log\left[1 + \frac{\Gamma}{H(T_3)}\frac{3}{8}r^{5/2}\left(1 - \frac{x_3^4}{x^4}\right)e^{(1-r)x'_3}\right]. \tag{5.103}$$

This analytic prediction in comparison with the numerical temperature evolution is shown in Fig. 5-16, showing good agreement until near the $2 \leftrightarrow 2$ freezeout, when μ_χ and $\mu_{A'}$ begin to diverge.

Summary of regimes and boundaries for $1 \lesssim r \lesssim 1.5$

Fig. 5-17 shows contours for fixed values of m_χ in the α_D–ϵ parameter space for which the observed relic abundance of $\Omega_\chi h^2 = 0.12$ is attained. We show the same set of contours for $r = 1.3$ and $r = 1.4$ as two representative values of r in the case of

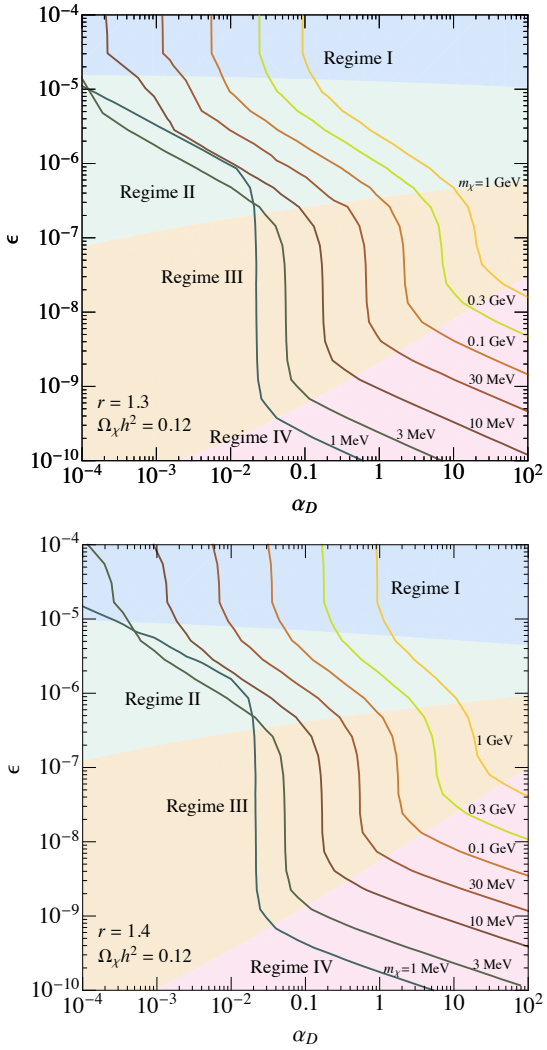

Figure 5-17: Contours of the observed relic abundance ($\Omega_\chi h^2 = 0.12$) in the α_D–ϵ plane for $1 \lesssim r \lesssim 1.5$, for *(left)* $r = 1.3$ and *(right)* $r = 1.4$, for various values of m_χ. The "classical forbidden" Regime I (blue), Regime II (green), Regime III (orange) and KINDER Regime IV (pink) are indicated, with the boundaries obtained using Eq. (5.89) between I/II, Eq. (5.90) between II/III and Eq. (5.94) between III/IV, all in conjunction with the relic abundance condition given in Eq. (5.10).

$1 \lesssim r \lesssim 1.5$. The four regimes can be made out by changes in behavior of the contour lines. Note that the boundary between the WIMP regime and Regime I occurs at ϵ values above the maximum ϵ shown in Fig. 5-17.

In Regime I, the relic abundance is controlled entirely by the $2 \leftrightarrow 2$ freezeout, which only depends on α_D, leading to vertical contours in the α_D–ϵ plane. Decreasing ϵ into Regime II, the relic abundance is controlled by when the freezeout of $3 \leftrightarrow 2$ and of $2 \leftrightarrow 2$ occur, as well as how efficiently $\chi f \leftrightarrow \chi f$ heats the dark sector and impedes the cooling due to $\chi\overline{\chi} \to A'A'$, leading to some nontrivial dependence on ϵ and α_D. Once we arrive at Regime III however, elastic scattering becomes extremely inefficient, and the rate of dark sector cooling after $3 \leftrightarrow 2$ freezeout depends only on the $3 \to 2$ rate itself. Since all of the physically important processes are purely dark sector processes, the contours are once again independent of ϵ. Finally, in Regime IV, the relic abundance is determined by when kinetic decoupling occurs, but also by the long power-law decrease in n_χ in phase B, which is dictated by dark-sector-only processes. This once again leads to contours that depend on both α_D and ϵ.

We note that the contour of $\Omega_\chi h^2 = 0.12$ for $m_\chi \lesssim 5\,\mathrm{MeV}$ shows an abrupt change in behavior in Regime II compared to higher DM masses. This occurs due to the fact that in Regime II thermodynamic phase C, DM particles with masses below $\sim 5\,\mathrm{MeV}$ undergo elastic scattering with nonrelativistic, rather than relativistic, electrons throughout most of the freezeout process. The Boltzmann suppression of nonrelativistic electrons leads to a sharp decrease in $\langle\sigma v \delta E\rangle$, which controls the cooling rate of the dark sector in this phase, and hence the relic abundance of DM. The correct relic abundance is thus achieved at a higher value of ϵ than expected, in order for the stronger mixing to compensate for the decrease in electron number density. We refer the reader to App. B.2 for more details on how $\langle\sigma v \delta E\rangle_{\chi f \to \chi f}$ is computed.

5.6 Experimental Probes and Constraints

There are significant constraints on dark photons from both terrestrial experiments and supernova observations. There are also cosmological constraints on the DM itself, from DM annihilation to electrons and positrons affecting the anisotropies of the CMB, and from modifications to the number of effective degrees of freedom during Big Bang nucleosynthesis (BBN) and the CMB epoch. DM self-interactions mediated by the dark photon exchange can be large, and can be probed by observations of galactic structure. Finally, a sufficiently warm dark sector can be constrained by measurements of the matter power spectrum. We will discuss these constraints in this section, and plot the results in Fig. 5-18.

5.6.1 Accelerator and Direct-Detection Experiments

For $1 \lesssim r \lesssim 2$ with a dark photon mass above 1 MeV, dark photons produced at beam experiments decay visibly into SM particles. The observational signatures of visibly decaying dark photons have been studied extensively in the literature [101–110]. In 5-18, we plot the region of parameter space excluded by these experiments. This excluded region covers considerable parameter space, extending down to $\epsilon \sim 10^{-7} - 10^{-8}$ for $m_\chi \lesssim 100\,\mathrm{MeV}$.

Direct-detection experiments can probe the scattering of the DM on both electrons and nucleons (including the Migdal effect [125, 126]) via dark photon exchange. In 5-18, we consider the constraints from DarkSide, Xenon 1T, SuperCDMS, and SENSEI [125–130]. In the parameter space we consider, nuclear scattering limits derived by exploiting the Migdal effect set the strongest bound. These limits are primarily sensitive to the high-mass, high-ϵ corner of our parameter space.

5.6.2 Supernova Constraints

The production and escape of dark sector particles during a core-collapse supernova can lead to cooling of the proto-neutron star that differs from the SM prediction [131, 132]. Such anomalous cooling is constrained by our observation of SN1987A [133,

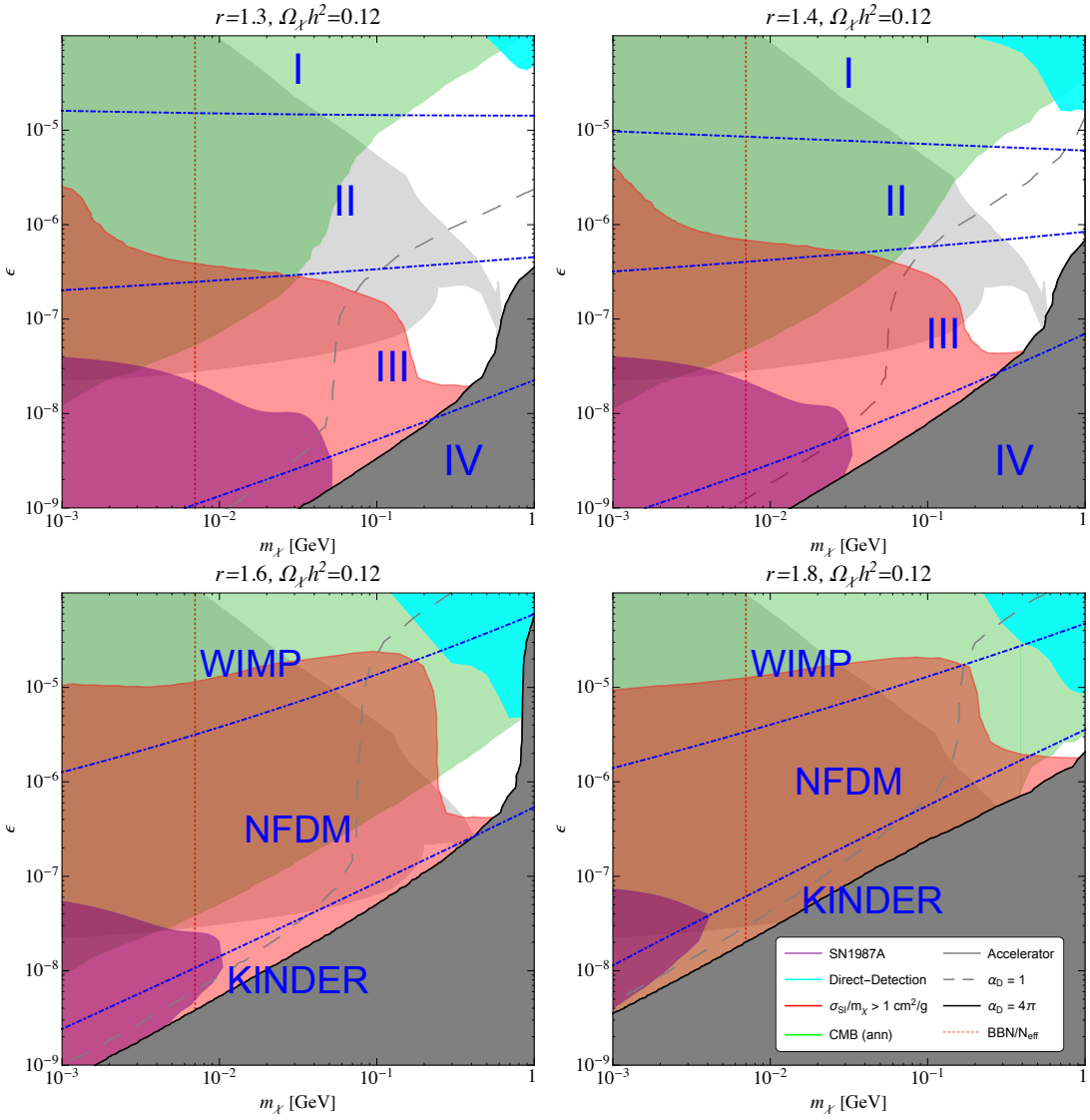

Figure 5-18: Constraints on our dark matter model for *(upper left)* $r = 1.3$, *(upper right)* $r = 1.4$, *(lower left)* $r = 1.6$, and *(lower right)* $r = 1.8$. The purple shaded regions are the constraints from SN1987A. The cyan regions on the upper-right corners of the plots are from direct detection experiments. We plot the self-interaction constraints as red shaded regions, and CMB s-wave annihilation limits with green. We also show the accelerator constraints as light gray. The constraint from BBN is shown as a red dotted line. Finally, we add the $\alpha_D = 1$ contour (gray dashed curve) and shade out the region where $\alpha_D > 4\pi$ with dark gray. For $r = 1.3$ and 1.4, the boundaries (blue dot-dashed lines) between Regimes I, II, III, and IV (see 5.5.2) are shown. For $r = 1.6$ and 1.8, the WIMP, NFDM (Section 5.4.1), and KINDER (Section 5.4.2) regimes are separated by blue dot-dashed lines.

134].[1]

Ref. [136] carefully derived constraints on the m_χ–ϵ plane in the vector-portal DM model using the SN1987A result for $m_\chi = 3m_{A'}$, and for two discrete α_D values, together with constraints for models with only A' and no DM. For fixed α_D, m_χ and $m_{A'}$, the excluded region is generally enclosed by two boundary values of ϵ. The lower boundary in ϵ is determined by the rate of production of the dark-sector particles from the SN core: models with smaller values of ϵ are allowed because they do not lead to enough production of dark-sector particles to modify the supernova evolution significantly. The upper boundary on ϵ is determined by whether the dark-sector particles will thermalize with the SM material in the proto-neutron star before escaping the SN, leaving these particles trapped; in models with larger values of ϵ, the dark-sector particles are thermalized efficiently and do not escape and cool the proto-neutron star, and hence these scenarios are unconstrained.

We now discuss how to recast the bounds in Ref. [136] for different values of α_D. The maximum value of m_χ is independent of α_D, being set by the kinematics of the supernova. The behavior of the lower bound is determined by the DM mass with respect to the plasma frequency of the interior, $\omega_p \sim 15\,\mathrm{MeV}$. For $2m_\chi > \omega_p$, the off-shell DM production via bremsstrahlung through virtual dark photons during neutron-proton collisions is suppressed, and the direct production of A' is more important. Consequently, the lower bound in ϵ is very similar to that in the dark-photon-only case, and is roughly independent of α_D. For $2m_\chi < \omega_p$, however, $\chi\overline{\chi}$-pairs can be produced through an on-shell A', and the production rate is fixed by the value of $\alpha_D\epsilon^2$. For a lower bound given at a reference value $\alpha_{D,\mathrm{ref}}$, we can therefore rescale to a new value of α_D by leaving the part of the bound where $2m_\chi > \omega_p$ constant, and rescaling the ϵ limit where $2m_\chi < \omega_p$ by $\sqrt{\alpha_{D,\mathrm{ref}}/\alpha_D}$.

The upper boundary of the limit on ϵ is determined by the dark-matter-proton scattering cross-section, and consequently varying α_D changes the asymptotically flat part of the upper boundary in ϵ such that $\alpha_D\epsilon^2$ is kept fixed, i.e. from a reference

[1] Alternative cooling models have also been proposed that cast doubt on the SN1987A bounds (see, e.g., Ref. [135]).

upper limit given for $\alpha_{D,\mathrm{ref}}$, we rescale by $\sqrt{\alpha_{D,\mathrm{ref}}/\alpha_D}$.

We find that for $\epsilon \gtrsim 10^{-9}$, the DM rate of production in the supernova in our model is always large enough for a significant amount to be produced; our limits are therefore set by the upper limit on ϵ, as determined by the thermalization condition. Note that this also happens for the lower boundary of our curves since there α_D is very large. The SN1987A constraints cover the low-ϵ and low-m_χ part of the parameter space, and generally lie entirely within the self-interaction constraints that we will describe next (albeit with different model-dependence).

5.6.3 DM Self-Interactions

The cross section for elastic DM-DM scattering is constrained by cluster mergers and halo shapes to satisfy $\sigma_{\mathrm{SI}}/m_\chi \leq 1\,\mathrm{cm}^2\,\mathrm{g}^{-1} \sim 5\times 10^3\,\mathrm{GeV}^{-3}$ [137]. The DM self-interaction rates for $\chi\chi \to \chi\chi$ and $\chi\bar{\chi} \to \chi\bar{\chi}$ (and their conjugate processes) are determined in Refs. [119, 120]. Including both s and t-channel tree level diagrams, the averaged cross section σ_{SI} is given by:

$$\begin{aligned}\frac{\sigma_{\mathrm{SI}}}{m_\chi} &= 3\pi h(r)\frac{\alpha_D^2}{m_\chi^3} \\ &= 1\,\mathrm{cm}^2\,\mathrm{g}^{-1}\left(\frac{h(r)}{1.2}\right)\left(\frac{10\,\mathrm{MeV}}{m_\chi}\right)^3\left(\frac{\alpha_D}{0.02}\right)^2 ,\end{aligned} \tag{5.104}$$

where

$$h(r) \equiv \frac{16-16r^2+5r^4}{r^4(r^2-4)^2}\,. \tag{5.105}$$

Typical values of $h(r)$ are $h(1.3) = 0.2$ and $h(1.8) = 2.7$.

As shown in Fig. 5-18, this constraint rules out a large fraction of the parameter space especially at low ϵ, generically excluding ϵ as high as 10^{-6}–10^{-5} depending on r; this behavior occurs because the values of α_D required to obtain the correct relic density are higher at small ϵ. In this sense the self-interaction bound is complementary to limits on the interactions with the SM, which are suppressed by small ϵ.

One possible way to evade this constraint is to consider a scenario where only some

subdominant fraction of the DM is produced by the mechanisms we have considered in this work, as this limit is rather sensitive to the fraction of DM that is self-interacting. For example, Ref. [138] shows that if the self-interacting component is less than 1% of the DM, these constraints become inapplicable. However, a full self-consistent treatment of fractionally abundance self-interacting dark matter constraints would require recalculation of the cosmological evolution in order to obtain a lower relic density, and is beyond the scope of this work.

5.6.4 CMB Constraints on DM Annihilation

During the post-recombination epoch, DM annihilation to e^+e^- leads to energy deposition into the baryonic gas; the resulting extra ionization can be constrained based on observations of the CMB anisotropy. We compare the annihilation cross section for $\chi\bar{\chi} \to f\bar{f}$ (see App. B) to the limits derived in Ref. [139] and updated in Ref. [115]. We plot the region excluded by this constraint in 5-18.

We observe that these CMB constraints provide some of the strongest bounds on models of this type for r close to 2, excluding most of the available parameter space. Even for smaller values of r, the CMB constraints provide stringent limits for models with low m_χ and high ϵ.

These limits could be lifted or relaxed if the dark-sector model were adjusted in order to suppress the DM annihilation to SM particles at low velocities. For example, this could be achieved if the DM was a scalar rather than a fermion, as then the leading-order annihilation through the dark photon would be p-wave and scale as $\langle\sigma v\rangle \propto v^2$.

5.6.5 Cosmological Constraints on Light Relics

Electromagnetically coupled DM with a mass of around 1 MeV can significantly affect the process of Big Bang Nucleosynthesis (BBN) by *(i)* directly increasing the expansion rate as a contribution to the energy density of the universe, and *(ii)* injecting entropy into the SM sector and changing the relative energy density of the

electromagnetic sector as compared to the neutrino sector, altering the temperature evolution of both sectors with respect to standard cosmology. These changes in turn alter the predicted abundance of light nuclei like deuterium and helium-4, which can then be compared with existing measurements of the abundances of these nuclei (see e.g. Refs. [140–144] for deuterium and helium-4). The injection of entropy from electromagnetically coupled DM can also decrease N_{eff} [145], the effective number of degrees of freedom, during the CMB epoch, which can then be constrained by the CMB anisotropy power spectrum [115].

Ref. [146] modelled the predicted primordial elemental abundances in the presence of an electromagnetically coupled dark matter particle; we adopt their results for our BBN constraints. They presented two constraints, depending on whether a prior was imposed on $\Omega_b h^2$ in the BBN calculations. When no prior was imposed, the bound is relatively weak, $m_\chi \gtrsim 0.7\,\text{MeV}$ for Dirac fermion DM. With a prior based on CMB observations, $\Omega_b h^2 = 0.02225 \pm 0.00066$ [115], this bound improves to $m_\chi \gtrsim 7\,\text{MeV}$, since the effect of entropy injection into the SM from the DM cannot be compensated for by lowering $\Omega_b h^2$ arbitrarily.

We note however that assuming the central value of $\Omega_b h^2$ from Planck leads to a standard BBN theoretical prediction of D/H that is roughly 2σ below the central measured value. This discrepancy may indicate an incomplete understanding of the process of BBN even in standard cosmology, which may therefore affect the bound given above.

As mentioned above, one can also consider the impact of electromagnetically coupled DM particles on the CMB anisotropy power spectrum. Electromagnetically coupled DM particles heat the electromagnetic sector as they become nonrelativistic, effectively decreasing the number of relativistic degrees of freedom at late times by increasing the ratio of photon to neutrino temperatures. The Planck 2018 measurement [115] sets a constraint on electrophilic Dirac fermions of $m_\chi \gtrsim 7.4\,\text{MeV}$. A joint constraint using both primordial elemental abundance and CMB data strengthens the constraint on electrophilic Dirac fermion to $m_\chi \gtrsim 10\,\text{MeV}$. However, CMB N_{eff} bounds are less robust than the BBN constraint, and can be overcome by e.g. adding

dark, relativistic degrees of freedom to compensate for the effect of the electromagnetically coupled DM [147].

Given the above consideration, we set a tentative constraint of $m_\chi > 7\,\mathrm{MeV}$ to indicate the potential constraint from BBN and CMB. Since the region with $m_\chi < 10\,\mathrm{MeV}$ is already strongly constrained by the CMB limits on DM s-wave annihilation, beam dump experiments and SN1987A, this constraint is not particularly important to understanding the viability of the model.

5.6.6 Warm Dark Matter

In the $1.5 \lesssim r \lesssim 2$ KINDER regime, the dark sector undergoes an early kinetic decoupling from the SM, after which the dark sector temperature T' evolves only logarithmically with respect to the SM temperature T until the $3 \to 2$ process freezes out. As a result, the dark sector temperature can be much higher than in the standard WIMP paradigm, where $T' = T$ until kinetic decoupling, after which $T' \propto (1+z)^2$. Models of warm dark matter (WDM) typically have suppressed structure on small scales [148, 149], and can be constrained by measurements of the matter power spectrum from the Lyman-α forest [72, 150], which are sensitive to modes with comoving wavenumber as large as $k_{\mathrm{max}} \sim 3\,h\,\mathrm{Mpc}^{-1}$.

To get an estimate for how important the WDM Lyman-α bounds are to the KINDER regime, we estimate the comoving Jeans length λ_J of DM, and compare this with $2\pi/k_{\mathrm{max}} \sim 2\,h^{-1}\,\mathrm{Mpc}$; for model parameters where $\lambda_J \ll 2\pi/k_{\mathrm{max}}$, the model is unlikely to leave a significant imprint on the matter power spectrum on scales currently probed by experiments. We leave a detailed analysis of such potential WDM constraints for future work.

The comoving Jeans length for the DM is given by [151]

$$\lambda_J(z) = (1+z)\sqrt{\frac{T'(z)}{m_\chi}}\frac{2\sqrt{2}\pi}{H(z)}\,. \tag{5.106}$$

After the dark sector completely freezes out, $T' \propto (1+z)^2$; in the radiation dominated era, λ_J stays roughly constant, while $\lambda_J \propto (1+z)^{1/2}$ during matter domination,

decreasing with time. To make a conservative estimate, we therefore want to compare $\lambda_J(z_{\rm eq})$ with $2\pi/k_{\rm max}$ at the redshift of matter-radiation equality, $z_{\rm eq}$.[2] We can estimate the temperature of the dark sector at $z_{\rm eq}$ as

$$T'(z_{\rm eq}) \simeq T'_3 \frac{(1+z_{\rm eq})^2}{(1+z_3)^2} \simeq T'_3 \frac{T_{\rm eq}^2}{T_3^2}\,, \tag{5.107}$$

where z_3 and T'_3 are the redshift and dark sector temperature at $3 \to 2$ freezeout respectively. With this approximation, we have

$$\lambda_J(z_{\rm eq}) \simeq (1+z_{\rm eq}) \frac{T_{\rm eq}}{T_3\sqrt{x'_3}} \frac{2\sqrt{2}\pi}{H(z_{\rm eq})}\,. \tag{5.108}$$

Taking $z_{\rm eq} = 3402$ and assuming a ΛCDM cosmology, we can obtain the following estimate for the Jeans length at matter-radiation equality:

$$\lambda_J(z_{\rm eq}) \simeq 0.2\,h^{-1}\,{\rm Mpc} \left(\frac{x_3}{10^3}\right) \left(\frac{\rm MeV}{m_\chi}\right) \left(\frac{10}{x'_3}\right)^{1/2}\,. \tag{5.109}$$

In the $1.5 \lesssim r \lesssim 2$ KINDER regime, we know that $x'_3 \sim x_d$, since x' evolves logarithmically with respect to x in thermodynamic phases B and C, while x_3 is largest when the $3 \leftrightarrow 2$ process freezes out at the latest possible time. We therefore find that $\lambda_J(z_{\rm eq})$ is largest at *(i)* small ϵ, so that decoupling occurs early, minimizing x_d and thus x'_3, and *(ii)* large α_D with small m_χ, so that the $3 \leftrightarrow 2$ cross section is large, and the process freezes out as late as possible, maximizing x_3. To maximize the impact on small-scale structure, we therefore take the smallest mass we consider $m_\chi = 1\,{\rm MeV}$, choose the largest perturbative value of $\alpha_D = 4\pi$, giving $\epsilon = 3.5 \times 10^{-9}$ to achieve the observed relic abundance for $r = 1.8$. We find that $x_3 = 5500$ and $x'_3 = 45$, leading to $\lambda_J(z_{\rm eq}) \simeq 0.5\,h^{-1}\,{\rm Mpc}$, which is still small enough to be consistent with probes of small-scale structure. Other parameter combinations that obtain the observed relic abundance lead to smaller values of $\lambda_J(z_{\rm eq})$.

For $1 \lesssim r \lesssim 1.5$, Regime I has $T' = T$ until freezeout of the dark sector, while in

[2]The Jeans scale at matter-radiation equality is on the same order as the free-streaming length of warm dark matter at matter-radiation equality, another common method of determining the length scale below which structure is damped [151].

Regimes II and III, the dark sector is actually colder than a dark sector that is thermally coupled to the SM until freezeout, easily avoiding these warm DM constraints. In Regime IV, a similar argument as above shows that $\lambda_J(z_{\rm eq})$ is given by Eq. (5.109) with x_3, x_3' replaced by x_2, x_2'. Once again, large values of α_D, small values of ϵ and small m_χ would lead to the largest impact on small-scale structure. SN1987A constraints and the requirement of a perturbative value of $\alpha_D < 4\pi$, however, are enough to constrain $\epsilon \gtrsim 10^{-9}$. Choosing $r = 1.4$, $\epsilon = 10^{-9}$, $m_\chi = 1\,\text{MeV}$ and $\alpha_D = 0.19$, we find $x_2 = 630$, $x_2' = 89$ and $\lambda_J(z_{\rm eq}) \simeq 5 \times 10^{-3}\,h^{-1}\,\text{Mpc}$, much smaller than would be observable. Larger values of m_χ require larger values of ϵ to meet the relic abundance criterion, and lead to even smaller values of $\lambda_J(z_{\rm eq})$. Similar results hold for $r = 1.3$ as well.

We therefore find that $\lambda_J(z_{\rm eq}) \ll 2\pi/k_{\rm max}$ is satisfied throughout all relevant parameter space, leaving our model unconstrained by small-scale structure observations. However, parts of the KINDER regime are close to being constrained by existing power spectrum measurements; future improvements in WDM constraints could potentially probe these models.

5.6.7 Summary of Constraints

Fig. 5-18 shows a plot of the constraints on the m_χ–ϵ plane with four different values of r, with α_D chosen at every point in parameter space such that the observed relic abundance of DM is attained, $\Omega_\chi h^2 = 0.12$. Regions ruled out by the constraints discussed above are marked in color; parts of the space that require $\alpha_D > 4\pi$ to obtain the correct relic abundance are also shaded gray, since perturbative control of our model breaks down there. The contour of $\alpha_D = 1$ is also shown for reference.

For $1.5 \lesssim r \lesssim 2$, we show the constraints for two representative values, $r = 1.6$ and $r = 1.8$. In both cases, a small region of open parameter space exists near $\epsilon \sim 10^{-6}$ and with DM masses of a few hundred MeV. For these values of r, the vector-portal DM model is bounded from below by the nonperturbative region, and is strongly constrained by the CMB s-wave annihilation bound and self-interaction limits. The available parameter space sits in the NFDM regime for $r = 1.6$, and in the

KINDER regime for $r = 1.8$. The unconstrained regions are similar to those obtained in Ref. [120] at the high-ϵ end, but differ at the low-ϵ end due to the KINDER regime that we have found in this chapter.

For $1 \lesssim r \lesssim 1.5$, we show the constraints for $r = 1.3$ and $r = 1.4$. Here, there are two viable regions of parameter space: both are in the range $m_\chi \gtrsim 100\,\mathrm{MeV}$, and are separated by the beam dump constraints: one region in Regime III is in the range $\epsilon \sim 10^{-8}$ – 10^{-7}, while the other is in Regime II and I in the range $\epsilon \sim 10^{-7}$ – 5×10^{-5}. In this range of r-values, both the self-interaction and CMB s-wave annihilation limits are less constraining, allowing more open parameter space than for $1.5 \lesssim r \lesssim 2$. These new limits represent an improved calculation over those found in Ref. [113]. In particular, most of the available parameter space is not in Regime I, as assumed by Ref. [113]. In contrast to that work, we find that there is a lower limit of $\epsilon \gtrsim 10^{-8}$ imposed by perturbativity and self-interaction constraints, since (in Regime IV) α_D needs to become very large at such small values of ϵ in order to achieve the correct DM relic abundance.

We emphasize that these constraints are derived assuming that the dark sector is in thermal equilibrium with the SM at $T \sim m_\chi$, which may not be a valid assumption for values smaller than ϵ_{eq} as defined in Eq. (5.37). For $\epsilon \sim 10^{-9}$ and below, other mechanisms such as freeze-in can potentially achieve the correct relic abundance without the dark sector ever being in thermal equilibrium with the SM.

5.6.8 Lifting CMB and Self-Interaction Constraints with Pseudo-Dirac DM

In the previous subsections, we have demonstrated that the bulk of the parameter space for this class of models with $1.2 < r < 1.8$ has been tested by existing observations and experiments, for the baseline scenario where the DM is a Dirac fermion. Narrow regions of parameter space remain open, but for example, Regime IV for $1 \lesssim r \lesssim 1.5$ appears to be fully excluded.

However, these exclusions rely critically on constraints from the CMB and from

self-interactions, both of which probe the behavior of the DM long after freezeout. This exclusion is model-dependent; it is possible to perturb our baseline model in ways that dramatically alleviate these constraints while leaving the cosmology during the freezeout epoch essentially unchanged.

As a specific example, suppose that the DM is a pseudo-Dirac fermion, where at low energies the DM is split into two nearly-degenerate Majorana mass eigenstates χ_1, χ_2 (see e.g. Refs. [152–154] for specific models). The gauge interaction between the DM and the A' ($\bar{\chi}\not{A}'\chi$) then gives rise to interactions of the form $\bar{\chi}_i\not{A}'\chi_j$, $i \neq j$. There is no $\bar{\chi}_i\not{A}'\chi_i$ vertex as Majorana fermions cannot carry a conserved dark charge. The heavier mass eigenstate χ_2 can thus decay to the lighter eigenstate χ_1 via emission of an off-shell A'.

When the temperature of the dark sector exceeds the mass splitting between the states, the DM will behave as a Dirac fermion, and thus for a mass splitting $\Delta m_\chi \ll T'$ throughout freezeout, our previous cosmological results will still hold. However, once $T' \ll \Delta m_\chi$, the DM will convert into the lighter mass eigenstate provided the lifetime of the heavier eigenstate is sufficiently short (even if the lifetime is long, DM-DM scattering can also efficiently deplete the heavier eigenstate). Thus during the recombination epoch and in galaxies at late times, any process requiring the presence of both mass eigenstates will be strongly suppressed.

This suppression applies to both the annihilation $\bar{\chi}\chi \to e^+e^-$ through an s-channel A', which determines the CMB constraint,[3] and to the contribution to the tree-level self-interaction cross section $\bar{\chi}\chi \to \bar{\chi}\chi$ from an s-channel A'. The contribution to the tree-level self-interaction cross sections from a t-channel A' exchange is suppressed for a related reason; if the initial state is $\chi_1\chi_1$ then the final state (at tree level) can only be $\chi_2\chi_2$, which is kinematically forbidden provided the kinetic energy of DM particles in the halo is much smaller than the mass splitting. There will still be a contribution to the self-interaction cross section at 1-loop order, and a CMB signal via t-channel annihilation of χ_1's to the 3-body final state $A' + e^+ + e^-$ [155] (as well as possible contributions from the residual χ_2 abundance), but these rates are

[3]The relaxation of CMB bounds for pseudo-Dirac DM is well-known, see e.g. Ref. [96].

parametrically suppressed compared to those relevant for the Dirac case.

Thus we expect both the CMB and self-interaction limits to be dramatically relaxed in the pseudo-Dirac case without changing the freezeout history, for mass splittings that are small compared to T' at freezeout, but large compared to the DM temperature during recombination and the kinetic energy of DM particles in present-day halos. This modification opens up allowed parameter space spanning all the freezeout regimes we have studied; we will present a detailed computation of the modified constraints in Chapter 6.

5.7 Conclusion

We have fully characterized the possible freezeout histories of the vector-portal DM model in Eq. (5.1), in the region of parameter space in which the DM is a thermal relic, and $1 \lesssim r \lesssim 2$. In this region, the $\chi\chi\bar{\chi} \leftrightarrow \chi A'$ $(3 \leftrightarrow 2)$ and kinematically suppressed $\chi\bar{\chi} \leftrightarrow A'A'$ $(2 \leftrightarrow 2)$ processes play important roles in the thermal freezeout of the DM. Extending beyond the scope of previous studies [113, 114], we explored this model for values of the kinetic mixing parameter ϵ where the dark and SM sectors do not remain in kinetic equilibrium throughout the process of DM thermal freezeout. Doing so reveals a rich set of novel thermal histories, leading to very different dependences of the DM relic abundance on the model parameters.

We have identified four novel pathways by which thermal freezeout of the dark sector can proceed, in addition to those identified in previous studies. Two of these pathways share key features, and represent a general class of freezeout histories that we dub the "KINetically DEcoupling Relic" (KINDER). In the KINDER scenario, the DM relic abundance is determined primarily by the kinetic decoupling of the dark and SM sectors. KINDER is realized through a process of dark sector cannibalization, which was previously invoked in the ELDER scenario [93, 94]. In this work, we have demonstrated that cannibalization can be supported by a $3 \to 2$ annihilation process involving multiple dark sector species, and can proceed even in the presence of nonzero dark sector chemical potentials. ELDER DM can be regarded as an example

of a KINDER scenario where the kinetic decoupling is controlled by elastic scattering between the DM and SM.

We have presented detailed numerical results for the thermal history of the dark sector in each of these new regimes. Additionally, in a number of cases we were able to analytically derive the evolution of the dark sector temperature T' and dark matter abundance Y_χ, throughout the freezeout of the DM; this allows us to analytically demonstrate the dependence of the DM relic abundance on the model parameters in much of parameter space.

The novel freezeout mechanisms we have characterized, and their corresponding distinct regimes of parameter space, can be separated into two main parameter regions in r. In the region $1.5 \lesssim r \lesssim 2$, in addition to the "classic not-forbidden" regime studied in Ref. [114], we have identified a realization of KINDER at low values of ϵ.

In the region $1 \lesssim r \lesssim 1.5$, in addition to the "classic forbidden" regime studied in Ref. [113] (Regime I), which is valid at high ϵ, we identify a second variation of KINDER at very low ϵ (Regime IV). At intermediate values of ϵ, we find two previously unrecognized parameter regimes with distinct freezeout histories (Regimes II and III). In Regimes II and III the $A' \to f\bar{f}$ process is fast enough to maintain $n_{A'} \approx n_{A',0}\,(T)$ until all number-changing processes have frozen out. However, during the period after $3 \to 2$ freezeout and before $2 \to 2$ freezeout, this process cannot maintain thermal equilibrium between the DM and SM sectors due to number and energy conservation requirements enforced by the Boltzmann equations. In these regimes the elastic scattering $\chi f \to \chi f$ process controls the heat exchange between the DM and SM sectors after the freezeout of the $3 \leftrightarrow 2$ process and before the freezeout of the $2 \leftrightarrow 2$ process, while the $2 \leftrightarrow 2$ process cools the dark sector.

The distinguishing feature between Regimes II and III is the efficiency with which the elastic scattering process heats the dark sector. In Regime III, elastic scattering is inefficient, the dark sector is cooled by the kinematically forbidden $\chi\overline{\chi} \to A'A'$ $(2 \to 2)$ process, and the chemical potential of the dark sector is such that the χ abundance no longer evolves appreciably after the $3 \leftrightarrow 2$ process freezes out. This leads to a DM relic abundance determined only by the freezeout of the $3 \leftrightarrow 2$ process, even though

the $2 \leftrightarrow 2$ process is significantly faster. In Regime II, in contrast, elastic scattering remains efficient after the freezeout of the $3 \leftrightarrow 2$ process, and can counteract the cooling of the dark sector, allowing continued evolution of the DM density. This leads to a DM relic abundance determined by the interplay of elastic scattering and dark sector processes.

The two variations of KINDER we have identified differ in their evolution at late times, after the slower dark sector process freezes out. For $1.5 \lesssim r \lesssim 2$, cannibalization continues through the $3 \to 2$ process until all number-changing processes have frozen out, ensuring a slow evolution of the DM number density after kinetic decoupling. In contrast, for $1 \lesssim r \lesssim 1.5$, the cannibalization is halted once the $3 \leftrightarrow 2$ process freezes out. The number-changing $2 \leftrightarrow 2$ process is still active at this point, and cools the dark sector; however, the chemical potential evolves such that the χ abundance remains constant regardless.

We have calculated the relevant experimental constraints on our model. Our results drastically modify those of Ref. [113] for $\epsilon \lesssim 10^{-5}$ (below Regime I) and those of Ref. [114] for $\epsilon \lesssim 10^{-7}$ (the NFDM and KINDER Regimes). The KINDER mechanism realized in our model implies large self-interaction rates, and a large s-wave annihilation signal in the CMB, for symmetric Dirac fermion DM; these limits are in tension with the KINDER regime, although a small window of open parameter space remains for $r = 1.8$. There is also available parameter space in Regimes II and III for DM masses $\sim (0.1 - 1)$ GeV where experiments have not yet explored. In these allowed regions of parameter space, self-interactions can be in the correct range ($0.1\ \mathrm{cm}^2/\mathrm{g} \lesssim \sigma_{\mathrm{SI}}/m_\chi \lesssim 1\ \mathrm{cm}^2/\mathrm{g}$) to have observable consequences for the small-scale structure of galaxies without being currently excluded. Our new calculations provide target regions that can be tested by future sub-GeV direct detection experiments and dark photon searches.

In this chapter we have presented the baseline scenario of this vector-portal model in which the DM χ is a Dirac fermion. In Chapter 6 we will present an alternative to this baseline scenario in which the DM is a pseudo-Dirac fermion which at low energies splits into two nearly-degenerate Majorana mass eigenstates. For the correct range of

values of the mass splitting this scenario shares essentially the same cosmology as the Dirac case, while modifying the late-time cosmology in a way that relaxes both CMB and self-interaction constraints, thus opening windows of parameter space spanning all the novel freezeout regimes we have presented.

Chapter 6

New Thermal Relic Targets for Inelastic Vector-Portal Dark Matter

This chapter is from [3], *by me, Hongwan Liu, Tracy R. Slatyer, and Yu-Dai Tsai.*

6.1 Introduction

In Chapter 5 we studied the $1 \lesssim r \lesssim 2$ regime of the vector or kinetic mixing portal DM model [156], lifting the assumption of efficient thermal contact between the dark sector and the SM which had been made in previous studies (e.g. Refs. [119, 120]). We showed that dark sector processes like $\chi\overline{\chi} \leftrightarrow A'A'$ (hereafter denoted $2 \leftrightarrow 2$) and $\chi\overline{\chi}\chi \leftrightarrow \chi A'$ (hereafter denoted $3 \leftrightarrow 2$) can generate or remove a significant amount of heat during freezeout; if the energy transfer rate between the SM and the dark sector is not large enough, this can lead to the dark sector kinetically decoupling from the SM before freezeout is complete, leading to a dark sector temperature T' that is different from the SM temperature T. Kinetic decoupling can occur before freezeout is complete for ϵ as large as 10^{-5} for $r \lesssim 1.5$, dramatically altering how the correct relic abundance is achieved. While some parameter space in the range $100\,\mathrm{MeV} \lesssim m_\chi \lesssim 1\,\mathrm{GeV}$ remains open, especially for $r \lesssim 1.5$, this model still faces strong constraints from limits on the self-interaction cross section of DM, as well as the aforementioned CMB power spectrum.

In this chapter, we investigate the experimental constraints on *inelastic* DM in the $1 \lesssim r \lesssim 2$ regime, where the dark matter is now made up of a Majorana ground state χ and excited state χ^*, with a small mass splitting between them [157], with coupling to the A' now occurring off-diagonally, i.e., only between the χ and χ^* states. The existence of the mass splitting suppresses the primordial abundance of χ^* relative to χ, decreasing the annihilation rate of χ during recombination and lifting the CMB power spectrum constraints. The off-diagonal nature of the coupling ensures that self-interaction between χ particles is forbidden at tree-level, reducing the severity of the self-interaction constraints. This model has a significantly enlarged range of experimentally allowed parameter space, including a new window at $m_\chi \sim 10\,\mathrm{MeV}$ and $\epsilon \sim 10^{-8}$. Our results motivate future beam experiments to close the full range of ϵ over which the dark sector can thermalize with the SM in the early Universe, as well as an improved understanding of supernova cooling constraints on light DM at small mixing.

In the remainder of this chapter, we will review the inelastic DM model, the predicted primordial abundance of the excited state, and the newly relevant CMB and self-interaction constraints that replace those applicable to a symmetric DM model. We will then conclude by examining the existing experimental constraints on this model. More details of the inelastic DM model, a discussion of subdominant or model-dependent constraints, and the details of our calculation of the $\chi\chi \to \chi\chi$ one-loop self-interaction cross section can be found in Appendix C.

6.2 Vector-portal inelastic dark matter

Our dark sector contains a dark photon A' with mass $m_{A'}$, the massive gauge boson of a broken U(1) gauge symmetry in the dark sector, as well as a pair of Majorana fermions χ and χ^* that makes up the DM, which we will refer to as the 'ground state' and 'excited state' respectively. Both states have similar masses m_χ and m_{χ^*}, with a small dimensionless mass splitting δ defined by $m_{\chi^*} - m_\chi \equiv \delta m_\chi$, and $\delta \ll 1$.

Due to the ultraviolet (UV) origins of the mass splitting and the symmetry break-

ing of the dark U(1), which we will describe in more detail below, the dark fermions couple off-diagonally to A'. Moreover, A' kinetically mixes with the Standard Model (SM) photon [156], generating a coupling between the SM electromagnetic current $J^{\mu}_{\rm EM}$ and A'. The terms in the Lagrangian chiefly responsible for the phenomenology of this model are thus

$$\mathcal{L} \supset g_D A'_{\mu} \overline{\chi^*} \gamma^{\mu} \chi + \epsilon e A'_{\mu} J^{\mu}_{\rm EM} \,, \tag{6.1}$$

where ϵ is the kinetic mixing parameter, g_D is the dark sector coupling and e is the electron charge.

The small mass splitting δ, the dark fermion masses and the breaking of the U(1) symmetry associated with A' can all be simultaneously achieved by a judicious choice of the symmetry breaking pattern and couplings between a dark Higgs field Φ and a Dirac fermion Ψ, which is split into χ and χ^* mass eigenstates after symmetry breaking. Many models to achieve this have been proposed; two specific example models [153, 154] are discussed in Appendix C. In the limit where the dark Higgs particle h_D is much more massive than χ, there is no significant difference in the phenomenology of the dark sector between these two models; by default we will assume this condition.

There are several tree-level processes present in the symmetric model considered in Ref. [2] that are absent in this inelastic case: DM ground-state annihilation into SM particles, tree-level scattering between the DM and SM fermions, and DM self-interactions. This relaxes the resulting constraints significantly. In Fig. 6-1, we show three processes that are of particular importance to the inelastic vector-portal DM model, which we will now consider in turn.

6.3 Freezeout and primordial excited state abundance

For $\epsilon \gtrsim 10^{-9}$, the dark sector achieves thermal equilibrium with the SM while dark sector particles are relativistic [2]. Furthermore, while $T'/m_{\chi} \gg \delta$ (where T' is the

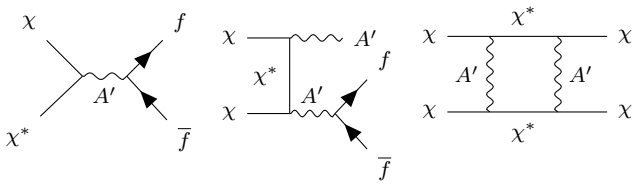

Figure 6-1: Feynman diagrams for important processes in the inelastic vector-portal dark matter model with $1 \lesssim r \lesssim 2$: (left) annihilation of DM with primordially produced excited states, $\chi^*\chi \to f\overline{f}$, where f is a SM fermion; (center) annihilation of DM with initial state radiation, $\chi\chi \to A'A'^*$, $A'^* \to f\overline{f}$, and (right) DM self-interaction, $\chi\chi \to \chi\chi$ (only one diagram shown here: see Appendix C for a complete discussion).

dark sector temperature), the mass splitting between χ and χ^* is irrelevant, and the dark fermions can equivalently be treated as part of the Dirac fermion Ψ. Taking $\delta \lesssim 10^{-3}$ ensures that this condition holds throughout freezeout. In this case, the results of Ref. [2] are fully applicable to the evolution of the dark sector until all DM number-changing processes have frozen out, fixing the DM abundance. After this point, the only remaining processes that are fast compared to cosmic expansion involve total-number-conserving conversions between χ and χ^* only.

After the freezeout of DM, χ and χ^* particles stay in chemical equilibrium as the dark sector cools, until $T'/m_\chi \sim \delta$, when the number density of χ^* starts becoming Boltzmann-suppressed relative to the number density of χ. Eventually, the comoving number density of χ^* freezes out when the number-changing process $\chi^*\chi^* \leftrightarrow \chi\chi$ becomes slow, i.e., when $n_{\chi^*}\langle\sigma v\rangle_{\chi^*\chi^*\to\chi\chi} \sim H$. The annihilation cross section is given by [158]

$$\langle\sigma v\rangle_{\chi^*\chi^*\to\chi\chi} \simeq \frac{8\sqrt{\pi}}{r^4}\frac{\alpha_D^2}{m_\chi^2} \times \begin{cases} \sqrt{1/x'}\,, & \delta \ll 1/x'\,, \\ \sqrt{\pi\delta/2}\,, & \delta \gg 1/x'\,, \end{cases} \tag{6.2}$$

where $x' \equiv m_\chi/T'$, and $\alpha_D \equiv g_D^2/(4\pi)$. In order for χ to make up all of the DM, $m_\chi n_\chi \sim T_{\rm eq}T^3$, where $T_{\rm eq}$ is the temperature of the CMB at matter-radiation equality.

Thus we can estimate the ratio n_{χ^*}/n_χ to be

$$\frac{n_{\chi^*}}{n_\chi} \sim \frac{1}{4\sqrt{2}\pi\sqrt{\delta}} \frac{r^4 x_*}{M_{\rm pl} T_{\rm eq}} \frac{m_\chi^2}{\alpha_D^2}\,, \tag{6.3}$$

where x_* (x'_*) is defined by the SM (dark sector) temperature at which $\chi^*\chi^* \leftrightarrow \chi\chi$ freezes out. Since the dark sector temperature redshifts as $(1+z)^2$ after DM freezes out, which occurs at SM and dark sector temperatures x_f and x'_f respectively, we can write $x_* \approx x_f (x'_*/x'_f)^{1/2}$, giving this parametric expression for n_{χ^*}/n_χ:

$$\begin{aligned}\frac{n_{\chi^*}}{n_\chi} \sim 10^{-7} \left(\frac{m_\chi}{10\,\text{MeV}}\right)^2 \left(\frac{r}{1.6}\right)^4 \left(\frac{0.1}{\alpha_D}\right)^2 \\ \times \left(\frac{x_f}{200}\right) \left(\frac{10}{x'_f}\right)^{1/2} \left(\frac{x'_* \delta}{20}\right)^{1/2} \left(\frac{10^{-3}}{\delta}\right)\,,\end{aligned} \tag{6.4}$$

keeping in mind that $x'_*\delta$ is the ratio of the mass splitting to the DM temperature at $\chi^*\chi^* \leftrightarrow \chi\chi$ freezeout. We see that n_{χ^*} can be easily suppressed by seven orders of magnitude relative to n_χ in our parameter space of interest.

Even with this small primordial abundance, $\chi^*\chi$ annihilations into energetic SM particles are potentially constrained by the CMB power spectrum [115]. The limit is given approximately as $f(m_\chi)\langle\sigma v\rangle_{\chi^*\chi\to e^+e^-} \lesssim 1.7\times10^{-30}\,\text{cm}^3\,\text{s}^{-1}(m_\chi/10\,\text{MeV}) n_\chi/n_{\chi^*}$, where $f(m_\chi)$ is an efficiency factor accounting for delayed absorption of energy injected through annihilations [115, 139].

The decays of χ^* through an off-shell A' to the ground state can produce high-energy particles that may be constrained by their effect on the CMB power spectrum and primordial elemental abundances from Big Bang Nucleosynthesis (BBN) [159]. For $\delta \lesssim 2\times10^{-4}$, these constraints are unimportant in the parameter space of interest, and so for convenience we choose $\delta = 10^{-4}$ in this Letter. For a discussion of these decays and the relevant constraints at larger values of δ, we refer the reader to Appendix C.

6.4 CMB initial state radiation limits

Even without an excited state population during recombination, DM self-annihilation with initial state radiation (ISR) $\chi\chi \to A'A'^*$, $A'^* \to f\overline{f}$ is kinematically allowed, with one A' produced off-shell. This process is particularly important in the forbidden regime [155], and may also be constrained by the CMB power spectrum (we also considered $\chi\chi \to f\overline{f}$ annihilations at one-loop, but find they are subdominant to both ISR and the $\chi\chi^*$ annihilations discussed above). In our model, the annihilation cross section into e^+e^- is given by [155]

$$\langle\sigma v\rangle_{\chi\chi\to A'e^+e^-} = 1.3\times10^{-35}\,\mathrm{cm}^3\,\mathrm{s}^{-1}\left(\frac{1.6}{r}\right)^2 \times\left(\frac{\alpha_D}{0.1}\right)^2\left(\frac{\epsilon}{10^{-8}}\right)^2\left(\frac{10\,\mathrm{MeV}}{m_\chi}\right)^2\left(\frac{I(r,\delta)}{10^{-2}}\right), \quad (6.5)$$

where $I(r,\delta)$ is a phase space integral that ranges from approximately 10^{-3} to 0.25 in the range $1.2 \lesssim r \lesssim 1.8$; the fit $\log_{10} I(r,\delta) \approx -0.19 + 5.0\log_{10}(2-r) + 1.7\log_{10}^2(2-r)$ is accurate to within 20% in the same range of r-values, as long as $\delta \lesssim 0.1$. Comparing this expression with the CMB limit of approximately $f(m_\chi)\langle\sigma v\rangle \lesssim 3.3\times10^{-30}\,\mathrm{cm}^3\,\mathrm{s}^{-1}(m_\chi/10\,\mathrm{MeV})$, we find that the CMB power spectrum limits on energy injection offer only mild constraints in our parameter space of interest.

6.5 Self-interaction limits

The other key constraint on this model comes from self-interaction between χ particles, which can modify the structure of galaxies; we adopt an upper limit of $\sigma/m_\chi < 1\,\mathrm{cm}^2\,\mathrm{g}^{-1}$ (e.g., Ref. [137]) on this cross section. $\chi\chi \to \chi\chi$ elastic scattering via A' exchange is forbidden at tree-level, but there is a non-zero contribution at one-loop order. We compute this one-loop cross section in the low-velocity limit; in the range $1 \lesssim r \lesssim 2$, the cross section is well-approximated numerically by

$$\sigma_{\chi\chi} \approx \frac{3}{4}\left(\frac{\pi}{2r^6} + \frac{18}{\pi r^4}\right)\frac{\alpha_D^4}{m_\chi^2}. \quad (6.6)$$

The complete expression for all r can be found in the Supplementary Materials; this expression agrees with the low-velocity cross section in the massless A' limit derived in Ref. [160].

Scattering processes involving initial χ^* particles can occur (e.g., $\chi\chi^* \to \chi\chi^*$, $\chi^*\chi^* \to \chi\chi$) but their rates are suppressed in galaxies by the small abundance of χ^* in the late Universe. Inelastic scattering can in principle occur from the ground state, $\chi\chi \to \chi^*\chi^*$, but will be kinematically forbidden provided that $\sqrt{2\delta}$ exceeds the local maximum velocity (e.g., for $\delta \gtrsim 10^{-4}$, $\sqrt{2\delta}$ is above the typical escape velocity of galaxy clusters). For these reasons, we take the overall self-interaction rate to be given by Eq. (6.6). Compared to the tree-level self-interaction cross section in the symmetric case [2], there is a parametric suppression of order α_D^2, which relaxes the constraints on the m_χ-ϵ parameter space significantly.

There can additionally be a model-dependent tree-level $\chi\chi$ self-interaction via the dark Higgs; the rate will be suppressed by $1/m_{h_D}^4$, where m_{h_D} is the dark Higgs mass, in addition to any model-dependent factors (e.g., one of our example models gives a δ^4 suppression). While this self-interaction rate can in principle dominate for specific models and a sufficiently light dark Higgs, we can safely choose model parameters for this to be subdominant to the one-loop expression in Eq. (6.6) without affecting the rest of the analysis. We leave a detailed discussion of the dark Higgs self-interaction rate to Appendix C.

