Scuola Internazionale Superiore di Studi Avanzati - Trieste

ASTROPARTICLE PHYSICS CURRICULUM

PHILOSOPHIAE DOCTOR THESIS

Tackling the dark matter problem: a phenomenologist's perspective

Supervisor:
Prof. Andrea De Simone

Candidate:
Alessandro Morandini

16th September 2020

SISSA - Via Bonomea 265 - 34136 TRIESTE - ITALY

Abstract

In this thesis, we will approach the dark matter problem from several different directions. This work is developed from the point of view of a dark matter phenomenologist and as such it will involve both theory and experiments.

We will first review the construction of the standard model and its problems. Among these problems, we focus our attention on the discussion of dark matter, revising the evidences, the candidates and the experiments currently looking for it.

We proceed by building two original models with a good dark matter candidate: the motivations for building the two models will be very different. First, we build a composite Higgs model driven by theoretical motivations. The construction of the model relies entirely on symmetry arguments, furthermore we have the possibility of solving more than one standard model problem at once. Then, we build a model where our dark matter candidate is accompanied by a colored particle and the motivation for such a model is tied to the experiments. In this second model the focus will be on the prospects for detectability and study at LHC, more specifically on the interplay between different signatures.

But this is not the end of the story. Finally, there is a third approach that is neither theoretical, nor experimental, but rather data-driven. Here we will try to investigate the possibility of finding dark matter by applying anomaly detection to the LHC data.

Acknowledgments

First of all, I would like to thank my family for always supporting me. Even though I have been far from my parents most of the time, I knew I could always count on them.

Then I want to thank Andrea, for always being able to guide and encourage me, without ever making my job stressful. Of course, I thank also the other people I had the pleasure to work with in these years and especially Alle: I cannot put into words how rewarding it was to work with him.

Next, I want to express my gratitude and my love to my close friends, my "cluster". Every step of the way, if I was ever feeling down, I could rely on them to cheer me up or at least to tell me that it be like that sometimes.

Finally, I want to thank everybody at SISSA. It is hard to describe how friendly this environment is, from the secretariat, to the professors, to the canteen staff. But the most important part are my colleagues, who have become like a second family to me. Ale, Alfredo, Lollone, Paolo, Shani, Uriel, Valerio, Warren: you have no idea how much I will miss you.

Contents

Ai miei genitori

Preface

In this thesis, we will explore the dark matter (DM) problem. Specifically, we will try to tackle the study of dark matter with the approaches available to a particle physics phenomenologist.

Two dark matter models will be built by us: one motivated by theoretical arguments, the other by experiments at LHC. A third, data-driven approach will be discussed at the end.

In chapter 1, we describe how to derive the Lagrangian for a model from its symmetries and in particular we build the standard model (SM) Lagrangian. Then, we will give a brief overview of the main SM problems.

In chapter 2, we will study in detail an experimental problem of the SM: the presence of dark matter. Here we explain why we believe there is DM and which candidates have been proposed, focusing on the weakly interacting massive particle (WIMP). We conclude the chapter with a discussion of the experiments that are looking for dark matter.

In chapter 3, we discuss our model of theoretical interest. The study is based on Ref. [1]. Here, the model follows directly from symmetry arguments. This model delivers a good DM candidate which is consistent with all experimental constraints. In addition to that, the issues of naturalness and Higgs hierarchy are addressed as well. For what concerns the DM phenomenology, the field content of the model allows both the thermal freeze-out and the late-time decay of a partner particle.

In chapter 4, we discuss our model of experimental interest. The study is based on Ref. [2] and has further led to a contribution to the white paper [3]. In this case we have a simplified model, whose interest comes from the presence of a colored partner. The role of this particle is two-fold: a colored partner boosts the production at LHC, which is dominated by QCD, and also permits the presence of a long-lived particle (LLP). We will discuss the detectability of the model, more specifically the interplay between the LLP signature and standard DM signatures at LHC, such as the monojet.

In chapter 5, we present a third approach, based on data analysis and in particular on density estimation. The study is based on Ref. [4] and has led to a contribution to a Les Houches Proceedings [5]. Starting from the physical concept of Markov chains, we develop a density estimator: this can be used in any density based approach to data science. We show its prospects for use in a local outlier detector applied to general scope datasets. We then look at what happens when applied to the high energy physics data from DarkMachines.

Finally, we conclude in chapter 6.

Chapter 1

Symmetries, Lagrangians and the Standard Model

The introduction of quantum mechanics reshaped the way scientists thought about physics. Among the several consequences of the change of paradigm, physicists addressing the quantum world had to leave behind many results learned in classical mechanics. The quantum world obeys different rules and different tools are needed to study it. This partially explains why the development of quantum field theory (QFT) has been so tricky.