6.6 Summary of constraints

Fig. 6-2 shows the experimentally allowed regions for $r = 1.3$, 1.4, 1.6 and 1.8 in the m_χ-ϵ plane, choosing α_D to obtain the correct relic abundance. We also show the different freezeout phases in the $1 \lesssim r \lesssim 2$ regime, derived in Ref. [2]. In addition to some of the new constraints unique to inelastic DM discussed above, additional constraints that apply identically to the symmetric Dirac fermion case studied in Ref. [2] are shown. These include: searches for the visible decay of A' at beam experiments [101–107, 109, 110, 161–167], limits on the emission of light

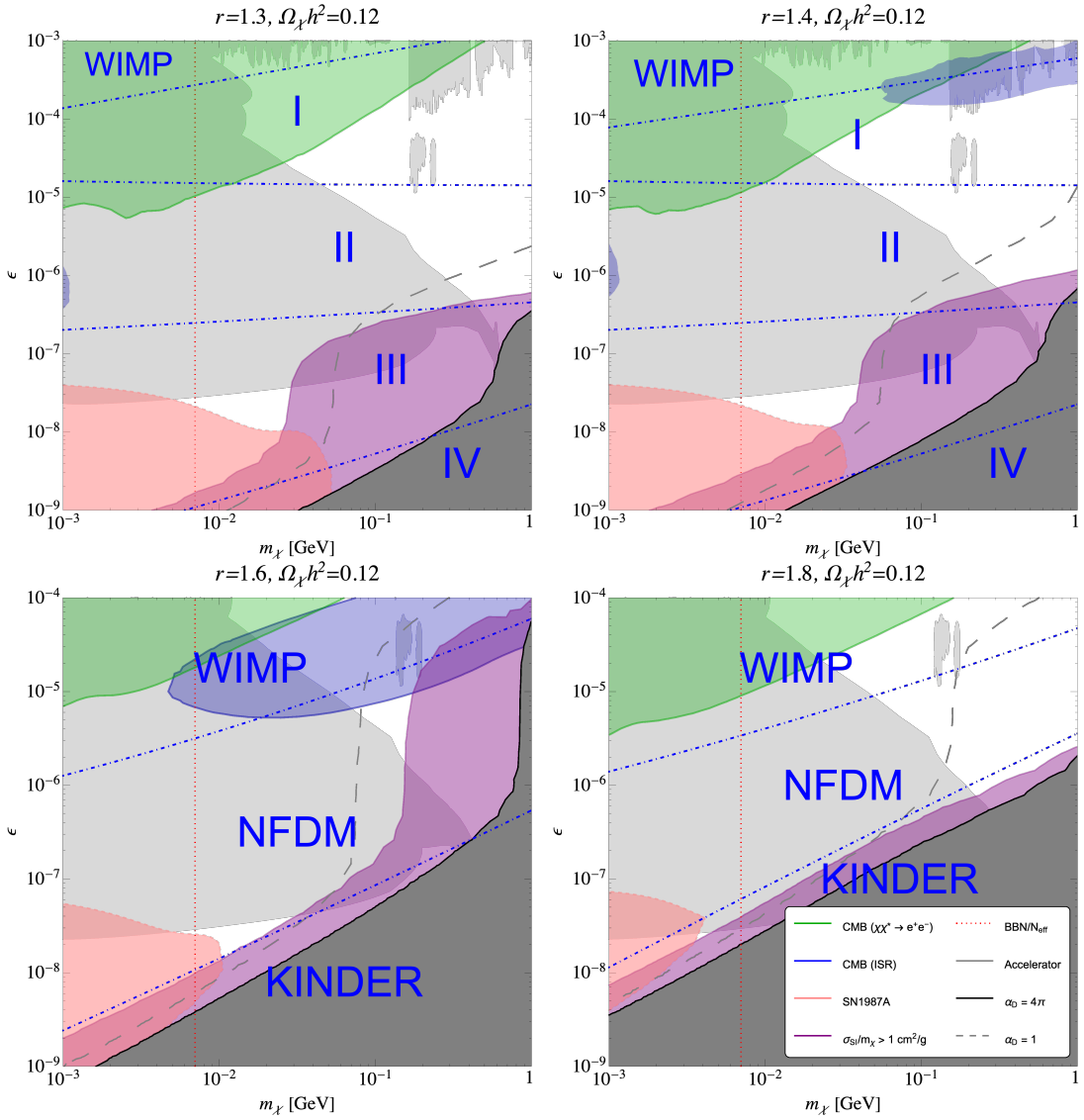

Figure 6-2: The m_χ-ϵ parameter space for (top left) $r = 1.3$, (top right) $r = 1.4$, (bottom left) $r = 1.6$ and (bottom right) $r = 1.8$, with $\delta = 10^{-4}$. The value of α_D that is needed to obtain a relic abundance of $\Omega_\chi h^2 = 0.12$ has been chosen for every point on the plot. Constraints on the parameter space from the cooling of SN1987a (red), $\chi\chi \to \chi\chi$ self-interaction (purple), CMB power spectrum constraints on $\chi\chi^* \to e^+e^-$ (green) (assuming $\delta = 10^{-4}$, with constraints weakening at larger δ) and $\chi\chi \to A'^*A' \to A'e^+e^-$ (blue), as well as beam experiments (light gray) are shown. A limit on electromagnetically coupled light dark matter from BBN and CMB is shown by the dotted red line; masses below the line are ruled out (assuming no other dark-sector effects). The region of the parameter space where nonperturbative values of $\alpha_D > 4\pi$ are needed for the right relic abundance is indicated in dark gray; the dashed gray line indicates $\alpha_D = 1$. Large labels corresponding to the various regimes discussed in Ref. [2] are shown for reference.

particles from the cooling of SN1987a [131–134, 136] (although arguments made in Refs. [135, 168] may alter or remove these limits), as well as joint BBN and CMB constraints on electromagnetically coupled DM, assuming no other dark sector effects on either of these observables [146]. We expect direct-detection limits to be irrelevant: DM scattering with electrons and nucleons occurs only at one-loop for the ground state (with upscattering being kinematically forbidden), and is suppressed by at least an additional factor of ϵ^2 relative to the symmetric scattering rate, while the scattering rate of excited states is suppressed by their tiny abundance [169, 170].

For the benchmark points where $r < 1.5$ and $r = 1.8$, a high-mass window with $30\,\mathrm{MeV} \lesssim m_\chi \lesssim 1\,\mathrm{GeV}$ and $10^{-6} \lesssim \epsilon \lesssim 10^{-4}$ remains mostly unconstrained, with the notable exception of the LHCb search for relatively long-lived A' decaying into muons [165, 167]. For $r = 1.6$, a combination of constraints from DM self-interaction and CMB constraints on DM self-annihilation with ISR closes most of the high mass window. In addition, for $r \leq 1.6$, a low-mass window near $m_\chi \sim 10\,\mathrm{MeV}$ and $\epsilon \sim 10^{-8}$ remains viable. For $r = 1.8$, the low mass window is almost completely closed due to the large dark sector couplings required for the correct relic abundance and consequently large self-interaction limits; however, this exclusion is dependent on the accuracy of the supernova constraints.

6.7 Conclusion

We have analyzed the experimental constraints on the vector-portal inelastic DM model in the regime where $1 \lesssim r \equiv m_{A'}/m_\chi \lesssim 2$, adopting the results of Ref. [2] for the relic abundance calculations in this regime. We have studied several possible DM annihilation and decay channels which are constrained by the CMB power spectrum, and have carefully examined the self-interaction constraints on this model, deriving in particular the one-loop self-scattering cross section at zero velocity for arbitrary A' masses. This model has significant regions of viable parameter space that are still unexplored, providing a simple and strong motivation for a range of future experimental searches for the A'. The high-mass window is a potential target for beam

dump experiments which are sensitive to relatively long-lived dark photons which decay visibly, including LHCb [164, 171], FASER [172], Belle II [173], DarkQuest/-LongQuest [110, 174], and many other future experiments [174–180]. The low-mass window also motivates a more complete study of SN1987a constraints in the $1 \lesssim r \lesssim 2$ regime.

Chapter 7

Dark-onium

7.1 Introduction

We might expect dark matter to form bound states, similar to the formation of bound states in the Standard Model. For example, dark matter may interact with itself via a massless force carrier, in which case it would form bound states analogous to the formation of bound states such as positronium and atoms in electromagnetism. In the case of the dark photon model of Chapters 5 and 6 this can be realized in the limit in which $r \equiv m_{A'}/m_\chi \ll 1$. If instead dark matter interacts via a massive force carrier, where in the nonrelativistic limit its interaction potential is given by the Yukawa potential, $V(r) = -\alpha_D \exp(-m_Y r)/r$, dark matter particles can capture into bound states when the potential is sufficiently long-range. The potential is sufficiently long-range to support the formation of bound states when the Bohr radius of the bound state, which is the size of the ground state, $a_0 \sim 1/(\alpha_D \mu_\chi)$, is smaller than the length scale of the mediator, $1/m_Y$:

$$\frac{1}{m_Y} \gtrsim \frac{1}{\alpha_D \mu_\chi}, \tag{7.1}$$

where μ_χ is the reduced mass of the two-particle system.

If the dark matter forms bound states then, as we'll see in this (and the next) chapters, this can lead to interesting phenomenological consequences with distinct experimental consequences. Before getting into that, let's warm up our understanding

of bound states with two useful examples from quantum electrodynamics.

7.2 Bound states in quantum field theory

Since usually in quantum field theory we are used to dealing with free particle states, as we've done throughout the work in this thesis so far, let's first start with a brief discussion of how bound states are treated in quantum field theory. It will be sufficient for us to consider the formation and decay of nonrelativistic bound states in which the internal motions of the constituent particles are slow. In this section I'll mostly follow the discussion in Chapter 5 of Michael Peskin and Daniel Schroeder's textbook "An Introduction to Quantum Field Theory" [181].

Let's start by writing a bound state of positronium in terms of free-particle electron and positron states in the center-of-mass frame with total momentum $\mathbf{K} = \mathbf{k}_1 + \mathbf{k}_2 = 0$:

$$|\mathcal{B}\rangle = \sqrt{2M_{\mathcal{B}}} \int \frac{d^3\mathbf{k}}{(2\pi)^3} \psi(\mathbf{k}) C_{ab} \frac{1}{\sqrt{2m}} |e_-^a(\mathbf{k})\rangle \frac{1}{\sqrt{2m}} |e_+^b(-\mathbf{k})\rangle, \tag{7.2}$$

where $\psi(\mathbf{k})$ is the Schrödinger wavefunction which solves the nonrelativistic Schrödinger equation describing the interaction producing the bound state, in the momentum-space representation:

$$\psi(\mathbf{k}) = \int d^3\mathbf{x} e^{i\mathbf{k}\cdot\mathbf{r}} \psi(\mathbf{r}), \tag{7.3}$$

which is normalized by:

$$\int \frac{d^3\mathbf{k}}{(2\pi)^3} |\psi(\mathbf{k})|^2 = 1. \tag{7.4}$$

$M_{\mathcal{B}}$ is the mass of the state, and the factors of $\sqrt{2m}$, where m is the mass of the electron/positron, are included to obtain the correct normalization of the bound state. a and b are spin labels, and the coefficient C_{ab} depends on the spin configuration of positronium. Equation 7.2 can easily be generalized to represent a bound state of any two fermions with non-zero total momentum. In Equation 7.2 we've written the bound state as a linear combination of free-particle states weighted by $\psi(\mathbf{k})$ over momentum $\mathbf{k}$. To get some intuition let's apply Equation 7.2 to a concrete example, positronium decay.

7.2.1 Positronium decay

Let's consider the decay of positronium to two photons: Ps $\to \gamma\gamma$. From Equation 7.2 we can immediately write down the amplitude for this process:

$$\mathcal{M}\left(\text{Ps} \to 2\gamma\right) = \sqrt{2M_{\mathcal{B}}} \int \frac{d^3\mathbf{k}}{\left(2\pi\right)^3} \psi\left(\mathbf{k}\right) C_{ab} \frac{1}{\sqrt{2m}} \frac{1}{\sqrt{2m}} \mathcal{M}\left(e_-^a\left(\mathbf{k}\right) e_+^b\left(-\mathbf{k}\right) \to 2\gamma\right), \tag{7.5}$$

So far we haven't specified the initial angular momentum and spin of the positronium state. Let's now specialize to the case of positronium in the $\ell = 0$ spin-0 (singlet) state, ${}^1S_0$. Note that the decay of positronium from the spin-1 (triplet), ${}^3S_1$ state is not allowed since it violates C-parity, which is conserved by quantum electrodynamics. As a quick aside, let's explicitly see this. Under charge-parity a positronium ${}^{2s+1}S_J$ state transforms as:

$$\text{C} : \left|{}^{2s+1}S_J\right\rangle \to \text{C}\left|{}^{2s+1}S_J\right\rangle = \left(-1\right)^{\ell+s} \left|{}^{2s+1}S_J\right\rangle, \tag{7.6}$$

from which we immediately see that the ${}^1S_0$ state has even C-parity, $\text{C}|{}^1S_0\rangle = \left(+1\right)|{}^1S_0\rangle$, and the ${}^3S_1$ state has odd C-parity, $\text{C}|{}^3S_1\rangle = \left(-1\right)|{}^3S_1\rangle$. Under C-parity the $n\gamma$ final state transforms as:

$$\begin{aligned} C|\gamma\gamma\rangle &= \left(-1\right)\left(-1\right)|\gamma\gamma\rangle = \left(+1\right)|\gamma\gamma\rangle \\ C|\gamma\gamma\gamma\rangle &= \left(-1\right)\left(-1\right)\left(-1\right)|\gamma\gamma\gamma\rangle = \left(-1\right)|\gamma\gamma\gamma\rangle, \end{aligned} \tag{7.7}$$

and so we see that the ${}^3S_1$ cannot decay to 2γ since it violates C-parity.

Now, the amplitude for the process $e_-^a\left(k_1\right) e_+^b\left(k_2\right) \to \gamma\left(p_1\right)\gamma\left(p_2\right)$, which proceeds through a t-channel and a u-channel diagram, is given by:

$$\begin{aligned} i\mathcal{M}\left(e_-^a\left(\mathbf{k}_1\right) e_+^b\left(\mathbf{k}_2\right) \to 2\gamma\right) = &\left(-ie\right)^2 \epsilon_\mu^*\left(p_1\right) \epsilon_\nu^*\left(p_2\right) \\ &\times \bar{v}\left(k_2\right) \left[\gamma^\nu \frac{i\left(\not{k}_1 - \not{p}_1 + m\right)}{\left(k_1 - p_1\right)^2 - m^2} \gamma^\mu + \gamma^\mu \frac{i\left(\not{k}_1 - \not{p}_2 + m\right)}{\left(k_1 - p_2\right)^2 - m^2} \gamma^\nu\right] u\left(k_1\right). \end{aligned} \tag{7.8}$$

We can write the Dirac spinors in terms of two-component spinors as:

$$\begin{aligned} u\left(k_1\right) &= \begin{pmatrix} \sqrt{k_1 \cdot \sigma}\xi \\ \sqrt{k_1 \cdot \bar{\sigma}}\xi \end{pmatrix} \\ v\left(k_2\right) &= \begin{pmatrix} \sqrt{k_2 \cdot \sigma}\xi' \\ \sqrt{k_2 \cdot \bar{\sigma}}\xi' \end{pmatrix}, \end{aligned} \tag{7.9}$$

where $\sigma^\mu = (1, \sigma^i)$ and $\bar{\sigma}^\mu = (1, -\sigma^i)$, where σ^i are the three Pauli matrices. We can also write γ^μ as:

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}. \tag{7.10}$$

Let's take the nonrelativistic limit of the initial electron and positron, so the momenta are given by:

$$\begin{aligned} k_1^\mu &= k_2^\mu = (m, 0, 0, 0) \\ p_1^\mu &= (m, 0, 0, m) \\ p_2^\mu &= (m, 0, 0, -m), \end{aligned} \tag{7.11}$$

and the photon polarization vectors are given by:

$$\begin{aligned} \epsilon_\pm^\mu\left(p_1\right) &= \frac{1}{\sqrt{2}}\left(0, 1, \pm i, 0\right) \\ \epsilon_\pm^\mu\left(p_2\right) &= \frac{1}{\sqrt{2}}\left(0, -1, \pm i, 0\right). \end{aligned} \tag{7.12}$$

Plugging Equations 7.9 through 7.12 into Equation 7.8 we obtain:

$$i\mathcal{M}\left(e_-^a\left(\mathbf{k}_1\right) e_+^b\left(\mathbf{k}_2\right) \to 2\gamma\right) = ie^2 \epsilon_\mu^*\left(p_1\right) \epsilon_\nu^*\left(p_2\right) \xi'^\dagger \left(\sigma^\nu \sigma^3 \sigma^\mu - \sigma^\mu \sigma^3 \sigma^\nu\right) \xi. \tag{7.13}$$

Now we specify the polarizations of the final state photons and substitute $\xi\xi'^\dagger = \frac{1}{\sqrt{2}}\mathbf{1}$ in the spin-singlet state, which gives:

$$\begin{aligned} i\mathcal{M}_{++} &= -i\mathcal{M}_{--} = i2\sqrt{2}e^2 \\ i\mathcal{M}_{+-} &= -i\mathcal{M}_{-+} = 0, \end{aligned} \tag{7.14}$$

where the subscripts indicate the final state photon polarizations.

Now we can plug all of this back into Equation 7.5 to obtain:

$$\mathcal{M}_{\text{pol}}\left({}^1S_0 \to 2\gamma\right) = \sqrt{2M_{\mathcal{B}}} \int \frac{d^3\mathbf{k}}{(2\pi)^3} \psi\left(\mathbf{k}\right) \frac{1}{\sqrt{2m}} \frac{1}{\sqrt{2m}} \mathcal{M}_{\text{pol}}, \tag{7.15}$$

where the subscript pol indicates a specific configuration of photon polarizations ($++$, $--$, $+-$ or $-+$). Since the free-particle amplitude $\mathcal{M}_{\text{pol}}$ is independent of the momentum $\mathbf{k}$ the integral over $\mathbf{k}$ picks out the wavefunction at the origin, $\psi\left(0\right)$. Also we assume that the bound state has mass roughly equal to its constituent particles, $M_{\mathcal{B}} \simeq 2m$. Then Equation 7.15 gives:

$$\mathcal{M}_{\text{pol}}\left({}^1S_0 \to 2\gamma\right) = \frac{\psi\left(\mathbf{x}=0\right)}{\sqrt{m}} \mathcal{M}_{\text{pol}}. \tag{7.16}$$

Summing over all photon polarization configurations and squaring gives:

$$\sum_{\text{pol}} \left|\mathcal{M}_{\text{pol}}\left({}^1S_0 \to 2\gamma\right)\right|^2 = \frac{\left|\psi\left(\mathbf{x}=0\right)\right|^2}{m} \left(\left|\mathcal{M}_{++}\right|^2 + \left|\mathcal{M}_{--}\right|^2\right), \tag{7.17}$$

and plugging in the value:

$$\psi_{n=1,\ell=0,m=0}\left(r\right) = \sqrt{\frac{\left(\alpha m/2\right)^3}{\pi}} \exp\left(-\alpha m r\right) \tag{7.18}$$

evaluated at the origin, we get:

$$\begin{aligned} \sum_{\text{pol}} \left|\mathcal{M}_{\text{pol}}\left({}^1S_0 \to 2\gamma\right)\right|^2 &= \frac{\left(\alpha m/2\right)^3}{\pi} \frac{1}{m} \left(\left|\mathcal{M}_{++}\right|^2 + \left|\mathcal{M}_{--}\right|^2\right) \\ &= \frac{\left(\alpha m\right)^3}{\pi} \frac{1}{m} \left(2e^4\right) \\ &= 32\pi\alpha^5 m^2. \end{aligned} \tag{7.19}$$

Now we can calculate the decay width of the ${}^1S_0 \to 2\gamma$ process:

$$\begin{aligned}\Gamma\left({}^1S_0 \to 2\gamma\right) &= \frac{1}{2}\frac{1}{4m}\int \frac{d^3\mathbf{p}_1 d^3\mathbf{p}_2}{(2\pi)^6\, 2E_1 2E_2} \sum_{\text{pol}} \left|\mathcal{M}_{\text{pol}}\left({}^1S_0 \to 2\gamma\right)\right|^2 (2\pi)^4\, \delta^4\left(p_{\text{Ps}} - p_1 - p_2\right) \\ &= \alpha^5 m.\end{aligned} \tag{7.20}$$

where the additional factor of $1/2$ comes from the two identical photons in the final state.

As a quick sanity check let's do a quick parametric estimate to see if Equation 7.20 makes sense. We know the decay rate should be equal to something like:

$$\Gamma\left({}^1S_0 \to 2\gamma\right) \sim n_e v\sigma\left(e^+e^- \to 2\gamma\right), \tag{7.21}$$

where n_e is the number density of electrons in the positronium atom. When $\psi(\mathbf{r})$ is properly normalized:

$$\int_0^\infty \psi(\mathbf{r})\, d^3\mathbf{r} = 1 \tag{7.22}$$

then we can think of it as being closely related to n_e:

$$|\psi(\mathbf{r})| \sim \frac{1}{R^{3/2}} \sim n_e^{1/2}. \tag{7.23}$$

We have to specify the volume scale, and it makes sense to choose the Bohr radius, r_B, since outside that radius $\psi(\mathbf{r})$ exponentially loses support, so we can write:

$$\int_0^{r_B} \psi(\mathbf{r})\, d^3\mathbf{r} \sim 1, \tag{7.24}$$

and

$$|\psi(0)| \sim \frac{1}{r_B^{3/2}} \sim n_e^{1/2}. \tag{7.25}$$

So far we have:

$$\begin{aligned}\Gamma\left({}^1S_0 \to 2\gamma\right) &\sim \frac{1}{r_B^3} v\sigma\left(e^+e^- \to 2\gamma\right) \\ &\sim (m\alpha)^3\, v\sigma\left(e^+e^- \to 2\gamma\right).\end{aligned} \tag{7.26}$$

If we take the limit of nonrelativistic electron and positron annihilating into photons, then we know the tree-level cross section scales like:

$$\sigma\left(e^+e^- \to 2\gamma\right) \sim \frac{\alpha^2}{s}\frac{|\mathbf{p}_f|}{|\mathbf{p}_i|} \sim \frac{\alpha^2}{m^2c^4}\frac{mc}{mv} \tag{7.27}$$

and the decay rate scales as:

$$\begin{aligned}\Gamma\left({}^1S_0 \to 2\gamma\right) &\sim (\alpha m)^3 v \frac{\alpha^2}{m^2}\frac{m}{mv} \\ &\sim \alpha^5 m,\end{aligned} \tag{7.28}$$

which is equal to what we get in Equation 7.20.

7.2.2 Positronium formation

Now let's think about the radiative capture of an electron and positron to form positronium. I'll start this section by describing the radiative capture of any two oppositely charged leptons to the bound state, say for example in the case of the capture of electron and muon into muonium, because it's useful to have those general formulas. Then it will be easy to specialize the result to the case of positronium formation. Because in this case there is no particle creation or destruction like there was in the previous case of positronium decay, in the nonrelativistic limit we can use single-particle quantum mechanics. In this section I'll follow [182].

The process we're considering is:

$$\mu^+\left(p_+\right) + e^-\left(p_-\right) \to \gamma\left(k\right) + \left(\mu^+e^-\right)\left(p\right). \tag{7.29}$$

The differential probability of the radiative transition is:

$$\begin{aligned}dW =& \frac{TV}{(2\pi)^2\, 2\omega}\delta\left(E_i - E_f - \omega\right)\delta^3\left(\mathbf{Q} - \mathbf{k} - \mathbf{p}\right) \times \\ & e^2\left|\bar{M}\right|^2 \omega^2 d\omega d\Omega d^3p,\end{aligned} \tag{7.30}$$

where T is the interaction time, V is the normalized volume, ω is the energy of the

radiated photon, E_i is the energy of the electron and muon at infinity, E_f is the final energy taking into account the binding energy of the muonium, $\mathbf{Q}$ is the center-of-mass momentum, $d\Omega$ is the solid angle element of the photon, and $\bar{M}$ is the reduced matrix element:

$$\bar{M} = \int d^3\mathbf{r}\Psi_f^*(\mathbf{r}) \left(\frac{\mathbf{a}\cdot\nabla}{m_e} e^{i\mathbf{k}_1\cdot\mathbf{r}} + \frac{\mathbf{a}\cdot\nabla}{m_\mu} e^{-i\mathbf{k}_2\cdot\mathbf{r}} \right) \Psi_i(\mathbf{r}), \tag{7.31}$$

where $\mathbf{a}$ is the photon polarization vector, m_e is the mass of the electron, m_μ is the mass of the muon, $\mathbf{k}_1 = m_\mu\mathbf{k}/(m_\mu + m_e)$, and $\mathbf{k}_2 = m_e\mathbf{k}/(m_\mu + m_e)$.

We are interested in the relative motion of the muon and electron so we can go to the center-of-mass frame where $\mathbf{Q} = 0$. Working in the center-of-mass frame, we will now relabel the initial electron momentum $\mathbf{p}_- \equiv \mathbf{p}$, and the initial muon momentum $\mathbf{p}_+ = -\mathbf{p}$.

The initial-state wavefunction Ψ_i describes an initially free electron and muon in the Coulomb field, and has the form of a superposition of the plane wave and an outgoing spherical Coulomb wave at large r, given by:

$$\Psi_i(\mathbf{p},\mathbf{r}) = \exp(\pi\zeta/2)\,\Gamma(1-i\zeta)\,\mathrm{F}(i\zeta,1,i(pr-\mathbf{p}\cdot\mathbf{r}))\,e^{i\mathbf{p}\cdot\mathbf{r}}, \tag{7.32}$$

where $\zeta = \alpha m_e m_\mu/p(m_\mu + m_e)$, $\alpha = 1/137$ is the fine structure constant, and F is the confluent hypergeometric function of the first kind, ${}_1F_1$. $\Psi_i(\mathbf{p},\mathbf{r})$ is normalized the same way as a plane wave:

$$\int d^2\mathbf{r}\Psi_i^*(\mathbf{p},\mathbf{r})\,\Psi_i\left(\mathbf{p}',\mathbf{r}\right) = (2\pi)^3\,\delta^3\left(\mathbf{p}-\mathbf{p}'\right). \tag{7.33}$$

We consider capture into the ground state of muonium, so the final-state wavefunction Ψ_f is given by:

$$\Psi_f(r) = \sqrt{\frac{\eta^3}{\pi}} e^{-r\eta}, \tag{7.34}$$

where $\eta = \frac{\alpha m_e m_\mu}{m_e+m_\mu} \equiv \zeta p$.

For a constant vector $\mathbf{a}$:

$$g\mathbf{a}\cdot\nabla f = \nabla\cdot(gf\mathbf{a}) - \nabla g f\cdot\mathbf{a}. \tag{7.35}$$

which we can use to recast the integrand of Equation 7.31 for $\bar{M}$:

$$\begin{aligned}
&\Psi_f^*(r)\,\mathbf{a}\cdot\nabla\left[\left(\frac{e^{i\mathbf{k}_1\cdot\mathbf{r}}}{m_e}+\frac{e^{-i\mathbf{k}_2\cdot\mathbf{r}}}{m_\mu}\right)e^{i\mathbf{p}\cdot\mathbf{r}}\mathrm{F}\left(i\zeta,1,i\left(pr-\mathbf{p}\cdot\mathbf{r}\right)\right)\right]\\
&=\nabla\cdot\left[\Psi_f^*\left(\frac{e^{i\mathbf{k}_1\cdot\mathbf{r}}}{m_e}+\frac{e^{-i\mathbf{k}_2\cdot\mathbf{r}}}{m_\mu}\right)e^{i\mathbf{p}\cdot\mathbf{r}}\mathrm{F}\left(i\zeta,1,i\left(pr-\mathbf{p}\cdot\mathbf{r}\right)\right)\mathbf{a}\right]\\
&-\nabla\Psi_f^*\left(\frac{e^{i\mathbf{k}_1\cdot\mathbf{r}}}{m_e}+\frac{e^{-i\mathbf{k}_2\cdot\mathbf{r}}}{m_\mu}\right)e^{i\mathbf{p}\cdot\mathbf{r}}\mathrm{F}\left(i\zeta,1,i\left(pr-\mathbf{p}\cdot\mathbf{r}\right)\right)\cdot\mathbf{a}.
\end{aligned} \tag{7.36}$$

The overall derivative term will vanish in the integral over $d^3\mathbf{r}$. Also, $\nabla\Psi_f^* = -\eta\sqrt{\frac{\eta^3}{\pi}}e^{-\eta r}\frac{\mathbf{r}}{r}$. Plugging Equations 7.36 and 7.34 into Equation 7.31 gives:

$$\begin{aligned}
\bar{M} =& N\int d^3\mathbf{r}e^{-\eta r}\frac{\mathbf{a}\cdot\mathbf{r}}{r}\left(\frac{1}{m_e}e^{-i\mathbf{k}_1\cdot\mathbf{r}}+\frac{1}{m_\mu}e^{i\mathbf{k}_2\cdot\mathbf{r}}\right)\times\\
&e^{i\mathbf{p}\cdot\mathbf{r}}\mathrm{F}\left(i\zeta,1,i\left(pr-\mathbf{p}\cdot\mathbf{r}\right)\right),
\end{aligned} \tag{7.37}$$

where I've defined $N=\sqrt{\eta^5/\pi}\exp\left(\pi\zeta/2\right)\Gamma\left(1-i\zeta\right)$.

We can make use of the following formula [182]:

$$\begin{aligned}
&\int d^3\mathbf{r}\frac{1}{r}e^{i(\mathbf{p}-\mathbf{k})\cdot\mathbf{r}-\eta r}\mathrm{F}\left(i\zeta,1,i\left(pr-\mathbf{p}\cdot\mathbf{r}\right)\right)=\\
&4\pi\frac{\left[k^2+(\eta-ip)^2\right]^{-i\zeta}}{\left[(\mathbf{p}-\mathbf{k})^2+\eta^2\right]^{1-i\zeta}},
\end{aligned} \tag{7.38}$$

and differentiating with respect to $\mathbf{p}$ we get:

$$\begin{aligned}
&\int d^3\mathbf{r}e^{-\eta r}\frac{\mathbf{a}\cdot\mathbf{r}}{r}e^{-i\mathbf{k}\cdot\mathbf{r}}e^{i\mathbf{p}\cdot\mathbf{r}}\mathrm{F}\left(i\zeta,1,i\left(pr-\mathbf{p}\cdot\mathbf{r}\right)\right)=\\
&8\pi\frac{i\left(1-i\zeta\right)\mathbf{a}\cdot\mathbf{p}}{\left[(\mathbf{p}-\mathbf{k})^2+\eta^2\right]^{2-i\zeta}}\left[k^2+(\eta-ip)^2\right]^{-i\zeta},
\end{aligned} \tag{7.39}$$

which we can use to write Equation 7.37 as:

$$\begin{aligned}\bar{M} = & i\left(1-i\zeta\right)N8\pi\left(\mathbf{a}\cdot\mathbf{p}\right)\\ & \left[\frac{\left[\mathbf{k}_1^2+\left(\eta-ip\right)^2\right]^{-i\zeta}}{m_e\left[\left(\mathbf{p}-\mathbf{k}_1\right)^2+\eta^2\right]^{2-i\zeta}}+\frac{\left[\mathbf{k}_2^2+\left(\eta-ip\right)^2\right]^{-i\zeta}}{m_\mu\left[\left(\mathbf{p}+\mathbf{k}_2\right)^2+\eta^2\right]^{2-i\zeta}}\right].\end{aligned} \tag{7.40}$$

Now let's simplify this in the case of positronium formation, where $m = m_\mu = m_e$, $\zeta = \frac{\alpha m}{2p}$, and $\eta = \frac{\alpha m}{2}$. In this case Equation 7.40 becomes:

$$\bar{M}_{\mathrm{Ps}} = i\left(1-i\frac{\alpha m}{2p}\right)N8\pi\left(\mathbf{a}\cdot\mathbf{p}\right)2\frac{\left[\frac{\mathbf{k}^2}{4}+\frac{\left(m\alpha\right)^2}{4}\right]^{-i\frac{\alpha m}{2p}}}{m\left[\frac{\mathbf{k}^2}{4}+\frac{\left(m\alpha\right)^2}{4}\right]^{2-i\frac{\alpha m}{2p}}}, \tag{7.41}$$

and this further reduces to:

$$\bar{M}_{\mathrm{Ps}} = \frac{\alpha m}{p}N8\pi\left(\mathbf{a}\cdot\mathbf{p}\right)\frac{1}{m\left[\frac{\mathbf{k}^2}{4}+\frac{\left(m\alpha\right)^2}{4}\right]^2}, \tag{7.42}$$

where:

$$N = \sqrt{\frac{\left(m\alpha\right)^5}{2^5\pi}}\exp\left(\frac{\pi}{2}\frac{m\alpha}{2p}\right)\Gamma\left(1-i\frac{m\alpha}{2p}\right). \tag{7.43}$$

The differential cross section for capture to the two-particle final state of positronium and the photon which is emitted during radiative capture is given by:

$$d\sigma = \frac{dw}{v_r} \tag{7.44}$$

where dw is the probability per unit volume per unit time of the radiative transition of electron and positron to positronium:

$$dw = \frac{dW}{TV}, \tag{7.45}$$

with dW given by Equation 7.30, and where v_r is the final-state reduced energy:

$$v_r = \frac{p\left(m+\mu\right)}{m\mu} = \frac{2p}{m}, \tag{7.46}$$

where in the second equality I applied the case of positronium ($\mu = m$). Plugging Equations 7.46, 7.45, 7.30, and 7.42 into Equation 7.44, integrating and taking the limit that the momentum of the initial particles is very small ($p \ll \alpha m/2$), we get:

$$\sigma = \frac{2^9 \pi^2 \alpha^3 e^{-4}}{3p^2} \left[1 - \frac{10p^2 + 12\omega^2}{15\eta^2}\right], \tag{7.47}$$

where the photon energy ω is defined from energy conservation:

$$\omega + \omega^2/4m = p^2/m + \alpha^2 m/4. \tag{7.48}$$

The result in Equation 7.47 will be a helpful comparison to gain insight into the more complicated case of radiative capture into bound states of dark matter particles which inhabit a more complex gauge group, which we'll get into in the next section and in Chapter 8.

7.3 Nonabelian "positronium"

Let's now generalize the discussion above to the case in which bound states can consist of multiple species of particles which inhabit a nonabelian gauge group. This will give us the tools we need to analyze the formation of bound states in more general dark-sector models. In this section I will leave the detailed structure of the gauge group which dark matter inhabits arbitrary, and work with wavefunctions Ψ which are general multiplets of two-particle states whose interactions may be described, in the nonrelativistic limit, by a general matrix potential V. Later in Chapter 8 we will apply the discussion in this section to a specific dark sector model.

Like we did in the last section, we'll work in the nonrelativistic limit in which the formation of bound states can be described by nonrelativistic quantum mechanics. In this limit the gauge bosons which mediate interactions between dark matter particles are either much lighter or much heavier than the energy scale of interest and can be integrated out to give a nonlocal potential describing the interactions. The nonlocal potential V in general describes interactions between multiple species of dark sector

particles and takes the form of a matrix. Under the interactions described by V the multiple species of dark sector particles transform as a part of a multiplet of states Ψ. Since we are interested in the formation of bound states it is most convenient to work with wavefunctions Ψ which are multiplets of two-particle states, and a matrix potential V which describes interactions between these two-particle states via gauge-boson exchange. To give a concrete example, in Chapter 8 we will apply this discussion to the case of SU (2) triplet dark matter (the wino) which has three species of dark sector particles (χ_0, χ_+, χ_-), and we will consider transitions between pair states $\chi_0\chi_0$ and $\chi_+\chi_-$, so we will consider two-component wavefunctions $\Psi = (\chi_0\chi_0, \chi_+\chi_-)^{\mathrm{T}}$ interacting via a 2×2 matrix potential V. But for now let's work with a general n-component multiplet of two-particle states Ψ interacting via a general $n \times n$ matrix potential V.

Let's also work in the "Coulombic" limit in which the gauge bosons are massless. In this limit the matrix potential V takes the form:

$$V(r) = \frac{\alpha}{r}\bar{V}, \tag{7.49}$$

where α is the dark sector coupling and $\bar{V}$ is an $n \times n$ matrix determined from the symmetry group under which Ψ transforms. The potential V enters the Hamiltonian via the Schrödinger equation:

$$i\partial_t\Psi = H^0\Psi = \left[-\frac{\nabla_X^2}{4M_\chi} - \frac{\nabla_r^2}{M_\chi} + V(r)\right]\Psi, \tag{7.50}$$

where X is the center-of-mass coordinate and r is the relative spatial coordinate. In the Coulombic limit, all species of dark sector particles included in Ψ are taken to have the same mass M_χ.

H^0 in Equation 7.50 is the unperturbed Hamiltonian before we take into account the interaction of pair states Ψ with the surrounding gauge field responsible for radiative transitions between continuum and bound states. The full Hamiltonian is given by:

$$H = H^0 + V_{\mathrm{rad}}, \tag{7.51}$$

where V_{rad} describes the interaction of particles Ψ with the gauge field associated with the radiated particle. For example, in the case of the wino which we will discuss in Chapter 8, one component of the pair states in Ψ contains particles charged under electromagnetism, and the particle being radiated in V_{rad} is the photon; V_{rad} in this case takes the form (in the broken regime in which the gauge bosons are not massless):

$$\begin{aligned} V_{\text{rad}} = &\left(-\sum_n \frac{e_n}{M_\chi} \mathbf{A}(\mathbf{x}_n) \cdot \mathbf{p}_n + \sum_n \frac{e_n^2}{2m_n} \mathbf{A}(\mathbf{x}_n)^2 \right) \mathbb{P}_{CC} \\ &+ \left(i\sqrt{2}\, e\, \alpha_W \mathbf{A}(0) \cdot \hat{\mathbf{r}}\, e^{-m_W r} \right) \mathbb{P}_{NC}, \end{aligned} \tag{7.52}$$

where n labels the relevant charged particle, e_n is the signed EM coupling, and e the coupling to a positive charge. The projectors $\mathbb{P}_{CC,NC}$ enforce that the interactions only couple the charged sector of the two-particle Hilbert space to itself, and the neutral sector to the charged, respectively. The two-particle states we're examining are all net-neutral, and here by "charged" I'm referring to a two-particle state composed of oppositely charged constituents. $\mathbb{P}_{CC}$ picks out the components where the radiated photon couples to the two-particle states directly, and the term in Equation 7.52 proportional to $\mathbb{P}_{CC}$ has the familiar form of the Hamiltonian for a charged particle moving in an electromagnetic field described by the vector potential $\mathbf{A}(\mathbf{r}, t)$ in the Coulomb gauge (where the electromagnetic scalar potential is set to zero), with additional nonabelian gauge structure contained in $\mathbb{P}_{CC}$.

The projector $\mathbb{P}_{NC}$ in Equation 7.52 picks out the components where the radiated photon couples to the potential line between any two two-particle states. The term proportional to $\mathbb{P}_{NC}$ accounts for the ability of the potential itself to emit radiation, and is an intrinsically nonabelian effect not present in the case of electromagnetism. In the case of the wino that term captures the ability of the W boson mediating the interaction between the components of Ψ picked out by $\mathbb{P}_{NC}$ to emit radiation.

Let's now return to discussing the general case of nonabelian positronium. In order to describe radiative transitions between two-particle states in Ψ we proceed like usual and treat V_{rad} like a time-dependent perturbation to the time-independent H^0.

Let's make the dipole approximation, which works in the limit that the wavelength of the radiated particle in radiative transition is much larger than the size of the atom, so that the gauge field $\mathbf{A}(\mathbf{r}, t)$ of the radiated particle can be taken to be uniform over the entire two-particle state. The dipole approximation also implies the weak field case in which the term proportional to $\mathbf{A}(\mathbf{x}_n)^2$ in Equation 7.52 is small compared to the term proportional to $\mathbf{A}(\mathbf{x}_n)$ and can be neglected. This means we'll only treat the emission or absorption of one radiated particle at a time. Later, in Appendix D.3.2 we discuss the case of two-particle emission.

The expression for the capture and transition cross sections in the dipole approximation is given by:

$$\begin{aligned}&\sigma v_{\rm rel}\,(\text{continuum} \to \text{bound}) \text{ or } \Gamma\,(\text{bound} \to \text{bound}) \\ &= \frac{2\,\alpha_{\rm rad}}{\pi}\frac{k}{M_\chi^2}\int d\Omega_k \left|\epsilon(\hat{k},\sigma)\cdot\int d^3r\left[\Psi_f^*(\mathbf{r})\hat{C}_1\nabla_{\mathbf{r}}\,\Psi_i(\mathbf{r}) + \frac{\alpha M_\chi}{2}\Psi_f^*(\mathbf{r})\hat{C}_2\hat{\mathbf{r}}\Psi_i(\mathbf{r})\right]\right|^2,\end{aligned} \tag{7.53}$$

where the coupling of the radiated particle, $\alpha_{\rm rad}$, may be different than the full gauge coupling α. $\hat{C}_1$ and $\hat{C}_2$ are the general analogues of $\mathbb{P}_{CC}$ and $\mathbb{P}_{NC}$, respectively, from the wino example in Equation 7.52 above. $\hat{C}_1$ and $\hat{C}_2$ describe the gauge structure of the capture operator (respectively, how the radiated particle couples to the two-particle states directly, and how it couples to the potential line between any two two-particle states); the term proportional to $\hat{C}_1$ in Equation 7.53 has the familiar form of an electric dipole transition between two charged states which directly couple to the gauge boson responsible for the radiative interaction. The term proportional to $\hat{C}_2$ accounts for the ability of the potential itself to emit radiation.

For example, in the wino case, as we'll see in Chapter 8, $\Psi(\mathbf{r})$ is a two-component vector, $(\psi_N(\mathbf{r}), \psi_C(\mathbf{r}))$; $\alpha_{\rm rad}$ is the coupling for electromagnetism, while the gauge coupling α becomes α_W. The photon couples only to the $\chi^+\chi^-$ component of the two-particle states, so $\hat{C}_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$; the photon also couples to the exchanged W boson, which couples the $\chi^0\chi^0$ component of one state to the $\chi^+\chi^-$ component of the

other state, and $\hat{C}_2 = \sqrt{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

The positronium case we looked at in Section 7.2.2 can be recovered by setting $\hat{C}_1 = 1$, $\hat{C}_2 = 0$, considering only the single two-particle state e^+e^-, and setting $\alpha_{\text{rad}} = \alpha_{\text{EM}}$.

The final ingredient we need in this generalized calculation will be the initial condition for the unscattered part of the continuum state, corresponding to a plane wave at large radii, which we'll describe by the unit vector $\mathbf{I}$. For example, in the wino case, where we are interested in the scattering of particles initially in the $\chi^0\chi^0$ state, $\mathbf{I} = (1, 0)$. In the case where the potential has a finite range, the unperturbed plane-wave will then have the asymptotic form $\mathbf{I}e^{i\mathbf{p}\cdot\mathbf{r}}$, up to (anti)symmetrization factors if the particles are identical. In the true Coulombic case where the potential has infinite range, the wavefunction is distorted by the potential even as $r \to \infty$; nonetheless, we can still describe the large-r boundary condition by the same constant vector, as we will discuss in Section D.3.4.

7.3.1 The scattering wavefunction

Given a potential matrix $V(r) = -\frac{\alpha}{r}\bar{V}$, suppose that diagonalizing $\bar{V}$ yields eigenvalues λ_i, $i = 1..n$, and corresponding orthonormal eigenvectors η_i. Then let us write $\Psi(\mathbf{r}) = \sum_i \eta_i \phi_i(\mathbf{r})$, where $\phi_i(\mathbf{r})$ is a scalar function. Then the Schrödinger equation separates into equations for each i:

$$\frac{p^2}{2\mu}\phi_i(\mathbf{r}) = -\frac{1}{2\mu}\nabla^2\phi_i(\mathbf{r}) - \frac{\lambda_i\alpha}{r}\phi_i(\mathbf{r}), \tag{7.54}$$

where μ is the reduced mass. That is, each ϕ_i is a solution to the Schrödinger equation with a Coulomb potential with coupling $\lambda_i\alpha$; if λ_i is positive, the potential is attractive, while if λ_i is negative, the potential is repulsive.

For a pair of distinguishable fermions with equal and opposite charges, interacting via an attractive Coulomb potential with coupling α, the solution to the Schrödinger equation with incoming wave corresponding to the plane wave $e^{i\mathbf{p}\cdot\mathbf{r}}$ is given by (e.g.

[182]):

$$\phi(\alpha;\mathbf{r}) = e^{\pi\zeta/2}\Gamma(1-i\zeta)\ {}_1F_1(i\zeta,1,i(pr-\mathbf{p}\cdot\mathbf{r}))e^{i\mathbf{p}\cdot\mathbf{r}} \tag{7.55}$$

(we recognize this from Equation 7.32 in the positronium case), where $\zeta = \alpha\mu/p = \alpha/v_{\rm rel}$ and F is the hypergeometric function. Then we can write the general solution to the Schrödinger equation as:

$$\Psi(\mathbf{r}) = \sum_i A_i\eta_i\phi(\lambda_i\alpha;\mathbf{r}) \tag{7.56}$$

Asymptotically, this solution will correspond to a plane wave and a scattered spherical wave, and by construction the plane wave will be given by $e^{i\mathbf{p}\cdot\mathbf{r}}\sum_i A_i\eta_i$.[1] Thus to impose the boundary condition that determines the initial state of the two interacting particles, we must impose $\mathbf{I} = \sum_i A_i\eta_i$. By orthonormality of the η_i, this is equivalent to requiring $A_i = \mathbf{I}\cdot\eta_i$. Thus our full solution has the form:

$$\Psi(\mathbf{r}) = \sum_i (\mathbf{I}\cdot\eta_i)\eta_i\phi(\lambda_i\alpha;\mathbf{r}) \tag{7.57}$$

For a pair of *identical* fermions this spatial wavefunction must be symmetrized (antisymmetrized) for spin-singlet (spin-triplet) configurations. This corresponds to omitting the terms in the partial-wave expansion with odd (spin-singlet) or even (spin-triplet) L, and multiplying the remaining wavefunction by $\sqrt{2}$. To make the connection to positronium clearer, we will generally use the normalization for distinguishable fermions throughout this section. When the initial state consists of indistinguishable fermions, as in the wino case discussed in Chapter 8, all cross sections should be multiplied by 2, but only contributions to the capture rate where the final state has odd $L+S$ should be included (since the photon emission changes $L+S$ by 1, and the initial state must be purely $L+S$-even by spin-statistics arguments).

[1] There is a subtlety here, as mentioned previously, in that the asymptotic solution is not exactly a plane wave when the potential has infinite range. See section D.3.4 for discussion.

7.3.2 The bound state wavefunction

A similar approach is valid for the bound state wavefunction, with the following exception: repulsive Coulomb potentials do not support bound states, so only positive eigenvalues will produce bound-state solutions. Again, in general we will have $\Psi(\mathbf{r}) = \sum_{\lambda_i>0} \eta_i \phi_i(\mathbf{r})$, and so a given bound state will be characterized by the usual quantum numbers n, l, m but also by λ_i. To compute the total capture rate, we should sum over all such possible final states, but we may also be interested in capture into a specific state.

We can therefore write the (properly normalized) bound state wavefunctions as:

$$\Psi^i_{nlm}(\mathbf{r}) = Y_{lm}(\theta,\phi)\eta_i R_{nl}(\lambda_i\alpha; r), \tag{7.58}$$

where $Y_{lm}(\theta,\phi)$ are the spherical harmonics and $R_{nl}(\lambda_i\alpha; r)$ is the hydrogenic radial wavefunction with α replaced by $\lambda_i\alpha$ and m_e replaced by μ, given by:

$$R_{nl}\left(\lambda_i\alpha; r\right) = \left[\left(\frac{2\alpha\lambda_i\mu}{n}\right)^3 \frac{(n-l-1)!}{2n\,(n+l)!}\right]^{1/2} e^{-\mu\lambda_i\alpha r/n}\left(\frac{2\alpha\lambda_i\mu r}{n}\right)^l \mathrm{L}^{2l+1}_{n-l-1}\left(\frac{2\alpha\,\lambda_i\mu r}{n}\right). \tag{7.59}$$

7.3.3 The capture rate

Let's define the reduced matrix element $\bar{M}$ for radiative capture following Section 7.2.2 and [182]; in the dipole approximation we have:

$$\begin{aligned}\bar{M} &\equiv \frac{1}{\mu}\epsilon\cdot\int d^3r\left[\Psi^*_f(\mathbf{r})\hat{C}_1\nabla_{\mathbf{r}}\,\Psi_i(\mathbf{r}) + \alpha\mu\Psi^*_f(\mathbf{r})\hat{C}_2\hat{\mathbf{r}}\Psi_i(\mathbf{r})\right]\\ &= \frac{1}{\mu}\epsilon\cdot\sum_i(\mathbf{I}\cdot\eta_i)\eta^\dagger_f\left[\hat{C}_1\int d^3rY^*_{lm}(\theta,\phi)R^*_{nl}(\lambda_f\alpha;r)\nabla_{\mathbf{r}}\phi(\lambda_i\alpha;\mathbf{r})\right.\\ &\left.+\alpha\mu\hat{C}_2\int d^3rY^*_{lm}(\theta,\phi)R^*_{nl}(\lambda_f\alpha;r)\hat{\mathbf{r}}\phi(\lambda_i\alpha;\mathbf{r})\right]\eta_i,\end{aligned} \tag{7.60}$$

where the final bound state is characterized by quantum numbers nlm and eigenvalue λ_f, and i sums over the eigenvalues of the potential experienced by the initial

continuum state. This reduced matrix element is related to the capture cross section by:

$$\sigma v_{\rm rel}\left(\text{continuum} \to \text{bound}\right) = \frac{\alpha_{\rm rad} k}{2\pi} \int d\Omega_k \left|\bar{M}\right|^2 . \tag{7.61}$$

To evaluate $\bar{M}$, we can substitute the explicit wavefunctions for the continuum and bound states (Eqs. 7.59 and 7.55) into eq. 7.60. Noting that:

$$\nabla\ {}_1F_1\left[i\zeta, 1, i(qr - \mathbf{q}\cdot\mathbf{r})\right] = -\zeta q(\hat{\mathbf{r}} - \hat{\mathbf{q}})\ {}_1F_1\left[1 + i\zeta, 2, i(qr - \mathbf{q}\cdot\mathbf{r})\right], \tag{7.62}$$

we obtain:

$$\bar{M} = \sum_i (\mathbf{I}\cdot\eta_i)\eta_f^\dagger \left[\alpha e^{\frac{\pi\alpha\lambda_i\mu}{2p}} \Gamma\left(1 - i\frac{\alpha\lambda_i\mu}{p}\right)\epsilon\right] \cdot \left[-\lambda_i \hat{C}_1 \mathbf{K}^1 + \hat{C}_2 \mathbf{K}^2\right] \eta_i \tag{7.63}$$

where $\mathbf{K}^1$, $\mathbf{K}^2$ are given by the integrals:

$$\begin{aligned}
\mathbf{K}^1 &= \int d^3r Y_{lm}^*(\theta,\phi) R_{nl}^*(\lambda_f\alpha; r) e^{i\mathbf{p}\cdot\mathbf{r}} (\hat{\mathbf{r}} - \hat{\mathbf{p}})\ {}_1F_1(1 + i\alpha\lambda_i\mu/p, 2, i(pr - \mathbf{p}\cdot\mathbf{r})), \\
\mathbf{K}^2 &= \int d^3r Y_{lm}^*(\theta,\phi) R_{nl}^*(\lambda_f\alpha; r) e^{i\mathbf{p}\cdot\mathbf{r}} \hat{\mathbf{r}}\ {}_1F_1(i\alpha\lambda_i\mu/p, 1, i(pr - \mathbf{p}\cdot\mathbf{r})).
\end{aligned} \tag{7.64}$$

Let us choose a coordinate system where $\mathbf{p}$ points in the z-direction. Then the angular integral will set to zero any transitions to states with $|m| > 1$, since the initial continuum state has only $m = 0$ components and the single-photon dipole transition

requires $|\Delta m| \leq 1$. Performing the angular integrals over ϕ for $m = 0, \pm 1$ yields:

$$\begin{aligned}
\mathbf{K}^1_{m=0} &= 2\pi\hat{z}\sqrt{\frac{(2l+1)}{4\pi}} \\
&\quad \times \int r^2 dr \sin\theta d\theta P_l^{0*}(\cos\theta) R^*_{nl}(\lambda_f \alpha; r) e^{ipr\cos\theta}(\cos\theta - 1)\ {}_1F_1\left(1 + i\frac{\alpha\lambda_i\mu}{p}, 2, ipr(1-\cos\theta)\right), \\
\mathbf{K}^1_{m=\pm 1} &= \pi\sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}}(\hat{x} \mp i\hat{y}) \\
&\quad \times \int r^2 dr d\theta \sin^2\theta P_l^{m*}(\cos\theta) R^*_{nl}(\lambda_f \alpha; r) e^{ipr\cos\theta}\ {}_1F_1\left(1 + i\frac{\alpha\lambda_i\mu}{p}, 2, ipr(1-\cos\theta)\right), \\
\mathbf{K}^2_{m=0} &= 2\pi\hat{z}\sqrt{\frac{(2l+1)}{4\pi}} \\
&\quad \times \int r^2 dr \sin\theta d\theta P_l^{0*}(\cos\theta) R^*_{nl}(\lambda_f \alpha; r) e^{ipr\cos\theta}\cos\theta\ {}_1F_1\left(\frac{i\alpha\lambda_i\mu}{p}, 1, ipr(1-\cos\theta)\right), \\
\mathbf{K}^2_{m=\pm 1} &= \pi\sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}}(\hat{x} \mp i\hat{y}) \\
&\quad \times \int r^2 dr \sin^2\theta d\theta P_l^{m*}(\cos\theta) R^*_{nl}(\lambda_f \alpha; r) e^{ipr\cos\theta}\ {}_1F_1\left(\frac{i\alpha\lambda_i\mu}{p}, 1, ipr(1-\cos\theta)\right).
\end{aligned} \tag{7.65}$$

If we restrict ourselves to $n = 1, 2$ bound states, which are the most deeply bound and generally have the largest capture cross sections, the required integrals can be evaluated in the low-$v_{\rm rel}$ limit using the results of appendix D.6. For the possible choices of nlm we obtain the following results for the $\mathbf{K}$ integrals:

$$\begin{aligned}
\mathbf{K}^1_{100}: &\quad -8\sqrt{\pi}(\alpha\lambda_f\mu)^{-3/2} e^{-2\lambda_i/\lambda_f}\hat{z} \\
\mathbf{K}^1_{200}: &\quad -32\sqrt{2\pi}(\alpha\lambda_f\mu)^{-3/2} e^{-4\lambda_i/\lambda_f}\left[\frac{2\lambda_i}{\lambda_f} - 1\right]\hat{z} \\
\mathbf{K}^1_{210}: &\quad -16\sqrt{2\pi}(\alpha\lambda_f\mu)^{-3/2} e^{-4\lambda_i/\lambda_f}\left[\frac{4\lambda_i}{\lambda_f} - 1\right]\hat{z} \\
\mathbf{K}^1_{21(\pm 1)}: &\quad 16\sqrt{2\pi}(\alpha\lambda_f\mu)^{-3/2} e^{-4\lambda_i/\lambda_f}\left(\frac{\mp\hat{x} + i\hat{y}}{\sqrt{2}}\right).
\end{aligned} \tag{7.66}$$

$$
\begin{aligned}
\mathbf{K}^2_{100}: &\quad 8\sqrt{\pi}(\alpha\lambda_f\mu)^{-3/2}e^{-2\lambda_i/\lambda_f}\frac{\lambda_i}{\lambda_f}\hat{z} \\
\mathbf{K}^2_{200}: &\quad 32\sqrt{2\pi}(\alpha\lambda_f\mu)^{-3/2}e^{-4\lambda_i/\lambda_f}\left[\frac{\lambda_i}{\lambda_f}\left(\frac{4\lambda_i}{\lambda_f}-3\right)\right]\hat{z} \\
\mathbf{K}^2_{210}: &\quad 16\sqrt{2\pi}(\alpha\lambda_f\mu)^{-3/2}e^{-4\lambda_i/\lambda_f}\left[1-\frac{4\lambda_i}{\lambda_f}+8\left(\frac{\lambda_i}{\lambda_f}\right)^2\right]\hat{z} \\
\mathbf{K}^2_{21(\pm1)}: &\quad 16\sqrt{2\pi}(\alpha\lambda_f\mu)^{-3/2}e^{-4\lambda_i/\lambda_f}\left[1-\frac{4\lambda_i}{\lambda_f}\right]\left(\frac{\mp\hat{x}+i\hat{y}}{\sqrt{2}}\right).
\end{aligned}
\tag{7.67}
$$

Accordingly we can write the matrix element for capture into the nlm bound state as:

$$
\begin{aligned}
\bar{M} = 8\sqrt{\pi}\alpha\epsilon\cdot(\hat{r}_m)^*\sum_i \mathbf{I}\cdot\eta_i(\alpha\lambda_f\mu)^{-3/2}e^{-2n\lambda_i/\lambda_f}e^{\frac{\pi\alpha\lambda_i\mu}{2p}}\Gamma\left(1-i\frac{\alpha\lambda_i\mu}{p}\right) \\
\times\eta_f^\dagger\left[\lambda_i\hat{C}_1 u_{nlm}+\hat{C}_2 v_{nlm}\right]\eta_i,
\end{aligned}
\tag{7.68}
$$

where $\hat{r}_0 = z$, $(\hat{r}_{-1}-\hat{r}_1)/\sqrt{2} = \hat{x}$, and $i\,(\hat{r}_{-1}+\hat{r}_1)/\sqrt{2} = \hat{y}$, and u_{nlm}, v_{nlm} are given by:

$$
\begin{aligned}
&u_{100}: 1, \quad u_{200}: 4\sqrt{2}\,[2\lambda_i/\lambda_f - 1]\,, \quad u_{210}: 2\sqrt{2}[4\lambda_i/\lambda_f-1], \quad u_{21(\pm1)}: -2\sqrt{2}, \\
&v_{100}: \lambda_i/\lambda_f, \quad v_{200}: 4\sqrt{2}\left[\frac{\lambda_i}{\lambda_f}\left(\frac{4\lambda_i}{\lambda_f}-3\right)\right], \\
&v_{210}: 2\sqrt{2}\left[1-\frac{4\lambda_i}{\lambda_f}+8\left(\frac{\lambda_i}{\lambda_f}\right)^2\right], \quad v_{21(\pm1)}: -2\sqrt{2}\left[\frac{4\lambda_i}{\lambda_f}-1\right].
\end{aligned}
\tag{7.69}
$$

It is useful to note how the gamma-function term in eq. 7.68 scales in the small-$\mathbf{p}$ limit. We use the relation:

$$
\left|\Gamma\left(1-i\zeta\right)e^{\frac{\pi\zeta}{2}}\right|^2 = \frac{\pi\zeta}{\sinh\left(\pi\zeta\right)}e^{\pi\zeta} \to \begin{cases} 2\pi\zeta,\ \zeta\gg 1 \\ -2\pi\zeta e^{2\pi\zeta},\ \zeta\ll -1 \end{cases}
\tag{7.70}
$$

It follows that at small momentum, where $\alpha\mu/p$ becomes large, all contributions to $\bar{M}$ from i corresponding to $\lambda_i < 0$ are suppressed by a factor of order $e^{\pi\alpha\mu\lambda_i/p}$. Physically, this is because $\lambda_i < 0$ indicates a repulsive potential, and at low velocities the resulting wavefunctions have very little overlap with the bound states. In the

calculations that follow we will neglect these exponentially suppressed contributions.

To compute the cross section from this matrix element, we need only to sum over the final photon polarization states, performing the integral $\int d\Omega_k |\bar{M}|^2$ for each. Writing $\bar{M} = A\epsilon \cdot \hat{r}_{-m}$, and following the same procedure as Eqs. 8.21-8.22 in Chapter 8 (and noting that $\int d\Omega_k \sin^2\theta_k = \int d\Omega_k \left(1 - \frac{1}{2}\sin^2\theta_k\right) = 8\pi/3$), we find that $\int d\Omega_k |\bar{M}|^2 = \frac{8\pi}{3}|A|^2$.

Thus finally we obtain the capture cross section:

$$\sigma v_{\rm rel}\,(\text{continuum} \to \text{bound}) = \frac{2^8 \pi \alpha_{\rm rad} k}{3} \times \left| \sum_i \mathbf{I} \cdot \eta_i \alpha (\alpha \lambda_f \mu)^{-3/2} e^{-2n\lambda_i/\lambda_f} e^{\frac{\pi\alpha\lambda_i\mu}{2p}} \Gamma\left(1 - i\frac{\alpha\lambda_i\mu}{p}\right) \eta_f^\dagger \left[\lambda_i \hat{C}_1 u_{nlm} + \hat{C}_2 v_{nlm}\right] \eta_i \right|^2 . \tag{7.71}$$

Recall this is the capture cross section for distinguishable fermions in the initial state; if the initial state consists of identical fermions, this cross section should be multiplied by 2 where the final state has $L + S$ odd, and set to zero when the final state has $L + S$ even. Since 1/4 of all particle pairs are spin-singlet ($S = 0$) and 3/4 are spin-triplet ($S = 1$), this corresponds to a spin-averaged cross section related to eq. 7.71 by a factor of 3/2 for final states of even L, and a factor of 1/2 for final states of odd L.

Chapter 8

Capture and Decay of Electroweak WIMPonium

This chapter is from [4], *by Pouya Asadi, Matthew Baumgart, me, Emmett Krupczak, and Tracy R. Slatyer. The original published version of* [4] *contained a sign error, we intend to publish an erratum in the future. We correct that error in this text.*

8.1 Introduction

Although the particle content of the Standard Model (SM) does not contain a DM candidate, it is straightforward to add new degrees of freedom with the necessary properties: the so-called Weakly Interacting Massive Particles (WIMPs) [183, 184]. The coldness of Cold DM implies we are immersed in a sea of slowly-moving particles, and giving the DM couplings of similar strength to the SM (perturbative but not ultra-weak), the correct DM relic abundance is naturally obtained for masses of $\mathcal{O}(1\text{ TeV})$ [185].

Thus, we are moved to consider the physics of heavy, slow particles, with simulations suggesting a mean velocity $\langle v \rangle \sim 10^{-3}$ [186]. We can therefore work in the nonrelativistic limit, setting up an effective field theory for the DM in analogy with NRQCD and NRQED [187–190]. In this limit, the interactions of the DM with long-range force carriers (e.g. electroweak bosons, dark-sector photons) are properly

treated as a nonperturbative, nonlocal, but instantaneous potential. This leads to the well-known phenomenon of Sommerfeld enhancement in DM annihilations [191–197]. The potential deforms the two-particle DM wavefunction near the origin, leading to large deviations from a calculation treating the initial state as a plane wave. Schematically, for annihilation of DM in an s-wave state, the annihilation rate goes as

$$\sigma v = \Gamma |\psi(0)|^2, \tag{8.1}$$

where Γ is the perturbatively-calculated, short-distance annihilation rate, and $\psi(0)$ is the wavefunction of the two-particle DM-DM state at the origin. In the limit that the potential turns off, $|\psi(0)| = 1$, and we recover the perturbative result.

The wavefunction in eq. 8.1 is for a positive-energy scattering state. However, the spectrum of the long-range potential may also include negative-energy bound states. When the binding energy for one of these states approaches zero, it induces a large resonant enhancement to the scattering-state wavefunction at the origin $\psi(0)$, and hence to the Sommerfeld enhancement [191, 198].

The presence of bound states in the spectrum can have effects beyond an enhanced Sommerfeld factor. In particular, capture of DM particles into these bound states gives rise to an alternative annihilation channel for the DM, analogous to formation and annihilation of positronium, which in some circumstances may dominate over the Sommerfeld-enhanced direct annihilation. Transitions into and between bound states can also produce particles at energies parametrically suppressed relative to the DM mass. There has been considerable interest in the literature in such WIMPonium states and their properties [199–213]; however, most of the work on indirect signatures to date has focused on models where the DM couples to only a single mediator (a dark photon or scalar), and where the mass of the mediator is sufficiently light that the resulting potential can be approximated by the Coulomb potential.

In this chapter, we extend these considerations to the electroweak potential, where these simplifying assumptions do not apply: the DM is generally part of a multiplet of states of similar masses, and these states may couple to both massive and massless

gauge bosons. DM transforming under $SU(2)_L \times U(1)_Y$ is known to receive large Sommerfeld corrections for masses above $\sim$1 TeV, with the first resonance – signaling the presence of a bound state – occurring for the SU(2) triplet, or wino, at a DM mass ~ 2.5 TeV. Interestingly, it is for similar wino masses (~ 3 TeV) that the present-day abundance of DM is naturally obtained, i.e. the wino is a thermal relic. Unfortunately, as several groups have independently shown [189, 190, 214–217], thermal wino DM is now in severe tension with constraints on gamma-ray lines from the HESS experiment [218]. Nonetheless, we will consider here the phenomenology of heavy wino bound states, with the following motivations:

- The results of any indirect detection experiment come with large astrophysical uncertainties due to the poorly-constrained DM halo density profile. Thus, we should continue to explore new phenomena that could allow for additional constraints.

- Even if the wino is not a thermal relic, nature could still realize a high-scale MSSM as means of resolving most of the hierarchy problem along with providing grand-unification. Current and future Cherenkov telescopes like CTA and HAWC will set limits on DM masses up to 100 TeV and 1000 TeV, respectively, albeit with sensitivity less than the rates predicted for electroweak DM [219–221]. We should explore the physics of electroweak WIMPs in this regime, even if the mechanism for providing their relic density is unspecified.

- Dark-sector models have provided a WIMP DM candidate unshackled by the specific couplings of the SM. It is worth considering scenarios where the hidden-sector gauge group is more complex than the dark U(1) of simple dark photon models (e.g. [193, 222]), and the DM can be part of a nontrivial multiplet. In such scenarios, the dark gauge group may feature large hierarchies between force carrier masses, just as we see in $SU(2)_L \times U(1)_Y$. Our wino calculations are therefore a toy model for studying bound state physics in the presence of nearly-degenerate matter fields that may experience both long- and short-range forces, where the particle radiated in the formation of the bound state may be

different from the force carrier primarily responsible for the potential. Lastly, the nonabelian potential contains richer structures, including the ability of force carriers to emit radiation and the possibility of multiple attractive and repulsive channels.

In section 8.2, we discuss winos in the nonrelativistic limit, the potential that governs their evolution, and its spectrum of bound and continuum states. In section 8.3, we develop the necessary formalism to calculate the rate of bound state formation by radiative capture (cf. figure 8-1) in the case of wino DM, as well as the rates for

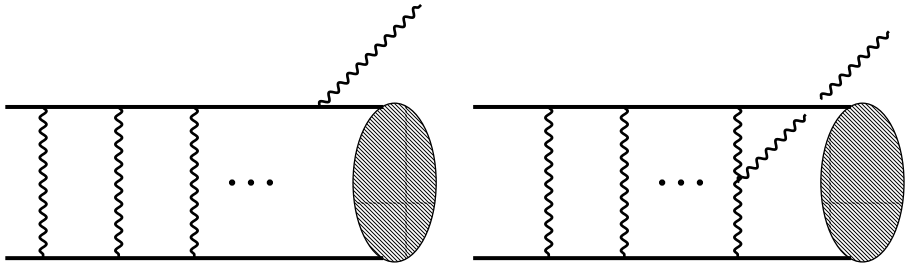

Figure 8-1: WIMPs exchange a ladder of weak gauge bosons, which gives rise to a non-local potential in the nonrelativistic limit. Finally, the dipole emission of a single photon can convert the initial, positive-energy scattering state to a negative-energy bound state, WIMPonium. (**Right:**) Since the potential contains charged force carriers, $W^{\pm}$, they can also emit radiation to capture into the bound state.

bound states to transition among themselves and annihilate to SM particles (cf. figure 8-2). In section 8.4, we apply the results of section 8.3, present our numerical results, and discuss observational possibilities, before presenting our conclusions in section 8.5.

Finally, in appendices D.1-D.2 we detail our numerical procedures for computing wavefunctions. In appendices D.3-D.4 we discuss two illuminating toy problems: (1) the nonabelian analogue of positronium, which we began discussing in Section 7.3, and which the wino case approaches in the limit of very high DM mass, and (2) the bound states of the Hulthén potential, which provides an analytically tractable approximation to the Yukawa potential and hence allows us to study the effects of reducing the range of the potential in a simple system. In appendix D.5 we discuss

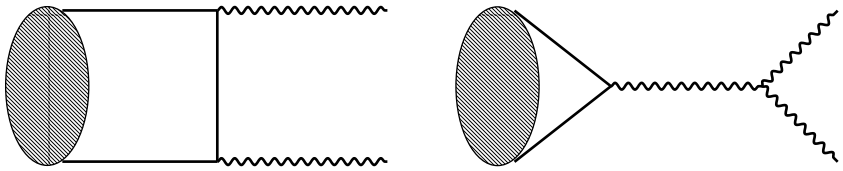

Figure 8-2: Since we consider WIMPonium constituents that are charged under the electroweak gauge group, its lifetime is set by weak-scale physics. Excited states typically transition to deeper bound states, but the deeper bound states will annihilate to SM particles. We note that if the WIMPonium is in a ${}^1S_0$ state, s-channel annihilation through a gauge boson is forbidden.

how to translate existing results in the literature to our formalism for WIMPonium annihilation, and in appendix D.6 we derive and present several useful integrals.

8.2 Winos in the nonrelativistic limit

The specific WIMP whose capture and annihilation we compute is an $\mathrm{SU}(2)_L$ triplet Majorana fermion, denoted χ^a, with mass M_χ:

$$\mathcal{L} = i\chi^{a\,\dagger}\left(\bar{\sigma}^\mu\partial_\mu + i\,g\,\bar{\sigma}^\mu W_\mu^b\,T_{ac}^b\right)\chi^c - \frac{1}{2}M_\chi(\chi^a\chi^a + \text{h.c.}) \tag{8.2}$$

We refer to it as the wino even though it is the only field beyond the SM we include. One can think of it as either a minimal extension of the SM to provide DM or as the lightest supersymmetric particle (LSP) of an otherwise decoupled SUSY sector. Although we are interested in the multi-TeV regime, it is necessary to include the effects of electroweak symmetry breaking in the W and Z masses and to work in the wino mass eigenstate basis, with the neutralino $\chi^0 = \chi^3$, and the chargino $\chi^\pm = \frac{1}{\sqrt{2}}(\chi^1 \mp i\,\chi^2)$. There is a small, but important, mass splitting between the charged and neutral states, arising from radiative corrections from SM fields:

$$\delta M \equiv M_{\chi^\pm} - M_{\chi^0} \;=\; 165\,\text{MeV}, \tag{8.3}$$

and we will take $M_{\chi^0} \equiv M_\chi$ [214, 223].

8.2.1 General considerations and symmetries

In the nonrelativistic limit of electroweak WIMPs, the interactions of the fermions with gauge bosons whose momenta have "potential" scaling, $(E, \mathbf{p}) \sim (M_\chi v^2,\, M_\chi v)$, can be integrated out to give a nonlocal potential. Furthermore, for all of our processes of interest – Sommerfeld-enhanced annihilation, capture into bound states, transitions between bound states, and annihilation of bound states to SM fields – it is more useful to work with two-particle states, rather than single-particle quantum fields. If the state has positive energy, it will be a plane-wave-normalized, two-particle state.[1] If it is a negative-energy bound state, then it will have the standard single-particle normalization (i.e. integrating over the norm-squared of the position-space wavefunction gives 1). We will detail a formalism below that can handle both cases.

Whether the state is positive or negative-energy, the potential due to gauge boson exchange experienced by a two-particle state with even-$L+S$ is:

$$V_{L+S\,\text{even}}(r) = \begin{pmatrix} 0 & -\sqrt{2}\alpha_W \frac{e^{-m_W r}}{r} \\ -\sqrt{2}\alpha_W \frac{e^{-m_W r}}{r} & 2\delta M - \frac{\alpha}{r} - \alpha_W c_W^2 \frac{e^{-m_Z r}}{r} \end{pmatrix}. \tag{8.4}$$

Here L and S denote the total orbital and spin angular momentum quantum numbers for the two-particle state, respectively (we will generally use upper-case letters to denote the quantum numbers of an arbitrary two-particle state, while using lower-case nlm to label the quantum numbers of the bound states). For a detailed derivation of this potential and the construction of two-body quantum-mechanical states starting from the fully relativistic quantum field theory, see [191, 216]. This potential enters the Hamiltonian via,

$$i\partial_t \Psi = H^0\,\Psi = \left[-\frac{\nabla_X^2}{4M_\chi} - \frac{\nabla_r^2}{M_\chi} + V(r)\right]\Psi, \tag{8.5}$$

[1] There is a subtlety in this normalization for states consisting of identical fermions, which must be appropriately antisymmetrized, as we will discuss below.

where X is the center of mass coordinate and Ψ is a two-component wavefunction,

$$\Psi = \begin{pmatrix} \psi_N \, (\equiv \chi^0\chi^0) \\ \psi_C \, (\equiv \chi^+\chi^-) \end{pmatrix}. \tag{8.6}$$

The nonzero off-diagonal terms in $V_{L+S\,\mathrm{even}}(r)$ mix the charged and neutral components, so we must evolve them simultaneously.

As noted by [194, 197], in this basis even the lowest-order nonrelativistic potential is dependent on the spin and angular momentum of the two-particle states. The potential of eq. 8.4 applies to spin-singlet states with even L and spin-triplet states with odd L. For spin-singlet states with odd L or spin-triplet states with even L, so $L + S$ is odd, the wavefunction is symmetric and there can be no two-particle state consisting of the identical neutral fermions $\chi^0\chi^0$; consequently, the potential is non-zero only for the charged two-particle state $\chi^+\chi^-$,

$$V_{L+S\,\mathrm{odd}}(r) = \begin{pmatrix} 0 & 0 \\ 0 & 2\delta M - \frac{\alpha}{r} - \alpha_W c_W^2 \frac{e^{-m_Z r}}{r} \end{pmatrix}. \tag{8.7}$$

We will not consider in this work the on-shell emission of W or Z bosons. Since the parametric size of the binding energy $E_n \lesssim \mathcal{O}(\alpha_W^2 M_\chi)$, this process is kinematically forbidden for DM lighter than ~ 100 TeV.[2] Off-shell production of W and Z bosons which subsequently decay is allowed, but will be strongly suppressed relative to processes involving the emission of a photon, by a factor $\sim \frac{\alpha_W}{\pi}\left(\frac{E_n}{m_W}\right)^4$. Accordingly, we only consider the $Q = 0$ sector of two-particle states (i.e. the total electric charge of the state is zero).

Electric dipole transitions with single-photon emission do not flip the spin of the two-particle state, but change its angular momentum by $\Delta L = \pm 1$. Since the initial

[2]We can estimate this a bit more precisely. In this high mass limit, electroweak symmetry is approximately restored. Thus, we just need the binding energy for a Coulomb potential with coupling α_W, $E_n = -\frac{\alpha_W^2 M_\chi}{4n^2}$. For our dominant single-photon capture to s-wave, $n = 1$, and sufficient energy to produce an on-shell Z requires $M_\chi = 321$ TeV. Numerically, we find with our full potential that the crossover occurs at 329 TeV. This is higher than the mass regime we study in detail, which goes up to 300 TeV.

two-particle state, far from the point of interaction, will consist of neutral identical fermions, it must have even $L+S$ (the s-wave piece is purely spin-singlet; the p-wave piece is purely spin-triplet, etc). The two-particle state resulting from a single photon emission will then have odd $L+S$, and so must be purely $\chi^+\chi^-$.

Computing the capture rate, $\sigma v\,(\chi^0\chi^0 \to \text{WIMPonium} + \gamma)$, will be very similar to the standard quantum-mechanical calculation of radiative transitions between hydrogenic bound states. Instead of our initial state being negative-energy with a compact wavefunction, it will be a positive-energy solution to the Schrödinger equation, eq. 8.5, with potential given by eq. 8.4, and energy $M_\chi v_{\text{rel}}^2/4$ in the center-of-momentum (CM) frame. Additionally, we will have to account for the fact that the potential itself is charged. Although our Hamiltonian requires numerical analysis due to the Yukawa terms, one can calculate analytically the pure QED process for e^+e^- to bind into positronium after electric dipole emission [182, 224]. In appendix D.3, we present exact analytic results for the SU(2) analog of positronium, with potentials corresponding to those in eqs. 8.4 and 8.7 in the limit δM, m_W, $m_Z \to 0$.

For the WIMP bound states, we will need to find the negative-energy solutions with the single-component potential in eq. 8.7. Bound states supported by the potential of eq. 8.4 do exist, but cannot be accessed from our initial state by single-photon emission; nonetheless, we will discuss their properties. We can obtain parametric intuition for the effect of short-range potentials by studying the Hulthén potential, a close cousin of the Yukawa. We collect detailed results on this potential in appendix D.4.

Note that our convention for zero energy is set by two χ^0 particles far apart at rest; the $2\delta M$ term in eq. 8.7 can therefore set the energies of some of the $\chi^+\chi^-$ bound states to be positive, although they would have negative energy in the alternate convention where zero is set by the constituents' rest masses at infinity. We will briefly discuss the behavior of these "positive-energy" bound-states, although we do not expect them to be important for generic parameters.

8.2.2 The bound state spectrum in the high-mass limit

Let us consider the spectrum of bound states present in the case where the $\mathrm{SU}(2)_L$ symmetry is unbroken, the force carriers are massless, and there is no mass splitting between the charginos and neutralinos. The potential matrices simplify to:

$$V(r) = \frac{\alpha_W}{r}\begin{pmatrix} 0 & -\sqrt{2} \\ -\sqrt{2} & -1 \end{pmatrix}, \quad L+S \text{ even}, \qquad V(r) = \frac{\alpha_W}{r}\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}, \quad L+S \text{ odd}. \tag{8.8}$$

In this limit, the Hamiltonian can be diagonalized and the solutions to the Schrödinger equation can be immediately written down in terms of the eigenstates of the Coulomb potential. For the case of odd $L+S$ this is trivial. For the case of even $L+S$, the matrix potential has eigenvalues $-\lambda_i \frac{\alpha_W}{r}$ where $\lambda_1 = 2$, $\lambda_2 = -1$; the corresponding orthonormal eigenvectors are $\eta_1 = \begin{pmatrix}\sqrt{\frac{1}{3}} & \sqrt{\frac{2}{3}}\end{pmatrix}$, $\eta_2 = \begin{pmatrix}-\sqrt{\frac{2}{3}} & \sqrt{\frac{1}{3}}\end{pmatrix}$. The general solution to the Schrödinger equation for even $L+S$ is given by (eq. 7.56):

$$\Psi(\mathbf{r}) = \sum_i A_i \eta_i \, \phi(\lambda_i \alpha_W; \mathbf{r}) \tag{8.9}$$

where $\phi(\lambda_i \alpha_W; \mathbf{r})$ is the scalar function solving the Schrödinger equation for a Coulomb potential, with coupling $\lambda_i \alpha_W$.

In particular, bound states cannot be supported by a repulsive Coulomb potential, so all bound states with even $L+S$ will be of the form $\phi(\lambda_1 \alpha_W; \mathbf{r})\eta_1$. In this case $\lambda_1 = 2$, so the states have binding energies corresponding to a Coulomb potential with coupling $2\alpha_W$ and reduced mass $\mu = M_\chi/2$, i.e. $E_n = \alpha_W^2 M_\chi / n^2$. The bound states with odd $L+S$ form a separate tower with wavefunctions of the form $\phi(\alpha_W; \mathbf{r}) \begin{pmatrix} 0 & 1 \end{pmatrix}$. Accordingly, their binding energies are $E_n = \alpha_W^2 M_\chi / 4n^2$.

This means, for example, that the lowest-lying spin-singlet $L = 1$ states are more weakly bound than the lowest-lying spin-singlet $L = 2$ states; the former have odd $L+S$ and so have energy $E_2 = \alpha_W^2 M_\chi/16$, whereas the latter have even $L+S$ and so have binding energy $E_3 = \alpha_W^2 M_\chi/9$. Consequently, $n = 2$ states may have multiple

open decay channels, to $n = 3$ states as well as $n = 1$. The low-lying states for both spin-singlet and spin-triplet configurations are summarized in figure 8-3.

	Spin-Singlet Spectrum	Spin-Triplet Spectrum
$\lvert\frac{E_n}{M_\chi \alpha_W^2}\rvert = \frac{1}{144}$		6D
$\lvert\frac{E_n}{M_\chi \alpha_W^2}\rvert = \frac{1}{100}$	5P	5D
$\lvert\frac{E_n}{M_\chi \alpha_W^2}\rvert = \frac{1}{64}$	4P	4S 4D
$\lvert\frac{E_n}{M_\chi \alpha_W^2}\rvert = \frac{1}{36}$	3P 6D	3S 3D
$\lvert\frac{E_n}{M_\chi \alpha_W^2}\rvert = \frac{1}{25}$	5D	5P
$\lvert\frac{E_n}{M_\chi \alpha_W^2}\rvert = \frac{1}{16}$	4S 2P 4D	2S 4P
$\lvert\frac{E_n}{M_\chi \alpha_W^2}\rvert = \frac{1}{9}$	3S 3D	3P
$\lvert\frac{E_n}{M_\chi \alpha_W^2}\rvert = \frac{1}{4}$	2S	1S 2P
$\lvert\frac{E_n}{M_\chi \alpha_W^2}\rvert = 1$	1S	

Figure 8-3: The energy spectrum of bound states for each spin configuration in the large M_χ limit. The lowest four states for $L = 0$ (*blue*), $L = 1$ (*red*), and $L = 2$ (*green*) are included. For each spin configuration, the couplings in the L-even and L-odd potentials differ by a factor of two in the high-mass limit. This distorts the order of the bound states compared to a hydrogen atom. For the spin-singlet tower the $L = 1$ bound states are pulled up to higher energies, while for spin-triplet they have been pushed down to lower energies.

8.2.3 The bound state spectrum for all masses

Beyond this high-mass limit, we must proceed numerically. We approximate the bound states as a linear combination of Coulombic wavefunctions, and solve for the coefficients of these basis states. We exploit the fact that our bound-state potential (eq. 8.7) is rotationally symmetric, and thus L is still a good quantum number. This allows us to expand the solution for the full potential with fixed quantum numbers (n, l) in terms of hydrogenic states with the same L, but summed over radial eigenvalues from $L - 1$ up to some $n_{\max}$, beyond which the calculation is numerically stable. Determining the coefficients of this expansion is a straightforward linear algebra exercise (cf. eq. D.31). Furthermore, in the limits $m_Z/M_\chi \to 0, \infty$, we recover

a Coulombic potential with coupling α_W, α, respectively. The details of our method are presented in appendix D.2.[3]

We display the resulting spectrum of bound states in figure 8-4. We will use these numerical wavefunctions to compute transition rates involving the bound states: between bound states, from bound states to SM particles, and from the initial free particles to the bound states.

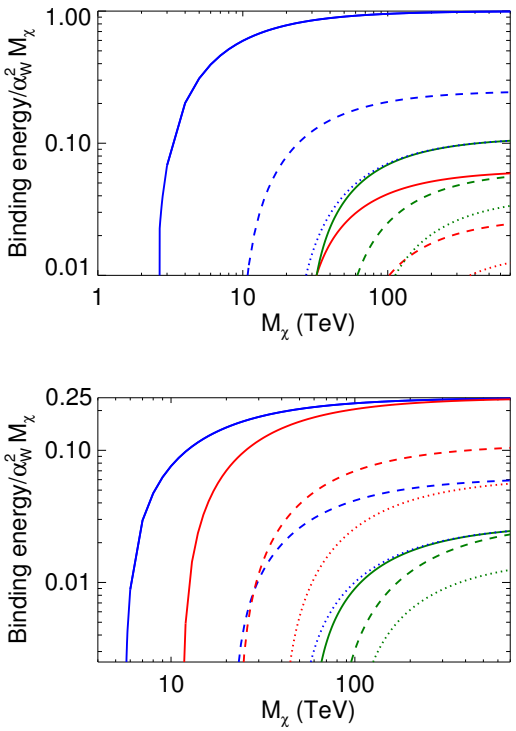

Figure 8-4: SU(2)-triplet WIMPonium binding energies relative to $\alpha_W^2 M_\chi$. **Left:** Spectrum of spin-singlet states. **Right:** Spectrum of spin-triplet states. *Blue*, *red*, and *green* lines denote s, p and d-wave bound states respectively, with solid, dashed, dotted lines denoting the ranking in n quantum number, where we have included only states with the three lowest n for each partial wave. In the high-mass limit, both potentials asymptote to Coulombic behavior, with effective coupling $2\alpha_W$ for $L+S$-even states and α_W for $L+S$-odd states, and we recover the expected binding energies.

We observe that the first negative-energy, spin-singlet bound state appears in the spectrum at $M_\chi \approx 2.6$ TeV, and the first negative-energy, spin-triplet bound state at

[3]We thank S. Thomas for his help in developing this numerical procedure.

$M_\chi \approx 5.6$ TeV. However, the spin-singlet bound state cannot be accessed by single-dipole-photon capture from the initial state, since the L=1, S=0 continuum state is not populated by the identical fermionic DM particles (due to Fermi statistics). Spin-singlet configurations thus do not contribute to the single-photon capture rate until $M_\chi \gtrsim 25$ TeV, where the first accessible spin-singlet p-wave state appears.

At high DM masses, the spectrum of bound states in figure 8-4 converges to the limiting Coulombic case discussed above and displayed in figure 8-3. At lower DM masses, however, the relative ordering of the states can shift.

8.3 Formation, transitions and annihilation of WIMP-onium

Given an initial population of free neutralinos, bound states can form via radiative capture with the emission of a photon. Those bound states may subsequently decay to lower-energy states in the spectrum, or annihilate into SM particles. In this section we will develop the formalism for computing the relevant rates.

8.3.1 Continuum-bound and bound-bound transitions

We calculate the rate for transitions between either continuum or bound states, with single photon emission, using time-ordered perturbation theory. Our discussion parallels the treatment of radiative transition rates in [225]. In the WIMP sector, our wavefunctions are eigenstates of the Hamiltonian constructed with $V(r)$ in eq. 8.4 for the initial state, and eq. 8.7 for the final bound state:

$$\begin{aligned} H^0_{L+S\,\text{even}}\,\Psi_i &= \frac{M_\chi v_{\text{rel}}^2}{4}\Psi_i \\ H^0_{L+S\,\text{odd}}\,\Psi_f\left[{}^{2S+1}L_J\right] &= E_n\,\Psi_f\left[{}^{2S+1}L_J\right], \end{aligned} \tag{8.10}$$

where E_n is the binding energy and v_{rel} is the relative velocity of the two particles.

Up to corrections that go like $M_\chi v_{\text{rel}}^4$, capture is kinematically possible if $E_n <$

$M_\chi v_{\rm rel}^2/4$. For the small velocities we consider, generally only bound states with $E_n < 0$ will be kinematically accessible. Accounting for the chargino's and W's ability to radiate an on-shell photon, we obtain our full Hamiltonian in Coulomb gauge,

$$\begin{aligned} H &= H^0 + V_{\rm rad.} \\ V_{\rm rad.} &= \left(-\sum_n \frac{e_n}{M_\chi} \mathbf{A}(\mathbf{x}_n) \cdot \mathbf{p}_n + \sum_n \frac{e_n^2}{2m_n} \mathbf{A}(\mathbf{x}_n)^2 \right) \mathbb{P}_{CC} \\ &\quad + \left(i\sqrt{2}\, e\, \alpha_W \mathbf{A}(0) \cdot \hat{\mathbf{r}}\, e^{-m_W r} \right) \mathbb{P}_{NC}, \end{aligned} \tag{8.11}$$

where n labels the relevant chargino, with e_n the signed EM coupling, and e the coupling to a positive charge. The relative spatial coordinate in our Hamiltonian, eq. 8.5, is given as

$$\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2, \tag{8.12}$$

and in the CM frame, $\mathbf{x}_1 + \mathbf{x}_2 = 0$. The projectors $\mathbb{P}_{CC,NC}$ enforce that the interactions only couple the charged sector of the two-particle Hilbert space to itself, and neutral sector to the charged, respectively. For example, in the two-component Hilbert space of the $L + S$ even sector, the $\mathbb{P}_{CC}$ term only acts on the charged component, ψ_C (*cf.* eq. 8.6) of Ψ_i, which we will denote $\psi_{i,C}$. This is just the standard, single-particle electric dipole coupling, familiar from atomic physics.

The $\mathbb{P}_{NC}$ term accounts for the ability of the potential itself to emit electric dipole radiation. The explicit α_W in this contribution makes it appear naively suppressed relative to the chargino dipole emission. However, the $\mathbf{p}_n/M_\chi$ in the $\mathbb{P}_{CC}$ term brings in an expectation value of the WIMP velocity, $v \sim \alpha_W$, where the matrix element is supported by the bound state wavefunction. Thus, both terms in Eq. 8.11 are *a priori* the same order and must be included. We find that the $\mathbb{P}_{NC}$ term is typically numerically dominant by a factor of a few. The dipole emission off the potential is an intrinsically nonabelian effect, known in the NRQCD literature, whose origin we now review [226–228]. It arises from the process shown in figure 8-5, since our constituent WIMPs exchange charged force carriers. This contributes to the electric

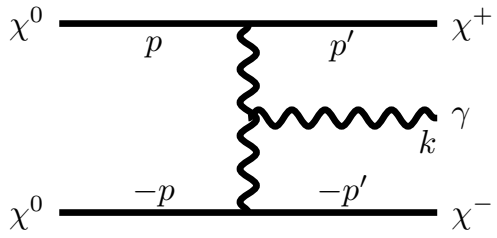

Figure 8-5: WIMPs exchange ladder $W^\pm$ bosons, which can radiate ultrasoft, electric dipole photons. This contributes to the $\Delta L = \pm 1$ capture rate to form WIMPonium. Integrating out the $W\pm$ in this amplitude generates the effective operators eq. 8.13, 8.14. In the quantum mechanical, two-body Hamiltonian this gives rise to the $\mathbb{P}_{NC}$ term in eq. 8.11.

dipole transition, and thus the Fermi statistics and angular momentum considerations in section 8.2.1 continue to hold. The $W^\pm$ exchange connects the $\chi^0\chi^0$ state to the $\chi^+\chi^-$. For the capture process, since the initial state contains both components, the amplitude for dipole emission off the potential involves only the neutral component of the initial wavefunction, $\psi_{i,N}$. Unsurprisingly, in the nonrelativistic effective field theory (NREFT) description, integrating out the potential gauge boson in figure 8-5 gives a nonlocal operator that resembles the potential, but with an additional ladder propagator and a dipole coupling to the photon,

$$\mathcal{L}_{pu} \;=\; \frac{2g^2}{\left((\mathbf{p}'-\mathbf{p})^2 + m_W^2\right)^2}(\mathbf{p}'-\mathbf{p})\cdot(eA_k)\left[\left(\chi_{-p}^{0\,c\,\dagger}\,\chi_{p'}^{+}\right)\left(\chi_{-p'}^{-\,\dagger}\,\chi_{-p}^{0}\right) \;+\; h.c.\right]. \quad (8.13)$$

The coupling g is that of $\mathrm{SU}(2)_L$, while e is that of electromagnetism. As these are nonrelativistic fields, each contains only creation or annihilation operators.[4] The term explicitly written destroys two χ^0s and creates a $\chi^\pm$ pair, while the conjugate term does the opposite. The field $\chi^{0\,c} = -i\sigma^2\chi^{0\,*}$. One can find a more complete description of the field content and how it connects to two-particle, quantum mechanical states in the appendix of [216]. This photon has $(k^0,\,\vec{k}) \;\sim\; (M_\chi v^2,\, M_\chi v^2)$,

[4]In NRQCD, the analogous operator describing gluon emission off the quark-antiquark potential is $\mathcal{L}_{pu}^{QCD} = \frac{2\,i\,g_s^2 f^{ABC}}{(\mathbf{p}'-\mathbf{p})^4}(\mathbf{p}'-\mathbf{p})\cdot(g_s A_k^C)\left[\psi_{p'}^\dagger\,T^A\psi_p\right]\left[\chi_{-p'}^\dagger\,\overline{T}^B\chi_{-p}\right]$, which in position space is $\mathcal{L}_{pu}^{QCD} = \alpha_s f^{ABC}\int d^3r\;\left[\psi^\dagger\,T^A\psi\right](x+\vec{r})\;\left[\chi^\dagger\,\overline{T}^B\chi\right](x)\,\hat{r}\cdot\big(g_s A^C(t,0)\big)$ [226–228].

and is thus "ultrasoft" in the NREFT terminology. Since our scattering and bound-state wavefunctions are written in position space, it is easier to work with the Fourier transform of the operator in eq. 8.13,

$$\mathcal{L}_{pu} = i\,\alpha_W \int d^3r \left[\chi^{0\,c\,\dagger}\chi^{+}\right](x+\vec{r})\ \left[\chi^{-\,\dagger}\chi^{0}\right](x)\,\hat{r}\,e^{-m_W r}\big(eA(t,0)\big)\ +\ \text{h.c.}, \qquad (8.14)$$

where we note that the softness of the photon spatial momentum sets its position coordinate to the origin of space. With the position-space operator, it is straightforward to use the quantum mechanical state definitions in the appendix of [216] to convert $\mathcal{L}_{pu}$ to a term in $V_{\text{rad.}}$, the perturbative Hamiltonian that acts on our two-particle states, eq. 8.11.

We treat $V_{\text{rad.}}$ as a perturbation, and capture from single-photon emission occurs at first order, with the following S-matrix element,

$$\begin{aligned} S_{i,\,f\gamma} \;=\; & 2\pi i\,\delta\big[M_\chi v^2/4 - E_n - k - P_{\text{BS}}^2/(4M_\chi)\big]\ \Bigg(\sum_n \frac{e_n}{M_\chi}\,\langle \Psi_f\big[^{2S+1}L_J\big]\ \gamma(k)|\ \mathbf{A}(\mathbf{x}_n)\cdot\mathbf{p}_n\,|\psi_{i,C}\rangle \\ & + i\sqrt{2}\,e\,\alpha_W\,\langle \Psi_f\big[^{2S+1}L_J\big]\ \gamma(k)|\ e^{-m_W r}\mathbf{A}(0)\cdot\hat{\mathbf{r}}\,|\psi_{i,N}\rangle\ \Bigg) \end{aligned} \qquad (8.15)$$

where P_{BS} is the momentum of the bound state and $\mathbf{k}$ is the momentum of the emitted photon. We use nonrelativistic normalization for the initial and final states. In the center of mass (CM) frame, $E_{\text{BS}} \sim k^2/4M_\chi \sim \alpha_W^4 M_\chi$, and is thus suppressed relative to the other energy scales; henceforth we drop this term. We can act with the photon field to obtain an overlap integral in terms of the WIMP wavefunctions. In the first term, we perform the sum over charginos. The matrix element in eq. 8.15 contains overlap integrals in $\mathbf{x}_{1,2}$. Changing variables to CM position, $\mathbf{X}$, and $\mathbf{r}$, the former trivially integrates to give a spatial-momentum δ-function, which we evaluate in the

CM frame with $\mathbf{P}_i = 0$. Together, these steps give[5]

$$\begin{aligned} S_{i,\,f\gamma} &= \frac{2\pi i}{\sqrt{2k(2\pi)^3}}\,\delta\big[M_\chi v^2/4 - E_n - k\big]\,\delta^{(3)}(\mathbf{k}+\mathbf{P}_{\rm BS})\epsilon(\hat{k},\sigma) \\ &\quad\cdot\left(\frac{e}{M_\chi}\frac{1}{(2\pi)^{3/2}}\int d^3r\,\psi^*_{f,C}\left[{}^{2S+1}L_J\right](\mathbf{r})\,(e^{-i\mathbf{k}\cdot\mathbf{r}/2}+e^{i\mathbf{k}\cdot\mathbf{r}/2})(-i\nabla_{\mathbf{r}})\,\psi_{i,C}(\mathbf{r})\right. \\ &\quad\left.+i\sqrt{2}\,e\,\alpha_W\frac{1}{(2\pi)^{3/2}}\int d^3r\,\psi^*_{f,C}\left[{}^{2S+1}L_J\right](\mathbf{r})\,e^{-m_W r}\,\hat{\mathbf{r}}\,\psi_{i,N}(\mathbf{r})\right). \end{aligned} \tag{8.16}$$

We now have a factor of the photon polarization, $\epsilon(\hat{k},\sigma)$, that we will ultimately sum over upon squaring the amplitude and obtaining the capture rate. The factor of $1/(2\pi)^{3/2}$ in front of the integral arises from our convention on wavefunction normalization.[6] We can make use of the dipole approximation, $e^{i\mathbf{k}\cdot\mathbf{r}/2} \approx 1$, which holds in our regime of interest. The bound state wavefunctions, $\psi_{^{2S+1}L_J}(r)$ die off exponentially after a few Bohr radii, $\sim 1/(\alpha_W M_\chi)$, while the photon energy is set by the binding energy, $\sim \alpha_W^2 M_\chi$. Thus, over the integral's domain of support, the exponent is small.

To get the differential rate to capture to the two-particle final state of photon and WIMPonium, we strip the δ-functions from the S matrix in Eqs. 8.15, 8.16 and integrate the bound-state phase space to get

$$\begin{aligned} (d\sigma)v_{\rm rel} &= (2\pi)^2\,\mu_f\,k\,|M|^2_{i,\,f\gamma}\,d\Omega_k, \quad \text{where} \\ S_{i,\,f\gamma} &= \delta\left(M_\chi v^2/4 - E_n - k\right)\delta^{(3)}(\mathbf{k}+\mathbf{P}_{\rm BS})\,M_{i,\,f\gamma}, \end{aligned} \tag{8.17}$$

and $\mu_f = k\,E_{\rm BS}/(k+E_{\rm BS}) \approx k$ is the final-state reduced energy, including the rest mass. When computing the rate for decay from one bound state to another through emission of a single dipole photon, the calculation is identical, except that we replace

[5]Eq. 8.16 contains a mild abuse of notation as the photon field from the $\mathbb{P}_{NC}$ term in eq. 8.11 is located at the spatial origin, and thus the ultrasoft photon spatial momentum does not give rise to the prefactor $\delta^{(3)}(\mathbf{k}+\mathbf{P}_{\rm BS})$. Operationally though, this δ-function just serves to remove the wimponium phase-space integral, $d^3P_{\rm BS}$, in the cross section and the end result is the same with the formally correct factor for this term, $\delta^{(3)}(\mathbf{P}_{\rm BS})$.

[6]In the free-theory limit, our continuum state would be a plane wave, $\psi_{i,C}(\mathbf{r}) = e^{i\mathbf{p}\cdot\mathbf{r}}$. The $(2\pi)^{-3/2}$ we have pulled out of the integral in eq. 8.16, is a factor giving the normalized continuum state $\Psi_i = (2\pi)^{-3/2}\psi_{i,C}(\mathbf{r})$. The benefit of this convention is that we get a simple inner product for our continuum states, $\int d^3r\,\Psi^\dagger_{i\,\mathbf{p}'}(\mathbf{r})\,\Psi_{i\,\mathbf{p}}(\mathbf{r}) = \delta^{(3)}(\mathbf{p}-\mathbf{p}')$, which one can check trivially holds for the plane-wave case.

$(d\sigma)v_{\rm rel}$ with $d\Gamma$. For capture, the initial state wavefunction is dimensionless, and as mentioned in the above footnote, normalized so that $\int d^3r\, \Psi_{\mathbf{p}'}(\mathbf{r})^\dagger \Psi_{\mathbf{p}}(\mathbf{r}) = \delta^{(3)}(\mathbf{p}-\mathbf{p}')$. For bound-bound state transitions, however, both initial and final state wavefunctions are normalized such that $\int d^3r |\Psi(\mathbf{r})|^2 = 1$, and thus the wavefunctions have units of $(\text{mass})^{3/2}$. Thus the matrix element $M_{i,\,f\gamma}$ has units of $(\text{mass})^{-2}$ in the case of capture into a bound state, and units of $(\text{mass})^{-1/2}$ in the case of transitions between bound states. This yields the correct dimensions for Γ and $\sigma v_{\rm rel}$ (mass and mass^{-2} respectively).

To summarize, in the dipole approximation we have:

$$\begin{aligned}&\sigma v_{\rm rel}\,(\text{continuum} \to \text{bound})\ \text{or}\ \Gamma\,(\text{bound} \to \text{bound})\\ &= \frac{2\,\alpha}{\pi}\frac{k}{M_\chi^2}\int d\Omega_k \left|\epsilon(\hat{k},\sigma)\cdot\int d^3r\,\left(\psi^*_{f,C}(\mathbf{r})\nabla_{\mathbf{r}}\,\psi_{i,C}(\mathbf{r}) - \frac{\alpha_W\, M_\chi\, e^{-m_W\, r}}{\sqrt{2}}\,\psi^*_{f,C}(\mathbf{r})\,\hat{\mathbf{r}}\,\psi_{i,N}(\mathbf{r})\right)\right|^2,\end{aligned} \tag{8.18}$$

where $\alpha = e^2/4\pi$, k is the energy of the emitted photon ($k = -E_n + M_\chi v_{\rm rel}^2/4$ in the case of capture, or the difference in binding energies in the case of a bound-bound state transition), and we have dropped the explicit spin and angular momentum labels on the final state.

For states of known initial and final angular momentum, we can perform the angular integral and reduce the necessary calculation to a one-dimensional integral over the radial wavefunctions, which we compute numerically as described in appendices D.1 and D.2. This procedure is particularly simple where either the initial or final state is s-wave, since (using integration by parts) we avoid the need to apply $\nabla_{\mathbf{r}}$ to a wavefunction with non-trivial angular dependence.[7] In particular, for illustration, let us consider transitions between (continuum or bound) s-wave and p-wave states,

[7] A common procedure in radiative transition calculations is to convert the expectation value of $\nabla_{\mathbf{r}}$ to $\mathbf{r}$ by the relation $\left[H^0, \mathbf{r}\right] = -i\,\mathbf{p}/M$, which converts $\langle f|-i\,\nabla_{\mathbf{r}}\,|i\rangle = M(E_f - E_i)\,\langle f|\,\mathbf{r}\,|i\rangle$. However, we cannot make use of this in a straightforward way in our capture or transition calculations as the Hamiltonian acting on our initial and final states is different.

where the integral in the first term of Eq. 8.18 to be computed takes the form:

$$\int d^3r\, \psi^*_{f,C}(\mathbf{r})\nabla_{\mathbf{r}}\,\psi_{i,C}(\mathbf{r}) \to \int d^3r\, \phi^*_{L=1}(r)\, Y^*_{1m}(\theta,\phi)\,\hat{\mathbf{r}}\frac{\partial}{\partial r}\left[Y_{00}(\theta,\phi)\phi_{L=0}(r)\right]$$
$$= \frac{1}{\sqrt{4\pi}}\int d\Omega\, Y_{1m}(\theta,\phi)^*\,\hat{\mathbf{r}}\int r^2 dr\,\phi^*_{L=1}(r)\phi'_{L=0}(r), \quad (8.19)$$

where we have written the full (charged-component) wavefunctions $\psi_C(r) = \phi(r)Y_{Lm}(\theta,\phi)$, using $\phi(r)$ to denote the radial wavefunctions, and $m = L_z$ labels the magnetic quantum number. The second term, arising from dipole emission off the potential, follows trivially from Eq. 8.19 by replacing $\partial_r \to \alpha_W M_\chi/\sqrt{2}$.

Since we are considering p-wave states, it is useful to write the unit vector $\hat{\mathbf{r}}$ in a basis of $L = 1$ spherical harmonics,

$$\hat{\mathbf{r}} = -\sqrt{\frac{4\pi}{3}}Y_{11}\,\hat{r}_{-1} - \sqrt{\frac{4\pi}{3}}Y_{1,-1}\,\hat{r}_1 + \sqrt{\frac{4\pi}{3}}Y_{10}\,\hat{r}_0, \quad (8.20)$$

where $\hat{r}_0 = z$, $(\hat{r}_{-1} - \hat{r}_1)/\sqrt{2} = \hat{x}$, and $i\,(\hat{r}_{-1} + \hat{r}_1)/\sqrt{2} = \hat{y}$. Thus, for a given m in the p-wave wavefunction, only one of the terms in eq. 8.20 will be nonvanishing. Additionally, since we will be squaring the matrix element and summing over photon polarizations, we can make use of the identity

$$\sum_\sigma \epsilon_i(\hat{k},\sigma)\epsilon^*_j(\hat{k},\sigma) = \delta_{ij} - \hat{k}_i\hat{k}_j. \quad (8.21)$$

Since the different m states sum incoherently, the following angular overlap integrals will enter into the final cross section:

$$(1-\hat{k}_0^2)\left[\int d\Omega\sqrt{\frac{4\pi}{3}}Y_{10}^2\right]^2 = \frac{4\pi}{3}\sin^2\theta_k \qquad m = 0$$
$$(1+\hat{k}_1\hat{k}_{-1})\left[\int d\Omega\sqrt{\frac{4\pi}{3}}Y_{11}\,Y_{1,-1}\right]^2 = \frac{4\pi}{3}\left(1-\frac{\sin^2\theta_k}{2}\right) \quad m = 1 \text{ or } -1, \quad (8.22)$$

where we have used the fact that $Y^*_{1\pm1} = -Y_{1\mp1}$ and $\hat{r}_{-1}\cdot\hat{r}_1 = -1$.

Accordingly, when summing over m states we obtain an overall factor of $8\pi/3$

from the angular integral including the insertion and sum over polarization vectors. For initial states other than s-wave, a difference arises between capture and transition involving which m states are included. For the capture process, our initial state is asymptotically an incoming plane wave, $\Psi_i \propto e^{ikz}$. This has no angular momentum about the direction of travel and therefore $m = 0$. Since our potential, eq. 8.4, is spherically symmetric, the full wavefunction only has a Y_{L0} component, and we do not average over initial polarizations. The on-shell photon emission breaks the rotational symmetry and we can therefore capture into bound states with arbitrary m. Thus, for any process with a WIMPonium initial state, we consider all Y_{Lm} and average over m, dividing by $1/(2L+1)$. In practice though, both processes just give a factor of $1/(2L+1)$ relative to the case of an initial s-wave state (in fact, the rate for transitions from a p-wave state to an s-wave state is independent of the initial value of m, so the average is trivial). Consequently transitions between s- and p-wave states have rates given by:

$$\begin{aligned}&\sigma v_{\rm rel}\,(\text{continuum} \to \text{bound}) \text{ or } \Gamma\,(\text{bound} \to \text{bound}) \\ &= \frac{16}{3}\frac{\alpha\, k}{M_\chi^2}\left|\int r^2 dr\, \phi^*_{L=1}(r)\left(\partial_r - (-1)^{S_i}\frac{\alpha_W M_\chi\, e^{-m_W r}}{\sqrt{2}}\right)\phi_{L=0}(r)\right|^2 \times \begin{cases} 1 & \text{initial } s\text{-wave} \\ 1/3 & \text{initial } p\text{-wave} \end{cases},\end{aligned} \tag{8.23}$$

where S_i is the spin quantum number of the initial state. As a reminder, this rate includes a summation over all possible values of m for the final state (this is the origin of the relative factor of 3 between the process with a p-wave final state and the one with an s-wave final state).

Repeating this calculation for transitions between p-wave and d-wave states yields:

$$\begin{aligned}&\sigma v_{\rm rel}\,(\text{continuum} \to \text{bound}) \text{ or } \Gamma\,(\text{bound} \to \text{bound}) \\ &= \frac{32}{3}\frac{\alpha\, k}{M_\chi^2}\left|\int r^2 dr \phi^*_{L=2}(r)\left[-\frac{1}{r} + \partial_r + (-1)^{S_i}\frac{\alpha_W M_\chi\, e^{-m_W r}}{\sqrt{2}}\right]\phi_{L=1}(r)\right|^2 \times \begin{cases} 1/3 & \text{initial } p\text{-wave} \\ 1/5 & \text{initial } d\text{-wave} \end{cases}.\end{aligned} \tag{8.24}$$

In this case, the transition rate does depend on m for the initial and final states. To obtain the quoted m-independent rate/cross section we have summed over final m and averaged over initial m (note that after summing over final m the transition rates are independent of initial m, and likewise after averaging over initial m the transition rates are independent of final m). In appendix D.3.3 we calculate the rate for a number of transitions, including $p \to d$ transitions, broken down by initial and final m.

These results all assume a specific spin state. This makes sense for bound-bound transitions, where states have definite total spin $S = 0$ or $S = 1$, but for the initial capture generically both spin-singlet and spin-triplet $\chi^0\chi^0$ pairs will be present, in a ratio of 1:3 (singlet:triplet). As discussed above, the initial state must have even $L + S$ to admit a $\chi^0\chi^0$ component, so once L for the initial state is specified, there are contributions to the capture rate only from the spin-singlet pairs (even L) or the spin-triplet pairs (odd L). To obtain the overall spin-averaged capture rate, the rates above should therefore be multiplied by 1/4 (even initial L) or 3/4 (odd initial L).

Let us briefly discuss the boundary condition on the radial continuum wavefunctions. The asymptotic incoming state should be a plane wave with unit normalization, with support only in $\psi_N(r)$ at sufficiently large r.[8] However, because our initial condition corresponds to a pair of identical Majorana fermions, the incoming plane wave state must be antisymmetrized appropriately. For spin-singlet states, the spatial wavefunction must be symmetric, while for spin-triplet states, it must be antisymmetric. Using the asymptotic expansion of a plane wave propagating in the z-direction:

$$e^{ikz} \to \frac{1}{2ikr} \sum_L (2L+1) P_L(\cos\theta) \left(e^{ikr} - (-1)^L e^{-ikr} \right), \tag{8.25}$$

[8] As we discuss in appendix D.1 on calculating the positive-energy wavefunctions, in most of the parameter space we consider, only the neutral component of Ψ, ψ_N, scales like a Bessel function at large radii. Because of the mass-shift, the charged component of the state ψ_C is always off-shell and decays exponentially with distance. It is straightforward to generalize to the case with non-decaying ψ_C, as the incoming, asymptotically plane-wave DM state is still purely in ψ_N.

we see that the appropriately (anti)symmetrized plane wave has the asymptotic form:

$$\begin{aligned}
\text{spin-singlet: } \frac{1}{\sqrt{2}}\left(e^{ikz}+e^{-ikz}\right) &\to \frac{1/\sqrt{2}}{2ikr}\sum_L (2L+1)P_L(\cos\theta)\left(e^{ikr}-e^{-ikr}\right)\left(1+(-1)^L\right) \\
&\to \sum_{L\,\text{even}} \sqrt{2}\sqrt{4\pi(2L+1)}Y_{L0}(\theta,\phi)\frac{\sin(kr)}{kr}, \\
\text{spin-triplet: } \frac{1}{\sqrt{2}}\left(e^{ikz}-e^{-ikz}\right) &\to \frac{1/\sqrt{2}}{2ikr}\sum_L (2L+1)P_L(\cos\theta)\left(e^{ikr}+e^{-ikr}\right)\left(1-(-1)^L\right) \\
&\to \sum_{L\,\text{odd}} \sqrt{2}\sqrt{4\pi(2L+1)}Y_{L0}(\theta,\phi)\frac{\cos(kr)}{ikr}, \qquad (8.26)
\end{aligned}$$

where we have used the fact that $P_L(\cos\theta) = \sqrt{\frac{4\pi}{2L+1}}Y_{L0}(\theta,\phi)$.