And yet one object that can be constructed in classical physics retains its potential even in the quantum world: the Lagrangian. This is easily justified, in quantum physics symmetries are a fundamental guiding principle and the Lagrangian is exactly encoding information about the symmetries of the system. The common lore is that a Lagrangian is all a particle physicist needs. While this is clearly over-simplistic, we can understand many of the successes and problems of the standard model just by looking at its Lagrangian. Building the SM Lagrangian and understanding its problems will be the main topics of this chapter.

1.1 Symmetries of the standard model

1.1.1 A brief group theory summary

In order to understand the construction of the standard model, it is useful to have a brief reminder about group theory. Many modifications of the SM rely on group theoretical arguments as well. This will be the case for instance of our models of theoretical and experimental interest.

The fundamental object we will have to deal with are Lie groups [6, 7]. For what concerns a particle physicist, it suffices to say that a Lie group is a group of continuous transformations which is also a smooth differentiable manifold. Given this latter property, it is always possible to consider the tangent space to the Lie group in order to study local properties. This will be denoted as the Lie algebra and in this tangent space we will be close to the identity transformation of the Lie group. This implies that a general transformation U of the Lie group can be expressed in terms of the algebra elements T^a as

$$U = e^{i\theta^a T^a}. \tag{1.1}$$

It is worth remembering that an algebra is a vector space V identified by its elements (here denoted with T^a) and a bilinear relation $V \times V \to V$. This bilinear relation takes the name of Lie bracket when talking about Lie algebras and it is represented by a commutator:

$$[T^a, T^b] = i f^{abc} T^c. \tag{1.2}$$

The algebra representation can then be identified by specifying the value of f^{abc}, which is known as the structure constant.

We have been talking about groups and algebras as if the two descriptions were equivalent. Now one may wonder whether the Lie group can be described by the Lie algebra in a unique way. Indeed, it is a known result that if two Lie groups have the same Lie algebra they are locally isomorphic to each other and share the same universal cover. This is a consequence of Lie's third theorem [8]. Furthermore, if a Lie group is simply connected, then it is in bi-univocal relation to the algebra.

This is good enough for us. Notably, this allows to simplify the classification of Lie groups by looking at the algebras. A result of primary importance by Cartan [9] is the classification of all possible simple algebras. The importance of these algebras for quantum physics cannot be overstated. Simple finite dimensional algebras and direct sums of simple finite dimensional algebras preserve unitarity, a fundamental requirement for quantum theories. The result from Cartan is that the only finite dimensional simple groups are the special unitary group SU(N), the special orthogonal group SO(N), the symplectic group Sp(N) and the exceptional groups G_2, F_4, E_6, E_7, E_8.

We now wish to use these groups and algebras in practice, so we need some kind of representation. Indeed, a representation can be obtained by embedding the generators inside matrices. The possible representations (which we will denote as R) are a priori infinite, since their dimensionality can be arbitrary large. However, from a minimalist perspective, it makes sense to consider only the lowest dimensional ones. As a matter of fact, in the SM the largest representation for a field is an octet. The only representations that will be needed to discuss the SM and our proposed extensions are the fundamental, anti-fundamental and adjoint.

The fundamental representation is the smallest non-trivial representation we can get, the anti-fundamental can be obtained by complex conjugation of the fundamental one (and it has of course the same dimensionality). A larger algebra representation is the adjoint one, which will have the same dimensionality of the group.

We conclude this brief summary by saying that a representation R can be characterized by some invariants: the index $C(R)$, the quadratic Casimir $C_2(R)$ and the anomaly coefficient $A(R)$. These can be calculated as (we denote the fundamental representation with $R = F$):

$$\begin{aligned} \mathrm{Tr}[T_R^a T_R^b] &= C(R)\delta^{ab}, & (1.3)\\ T_R^a T_R^a &= C_2(R), & (1.4)\\ \mathrm{Tr}[T_R^a \{T_R^b, T_R^c\}] &= A(R)\,\mathrm{Tr}[T_F^a \{T_F^b, T_F^c\}]. & (1.5) \end{aligned}$$

1.1.2 From symmetries to Lagrangians

Our aim now is constructing a theory starting from some symmetries [10, 11]. The core distinction we have to make here is the one between local and global symmetries. In eq. (1.1)

we introduced some transformation parameters θ^a, these parameters can either be global or local. If they are indeed local, this means that $\theta^a = \theta^a(x)$. In the case of theories with local symmetries we will use equivalently the term "gauge theory" and quantities respecting this symmetry will be called "gauge invariant". The idea of building a Lagrangian in a way that is gauge invariant has been guiding model builders for dozens of years. We are here no exceptions, as gauge invariance of the Lagrangian will be central in both the models we will study in detail.