Thus at large r, the incoming piece of our continuum wavefunction for fixed L should be normalized as

$$\psi_N(\mathbf{r}) \to Y_{L0}(\theta,\phi)\left[\sqrt{2}\sqrt{4\pi\,(2L+1)}\,\sin(p\,r)/(p\,r)\right], \quad r\to\infty, \qquad (8.27)$$

for even L, and with the same normalization except with a phase shift for odd L. Here $p = M_\chi v_{\rm rel}/2$. Note this normalization is a factor of $\sqrt{2}$ higher than the standard normalization for the partial-wave components of the e^{ikz} plane wave, because only half the partial waves are non-zero as a result of spin statistics.

8.3.2 Decay through annihilation to SM final states

The bound states can also decay through annihilation to SM final states. We will proceed by writing the bound states in terms of free-particle states, but the normalization factor for the states depends on whether the particles involved are distinguishable or indistinguishable. In the center-of-mass frame we have:

$$\begin{aligned}
|\psi\rangle &= \sqrt{\frac{1}{2\mu}}\int \frac{d^3p}{(2\pi)^3}\psi(p)|\mathbf{p},-\mathbf{p}\rangle \quad \text{(distinguishable particles)}, \\
&= \sqrt{\frac{1}{4\mu}}\int \frac{d^3p}{(2\pi)^3}\psi(p)|\mathbf{p},-\mathbf{p}\rangle \quad \text{(identical particles)}, \qquad (8.28)
\end{aligned}$$

where $\mu = M_\chi/2$ is the reduced mass of the two-particle state.

The tree-level annihilation cross sections for wino DM to SM final states have been computed previously in the literature, including the separate s-wave and p-wave contributions [229]. The standard calculation assumes plane-wave initial states; in order to determine the decay rate of the bound states via annihilation, we will write the matrix element for the bound state decay in terms of the matrix elements for free-particle annihilation, following the standard procedure (e.g. [230]). To wit, for a bound state B and final state f, and working in the center-of-momentum frame, we write:

$$\begin{aligned}\mathcal{M}(B \to f) &= \sqrt{\frac{1}{2\mu}} \int \frac{d^3p}{(2\pi)^3} \psi(p) \mathcal{M}(\chi(\mathbf{p})\chi(-\mathbf{p}) \to f) \quad \text{(distinguishable particles)}, \\ &= \sqrt{\frac{1}{4\mu}} \int \frac{d^3p}{(2\pi)^3} \psi(p) \mathcal{M}(\chi(\mathbf{p})\chi(-\mathbf{p}) \to f) \quad \text{(identical particles)}, \end{aligned} \tag{8.29}$$

where $\mathcal{M}(\chi(\mathbf{p_1})\chi(\mathbf{p_2}) \to f)$ is the matrix element for annihilation of free particles with momenta $\mathbf{p}$, $-\mathbf{p}$ to final state f. The differing normalizations for identical and non-identical particles arise from the differing normalizations of the bound states (eq. 8.28).

In the case of states with odd $L + S$, the bound state is composed purely of the $\chi^+\chi^-$ two-particle state, and we need only use the result for distinguishable particles. For even $L+S$, the annihilation may proceed from either the $\chi^0\chi^0$ or $\chi^+\chi^-$ components of the bound state, and the matrix elements will add coherently. Thus, we should write:

$$\mathcal{M}(B \to f) = \sqrt{\frac{1}{2\mu}} \int \frac{d^3p}{(2\pi)^3} \left[\frac{1}{\sqrt{2}} \psi_N(p) \mathcal{M}(\chi^0(\mathbf{p})\chi^0(-\mathbf{p}) \to f) + \psi_C(p) \mathcal{M}(\chi^+(\mathbf{p})\chi^-(-\mathbf{p}) \to f) \right]. \tag{8.30}$$

In the more general case where the bound state is composed of more than two distinct two-particle states, one should add all the matrix elements coherently, with normalizations determined by whether the particles are identical or not.

Now let us consider the two simplifying cases where the bound state is s-wave or p-wave. In the case of s-wave annihilation, the matrix element for free-particle annihilation is independent of $\mathbf{p}$ in the small-p nonrelativistic limit (and the wavefunction, which weights the integral, is suppressed for large p), and thus we can take it outside the integral. Since $\psi(\mathbf{r}) = \int \frac{d^3p}{(2\pi)^3}\psi(\mathbf{p})e^{i\mathbf{p}\cdot\mathbf{r}}$, it follows that $\int \frac{d^3p}{(2\pi)^3}\psi(\mathbf{p}) = \psi(\mathbf{r}=0)$. Thus we have, in the nonrelativistic limit:

$$\mathcal{M}^{L=0}(B \to f) = \sqrt{\frac{1}{2\mu}}\left[\frac{1}{\sqrt{2}}\psi_N(\mathbf{r}=0)\mathcal{M}^{L=0}(\chi^0\chi^0 \to f) + \psi_C(\mathbf{r}=0)\mathcal{M}^{L=0}(\chi^+\chi^- \to f)\right]. \tag{8.31}$$

In the p-wave case, the matrix element for free-particle annihilation scales linearly with $\mathbf{p}$ in the limit of small p. Thus, the integrals over d^3p will take the form:

$$\int \frac{d^3p}{(2\pi)^3}\psi(\mathbf{p})\mathbf{p} = \lim_{\mathbf{r}\to\mathbf{0}}\left(i\nabla_{\mathbf{r}}\int \frac{d^3p}{(2\pi)^3}\psi(\mathbf{p})e^{-i\mathbf{p}\cdot\mathbf{r}}\right) = i\lim_{\mathbf{r}\to\mathbf{0}}\nabla_{\mathbf{r}}\psi(\mathbf{r}). \tag{8.32}$$

Furthermore, the $L=1$ wavefunctions have a universal form at small r:

$$\begin{aligned}\psi_N(\mathbf{r}) &= \sqrt{\frac{4\pi}{3}}A_N\, r\, Y_{Lm}(\theta,\phi),\\ \psi_C(\mathbf{r}) &= \sqrt{\frac{4\pi}{3}}A_C\, r\, Y_{Lm}(\theta,\phi).\end{aligned} \tag{8.33}$$

Accordingly, $\nabla_{\mathbf{r}}\psi_N(\mathbf{r}) = A_N\,\hat{r}_m$, and similarly $\nabla_{\mathbf{r}}\psi_C(\mathbf{r}) = A_C\,\hat{r}_m$. Thus, we can write the matrix element for annihilation from the p-wave bound state in the form:

$$\begin{aligned}\mathcal{M}^{L=1}(B \to f) &= \sqrt{\frac{1}{2\mu}}\left[\frac{1}{\sqrt{2}}\mathcal{M}^{L=1}(\chi^0(\mathbf{p})\chi^0(-\mathbf{p}) \to f)|_{\mathbf{p}\to i\nabla_{\mathbf{r}}\psi_N(\mathbf{r})|_{\mathbf{r}=0}}\right.\\ &\quad \left. + \mathcal{M}^{L=1}(\chi^+(\mathbf{p})\chi^-(-\mathbf{p}) \to f)|_{\mathbf{p}\to i\nabla_{\mathbf{r}}\psi_N(\mathbf{r})|_{\mathbf{r}=0}}\right],\\ &= i\sqrt{\frac{1}{2\mu}}\left(\tfrac{1}{\sqrt{2}}\hat{r}_m\cdot\mathcal{M}_0^{L=1}(\chi^0\chi^0 \to f),\quad \hat{r}_m\cdot\mathcal{M}_0^{L=1}(\chi^+\chi^- \to f)\right)\begin{pmatrix}A_N\\ A_C\end{pmatrix}\end{aligned} \tag{8.34}$$

where the $\mathcal{M}_0$ matrix elements are vectorial but momentum-independent, and satisfy

$\mathcal{M}^{L=1}(\chi\chi \to f) = \mathbf{p} \cdot \mathcal{M}_0^{L=1}(\chi\chi \to f)$. We can use eq. 8.32 to replace $\mathbf{p}$ in eq. 8.34 as the dependence of the matrix element, $\mathcal{M}^{L=1}$ is linear in $\mathbf{p}$.

The decay width for the bound state due to these annihilation processes is:

$$\Gamma = \frac{1}{2M_B} \int d\Pi_n \left|\mathcal{M}(B \to f)\right|^2, \tag{8.35}$$

where $M_B \approx 2M_\chi$ is the mass of the bound state and Π_n denotes the final state integral over phase space.

For s-wave annihilation we can therefore write:

$$\begin{aligned}
\Gamma^{L=0} &= \frac{1}{4M_\chi^2} \int d\Pi_n \left| \frac{1}{\sqrt{2}} \psi_N(\mathbf{r}=0) \mathcal{M}^{L=0}(\chi^0\chi^0 \to f) + \psi_C(\mathbf{r}=0) \mathcal{M}^{L=0}(\chi^+\chi^- \to f) \right|^2, \\
&= \begin{pmatrix} \psi_N^*(\mathbf{r}=0) & \psi_C^*(\mathbf{r}=0) \end{pmatrix} \Sigma_{L=0}(f) \begin{pmatrix} \psi_N(\mathbf{r}=0) \\ \psi_C(\mathbf{r}=0) \end{pmatrix}, \\
\Sigma_{L=0}(f) &\equiv \frac{1}{(2M_\chi)^2} \begin{pmatrix} \frac{1}{2}\int d\Pi_n |\mathcal{M}(\chi^0\chi^0 \to f)|^2 & \Sigma^{12}_{L=0}(f) \\ (\Sigma^{12}_{L=0}(f))^* & \int d\Pi_n |\mathcal{M}(\chi^+\chi^- \to f)|^2 \end{pmatrix}, \\
\Sigma^{12}_{L=0}(f) &\equiv \frac{1}{\sqrt{2}} \int d\Pi_n \mathcal{M}^*(\chi^0\chi^0 \to f) \mathcal{M}(\chi^+\chi^- \to f)),
\end{aligned} \tag{8.36}$$

where in the last line all matrix elements are s-wave but we have omitted the $L = 0$ superscripts for notational convenience.

Similarly, the decay rate corresponding to p-wave annihilation is:

$$\begin{aligned}
\Gamma^{L=1} &= \frac{1}{M_\chi^2} \begin{pmatrix} A_N^* & A_C^* \end{pmatrix} \Sigma_{L=1}(f) \begin{pmatrix} A_N \\ A_C \end{pmatrix}, \\
\Sigma_{L=1}(f) &\equiv \frac{1}{(2M_\chi)^2} \begin{pmatrix} \frac{1}{2}\int d\Pi_n |M_\chi \hat{r}_m \cdot \mathcal{M}_0(\chi^0\chi^0 \to f)|^2 & \Sigma^{12}_{L=1}(f) \\ (\Sigma^{12}_{L=1}(f))^* & \int d\Pi_n |M_\chi \hat{r}_m \cdot \mathcal{M}_0(\chi^+\chi^- \to f)|^2 \end{pmatrix}, \\
\Sigma^{12}_{L=1}(f) &\equiv \frac{1}{\sqrt{2}} \int d\Pi_n \left[(M_\chi \hat{r}_m) \cdot \mathcal{M}_0(\chi^0\chi^0 \to f)\right]^* \left[(M_\chi \hat{r}_m) \cdot \mathcal{M}_0(\chi^+\chi^- \to f)\right],
\end{aligned} \tag{8.37}$$

where all matrix elements are for $L = 1$, but again we have omitted the superscripts for notational convenience. Note we have included factors of M_χ in $\Sigma_{L=1}(f)$ so that

it retains the dimensions of a cross section.

The diagonal elements of $\Sigma(f)$ give the cross sections for free-particle annihilation for distinguishable particles, and the cross sections multiplied by a factor of 1/2 for identical particles (as in the annihilation matrices of [197]), except that in the p-wave case, in all cross sections $\mathbf{p}$ has been replaced with $M_\chi \hat{r}_m$. After integrating over the final-state phase space and performing all spin and polarization sums/averages, this amounts to multiplying all cross sections by M_χ^2/p^2. If we set, for example, $A_C = p$ and $A_N = 0$ (as appropriate for a plane wave purely in the $\chi^+\chi^-$ state), then we recover the rate for free-particle annihilation from the chargino-chargino state. These precise annihilation matrices Σ, up to trivial prefactors, have already been computed in the literature [197, 229] for general electroweakly interacting DM. To facilitate extension of our results to other models, in appendix D.5 we provide a general algorithm for determining the Σ matrices from existing results. We have also independently derived several of the results presented below (all for the spin-singlet case, and the channels with the largest branching ratios for the s-wave spin-triplet case).

In the particular case of the wino, we have for the $L = 0$ spin-singlet bound states [192, 197]:

$$\Sigma(W^+W^-) = \frac{4\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} \end{pmatrix}, \quad \Sigma(ZZ) = \frac{4\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ 0 & c_W^4 \end{pmatrix}$$
$$\Sigma(Z\gamma) = \frac{4\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ 0 & 2c_W^2 s_W^2 \end{pmatrix}, \quad \Sigma(\gamma\gamma) = \frac{4\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ 0 & s_W^4 \end{pmatrix}, \tag{8.38}$$

and for the $L = 1$ spin-triplet bound states the similar result (see appendix D.5):

$$\Sigma(W^+W^-) = \frac{28\pi\alpha_W^2}{9M_\chi^2}\begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} \end{pmatrix}, \quad \Sigma(ZZ) = \frac{28\pi\alpha_W^2}{9M_\chi^2}\begin{pmatrix} 0 & 0 \\ 0 & c_W^4 \end{pmatrix}$$
$$\Sigma(Z\gamma) = \frac{28\pi\alpha_W^2}{9M_\chi^2}\begin{pmatrix} 0 & 0 \\ 0 & 2c_W^2 s_W^2 \end{pmatrix}, \quad \Sigma(\gamma\gamma) = \frac{28\pi\alpha_W^2}{9M_\chi^2}\begin{pmatrix} 0 & 0 \\ 0 & s_W^4 \end{pmatrix}. \tag{8.39}$$

As discussed above, p-wave bound states in the spin-singlet configuration and s-wave

bound states in the spin-triplet configuration (odd $L+S$) are only composed of $\chi^+\chi^-$, with no $\chi^0\chi^0$ component. Thus the Σ matrix now only has one non-zero component, namely the diagonal entry corresponding to $\chi^+\chi^-$ annihilation. Furthermore, for the p-wave state the Landau-Yang theorem forbids the decay into massless neutral vector bosons. However, for the s-wave bound state, the s-channel annihilation is open and permits decays into all SM final states. As calculated in appendix D.5, the non-zero annihilation matrices for these bound states are given by:

$$\Sigma(W^+W^-) = \frac{2}{3}\frac{\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{8.40}$$

for the spin-singlet p-wave states, and

$$\begin{aligned} \Sigma(W^+W^-) &= \frac{1}{12}\frac{\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Sigma(Zh^0) = \frac{1}{12}\frac{\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \\ \Sigma(q\bar{q}) &= \frac{1}{2}\frac{\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Sigma(l^+l^-, \nu\bar{\nu}) = \frac{1}{6}\frac{\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned} \tag{8.41}$$

for the spin-triplet s-wave states.

In principle, one could hope to detect an energetic, monochromatic photon line from the bound state's annihilation to $\gamma\gamma$ or γZ. However, only the $L+S$-even bound states have a sizable branching ratio to photons, but, as discussed above, we directly capture only to states with odd $L+S$. Thus, line-photon annihilation events will require that capture occurs into an excited state that can decay by dipole emission to a state with even $L+S$, e.g. the free winos capture into the spin-singlet $2p$ state, which subsequently transitions to d or s-wave. As we show in section 8.4.2, for winos in the Milky Way halo, capture into the excited, p-wave states is dominated by the direct rate for WIMPs to annihilate to $\gamma + X$. Thus we expect a small branching ratio for monochromatic gamma-ray annihilation lines from bound states.

Note that by dimensional analysis, we naively expect $|\Psi(0)| \sim (\alpha_W M_\chi)^{3/2}$ for s-wave bound states, and $A_{N,C} \sim (\alpha_W M_\chi)^{5/2}$ for p-wave bound states. Thus we expect

the decay width for annihilations from s-wave bound states to scale as $\alpha_W^5 M_\chi$, and for annihilations from p-wave bound states to scale as $\alpha_W^7 M_\chi$. More generally (as also noted in [210]), the width for decay via annihilation will scale as $\Gamma \propto \alpha_W^{5+2L} M_\chi$.

8.4 Analytic and numerical results

In this section we apply the results of section 8.3; we first consider the fate of bound states once they form, and then move on to discuss the capture cross section, which primarily determines the overall importance of bound state formation relative to direct annihilation.

8.4.1 WIMPonium decays

As discussed above, the WIMPonium bound states may decay to lower-energy states in the spectrum by emission of photons, or annihilate to SM particles. As we will demonstrate, the former generally dominate for $L > 0$ if electric-dipole transitions are allowed, with widths scaling as $\Gamma \propto \alpha\, \alpha_W^4 M_\chi$ in the unbroken SU(2) limit. Let us begin by considering the circumstances under which such transitions can occur.

Unsuppressed (single-photon electric-dipole) transitions, either between continuum states or bound states, require $\Delta L = \pm 1$, e.g. p-wave states can decay to s-wave or d-wave bound states. As discussed above, bound states populated by single-photon capture will have odd $L + S$, and so will be purely comprised of chargino pairs. The states to which they can decay by single-photon emission will have even $L + S$; they will consequently tend to have larger binding energies for a given principal quantum number (since the potential for even $L + S$ has a larger eigenvalue for its attractive component, in the illustrative "SU(2) positronium" limit). This means that, for example, the $1s$ and $2s$ states with odd $L + S$ (i.e. spin-triplet states) may in some circumstances have available single-photon dipole decays to a $2p$ state with even $L+S$, and so need not be (meta)stable as they are in the hydrogen atom. Similarly, the $2p$ state with odd $L + S$ (spin-singlet) may have open decay channels to the $3d$ and $3s$ states as well as the $1s$ state. This point is illustrated in figure 8-3.

As we will demonstrate in the next subsection, at low DM masses the dominant capture process populates the spin-triplet $1s$ state, which is generically the lowest-lying spin-triplet bound state; consequently this process gives rise to no subsequent transitions, which would require an α_W-suppressed spin-flip.[9] At higher DM masses the most important capture processes dominantly populate the lowest-lying spin-singlet states with odd L, i.e. the $2p$ states, via capture from the s- and d-wave parts of the original plane wave. These statements assume the typical velocity of the Milky Way halo $v \sim \mathcal{O}(10^{-3})$; at low velocities where $M_\chi v \lesssim m_W$, then the system is still in the Yukawa regime and the contributions from higher-L partial waves have velocity suppression due to the short-range nature of the potential (see appendix D.4 for a discussion of the scaling). The $2p$ states generically have open and unsuppressed decays to the spin-singlet $1s$, $2s$, $3s$ and $3d$ states. The $3d$, $3s$ and $2s$ spin-singlet states are themselves metastable, as there are no lower-lying spin-singlet states with odd L, so they can decay only through annihilation to SM states or:

- two-photon transitions to the $1s$ spin-singlet ground state,
- electric quadrupole transitions to the $1s$ spin-singlet ground state with even $L + S$ (available for d-wave states only),
- magnetic dipole, spin-flip transitions to the spin-triplet states, induced by relativistic effects.

In order to develop intuition, let us first consider the allowed transitions from the spin-singlet $2p$ states in the high-energy limit where the SU(2) is unbroken. From figure 8-3, we see that the 1–3s and $3d$ spin-singlet states are at lower energies. Note that while we formally set the masses of all force carriers to zero for this analysis, we still examine only capture through emission of W^3, which will correspond to photon emission after SU(2) is broken. We obtain the total decay rates for each of the spin-singlet $nlm = 21m$ as $\Gamma \approx 0.16\,\alpha\,\alpha_W^4 M_\chi$ (appendix D.3.3). Approximately half the total branching ratio is to the ground state, followed closely by decays to the $2s$

[9] As we see in figure 8-3, capture to s-wave states with $n \geq 5$ can have dipole transitions to the $2p$ state or others for sufficiently high n.

excited state; decays to the $3s$ and $3d$ excited states contribute only $\sim 6\%$ of the total rate.

For annihilation, let us consider the same unbroken limit and compute the annihilation rates for the $2p$ and $1s$ spin-singlet states, and the $1s$ spin-triplet state, using the explicit form of the Coulombic bound state wavefunction (see discussion in appendix 7.3.2):

$$\Psi^i_{nlm}(\mathbf{r}) = Y_{lm}(\theta,\phi)\eta_i R_{nl}(\lambda_i\alpha; r), \tag{8.42}$$

$$R_{nl}\left(\lambda_i\alpha; r\right) = \left[\left(\frac{2\alpha\lambda_i\mu}{n}\right)^3 \frac{(n-l-1)!}{2n\,(n+l)!}\right]^{1/2} e^{-\mu\lambda_i\alpha r/n}\left(\frac{2\alpha\lambda_i\mu r}{n}\right)^l \mathrm{L}^{2l+1}_{n-l-1}\left(\frac{2\alpha\,\lambda_i\mu r}{n}\right). \tag{8.43}$$

For the spin-singlet $1s$ state, which has $\lambda_i = 2$ and $\eta_i = \left(\sqrt{\frac{1}{3}} \quad \sqrt{\frac{2}{3}}\right)$, we obtain $\Psi^i_{100}(0) = \eta_i(\alpha_W M_\chi)^{3/2}/\sqrt{\pi}$, i.e. $\psi_C(0) = \sqrt{2/3}\,(\alpha_W M_\chi)^{3/2}/\sqrt{\pi}$, $\psi_N(0) = \sqrt{1/3}\,(\alpha_W M_\chi)^{3/2}/\sqrt{\pi}$. Thus the decay rates to different final states, $\Gamma(f)$, from the spin-singlet $1s$ state, are given by:

$$\begin{aligned} \Gamma(W^+W^-) &= \frac{16}{3}\alpha_W^5 M_\chi, & \Gamma(ZZ) &= \frac{8}{3}c_W^4\alpha_W^5 M_\chi, \\ \Gamma(Z\gamma) &= \frac{16}{3}c_W^2 s_W^2\alpha_W^5 M_\chi, & \Gamma(\gamma\gamma) &= \frac{8}{3}s_W^4\alpha_W^5 M_\chi. \end{aligned} \tag{8.44}$$

For the spin-triplet $1s$ state we have $\psi_N(0) = 0$, $\psi_C(0) = (\alpha_W M_\chi/2)^{3/2}/\sqrt{\pi}$. The decay rates then become:

$$\begin{aligned} \Gamma(W^+W^-) &= \frac{1}{96}\alpha_W^5 M_\chi, & \Gamma(Zh^0) &= \frac{1}{96}\alpha_W^5 M_\chi, \\ \Gamma(q\bar{q}) &= \frac{1}{16}\alpha_W^5 M_\chi, & \Gamma(l^+l^-, \nu\bar{\nu}) &= \frac{1}{48}\alpha_W^5 M_\chi. \end{aligned} \tag{8.45}$$

Finally, for annihilation from the $2p$ spin-singlet states, we have $|\nabla\psi_C(0)|^2 = (\alpha_W\mu)^5/(32\pi)$ (for all m), where $\mu = M_\chi/2$, so therefore:

$$\Gamma(W^+W^-) = \frac{1}{3\times 2^9}\,\alpha_W^7 M_\chi \tag{8.46}$$

We see that, as claimed earlier, the annihilation rate for the $2p$ state is parametri-

cally suppressed relative to the electric-dipole single-photon transitions to lower-lying s and d states, which have rates $\sim 10^{-1}\alpha\,\alpha_W^4 M_\chi$; consequently, we can safely approximate that any capture into the $2p$ spin-singlet state results in a transition to an ns (n=1–3) or $3d$ spin-triplet state, followed by annihilation or a suppressed decay as appropriate. The principal decay channel will be to the $1s$ spin-singlet state, and so in this unbroken limit, we expect capture to the $2p$ state to eventually result in annihilation decays to the SM with approximately the branching ratios in eq. 8.44. For capture to the $1s$ spin-triplet state, a wide range of SM final states can be produced due to the presence of an s-channel annihilation; most of the branching ratio is to hadronic channels (quarks), and thus the resulting decay annihilation signal would be rich in continuum photons and charged particles, but with no appreciable gamma-ray line at the DM mass.

Moving beyond the SU(2) symmetric limit, we can use the numerical method introduced in appendix D.2 to calculate the bound-state wavefunctions and then use eq. 8.23 to find the spin-singlet, $2p$ to ns transitions. The results are shown in Figure 8-6, together with the analytic results presented above, which approach validity in the limit of high DM mass. As expected, the decay via annihilation is suppressed by a few orders of magnitude compared to the transition to lower s- and d-wave bound states.

Finally, since the bound-state Hamiltonian includes the positive mass-shift, some of the bound states (in the sense that their wavefunctions are exponentially suppressed at large r) will have positive energy, according to our definition of zero energy. Such states could therefore decay into lower-energy unbound states (corresponding to free $\chi^0\chi^0$ pairs at large r) through the emission of a photon, which changes $L+S$ from odd to even. It is a quirk of our Hamiltonian with the mass-shift term that bound and continuum states overlap in the spectrum between $E=0$ and $E=+2\delta M$.

However, we expect the impact of these positive-energy bound states to be small, and neglect them in our calculations. As we see in figure 8-4, for $M_\chi > 6$ TeV, there are negative-energy bound states in the $L+S$-odd Hamiltonian spectrum available for capture. Capture rates are typically dominated by the deepest-available bound

states, with the rates smoothly turning off as the binding energy approaches zero from below. Furthermore, capturing into the full range of positive-energy bound states requires sufficient kinetic energy from the initial $\chi^0\chi^0$. For example, with our standard mean velocity, $v = 10^{-3}$, we would need $M_\chi > 1320$ TeV to capture into all $L+S$-odd, $\chi^+\chi^-$ bound states; if the kinetic energy is much smaller than this value, positive-energy bound states will only be available for capture in the fine-tuned case where their binding energy relative to the free $\chi^+\chi^-$ state is very close to the mass splitting $2\delta M$.

If we do form such a WIMPonium, its "fall-apart" transition back to free $\chi^0\chi^0$ (with emission of another photon) is kinematically suppressed relative to standard dipole-emission decay to a negative-energy bound state, if such an accessible state exists in the spectrum. As stated above, the bound-state to bound-state transition rate scales like $\alpha\,\alpha_W^4 M_\chi$. We can estimate the rate of WIMPonium $\to \chi^0\chi^0\,\gamma$ from the capture rate in the Coulomb limit, which scales as $\sigma v \sim \alpha E_\gamma/v$, as the overlap integral and thus the squared matrix elements are the same. However, to convert σv to Γ, we need an additional factor of the phase space for the relative WIMP momentum, $p = M_\chi v_{\rm rel}/2$. The positive powers of v from this measure will (more than) cancel the $1/v$ that gave the Sommerfeld enhancement for capture. We thus get a factor $(E_\gamma v_{\rm rel}^2)/M_\chi$. For the highest-energy bound states, $E_\gamma \sim \delta M$ and $v_{\rm rel}^2 \sim \delta M/M_\chi$, for an overall scaling like $\alpha(\delta M/M_\chi)^2$. Thus, making this process competitive with the dipole transition rate to another bound state would require $(\delta M/M_\chi) > \alpha_W^2$ and thus $M_\chi < 300$ GeV. However, this is outside the regime where Sommerfeld and electroweak bound state effects occur, which requires $M_\chi \alpha_W/m_W \gtrsim 1$.

8.4.2 WIMPonium formation

Again, we will begin by considering the symmetric limit where SU(2) is unbroken in order to build intuition, as in this limit the single-photon capture rates can be calculated analytically from the formulae derived in section 8.3.1. The spin-averaged cross sections for radiative capture into the first few bound states from the full initial (asymptotically plane-wave) state, by single-photon emission in the dipole approxi-

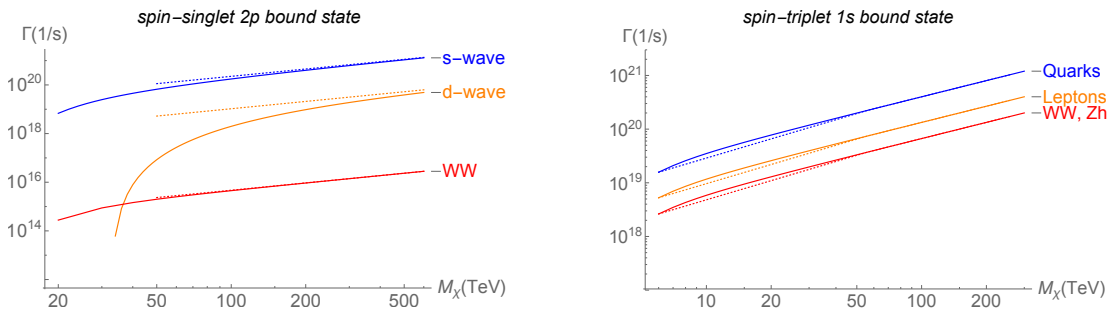

Figure 8-6: Our numerical (solid) and exact results in the SU(2)-symmetric limit (dotted) for the decay channels of the lowest-energy bound states we capture into in each spin configurations. **Left:** Decay rates of the $2p$ spin-singlet bound state. The *blue* curve shows the inclusive decay rate to the three lower s-wave bound states available (1–3s); the *yellow* (*red*) curve is the rate to the $3d$ bound state (W^+W^-). **Right:** Decay rates of the $1s$, spin-triplet bound state, given by the annihilation matrices in eq. 8.41.

mation and in the limit of small initial momenta, are given by (eq. D.32):

$$\sigma v_{\rm rel} = \frac{2^{13}\,\pi^2}{3^3}\,\frac{\alpha\,\alpha_W^2}{M_\chi^2\,v_{\rm rel}}\,\frac{1}{n^2}\,e^{-8n}\,f_{nlm}. \tag{8.47}$$

where the f_{nlm} coefficients are given by:

$$f_{100} = 0, \quad f_{200} = 384, \quad f_{210} = 242, \quad f_{21\pm1} = 50. \tag{8.48}$$

As discussed above, here we have multiplied the cross sections for even-L final states by 3/4 to account for the fact that the initial state must be odd-L and hence spin-triplet, and likewise we have multiplied the cross sections for odd-L final states by 1/4.

Consider the more general case where the incoming two-particle state experiences a Coulomb potential with coupling $\lambda_i\alpha_W$ and corresponding eigenvector η_i, and the final bound state is supported by a Coulomb potential with coupling $\lambda_f\alpha_W$ and corresponding eigenvector η_f. We find that (at least for these low-lying states) there is a generic accidental suppression in the cross section of the form $e^{-4n\lambda_i/\lambda_f}$, arising from the overlap between the wavefunctions with different eigenvalues (see appendix

7.3.3 for the derivation). Since for the wino-like case, $\lambda_i = 2$ for the attracted component (since the attracted component must have even $L + S$ to allow mixing between the $\chi^0\chi^0$ and $\chi^+\chi^-$ two-particle states), and $\lambda_f = 1$, this suppression is e^{-8n}, and acts quite strongly to suppress capture into higher-n bound states. For positronium, where there is only one relevant Coulomb potential, this factor is only e^{-4n}. As we will see, the single-photon capture cross section for the wino is generically well below the direct annihilation cross section, which is not the case for positronium (we discuss this point further in section 8.4.3).

In this regime, where the potential has infinite range, there is no velocity suppression of terms corresponding to higher partial waves in the incoming two-particle state. However, we expect such a velocity suppression to occur once the relative particle velocity is comparable to m_W/M_χ. As a crude estimate of the effects on the cross section, we can separate out the contribution to eq. 8.47 originating purely from the $L = 0$ partial wave, setting all other contributions to zero. For capture from the s-wave piece of the initial state to the $n = 2$, $l = 1$ bound states, we find (eq. D.41):

$$\sigma v_{\rm rel} = \frac{2^{12}\,\pi^2}{3^4} e^{-16} \frac{\alpha\,\alpha_W^2}{M_\chi^2\,v_{\rm rel}}. \tag{8.49}$$

Note that the contributions to capture rates into the $nlm = 210$, 211 and 21-1 states are identical for this case; here we have summed the cross sections together. We have also averaged over the spin configuration, which amounts to dividing the cross section for the spin-singlet case by 4, since there is no s-wave component of the spin-triplet state due to Fermi statistics. However, as we show in appendix D.4, the anticipated velocity suppression of the higher partial waves is of order $(M_\chi v/m_W)^{2L}$, rather than simply v^{2L}. Consequently, so long as v is not too small compared to m_W/M_χ, the contributions from higher partial waves may still be non-negligible and even dominate. However, even in the limit of unbroken $SU(2)$ there can be cases where an accidental cancellation sets the rate for a particular capture channel to zero. For example, for the wino this occurs for the spin-averaged capture rate from the p-wave piece of the initial state to the spin-triplet $1s$ state.

To compute the full capture cross section, we numerically solve for the radial wavefunctions for both the continuum and bound states, using the methods presented in appendices D.1 and D.2. In figure 8-7, we plot the capture rate to the $2p$ spin-singlet bound state with the rate given in eq. 8.23.[10]

The capture to the spin-triplet $1s$ state, which at the order we have worked originates entirely from the p-wave part of the initial continuum state, is the only available channel at low DM masses, is competitive with capture to the $2p$ states at intermediate DM masses, and becomes subdominant at high DM masses because it vanishes in the Coulomb limit.

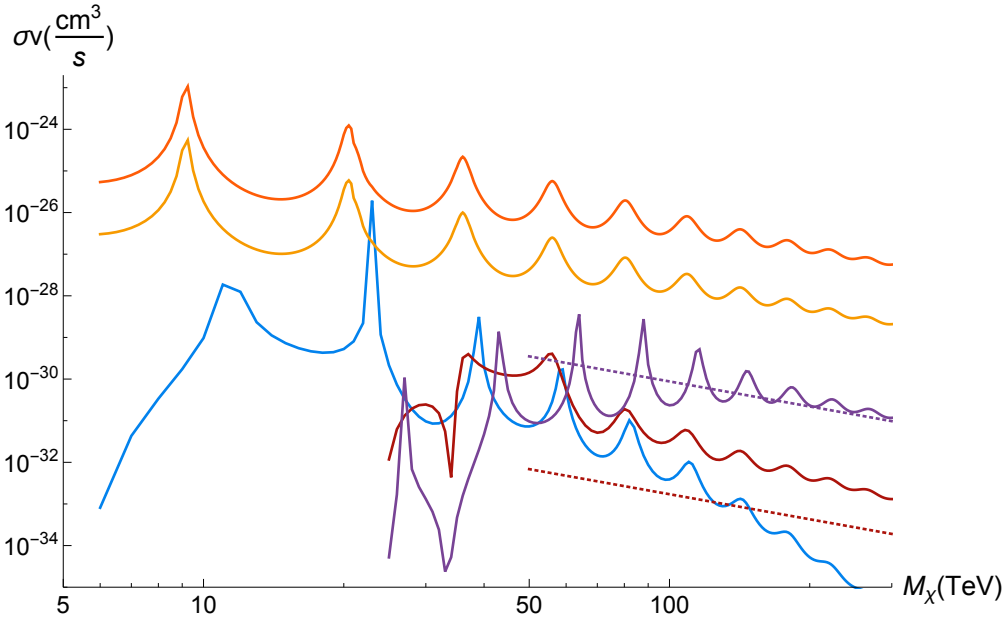

Figure 8-7: Rates for initial wino dark matter state, $\chi^0\chi^0$, to capture to WIMPonium or annihilate directly to SM bosons with $v_{\rm rel} = 10^{-3}$. *Dark Orange*: Tree-level inclusive annihilation rate to W^+W^-, γZ, and $\gamma\gamma$. *Yellow*: Semi-inclusive observable to rate to a monochromatic photon, $\chi^0\chi^0 \to \gamma + X$, assuming a HESS-like detector, using the results of [217]. *Blue*: $\chi^0\chi^0(p\text{–wave}) \to {}^3S_1 + \gamma$, lowest–$n$ bound state, n=1. *Purple*: $\chi^0\chi^0(d\text{–wave}) \to {}^1P_1 + \gamma$ ($n = 2$). *Maroon*: $\chi^0\chi^0(s\text{–wave}) \to {}^1P_1 + \gamma$, $n = 2$. *Dashed* lines indicate $3\times\sigma v_{\rm rel}$ computed analytically in the SU(2)-symmetric limit for $s \to p$ (*maroon*, eq. D.41) and $d \to p$ (*purple*, as discussed in Appendix D), capturing to the lowest-n bound states only (see the text for an explanation of the factor of 3).

[10]We note that for our numerical analysis, we have taken the parameters of the electroweak potential at their PDG m_Z values [231]. Since the proper scale is given by value of order the momentum transfer, in the potential, this is $\max(m_W, M_\chi v_{\rm rel})$ and for the photon emission is $\max(E_n, M_\chi v_{\rm rel}^2)$, where E_n is the binding energy, typically $\mathcal{O}$(few $\times$ GeV). Summing the logarithms associated with the hierarchy is beyond our scope.

Our numerical results in figure 8-7 only include capture to the relevant state with the lowest n quantum number. Generically the capture to excited states is suppressed by the e^{-8n} factor (although as discussed above accidental cancellations can change this hierarchy).

At low masses, the capture cross section experiences a pattern of resonances similar to that for Sommerfeld-enhanced direct annihilation, due to the enhancement of the continuum-state wavefunction close to the origin when a bound state in the spectrum passes through zero energy (note that these are bound states for the potential with *even* $L+S$, whereas the bound states produced by the single-photon-mediated capture necessarily have odd $L + S$). At high masses, the resonance peaks diminish and the result approaches our analytic calculation for the unbroken SU(2) limit, up to an overall factor; when we test very high masses beyond the reach of Figure 8-7 our numerical capture rate consistently exceeds the analytical prediction by a factor of 3 for $v \lesssim 10^{-3}$.

We attribute this factor of 3 to a somewhat subtle effect discussed in appendix E2 of [160]; the issue is that since the chargino states are not kinematically accessible, we are not truly in the limit of unbroken SU(2), as the mass splitting between the neutralino and chargino states is large compared to other energy scales in the problem (i.e. the kinetic energy of the particles). The transition between the large-r regime, where the mass splitting dominates the potential, and the small-r regime, where the potential is approximately Coulombic, can give rise to effects that do not appear in the unbroken-SU(2) limit.

Our cross section result in the unbroken SU(2) limit includes a factor of 1/3 from the overlap between our initial condition (particles begin as neutralinos) and the eigenvector of the potential matrix that experiences an attractive interaction; in the language of appendix D.3, and particularly eq. 7.71, this factor appears in the matrix element as $\mathbf{I} \cdot \eta_i$ ($= 1/\sqrt{3}$ for the wino). More specifically, it appears in the contribution to the matrix element from each component of the continuum wavefunction that experiences an attractive interaction. In the low-velocity limit, it is these contributions that control the overall capture rate, since any component of the wave-

function that experiences a repulsive interaction is suppressed toward the origin and its contribution to the capture rate is exponentially suppressed (as we demonstrate in appendix D.3). In the unbroken SU(2) limit, the wavefunctions associated with the various eigenvectors of the potential obey decoupled Schrödinger equations, and the eigenvectors themselves are independent of r; thus we can determine the wavefunctions associated with the various eigenvectors at some large r (i.e. by setting an initial condition), and then evolve them straightforwardly for all r.

However, in the more general case where SU(2) is broken, the fraction of the wavefunction corresponding to each of the r-dependent eigenvectors will evolve with r in a non-trivial way. In particular, when the $\chi^0\chi^0$ and $\chi^+\chi^-$ states have different masses – that is, they have different energies as $r \to \infty$ – this mass splitting defines the two eigenvectors of the potential matrix at large r, whereas at small r Coulomb-like behavior is recovered and the eigenvectors of the matrix correspond to states experiencing attractive (lower energy) or repulsive (higher energy) Coulomb potentials. One particularly simple case occurs when the transition between the two regimes is sufficiently slow and adiabatic: then if the wavefunction is purely in the lower-energy eigenstate at large r (i.e. the $\chi^0\chi^0$ state), it will entirely populate the lower-energy (attracted) eigenstate at small r also. Consequently, the $\chi^0\chi^0$ state effectively feels a purely attractive interaction, and there is no suppression factor in the matrix element to account for the fraction of the state that experiences repulsion and does not contribute to the capture rate.

If this adiabatic approximation is valid, then the factors of $\mathbf{I} \cdot \eta_i$ appearing in eq. 7.71 should be replaced by 1 for the lowest-energy eigenstate at small r – i.e. the eigenstate η_i corresponding to the largest value of λ_i – and by 0 for all other eigenstates. However, caution is warranted when applying this naive estimate to systems with multiple eigenstates with $\lambda_i > 0$, as eigenstates with smaller values of λ_i can yield exponentially larger contributions to the capture cross section (via the $e^{-2n\lambda_i/\lambda_f}$ factor of eq. 7.71), and to our knowledge this behavior has only been studied in systems with a single attracted eigenstate.

In a simpler multi-state model, with only a single force carrier with mass m_A and

coupling α_A, [160] gave the criterion for this adiabatic rotation to occur as $m_A v_{\rm rel} \lesssim 2\delta$. In our model, the equivalent criterion would be $m_W v_{\rm rel} \lesssim 2\delta$, which is generically true for $v_{\rm rel} \lesssim 2\delta/m_W \sim 4 \times 10^{-3}$, independent of the DM mass.

For the detailed analysis above, we have taken the relative WIMP velocity to be $v_{\rm rel} = 10^{-3}$, typical of DM velocities in the Milky Way halo [186]. However, the capture rates are not velocity-independent in general. It is interesting to scan in $v_{\rm rel}$ both because the true WIMP velocity has a Maxwellian distribution and thus will have support at other values, and as a check on our expectations for scaling of the rates with velocity; the latter will be particularly important when considering signals from e.g. clusters, dwarf galaxies, or substructure in the Milky Way halo. In figure 8-8 we plot the effects of varying $v_{\rm rel}$ by an order of magnitude for the $s \to p$–wave and $p \to s$–wave capture rates, to the deepest bound states available in both cases.

In these figures, we only considered capture via photon emission, even though for some of the parameter space on both plots, on-shell Z-emission is also allowed. We did however, take into account that for $\frac{8\delta M}{M_\chi v^2} < 1$, the charged component of the wavefunction, ψ_C, is no longer exponentially suppressed at large radii. For example, this covers the $M_\chi > 13$ TeV range of $v_{\rm rel} = 10^{-2}$ in both plots. We see that as expected for a short-range potential, at lower M_χ we have a pronounced velocity suppression for the p-wave initial state. However, at higher masses, where we are in the SU(2)-symmetric limit, the velocity suppression is lifted, and we expect the slower WIMPs to cross over to having a larger capture rate, scaling like $1/v_{\rm rel}$, which saturates once $v_{\rm rel} < m_W/M_\chi$.

We observe that the $v_{\rm rel} = 10^{-2}$ case does not respect this scaling, with rates that can be much higher than the $v_{\rm rel} = 10^{-3}$ case. Our simple analytic results neglect the contribution from the continuum states that experience a *repulsive* Coulomb potential, as discussed in appendix 7.3.3, on the grounds that this contribution is exponentially suppressed at low velocities. The suppression scales as $e^{-2\pi\alpha_W/v_{\rm rel}}$ in the cross section (see eq. 7.70), so when $v_{\rm rel}$ becomes comparable to α_W, this term can no longer be clearly neglected. More generally, we have worked in the limit of small $v_{\rm rel} \ll \alpha_W$ throughout this calculation; as $v_{\rm rel}$ becomes large our analytic results

should not be expected to describe the full behavior of the system.

8.4.3 Capture vs direct annihilation

From figure 8-7, we see that in the wino case the capture rate is quite suppressed relative to direct annihilation, across almost the whole range of possible DM masses. This contrasts with the case of e^+e^- annihilation, where capture into positronium dominates direct annihilation at low relative velocities. In this subsection we explore the origin of this difference, and how it might generalize to other complex dark sectors. As previously, we proceed by examining the limit where SU(2) is unbroken.

For positronium, where $\lambda_i = \lambda_f = 1$, all the gauge factors are trivial, diagrams of the form shown in figure 8-5 are forbidden (i.e. $\hat{C}_2 = 0$, in the notation of appendix D.3), and there is only a single relevant two-body state (e^+e^-), we obtain the cross section for capture into the positronium ground state from eq. 7.71 as:

$$\sigma v_{\rm rel} = \frac{2^{10}\pi^2}{3} e^{-4}\, \alpha^3 \frac{1}{m_e^2} \frac{1}{v_{\rm rel}}. \tag{8.50}$$

For the wino, as discussed above, the high-mass limit of the cross section for capture into the ground state is zero. The cross section for capture into the $2p$ states is:

$$\sigma v_{\rm rel} = \frac{2^{12}\pi^2 \times 19}{3} e^{-16} \alpha\, \alpha_W^2 \frac{1}{M_\chi^2} \frac{1}{v_{\rm rel}}. \tag{8.51}$$

We see that the numerical prefactor is smaller by a factor of $\sim 5 \times 10^{-4}$ for the wino compared to positronium; the factor of e^{-16} vs. e^{-4} from the overlap integral suppresses the rate for the wino, and is not fully compensated by other numerical prefactors.

Now let us consider the rate for direct annihilation. The Sommerfeld enhancement at low velocities and for massless force carriers is $2\pi\alpha/v_{\rm rel}$. The spin-averaged annihilation cross section for $e^+e^- \to \gamma\gamma$ *without* accounting for Sommerfeld enhancement

is $\sigma v_{\rm rel} = \pi\alpha^2/m_e^2$ (e.g. [224]). Thus the enhanced cross section is:

$$\sigma v_{\rm rel} = \frac{2\pi^2\alpha^3}{m_e^2\, v_{\rm rel}}. \tag{8.52}$$

For direct annihilation (into all channels), on the other hand, the leading-order spin-averaged s-wave annihilation rate for the wino is:

$$\sigma v_{\rm rel} = \frac{2\pi\alpha_W^2}{M_\chi^2} \begin{pmatrix} \psi_N^*(0) & \psi_C^*(0) \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{3}{2} \end{pmatrix} \begin{pmatrix} \psi_N(0) \\ \psi_C(0) \end{pmatrix}. \tag{8.53}$$

Assuming that only the attracted eigenstate gives a non-negligible contribution to the wavefunctions at the origin, and that this eigenstate has eigenvector η and eigenvalue λ, we obtain:

$$\begin{aligned} \sigma v_{\rm rel} &= \frac{2\pi\alpha_W^2}{M_\chi^2} |\mathbf{I}\cdot\eta|^2 \, |\phi(\lambda\alpha_W;0)|^2 \, \eta^\dagger \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{3}{2} \end{pmatrix} \eta \\ &= \frac{2\pi\alpha_W^2}{M_\chi^2} \frac{1}{3} \, |\phi(2\alpha_W;0)|^2 \begin{pmatrix} \sqrt{\frac{1}{3}} & \sqrt{\frac{2}{3}} \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{3}{2} \end{pmatrix} \begin{pmatrix} \sqrt{\frac{1}{3}} \\ \sqrt{\frac{2}{3}} \end{pmatrix}. \end{aligned} \tag{8.54}$$

We have $|\phi(\lambda\alpha;0)|^2 \approx 2\pi\lambda\alpha/v_{\rm rel}$ for small velocities, from our earlier results (this also cross-checks our Sommerfeld enhancement formula for positronium). Thus overall we obtain:

$$\sigma v_{\rm rel} = \frac{2^4\pi^2\alpha_W^3}{3M_\chi^2 v_{\rm rel}}. \tag{8.55}$$

We see that the cross section is slightly larger for the wino than one would expect from a naive extrapolation from positronium; the presence of multiple channels and the stronger coupling (since $\lambda = 2$) outweighs the penalty factor from the non-trivial overlap between the initial conditions and the attracted state. While for positronium, the capture/annihilation ratio is $2^9 e^{-4}/3 \approx 3$, for the wino we expect it to be $2^8 \times 19 e^{-16}(\alpha/\alpha_W) \approx 5 \times 10^{-4}(\alpha/\alpha_W) \approx 10^{-4}$.

With regard to general dark sectors, we see that a large attractive eigenvalue for

the initial state boosts the rate for direct annihilation by a factor λ, but suppresses the capture rate by an exponential factor (that depends on the ratio of this eigenvalue to the attractive eigenvalue of the potential supporting the final state). Thus in general, smaller attractive eigenvalues for the initial state (and also larger attractive eigenvalues for the bound state) will tend to boost the capture/annihilation ratio.

8.4.4 Discussion

The capture rate we have derived for the wino is very small, consistently well below the direct annihilation cross section. Furthermore, at low masses the dominant capture mode for velocities typical of the Milky Way halo is to the $1s$ spin-triplet state, which is a pure-chargino bound state that subsequently decays dominantly through s-channel annihilation to SM quarks. Thus the presence of bound states will not directly enhance the annihilation rate by a significant fraction, and in particular will not enhance the gamma-ray line cross section. Previous calculations of the gamma-ray line signal, neglecting the impact of radiative capture into bound states, thus remain valid.

One might ask to what degree this conclusion is generic to complex dark sectors, where the DM interacts through the exchange of multiple force carriers and may have nearly-degenerate partner particles. Compared to the case of positronium, where the capture rate dominates the direct annihilation rate by a factor of a few at low velocities, there are three principal sources of suppression of the capture cross section for the wino:

- Only some fraction of the propagating two-particle state couples to the radiated particle (the photon, in our case), leading to $\mathcal{O}(1)$ suppression factors. Thus, for example, the capture cross section scales as $\alpha\,\alpha_W^2$ in the high-mass limit, whereas the direct annihilation cross section scales as α_W^3. Factors of this form will be generic in complex dark sector models, although their exact size will vary.

- For positronium the capture into the ground state dominates but for the wino

this capture rate is generically suppressed, as it vanishes in the Coulombic limit due to an accidental cancellation. This suppression is not universal to other dark sector models. For models where this term does not vanish in the Coulomb limit it still may be suppressed at low velocities since it involves a p-wave initial state; this suppression does not affect the s-wave direct annihilation cross section. This factor depends on the mass of the force carriers relative to the mass of the DM; in non-electroweakino DM models, there is much greater freedom to adjust the force carrier mass and hence the degree of velocity suppression. For example, lowering the force carrier masses will reduce the effect of the velocity suppression on the capture rate from higher-partial-wave components of the continuum wavefunction, since this velocity suppression scales as $(M_\chi v/m_W)^{2L}$. Also, for $m_W \ll M_\chi$, we enter the Coulombic regime, where there is no velocity suppression for higher partial waves, and we recover a $1/v$ scaling in the capture rate.

- There is also the apparently accidental $e^{-4n\lambda_i/\lambda_f}$ factor appearing in the cross section for capture, arising from the overlap integral between the continuum and bound states, which does not affect the direct annihilation cross section. For positronium and the case of capture to the ground state, $\lambda_i = \lambda_f$ and this factor is just e^{-4}. For the wino case, where $\lambda_i = 2 = 2\lambda_f$, this factor is e^{-8} at most, and it increasingly suppresses capture into bound states with higher principal quantum number. This factor is *not* universal; for example, for a simple two-state model coupled to a single force carrier [196] we have $\lambda_i/\lambda_f = 1$, for fermionic DM transforming as a SU(2) doublet (quintuplet) we find $\lambda_i/\lambda_f = 1$ ($\frac{6}{5}$ and $\frac{3}{5}$, as the quintuplet has two eigenvectors that experience an attractive potential, both of which can contribute to capture) [194].

Dark sectors where the bound states experience a stronger attractive potential than the continuum states, or where the ratio of force carrier mass to DM mass is not much larger than typical velocities in the Milky Way halo, are therefore more likely to have large cross sections for capture relative to direct annihilation.

One might also ask whether the photons radiated on capture *themselves* constitute a detectable signal. In principle, detecting lines from capture and/or transitions between bound states could allow study of the quantum numbers of the DM. However, because the capture rate for wino-like DM is so low and the mass scales where bound states occur are quite high, for this particular toy model this would be an extremely challenging search. Assuming an NFW DM profile with local DM density $\rho(8.5\text{kpc}) = 0.4\,\text{GeV/cm}^3$ and scale radius 20 kpc, and (as a benchmark) 10 TeV DM with a capture cross section of 10^{-29} cm^3/s, we find that on average one would receive $\mathcal{O}(10^{-4})$ photons/m^2/yr at Earth from the whole Milky Way halo. From the region within 1 degree of the Galactic center, the rate is instead $\mathcal{O}\,(10^{-5})$ photons/m^2/yr. This rate is prohibitively small for any reasonable space-based telescope. Ground-based gamma-ray telescopes, on the other hand, can have effective areas of $\sim 10^{5-6}$ m^2 and so might be able to observe a very small number of capture photons – but current and near-future ground-based telescopes have low-energy thresholds in the $10 - 20$ GeV range or higher, which would need to be lowered by an order of magnitude to observe capture and transition photons from $\mathcal{O}(10)$ TeV DM (for which the deepest bound states accessible by capture have $E_n \sim 1$ GeV), and would likely also need excellent energy resolution in order to isolate such a small line signal from substantial astrophysical backgrounds. Higher DM masses would produce capture line photons with higher energies – e.g. 25 GeV for 100 TeV DM – but would also correspond to a much lower DM number density, suppressing the already-low rate of possible detections. However, if an annihilation signal had already been detected, such a search would be well-motivated, and might provide one of the only ways to probe the particle properties of the DM in the absence of a discovery at a collider. Detection of a high-energy annihilation signal would also open up other options in searching for the capture transition lines, for example by examining cross-correlations with the DM annihilation spatial distribution.

8.5 Conclusions

We have computed the rate for formation of wino-onium bound states, and their subsequent decays to lower-energy states or SM particles. We find that bound state formation by single photon emission is possible for large wino masses, $M_\chi \gtrsim 5.6$ TeV, but in general, and in contrast to the case of positronium, the capture rate is subdominant to direct annihilation. Consequently, previous calculations of the detectability of e.g. high-energy gamma-ray lines from wino DM should not require significant modification in most of parameter space.

This scenario has several novel features relative to the case of positronium, or dark-sector configurations where there is only one DM state and the potential is mediated by a single dark photon. Many of these features will generalize to any complex dark sector where the gauge group is nonabelian and the potential couples together several nearly-degenerate dark-matter-like states.

Spin statistics demands that only two-particle states with even $L+S$ can possess a $\chi^0\chi^0$ component; states with odd $L+S$ must in this case be entirely comprised of $\chi^+\chi^-$. Consequently, states with odd vs even $L+S$ experience different effective potentials and form distinct towers of bound states, which will generically be displaced from each other in energy. The unsuppressed decay channels to lower-energy bound states may thus be very different to the familiar case of hydrogen-like atoms. The annihilation channels of the two towers of states are also quite different; for the wino, states with even $L+S$ decay primarily to gauge bosons, whereas those with odd $L+S$ decay primarily through an s-channel diagram to quarks and leptons.

The presence of massive force carriers generically suppresses the capture cross section at low velocities, by suppressing all contributions from initial states with $L > 0$. However, the distortion of the continuum wave functions due to the presence of near-threshold bound states can lead to resonant enhancement of the capture cross section, in the same way that resonant Sommerfeld enhancement leads to a larger direct annihilation cross section. Furthermore, for the wino and for velocities typical of the Milky Way halo, the capture rate can have a significant velocity dependence,

in contrast to direct annihilation.

Detection of the low-energy photon lines ($\mathcal{O}$(GeV) energies for 10 TeV+ DM) from radiative capture and transitions between bound states could potentially provide a unique probe into the gauge structure of the dark sector. However, for the heavy wino this search appears very challenging, due to the low number density of multi-TeV DM; experiments designed to search for high-energy gamma rays have large enough effective areas to observe these photons, but their energy threshold is presently too high to have sensitivity, and furthermore the gamma-ray backgrounds at these low energies are substantial.

In contrast to the features discussed above, the factors which suppress the winoonium capture cross section are *not* generic; they depend sensitively on the representation of the DM under the gauge group, and the relative masses of the DM and force carriers. Thus the formation of bound states cannot be safely ignored in models with non-trivial dark sectors. We have presented general analytic results for the capture rate into DM bound states in the limit where the force carriers are very light and the gauge symmetry is approximately unbroken, to facilitate estimates of whether the capture rate can be important for a given dark-sector model. In such models, the presence of bound states could enhance the capture rate, change the branching ratio to different SM final states, and perhaps generate non-negligible transition lines – although if the dark gauge group is not the electroweak gauge group, the transition lines would presumably be comprised of "dark photons", and their observable signatures would depend on the coupling of those dark photons to the SM.

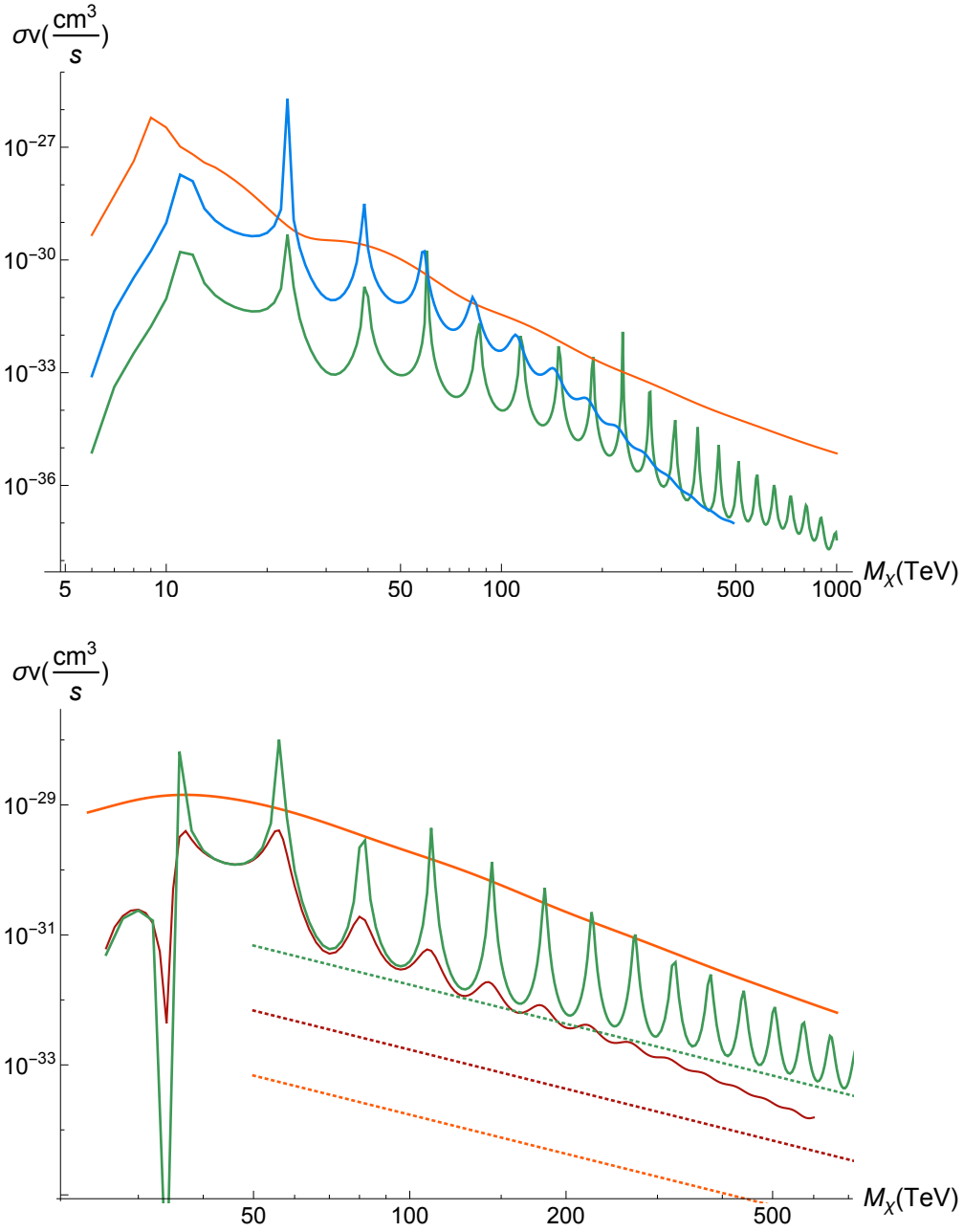

Figure 8-8: **Top:** Rate for $\chi^0\chi^0(p\text{–wave}) \to {}^3S_1 + \gamma$, considering only capture into the deepest bound state. **Bottom:** Rate for $\chi^0\chi^0(s\text{–wave}) \to {}^1P_1 + \gamma$, considering only capture into the deepest bound states. Velocities are $v_{\rm rel} = 10^{-2}$ (*orange*), $v_{\rm rel} = 10^{-3}$ (*blue*: *p*-wave; *maroon*: *s*-wave), and $v_{\rm rel} = 10^{-4}$ (*green*). All colored *dotted* lines display $3 \times \sigma v$, the analytic predictions from the Coulomb limit (see the text for an explanation of the factor of 3).

Chapter 9

Conclusion

In this thesis I presented a framework to numerically solve the equations of motion which describe the evolution of inhomogeneities in the early universe, in a way that is much more efficient than simulations involving full numerical general relativity. This framework captures the dominant nonlinear effect of inhomogeneities influencing the surrounding spacetime, and vice-versa. The results I presented address to what extent inflation can occur under inhomogeneous initial conditions in the very early universe. The efficiency of this approach allows us to simulate the onset of inflation in a wider range of models and over more broad regions of the phase space of initial conditions than the numerical-relativity simulations can probe. I presented results which demonstrate this method on a simple inflationary model in a spacetime with zero spatial curvature, where we find that cosmic inflation is robust in the face of significant initial inhomogeneities.

This framework can be extended to investigate the robustness of cosmic inflation in a wide variety of models. For example, one interesting direction forward is to investigate the onset of inflation amid inhomogeneous initial conditions in models in which cosmic inflation is driven by multiple interacting matter fields. Multi-field inflation is well-motivated by particle physics and string theory, and gives rise to distinctive experimental signatures. The close agreement between the spectrum of primordial inhomogeneities predicted by inflation and precision measurements of the CMB is one of the most important recent accomplishments of cosmology. Multi-field

inflation can generate unique features in the primordial inhomogeneities which are highly constrained by CMB measurements, providing a powerful discriminating tool for constraining models of inflation. For example, string theory leads not to one fully specified inflationary model, but to an ensemble of possible models; comparing their experimental signatures with observations and testing their robustness to inhomogeneous initial conditions can allow us to distinguish which inflationary models may be realistic descriptions of nature.

In this thesis I also fully explored the vector-portal dark matter model, for both symmetric and inelastic DM models, in the regime where the ratio of the dark photon mass $m_{A'}$ to the dark matter mass m_χ is in the range $1 \lesssim m_{A'}/m_\chi \lesssim 2$, carefully tracking the separate dark sector and SM temperatures throughout freezeout. This reveals a rich set of new mechanisms for obtaining the correct DM relic density, which lead to interesting cosmological histories in which the separate temperature evolution of the dark sector plays an important role, and which open new windows of parameter space for the DM where experiments have not yet explored.

This demonstrates the importance of carefully searching even well-understood hidden sector models, as well as novel DM candidates, taking into account the important effects that the temperature evolution of the dark sector can have. This is a lesson to take forward as we continue exploring new mechanisms for producing dark matter. It is now a particularly important time to continue exploring the landscape of models which produce DM at light (sub-GeV) masses as new direct detection experiments will soon explore the parameter space of light DM with unprecedented sensitivity [96].

In this thesis I also presented a calculation of the rate for the wino to bind into WIMPonium and to subsequently decay to lower-energy states or SM particles. Extending beyond the case of electromagnetism to the formation of bound states of dark matter particles which inhabit a more general nonabelian dark sector reveals rich structures which are not present in the case of electromagnetism, and which could be revealed by astrophysical detection of photon lines from radiative capture and transitions between WIMPonium bound states. For the wino such photons are rare, but might be observable by a future ground-based gamma-ray telescope combin-

ing large effective area and a low energy threshold. WIMPonium can also decay into radiation which provides an additional astrophysical detection signal. In the specific case of the wino the rate for bound state formation provides only a small correction to the overall annihilation rate.

There are many open directions to explore about dark sector bound states and their implications for cosmology, such as scenarios where forces between DM particles in a bound state are mediated by a pseudo-scalar or axial vector, and the discovery potential for DM bound states at the Large Hadron Collider.

The work presented in this thesis represents only a small part of what we think we understand about cosmology, and our understanding is changing daily.

Appendix A

Appendix for Chapter 3

A.1 Discrete Spectrum for Mode Functions

As described in Section 3.3, we expand quantized field fluctuations $\delta\hat{\phi}\,(x^\mu)$ and metric perturbations $\hat{\Psi}\,(x^\mu)$ in eigenfunctions $Z_{n\ell m}\,(\mathbf{x})$ of the comoving spatial Laplacian. Within a comoving spatial volume of finite size, Eq. (3.12) then takes the form

$$\nabla^2 Z_{n\ell m}(r,\theta,\phi) = -k_{n\ell}^2\, Z_{n\ell m}(r,\theta,\phi), \tag{A.1}$$

with positive integer $n \geq 1$. We restrict attention to a finite sphere of comoving radius R within a spatially flat FLRW background spacetime ($K = 0$). We select Dirichlet boundary conditions $Z_{n\ell m}(R,\theta,\phi) = 0$, which causes ∇^2 to have a negative definite spectrum as desired, and for the resulting Sturm-Liouville system to yield a complete, orthonormal basis:

$$\int_{r<R} d^3\mathbf{x}\sqrt{h(\mathbf{x})}\, Z_{n\ell m}(\mathbf{x})\, Z^*_{n'\ell'm'}(\mathbf{x}) = \delta_{nn'}\delta_{\ell\ell'}\delta_{mm'}. \tag{A.2}$$

In spherical polar coordinates, Eq. (A.1) becomes

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial Z_{n\ell m}}{\partial r}\right) - \frac{1}{r^2}\mathbf{L}^2 Z_{n\ell m} + k_{n\ell}^2 Z_{n\ell m} = 0, \tag{A.3}$$

where

$$\mathbf{L}^2 = -\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) - \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}. \tag{A.4}$$

The eigenfunctions of the operator $\mathbf{L}^2$ are the familiar spherical harmonics $Y_{\ell m}(\theta,\phi)$, which satisfy $\mathbf{L}^2 Y_{\ell m} = \ell(\ell+1)\,Y_{\ell m}$. Solutions to Eq. (A.3) may then be written in the form

$$Z_{n\ell m}(r,\theta,\phi) = N_{n\ell m}\, j_\ell(k_{n\ell} r)\, Y_{\ell m}(\theta,\phi), \tag{A.5}$$

where $N_{n\ell m}$ is a normalization constant, $j_\ell(x)$ is a spherical Bessel function of order ℓ, and the boundary conditions require

$$j_\ell(k_{n\ell} R) = 0. \tag{A.6}$$

The requirement of Eq. (A.6) yields a discrete spectrum of allowable wavenumbers,

$$k_{n\ell}(R) = \frac{x_{n\ell}}{R} \tag{A.7}$$

where, as noted below Eq. (3.33), $x_{n\ell}$ is the nth zero of the Bessel function $j_\ell(x)$ for $n \geq 1$. Making use of the orthonormality properties of the $Y_{\ell m}(\theta,\phi)$, Eq. (A.2) then becomes

$$N_{n\ell m}\, N_{n'\ell m} \int_0^R dr\, r^2\, j_\ell(k_{n\ell} r)\, j_\ell(k_{n'\ell} r) = \delta_{nn'}. \tag{A.8}$$

Upon using Eqs. (11.49) and (11.50) of Ref. [232], we find

$$N_{n\ell m} = \frac{\sqrt{2}}{R^{3/2}} \frac{1}{|j_{\ell+1}(k_{n\ell} R)|}, \tag{A.9}$$

with (as usual) $j_\ell(k_{n\ell} R) = 0$. Note that the normalization $N_{n\ell m}$ is independent of m. The basis functions $Z_{n\ell m}(r,\theta,\phi)$ provide spectral convergence to any spatial configuration as we take $n_{\max} \to \infty$.

Given the spatial eigenfunctions $Z_{n\ell m}(\mathbf{x})$ and the properties of the operators $\hat{a}_{n\ell m}$,

$\hat{a}^\dagger_{n\ell m}$, we may evaluate various two-point functions. For example, we have

$$\langle 0|\delta\hat{\phi}(t,\mathbf{x})\,\delta\hat{\phi}(t,\mathbf{y})|0\rangle = \sum_{n\ell m} |\delta\phi_{n\ell m}(t)|^2 Z_{n\ell m}(\mathbf{x})\, Z^*_{n\ell m}(\mathbf{y}). \tag{A.10}$$

The term $\langle(\delta\hat{\phi})^2\rangle$, which appears throughout the equations of motion in the Hartree approximation, is Eq. (A.10) in the limit $\mathbf{x} \to \mathbf{y}$. As our slicing of spacetime has been chosen such that spatial slices are homogeneous and isotropic (this holds even inside the finite sphere due to completeness), we may evaluate Eq. (A.10) at $\mathbf{x} = \mathbf{y} = \mathbf{0}$. From

$$j_\ell(z) \to \frac{z^\ell}{1\cdot 3\cdot 5\cdot \ldots \cdot (2\ell+1)} \quad \text{for } z \to 0 \tag{A.11}$$

we note that only the $\ell = 0$ mode remains nonzero at the origin, with amplitude $j_0(0) = 1$. We further note that $Y_{00}(\theta,\phi) = 1/\sqrt{4\pi}$ for $\ell = m = 0$, and hence we find

$$\langle(\delta\hat{\phi})^2\rangle = \frac{\pi}{2R^3}\sum_n n^2 |\delta\phi_{n00}(t)|^2, \tag{A.12}$$

as in Eq. (3.38), prior to applying the UV regularization. The same steps yield $\langle(\delta\dot{\phi})^2\rangle$ as in Eq. (3.38).

The remaining two-point function of interest is the contribution to $\delta\rho_{(2)}$ from the spatial gradients. We find

$$\begin{aligned}\langle(\nabla\delta\hat{\phi})^2\rangle = \sum_{n\ell m} &|\delta\phi_{n\ell m}(t)|^2 \\ &\times \Big\{ h^{rr}\partial_r Z_{n\ell m}(\mathbf{0})\,\partial_r Z^*_{n\ell m}(\mathbf{0}) \\ &\quad + h^{\theta\theta}\partial_\theta Z_{n\ell m}(\mathbf{0})\,\partial_\theta Z^*_{n\ell m}(\mathbf{0}) \\ &\quad + h^{\phi\phi}\partial_\phi Z_{n\ell m}(\mathbf{0})\,\partial_\phi Z^*_{n\ell m}(\mathbf{0}) \Big\}.\end{aligned} \tag{A.13}$$

Using Eq. (A.11), we see that only the term with $\ell = 1$ will contribute to the first term in brackets within Eq. (A.13). From the properties of $Y_{1m}(0,\phi)$ and $N_{n\ell m}$ in

Eq. (A.9), we then find

$$h^{rr}\partial_r Z_{n\ell m}(\mathbf{0})\,\partial_r Z^*_{n\ell m}(\mathbf{0}) = \frac{1}{6\pi R^3}\frac{k_{n1}^2}{|j_2(k_{n1}R)|^2}\delta_{\ell,1}\,\delta_{m,0}. \tag{A.14}$$

For the last two terms in brackets in Eq. (A.13), we require that the terms be well-behaved in the vicinity of the origin. Given $h^{\theta\theta} = 1/r^2$ and the properties of the $Y_{\ell m}(\theta,\phi)$, the only contribution to the second term in brackets that will remain regular (and nonzero) near $\mathbf{x} \to \mathbf{0}$ comes from $\ell = 1$, and we find

$$\begin{aligned}\sum_{n\ell m} h^{\theta\theta}\,\partial_\theta Z_{n\ell m}(\mathbf{0})\,\partial_\theta Z^*_{n\ell m}(\mathbf{0}) \\ = \sum_n \frac{1}{12\pi R^3}\frac{k_{n1}^2}{|j_2(k_{n1}R)|^2}\delta_{\ell,1}\left[\delta_{m,1}+\delta_{m,-1}\right].\end{aligned} \tag{A.15}$$

Proceeding similarly, we find

$$\begin{aligned}\sum_{n\ell m} h^{\phi\phi}\,\partial_\phi Z_{n\ell m}(\mathbf{0})\,\partial_\phi Z^*_{n\ell m}(\mathbf{0}) \\ = \sum_n \frac{1}{12\pi R^3}\frac{k_{n1}^2}{|j_2(k_{n1}R)|^2}\delta_{\ell,1}\left[\delta_{m,1}+\delta_{m,-1}\right].\end{aligned} \tag{A.16}$$

Combining Eqs. (A.13) - (A.16), we find

$$\langle(\nabla\delta\hat{\phi})^2\rangle = \frac{1}{6\pi R^3}\sum_n \sum_{m=-1,0,1}\frac{k_{n1}^2}{|j_2(k_{n1}R)|^2}|\delta\phi_{n1m}(t)|^2, \tag{A.17}$$

as in Eq. (3.38).

A.2 Initial Conditions for the Field Fluctuations

To establish the initial conditions for the mode functions $\delta\phi_{n\ell m}(t_0)$ and $\delta\dot{\phi}_{n\ell m}(t_0)$, we first consider the case of $R \to \infty$, for which the spectrum of allowable wavenumbers is continuous, with $0 \le k < \infty$. As noted in Section 3.3, within the Hartree approximation higher-order interaction terms among the fluctuations $\delta\hat{\phi}$ are replaced by an effective mass, so we may proceed by following most of the steps for quantizing a free,

massive scalar field in FLRW spacetime.

The equal-time commutation relation for a free scalar field stipulates

$$[\delta\hat{\phi}(t,\mathbf{x}), \delta\hat{\Pi}(t,\mathbf{y})] = i\delta^{(3)}(\mathbf{x}-\mathbf{y}), \tag{A.18}$$

where

$$\delta\hat{\Pi} \equiv \frac{\partial\mathcal{L}}{\partial\,\dot{\delta\hat{\phi}}} = a^3(t)\,\dot{\delta\hat{\phi}} \tag{A.19}$$

is the momentum canonically conjugate to $\delta\hat{\phi}$. Upon expanding $\delta\hat{\Pi}(x^\mu)$ in a series of complex mode functions $\delta\Pi_{k\ell m}(t)$ and creation and annihilation operators akin to Eq. (3.10), and using the commutation relations for $\hat{a}_{k\ell m}$ and $\hat{a}^\dagger_{k\ell m}$ in Eq. (3.13), Eq. (A.18) imposes a constraint on the mode functions:

$$\delta\phi_{k\ell m}\,\delta\Pi^*_{k\ell m} - \delta\phi^*_{k\ell m}\,\delta\Pi_{k\ell m} = i. \tag{A.20}$$

We rescale the modes $\delta\phi_{k\ell m}(t) = v_{k\ell m}(t)/a(t)$ and introduce conformal time, $d\tau \equiv dt/a(t)$, so that

$$\delta\Pi_{k\ell m}(\tau) = a(\tau)(v'_{k\ell m} - \mathcal{H}v_{k\ell m}), \tag{A.21}$$

where primes denote $d/d\tau$ and $\mathcal{H} \equiv a'/a$. Then Eq. (A.20) becomes

$$v_{k\ell m}\,v^{*\prime}_{k\ell m} - v^*_{k\ell m}\,v'_{k\ell m} = i, \tag{A.22}$$

a Wronskian condition that we will use when setting initial conditions at τ_0.

In terms of $v_{k\ell m}(\tau)$, Eq. (3.18) takes the form

$$v''_{k\ell m} + \omega_k^2(\tau)\,v_{k\ell m} = S_{k\ell m}(\tau), \tag{A.23}$$

where the source term $S_{k\ell m}$ depends on $\Psi_{k\ell m}$ and $\Psi'_{k\ell m}$, and the frequency is given by

$$\omega_k^2(\tau) = k^2 + a^2(\tau)\,m^2_{\rm eff}(\tau) - \frac{a''(\tau)}{a(\tau)}, \tag{A.24}$$

with the effective mass $m^2_{\rm eff}$ given in Eq. (3.34). The effective mass is suppressed

by a small coupling constant λ and hence we expect it to remain subdominant around τ_0, while, on dimensional grounds, we expect $a''/a \sim \mathcal{O}(H^2)$ around τ_0. When we numerically evolve the modes within a region of finite comoving radius R in our simulation, we introduce an infrared cut-off given by $k_{\rm min} = \pi/R$, with $R \sim 1/[a(\tau_0)H(\tau_0)]$. Hence even for the longest-wavelength modes in our simulation, we expect $k^2 \gtrsim \{a^2(\tau_0)\, m^2_{\rm eff}(\tau_0), a''(\tau_0)/a(\tau_0)\}$. For setting initial conditions, we therefore approximate $\omega_k(\tau_0) \sim k$.

We do not assume that the system has attained its minimum energy state at the initial time τ_0, and hence we consider initial conditions for the modes that could depart from the usual Bunch-Davies vacuum state. (Note that as we are describing a weakly interacting system, the Bunch-Davies state that we construct is only an approximation of the lowest-energy state, which we do not attempt to find.) We parameterize the initial conditions for the modes as

$$\begin{aligned} v_{k\ell m}(\tau_0) &= \frac{1}{\sqrt{2k}}\,(\alpha_{k\ell m} + i\beta_{k\ell m})\,, \\ v'_{k\ell m}(\tau_0) &= -i\sqrt{\frac{k}{2}}\,(\gamma_{k\ell m} + i\delta_{k\ell m})\,, \end{aligned} \tag{A.25}$$

where $\alpha_{k\ell m}$, $\beta_{k\ell m}$, $\gamma_{k\ell m}$, and $\delta_{k\ell m}$ are each real-valued, dimensionless constants. (In the limit $\omega_k \to k$, the Bunch-Davies state corresponds to $\alpha_{k\ell m} = \gamma_{k\ell m} = 1$, $\beta_{k\ell m} = \delta_{k\ell m} = 0$.) The Wronskian condition of Eq. (A.22) then becomes

$$\alpha_{k\ell m}\gamma_{k\ell m} + \beta_{k\ell m}\delta_{k\ell m} = 1. \tag{A.26}$$

Without loss of generality, we may set $\beta_{k\ell m} = 0$ for all modes; then Eq. (A.26) fixes $\alpha_{k\ell m} = 1/\gamma_{k\ell m}$. Setting $a(\tau_0) = 1$, Eqs. (A.21) and (A.25) yield

$$\begin{aligned} \delta\phi_{k\ell m}(t_0) &= \frac{\alpha_{k\ell m}}{\sqrt{2k}}, \\ \delta\dot{\phi}_{k\ell m}(t_0) &= \sqrt{\frac{k}{2}}\left(-i\gamma_{k\ell m} + \delta_{k\ell m} - \frac{\alpha_{k\ell m}\bar{H}_0}{k}\right). \end{aligned} \tag{A.27}$$

We use $\bar{H}_0$ (rather than H_0) in Eq. (A.27) because, as a practical matter, we must

first select a point in $(\varphi_0, \dot{\varphi}_0)$ and then initialize the fluctuations. The parameter $\bar{H}_0$, defined in Eq. (3.36), is determined by the selection $(\varphi_0, \dot{\varphi}_0)$; only after the fluctuations are initialized can we evaluate $\delta\rho_{(2)}(t_0)$ and thereby include their contribution to H_0. (For the regimes of interest, $H_0/\bar{H}_0 \lesssim 1.4$, so the difference is not significant.) When we evolve the system within a finite sphere of comoving radius R, the continuous variable k in Eqs. (A.26) and (A.27) is replaced by the discrete spectrum of wavenumbers $k_{n\ell}$ with $n \geq 1$, which yields Eq. (3.35).

Given these initial conditions for the modes, we may estimate quantities of interest such as $\langle(\delta\hat{\phi}(t_0))^2\rangle$ and $\delta\rho_{(2)}(t_0)$. We derive our expressions in the continuum limit using the UV regulator function $F_{n\ell}(\kappa, R)$ of Eq. (3.39). For the initial value of the two-point function, we begin with the expansion for $\delta\hat{\phi}(x^\mu)$ in Eq. (3.10), use Eq. (3.13), make use of the properties of $Z_{k\ell m}(\mathbf{x})$ at $\mathbf{x} \to \mathbf{0}$, and substitute $\delta\phi_{k\ell m}(t_0)$ from Eq. (A.27) to write

$$\begin{aligned}
&\langle(\delta\hat{\phi}(t_0))^2\rangle \\
&\quad = \int dk \sum_{\ell,m} |\delta\phi_{k\ell m}(t_0)|^2 |Z_{k\ell m}(\mathbf{0})|^2 \, e^{-k^2/(2\kappa^2)} \\
&\quad = \int_0^\infty \frac{dk}{2\pi^2} k^2 \left(\frac{\alpha_{k00}^2}{2k}\right) e^{-k^2/(2\kappa^2)} \\
&\quad = \left(\alpha^2\right)_{\text{avg}} \frac{\kappa^2}{4\pi^2}.
\end{aligned} \tag{A.28}$$

The third line follows because the random coefficients $\alpha_{k\ell m}$ vary independently of k, so we may take the term α_{k00}^2 outside the integral and replace it with an average value; moreover, since we draw the variables $\gamma_{k\ell m} = 1/\alpha_{k\ell m}$ from the same distribution for all $(k\ell m)$, the same average value $(\alpha^2)_{\text{avg}}$ holds for any k, ℓ, and m. Upon using $\kappa = b\bar{H}_0$, we arrive at Eq. (3.41).

We proceed similarly to estimate $\delta\rho_{(2)}(t_0)$. Using Eqs. (3.10), (3.13), (3.22), (3.39),

and (A.27), we find

$$
\begin{aligned}
&\delta\rho_{(2)}(t_0) \\
&\quad = \frac{1}{2}\int dk \sum_{\ell,m} |\delta\dot{\phi}_{k\ell m}(t_0)|^2 |Z_{k\ell m}(\mathbf{0})|^2 \, e^{-k^2/(2\kappa^2)} \\
&\qquad + \frac{1}{2a^2(t_0)}\int dk \sum_{\ell,m} |\delta\phi_{k\ell m}(t_0)|^2 |\nabla Z_{k\ell m}(\mathbf{0})|^2 \, e^{-k^2/(2\kappa^2)} \\
&\quad = \frac{1}{2}\int dk \frac{k^2}{2\pi^2}\frac{k}{2}\Bigg\{ \gamma_{k00}^2 + \delta_{k00}^2 \\
&\qquad\qquad - \frac{2\alpha_{k00}\delta_{k00}\bar{H}_0}{k} + \frac{\alpha_{k00}^2\bar{H}_0^2}{k^2} \Bigg\} e^{-k^2/(2\kappa^2)} \\
&\qquad + \frac{1}{2a^2(t_0)}\int dk \frac{k^2}{6\pi^2} \sum_{m=-1,0,1} k^2 \left(\frac{\alpha_{k1m}^2}{2k}\right) e^{-k^2/(2\kappa^2)} \\
&\quad = \frac{1}{8\pi^2}\int dk \Bigg\{ \left(\alpha^2 + \gamma^2 + \delta^2\right)_{\rm avg} k^3 \\
&\qquad\qquad - 2\left(\alpha\delta\right)_{\rm avg} \bar{H}_0 k^2 + \left(\alpha^2\right)_{\rm avg} \bar{H}_0^2 k \Bigg\} e^{-k^2/(2\kappa)^2} \\
&\quad = \frac{\kappa^4}{4\pi^2}\Bigg\{ \left(\alpha^2 + \gamma^2 + \delta^2\right)_{\rm avg} \\
&\qquad\qquad - \left(\alpha\delta\right)_{\rm avg} \sqrt{\frac{\pi}{2}} \left(\frac{\bar{H}_0}{\kappa}\right) + \frac{\left(\alpha^2\right)_{\rm avg}}{2} \left(\frac{\bar{H}_0}{\kappa}\right)^2 \Bigg\},
\end{aligned}
\tag{A.29}
$$

where we have used $a(t_0) = 1$ and assumed that $m_{\rm eff}^2(t_0) \ll H_0^2$ in the regimes of interest, so that the contributions to $\delta\rho_{(2)}(t_0)$ are dominated by the kinetic energy and spatial-gradient energy of the fluctuations. Given $\kappa = b\bar{H}_0 > \bar{H}_0$, we find Eq. (3.42).

The magnitudes of $\langle(\delta\hat{\phi}(t_0))^2\rangle$ and $\delta\rho_{(2)}(t_0)$ depend on the random variables $\alpha_{n\ell m}$, $\gamma_{n\ell m}$, and $\delta_{n\ell m}$ for each mode. We draw $\gamma_{n\ell m}$ and $\delta_{n\ell m}$ from flat distributions within specific ranges; once $\gamma_{n\ell m}$ is selected, $\alpha_{n\ell m}$ is fixed by the quantization condition to be $\alpha_{n\ell m} = 1/\gamma_{n\ell m}$. If we draw $\gamma_{n\ell m}$ from the range $\{A, B\}$ and $\delta_{n\ell m}$ from the range

$\{-C, C\}$, then we expect

$$\begin{aligned}
\left(\gamma^2\right)_{\rm avg} &= \frac{1}{(B-A)}\int_A^B dx\, x^2 = \frac{1}{3}\left(A^2 + AB + B^2\right), \\
\left(\alpha^2\right)_{\rm avg} &= \frac{1}{(B-A)}\int_A^B dx \frac{1}{x^2} = \frac{1}{AB}, \\
\left(\delta^2\right)_{\rm avg} &= \frac{1}{2C}\int_{-C}^{C} dx\, x^2 = \frac{C^3}{3}.
\end{aligned} \tag{A.30}$$

In order to study large initial fluctuations, with $\mathcal{C} \equiv \left(\alpha^2 + \gamma^2 + \delta^2\right)_{\rm avg} \simeq 20$, we therefore use the ranges in Eq. (3.44).

A.3 Initializing the Metric Perturbations

In this appendix, we discuss subtleties in the initialization of $\Psi_{n\ell m}(t_0)$, and the techniques we use to avoid these issues numerically.

Given an instantiation of $\delta\phi_{n\ell m}(t_0)$ and $\delta\dot{\phi}_{n\ell m}(t_0)$, the metric perturbations $\Psi_{n\ell m}(t_0)$ are initialized using Eq. (3.29). For our numerical simulations, we evolve the system within a sphere of comoving radius R, within which Eq. (3.29) takes the form

$$\begin{aligned}
&\left[\dot{H} + \frac{2}{3M_{\rm pl}^2 a^2}\langle(\nabla\delta\hat{\phi})^2\rangle + \frac{1}{a^2}\left(k_{n\ell}^2 - 3K\right)\right]\Psi_{n\ell m} \\
&\qquad = \frac{1}{2M_{\rm pl}^2}\left[\ddot{\varphi}\,\delta\phi_{n\ell m} - \dot{\varphi}\,\delta\dot{\phi}_{n\ell m}\right].
\end{aligned} \tag{A.31}$$

As $\dot{H}$ tends to be large and negative at early times, it is common for there to exist a value for $k_{n\ell}$ at which the coefficient of $\Psi_{n\ell m}$ vanishes at t_0: we dub the $k_{n\ell}$ value at which this occurs the "Ψ pole." For $k_{n\ell}$ near the pole, $\Psi_{n\ell m}(t_0)$ would be initialized with an artificially large initial amplitude, which can cause a single mode to dominate the nonlinear contributions.

The pole exists as an artifact of the longitudinal gauge, and arises because the four-dimensional phase space ($\delta\phi_{n\ell m}$, $\delta\dot{\phi}_{n\ell m}$, $\Psi_{n\ell m}$ and $\dot{\Psi}_{n\ell m}$) has two constraints, Eqs. (3.21) and (3.29). When $k_{n\ell}$ is on the Ψ pole, the constraint surface in phase space forces $\delta\phi_{n\ell m}$ and $\delta\dot{\phi}_{n\ell m}$ to be strictly proportional to each other. In this situation, we

do not lose a degree of freedom (the constraint surface remains two-dimensional); it is simply that $\delta\phi_{n\ell m}$ and $\delta\dot{\phi}_{n\ell m}$ cease to be appropriate coordinates on the constraint surface.

We emphasize that this does not arise due to our use of the Hartree approximation, but rather through imposing canonical commutation relations on $\delta\phi_{n\ell m}$ and $\delta\dot{\phi}_{n\ell m}$ at t_0 under the assumption that they are free fields. Within linear perturbation theory, this issue is typically overcome by using Mukhanov-Sasaki variables, $Q_{n\ell m} \equiv \delta\phi_{n\ell m} + (\dot{\varphi}/H)\Psi_{n\ell m}$, which are good coordinates on the phase space. However, the $Q_{n\ell m}$ are no longer gauge-invariant beyond linear order in perturbations. If one were to apply the (nonlinear) Hartree corrections to the $Q_{n\ell m}$, the resulting expressions that went beyond $\mathcal{O}(\Psi)$ would not bear any clear relationship to an expansion of the Einstein tensor $G_{\mu\nu}$ beyond $\mathcal{O}(\Psi)$.

The most correct way to approach this issue would be to employ Dirac's constrained Hamiltonian formalism to the constrained phase space, but such an analysis is beyond the scope of this work. For our present purposes, we adopt methods to numerically alleviate the issue of the Ψ pole.

We can compute the boost given to a particular mode $\Psi_{n\ell m}(t_0)$ as a result of being near the pole as follows. Taking $K = 0$, $a(t_0) = 1$, and identifying the pole location as

$$k_0^2 \equiv -\dot{H} - \frac{2}{3M_{\rm pl}^2}\langle(\nabla\delta\hat{\phi})^2\rangle, \tag{A.32}$$

we have

$$\begin{aligned}\Psi_{n\ell m}(t_0) = \frac{\bar{H}_0^2}{k_{n\ell}^2 - k_0^2}\frac{1}{2M_{\rm pl}^2\bar{H}_0^2} \\ \times\left[\ddot{\varphi}_0\,\delta\phi_{n\ell m}(t_0) - \dot{\varphi}_0\,\delta\dot{\phi}_{n\ell m}(t_0)\right],\end{aligned} \tag{A.33}$$

where factors of $\bar{H}_0$ (the relevant scale) have been inserted to make the coefficient dimensionless. We set a threshold such that any instantiation with a boost of $\bar{H}_0^2/(k_{n\ell}^2 - k_0^2) > 10$ for any $\ell = 0$ or $\ell = 1$ modes was too close to the pole for reliable results.

For such instantiations, we randomly changed $R = 1.5\pi/\bar{H}_0$ up or down by up to 10% (which changes the $k_{n\ell}(R)$ spectrum of the modes), and re-initialized all variables. The new position of the $k_{n\ell}$ values typically meant that no modes fell too close to the pole; if necessary, we repeated the process. We emphasize that these shifts to avoid the Ψ pole are only necessary for setting initial conditions at t_0. Dynamically, we evolve the modes $\Psi_{n\ell m}(t)$ according to Eq. (3.21), which is well behaved.

When constructing initial surfaces for $\Psi(t_0, \mathbf{x})$ (as in Fig. 3-1), this technique for avoiding the pole proved insufficient, as such surfaces required modes for $\ell = 0, 1, \ldots, \ell_{\max}$, where $\ell_{\max}$ was typically taken to be around 12. (Every mode $\Psi_{n\ell m}(t)$ evolves according to Eq. (3.21), though only modes $\delta\phi_{n\ell m}$ with $\ell = 0, 1$ contribute to the Hartree corrections.) The spectrum of all modes in such cases forms a rather dense forest, so shifting the spectrum (by adjusting R) simply shifts the pole from one mode to another. To ameliorate the pole in this situation, we multiplied $\bar{H}_0^2/(k_{n\ell}^2 - k_0^2)$ by a regulating function of $k_{n\ell}^2$ with a double zero at k_0^2 that takes a value of unity away from the pole. The transition width was chosen to be sufficiently narrow so that the complete factor never grew sufficiently large for one mode to dominate the plots. This is not a physical regularization, but serves to suppress modes near the pole for the purposes of visualization, and is used only to construct $\Psi(t_0, \mathbf{x})$.

In order to construct an initial spatial representation of $\Psi(t_0, r, \theta, \phi)$, we begin by sampling the initial conditions for $\delta\phi_{n\ell m}(t_0)$ and $\delta\dot{\phi}_{n\ell m}(t_0)$, and use these to construct $\Psi_{n\ell m}(t_0)$ (with regulator, if necessary). We then turn to the mode expansion for $\hat{\Psi}(x^\mu)$,

$$\hat{\Psi}(x^\mu) = \sum_{n,\ell,m} \left[\Psi_{n\ell m}(t)\, \hat{a}_{n\ell m}\, Z_{n\ell m}(r, \theta, \phi) + H.c.\right]. \tag{A.34}$$

This is a quantum-mechanical expansion (as evidenced by the $\hat{a}_{n\ell m}$ and $\hat{a}^\dagger_{n\ell m}$ operators). To construct a classical realization, we demote the operators to classical complex random variables $\tilde{a}_{n\ell m}$ and $\tilde{a}^*_{n\ell m}$ with the same statistics as the quantum operators.

For a given mode, the quantum operator is $\hat{\Psi}_{n\ell m} = \Psi_{n\ell m}(t)\, Z_{n\ell m}\, \hat{a}_{n\ell m} + H.c.$, so

we have $\langle\hat{\Psi}_{n\ell m}\rangle = 0$ and $\langle\hat{\Psi}^2_{n\ell m}\rangle = |Z_{n\ell m}|^2|\Psi_{n\ell m}(t)|^2$. (We only need to go to the two-point function as we have Gaussian statistics.) Demanding the same statistics for a classical function $\tilde{\Psi}_{n\ell m} = \Psi_{n\ell m}(t)\, Z_{n\ell m}\, \tilde{a}_{n\ell m} + c.c.$, we require

$$\langle\tilde{a}^2_{n\ell m}\rangle = \langle\tilde{a}^{*2}_{n\ell m}\rangle = 0, \qquad \langle\tilde{a}_{n\ell m}\tilde{a}^*_{n\ell m}\rangle = \frac{1}{2}. \tag{A.35}$$

This is achieved by constructing $\tilde{a}_{n\ell m} = x + iy$ for each mode, where x and y are independent Gaussian random variables with zero mean and standard deviation $\sigma_x = \sigma_y = 1/2$. Constructing the spatial slice is then straightforwardly accomplished by randomly sampling $\tilde{a}_{n\ell m}$, after which Eq. (A.34) may be evaluated. Identical techniques can be used to construct initial surfaces for $\delta\phi(t, \mathbf{x})$.

A.4 Numerical Convergence Tests

In this appendix, we discuss the accuracy of our simulations.

First, consider the accuracy of our numerical integration. The RK45 integrator uses a variable time step; early on in the integration, it takes small steps so as to capture the oscillations of the modes, and after the last mode freezes out, it takes increasingly large steps. We cap the maximum time step to have an estimated ΔN of 0.1, in terms of efolds $N = \ln a$. We employed relative and absolute tolerances of 10^{-10}. Investigations showed this to give a good trade-off between accuracy, numerical precision and computational time. Our measurements of the duration of inflation, $N_{\rm infl}$, for a given run are completely dominated by uncertainty as to when to start measuring $N_{\rm infl}$ rather than precision from the numerical evolution.

Second, our simulations cannot simulate the entire continuous universe, but only a discrete portion thereof. Our analysis essentially consists of three tuneable parameters describing this discretization: κ, $k_{\rm max}$ and R. The continuum limit takes each of these parameters to approach infinity, but such lies beyond our computational power. As such, we now discuss the convergence of our simulations as each of these parameters is increased. For each of these convergence tests, we look for convergence

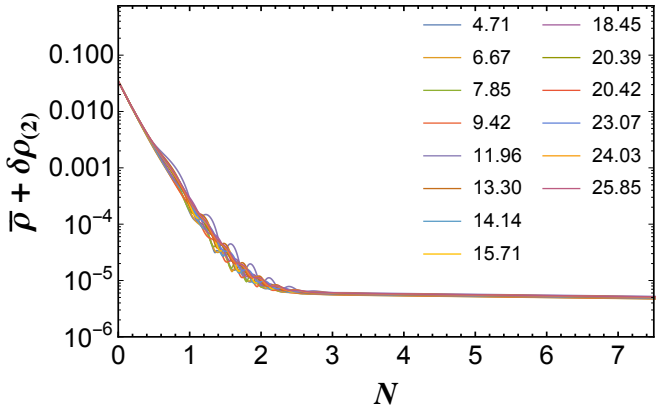

Figure A-1: The quantity $\rho_{\text{total}} = \bar{\rho} + \delta\rho_{(2)}$ versus N as we vary $R = c/\bar{H}_0$. We fix $(\varphi_0, \dot{\varphi}_0) = (25\, M_{\text{pl}}, -0.25\, M_{\text{pl}}^2)$ and initialize the field fluctuations in the Bunch-Davies state. Throughout the simulations described in our paper, we fixed $c = 1.5\pi \approx 4.71$. We found similar results as we increased c. In particular, for these initial conditions we found $N_{\text{total}} = 63.01 \pm 0.78$ as we varied $4.71 \leq c \leq 25.85$, indicating minimal dependence of the evolution of the dynamical system on our choice of R.

in the two-point functions and overall efolds for Bunch-Davies simulations, for which comparisons with analytic results are possible.

The outer boundary R controls both the spectrum of the modes and the weight for each mode's contribution to the two-point functions. As the two-point functions are computed by summing over k modes, taking $R \to \infty$ corresponds to taking the limit of the Riemann sum, yielding the continuous limit. Holding the upper limit k_{max} fixed, we expect spectral convergence with increasing R based on the properties of the basis functions we employ. Unfortunately, the presence of the Ψ pole (see Appendix A.3) means that as R increases, modes get pushed closer to the pole, eventually leading to divergent values for the two-point functions. Although our trick of shifting R to avoid the pole provides reasonable initializations (see Appendix A.3), we see consistency in our results as R increases, but not convergence. To properly observe convergence will require resolving the issue of the Ψ pole. As such, we need to find a balance between avoiding the Ψ pole and having sufficient modes both inside and outside the horizon at t_0.

Nonetheless, we can test for the dependence of various numerical quantities on our selection of R, keeping $\lambda = 10^{-10}$, $\kappa = 5\bar{H}_0$, and $k_{\text{max}} = 4\kappa$ fixed. When we initialize the fluctuations in the Bunch-Davies initial state and parameterize $R = c/\bar{H}_0$, we

find that quantities such as $\rho_{\rm total} = \bar{\rho} + \delta\rho_{(2)}$ show only modest variation as we vary c between the fiducial value we used throughout our simulations, $c = 1.5\pi = 4.71$, up through $c = 25.85$, as shown in Fig. A-1. For initial conditions $(\varphi_0, \dot{\varphi}_0) = (25\, M_{\rm pl}, -0.25\, M_{\rm pl}^2)$, we found $N_{\rm total} = 63.01 \pm 0.78$ as we varied $4.71 \leq c \leq 25.85$, indicating minimal dependence of the evolution of the dynamical system on our choice of R.

The regulator $\kappa = b\bar{H}_0$ is used to constrain the contribution to the two-point functions from small wavelengths. From theoretical analysis, we expect the two-point functions to diverge with increasing b, as shown in Section 3.3. The expected divergence is observed numerically. In order to demonstrate results that are immune to changing κ requires the implementation of a renormalization scheme, which is beyond the scope of this paper.

Finally, the wavenumber cutoff $k_{\rm max}$ represents the shortest length-scale that we allow to contribute to the two-point functions. So long as $k_{\rm max}$ is chosen to be sufficiently larger than κ, we observe the expected exponential convergence with increasing $k_{\rm max}$. We chose $k_{\rm max} = 4\kappa$, which corresponds to suppressing the initial contributions of the shortest wavelength modes to the various two-point functions by a factor of $\sim 3 \times 10^{-4}$.

Appendix B

Appendix for Chapter 5

B.1 Cross Sections and Decay Widths

The thermally averaged cross sections and decay widths are computed using the same conventions as in Ref. [120]. Table B.1 gives a list of the relevant cross sections and decay widths used throughout this paper.

B.2 Elastic Scattering Energy Transfer Rate

In this section, we outline the derivation of the elastic scattering energy transfer cross section $\langle \sigma v \delta E \rangle_{\chi f \to \chi f}$ that appears in the energy density Boltzmann equation, Eq. (5.16). For consistency with existing literature, we switch notations within this section so that a subscript $\chi, \overline{\chi}$ denotes quantities for both the DM particle and antiparticle, while a single subscript χ denotes a quantity associated only with the DM particle alone.

Following Ref. [94], the Boltzmann equation for the phase space distribution $f_{\chi,\overline{\chi}}(\vec{p}_1, t)$ of DM (both χ and $\overline{\chi}$) is

$$\frac{\partial f_{\chi,\overline{\chi}}}{\partial t} - H \frac{\vec{p}_1^{\,2}}{E_1} \frac{\partial f_{\chi,\overline{\chi}}}{\partial E_1} = \frac{C[f_\chi]}{E_1}\,, \tag{B.1}$$

where $C[f_\chi]$ is the collision term, and $E_1^2 = \vec{p}_1^{\,2} + m_\chi^2$. Here, we will focus on the elastic

Process	Cross Section or Decay Width
$\chi\overline{\chi}\chi \to A'\chi$	$\langle\sigma v^2\rangle = \frac{g_D^6(r-4)(r+4)(-32r^8+167r^6-534r^4+668r^2-512)}{36m_\chi^2(r^2-4)^4(r^2+2)^2}\frac{\sqrt{r^4-20r^2+64}}{96\pi m_\chi^3}$
$A'A' \to \chi\overline{\chi}$	$\langle\sigma v\rangle = \frac{32g_D^4(r^4-1)}{9r^4}\frac{\sqrt{r^2-1}}{8\pi m_\chi^2 r^3}$
$\chi\overline{\chi} \to e^+e^-$	$\langle\sigma v\rangle = \frac{4e^2\epsilon^2 g_D^2\left(2+m_e^2/m_\chi^2\right)}{(r^2-4)^2}\frac{\sqrt{1-m_e^2/m_\chi^2}}{8\pi m_\chi^2}$
$A' \to f\overline{f}$	$\Gamma = \frac{\epsilon^2 e^2}{12\pi}\left(1+\frac{2m_f^2}{m_{A'}^2}\right)\sqrt{m_{A'}^2-4m_f^2}$

Table B.1: List of cross sections and decay widths for the dark sector processes considered in this paper. All quantities are evaluated at the kinematic threshold.

scattering collision term, which includes $\chi f \to \chi f$, $\overline{\chi} f \to \overline{\chi} f$, $\chi\overline{f} \to \chi\overline{f}$ and $\overline{\chi}\overline{f} \to \overline{\chi}\overline{f}$ scatterings, where f is an SM fermion, which we take to be the electron throughout this paper for simplicity. From here on, $\chi f \to \chi f$ should be taken as shorthand for all four of these processes. Explicitly, taking the indices 1 and 3 for incoming and outgoing dark sector particles, and the indices 2 and 4 for incoming and outgoing SM particles, the collision operator is

$$\begin{aligned} C[f_\chi] = \frac{1}{2}\int \frac{d^3\vec{p}_2}{(2\pi)^3 2E_2}\int \frac{d^3\vec{p}_3}{(2\pi)^3 2E_3}\int \frac{d^3\vec{p}_4}{(2\pi)^3 2E_4} \\ \times (2\pi)^4\delta^4(p_1+p_2-p_3-p_4)\overline{|\mathcal{M}|^2} \\ \times \left[f_3 f_4(1\mp f_1)(1\mp f_2)\right. \\ \left. - f_1 f_2(1\mp f_3)(1\mp f_4)\right], \end{aligned} \tag{B.2}$$

with f_i denoting the phase space distribution of the particle indexed by i, We follow the conventions of Ref. [233], where the number density of a particle i is related to its phase space distribution via

$$n_i = \int \frac{d^3\vec{p}}{(2\pi)^3} f_i(\vec{p}), \tag{B.3}$$

with the number of degrees of freedom of particle i absorbed into the definition

of f_i. Furthermore, the matrix element squared is summed over final states but averaged over initial states, and is a sum of all four matrix elements squared for the four conjugate processes. The number of degrees of freedom for Dirac fermions χ is $g_\chi = 2$, and likewise for the SM fermion f; the degrees of freedom of particles and antiparticles are always counted separately. We note that all of these conventions are different from those used in Ref. [94], but the final results are equivalent.

Following Ref. [94] Eq. (B1), we can multiply Eq. (B.1) by E and integrate over all momenta to obtain

$$\dot{\rho}_{\chi,\overline{\chi}} + 3H(\rho_{\chi,\overline{\chi}} + P_{\chi,\overline{\chi}}) \simeq \int \frac{d^3\vec{p}_1}{(2\pi)^3 2E_1} \frac{\vec{p}_1^{\,2}}{m_\chi} C[f_\chi] \tag{B.4}$$

after taking the nonrelativistic approximation for χ, as found in Ref. [94] Eq. (B1).

Ref. [94] provides an expression assuming that f is relativistic; however, for $m_\chi \sim \mathcal{O}(\text{MeV})$, electrons are nonrelativistic before the dark sector decouples from the SM. We present here a compact expression for $C[f_\chi]$ that applies for electrons in all regimes. Following Ref. [234], we can write the collision term as

$$C[f_\chi] \simeq c(T) m_\chi^2 \left[m_\chi T \Delta_{\vec{p}_1} + \vec{p}_1 \cdot \nabla_{\vec{p}_1} + 3 \right] f_\chi \,, \tag{B.5}$$

where $\Delta_{\vec{p}}$ and $\nabla_{\vec{p}}$ are the Laplacian and del operators with respect to $\vec{p}$ respectively. Again, the only approximation made is that χ is nonrelativistic. The expression for $c(T)$ is given as [234]

$$c(T) = \frac{g_f}{12(2\pi)^3 m_\chi^4 T} \int dp_2 \, p_2^5 E_2^{-1} g^\pm \left(1 \mp g^\pm\right) \overline{|\mathcal{M}|^2}_{\substack{t=0 \\ s=s_0}} \,, \tag{B.6}$$

where $g^\pm \equiv [\exp(E_2/T) \pm 1]^{-1}$, taking the plus sign when f is a fermion,[1] and the matrix element squared is to be evaluated at $t = 0$ and $s = s_0 \equiv m_\chi^2 + 2m_\chi E_2 + m_f^2$. Ref. [233] found an analytic expression for $c(T)$ by making the relativistic approxi-

[1] We do not choose a sign for $g^\pm$ to be as general as possible, since the calculation follows equally easily for scattering off a boson.

mation for f, i.e. $E_2 \simeq p_2$, and writing

$$\overline{|\mathcal{M}|^2}_{\substack{t=0 \\ s=s_0}} = c_n \left(\frac{E_2}{m_\chi}\right)^n + \mathcal{O}\left[\left(\frac{E_2}{m_\chi}\right)^{n+1}\right] \tag{B.7}$$

and keeping only the leading order term. For $\chi f \to \chi f$ in our model, i.e. with χ and f both Dirac fermions mediated by A', this approximation is actually exact, with $n = 2$ being the only term in the expansion,

$$c_2 = \frac{16(4\pi)^2 \alpha_{\rm EM} \alpha_D \epsilon^2}{r^4} \qquad (\chi f \to \chi f)\,. \tag{B.8}$$

In fact, the integral in Eq. (B.6) can be performed analytically without making the relativistic approximation for the SM fermions; the result is

$$c(T) = \frac{g_f c_n T^{4+n}}{12(2\pi)^3 m_\chi^{4+n}} R_\pm(n+3, \xi)\,, \tag{B.9}$$

where $\xi = m_f/T$. The function $R_\pm$ is defined as

$$R_\pm(q,\xi) \equiv \big[(q+1)Q_\pm(q,\xi) - 2(q-1)\xi^2 Q_\pm(q-2,\xi) + (q-3)\xi^4 Q_\pm(q-4,\xi)\big]\,, \tag{B.10}$$

with

$$Q_\pm(q,\xi) \equiv \mp q! \sum_{s=0}^{q} \frac{\xi^s}{s!} \mathrm{Li}_{q-s+1}\left(\mp e^{-\xi}\right)\,, \tag{B.11}$$

where $\mathrm{Li}_m(z)$ is the polylogarithm [235]. We can check that in the relativistic limit for the fermion $\xi \to 0$, we get

$$\begin{aligned} R_+(q,0) &\equiv N_q^+ = (q+1)!(1-2^{-q})\zeta(q+1) \\ R_-(q,0) &\equiv N_q^- = (q+1)!\zeta(q+1)\,, \end{aligned} \tag{B.12}$$

where $\zeta(n)$ is the Riemann zeta function. This recovers the expression presented in

Eq. (B22) of Ref. [233] after substituting these expressions into Eq. (B.9).

In the nonrelativistic limit, $\xi \to \infty$, we obtain

$$Q_{\pm}(q, \xi \gg 1) \simeq \Gamma(q+1, \xi) \mp 2^{-(q+1)} \Gamma(q+1, 2\xi)\,, \tag{B.13}$$

where $\Gamma(n, y)$ is the incomplete Gamma function,

$$\Gamma(n, y) \equiv \int_y^\infty dt\, t^{n-1} e^{-t}\,. \tag{B.14}$$

With the expression for $c(T)$ in Eq. (B.9), we can substitute Eq. (B.5) into Eq. (B.4), and perform the momentum integral, assuming the dark matter follows a Maxwell-Boltzmann phase space distribution. The result is

$$\int \frac{d^3\vec{p}_1}{(2\pi)^3 2E_1} \frac{\vec{p}_1^2}{m_\chi} C[f_\chi] \simeq \frac{g_f c_n}{64\pi^3} m_\chi n_{\chi,\overline{\chi}} \times \left(\frac{T}{m_\chi}\right)^{4+n} R_{\pm}(n+3, \xi)(T - T')\,. \tag{B.15}$$

We can now define the right-hand side of Eq. (B.4) as $-n_{\chi,\overline{\chi}} n_{f,\overline{f}} \langle \sigma v \delta E \rangle_{\chi f \to \chi f}$, giving

$$\langle \sigma v \delta E \rangle_{\chi f \to \chi f} \simeq \frac{g_f c_n m_\chi}{64\pi^3 n_{f,\overline{f}}} \left(\frac{T}{m_\chi}\right)^{4+n} R_{\pm}(n+3, \xi)(T' - T)\,. \tag{B.16}$$

To our knowledge, this result is new, and is accurate for SM fermions in both the relativistic and nonrelativistic regimes, assuming nonrelativistic DM. In the limit where SM fermions are relativistic, we find for $\chi f \to \chi f$ mediated by A'

$$\langle \sigma v \delta E \rangle_{\chi f \to \chi f} \simeq \frac{8}{\pi} \frac{\alpha_{\rm EM} \alpha_D \epsilon^2}{r^4 x^6} 6!(1 - 2^{-5}) \zeta(6)(T' - T)\,. \tag{B.17}$$

Appendix C

Appendix for Chapter 6

C.1 Inelastic Dark Matter Models and Dark Higgs Self-Interaction

In this section, we provide two dark sector vector-portal models that lead to the inelastic DM model described in the main text, after dark U(1) symmetry breaking. The different symmetry breaking patterns in each model lead to different self-interaction strengths through a dark Higgs exchange, and a careful examination is required to ensure that self-interaction through an exchange of a dark Higgs can be neglected, as claimed in the main Letter.

C.1.1 Models

At high energies, the dark sector has a U(1) gauge symmetry with an associated gauge boson A', a single Dirac fermion Ψ charged under the U(1) gauge group and a Dirac mass m_D, as well as a dark Higgs field Φ, which also carries a dark U(1) charge. The dark sector A' kinetically mixes with the SM photon with kinetic mixing parameter ϵ. The important terms in the Lagrangian describing this model are

$$\mathcal{L} \supset -\frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} - \frac{\epsilon}{2\cos\theta_w}F'_{\mu\nu}B^{\mu\nu} + \overline{\Psi}\left(i\gamma^\mu(\partial_\mu + ig_D A'_\mu) - m_D\right)\Psi + |D_\mu\Phi|^2 + V(\Phi) + \mathcal{L}_{\Phi\Psi}\,. \tag{C.1}$$

Here, $B^{\mu\nu}$ is the $\mathrm{U}(1)_\mathrm{Y}$ hypercharge field tensor, which gives the gauge invariant kinetic mixing term at high energies, and the kinetic mixing is normalized so that after electroweak symmetry breaking, the mixing becomes $(\epsilon/2)F'_{\mu\nu}F^{\mu\nu}$ (θ_w is the weak mixing angle). $V(\Phi)$ is the potential in the Higgs sector, which we assume to be of a suitable form to break the U(1) symmetry prior to the start of freezeout, with the dark Higgs particle h_D having a mass m_{h_D} heavier than the other dark sector particles. The covariant derivative $D_\mu \equiv \partial_\mu + ig_D Q_\Phi A_\mu$ depends on the charge of Q_Φ, which we will choose appropriately below. Finally, the term $\mathcal{L}_{\Phi\Psi}$ contains interaction terms between Φ and Ψ.

With an appropriate choice of m_D, Q_Φ and $\mathcal{L}_{\Phi\Psi}$, the symmetry breaking of the U(1) symmetry also splits the Dirac fermion into two Majorana fermions with slightly different masses. We consider two example choices that have been studied in the literature:

1. Large m_D, $Q_\Phi = 1$ and $\mathcal{L}_{\Phi\Psi} = (1/2)(y_D/\Lambda)(\overline{\Psi^C}\Psi\Phi^*\Phi^* + \text{h.c.})$, a dimension-5 interaction term suppressed by the scale Λ, between Ψ and Φ [153];

2. Small m_D, $Q_\Phi = 2$ and $\mathcal{L}_{\Phi\Psi} = y_D(\overline{\Psi^C}\Psi\Phi^* + \text{h.c.})$, a Yukawa interaction with a charge-2 Φ-field is introduced instead [154].

In both cases, working in unitary gauge where after symmetry breaking, we have $\Phi \to (v_D + h_D)/\sqrt{2}$, we see that a Majorana mass term $(m_M/2)(\overline{\Psi^C}\Psi + \text{h.c.})$ is generated, with $m_M = y_D v_D^2/(2\Lambda)$ in Model (1), and $m_M = \sqrt{2}v_D y_D$ in Model (2). In Model (1), m_M is parametrically suppressed by a very large scale, and naturally gives a small Majorana mass compared to the Dirac mass, i.e., $m_M \ll m_D$: the dark matter particle is a pseudo-Dirac fermion. Conversely, in the second model, m_M can be as large as v_D, leading naturally to a large Majorana mass; in this case, we will choose $m_D \ll m_M$ instead.

In either model, we can now diagonalize the fermion mass matrix to obtain the mass eigenstates. Writing Ψ as a pair of Weyl fermions $(\xi, \eta^\dagger)$, the fermion mass

terms can be written as

$$\frac{1}{2}\begin{pmatrix} \xi & \eta^\dagger \end{pmatrix}\begin{pmatrix} m_M & m_D \\ m_D & m_M \end{pmatrix}\begin{pmatrix} \xi \\ \eta^\dagger \end{pmatrix} + \text{h.c.} \tag{C.2}$$

For Model (1), with $m_D \gg m_M$, the mass eigenstates are $\chi^* = (\eta + \xi)/\sqrt{2}$ with mass $m_M + m_D$, and $\chi = i(\eta - \xi)/\sqrt{2}$ with mass $m_\chi \equiv |m_M - m_D|$. The splitting between the two states is $m_\chi \delta = y_D v_D^2/\Lambda$. In Model (2), with $m_M \gg m_D$, the mass eigenstates are instead $\chi^* = (\eta + \xi)/\sqrt{2}$ with mass $m_M + m_D$, and $\chi = (\eta - \xi)/\sqrt{2}$ with mass $m_\chi \equiv m_M - m_D$. Here, the splitting between the two states is $m_\chi \delta = 2m_D$.

With a slight abuse of notation, we can define four-component Majorana spinors χ and χ^*, which are simply constructed from their respective, previously defined two-component versions. The coupling to the dark photon can then be simply written as

$$g_D A'_\mu \overline{\Psi}\gamma^\mu \Psi = g_D A'_\mu \overline{\chi^*}\gamma^\mu \chi \,, \tag{C.3}$$

giving the off-diagonal interaction in the part of the Lagrangian shown in the main Letter.

The interaction terms between the dark Higgs and the DM states depend on the details of the model. In Model (1), we have

$$\mathcal{L}_{\Phi\Psi} \supset \frac{y_D v_D}{2\Lambda}\left(\overline{\chi^*}\chi^* h_D - \overline{\chi}\chi h_D\right) + \frac{y_D}{4\Lambda}\left(\overline{\chi^*}\chi^* h_D^2 - \overline{\chi}\chi h_D^2\right) . \tag{C.4}$$

On the other hand, for Model (2), we obtain

$$\mathcal{L}_{\Phi\Psi} \supset \frac{y_D}{\sqrt{2}}\left(\overline{\chi^*}\chi^* h_D + \overline{\chi}\chi h_D\right) . \tag{C.5}$$

Finally, the vev of the dark Higgs breaks the U(1) symmetry and gives the dark photon a mass of $m_{A'} = Q_\Phi g_D v_D$. We can therefore re-express v_D in terms of the

phenomenologically important parameters for DM freezeout and detection. In both models, we can also express g_{h_D}, the coupling constant between χ, χ^* and h_D, in terms of these parameters.

In Model (1), m_χ is a free parameter, set approximately by the bare Dirac mass in the Lagrangian. The splitting on the other hand is given by $\delta m_\chi = y_D m_{A'}^2/(2g_D^2\Lambda)$, which can be suitably adjusted by choosing y_D/Λ appropriately. Once the splitting is chosen, however, g_{h_D} is fixed for constant g_D and r, since $g_{h_D} = y_D m_{A'}/(2g_D\Lambda) = g_D\delta/r$.

In Model (2), we find that $m_\chi \approx y_D m_{A'}/(\sqrt{2}g_D)$, while the splitting is a free parameter set by the bare Dirac mass. The expression for m_χ also sets $g_{h_D} = y_D \approx \sqrt{2}g_D/r$. Unlike Model (1), where the dark Higgs interactions originate from an irrelevant operator and are necessarily suppressed by a small parameter, there is no natural suppression on the coupling in Model (2).

C.1.2 Dark Higgs Self-Interaction

The difference in the coupling constant between χ, χ^* and h_D between these two models leads to slightly different expressions for the dark-Higgs-mediated self-interaction rate between χ particles. We will now derive the scattering cross section in each model, and discuss the conditions under which dark Higgs self-interactions can be neglected, as is done in the Letter.

In the limit of low velocity, the self-interaction scattering cross section via an exchange of a dark Higgs is

$$\sigma = \frac{g_{h_D}^4}{8\pi}\frac{m_\chi^2}{m_{h_D}^4}\,, \tag{C.6}$$

where m_{h_D} is the mass of h_D. For Model (1), we have $g_{h_D} = g_D\delta/r$, and so the scattering cross section takes the form:

$$\sigma_{(1)} = 2\pi\alpha_D^2\left(\frac{\delta}{r}\right)^4\frac{m_\chi^2}{m_{h_D}^4}\,. \tag{C.7}$$

We see that in this model the elastic scattering cross-section due to Higgs exchange is parametrically suppressed by $(\delta/r)^4$, as well as by $1/m_{h_D}^4$, since the mass splitting controls the size of the Yukawa coupling. Consequently we expect this rate to be very subdominant for $\delta \ll 10^{-1}$, even if m_{h_D} is not much larger than m_χ.

For Model (2), we have $g_{h_D} = \sqrt{2} g_D / r$, and we find instead

$$\sigma_{(2)} = 32\pi \frac{\alpha_D^2}{r^4} \frac{m_\chi^2}{m_{h_D}^4} \,. \tag{C.8}$$

We see that in this case the self-interaction cross-section is expected to be parametrically of order α_D^2/m_χ^2, unless $m_{h_D} \gg m_\chi$, and so may dominate over the 1-loop elastic scattering cross section, depending on α_D and m_{h_D}. For simplicity, we neglect all dark Higgs self-interaction diagrams, either by adopting Model (1), or by assuming that m_{h_D} is large enough to neglect such interactions in Model (2).

C.2 Additional Constraints on Dark Matter Energy Injection

We discuss two main sources of energy injection from our dark sector that may face additional constraints: excited state decays, and the one-loop annihilation of $\chi\chi \to \ell^+\ell^-$, where ℓ is either an electron or a muon. We will first discuss each process in turn, and then end the section by showing that these processes have limits that are subleading to those discussed in the main Letter, or are model dependent.

C.2.1 Excited State Decays

Due to the kinetic mixing between A' and Standard Model particles, the excited state can decay through the emission of an off-shell A' into various Standard Model states. Depending on the size of the splitting between ground and excited states, different channels are kinematically allowed, and result in important differences in the primordial abundance of excited states.

If $m_\chi \delta > 2m_e$, the excited state can decay into an electron/positron pair. The lifetime in the limit where $2m_e/m_\chi \ll \delta \ll 1$ is:

$$\tau_{\chi^* \to \chi e^+ e^-} \simeq \frac{15\pi m_{A'}^4}{4\epsilon^2 \alpha_{\rm EM} \alpha_D m_\chi^5 \delta^5} = 7.0 \times 10^{-5}\,{\rm s} \left(\frac{r}{1.6}\right)^4 \left(\frac{10^{-3}}{\epsilon}\right)^2 \times \left(\frac{1.0}{\alpha_D}\right) \left(\frac{1\,{\rm GeV}}{m_\chi}\right) \left(\frac{10^{-2}}{\delta}\right)^5 . \tag{C.9}$$

With this decay channel open, any primordial population of χ^* produced after the freezeout of $\chi^*\chi^* \to \chi\chi$ is depleted on the timescale of $\tau_{\chi^* \to \chi e^+ e^-}$, leading to a negligible primordial population of χ^*. The lack of any primordial χ^* removes the CMB constraint on $\chi^*\chi \to e^+e^-$ discussed in the main text, while the other constraints mentioned in the main text remain unchanged. However, in order to avoid constraints from energy injection during the BBN epoch, the decay lifetime must be sufficiently short (roughly less than 1 second), which can be achieved if δ varies with m_χ appropriately. While allowing $\chi^* \to \chi e^+ e^-$ is likely possible for this model, we choose $m_\chi \delta < 2m_e$ for the sake of simplicity, allowing us to fix a value of δ for all m_χ with ease.

For $m_\chi \delta < 2m_e$, the allowed decay processes are significantly more suppressed. The kinetic mixing term between A' and photons is achieved in a UV-complete manner through the following kinetic mixing term above the electroweak symmetry breaking scale (see e.g., Ref. [236]):

$$\mathcal{L} \supset -\frac{\epsilon}{2\cos\theta_w} F'_{\mu\nu} B^{\mu\nu} , \tag{C.10}$$

where $B_{\mu\nu}$ is the field strength tensor of the U(1) hypercharge gauge boson. After electroweak symmetry breaking, this term generates both the kinetic mixing term between A' and photons discussed in the main text and a mixing between A' and the Z-boson. This mixing allows the decay of the excited state into the ground state and two neutrinos, $\chi^* \to \chi\nu\overline{\nu}$. We calculated the decay width in the mass basis using the

results of Ref. [236], obtaining (see also Refs. [237, 238] for similar calculations)

$$\tau_{\chi^* \to \chi \nu \bar{\nu}} \simeq \frac{945\pi m_{A'}^4 m_Z^4 \cos^4\theta_w}{4\epsilon^2 \alpha_{\rm EM} \alpha_D \delta^9 m_\chi^9} = 1.8 \times 10^{27}\,{\rm s} \left(\frac{r}{1.6}\right)^4 \left(\frac{10^{-3}}{\epsilon}\right)^2 \times \left(\frac{1.0}{\alpha_D}\right) \left(\frac{100\,{\rm MeV}}{m_\chi}\right)^5 \left(\frac{10^{-3}}{\delta}\right)^9 . \tag{C.11}$$

Another possible decay channel is $\chi^* \to \chi + 3\gamma$, again through an off-shell photon A' and an electron loop. Following Ref. [237], we estimate this decay lifetime in the limit where $\delta m_\chi \lesssim m_e$ to be

$$\tau_{\chi^* \to \chi + 3\gamma} \sim \frac{2^7 3^6 5^3 \pi^3 m_e^8}{17 \alpha_{\rm EM}^3 \alpha_D \epsilon^2 \delta^9 m_\chi^9} \frac{m_{A'}^4}{\alpha_D \delta^4 m_\chi^4} = 1.1 \times 10^{18}\,{\rm s} \left(\frac{r}{1.6}\right)^4 \left(\frac{10^{-3}}{\epsilon}\right)^2 \times \left(\frac{1.0}{\alpha_D}\right)^2 \left(\frac{100\,{\rm MeV}}{m_\chi}\right)^9 \left(\frac{10^{-3}}{\delta}\right)^{13} . \tag{C.12}$$

Eqs. (C.11) and (C.12) show that the lifetime of χ^* is generically much longer than the age of the Universe for $\delta \leq 10^{-3}$, and so the primordially produced χ^* population is not significantly depleted by such processes.

C.2.2 One-Loop Ground-State Annihilation

The ground state χ can annihilate to a pair of fermions at the one-loop level. We have verified that in the limit of zero χ momentum and zero electron mass, the one-loop amplitude vanishes, indicating that the velocity-averaged annihilation cross section $\langle \sigma v \rangle_{\chi\chi \to f\bar{f}}$ is either p-wave suppressed or helicity suppressed (or possibly both, leading to an even larger suppression). We can therefore write $\langle \sigma v \rangle$ parametrically as

$$\langle \sigma v \rangle_{\chi\chi \to f\bar{f}} \sim \frac{g_D^4 \epsilon^4 e^4}{16\pi^2} \frac{h(m_f^2/m_\chi^2, v^2)}{8\pi m_\chi^2} , \tag{C.13}$$

where $h(m_f^2/m_\chi^2, v^2)$ is a suppression factor of either v^2 or m_f^2/m_χ^2. The first factor comprises the coupling constants in the loop diagram, as well as a loop factor. The factor of $(8\pi m_\chi^2)^{-1}$ is simply the phase space available to the final states [120].

C.2.3 Additional Constraints

The excited state decay $\chi^* \to \chi + 3\gamma$ leads to the injection of high-energy photons into the Universe during the cosmic dark ages; such processes are constrained by the CMB power spectrum, since they can increase the ionization fraction during the cosmic dark ages [239]. We can estimate the constraints on the χ^* lifetime from these results as

$$\tau_{\chi^* \to \chi + 3\gamma} > 10^{25}\,\mathrm{s} \left(\frac{n_{\chi^*}}{n_\chi} \right) . \tag{C.14}$$

In Fig. C-1, we show the region of parameter space ruled out by this requirement. Since the decay lifetime is extremely sensitive to δ, we can see that the constraints vary rapidly as δ changes over an order of magnitude. However, the constraints are mostly confined to large values of m_χ and ϵ, leaving the overall viability of this model unchanged even for $\delta \sim 10^{-3}$; there are no constraints from this process for $\delta \leq 10^{-4}$.

Note that Eq. (C.14) is not strictly correct: CMB power spectrum constraints weaken significantly once the decay lifetime is shorter than the age of the Universe at recombination. Limits from CMB spectral distortion [240–243] and Big Bang Nucleosynthesis [159] become relevant, but are much less sensitive to scenarios where only a small subcomponent of DM decays, releasing energy that is much lower than its rest mass. We neglect this possibility in this discussion for simplicity, since it would simply remove the constraints at large values of ϵ and m_χ, with no real impact on the overall viability of the model.

The other process that may present additional constraints is the one-loop annihilation $\chi\chi \to f\overline{f}$, with cross section given in Eq. (C.13), which may be constrained by the CMB power spectrum limits on DM annihilation. To check if such processes are important, we compare the one-loop rate with the $\chi\chi^* \to f\overline{f}$ annihilation process considered in the main text. This sets the following limit for when the one-loop

$\chi\chi \to f\overline{f}$ is important:

$$\alpha_D \alpha_{\rm EM} \epsilon^2 h(m_f^2/m_\chi^2, v^2) > \frac{n_{\chi^*}}{n_\chi}\,. \tag{C.15}$$

Across our entire parameter space, this requirement is never met regardless of the form of h, and thus $\chi\chi \to f\overline{f}$ can be safely ignored in favor of the other constraints examined in the main Letter.

C.3 One-Loop Self-Interaction Cross Section

Neglecting dark Higgs self-scattering, the self-interaction scattering process $\chi\chi \to \chi\chi$ occurs at lowest order at the one-loop level, since the A' couples off-diagonally to χ and χ^*. There are six Feynman diagrams contributing to the amplitude of this process; three of these diagrams are shown in Fig. C-2, while the remaining three are related to these diagrams by interchanging the fermionic final states. We follow Refs. [244, 245] and their treatment of Majorana fermions and their interactions.

We compute the one-loop self-interaction cross section σ in the limit of zero momentum for the particles; this makes the computation of the one-loop diagrams more tractable, and is also a reasonable assumption, since we expect finite momentum corrections on the order of $v \sim 10^{-3}$ in typical dark matter structures today. We also neglect the splitting between the ground and excited states for the excited states in the loop, which would produce order δ corrections. We use `FeynCalc` [246–248] for symbolic manipulation of our amplitudes, and `PackageX` [249, 250] for the loop integral itself. Ultimately, we find

$$\sigma = \frac{\overline{|\mathcal{M}|^2}}{128\pi m_\chi^2}\,, \tag{C.16}$$

where $\overline{|\mathcal{M}|^2}$ is the squared amplitude, averaged over initial spin states and summed over final spin states. The full expressions for $\overline{|\mathcal{M}|^2}$ in four different regimes of r are given in Eqs. (C.19)–(C.22), located at the end of this appendix. Fig. C-3 shows σ in

units of α_D^4/m_χ^2, as a function of r. Although the analytic expressions for $\overline{|\mathcal{M}|^2}$ have to be written in different forms for different r regimes, the function is actually smooth over all values of r. For comparison, we include the second Born approximation of σ at zero momentum computed using nonrelativistic quantum mechanics techniques in Ref. [160] in the limit where $r \ll 1$, which found a cross section of $\sigma = (\pi/2)r^{-6}$, neglecting δ and applying a correction factor to account for identical particles in the initial and final states (Ref. [160] implicitly assumed distinguishable particles). The one-loop computation performed here is equivalent to a second Born approximation of σ at zero momentum, and for $r \ll 1$, we find

$$\lim_{r\to 0}\sigma(r) \simeq \frac{\alpha_D^4}{m_\chi^2}\left[\frac{\pi}{2r^6} + \frac{7\pi}{8r^4} - \frac{\log(2r)+5/3}{r^3} + \frac{23\pi}{16r^2} + \mathcal{O}(r^{-1})\right], \tag{C.17}$$

in excellent agreement with the nonrelativistic quantum mechanical result; in the range of interest to this paper, $1 \lesssim r \lesssim 2$, we obtain a cross section that is 3 – 12 times larger than the $r \to 0$ limit would predict. For $r \gg 1$, we find

$$\lim_{r\to\infty}\sigma(r) \simeq \frac{\alpha_D^4}{m_\chi^2}\left[\frac{18}{\pi r^4} + \frac{118 - 96\log r}{\pi r^6} + \mathcal{O}(r^{-8})\right]. \tag{C.18}$$

The full expressions for $\overline{|\mathcal{M}|^2}$ are as follows:

1. $r < 1$:

$$\begin{aligned}
\frac{\overline{|\mathcal{M}|^2}}{256\pi^4\alpha_D^4} = &-\frac{\log^2\left(-r^2-2\sqrt{1-r^2}+2\right)\left(r^2-1\right)^3}{\pi^4\left(r^2-2\right)^4} \\
&+\left(\frac{p(r)}{\pi^4\left(r^2-2\right)^4}-\frac{4\sqrt{1-r^2}\left(r^2-1\right)}{\pi^4\left(r^2-2\right)^2}\right)\log^2(r) \\
&+\frac{16r^2\left(r^2-4\right)\left(r^2-2\right)^2\left(r^2-1\right)^2-\pi^2k(r)}{4\pi^4r^6\left(r^2-4\right)^3\left(r^2-2\right)^2} \\
&+\left[\frac{f(r)^2}{\pi^3r^6\left(r^2-4\right)^3\left(r^2-2\right)^4}+\left(\frac{4g(r)\sqrt{1-r^2}\sqrt{4-r^2}}{\pi^4r^3\left(r^2-4\right)^2\left(r^2-2\right)^4}-\frac{2f(r)\sqrt{4-r^2}}{\pi^4r^3\left(r^2-4\right)^2\left(r^2-2\right)^2}\right)\log(r)\right. \\
&\left.+\frac{4g(r)\sqrt{4-r^2}}{\pi^4r^5\left(r^2-4\right)^3\left(r^2-2\right)^2}\right]\arctan\left(\frac{r}{\sqrt{4-r^2}}\right) \\
&+\left[-\frac{4\left(r^2-1\right)}{\pi^4r^2\left(r^2-4\right)}+\left(\frac{8\left(r^2-1\right)^2}{\pi^4r^2\left(r^2-4\right)\left(r^2-2\right)^2}-\frac{2g(r)\sqrt{4-r^2}}{\pi^3r^3\left(r^2-4\right)^2\left(r^2-2\right)^4}\right)\sqrt{1-r^2}\right. \\
&\left.+\frac{f(r)\sqrt{4-r^2}}{\pi^3r^3\left(r^2-4\right)^2\left(r^2-2\right)^2}\right]\log(r) \\
&+\log\left(-r^2-2\sqrt{1-r^2}+2\right)\left[\left(\frac{4\left(r^2-1\right)^3}{\pi^4\left(r^2-2\right)^4}+\frac{2\sqrt{1-r^2}\left(r^2-1\right)}{\pi^4\left(r^2-2\right)^2}\right)\log(r)\right. \\
&+\left(\frac{g(r)\sqrt{4-r^2}}{\pi^3r^3\left(r^2-4\right)^2\left(r^2-2\right)^4}-\frac{4\left(r^2-1\right)^2}{\pi^4r^2\left(r^2-4\right)\left(r^2-2\right)^2}\right)\sqrt{1-r^2} \\
&\left.-\frac{2f(r)\sqrt{1-r^2}\sqrt{4-r^2}\left(r^2-1\right)\arctan\left(\frac{r}{\sqrt{4-r^2}}\right)}{\pi^4r^3\left(r^2-4\right)^2\left(r^2-2\right)^4}\right] \\
&-\frac{2g(r)\sqrt{4-r^2}}{\pi^3r^5\left(r^2-4\right)^3\left(r^2-2\right)^2}-\frac{f(r)^2\arctan^2\left(\frac{r}{\sqrt{4-r^2}}\right)}{\pi^4r^6\left(r^2-4\right)^3\left(r^2-2\right)^4}\,. \qquad \text{(C.19)}
\end{aligned}$$

2. $1 \le r < \sqrt{2}$:

$$\frac{\overline{|\mathcal{M}|^2}}{256\pi^4\alpha_D^4} = \arctan\left(\frac{r}{\sqrt{4-r^2}}\right)\left[\frac{f(r)^2}{\pi^3 r^6\left(r^2-4\right)^3\left(r^2-2\right)^4} - \frac{2f(r)\sqrt{4-r^2}\log(r)}{\pi^4 r^3\left(r^2-4\right)^2\left(r^2-2\right)^2}\right.$$
$$-\frac{2f(r)\sqrt{4-r^2}\left(r^2-1\right)^{3/2}\arctan\left(\frac{2\sqrt{r^2-1}}{2-r^2}\right)}{\pi^4 r^3\left(r^2-4\right)^2\left(r^2-2\right)^4} + \frac{4g(r)\sqrt{4-r^2}}{\pi^4 r^5\left(r^2-4\right)^3\left(r^2-2\right)^2}$$
$$\left.+\frac{2g(r)\sqrt{4-r^2}\sqrt{r^2-1}}{\pi^3 r^3\left(r^2-4\right)^2\left(r^2-2\right)^4}\right]$$
$$+\left(\frac{f(r)\sqrt{4-r^2}}{\pi^3 r^3\left(r^2-4\right)^2\left(r^2-2\right)^2} - \frac{2\left(r^2-1\right)^{3/2}}{\pi^3\left(r^2-2\right)^2} - \frac{4\left(r^2-1\right)}{\pi^4 r^2\left(r^2-4\right)}\right)\log(r)$$
$$-\frac{2g(r)\sqrt{4-r^2}}{\pi^3 r^5\left(r^2-4\right)^3\left(r^2-2\right)^2}+\sqrt{r^2-1}\left(\frac{4\left(r^2-1\right)^2}{\pi^3 r^2\left(r^2-4\right)\left(r^2-2\right)^2} - \frac{g(r)\sqrt{4-r^2}}{\pi^2 r^3\left(r^2-4\right)^2\left(r^2-2\right)^4}\right)$$
$$+\arctan\left(\frac{2\sqrt{r^2-1}}{2-r^2}\right)\left[\sqrt{r^2-1}\left(\frac{g(r)\sqrt{4-r^2}}{\pi^3 r^3\left(r^2-4\right)^2\left(r^2-2\right)^4} - \frac{4\left(r^2-1\right)^2}{\pi^4 r^2\left(r^2-4\right)\left(r^2-2\right)^2}\right)\right.$$
$$\left.-\frac{2\left(r^2-1\right)^3}{\pi^3\left(r^2-2\right)^4} + \frac{2\left(r^2-1\right)^{3/2}\log(r)}{\pi^4\left(r^2-2\right)^2}\right] + \frac{16r^2\left(r^2-4\right)\left(r^2-1\right)^2 - \frac{\pi^2 q(r)}{\left(r^2-2\right)^4}}{4\pi^4 r^6\left(r^2-4\right)^3}$$
$$-\frac{f(r)^2\arctan^2\left(\frac{r}{\sqrt{4-r^2}}\right)}{\pi^4 r^6\left(r^2-4\right)^3\left(r^2-2\right)^4} + \frac{\left(r^2-1\right)^3\arctan^2\left(\frac{2\sqrt{r^2-1}}{2-r^2}\right)}{\pi^4\left(r^2-2\right)^4} + \frac{\log^2(r)}{\pi^4}\,. \quad \text{(C.20)}$$

3. $\sqrt{2} \leq r < 2$:

$$\frac{\overline{|\mathcal{M}|^2}}{256\pi^4\alpha_D^4} = \arctan\left(\frac{r}{\sqrt{4-r^2}}\right)\left[\frac{f(r)^2}{\pi^3 r^6\left(r^2-4\right)^3\left(r^2-2\right)^4} - \frac{2f(r)\sqrt{4-r^2}\log(r)}{\pi^4 r^3\left(r^2-4\right)^2\left(r^2-2\right)^2}\right.$$
$$\left. - \frac{2f(r)\sqrt{4-r^2}\left(r^2-1\right)^{3/2}\arctan\left(\frac{2\sqrt{r^2-1}}{2-r^2}\right)}{\pi^4 r^3\left(r^2-4\right)^2\left(r^2-2\right)^4} + \frac{4g(r)\sqrt{4-r^2}}{\pi^4 r^5\left(r^2-4\right)^3\left(r^2-2\right)^2}\right]$$
$$+\left(\frac{f(r)\sqrt{4-r^2}}{\pi^3 r^3\left(r^2-4\right)^2\left(r^2-2\right)^2} - \frac{4\left(r^2-1\right)}{\pi^4 r^2\left(r^2-4\right)}\right)\log(r) - \frac{2g(r)\sqrt{4-r^2}}{\pi^3 r^5\left(r^2-4\right)^3\left(r^2-2\right)^2}$$
$$+\arctan\left(\frac{2\sqrt{r^2-1}}{2-r^2}\right)\left[\sqrt{r^2-1}\left(\frac{g(r)\sqrt{4-r^2}}{\pi^3 r^3\left(r^2-4\right)^2\left(r^2-2\right)^4} - \frac{4\left(r^2-1\right)^2}{\pi^4 r^2\left(r^2-4\right)\left(r^2-2\right)^2}\right)\right.$$
$$\left. + \frac{2\left(r^2-1\right)^{3/2}\log(r)}{\pi^4\left(r^2-2\right)^2}\right] + \frac{16r^2\left(r^2-4\right)\left(r^2-1\right)^2 - \frac{\pi^2 f(r)^2}{\left(r^2-2\right)^4}}{4\pi^4 r^6\left(r^2-4\right)^3}$$
$$- \frac{f(r)^2\arctan^2\left(\frac{r}{\sqrt{4-r^2}}\right)}{\pi^4 r^6\left(r^2-4\right)^3\left(r^2-2\right)^4} + \frac{\left(r^2-1\right)^3\arctan^2\left(\frac{2\sqrt{r^2-1}}{2-r^2}\right)}{\pi^4\left(r^2-2\right)^4} + \frac{\log^2(r)}{\pi^4}\,. \quad \text{(C.21)}$$

4. $r \geq 2$:

$$\frac{\overline{|\mathcal{M}|^2}}{256\pi^4\alpha_D^4} = \log(r)\left[\frac{2f(r)\log\left(\frac{2}{\sqrt{r^2-4}+r}\right)}{\pi^4 r^3\left(r^2-4\right)^{3/2}\left(r^2-2\right)^2} - \frac{4\left(r^2-1\right)}{\pi^4 r^2\left(r^2-4\right)}\right.$$
$$\left. - \frac{2\left(r^2-1\right)^{3/2}\arctan\left(\frac{2\sqrt{r^2-1}}{r^2-2}\right)}{\pi^4\left(r^2-2\right)^2}\right]$$
$$+\sqrt{r^2-1}\left[\frac{4\left(r^2-1\right)^2\arctan\left(\frac{2\sqrt{r^2-1}}{r^2-2}\right)}{\pi^4 r^2\left(r^2-4\right)\left(r^2-2\right)^2} - \frac{2g(r)\log\left(\frac{2}{\sqrt{r^2-4}+r}\right)\arctan\left(\frac{2\sqrt{r^2-1}}{r^2-2}\right)}{\pi^4 r^3\left(r^2-4\right)^{3/2}\left(r^2-2\right)^4}\right]$$
$$- \frac{4g(r)\log\left(\frac{2}{\sqrt{r^2-4}+r}\right)}{\pi^4 r^5\left(r^2-4\right)^{5/2}\left(r^2-2\right)^2} + \frac{(r^2-1)^3\arctan^2\left(\frac{2\sqrt{r^2-1}}{r^2-2}\right)}{\pi^4(r^2-2)^4}$$
$$+ \frac{4(r^2-1)^2}{\pi^4 r^4(r^2-4)^2} + \frac{f(r)^2\log^2\left(\frac{2}{\sqrt{r^2-4}+r}\right)}{\pi^4 r^6(r^2-4)^3(r^2-2)^4} + \frac{\log^2(r)}{\pi^4}\,. \quad \text{(C.22)}$$

where

$$
\begin{aligned}
f(r) &\equiv r^{10} - 10r^8 + 26r^6 - 16r^4 + 16r^2 - 32\,, \\
g(r) &\equiv r^{12} - 11r^{10} + 36r^8 - 42r^6 + 32r^4 - 48r^2 + 32\,, \\
k(r) &\equiv r^{16} - 12r^{14} + 40r^{12} + 4r^{10} - 96r^8 - 224r^6 + 256r^4 + 256\,, \\
p(r) &\equiv r^8 - 12r^6 + 36r^4 - 44r^2 + 20\,, \\
q(r) &\equiv r^{20} - 24r^{18} + 212r^{16} - 900r^{14} + 2008r^{12} \\
&\quad - 2608r^{10} + 2688r^8 - 2432r^6 + 1280r^4 - 1024r^2 + 1024\,. \qquad \text{(C.23)}
\end{aligned}
$$

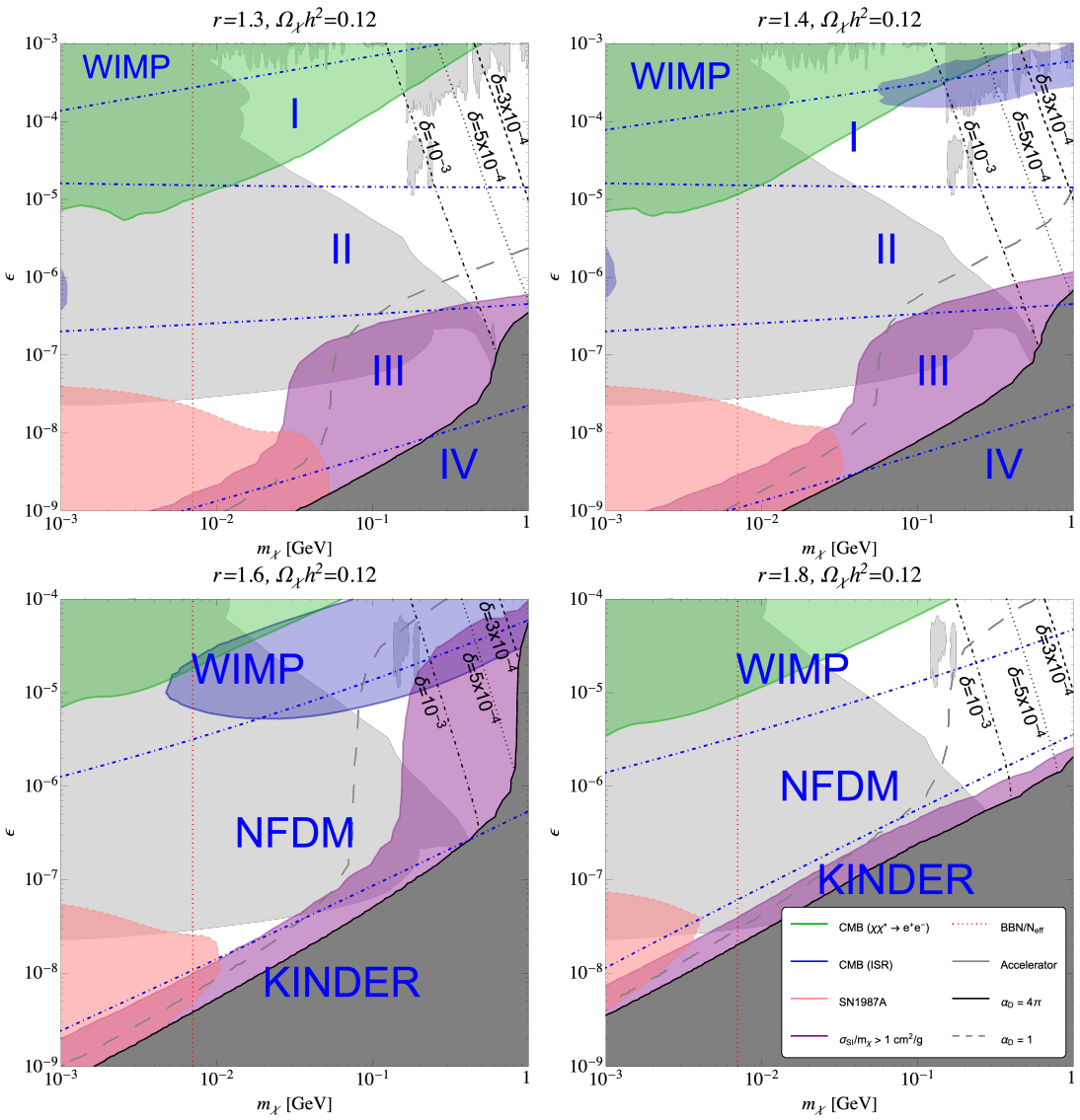

Figure C-1: Lines indicating CMB power spectrum constraints on $\chi^* \to \chi + 3\gamma$ for (black dashed) $\delta = 3 \times 10^{-4}$, (black dotted) $\delta = 5 \times 10^{-4}$ and (black dot-dashed) $\delta = 10^{-3}$ for (top left) $r = 1.3$, (top right) $r = 1.4$, (bottom left) $r = 1.6$ and (bottom right) $r = 1.8$ (regions above the lines are ruled out). The value of α_D that is needed to obtain a relic abundance of $\Omega_\chi h^2 = 0.12$ has been chosen for every point on the plot. Constraints on the parameter space from the cooling of SN1987a (red), $\chi\chi \to \chi\chi$ self-interaction (purple), CMB power spectrum constraints on $\chi\chi^* \to e^+e^-$ (green) and $\chi\chi \to A'^*A' \to A'e^+e^-$ (blue), as well as beam experiments (light gray) are shown. A limit on electromagnetically coupled light dark matter from BBN and CMB is shown by the dotted red line; masses below the line are ruled out (assuming no other dark-sector effects). Regions of the parameter space where nonperturbative values of $\alpha_D > 4\pi$ are needed for the right relic abundance is indicated in dark gray. The dashed gray line indicates $\alpha_D = 1$. Large labels corresponding to the various regimes discussed in Ref. [2] are shown for reference.

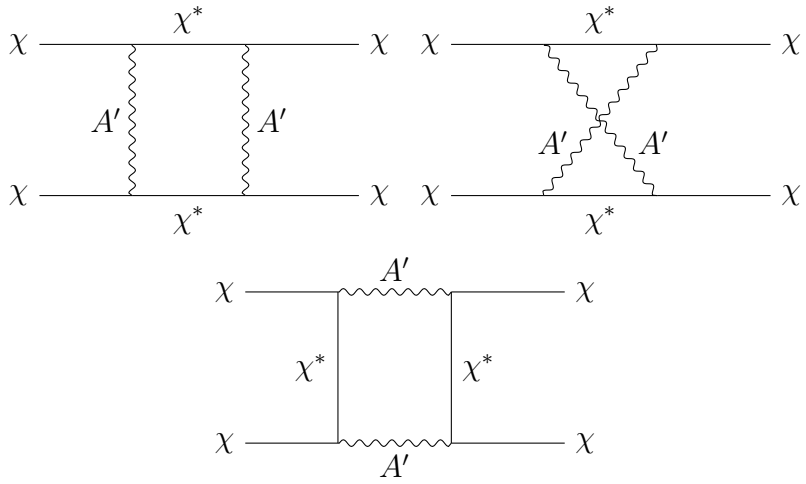

Figure C-2: One-loop Feynman diagrams for the self-interaction $\chi\chi \to \chi\chi$ scattering. There are three more diagrams related to these three by interchanging the final fermionic states, giving a total of six diagrams. We follow the conventions of Refs. [244, 245] and do not put arrows on Majorana fermion propagators and external states.

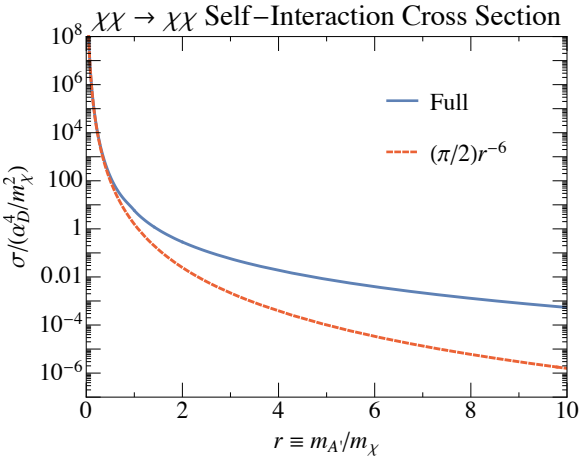

Figure C-3: Self-interaction cross section σ of $\chi\chi \to \chi\chi$ at zero momentum in units of α_D^4/m_χ^2 as a function of $r \equiv m_{A'}/m_\chi$ (blue). We also show the second Born approximation result at zero momentum computed assuming $r \ll 1$ in Ref. [160] for comparison (red, dashed), taking into account identical particles in the initial and final fermionic states.

Appendix D

Appendix for Chapter 8

D.1 Numerical method for computation of scattering states

Our initial-state wavefunctions are positive-energy solutions of the Schrödinger equation,

$$\begin{aligned} H^0_{L+S\,\mathrm{even}}\,\Psi &= E\,\Psi, \\ E &= \frac{M_\chi v^2}{4} \end{aligned} \tag{D.1}$$

with

$$V^0_{L+S\,\mathrm{even}}(r) = \begin{pmatrix} 0 & -\sqrt{2}\alpha_W \frac{e^{-m_W r}}{r} \\ -\sqrt{2}\alpha_W \frac{e^{-m_W r}}{r} & 2\delta M - \frac{\alpha}{r} - \alpha_W c_W^2 \frac{e^{-m_Z r}}{r} \end{pmatrix}. \tag{D.2}$$

Asymptotically, we are describing a state of two, free neutral WIMPs, $\chi^0\chi^0$, and thus know the energy eigenvalue in eq. D.1. The fact that our state contains two identical Majorana fermions fixes $L+S$ to be even in order to have a globally antisymmetric wavefunction. Since V is spherically symmetric, we can expand the general solution

in Legendre polynomials,

$$\Psi(\mathbf{r})_a \;=\; \sum_L \frac{[u_L(r)]_a}{r}\, A_{L\,a}\, P_L(\cos\theta). \tag{D.3}$$

The $A_{L\,a}$ will ultimately be fixed by normalization considerations, leaving the non-trivial task of determining the reduced wavefunctions, $u_L(r)$.[1] The behavior of $u_L(r)$ near the origin strongly deviates strongly from that of a plane wave. This leads to the well-known Sommerfeld enhancement in the direct annihilation of $\chi^0\chi^0 \to \gamma + X$ [191]. This nonperturbative effect can only be treated numerically and there is now a well developed literature on computing the wavefunction at the origin [196, 197]. Since annihilation of the incoming state proceeds via a highly-off-shell WIMP, to leading power in the velocity expansion, only $\Psi(0)$ is needed. As seen in section 8.3, the rate to capture to a bound state requires an overlap integral with the bound-state wavefunction (cf. eqs. 8.23 and 8.24). Since the bound states are spatially compact, their wavefunctions will decay exponentially past some number of Bohr radii, and as a practical matter, we only need the initial, scattering states out to this distance. Nonetheless, we are still responsible for determining the function $\Psi(\mathbf{r})$ (or $u_L(r)$) over a range of values. We cannot simply quantify the non-perturbative physics with a single number as in the annihilation problem.

We will find it useful to work with a dimensionless radial variable, $x(\equiv p\,r = M_\chi v_{\rm rel}\, r/2)$. Thus, we are solving the following reduced-wavefunction problem:

$$[u_L''(x)]_a + \left[\left(1 - \frac{L(L+1)}{x^2}\right)\delta^{ab} - \frac{V^{ab}(x)}{E}\right]\,[u_L''(x)]_b\,. \tag{D.4}$$

[1] Although the potential has spherical symmetry, our asymptotic solution contains an incoming plane wave. General scattering theory dictates that at large r, $\Psi(\mathbf{r})_a = e^{ik_a z} + f_a(\theta)e^{ik_a r}/r$. Thus, the solution still possesses cylindrical symmetry, justifying the independence of the general form, eq. D.3, on the azimuthal angle, ϕ.

For completeness, the rescaled potential term is

$$\frac{V(x)}{E} = \begin{pmatrix} 0 & -2\sqrt{2}\alpha_W \frac{e^{-2m_W x/(M_\chi v_{\rm rel})}}{x\, v_{\rm rel}} \\ -2\sqrt{2}\alpha_W \frac{e^{-2m_W x/(M_\chi v_{\rm rel})}}{x\, v_{\rm rel}} & \frac{8\delta M}{M_\chi v_{\rm rel}^2} - \frac{2\alpha}{x\, v_{\rm rel}} - 2\alpha_W c_W^2 \frac{e^{-2m_Z x/(M_\chi v_{\rm rel})}}{x\, v_{\rm rel}} \end{pmatrix}. \quad \text{(D.5)}$$

Although $V(x)$ creates significant distortion at small x, at large x we recover a free theory, but with the charged component exponentially decaying due to the mass-shift term, $\frac{8\delta M}{M_\chi v_{\rm rel}^2}$. It is straightforward to write the asymptotic solutions to eq. D.4 in terms of the normalized wavenumbers, $\hat{k}_a = k_a/p$, with

$$\begin{aligned} \hat{k}_1 &= 1 \\ \hat{k}_2 &= i\sqrt{\frac{8\delta M}{M_\chi v_{\rm rel}^2} - 1}. \end{aligned} \quad \text{(D.6)}$$

At the level of pure math, at large x we get solutions that are linear combinations of $e^{\pm i\,\hat{k}_i\, x}$. For the neutral component, we just need to determine the appropriate phase in the sinusoid. The charged component, however, contains an intrinsic instability for numerical evaluation. On physical grounds, by the (plane-wave) normalizability of the wavefunction, at long distances we will only get an exponentially-decaying term.[2] However, in practice it is intractable to set the boundary conditions for this second-order equation precisely enough to obtain a numerical solution that is purely-decaying. Tiny errors will generate an exponentially-growing term that eventually dominates and spoils the wavefunction. For this reason, in the current implementation of Mathematica's `NDSolve`, we can only obtain a solution to the direct Schrödinger equation (D.4) for $M_\chi \lesssim 100$ TeV. We will thus employ a numerical technique developed in the nuclear community [251] known as the Variable Phase Method. Much of the setup

[2]For a sufficiently massive wino, $M_\chi > 1320$ TeV for $v = 10^{-3}$, we get a real-valued $\hat{k}_2$. Thus, the charged component can also propagate to spatial infinity. This marks a qualitative change to the problem we are considering, as we must now consider the full SU(2) triplet throughout the calculation. This is an interesting regime, and one much closer to the Coulomb-limit cases we have discussed. Its full treatment is nonetheless beyond our scope.

follows [197], who developed this procedure in a DM context in order to compute the value of the wavefunction at the origin. We will extend their approach to find the wavefunction in a range out to several hundred Bohr radii. To our knowledge, the details of our method are novel and provide an efficient, powerful means to calculate positive-energy, multi-component DM wavefunctions.

The Variable Phase Method treats the reduced wavefunction as a spatially-varying linear combination of solutions to the free Schrödinger equation with the appropriate normalized wavenumber,

$$\left[u_L''(x)\right]_a + \left[\hat{k}_a^2 - \frac{L(L+1)}{x^2}\right]\left[u_L(x)\right]_a \;=\; 0. \tag{D.7}$$

Following [197], we take the following free reduced wavefunctions,

$$\begin{aligned} f_a(x) &= \sqrt{\frac{\pi x}{2}} J_{L+\frac{1}{2}}(\hat{k}_a x) \\ g_a(x) &= -\sqrt{\frac{\pi x}{2}}\left[Y_{L+\frac{1}{2}}(\hat{k}_a x) \;-\; i\, J_{L+\frac{1}{2}}(\hat{k}_a x)\right], \end{aligned} \tag{D.8}$$

which have the appropriately-normalized Wronskian, $f'\,g - f\,g' = 1$. We can now write the solution to the full problem, eq. D.4, as

$$\left[u_L(x)\right]_a \;=\; f_a(x)\,\alpha_a(x) \;-\; g_a(x)\,\beta_a(x). \tag{D.9}$$

Needing $\alpha_a(x)$ and $\beta_a(x)$ doubles the degrees of freedom in the solution, so we eliminate this redundancy by imposing the normalization

$$f_a(x)\alpha_a'(x) \;-\; g_a(x)\beta_a'(x) = 0. \tag{D.10}$$

Instead of using the parameterization of the wavefunction in eq. D.9, we define

$$N_{ab} \;=\; f_a\, g_a \delta_{ab} - g_a O_{ab}\, g_b, \tag{D.11}$$

where $\beta_a = O_{ab}\,\alpha_b$, and

$$\tilde{\alpha}_a = \frac{\alpha_a}{g_a}. \tag{D.12}$$

These combine so that

$$u_a(x) = N_{ab}\,\tilde{\alpha}_b. \tag{D.13}$$

The advantage of writing the reduced wavefunction in this way is that we can now determine it from a pair of nested, first-order equations,

$$\begin{aligned} N'_{ab} &= \delta_{ab} + \left(\frac{g'_a}{g_a} + \frac{g'_n}{g_b}\right) N_{ab} - N_{ac}\frac{\hat{V}_{cd}}{E} N_{db} \\ \tilde{\alpha}' &= \left(-\frac{g'_a}{g_a}\delta_{ab} + \frac{\hat{V}_{ac}}{E} N_{cb}\right) \tilde{\alpha}_b. \end{aligned} \tag{D.14}$$

We get $\hat{V}(x)$ from the full potential, eq. D.5 by only keeping those terms which vanish as $x \to \infty$, i.e. $V(x) = V_{\text{inf.}} + \hat{V}(x)$, and $\lim_{x\to\infty} V(x) = V_{\text{inf.}}$. Furthermore, we know that in the limit of very small x, our problem just becomes that of a Coulomb potential. Thus, near the origin $u_a(x) \propto x^{L+1}$. However, we do not know the appropriate prefactor. In the $L = 0$ case, this is just the Sommerfeld factor, or wavefunction at the origin, for which many numerical determinations exist. We could employ such a method, and then take an appropriate linear combination of solutions determined with independent sets of boundary conditions to determine the full wavefunction. However, we have found a way to impose a boundary condition on eq. D.14 that requires no additional inputs and finds the correct wavefunction. Following the method of [197] for finding the Sommerfeld factor, for some small x (which we generally take as $x_0 = 10^{-6}$), we demand

$$N_{ab}(x_0) = \frac{x_0}{2L+1}\delta_{ab}. \tag{D.15}$$

We found that in practice, imposing $\tilde{\alpha}(x)$ boundary conditions also at x_0 led to numerical instability beyond $\mathcal{O}(10\text{s})$ of Bohr radii, which was insufficient for the overlap integral needed to compute the capture rate. Just as in other numerical routines, the exponentially growing solution of the charged component eventually overwhelms the solution.

Instead, we impose the following boundary condition at relatively large x_f,[3]

$$\begin{aligned} \tilde{\alpha}_1(x_f) &= 1, \\ \tilde{\alpha}_2(x_f) &= 0. \end{aligned} \tag{D.16}$$

Fixing the charged component to vanish at large x eliminates the exponentially growing solution from our regime of interest. In fact, we imposed a similar boundary condition on the charged component when solving for the full $u(r)$ directly with `NDSolve`. The advantage of the Variable Phase Method over the direct approach is that the equations are first-order. Thus, `Mathematica` is not solving a Boundary Value Problem, shooting solutions from one boundary to the other and attempting to line them up. A key signal of breakdown in the direct method was the inability to satisfy the boundary conditions at both ends. This issue simply does not arise with $\tilde{\alpha}(x)$.

The further $\tilde{\alpha}_1(x_f) = 1$ condition, along with those imposed on N_{ab} (*cf.* eq. D.15), is actually sufficient to determine the physical solution. To see this it is useful to rewrite eq. D.14 in the limits of both small and large x, given the boundary conditions for N_{ab} in eq. D.15. These decouple $N_{12(21)}$ at small x and fix $N_{11(22)}(x_0) \sim x_0$ in the region of some arbitrarily small x_0. Thus, we get

$$\begin{aligned} N'_{aa} &= 1 + 2\,\frac{g'_a}{g_a} N_{aa} \\ \tilde{\alpha}'_a &= -\frac{g'_a}{g_a}\tilde{\alpha}_i, \end{aligned} \tag{D.17}$$

[3]We found minimal sensitivity to the exact location of x_f. Typical values taken were $x \in (50, 100)$.

where there is no sum over a. For our choice of g_a in eq. D.8,

$$\lim_{x\to 0} \frac{g_a'}{g_a} = \begin{cases} i\,\hat{k}_a & L = 0 \\ \\ -\frac{L}{x} & L \neq 0 \end{cases}. \tag{D.18}$$

The functional dependence is different for $L = 0$, but the scaling for N_{aa} and $\tilde{\alpha}_a$ will be the same as for higher L. We get for small x,

$$\begin{aligned} N_{aa} &\propto x \\ \tilde{\alpha}_a &\propto x^L. \end{aligned} \tag{D.19}$$

This sets $u(x) \propto x^{L+1}$, which is exactly the scaling of the regular solution to the Schrödinger equation with a Coulomb potential. Thus, we have obtained the correct physical solution up to an overall scaling factor. However, we know that at sufficiently large x, the potential is negligible, and thus the charged component will be exponentially suppressed, and the neutral component will just be a unit amplitude sinusoid, up to known factors that depend only on L and the number of identical particles in the initial state (see eq. 8.27):

$$u_1(x) = \sin\left(x - \frac{\pi\, L}{2} + \delta_L(x)\right), \tag{D.20}$$

with $\delta_L \ll x$. Setting $\tilde{\alpha}_2(x_f) = 0$ is a good approximation to the exponential suppression for sufficiently large x_f.[4] We will now show how $\tilde{\alpha}_1(x_f) = 1$ reproduces the normalization of eq. D.20. To start, we get the N, $\tilde{\alpha}$ equations in the large-x limit. This behavior is determined by the log derivative of g_a,

$$\lim_{x\to\infty} \frac{g_a'}{g_a} = i\,\hat{k}_a. \tag{D.21}$$

[4]Our method trivially generalizes to the case where u_2 is non-decaying at large x. Using the normalization conditions spelled out in [160], we can still impose that the radial component of u_2 is purely outgoing. Operationally, this is realized with the same boundary conditions as in eqs. D.15 and D.16.

This result is L-independent, because for sufficiently large x, the centrifugal term drops out of eq. D.7. We therefore get the following equations for N_{ab},

$$\begin{aligned} N'_{11} &= 1 + 2\,i\,N_{11} \\ N'_{12} &= -\left(\sqrt{\frac{8\delta M}{M_\chi v_{\rm rel}^2} - 1} - i\right) N_{12} \\ N'_{22} &= 1 - \left(2\sqrt{\frac{8\delta M}{M_\chi v_{\rm rel}^2} - 1}\right) N_{22}, \end{aligned} \tag{D.22}$$

and for our potential and N_{ab} boundary conditions, $N_{21} = N_{12}$. These have the following solutions,

$$\begin{aligned} N_{11} &= \frac{i}{2} + A_{11}\,e^{2i\,x} \\ N_{12} &= A_{12}\exp\left[\left(-\sqrt{\frac{8\delta M}{M_\chi v_{\rm rel}^2} - 1} + i\right) x\right] \\ N_{22} &= \frac{1}{2\sqrt{\frac{8\delta M}{M_\chi v_{\rm rel}^2} - 1}} + A_{22}\exp\left[-\left(\sqrt{\frac{8\delta M}{M_\chi v_{\rm rel}^2} - 1}\right) x\right]. \end{aligned} \tag{D.23}$$

We see that A_{12} and A_{22} decay exponentially, the latter to a predetermined, nonzero value. Since we are further setting $\tilde{\alpha}_2(x_f) = 0$, it is just the solution to N_{11} we need to examine in detail to determine the proper normalization of $u_1(x)$. The N_{ab} satisfy a first-order equation. Therefore, having imposed the boundary condition at small x, the value of A_{11} is predetermined, but its value is unknown based on the equation's behavior at large x. To constrain it, we now examine the behavior of $\tilde{\alpha}_1$ at large x. It obeys the trivial equation,

$$\begin{aligned} \tilde{\alpha}'_1 &= -i\,\tilde{\alpha}_1. \quad \text{Thus,} \\ \tilde{\alpha}_1 &= C\,e^{-i\,x}, \end{aligned} \tag{D.24}$$

where C is the free parameter we can set with the boundary condition at large x.

Combining N and $\tilde{\alpha}$ to get $u_1(x)$, we find

$$\begin{aligned} u_1(x) &= C\left[A_{11}\, e^{i\,x} - \frac{e^{-i\,x}}{2i}\right] \\ &= B\,\sin\left(x - \frac{\pi\, L}{2} + \delta_L(x)\right), \end{aligned} \tag{D.25}$$

where on the second line we have used the fact that we know our small-x behavior has given us the physical solution up to an overall scaling factor. We see that this fact fixes the value of $|A_{11}| = 1/2$, since $|\alpha e^{i\,x} + \beta e^{-i\,x}|$ varies between 0 and 1, as needed for sinusoidal behavior for $|\alpha| = |\beta| = 1/2$. Thus, having a properly normalized wavefunction at large x requires $|B| = 1$ and therefore fixes $|C| = 1$. Since our boundary condition sets $\tilde{\alpha}_1(x_f) = 1$, it means $C = e^{i\,x_f}$, thus obtaining the right asymptotic behavior. As a cross-check, we verified that the wavefunction-at-the-origin (Sommerfeld factor) that we obtained via this method agreed with those factors obtained via the Variable Method employing $\tilde{\alpha}^{-1}$ used in [197], and the Wronskian method of [196], wherever the latter two methods were numerically stable. An advantage of our Variable Phase Method is that it gives numerical stability as high as 1000 TeV (and possibly farther), whereas we found the others became unstable around 300 TeV ([197] Variable Phase) and 100 TeV (Wronskian). We also verified that the capture rate we find with our Variable Phase Method agrees very well with that obtained from calculating the reduced wavefunction $u_a(x)$ directly with `NDSolve`.[5]

D.2 Numerical method for computation of bound states

The bound states wavefunctions solve a Schrödinger equation with the appropriate potential depending on the $L + S$ of the DM pair. Generically, the neutralino ($\chi^0\chi^0$)

[5]For S and P-wave initial states, we find the direct `NDSolve` method is stable up to 110 TeV. Our procedure sets $u_{1,2}(x_0) = 0$ near the origin, and $u_2(x_f) \approx 0$ and $u_1(x_f) = 1$, for large x_f. In the numerically stable region, we find insensitivity to setting u_2 to a range exponentially small numbers. Unlike the Variable Phase Method, setting $u_1(x_f) = 1$ is insufficient to get an appropriately normalized solution. However, we can normalize the wavefunction after the fact to have a unit amplitude sinusoid at large x.

and chargino ($\chi^+\chi^-$) pairs can turn into one another by exchange of $W^\pm$ bosons. However, since the neutralinos are identical fermions, they require a globally antisymmetric wavefunction and therefore $L+S$-even. Thus, $V_{L+S\,\text{even}}$ is a matrix potential with off-diagonal entries between these pairs, while $V_{L+S\,\text{odd}}$ only evolves $\chi^+\chi^-$. As in the main text, we put the neutralino and the chargino components of the bound state's wavefunctions into a vector $\Psi^\top = (\psi_N\, \psi_C)$. The reduced radial wavefunction, $u_i(r) = r\,\psi_i(r)$, solves the Schrödinger equation

$$-\frac{1}{M_\chi}\partial_r^2 \begin{pmatrix} u_N(r) \\ \\ u_C(r) \end{pmatrix} + \begin{pmatrix} \frac{L(L+1)}{M_\chi r^2} + V_{11} & V_{12} \\ \\ V_{21} & \frac{L(L+1)}{M_\chi r^2} + 2\delta M + V_{22} \end{pmatrix} \begin{pmatrix} u_N(r) \\ \\ u_C(r) \end{pmatrix} = \frac{M_\chi v_{\text{rel}}^2}{4} \begin{pmatrix} u_N(r) \\ \\ u_C(r) \end{pmatrix}, \tag{D.26}$$

where the energy of the state is defined as $E = M_\chi v_{\text{rel}}^2/4$.

To simplify our analysis, we use the variable $z \equiv M_\chi \alpha_W\, r$, which normalizes the physical distance r to half the Bohr radius of WIMPonium in the high-mass limit. Solving the radial Schrödinger equation for the reduced wavefunction $u(z) = z\psi(z)$ with $V_{L+S-\text{even}}$, we have

$$\begin{pmatrix} -\partial_z^2 + \frac{L(L+1)}{z^2} & -\sqrt{2}\frac{e^{-\epsilon_\phi z}}{z} \\ \\ -\sqrt{2}\frac{e^{-\epsilon_\phi z}}{z} & -\partial_z^2 + \epsilon_\delta^2 + \frac{L(L+1)}{z^2} - \frac{c_W^2 e^{-\frac{\epsilon_\phi}{c_W} z}}{z} - \frac{s_W^2}{z} \end{pmatrix} \begin{pmatrix} u_N(z) \\ \\ u_C(z) \end{pmatrix} = \epsilon_v^2 \begin{pmatrix} u_N(z) \\ \\ u_C(z) \end{pmatrix}, \tag{D.27}$$

where c_W (s_W) stands for cosine (sine) of Weinberg angle, $\epsilon_v = v_{\text{rel}}/2\alpha_W$, $\epsilon_\phi = m_W/(\alpha_W M_\chi)$, and $\epsilon_\delta = \sqrt{2\delta M/M_\chi}/\alpha_W$ with $\delta M = M_{\chi^\pm} - M_\chi^0$.

If $L+S$ is odd, our bound state only has a chargino component, but we keep the

matrix notation, giving

$$\begin{pmatrix} 0 & 0 \\ 0 & -\partial_z^2 + \epsilon_\delta^2 + \frac{L(L+1)}{z^2} - \frac{c_W^2 e^{-\frac{\epsilon_\phi}{c_W} z}}{z} - \frac{s_W^2}{z} \end{pmatrix} \begin{pmatrix} 0 \\ u_C(z) \end{pmatrix} = \epsilon_v^2 \begin{pmatrix} 0 \\ u_C(z) \end{pmatrix}. \tag{D.28}$$

A detailed derivation of these potentials can be found in [191, 216]. In both eqs. D.27 and D.28 the 22 component has a Coulomb term corresponding to the photon. Thus, the $\chi^+\chi^-$ will form bound states at any value of $M_{\chi^\pm}$, in analogy to positronium; in the limit $M_{\chi^\pm} \to 0$, the charged component will dominate the $L+S$-even wavefunction, as the off-diagonal mixing terms become exponentially suppressed and the neutralinos decouple from the attractive potential.

We are dealing with a mixture of Coulomb and Yukawa potentials. In both the high and low-mass limits, the non-zero components of the potential will become purely Coulombic with coupling α (low-mass) and $2\alpha_W$ (α_W) for the high-mass $L+S$-even (odd) potentials. Using this insight, we expand our wavefunctions for a given L in a basis of solutions to the Coulomb Schrödinger problem with the same L, exploiting the fact that we still have rotational symmetry in the full problem. This will turn the Schrödinger equation into a finite-dimensional, matrix eigenvalue problem. We denote the Coulombic wave functions in the basis by $|n,l\rangle$ and the eigenfunctions of the full problem by $|\phi_l\rangle$. The Schrödinger equations thus becomes

$$\hat{H}|\phi_l\rangle \equiv \sum_{n=1}^{N} c_{n,l}\, \hat{H}|n,l\rangle, \tag{D.29}$$

where the c coefficients are used to write the eigenfunctions of the Hamiltonian in terms of N Coulomb wavefunctions and $\hat{H}$ is the Hamiltonian. Inserting a unit operator in the space of wavefunctions with fixed L, we have the following eigenvalue problem

$$\hat{H}|\phi_l\rangle = \sum_n \langle m,l|\hat{H}|n,l\rangle c_{n,l}|m,l\rangle = E_a\, c_{m,l}|m,l\rangle, \tag{D.30}$$

where the a in E_a refers to the radial quantum number. Hence, the Schrödinger equation reduces to the matrix equation,

$$H_{mn}c_{n,l} = E_a c_{m,l}. \tag{D.31}$$

where $H_{mn} = \langle m, l|\hat{H}|n, l\rangle$. We thus have to find the overlap integrals H_{mn} for the Hamiltonians corresponding to Eqs. D.27 and D.28.

For odd $L + S$ configurations (eq. D.28) we use a basis of functions solving the Coulomb potential $-\alpha_W/z$, corresponding to the high-mass limit of $V_{L+S\mathrm{odd}}$. For even $L + S$ configurations (eq. D.27), the potential matrix is off-diagonal; however, in the $M_\chi \gg m_W$ limit we can perform an r-independent diagonalization of the potential to obtain a repulsive Coulomb potential with $+\alpha_W/z$ (which does not contribute to the formation of bound states) and an attractive one with $-2\alpha_W/z$, as discussed in section 8.2.2. As a result, to find the eigenfunctions for eq. D.27 we choose a basis of functions solving the Coulomb potential, $-2\alpha_W/z$. We noted empirically that in order for this method to work satisfactorily, the coupling of the Coulomb potential used for the expansion in eq. D.30 needs to be as large or larger than that of the full problem. The choices above satisfy this criterion as it is only in the $M_\chi \to \infty$ limit that the potential has full α_W or $2\alpha_W$ strength. At any finite mass, the Yukawa terms lead to a weaker attractive force. It makes sense that the expansion coupling needs to be larger than the true potential. The true wavefunction will have support to a particular Bohr radius. If every term in the Coulomb expansion is wider, it is difficult to recover the correct behavior. Starting with Coulomb wavefunctions that are more compact than that of the full problem, we can build up the approximate wavefunction by including more and more radially-excited states. We find stability across our mass regime of interest when the number of Coulomb eigenfunctions in the expansion of eq. D.30 satisfies $N \gtrsim 15$, and we typically take N=30 for our capture rate plots.

We have performed several consistency checks of our numerical routine. We recover the Coulomb limits at high and low masses with couplings α_W, α, respectively

for $L+S$ odd, and $2\alpha_W$ in the high-mass $L+S$ even limit. We also recover some known results from the literature on the numerical study of solutions of Yukawa potentials [252, 253]. The results therein suggest that for finite mediator mass, as we increase the ratio ϵ_ϕ, the bound states cease to exist. Taking the limit $\delta M \to 0$ and $\mathrm{s}_W^2 \to 0$, which corresponds to a pure-Yukawa problem with either a 2D (eq. D.27) or 1D (eq. D.28) potential and comparing the value of ϵ_ϕ where this transition happens, we find an agreement between our results and [252]. For instance, the maximum value of ϵ_ϕ for which we found a $1s$ bound state in eq. D.28 is $\epsilon_\phi \approx 0.595$ which matches the values reported in Table IV of Ref. [252] with $Z=1$ in their notation. For other bound states, this upper-bound on ϵ_ϕ is decreasing, in accord with Table III in Ref. [252]. For fixed mediator masses (W and Z bosons) and fixed mass splitting $\delta M = 0.165$ GeV, the energy of the bound states as a function of DM mass is studied in figure D-1.

D.3 The Coulombic limit: nonabelian "positronium"

In the case where the gauge symmetry is approximately unbroken, the force carriers can be treated as massless and the interacting DM states can be approximated as degenerate. The matrix potential then takes a simple form and it is possible to solve analytically for the bound and continuum wavefunctions, along with the rates for bound state formation and decay. This limiting case provides intuition and furnishes a useful test of the detailed numerical results. In this appendix we leave the detailed structure of the potential matrix arbitrary, so that the results we present here can easily be generalized to other gauge groups or representations, facilitating quick estimates of when bound state formation is important in other dark-sector models.

In this section we pick up where we left off in Section 7.3. In that section we presented the expression for the capture and transition cross sections in the dipole approximation (Equation 7.53), derived the form of the general scattering wavefunction (Equation 7.57), derived the form of the general bound state wavefunction (Equation 7.59), and derived the cross section for radiative capture from the continuum to the

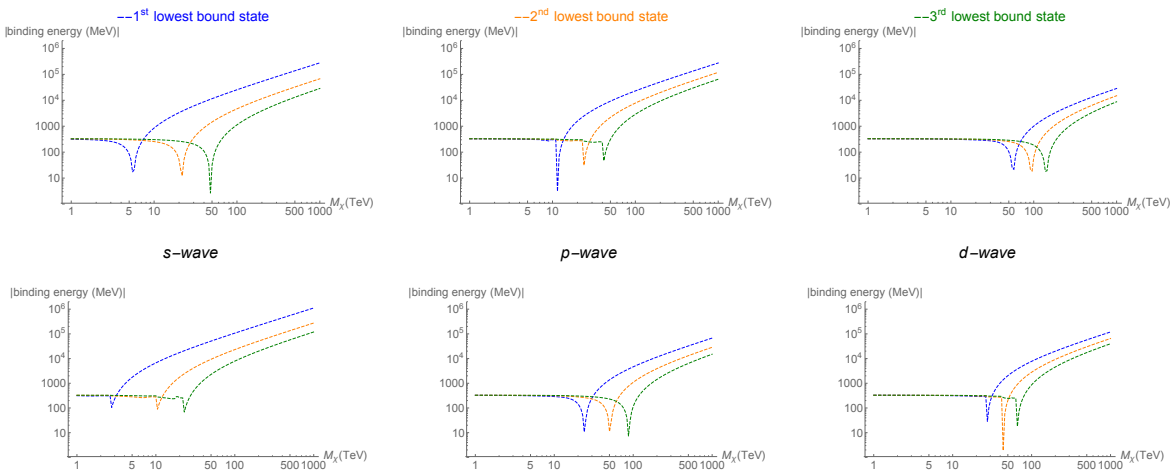

Figure D-1: The absolute value of the bound-state energy of the first three bound states, labeled by *blue*, *orange*, *green* in order of increasing n. **Top row:** spin-triplet configurations. **Bottom row:** spin-singlet configurations. From **left** to **right**, panels correspond to $L = 0, 1, 2$. All the diagrams follow a general trend: At low DM masses, where ϵ_ϕ is large, the Yukawa potential is too short-range to hold the particles in a bound state. The energy of the mostly $\chi^+\chi^-$ bound state – which is held together by photon exchange down to $M_{\chi^\pm} \to 0$ – is dominated by the mass splitting term, giving these states a positive binding energy in our convention. As we go to higher masses, corresponding to smaller ϵ_ϕ, the potential becomes long-range enough to form a negative-energy bound state. The dip on each line shows where the energy goes from being positive (compared to a neutralino pair at infinity) to being negative, as the DM mass increases. At large enough masses, the potentials are effectively all Coulombic, and the binding energy becomes linearly dependent on DM mass, as expected.

bound state (Equation 7.71).

Note that when the gauge symmetry is broken, radiation of massive gauge bosons is forbidden when the boson mass $m_V \gtrsim \alpha^2 M_\chi$, whereas the impact of the shortened range of the potential depends on the ratio of m_V to $\alpha\, M_\chi$; thus there is parametrically a region where the main difference between the broken and unbroken cases is simply that radiation of particular (linear combinations of) gauge bosons is kinematically forbidden in the broken case. Thus it can be productive to explicitly separate the bound-state formation induced by radiation of different gauge bosons; for the wino, we will calculate the capture/decay/annihilation rates with Coulombic wavefunctions, but still only consider radiation of the gauge boson combination that maps to the

photon in the unbroken theory.

When the gauge symmetry is truly unbroken, the incoming state could be in any linear combination of the (degenerate) two-body states; however, once the gauge symmetry is broken, when considering the case of annihilation of DM in the present day, the incoming particles will all (to a good approximation) be in the DM state, rather than one of the heavier states in the multiplet. Thus for ease of comparison to the broken case, it will be particularly useful to study the capture rate from the initial state that will correspond to the DM-DM state after symmetry breaking.

D.3.1 Capture cross sections for the wino case

As discussed in the main text, single-photon dipole emission changes angular momentum by $\Delta L = 1$ from an even $L + S$ state to an odd $L + S$ state, leading to different eigenvalues for the initial and final states. Neglecting the exponentially suppressed contributions from the negative eigenvalue of the even-$L + S$ potential matrix, for single-photon capture we need only consider the case with $\lambda_i = 2$, $\lambda_f = 1$. The associated eigenvectors are given in section 8.2.2.

If we impose the initial condition $I = \begin{pmatrix} 1 \\ \\ 0 \end{pmatrix}$ as discussed earlier (i.e. far from the interaction site, all particles are neutralinos rather than charginos), then $(\mathbf{I} \cdot \eta_i)(\eta_f^\dagger \hat{C}_1 \eta_i) = \begin{pmatrix} 1 \\ \\ 0 \end{pmatrix} \cdot \begin{pmatrix} \sqrt{\frac{1}{3}} \\ \\ \sqrt{\frac{2}{3}} \end{pmatrix} \begin{pmatrix} & \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{\frac{1}{3}} \\ \\ \sqrt{\frac{2}{3}} \end{pmatrix} = \frac{\sqrt{2}}{3}$, and likewise $(\mathbf{I} \cdot \eta_i)(\eta_f^\dagger \hat{C}_2 \eta_i) = -\frac{\sqrt{2}}{3}$.

Using eq. 7.71, replacing $\alpha_{\text{rad}} \to \alpha$, $\alpha \to \alpha_W$, and including the factor of 2 discussed above for identical particles in the initial state, we then obtain for the

capture cross sections:

$$\sigma v_{\rm rel} = \frac{2^{15}\pi^2}{3^3}\frac{\alpha\alpha_W^2}{M_\chi^2 v_{\rm rel}}\frac{1}{n^2}e^{-8n}f_{nlm}, \tag{D.32}$$

where the f_{nlm} are given by:

$$f_{100} = 0, \quad f_{200} = 128, \quad f_{210} = 242, \quad f_{21\pm1} = 50. \tag{D.33}$$

These cross sections are for initial states of definite spin (singlet or triplet). When averaging over all possible spin states, this contribution should be divided by 4 before adding it to the total for capture into states with odd L (where the leading contribution comes from even-L initial states, which must be spin-singlet), and multiplied by 3/4 for capture into states with even L (where the leading contribution comes from odd-L initial states, which must be spin-triplet). We expect the latter processes to be suppressed at low velocities, once the SU(2) symmetry is broken and the masses of the force carriers are non-negligible, as discussed in appendix D.4.

D.3.2 Capture cross sections separated by partial wave

We have so far computed the cross section for capture into the bound states from an incoming plane wave. However, for comparison to the regime where the force carrier masses are non-negligible, it is useful to separate out the contributions from different partial waves in the initial state, since we expect the higher partial wave contributions to be velocity-suppressed when the potential has a finite range (see appendix D.4).

For positronium, the properly normalized solution to the Schrödinger equation is given by eq. 7.55 (for distinguishable fermions; see the discussion following that equation for the case of indistinguishable fermions). Taking $\mathbf{p}$ in the z-direction, we can separate this into partial waves as follows:

$$\phi(\alpha;\mathbf{r}) = \sum_L (2L+1)P_L(\cos\theta)\frac{\Gamma(1+L-i\zeta)}{2ipr\Gamma(2(L+1))}e^{\pi\zeta/2}M\left[-i\zeta, \frac{1}{2}(1+2L), 2ipr\right], \tag{D.34}$$

where M is the Whittaker function, which is related to the hypergeometric function by:

$$M(a,b,z) = e^{-z/2} z^{b+1/2} \, {}_1F_1(1/2+b-a, 1+2b, z). \tag{D.35}$$

In terms of the confluent hypergeometric function we can write:

$$\phi(\alpha;\mathbf{r}) = \sum_L (2L+1) P_L(\cos\theta) \frac{\Gamma(1+L-i\zeta)}{\Gamma(2(L+1))} e^{\pi\zeta/2} e^{-ipr} (2ipr)^L \, {}_1F_1\left[1+L+i\zeta, 2(L+1), 2ipr\right]. \tag{D.36}$$

Capture from s-wave incoming state into p-wave bound state

Let us now consider the contribution to the capture cross section from the s-wave part of the incoming wino pair state, where the final state is one of the $n=2$, $l=1$ states. We will use this to cross-check our numerical results.,

From eq. D.34 the s-wave contribution to the continuum state, for a given eigenvalue λ_i, is given by:

$$\phi_s(\lambda_i\alpha; r) = \frac{\Gamma(1-i\lambda_i\alpha\mu/p)}{2ipr} e^{\pi\lambda_i\alpha\mu/2p} M\left[-i\lambda_i\alpha\mu/p, \frac{1}{2}, 2ipr\right]. \tag{D.37}$$

Repeating our previous analysis but projecting out the s-wave piece (i.e. replacing $\phi(\lambda_i\alpha;\mathbf{r})$ with $\phi_s(\lambda_i\alpha; r)$), we find for the reduced matrix element:

$$\begin{aligned}\bar{M} = \epsilon\cdot(\hat{r}_m)^* \frac{16\sqrt{2\pi}}{3} \alpha^{-1/2}\mu^{-3/2}\lambda_f^{-3/2} \sum_i (\mathbf{I}\cdot\eta_i)\Gamma(1-i\lambda_i\alpha\mu/p) e^{\pi\lambda_i\alpha\mu/2p} e^{-4\lambda_i/\lambda_f} \\ \times \eta_f^\dagger \left[\hat{C}_1\lambda_i\left(\frac{4\lambda_i}{\lambda_f}-3\right) + \hat{C}_2\left(3-12\frac{\lambda_i}{\lambda_f}+8\left(\frac{\lambda_i}{\lambda_f}\right)^2\right)\right]\eta_i.\end{aligned} \tag{D.38}$$

For the wino we have $(\mathbf{I}\cdot\eta_i)\eta_f^\dagger\hat{C}_1\eta_i = -(\mathbf{I}\cdot\eta_i)\eta_f^\dagger\hat{C}_2\eta_i = \sqrt{2}/3$, $\lambda_i = 2$, $\lambda_f = 1$. Substituting these into eq. D.38, we obtain:

$$\bar{M}_{s\to p} = -\epsilon\cdot(\hat{r}_m)^* \frac{2^5\sqrt{\pi}}{3^2} e^{-8} \alpha_W^{-1/2}\mu^{-3/2}\Gamma\left(1-\frac{2i\alpha_W\mu}{p}\right) e^{\pi\alpha_W\mu/p}, \tag{D.39}$$

for capture into any one of the $n=2$, $l=1$ states. (Here we have not inserted the

factor of $\sqrt{2}$ for the normalization of the initial state containing identical fermions, consistent with our previous approach of inserting it in the cross section.)

Evaluating eq. 7.68 for capture from an initial plane wave into the $n = 2$, $l = 1$ states, we obtain the matrix elements:

$$\bar{M}_{210} = -\epsilon \cdot (\hat{r}_0)^* \frac{2^5 \times 11}{3} \sqrt{\pi} e^{-8} \alpha_W^{-1/2} \mu^{-3/2} \Gamma \left(1 - \frac{2i\alpha_W \mu}{p}\right) e^{\pi \alpha_W \mu / p},$$
$$\bar{M}_{21\pm1} = \epsilon \cdot (\hat{r}_{\pm1})^* \frac{2^5 \times 5}{3} \sqrt{\pi} e^{-8} \alpha_W^{-1/2} \mu^{-3/2} \Gamma \left(1 - \frac{2i\alpha_W \mu}{p}\right) e^{\pi \alpha_W \mu / p}. \tag{D.40}$$

Comparing to eq. 7.68 for capture from an initial plane wave into these states, for the 210 state the s-wave contribution is $1/33$ of the total, whereas for the $m = \pm 1$ states it is $-1/15$ of the total (here the d-wave piece has the opposite sign, with magnitude $16/15$ of the total, and they add destructively). If the higher partial-wave components are suppressed, averaging over initial spin and summing over final states $m = 0, \pm 1$, and including the factor of 2 for the identical particles in the initial state, we obtain the total cross section for capture from the s-wave piece of the incoming state into the $n = 2$, $l = 1$ bound states:

$$\sigma v_{\rm rel} = \frac{2^{12} \pi^2}{3^4} e^{-16} \frac{\alpha \, \alpha_W^2}{M_\chi^2 v_{\rm rel}}. \tag{D.41}$$

Capture from p-wave incoming state into s-wave bound state

Capture into an s-wave state through single-photon dipole emission occurs only from p-wave states. Therefore the rates for p-to-s capture are just given by the full capture rates from the incoming plane wave into s-wave states (see Equation D.32). In particular, the capture rate into the ground state is zero through single-photon processes in the Coulomb limit. Thus it is small at high masses as we approach the Coulomb limit. One might wonder if two-photon processes can give an appreciable capture rate into the ground state. This has been studied by Refs. [254, 255] and found to be negligible.

D.3.3 Transition rates

We can likewise compute bound→bound rates using eq. 7.53 and the bound-state wavefunctions presented earlier; the decay rates are much simpler to calculate and can be done analytically using e.g. `Mathematica`, since both initial and final-state wavefunctions have fairly simple radial dependence (exponential functions of r multiplying polynomials). Accordingly, we will not detail the calculation here, but simply present results. We will be primarily interested in decays of the $n = 2$, $l = 1$ spin-singlet bound states, since these are populated by capture from the s-wave part of the initial continuum state.

The dominant decay of the $n = 2$, $l = 1$ states is to the spin-singlet ground state with $nlm = 100$. For general initial and final eigenvalues, we find a total decay rate of:

$$\Gamma = \frac{2^7}{3}\alpha\alpha_W^4 M_\chi \frac{\lambda_i(2\lambda_f - \lambda_i)(\lambda_i/\lambda_f)^4}{\left(2 + \frac{\lambda_i}{\lambda_f}\right)^7} \left|\lambda_f \eta_f^\dagger \hat{C}_1 \eta_i + \eta_f^\dagger \hat{C}_2 \eta_i\right|^2 . \tag{D.42}$$

For the wino, and assuming both states are spin-singlet, we have $\eta_f^\dagger \hat{C}_1 \eta_i = \eta_f^\dagger \hat{C}_2 \eta_i = \sqrt{\frac{2}{3}}$, $\lambda_i = 1$, and $\lambda_f = 2$, which yields:

$$\Gamma = \frac{2^{11} \times 3}{5^7}\, \alpha\, \alpha_W^4 M_\chi, \tag{D.43}$$

for decay from each of the initial states $n_i = 2$, $l_i = 1$, $m_i = 0, \pm 1$.

Considering more general decays from initial $n_i = 2$, $l_i = 1$ states, the rate is always proportional to $\alpha\, \alpha_W^4 M_\chi$, with proportionality factors given in table D.1.

The overall decay rates for each of the three initial states are thus,

$$\Gamma_{210} = \Gamma_{211} = \Gamma_{21-1} = 0.16\, \alpha\, \alpha_W^4 M_\chi, \tag{D.44}$$

with the main decays being to the ground state and the first excited s-wave state ($0.15/0.16 \approx 94\%$ of the total, with a 49% branching ratio to the ground state and a 44% branching ratio to the first excited state).

final nlm	$m_i = 0$	$m_i = 1$	$m_i = -1$	$m_i = 0$	$m_i = 1$	$m_i = -1$
100	$\frac{2^{11}\times 3}{5^7}$			7.9×10^{-2}		
200	$\frac{2^8\times 7^2}{3^{11}}$			7.1×10^{-2}		
300	$\frac{2^{11}\times 89^2}{3\times 7^{11}}$			2.7×10^{-3}		
320	$\frac{2^{24}}{3\times 7^{11}}$	$\frac{2^{22}}{3\times 7^{11}}$	$\frac{2^{22}}{3\times 7^{11}}$	2.8×10^{-3}	7.1×10^{-4}	7.1×10^{-4}
321	$\frac{2^{22}}{7^{11}}$	$\frac{2^{22}}{7^{11}}$	-	2.1×10^{-3}	2.1×10^{-3}	-
32-1	$\frac{2^{22}}{7^{11}}$	-	$\frac{2^{22}}{7^{11}}$	2.1×10^{-3}	-	2.1×10^{-3}
322	-	$\frac{2^{23}}{7^{11}}$	-	-	4.2×10^{-3}	-
32-2	-	-	$\frac{2^{23}}{7^{11}}$	-	-	4.2×10^{-3}

Table D.1: Exact (left) and approximate (right) numerical prefactors for the decay rate Γ of spin-singlet bound states with $n = 2$ and $l = 1$, which is given by $\alpha\,\alpha_W^4 M_\chi$ multiplied by this prefactor. Different columns indicate different initial states ($nlm = 21m_i$) while the rows label different final states.

D.3.4 Subtleties in imposing plane-wave boundary conditions on an infinite-range potential

When we earlier imposed the boundary condition that the original plane wave should be described by the vector $\mathbf{I}$, we tacitly assumed that the asymptotic solution *was* a plane wave, or at least that it could be characterized by a single r-independent vector multiplying a (scalar) modified plane wave. This is not correct for potentials with infinite range.

The true plane-wave-like solution to the Schrödinger equation at large range, in the presence of the matrix potential we consider, has the form:

$$\Psi(r) = e^{i\mathbf{p}\cdot\mathbf{r}} \sum_i A_i \eta_i e^{-i\frac{\alpha\lambda_i\mu}{p}\ln(pr-\mathbf{p}\cdot\mathbf{r})}. \tag{D.45}$$

This follows directly from the large-r solution of the scalar Schrödinger equations corresponding to the various eigenvalues. We can equivalently write this expression, more compactly, as:

$$\Psi(r) = e^{i\mathbf{p}\cdot\mathbf{r}} e^{-i\frac{\mu}{p}\alpha\bar{V}\ln(pr-\mathbf{p}\cdot\mathbf{r})} \sum_i A_i \eta_i. \tag{D.46}$$

From eq. D.46, it is apparent that the asymptotic coefficient of the distorted plane wave in this state should be considered to be $\mathbf{I} = \sum_i A_i\eta_i$, as in eq. 7.57. This is the constant quantity on which we can impose our boundary condition that the incoming state is in the DM-DM two-body state. But the relative probability assigned to the various two-body states (in the default basis) at large r will not be described by $\mathbf{I}$ alone, but by $e^{-i\frac{\mu}{p}\alpha\bar{V}\ln(pr-\mathbf{p}\cdot\mathbf{r})}\mathbf{I}$, where the exponential phase has non-trivial gauge structure through the $\bar{V}$ matrix.

It remains to check that the solution given by eq. 7.55 has the exact asymptotic plane-wave-like component we assumed above (in addition, it includes a scattered spherical wave):

$$\phi(\alpha;\mathbf{r}) \to e^{i\mathbf{p}\cdot\mathbf{r}} e^{-i\frac{\mu}{p}\alpha\ln(pr-\mathbf{p}\cdot\mathbf{r})} + \text{spherical scattered wave}. \tag{D.47}$$

To show this we use the well-known asymptotic form of the hypergeometric function:

$$\lim_{|z|\to\infty} {}_1F_1(a,b,z) = \frac{e^z z^{a-b}}{\Gamma(a)} + \frac{e^{i\pi a} z^{-a}}{\Gamma(b-a)}, \tag{D.48}$$

to obtain the asymptotic form of eq. 7.55:

$$\phi(\alpha;\mathbf{r}) \to e^{\pi\zeta/2}\Gamma(1-i\zeta)\left[\frac{e^{i(pr-\mathbf{p}\cdot\mathbf{r})}(i(pr-\mathbf{p}\cdot\mathbf{r}))^{1-i\zeta}}{\Gamma(i\zeta)} + \frac{e^{-\pi\zeta}(i(pr-\mathbf{p}\cdot\mathbf{r}))^{-i\zeta}}{\Gamma(1-i\zeta)}\right] e^{i\mathbf{p}\cdot\mathbf{r}}. \tag{D.49}$$

We identify the first term as the scattered wave and the second term as the plane-wave-like part; focusing on this second term, we find:

$$\begin{aligned}
\phi(\alpha;\mathbf{r}) &\to e^{-\pi\zeta/2}(e^{i\pi/2}(pr-\mathbf{p}\cdot\mathbf{r}))^{-i\zeta} e^{i\mathbf{p}\cdot\mathbf{r}} + \text{spherical scattered wave,} \\
&= e^{-\pi\zeta/2} e^{(i\pi/2)(-i\zeta)} e^{-i\zeta\ln(pr-\mathbf{p}\cdot\mathbf{r})} e^{i\mathbf{p}\cdot\mathbf{r}} + \text{spherical scattered wave,} \\
&= e^{-i\frac{\alpha\mu}{p}\ln(pr-\mathbf{p}\cdot\mathbf{r})} e^{i\mathbf{p}\cdot\mathbf{r}} + \text{spherical scattered wave,}
\end{aligned} \tag{D.50}$$

as required.

D.4 Effects of a massive force carrier

To understand the qualitative effects of giving a mass to (some of) the force carriers, so that the potential has a finite range, it is useful to consider the simpler case of a single DM state, with DM particles (of mass M_χ as previously) interacting via the attractive Hulthén potential,

$$V(r) = -\frac{\alpha_H m_H}{e^{m_H r} - 1}. \tag{D.51}$$

This potential has been considered as an approximation to the Yukawa potential $V(r) = -\frac{\alpha}{r}e^{-m_A r}$, in the context of the Sommerfeld enhancement [195, 256]. It has the advantage that the Schrödinger equation is exactly solvable for the s-wave states, and approximately solvable for higher partial waves; both continuum-state solutions [195] and bound-state solutions [256] have been presented in the literature,

for arbitrary l. In the calculation of the Sommerfeld enhancement, good agreement between the analytical results for the Hulthén potential and numerical results for the Yukawa potential is obtained with the identifications $\alpha = \alpha_H$, $m_H = \frac{\pi^2}{6} m_A$ [195].

In this appendix we briefly summarize the key results for the Hulthén potential, and then estimate the scaling of the bound state formation rate with m_H.

D.4.1 Wavefunctions for the Hulthén potential

In our notation, the radial wavefunctions for bound states of the Hulthén potential are approximately given by [256]:

$$R_{nl}(r) = \frac{N_{nl}}{r} e^{-\kappa r} (1 - e^{-m_H r})^{l+1} P_{n-l-1}^{(2\kappa/m_H, 2l+1-\kappa/m_H)} \left(1 - 2e^{-m_H r}\right), \tag{D.52}$$

where the bound state energy is $E_n = -\kappa^2/M_\chi$, with κ for a particular choice of principal quantum number n ($n = 1, 2, 3, ...$) given by:

$$\kappa_n = \frac{1}{2} \left(\frac{\alpha_H M_\chi - n^2 m_H}{n} \right). \tag{D.53}$$

Here N_{nl} is a (dimensionful) normalization factor, chosen to give the correct normalization for the bound state (it is straightforward to calculate for any given state, but the expressions are cumbersome), and $P_n^{(a,b)}(x)$ is the Jacobi P-function. These results are exact when $l = 0$ or $m_H = 0$; otherwise, they require use of an approximate form for the centrifugal term in the radial Schrödinger equation (see [256] for details).

Note that with the definitions above, we must have $\kappa_n > 0$ for bound states with principal quantum number n to exist, i.e. we must have $n < \sqrt{\alpha_H M_\chi / m_H}$. Thus when $m_H > \alpha_H M_\chi$, the potential supports no bound states; when $\alpha_H M_\chi / 4 < m_H < \alpha_H M_\chi$, only the $n = 1$ bound state exists; and so on. In the Coulomb limit where $m_H \to 0$, we recover the usual expression for the hydrogen-like bound state energies (recalling that the reduced mass of the system is $M_\chi/2$).

The partial-wave wavefunctions for continuum states with $m = 0$ are given by

[195]:

$$\psi_L(\mathbf{r}) = \frac{P_L(\cos\theta)}{(2L)!}\frac{\left(1-e^{-m_H r}\right)^{L+1}}{M_\chi r}e^{-i(M_\chi v_{\rm rel}/2)r}\ {}_2F_1\left[a^-, a^+, 2(L+1), 1-e^{-m_H r}\right]$$
$$\times\sqrt{\frac{\frac{2\pi}{\alpha_H v_{\rm rel}}\sinh\left(\frac{\pi v_{\rm rel}M_\chi}{m_H}\right)}{\cosh\left(\frac{\pi v_{\rm rel}M_\chi}{m_H}\right)-\cosh\left(\frac{\pi v_{\rm rel}M_\chi}{m_H}\sqrt{1-\frac{4\alpha_H m_H}{M_\chi v_{\rm rel}^2}}\right)}}$$
$$\times\sqrt{\prod_{k=0}^{L}\left(k^4+\left(\frac{\alpha M_\chi}{m_H}\right)^2+k^2\left(\frac{M_\chi v_{\rm rel}}{m_H}\right)^2\left(1-\frac{2\alpha_H m_H}{M_\chi v_{\rm rel}^2}\right)\right)}$$
$$a^\pm = 1+L+\frac{iM_\chi v_{\rm rel}}{2m_H}\left(1\pm\sqrt{1-\frac{4\alpha_H m_H}{M_\chi v_{\rm rel}^2}}\right). \tag{D.54}$$

In deriving these results from those of [195], we have used the identity $\Gamma(iz)\Gamma(-iz) = \pi/(z\sinh(\pi z))$. These wavefunctions are normalized to correspond to the L-wave components of a unit-normalized plane wave propagating in the z-direction. (Some caution must be taken in adding wavefunctions of different L together to re-form a plane wave, as the normalization condition does not fix the relative phases, but these wavefunctions will suffice for considering the scaling of the contributions to the capture rate by states of different L.)

There is an additional subtlety in this case; the approximate centrifugal term used by [195] to obtain the correct wavefunctions is exponentially suppressed at large r, whereas the real centrifugal term scales as $1/r^2$, and consequently the wavefunctions do not properly match onto the correct asymptotic solution at large r. (This is not a problem for the bound state wavefunctions, because the bound state wavefunctions are exponentially suppressed at large r independent of the centrifugal term.) This issue is independent of the details of any short-range potential (although it depends on m_H because the approximate form for the centrifugal term involves m_H), and so can be compensated by examining the effect of using the "wrong" centrifugal potential for plane waves; the prescription presented in [195] multiplies the short-range annihilation

rate derived using these wavefunctions by the factor:

$$C = \frac{w^{2L}}{\Pi_{k=0}^{L-1}[(L-k)^2 + w^2]}, \quad w = M_\chi v_{\rm rel}/m_H. \tag{D.55}$$

When using these approximate continuum wavefunctions to estimate scaling relationships, we will multiply the final capture rate by this correction factor. We do not expect this approach to be accurate in detail (for $L > 0$), because the capture process is not localized at very short distances in the same way as annihilation – indeed, the scale of the potential is the same as the scale of the (incorrect) centrifugal potential. However, it will suffice to estimate scaling relations for the capture rate; we will then confirm these scaling relations using numerically computed continuum wavefunctions.

D.4.2 Estimating the capture rate

Let us focus here on the cases of $s \to p$ and $p \to s$ capture, into the most deeply-bound available states in both cases; this will serve to illustrate the essential points. From eq. 8.23, the capture rate in this case will be given by:

$$\sigma v_{\rm rel}\,(\text{continuum} \to \text{bound}) \;= \frac{16}{3}\frac{\alpha E_n}{M_\chi^2}\left|\int r^2 dr\, \phi^*_{L=1}(r)\phi'_{L=0}(r)\right|^2 \times \begin{cases} 1 & \text{initial } s\text{-wave} \\ 1/3 & \text{initial } p\text{-wave} \end{cases}, \tag{D.56}$$

where $\psi(\mathbf{r}) = Y_{lm}(\theta,\phi)\phi(r)$. To make it easier to extract the scaling of $\sigma v_{\rm rel}$ with α_H, M_χ and m_H, let us define the dimensionless radial coordinate $x = \alpha_H M_\chi r$. Then $\sigma v_{\rm rel}$ scales as $(\alpha E_n/M_\chi^2)(\alpha_H M_\chi)^{-4}\left|\int x^2 dx\, \phi^*_{L=1}(x)\phi'_{L=0}(x)\right|^2$.

We will be especially interested in the regime of low velocity; the usual $1/v_{\rm rel}$ divergence as $v_{\rm rel} \to 0$ will in this case be regulated by the non-zero m_H.

Let us define the parameter $\xi = \alpha_H M_\chi/m_H$; note $\xi \geq 1$ if any bound states exist. In terms of the x coordinate and this parameter, the radial wavefunctions we need are given by:

- $nlm = 100$ bound state:

  $\phi(x) = (\alpha_H M_\chi)^{3/2} \sqrt{\frac{1}{2}\left(1 - \frac{1}{\xi^2}\right)} \frac{(1-e^{-x/\xi})}{x/\xi} e^{-(1-1/\xi)x/2}$

- $nlm = 210$ bound state (the scaling is identical for the $m = \pm 1$ states):

  $\phi(x) = (\alpha_H M_\chi)^{3/2} \frac{\sqrt{1 - \frac{20}{\xi^2} + \frac{64}{\xi^4}}}{16\sqrt{3}} x e^{-(1/4-1/\xi)x} \left(\frac{1-e^{-x/\xi}}{x/\xi}\right)^2$

- $L = 0$ contribution to continuum wavefunction in the low-velocity limit:

$$\phi(x) = \sqrt{\frac{(2\pi)^3 \xi}{1 - \cos\left(2\pi\sqrt{\xi}\right)}} \left(\frac{1 - e^{-x/\xi}}{x/\xi}\right) \, {}_2F_1\left[1 + \sqrt{\xi}, 1 - \sqrt{\xi}, 2, 1 - e^{-x/\xi}\right]. \tag{D.57}$$

- $L = 1$ contribution to continuum wavefunction in the low-velocity limit:

$$\phi(x) = \sqrt{\frac{\frac{2}{3}\pi^3 \xi}{1 - \cos\left(2\pi\sqrt{\xi}\right)}} (\xi - 1) \frac{\left(1 - e^{-x/\xi}\right)^2}{x/\xi} \, {}_2F_1\left[2 + \sqrt{\xi}, 2 - \sqrt{\xi}, 2, 1 - e^{-x/\xi}\right]. \tag{D.58}$$

Let us consider the limit where $\xi \gg 1$. In this case, the bound state wavefunctions have support for (and only for) $x \lesssim 1$. Thus in the region which gives a non-negligible contribution to the overlap integral that determines the capture rate, $\int x^2 \phi^*_{L=1}(x) \phi'_{L=0}(x) dx$, it follows that $x/\xi \ll 1$, and so we can safely approximate $1 - e^{-x/\xi} \approx x/\xi$ in computing this integral.

Using the identity:

$$ {}_2F_1(a, b, c, z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k k!} z^k, \tag{D.59}$$

and the approximation $(a)_k \sim a^n$ for $|a| \gg 1$, it follows that in the range of x relevant to the overlap integral,

$$ {}_2F_1[1+L+\sqrt{\xi}, 1+L-\sqrt{\xi}, 2(L+1), 1-e^{-x/\xi}] \approx \sum_{k=0}^{\infty} \frac{(-\sqrt{\xi})^k (\sqrt{\xi})^k}{(2(L+1))_k k!} (x/\xi)^k \approx \sum_{k=0}^{\infty} \frac{(-x)^k}{(2(L+1))_k k!}, \tag{D.60}$$

and consequently, this term in the continuum wavefunctions is approximately ξ-independent.

It follows that within the range where the integrand is non-negligible, provided $\xi \gg 1$, the bound-state wavefunctions $\phi(x)$ scale as $(\alpha_H M_\chi)^{3/2}\times$ (function of x only), with no leading-order ξ-dependence. The continuum wavefunctions scale approximately as $\sqrt{\xi/(1-\cos(2\pi\sqrt{\xi}))}\times$ (function of x only). Thus $\int x^2 \phi^*_{L=1}(x)\phi'_{L=0}(x)dx \propto (\alpha_H M_\chi)^{3/2}\sqrt{\xi/(1-\cos(2\pi\sqrt{\xi}))}$, where the proportionality factor arises from an integral over functions of x only, and is independent of all the parameters of the problem.

We can now estimate the scaling of the capture rate, for the limit as $v_{\rm rel} \to 0$ and assuming $m_H \ll \alpha_H M_\chi$:

$$\begin{aligned}\sigma v_{\rm rel}\,(\text{continuum} \to \text{bound}) &\propto \frac{\alpha E_n}{M_\chi^2}(\alpha_H M_\chi)^{-4}(\alpha_H M_\chi)^3 \frac{\xi}{1-\cos(2\pi\sqrt{\xi})} \\ &\propto \frac{\alpha E_n}{m_H}\frac{1}{M_\chi^2}\frac{1}{1-\cos(2\pi\sqrt{\alpha_H M_\chi/m_H})}. \end{aligned} \tag{D.61}$$

However, recall that – as discussed above – we must add a correction factor in the case where our initial continuum wavefunction has $L > 0$, to compensate for the incorrect asymptotic behavior of the centrifugal term and hence the continuum wavefunction. In the same limit as the one we are currently working in, where $v_{\rm rel}$ is taken to zero, the correction factor $C = (1/L!)^2(M_\chi v_{\rm rel}/m_H)^{2L}$ (eq. D.55). This is the *only* source of direct L dependence in the scaling. Thus our final scaling estimate becomes:

$$\sigma v_{\rm rel}\,(\text{continuum} \to \text{bound}) \propto \frac{\alpha E_n}{m_H}\frac{1}{M_\chi^2}\left(\frac{M_\chi v_{\rm rel}}{m_H}\right)^{2L}\frac{1}{1-\cos(2\pi\sqrt{\alpha_H M_\chi/m_H})}, \tag{D.62}$$

where L is the angular momentum quantum number for the initial state.

For comparison, in the Coulombic low-velocity limit we have:

$$\sigma v_{\rm rel}\,(\text{continuum} \to \text{bound}) \propto \frac{E_n}{M_\chi}\frac{1}{M_\chi^2}\frac{\alpha}{v_{\rm rel}}. \tag{D.63}$$

If the prefactors in both cases are similar, and away from the resonance regions where the $1/(1-\cos\theta)$ term can lead to a large enhancement in the case with $m_H \neq 0$, we would expect the two curves to intersect where $\frac{1}{M_\chi v_{\rm rel}} \sim \frac{1}{m_H}\left(\frac{M_\chi v_{\rm rel}}{m_H}\right)^{2L}$, i.e. when $\frac{M_\chi v_{\rm rel}}{m_H} \sim 1$.

Thus we expect to observe Coulomb-like behavior, and in particular the universal (independent of partial wave) $1/v_{\rm rel}$ velocity scaling in the capture rate, down to a "saturation velocity" $v_{\rm rel} = m_H/M_\chi$. At this point, the capture rate $\sigma v_{\rm rel}$ instead begins scaling as $v_{\rm rel}^{2L}$; accordingly, capture from higher partial waves is expected to be suppressed relative to capture from the s-wave part of the initial continuum state, by a factor of order $(v_{\rm rel} M_\chi/m_H)^{2L}$.

This is very similar to the parametric scaling of the Sommerfeld-enhanced annihilation rate [195]; the enhancement factor scales as $1/v_{\rm rel}^{2L+1}$ for velocities larger than the saturation velocity, but saturates at a constant value below this velocity. Since the bare un-enhanced annihilation rate ($\sigma v_{\rm rel}$) scales as $v_{\rm rel}^{2L}$, the enhanced annihilation rate scales as $1/v_{\rm rel}$ above the saturation velocity m_H/M_χ, but as $v_{\rm rel}^{2L}$ below this velocity scale, just as we find for the capture process. The resonance peaks in the capture rate that we expect for $v_{\rm rel} \lesssim m_H/M_\chi$ are also observed in the Sommerfeld enhancement. At these resonance locations, where $\sqrt{\alpha_H M_\chi/m_H}$ is close to an integer value, both Sommerfeld enhancement and capture rates pick up an extra scaling of $1/v_{\rm rel}^2$, down to a saturation velocity that depends on proximity to the resonance.

[201] presents an expression for a universal capture rate into near-threshold s-wave states, analogous to the result for capture of a proton and neutron into a deuteron bound state:

$$\sigma_c \sim \frac{p}{\gamma^2 + p^2}, \tag{D.64}$$

where the binding energy is γ^2/M_χ, and $p \propto M_\chi v_{\rm rel}$ is the incoming momentum. This result is applicable to our analysis when the range of the potential is short relative to the wavelength of the incoming state (i.e. $m_H \lesssim M_\chi v_{\rm rel}$), and the binding energy is small compared to m_H^2/M_χ. It predicts that $\sigma_c v_{\rm rel} \propto p^2$ when $\gamma \gg p$, and $\sigma_c v_{\rm rel}$ roughly p-independent when $\gamma \ll p$. These scalings agree with our results above; they

correspond to non-resonant and resonant capture into the s-wave ground state (from the p-wave part of the initial state), respectively, in the saturation regime where m_H cannot be neglected.

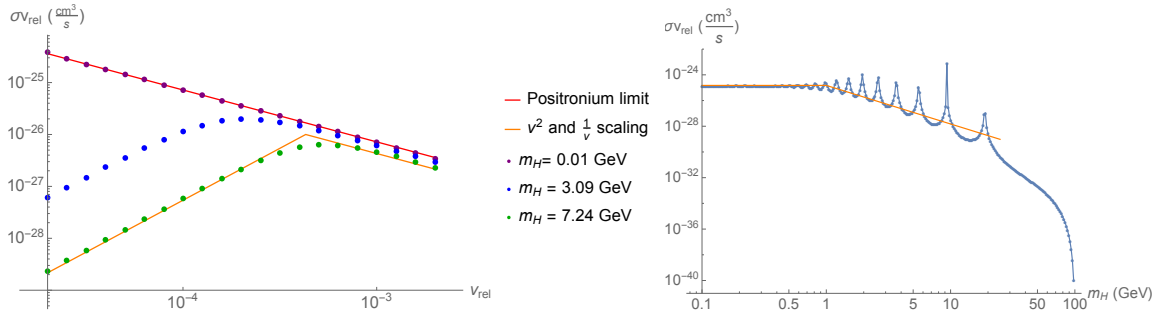

Figure D-2: $\sigma v_{\rm rel}$ for capture into the s-wave ground state in the Hulthén potential from a p-wave initial state, for $M_\chi = 10$ TeV, $\alpha_H = 0.01$. **Left:** $\sigma v_{\rm rel}$ as a function of $v_{\rm rel}$, for $m_H = 10$ MeV (*purple dots*), 3.09 GeV (*blue dots*) and 7.24 GeV (*green dots*). The *red* line indicates the analytic positronium-like limit where $m_H \to 0$; the *orange* line is proportional to $v_{\rm rel}^2$ ($v_{\rm rel}^{-1}$) below (above) the break, and is to guide the eye. **Right:** $\sigma v_{\rm rel}$ as a function of m_H, for $v_{\rm rel} = 15$ km/s $= 5 \times 10^{-5} c$ (*blue dots and line*). The *orange* line is proportional to m_H^0 (m_H^{-3}) below (above) the break, and is to guide the eye. Note we have chosen a small $v_{\rm rel}$ in order to display a large region with saturation and resonance effects, before m_H becomes too large to support a bound state.

To test our arguments above (and especially the scaling of the p-wave continuum states, since this relied on a correction factor that strictly only applies for short-range processes), we solved the Schrödinger equation with the Hulthén potential numerically (since this is a single-state problem, the numerical issues discussed in appendix D.1 are not a concern), computing the p-wave part of the initial plane wave. We used this numerical wavefunction and the analytic wavefunction for the ground state (for which the equations above are exact) to compute the capture rate into the ground state. Figure D-2 demonstrates the scaling of the resulting cross section with both $v_{\rm rel}$ and m_H, for sample parameters $\alpha_H = 0.01$, $M_\chi = 10$ TeV. We see the expected convergence to the positronium-like limit, with scaling $1/v_{\rm rel}$, for $m_H \ll M_\chi v_{\rm rel}$, and the resonance behavior and overall $v_{\rm rel}^2/m_H^3$ scaling for $m_H \gtrsim M_\chi v_{\rm rel}$. As m_H approaches $\alpha_H M_\chi = 100$ GeV, the cross section is suppressed and eventually vanishes

as the bound state energy goes to zero.

D.5 Conversion from Wilson coefficients to annihilation rates

In this section we demonstrate how to use the results in the literature on Wilson coefficients for processes of the form $\chi_{e_1}\chi_{e_2} \to XX \to \chi_{e_4}\chi_{e_3}$, as derived in [197, 229], in order to compute annihilation rates for WIMPonium bound states.

Hellmann and Ruiz-Femenia ([229], hereafter "HRF") write the perturbative, spin-averaged annihilation cross section in the form:

$$\begin{aligned}
\sigma^{\chi_{e_1}\chi_{e_2}\to X_A X_B} v_{\rm rel} &= a + (b_P + b_S)\frac{4p^2}{M_\chi^2} + \mathcal{O}(v_{\rm rel}^4), \\
a &= \hat{f}(^1S_0) + 3\hat{f}(^3S_1) \\
b_P &= \frac{1}{16}\left(\hat{f}(^1P_1) + \hat{f}(^3P_J)\right),
\end{aligned} \tag{D.65}$$

where the $\hat{f}$ terms are the Wilson coefficients derived in that work, we have set the reduced mass to $M_\chi/2$ and the mass of the two-particle state to $2M_\chi$, and we have written $v_{\rm rel} = 2p/M_\chi$, where p is the momentum of one of the initial particles in the COM frame. As discussed above, the non-spin-averaged leading-order perturbative annihilation rates from spin-triplet and spin-triplet initial states are then given by:

$$\begin{aligned}
&\text{singlet } L=0 : \sigma^{\chi_{e_1}\chi_{e_2}\to X_A X_B} v_{\rm rel} = 4\hat{f}(^1S_0), \\
&\text{triplet } L=0 : \sigma^{\chi_{e_1}\chi_{e_2}\to X_A X_B} v_{\rm rel} = 4\hat{f}(^3S_1), \\
&\text{singlet } L=1 : \sigma^{\chi_{e_1}\chi_{e_2}\to X_A X_B} v_{\rm rel} = \frac{p^2}{M_\chi^2}\hat{f}(^1P_1) \\
&\text{triplet } L=1 : \sigma^{\chi_{e_1}\chi_{e_2}\to X_A X_B} v_{\rm rel} = \frac{1}{3}\frac{p^2}{M_\chi^2}\hat{f}(^3P_J).
\end{aligned} \tag{D.66}$$

These Wilson coefficients are derived from the processes $\chi_{e_1}\chi_{e_2} \to XX \to \chi_{e_1}\chi_{e_2}$. In terms of the matrix element for $\chi_{e_1}\chi_{e_2} \to XX$, in our case where we can as-

sume the masses of the initial particle to be identical, they are proportional to $\int d\Pi_2 \, |\mathcal{M}(\chi_{e_1}(\mathbf{p})\chi_{e_2}(-\mathbf{p}) \to XX)|^2$ (with proportionality factors that can be trivially extracted from the expressions for the cross section just given). However, to compute the annihilation rates for bound states that mix the $\chi^0\chi^0$ and $\chi^+\chi^-$ states, we also need the terms of the form $\int d\Pi_2 \mathcal{M}^*(\chi_{e_1}(\mathbf{p})\chi_{e_2}(-\mathbf{p}) \to XX)\mathcal{M}(\chi_{e_4}(\mathbf{p})\chi_{e_3}(-\mathbf{p}) \to XX)$, where the initial states $\chi_{e_4}\chi_{e_3}$ and $\chi_{e_1}\chi_{e_2}$ need not be identical. These terms are given by the (absorptive part of the) off-diagonal Wilson coefficients extracted from the processes $\chi_{e_1}\chi_{e_2} \to XX \to \chi_{e_4}\chi_{e_3}$. We can thus promote the Wilson coefficients $\hat{f}$ for a given final state to a matrix $\hat{f}_{\{e_1e_2\}\{e_4e_3\}}$, where $\{e_1e_2\}$ and $\{e_4e_3\}$ label the relevant two-particle states. For the bound states we are interested in, the possible two-particle states are $\chi^0\chi^0$ and $\chi^+\chi^-$, and $\hat{f}$ is a matrix akin to the potential matrix defined in eq. 8.4.

Our situation is almost identical to the one we face in computing the Sommerfeld-enhanced annihilation cross section from the Wilson coefficients, which is discussed in detail in [197]; the only difference is a subtlety in how one normalizes the initial state (which in this case is a bound state, rather than a continuum state). For the s-wave case with Sommerfeld enhancement, the annihilation matrix – which we denote Σ – is to be contracted with the vector wavefunction at the origin to obtain the enhanced annihilation rate:

$$(\sigma v)_i = c_i \Psi^\dagger(0)\Sigma\Psi(0). \tag{D.67}$$

The prefactor c_i is 2 if the particles in the initial state are identical, and 1 otherwise. For the bound-state calculation, there is no such prefactor, as the different normalization of the bound state for identical particles cancels it out.

As discussed in [197], the annihilation matrix Σ is built from the Wilson coefficients, supplemented by factors of $1/\sqrt{2}$ if either the $\{e_1e_2\}$ or $\{e_4e_3\}$ states are comprised of identical particles, or by a factor of $1/2$ if both two-particle states are comprised of identical particles. In our analysis, we see that these factors of $1/\sqrt{2}$ arise naturally from the differing normalization of the bound states comprised of iden-

tical vs distinguishable particles (see section 8.3.2). There is also an overall prefactor relating the Wilson coefficients to the annihilation cross section, as discussed above.

Consequently, our algorithm for defining the general annihilation matrix (without spin averaging) is:

- Write down the Wilson coefficients $\hat{f}_{\{e_1e_2\}\{e_4e_3\}}$ for a specific final state, as calculated in [229] (these are given explicitly for the pure wino in appendix C3 of that work).

- Construct the annihilation matrix by:

$$\Sigma_{\{e_1e_2\}\{e_4e_3\}} = (1/\sqrt{2})^{n_i}\, c(L,S) \hat{f}_{\{e_1e_2\}\{e_4e_3\}}, \tag{D.68}$$

  where $n_i = 0$ if both $\{e_1e_2\}$ and $\{e_4e_3\}$ correspond to pairs of distinguishable particles, $n_i = 1$ if one pair is identical and the other distinguishable, and $n_i = 2$ if both pairs are comprised of identical particles (although the pairs may be different from each other). The constant prefactor $c(L,S)$, as discussed above, is 4 for s-wave states, 1 for spin-singlet p-wave states, 1/3 for spin-triplet p-wave states.

For $L = 0$, the diagonal elements of this matrix give the annihilation cross sections $\sigma v_{\rm rel}$ for particles initially in the appropriate two-particle continuum state in the absence of any potential, up to the factors of c_i discussed above. For $L = 1$, the diagonal matrix elements must be multiplied by an additional factor of (p^2/M_χ^2) to obtain the annihilation cross sections. We have stripped this latter factor out of the matrix Σ because it is precisely this factor that will be altered when the initial state is a bound state, rather than a free-particle continuum state.

The resulting annihilation matrix is precisely equivalent to the annihilation matrices defined in the main text in Eqs. 8.36-8.37. To check the normalization, note that for free-particle annihilation of nonrelativistic particles of equal mass M_χ, we have:

$$\sigma v_{\rm rel} = \frac{1}{(2M_\chi)^2} \int d\Pi_n \left|\mathcal{M}(\chi(\mathbf{p})\chi(-\mathbf{p}) \to f)\right|^2 . \tag{D.69}$$

Using the pure wino case as an example, from [229] we have the non-zero Wilson coefficients for the various cases:

Spin-singlet $L = 0$ ($^1S_0$)

- Final state W^+W^-: $\hat{f}_{\{\chi^0\chi^0\}\{\chi^0\chi^0\}} = \frac{2\pi\alpha_W^2}{M_\chi^2}$, $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = \frac{\pi\alpha_W^2}{2M_\chi^2}$, $\hat{f}_{\{\chi^+\chi^+\}\{\chi^0\chi^0\}} = \frac{\pi\alpha_W^2}{M_\chi^2}$,
- Final state ZZ: $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = c_W^4 \frac{\pi\alpha_W^2}{M_\chi^2}$,
- Final state $Z\gamma$: $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = 2c_W^2 s_W^2 \frac{\pi\alpha_W^2}{M_\chi^2}$,
- Final state $\gamma\gamma$: $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = s_W^4 \frac{\pi\alpha_W^2}{M_\chi^2}$,

Thus by the recipe above we obtain:

$$\Sigma(W^+W^-) = \frac{4\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} \end{pmatrix}, \quad \Sigma(ZZ) = \frac{4\pi\alpha_W^2}{M_\chi^2} c_W^4 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\Sigma(Z\gamma) = \frac{4\pi\alpha_W^2}{M_\chi^2} 2c_W^2 s_W^2 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Sigma(\gamma\gamma) = \frac{4\pi\alpha_W^2}{M_\chi^2} s_W^4 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{D.70}$$

Spin-triplet $L = 0$ ($^3S_1$)

- Final state W^+W^-: $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = \frac{1}{48}\frac{\pi\alpha_W^2}{M_\chi^2}$,
- Final state Zh^0: $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = \frac{1}{48}\frac{\pi\alpha_W^2}{M_\chi^2}$,
- Final state $q\bar{q}$ (for each individual quark species; multiply by 6 to get total rate): $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = \frac{1}{8}\frac{\pi\alpha_W^2}{M_\chi^2}$,
- Final state $l^+l^-, \nu\bar{\nu}$ (for each individual lepton or neutrino flavor; multiply by 6 to get total rate): $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = \frac{1}{24}\frac{\pi\alpha_W^2}{M_\chi^2}$,

By the recipe above we have:

$$\Sigma(W^+W^-) = \frac{1}{12}\frac{\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ & \\ 0 & 1 \end{pmatrix}, \quad \Sigma(Zh^0) = \frac{1}{12}\frac{\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ & \\ 0 & 1 \end{pmatrix},$$

$$\Sigma(q\bar{q}) = \frac{1}{2}\frac{\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ & \\ 0 & 1 \end{pmatrix}, \quad \Sigma(l^+l^-, \nu\bar{\nu}) = \frac{1}{6}\frac{\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ & \\ 0 & 1 \end{pmatrix}. \tag{D.71}$$

Spin-singlet $L = 1$ ($^1P_1$)

- Final state W^+W^-: $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = \frac{2}{3}\frac{\pi\alpha_W^2}{M_\chi^2}$

The only non-zero annihilation matrix is thus $\Sigma(W^+W^-) = \frac{2}{3}\frac{\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 0 & 0 \\ & \\ 0 & 1 \end{pmatrix}$.

Spin-triplet $L = 1$ (^3P_J)

- Final state W^+W^-: $\hat{f}_{\{\chi^0\chi^0\}\{\chi^0\chi^0\}} = \frac{56}{3}\frac{\pi\alpha_W^2}{M_\chi^2}$, $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = \frac{14}{3}\frac{\pi\alpha_W^2}{M_\chi^2}$, $\hat{f}_{\{\chi^+\chi^+\}\{\chi^0\chi^0\}} = \frac{28}{3}\frac{\pi\alpha_W^2}{M_\chi^2}$,

- Final state ZZ: $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = c_W^4\frac{28}{3}\frac{\pi\alpha_W^2}{M_\chi^2}$,

- Final state $Z\gamma$: $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = 2c_W^2 s_W^2\frac{28}{3}\frac{\pi\alpha_W^2}{M_\chi^2}$,

- Final state $\gamma\gamma$: $\hat{f}_{\{\chi^+\chi^-\}\{\chi^+\chi^-\}} = s_W^4\frac{28}{3}\frac{\pi\alpha_W^2}{M_\chi^2}$,

The annihilation matrices for this case are given by:

$$\Sigma(W^+W^-) = \frac{28}{9}\frac{\pi\alpha_W^2}{M_\chi^2}\begin{pmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} \end{pmatrix}, \quad \Sigma(ZZ) = \frac{28}{9}\frac{\pi\alpha_W^2}{M_\chi^2}c_W^4\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\Sigma(Z\gamma) = \frac{28}{9}\frac{\pi\alpha_W^2}{M_\chi^2}2c_W^2 s_W^2\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Sigma(\gamma\gamma) = \frac{28}{9}\frac{\pi\alpha_W^2}{M_\chi^2}s_W^4\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{D.72}$$

D.6 Useful integrals

In computing the continuum→bound capture rates to the lowest-lying s- and p-wave bound states in the Coulombic limit, we need to evaluate several non-trivial integrals; we collect the required results here for reference. Similar calculations have been presented elsewhere in the literature; e.g. [207] computed the integrals needed for capture to the ground state for both scalar and vector mediators, and for capture to the first excited p-wave state in the case of scalar mediators.

The integrals in question are:

$$\begin{aligned} I_1 &= \int_0^\infty r^2 dr \int_{-1}^1 dx \left[(x-1)e^{-\eta r}e^{iprx}\ {}_1F_1\left(1+i\zeta, 2, ipr(1-x)\right)\right] \\ I_2 &= \int_0^\infty r^2 dr \int_{-1}^1 dx \left[r(x-1)e^{-\eta r}e^{iprx}\ {}_1F_1\left(1+i\zeta, 2, ipr(1-x)\right)\right] \\ I_3 &= \int_0^\infty r^2 dr \int_{-1}^1 dx \left[rx(x-1)e^{-\eta r}e^{iprx}\ {}_1F_1\left(1+i\zeta, 2, ipr(1-x)\right)\right]. \end{aligned} \tag{D.73}$$

We will be particularly interested in the limit where $p \to 0$ while $p\,\zeta$ is held fixed, since this corresponds to the low-velocity limit of the capture rates.

Our starting point is the identities [182]:

$$\int_0^\infty r^2 dr \int_{-1}^1 dx \left[x e^{ipr x - \eta r} \; {}_1F_1(i\zeta, 1, ipr(1-x)) \right] = 4ip(1-i\zeta) \frac{(\eta - ip)^{-2i\zeta}}{(p^2+\eta^2)^{2-i\zeta}},$$
$$\int_0^\infty r^2 dr \int_{-1}^1 dx \left[\frac{1}{r} e^{ipr x - \eta r} \; {}_1F_1(i\zeta, 1, ipr(1-x)) \right] = 2 \frac{(\eta - ip)^{-2i\zeta}}{(p^2+\eta^2)^{1-i\zeta}}. \tag{D.74}$$

Differentiating these expressions with respect to p, we find:

$$\int_0^\infty r^2 dr \int_{-1}^1 dx e^{ipr x - \eta r} \left[\zeta r(x-1) x \; {}_1F_1(1+i\zeta, 2, ipr(1-x)) + irx^2 \; {}_1F_1(i\zeta, 1, ipr(1-x)) \right]$$
$$= 4i \frac{d}{dp} \left[p(1-i\zeta) \frac{(\eta - ip)^{-2i\zeta}}{(p^2+\eta^2)^{2-i\zeta}} \right],$$
$$\int_0^\infty r^2 dr \int_{-1}^1 dx e^{ipr x - \eta r} \left[-(1-x)\zeta \; {}_1F_1(1+i\zeta, 2, ipr(1-x)) + ix \; {}_1F_1(i\zeta, 1, ipr(1-x)) \right]$$
$$= 2 \frac{d}{dp} \left[\frac{(\eta - ip)^{-2i\zeta}}{(p^2+\eta^2)^{1-i\zeta}} \right], \tag{D.75}$$

The second line of eq. D.75 can be rewritten as:

$$\begin{aligned} I_1 &= \int_0^\infty r^2 dr \int_{-1}^1 dx e^{ipr x - \eta r} (x-1) \; {}_1F_1(1+i\zeta, 2, ipr(1-x)) \\ &= \frac{1}{\zeta} \left\{ 2 \frac{d}{dp} \left[\frac{(\eta - ip)^{-2i\zeta}}{(p^2+\eta^2)^{1-i\zeta}} \right] + 4p(1-i\zeta) \frac{(\eta - ip)^{-2i\zeta}}{(p^2+\eta^2)^{2-i\zeta}} \right\} \\ &= -4 \frac{(\eta - ip)^{-1-2i\zeta}}{(\eta^2+p^2)^{1-i\zeta}} \end{aligned} \tag{D.76}$$

Differentiating with respect to η then gives us:

$$\begin{aligned} I_2 &= \int_0^\infty r^2 dr \int_{-1}^1 dx e^{ipr x - \eta r} r(x-1) \; {}_1F_1(1+i\zeta, 2, ipr(1-x)) \\ &= -4(3\eta + ip - 2p\zeta) \frac{(\eta - ip)^{-1-2i\zeta}}{(\eta^2+p^2)^{2-i\zeta}} \end{aligned} \tag{D.77}$$

Finally, if ζ is large, the first line of eq. D.75 yields:

$$I_3 = \int_0^\infty r^2 dr \int_{-1}^1 dx e^{ipr x - \eta r} r x(x-1)\ {}_1F_1(1+i\zeta, 2, ipr(1-x))$$
$$\approx 4(\eta^2 - 3p^2 - 2\eta p\zeta)\frac{(\eta - ip)^{-2i\zeta}}{(\eta^2+p^2)^{3-i\zeta}}. \tag{D.78}$$

We can further simplify these integrals by taking the limit $p \to 0$, $p\,\zeta \to$ constant:

$$I_1 \to -4e^{-2p\zeta/\eta}/\eta^3, \quad I_2 \to -4(3\eta - 2p\zeta)e^{-2p\zeta/\eta}/\eta^5, \quad I_3 \to 4(\eta - 2p\zeta)e^{-2p\zeta/\eta}/\eta^5. \tag{D.79}$$

Finally, note also that:

$$\int_0^\infty r^2 dr \int_{-1}^1 dx \left[r(x^2-1)e^{-\eta r}e^{ipr x}\ {}_1F_1\left(1+i\zeta, 2, ipr(1-x)\right)\right] = I_2 + I_3 \to -8e^{-2p\zeta/\eta}/\eta^4. \tag{D.80}$$

In computing the integrals associated with the $\hat{C}_2$ structure, and with the capture rates separated by initial-state partial wave, we have employed the following integrals, which can all be performed by `Mathematica`:

$$\int r^2 dr e^{-\eta r}\ {}_1F_1\left(a, 1, iqr\right) \approx e^{iaq/\eta}\frac{2\eta^2 + 4i\eta aq - (aq)^2}{\eta^5},$$
$$\int r^3 dr e^{-\eta r}\ {}_1F_1\left(a, 1, iqr\right) \approx e^{iaq/\eta}\frac{6\eta^3 + 18i\eta^2 aq - 9\eta(aq)^2 - i(aq)^3}{\eta^7},$$
$$\int r^3 dr e^{-\eta r}\ {}_1F_1\left(a, 2, iqr\right) \approx e^{iaq/\eta}\frac{6\eta^2 + 6i\eta aq - (aq)^2}{\eta^6}$$
$$\int r^3 dr e^{-\eta r}\ {}_1F_1\left(a, 3, iqr\right) \approx 2e^{iaq/\eta}\frac{3\eta + iaq}{\eta^5} \tag{D.81}$$

Here we have evaluated the integrals in the limit $q \to 0$, but with $a\,q \to$ constant.

Bibliography

[1] J. K. Bloomfield, P. Fitzpatrick, K. Hilbert, and D. I. Kaiser, Phys. Rev. D **100**, 063512 (2019), arXiv:1906.08651 [astro-ph.CO] .

[2] P. J. Fitzpatrick, H. Liu, T. R. Slatyer, and Y.-D. Tsai, (2020), arXiv:2011.01240 [hep-ph] .

[3] P. J. Fitzpatrick, H. Liu, T. R. Slatyer, and Y.-D. Tsai, (2021), arXiv:2105.05255 [hep-ph] .

[4] P. Asadi, M. Baumgart, P. J. Fitzpatrick, E. Krupczak, and T. R. Slatyer, JCAP **02**, 005 (2017), arXiv:1610.07617 [hep-ph] .

[5] D. Kaiser, *Cosmic Inflation: An Informal Primer*, MIT Center for Theoretical Physics (2011).

[6] D. Kaiser, *FRW Spacetime: An Informal Primer*, MIT Center for Theoretical Physics (2011).

[7] D. Kaiser, *Gauge-Invariant Metric Perturbations: An Informal Primer*, MIT Center for Theoretical Physics (2011).

[8] D. Kaiser, *Primordial Perturbations from Inflation: An Informal Primer*, MIT Center for Theoretical Physics (2011).

[9] D. Baumann, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics of the Large and the Small* (2009) arXiv:0907.5424 [hep-th] .

[10] A. H. Guth and D. I. Kaiser, Science **307**, 884 (2005), arXiv:astro-ph/0502328 .

[11] V. Mukhanov, *Physical Foundations of Cosmology* (Cambridge University Press, Oxford, 2005).

[12] N. Aghanim *et al.* (Planck), Astron. Astrophys. **641**, A6 (2020), arXiv:1807.06209 [astro-ph.CO] .

[13] A. H. Guth, Phys. Rev. D **23**, 347 (1981).

[14] A. H. Guth and D. I. Kaiser, Science **307**, 884 (2005), arXiv:astro-ph/0502328 .

[15] V. Mukhanov, *Physical Foundations of Cosmology* (Cambridge University Press, Oxford, 2005).

[16] B. A. Bassett, S. Tsujikawa, and D. Wands, Rev. Mod. Phys. **78**, 537 (2006), arXiv:astro-ph/0507632 .

[17] D. H. Lyth and A. R. Liddle, *The primordial density perturbation: Cosmology, inflation and the origin of structure* (2009).

[18] J. Martin, C. Ringeval, and V. Vennin, Phys. Dark Univ. **5-6**, 75 (2014), arXiv:1303.3787 [astro-ph.CO] .

[19] A. H. Guth, D. I. Kaiser, and Y. Nomura, Phys. Lett. B **733**, 112 (2014), arXiv:1312.7619 [astro-ph.CO] .

[20] A. Linde, in *100e Ecole d'Ete de Physique: Post-Planck Cosmology* (2015) pp. 231–316, arXiv:1402.0526 [hep-th] .

[21] D. Baumann and L. McAllister, *Inflation and String Theory*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2015) arXiv:1404.2601 [hep-th] .

[22] J. Martin, (2019), arXiv:1902.05286 [astro-ph.CO] .

[23] D. S. Goldwirth and T. Piran, Phys. Rept. **214**, 223 (1992).

[24] R. Brandenberger, Int. J. Mod. Phys. D **26**, 1740002 (2016), arXiv:1601.01918 [hep-th] .

[25] D. S. Salopek and J. R. Bond, Phys. Rev. D **42**, 3936 (1990).

[26] R. H. Brandenberger and J. H. Kung, Phys. Rev. D **42**, 1008 (1990).

[27] V. Muller, H. J. Schmidt, and A. A. Starobinsky, Class. Quant. Grav. **7**, 1163 (1990).

[28] A. R. Liddle, P. Parsons, and J. D. Barrow, Phys. Rev. D **50**, 7222 (1994), arXiv:astro-ph/9408015 .

[29] K. Tzirakis and W. H. Kinney, Phys. Rev. D **75**, 123510 (2007), arXiv:astro-ph/0701432 .

[30] G. N. Remmen and S. M. Carroll, Phys. Rev. D **88**, 083518 (2013), arXiv:1309.2611 [gr-qc] .

[31] F. Azhar and D. I. Kaiser, Phys. Rev. D **98**, 063515 (2018), arXiv:1807.02088 [astro-ph.CO] .

[32] D. Chowdhury, J. Martin, C. Ringeval, and V. Vennin, Phys. Rev. D **100**, 083537 (2019), arXiv:1902.03951 [astro-ph.CO] .

[33] A. Albrecht, R. H. Brandenberger, and R. Matzner, Phys. Rev. D **32**, 1280 (1985).

[34] A. Albrecht, R. H. Brandenberger, and R. Matzner, Phys. Rev. D **35**, 429 (1987).

[35] H. Kurki-Suonio, R. A. Matzner, J. Centrella, and J. R. Wilson, Phys. Rev. D **35**, 435 (1987).

[36] R. H. Brandenberger and H. A. Feldman, Phys. Lett. B **220**, 361 (1989).

[37] D. S. Goldwirth and T. Piran, Phys. Rev. D **40**, 3263 (1989).

[38] D. S. Goldwirth and T. Piran, Phys. Rev. Lett. **64**, 2852 (1990).

[39] D. S. Goldwirth, Phys. Rev. D **43**, 3204 (1991).

[40] P. Laguna, H. Kurki-Suonio, and R. A. Matzner, Phys. Rev. D **44**, 3077 (1991).

[41] H. Kurki-Suonio, P. Laguna, and R. A. Matzner, Phys. Rev. D **48**, 3611 (1993), arXiv:astro-ph/9306009 .

[42] T. Vachaspati and M. Trodden, Phys. Rev. D **61**, 023502 (1999), arXiv:gr-qc/9811037 .

[43] A. Ijjas, P. J. Steinhardt, and A. Loeb, Phys. Lett. B **723**, 261 (2013), arXiv:1304.2785 [astro-ph.CO] .

[44] L. Berezhiani and M. Trodden, Phys. Lett. B **749**, 425 (2015), arXiv:1504.01730 [hep-th] .

[45] W. E. East, M. Kleban, A. Linde, and L. Senatore, JCAP **09**, 010 (2016), arXiv:1511.05143 [hep-th] .

[46] K. Clough, E. A. Lim, B. S. DiNunno, W. Fischler, R. Flauger, and S. Paban, JCAP **09**, 025 (2017), arXiv:1608.04408 [hep-th] .

[47] K. Clough, R. Flauger, and E. A. Lim, JCAP **05**, 065 (2018), arXiv:1712.07352 [hep-th] .

[48] C. Joana and S. Clesse, Phys. Rev. D **103**, 083501 (2021), arXiv:2011.12190 [astro-ph.CO] .

[49] M. C. D. Marsh, J. D. Barrow, and C. Ganguly, JCAP **05**, 026 (2018), arXiv:1803.00625 [gr-qc] .

[50] R. Easther, L. C. Price, and J. Rasero, JCAP **08**, 041 (2014), arXiv:1406.2869 [astro-ph.CO] .

[51] R. Jackiw, Phys. Rev. D **9**, 1686 (1974).

[52] L. Dolan and R. Jackiw, Phys. Rev. D **9**, 3320 (1974).

[53] J. M. Cornwall, R. Jackiw, and E. Tomboulis, Phys. Rev. D **10**, 2428 (1974).

[54] S.-J. Chang, Phys. Rev. D **12**, 1071 (1975).

[55] D. Boyanovsky, H. J. de Vega, and R. Holman, Phys. Rev. D **49**, 2769 (1994), arXiv:hep-ph/9310319 .

[56] M. Kleban and L. Senatore, JCAP **10**, 022 (2016), arXiv:1602.03520 [hep-th] .

[57] S. Dodelson and L. Hui, Phys. Rev. Lett. **91**, 131301 (2003), arXiv:astro-ph/0305113 .

[58] A. R. Liddle and S. M. Leach, Phys. Rev. D **68**, 103503 (2003), arXiv:astro-ph/0305263 .

[59] J. T. Giblin, J. B. Mertens, G. D. Starkman, and C. Tian, Phys. Rev. D **99**, 023527 (2019), arXiv:1810.05203 [astro-ph.CO] .

[60] R. D. Mattuck, *A guide to Feynman diagram in the many-body problem*, European physics series (McGraw-Hill Pub. Co., London, New York [etc, 1967).

[61] S. Y. Khlebnikov and I. I. Tkachev, Phys. Lett. B **390**, 80 (1997), arXiv:hep-ph/9608458 .

[62] B. A. Bassett and S. Tsujikawa, Phys. Rev. D **63**, 123503 (2001), arXiv:hep-ph/0008328 .

[63] B. A. Bassett, F. Tamburini, D. I. Kaiser, and R. Maartens, Nucl. Phys. B **561**, 188 (1999), arXiv:hep-ph/9901319 .

[64] H. Collins and R. Holman, Phys. Rev. D **74**, 045009 (2006), arXiv:hep-th/0605107 .

[65] H. Collins, R. Holman, and T. Vardanyan, JHEP **10**, 124 (2014), arXiv:1408.4801 [hep-th] .

[66] D. Seery, Class. Quant. Grav. **27**, 124005 (2010), arXiv:1005.1649 [astro-ph.CO] .

[67] L. Senatore and M. Zaldarriaga, JHEP **09**, 148 (2013), arXiv:1210.6048 [hep-th] .

[68] H. Andernach and F. Zwicky, "English and spanish translation of zwicky's (1933) the redshift of extragalactic nebulae," (2017), arXiv:1711.01693 [astro-ph.IM] .

[69] V. C. Rubin, J. Ford, W. K., and N. Thonnard, The Astrophysical Journal **238**, 471 (1980).

[70] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, The Astrophysical Journal **648**, L109 (2006).

[71] M. Tegmark *et al.* (SDSS), Astrophys. J. **606**, 702 (2004), arXiv:astro-ph/0310725 .

[72] V. Iršič *et al.*, Phys. Rev. D **96**, 023522 (2017), arXiv:1702.01764 [astro-ph.CO] .

[73] M. Schumann, J. Phys. G **46**, 103003 (2019), arXiv:1903.03026 [astro-ph.CO] .

[74] F. Kahlhoefer, Int. J. Mod. Phys. A **32**, 1730006 (2017), arXiv:1702.02430 [hep-ph] .

[75] T. Marrodán Undagoitia and L. Rauch, J. Phys. G **43**, 013001 (2016), arXiv:1509.08767 [physics.ins-det] .

[76] D. E. Gruber, J. L. Matteson, L. E. Peterson, and G. V. Jung, Astrophys. J. **520**, 124 (1999), arXiv:astro-ph/9903492 .

[77] L. Bouchet, E. Jourdain, J. P. Roques, A. Strong, R. Diehl, F. Lebrun, and R. Terrier, Astrophys. J. **679**, 1315 (2008), arXiv:0801.2086 [astro-ph] .

[78] C. Johnson, R. Caputo, C. Karwin, S. Murgia, S. Ritz, and J. Shelton, Phys. Rev. D **99**, 103007 (2019), arXiv:1904.06261 [astro-ph.HE] .

[79] M. Ajello *et al.* (Fermi-LAT), Astrophys. J. **819**, 44 (2016), arXiv:1511.02938 [astro-ph.HE] .

[80] M. Ackermann *et al.* (Fermi-LAT), Astrophys. J. **840**, 43 (2017), arXiv:1704.03910 [astro-ph.HE] .

[81] L. Rinchiuso (H.E.S.S.), EPJ Web Conf. **209**, 01023 (2019), arXiv:1901.05299 [astro-ph.HE] .

[82] S. Archambault *et al.* (VERITAS), Phys. Rev. D **95**, 082001 (2017), arXiv:1703.04937 [astro-ph.HE] .

[83] M. G. Aartsen *et al.* (IceCube), Eur. Phys. J. C **77**, 146 (2017), [Erratum: Eur.Phys.J.C 79, 214 (2019)], arXiv:1612.05949 [astro-ph.HE] .

[84] A. Albert *et al.* (ANTARES), Phys. Dark Univ. **16**, 41 (2017), arXiv:1612.06792 [hep-ex] .

[85] M. G. Aartsen *et al.* (IceCube), Eur. Phys. J. C **78**, 831 (2018), arXiv:1804.03848 [astro-ph.HE] .

[86] V. Barger, W. Y. Keung, D. Marfatia, and G. Shaughnessy, Phys. Lett. B **672**, 141 (2009), arXiv:0809.0162 [hep-ph] .

[87] M.-Y. Cui, Q. Yuan, Y.-L. S. Tsai, and Y.-Z. Fan, Phys. Rev. Lett. **118**, 191101 (2017), arXiv:1610.03840 [astro-ph.HE] .

[88] P. von Doetinchem, T. Aramaki, S. Boggs, H. Fuke, C. Hailey, I. Mognet, R. A. Ong, K. Perez, and J. Zweerink (GAPS), PoS **ICRC2015**, 1219 (2016), arXiv:1507.02717 [astro-ph.IM] .

[89] M. Lisanti, in *Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings* (2016) arXiv:1603.03797 [hep-ph] .

[90] E. W. Kolb and M. S. Turner, *The Early Universe*, Vol. 69 (1990).

[91] T. R. Slatyer, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Anticipating the Next Discoveries in Particle Physics* (2017) arXiv:1710.05137 [hep-ph] .

[92] Y. Hochberg, E. Kuflik, T. Volansky, and J. G. Wacker, Physical Review Letters **113** (2014), 10.1103/physrevlett.113.171301.

[93] E. Kuflik, M. Perelstein, N. R.-L. Lorier, and Y.-D. Tsai, Phys. Rev. Lett. **116**, 221302 (2016), arXiv:1512.04545 [hep-ph] .

[94] E. Kuflik, M. Perelstein, N. R.-L. Lorier, and Y.-D. Tsai, JHEP **08**, 078 (2017), arXiv:1706.05381 [hep-ph] .

[95] G. Steigman, B. Dasgupta, and J. F. Beacom, Phys. Rev. D **86**, 023506 (2012), arXiv:1204.3622 [hep-ph] .

[96] M. Battaglieri *et al.*, in *U.S. Cosmic Visions: New Ideas in Dark Matter* (2017) arXiv:1707.04591 [hep-ph] .

[97] D. Akerib *et al.* (LUX), Phys. Rev. Lett. **118**, 021303 (2017), arXiv:1608.07648 [astro-ph.CO] .

[98] E. Aprile *et al.* (XENON), Phys. Rev. Lett. **121**, 111302 (2018), arXiv:1805.12562 [astro-ph.CO] .

[99] R. Agnese *et al.* (SuperCDMS), Phys. Rev. Lett. **120**, 061802 (2018), arXiv:1708.08869 [hep-ex] .

[100] Q. Wang *et al.* (PandaX-II), (2020), arXiv:2007.15469 [astro-ph.CO] .

[101] F. Bergsma *et al.* (CHARM), Phys. Lett. **B166**, 473 (1986).

[102] F. Bergsma *et al.* (CHARM), Phys. Lett. **157B**, 458 (1985).

[103] A. Konaka *et al.*, *Proceedings, 23RD International Conference on High Energy Physics, JULY 16-23, 1986, Berkeley, CA*, Phys. Rev. Lett. **57**, 659 (1986).

[104] J. D. Bjorken, S. Ecklund, W. R. Nelson, A. Abashian, C. Church, B. Lu, L. W. Mo, T. A. Nunamaker, and P. Rassmann, Phys. Rev. **D38**, 3375 (1988).

[105] M. Davier and H. Nguyen Ngoc, Phys. Lett. **B229**, 150 (1989).

[106] J. Blümlein *et al.*, Z. Phys. **C51**, 341 (1991).

[107] J. Blümlein *et al.*, Int. J. Mod. Phys. **A7**, 3835 (1992).

[108] D. Banerjee *et al.* (NA64), Phys. Rev. Lett. **120**, 231802 (2018), arXiv:1803.07748 [hep-ex] .

[109] J. R. Batley *et al.* (NA48/2), Phys. Lett. **B746**, 178 (2015), arXiv:1504.00607 [hep-ex] .

[110] Y.-D. Tsai, P. deNiverville, and M. X. Liu, (2019), arXiv:1908.07525 [hep-ph] .

[111] E. D. Carlson, M. E. Machacek, and L. J. Hall, The Astrophysical Journal **398**, 43 (1992).

[112] K. Griest and D. Seckel, Phys. Rev. D **43**, 3191 (1991).

[113] R. T. DâĂŹAgnolo and J. T. Ruderman, Physical Review Letters **115** (2015), 10.1103/physrevlett.115.061301.

[114] J. M. Cline, H. Liu, T. R. Slatyer, and W. Xue, Physical Review D **96** (2017), 10.1103/physrevd.96.083521.

[115] N. Aghanim *et al.* (Planck), Astron. Astrophys. **641**, A6 (2020), arXiv:1807.06209 [astro-ph.CO] .

[116] T. Gherghetta, J. Kersten, K. Olive, and M. Pospelov, Phys. Rev. D **100**, 095001 (2019), arXiv:1909.00696 [hep-ph] .

[117] M. Cirelli, P. Panci, K. Petraki, F. Sala, and M. Taoso, Journal of Cosmology and Astroparticle Physics **2017**, 036 (2017).

[118] J. Edsjo and P. Gondolo, Phys. Rev. D **56**, 1879 (1997), arXiv:hep-ph/9704361 .

[119] R. T. D'Agnolo and J. T. Ruderman, Phys. Rev. Lett. **115**, 061301 (2015).

[120] J. M. Cline, H. Liu, T. Slatyer, and W. Xue, Phys. Rev. **D96**, 083521 (2017), arXiv:1702.07716 [hep-ph] .

[121] E. D. Carlson, M. E. Machacek, and L. J. Hall, Astrophys. J. **398**, 43 (1992).

[122] D. Pappadopulo, J. T. Ruderman, and G. Trevisan, Phys. Rev. D **94**, 035005 (2016), arXiv:1602.04219 [hep-ph] .

[123] M. Farina, D. Pappadopulo, J. T. Ruderman, and G. Trevisan, JHEP **12**, 039 (2016), arXiv:1607.03108 [hep-ph] .

[124] J. A. Evans, S. Gori, and J. Shelton, JHEP **02**, 100 (2018), arXiv:1712.03974 [hep-ph] .

[125] E. Aprile *et al.* (XENON), Phys. Rev. Lett. **123**, 251801 (2019), arXiv:1907.11485 [hep-ex] .

[126] L. Barak *et al.* (SENSEI), (2020), arXiv:2004.11378 [astro-ph.CO] .

[127] P. Agnes *et al.* (DarkSide), Phys. Rev. Lett. **121**, 111303 (2018), arXiv:1802.06998 [astro-ph.CO] .

[128] E. Aprile *et al.* (XENON), Phys. Rev. Lett. **123**, 241803 (2019), arXiv:1907.12771 [hep-ex] .

[129] D. Baxter, Y. Kahn, and G. Krnjaic, Phys. Rev. D **101**, 076014 (2020), arXiv:1908.00012 [hep-ph] .

[130] D. Amaral *et al.* (SuperCDMS), (2020), arXiv:2005.14067 [hep-ex] .

[131] G. Raffelt and D. Seckel, Phys. Rev. Lett. **60**, 1793 (1988).

[132] G. G. Raffelt, *Stars as laboratories for fundamental physics* (University of Chicago Press, 1996).

[133] A. Burrows and J. M. Lattimer, The Astrophysical Journal **307**, 178 (1986).

[134] A. Burrows and J. M. Lattimer, Astrophys. J. Lett. **318**, L63 (1987).

[135] N. Bar, K. Blum, and G. D'Amico, Phys. Rev. D **101**, 123025 (2020), arXiv:1907.05020 [hep-ph] .

[136] J. H. Chang, R. Essig, and S. D. McDermott, JHEP **09**, 051 (2018), arXiv:1803.00993 [hep-ph] .

[137] K. Bondarenko, A. Sokolenko, A. Boyarsky, A. Robertson, D. Harvey, and Y. Revaz, JCAP **01**, 043 (2021), arXiv:2006.06623 [astro-ph.CO] .

[138] J. Pollack, *Supermassive Black Holes from Gravothermal Collapse of Fractional Self-Interacting Dark Matter halos*, Bachelor thesis, Princeton U. (2012).

[139] T. R. Slatyer, Phys. Rev. D **93**, 023527 (2016), arXiv:1506.03811 [hep-ph] .

[140] Y. Izotov, T. Thuan, and N. Guseva, Mon. Not. Roy. Astron. Soc. **445**, 778 (2014), arXiv:1408.6953 [astro-ph.CO] .

[141] E. Aver, K. A. Olive, and E. D. Skillman, JCAP **07**, 011 (2015), arXiv:1503.08146 [astro-ph.CO] .

[142] R. J. Cooke, M. Pettini, and C. C. Steidel, Astrophys. J. **855**, 102 (2018), arXiv:1710.11129 [astro-ph.CO] .

[143] E. Zavarygin, J. Webb, S. Riemer-Sørensen, and V. Dumont, J. Phys. Conf. Ser. **1038**, 012012 (2018), arXiv:1801.04704 [astro-ph.CO] .

[144] M. Valerdi, A. Peimbert, M. Peimbert, and A. Sixtos, Astrophys. J. **876**, 98 (2019), arXiv:1904.01594 [astro-ph.GA] .

[145] C. Boehm, M. J. Dolan, and C. McCabe, JCAP **08**, 041 (2013), arXiv:1303.6270 [hep-ph] .

[146] N. Sabti, J. Alvey, M. Escudero, M. Fairbairn, and D. Blas, JCAP **01**, 004 (2020), arXiv:1910.01649 [hep-ph] .

[147] K. M. Nollett and G. Steigman, Phys. Rev. D **89**, 083508 (2014), arXiv:1312.5725 [astro-ph.CO] .

[148] K. N. Abazajian, Phys. Rept. **711-712**, 1 (2017), arXiv:1705.01837 [hep-ph] .

[149] M. Drewes *et al.*, JCAP **01**, 025 (2017), arXiv:1602.04816 [hep-ph] .

[150] N. Palanque-Delabrouille, C. Yèche, N. Schöneberg, J. Lesgourgues, M. Walther, S. Chabanier, and E. Armengaud, JCAP **04**, 038 (2020), arXiv:1911.09073 [astro-ph.CO] .

[151] A. Schneider, R. E. Smith, and D. Reed, Mon. Not. Roy. Astron. Soc. **433**, 1573 (2013), arXiv:1303.0839 [astro-ph.CO] .

[152] D. P. Finkbeiner and N. Weiner, Phys. Rev. D **76**, 083519 (2007), arXiv:astro-ph/0702587 .

[153] D. P. Finkbeiner, L. Goodenough, T. R. Slatyer, M. Vogelsberger, and N. Weiner, JCAP **05**, 002 (2011), arXiv:1011.3082 [hep-ph] .

[154] G. Elor, H. Liu, T. R. Slatyer, and Y. Soreq, Phys. Rev. D **98**, 036015 (2018), arXiv:1801.07723 [hep-ph] .

[155] T. G. Rizzo, (2020), arXiv:2006.08502 [hep-ph] .

[156] B. Holdom, Phys. Lett. B **166**, 196 (1986).

[157] D. Tucker-Smith and N. Weiner, Phys. Rev. D **64**, 043502 (2001), arXiv:hep-ph/0101138 .

[158] M. Baryakhtar, A. Berlin, H. Liu, and N. Weiner, (2020), arXiv:2006.13918 [hep-ph] .

[159] V. Poulin, J. Lesgourgues, and P. D. Serpico, JCAP **03**, 043 (2017), arXiv:1610.10051 [astro-ph.CO] .

[160] K. Schutz and T. R. Slatyer, JCAP **1501**, 021 (2015), arXiv:1409.2867 [hep-ph] .

[161] S. Gninenko, Phys. Rev. **D85**, 055027 (2012), arXiv:1112.5438 [hep-ph] .

[162] S. Gninenko, Phys. Lett. **B713**, 244 (2012), arXiv:1204.3583 [hep-ph] .

[163] D. Banerjee *et al.* (NA64), Phys. Rev. Lett. **118**, 011802 (2017), arXiv:1610.02988 [hep-ex] .

[164] P. Ilten, Y. Soreq, J. Thaler, M. Williams, and W. Xue, Phys. Rev. Lett. **116**, 251803 (2016), arXiv:1603.08926 [hep-ph] .

[165] R. Aaij *et al.* (LHCb), (2017), arXiv:1710.02867 [hep-ex] .

[166] P. Ilten, Y. Soreq, M. Williams, and W. Xue, JHEP **06**, 004 (2018), arXiv:1801.04847 [hep-ph] .

[167] R. Aaij *et al.* (LHCb), Phys. Rev. Lett. **124**, 041801 (2020), arXiv:1910.06926 [hep-ex] .

[168] A. Sung, G. Guo, and M.-R. Wu, (2021), arXiv:2102.04601 [hep-ph] .

[169] Y. Zhang, Phys. Dark Univ. **15**, 82 (2017), arXiv:1611.03492 [hep-ph] .

[170] M. Blennow, S. Clementz, and J. Herrero-Garcia, JCAP **03**, 048 (2017), arXiv:1612.06681 [hep-ph] .

[171] P. Ilten, J. Thaler, M. Williams, and W. Xue, Phys. Rev. D **92**, 115017 (2015), arXiv:1509.06765 [hep-ph] .

[172] J. Feng, I. Galon, F. Kling, and S. Trojanowski, Phys. Rev. **D97**, 035001 (2018), arXiv:1708.09389 [hep-ph] .

[173] W. Altmannshofer *et al.* (Belle-II), PTEP **2019**, 123C01 (2019), [Erratum: PTEP 2020, 029201 (2020)], arXiv:1808.10567 [hep-ex] .

[174] A. Berlin, S. Gori, P. Schuster, and N. Toro, Phys. Rev. D **98**, 035011 (2018), arXiv:1804.00661 [hep-ph] .

[175] S. Gninenko, *Addendum to the NA64 Proposal: Search for the $A' \to$ invisible and $X \to e^+e^-$ decays in 2021*, Tech. Rep. (CERN, Geneva, 2018).

[176] P. H. Adrian *et al.* (HPS), Phys. Rev. D **98**, 091101 (2018), arXiv:1807.11530 [hep-ex] .

[177] A. Caldwell *et al.*, (2018), arXiv:1812.11164 [physics.acc-ph] .

[178] L. Doria, P. Achenbach, M. Christmann, A. Denig, and H. Merkel, PoS **ALPS2019**, 022 (2020), arXiv:1908.07921 [hep-ex] .

[179] M. D'Onofrio, O. Fischer, and Z. S. Wang, Phys. Rev. D **101**, 015020 (2020), arXiv:1909.02312 [hep-ph] .

[180] N. Collaboration (NA62 Collaboration), *ADDENDUM I TO P326 Continuation of the physics programme of the NA62 experiment*, Tech. Rep. (CERN, Geneva, 2019).

[181] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory* (Addison-Wesley, Reading, USA, 1995).

[182] A. I. Akhiezer and N. P. Merenkov, Journal of Physics B Atomic Molecular Physics **29**, 2135 (1996).

[183] M. Cirelli, N. Fornengo, and A. Strumia, Nucl. Phys. **B753**, 178 (2006), arXiv:hep-ph/0512090 [hep-ph] .

[184] G. Jungman, M. Kamionkowski, and K. Griest, Phys. Rept. **267**, 195 (1996), arXiv:hep-ph/9506380 [hep-ph] .

[185] S. Dimopoulos, *ESO-CERN Topical Workshop on LEP and the Universe Geneva Switzerland, April 5-6, 1990*, Phys. Lett. **B246**, 347 (1990).

[186] J. Bovy, D. W. Hogg, and H.-W. Rix, The Astrophysical Journal **704**, 1704 (2009), arXiv:0907.5423 [astro-ph.GA] .

[187] M. Beneke, C. Hellmann, and P. Ruiz-Femenia, JHEP **03**, 148 (2013), [Erratum: JHEP10,224(2013)], arXiv:1210.7928 [hep-ph] .

[188] M. Baumgart, I. Z. Rothstein, and V. Vaidya, Phys. Rev. Lett. **114**, 211301 (2015), arXiv:1409.4415 [hep-ph] .

[189] M. Bauer, T. Cohen, R. J. Hill, and M. P. Solon, *Proceedings, Meeting of the APS Division of Particles and Fields (DPF 2015)*, JHEP **01**, 099 (2015), arXiv:1409.7392 [hep-ph] .

[190] G. Ovanesyan, T. R. Slatyer, and I. W. Stewart, Phys. Rev. Lett. **114**, 211302 (2015), arXiv:1409.8294 [hep-ph] .

[191] J. Hisano, S. Matsumoto, M. M. Nojiri, and O. Saito, Phys. Rev. **D71**, 063528 (2005), arXiv:hep-ph/0412403 [hep-ph] .

[192] J. Hisano, S. Matsumoto, M. Nagai, O. Saito, and M. Senami, Phys. Lett. **B646**, 34 (2007), arXiv:hep-ph/0610249 [hep-ph] .

[193] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, Phys. Rev. **D79**, 015014 (2009), arXiv:0810.0713 [hep-ph] .

[194] M. Cirelli, A. Strumia, and M. Tamburini, Nucl. Phys. **B787**, 152 (2007), arXiv:0706.4071 [hep-ph] .

[195] S. Cassel, J. Phys. **G37**, 105009 (2010), arXiv:0903.5307 [hep-ph] .

[196] T. R. Slatyer, JCAP **1002**, 028 (2010), arXiv:0910.5713 [hep-ph] .

[197] M. Beneke, C. Hellmann, and P. Ruiz-Femenia, JHEP **05**, 115 (2015), arXiv:1411.6924 [hep-ph] .

[198] J. Hisano, S. Matsumoto, and M. M. Nojiri, Phys. Rev. Lett. **92**, 031303 (2004), arXiv:hep-ph/0307216 [hep-ph] .

[199] J. D. March-Russell and S. M. West, Phys. Lett. **B676**, 133 (2009), arXiv:0812.0559 [astro-ph] .

[200] W. Shepherd, T. M. P. Tait, and G. Zaharijas, Phys. Rev. **D79**, 055022 (2009), arXiv:0901.2125 [hep-ph] .

[201] E. Braaten and H. W. Hammer, Phys. Rev. **D88**, 063511 (2013), arXiv:1303.4682 [hep-ph] .

[202] R. Laha and E. Braaten, Phys. Rev. **D89**, 103510 (2014), arXiv:1311.6386 [hep-ph] .

[203] M. B. Wise and Y. Zhang, JHEP **02**, 023 (2015), [Erratum: JHEP10,165(2015)], arXiv:1411.1772 [hep-ph] .

[204] M. B. Wise and Y. Zhang, Phys. Rev. **D90**, 055030 (2014), [Erratum: Phys. Rev.D91,no.3,039907(2015)], arXiv:1407.4121 [hep-ph] .

[205] K. Petraki, L. Pearce, and A. Kusenko, JCAP **1407**, 039 (2014), arXiv:1403.1077 [hep-ph] .

[206] B. von Harling and K. Petraki, JCAP **1412**, 033 (2014), arXiv:1407.7874 [hep-ph] .

[207] K. Petraki, M. Postma, and M. Wiechers, JHEP **06**, 128 (2015), arXiv:1505.00109 [hep-ph] .

[208] Y. Tsai, L.-T. Wang, and Y. Zhao, Phys. Rev. **D93**, 035024 (2016), arXiv:1511.07433 [hep-ph] .

[209] H. An, B. Echenard, M. Pospelov, and Y. Zhang, Phys. Rev. Lett. **116**, 151801 (2016), arXiv:1510.05020 [hep-ph] .

[210] H. An, M. B. Wise, and Y. Zhang, Phys. Rev. **D93**, 115020 (2016), arXiv:1604.01776 [hep-ph] .

[211] H. An, M. B. Wise, and Y. Zhang, (2016), arXiv:1606.02305 [hep-ph] .

[212] X.-J. Bi, Z. Kang, P. Ko, J. Li, and T. Li, (2016), arXiv:1602.08816 [hep-ph] .

[213] C. Kouvaris, K. Langaeble, and N. G. Nielsen, JCAP **1610**, 012 (2016), arXiv:1607.00374 [hep-ph] .

[214] T. Cohen, M. Lisanti, A. Pierce, and T. R. Slatyer, JCAP **1310**, 061 (2013), arXiv:1307.4082 .

[215] J. Fan and M. Reece, JHEP **10**, 124 (2013), arXiv:1307.4400 [hep-ph] .

[216] M. Baumgart, I. Z. Rothstein, and V. Vaidya, JHEP **04**, 106 (2015), arXiv:1412.8698 [hep-ph] .

[217] M. Baumgart and V. Vaidya, JHEP **03**, 213 (2016), arXiv:1510.02470 [hep-ph] .

[218] A. Abramowski *et al.* (HESS), Phys. Rev. Lett. **110**, 041301 (2013), arXiv:1301.1173 [astro-ph.HE] .

[219] M. Doro *et al.* (CTA Consortium), Astropart. Phys. **43**, 189 (2013), arXiv:1208.5356 [astro-ph.IM] .

[220] A. U. Abeysekara *et al.* (HAWC), Phys. Rev. **D90**, 122002 (2014), arXiv:1405.1730 [astro-ph.HE] .

[221] J. P. Harding and B. Dingus (HAWC), in *Proceedings, 34th International Cosmic Ray Conference (ICRC 2015)* (2015) arXiv:1508.04352 [astro-ph.HE] .

[222] F. Chen, J. M. Cline, and A. R. Frey, Phys. Rev. **D80**, 083516 (2009), arXiv:0907.4746 [hep-ph] .

[223] M. Ibe, S. Matsumoto, and R. Sato, Phys. Lett. **B721**, 252 (2013), arXiv:1212.5989 [hep-ph] .

[224] M. Pospelov and A. Ritz, Phys. Lett. **B671**, 391 (2009), arXiv:0810.1502 [hep-ph] .

[225] S. Weinberg, *Lectures on Quantum Mechanics* (Cambridge University Press, 2013).

[226] M. Beneke, *Proceedings, 8th International Symposium on Heavy Flavor Physics (Heavy Flavors 8)*, (1999), [PoShf8,009(1999)], arXiv:hep-ph/9911490 [hep-ph] .

[227] A. V. Manohar and I. W. Stewart, Phys. Rev. **D62**, 014033 (2000), arXiv:hep-ph/9912226 [hep-ph] .

[228] A. V. Manohar and I. W. Stewart, Phys. Rev. **D63**, 054004 (2001), arXiv:hep-ph/0003107 [hep-ph] .

[229] C. Hellmann and P. Ruiz-Femenia, JHEP **08**, 084 (2013), arXiv:1303.0200 [hep-ph] .

[230] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory; 1995 ed.* (Westview, Boulder, CO, 1995) includes exercises.

[231] K. A. Olive *et al.* (Particle Data Group), Chin. Phys. **C38**, 090001 (2014).

[232] G. B. Arfken and H. J. Weber, *Mathematical methods for physicists; 4th ed.* (Academic Press, San Diego, CA, 1995).

[233] T. Bringmann and S. Hofmann, JCAP **04**, 016 (2007), [Erratum: JCAP 03, E02 (2016)], arXiv:hep-ph/0612238 .

[234] T. Bringmann, New J. Phys. **11**, 105027 (2009), arXiv:0903.0189 [astro-ph.CO] .

[235] H. Liu, G. W. Ridgway, and T. R. Slatyer, Phys. Rev. D **101**, 023530 (2020), arXiv:1904.09296 [astro-ph.CO] .

[236] A. Berlin and F. Kling, Phys. Rev. D **99**, 015021 (2019), arXiv:1810.01879 [hep-ph] .

[237] B. Batell, M. Pospelov, and A. Ritz, Phys. Rev. D **79**, 115019 (2009), arXiv:0903.3396 [hep-ph] .

[238] D. P. Finkbeiner, T. R. Slatyer, N. Weiner, and I. Yavin, JCAP **09**, 037 (2009), arXiv:0903.1037 [hep-ph] .

[239] T. R. Slatyer and C.-L. Wu, Phys. Rev. D **95**, 023010 (2017), arXiv:1610.06933 [astro-ph.CO] .

[240] W. Hu and J. Silk, Phys. Rev. Lett. **70**, 2661 (1993).

[241] J. R. Ellis, G. B. Gelmini, J. L. Lopez, D. V. Nanopoulos, and S. Sarkar, Nucl. Phys. B **373**, 399 (1992).

[242] J. Chluba, Mon. Not. Roy. Astron. Soc. **436**, 2232 (2013), arXiv:1304.6121 [astro-ph.CO] .

[243] J. Chluba and D. Jeong, Mon. Not. Roy. Astron. Soc. **438**, 2065 (2014), arXiv:1306.5751 [astro-ph.CO] .

[244] A. Denner, H. Eck, O. Hahn, and J. Kublbeck, Phys. Lett. B **291**, 278 (1992).

[245] A. Denner, H. Eck, O. Hahn, and J. Kublbeck, Nucl. Phys. **B387**, 467 (1992).

[246] R. Mertig, M. Bohm, and A. Denner, Comput. Phys. Commun. **64**, 345 (1991).

[247] V. Shtabovenko, R. Mertig, and F. Orellana, (2016), arXiv:1601.01167 [hep-ph] .

[248] V. Shtabovenko, R. Mertig, and F. Orellana, Comput. Phys. Commun. **256**, 107478 (2020), arXiv:2001.04407 [hep-ph] .

[249] H. H. Patel, Comput. Phys. Commun. **197**, 276 (2015), arXiv:1503.01469 [hep-ph] .

[250] H. H. Patel, Comput. Phys. Commun. **218**, 66 (2017), arXiv:1612.00009 [hep-ph] .

[251] S. N. Ershov, J. Vaagen, and M. Zhukov, Phys. Rev. **C84**, 064308 (2011).

[252] F. J. Rogers, H. C. Graboske, and D. J. Harwood, Phys. Rev. A **1**, 1577 (1970).

[253] K. M. Roussel and R. F. O'Connell, Phys. Rev. A **9**, 52 (1974).

[254] E. Johnson, (2016), `http://indico.cern.ch/event/432527/contributions/1071485/attachments/1320722/1980515/Enhancement_of_Wino_Dark_Matter_Annihilation_through_the_Radiative_Formation_of_Bound_States.pdf`.

[255] E. Johnson, E. Braaten, and H. Zhang, in *38th International Conference on High Energy Physics (ICHEP 2016) Chicago, IL, USA, August 03-10, 2016* (2016) arXiv:1611.06212 [hep-ph] .

[256] M. Hamzavi, M. Movahedi, K.-E. Thylwe, and A. A. Rajabi, Chinese Physics Letters **29**, 080302 (2012).