Building a Lagrangian that is globally symmetric under the wanted symmetry is easier than building a Lagrangian that is locally symmetric. The starting point will then be constructing all the terms that are globally symmetric given a specific field content. An infinite series of term would be allowed, so people consider terms with the lowest dimensions. In particular, if we only consider terms with dimensionality less than four, the theory will (usually) be renormalizable. The theory constructed in this way is consistent and globally invariant.

If you have a theory with some global symmetries and you wish to promote a global symmetry to a local one, a few adjustments have to be made. A local field A^a_μ needs to be introduced. In order to retain gauge invariance, this field, called the gauge boson, will need to transform in the adjoint representation. These fields can be used to make the terms involving derivatives gauge invariant, by making the derivative covariant:

$$\partial_\mu \quad \rightarrow \quad D_\mu = \partial_\mu - igT^a A^a_\mu. \tag{1.6}$$

This change is necessary because otherwise the derivative of the transformed field would have extra terms coming from the local transformations parameters: $\partial_\mu U(x)\psi = U(x)\partial_\mu\psi - i\partial_\mu\theta^a(x)T^aU(x)\psi$.

A new field is present, so it needs to have a kinetic term. We build the field strength $F_{\mu\nu}\psi = i[D_\mu, D_\nu]\psi$. We will need to add to the previous Lagrangian (the one that was globally symmetric) a new term, called the Yang-Mills term:

$$\mathcal{L} \quad \rightarrow \quad \mathcal{L} - \frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a. \tag{1.7}$$

After making the derivatives covariant and adding the Yang-Mills term, the Lagrangian will be locally symmetric. Notice that in our treatment we have not pursued the usual pedagogical way of first introducing the abelian gauge symmetries and then the non-abelian ones. The case of abelian symmetries, that is symmetries for which we have $[T^a, T^b] = 0$, can be easily recovered from the general one.

Symmetries are a guiding principle in physics, but in practice they are often broken and the breaking leads to new, interesting effects. The breaking can occur at different levels. If the Lagrangian itself is not invariant, then the breaking is explicit. In this case we may still consider the Lagrangian as approximately symmetric and use perturbation theory.

It is also possible that the Lagrangian retains its symmetry completely, but the vacuum states are changed upon transforming. In this case, starting from a symmetry group $\mathcal{G}$, we can decompose in the invariant subgroup and the broken one:

$$U_{\mathcal{H}}\left|0\right\rangle = \left|0\right\rangle, \qquad U_{\mathcal{G}-\mathcal{H}}\left|0\right\rangle \neq \left|0\right\rangle. \tag{1.8}$$

What does this imply physically?

If the symmetry we are considering is a continuous, global one, then from the Goldstone theorem we know that dim($\mathcal{G}$)-dim($\mathcal{H}$) massless particles will appear in the theory [12, 13]. If in addition to the spontaneous breaking there is also a small explicit breaking of the symmetry, then the Goldstone bosons will acquire a mass and become pseudo Nambu-Goldstone bosons (pNGB). This is the case for instance of chiral QCD.

If the symmetry we are considering is a gauge symmetry, then things are different. This is because local symmetries have a different interpretation, they are actually not even symmetries, but redundancies in the description. As a consequence of this, it is always possible to recover the gauge symmetry after the spontaneous breaking. This is because the symmetry has just been hidden, it is now realized in a non-linear way. Although the symmetry has not actually disappeared, the consequences on the theory need to be investigated. Let us consider a scalar field ϕ whose expectation value is different from 0. This field is in some representation of $\mathcal{G}$ and its expectation value is invariant under $\mathcal{H}$. The kinetic term of the scalar field is

$$\frac{1}{2}D_\mu\phi D^\mu\phi = \frac{1}{2}(\partial_\mu\phi - igA^a_\mu T^a\phi)^2. \tag{1.9}$$

We are saying that ϕ has a non-invariant vacuum, this means that $U_\mathcal{G}\langle\phi\rangle \neq \langle\phi\rangle \neq 0$. Let us parametrize the vacuum as $\langle\phi\rangle = |\langle\phi\rangle|\langle\tilde{\phi}\rangle = v\langle\tilde{\phi}\rangle$, then the kinetic term of eq. (1.9) gives rise to

$$\frac{1}{2}g^2v^2(A^a_\mu T^a\langle\tilde{\phi}\rangle)^2 = \frac{1}{2}g^2v^2A^a_\mu A^{\mu,b}(T^a\langle\tilde{\phi}\rangle)(T^b\langle\tilde{\phi}\rangle). \tag{1.10}$$

This is clearly a mass term for the gauge bosons A^a_μ. How many gauge bosons will acquire a mass depends on the breaking pattern of the symmetry due to the non-zero vacuum, described by the term $(T^a\langle\tilde{\phi}\rangle)(T^b\langle\tilde{\phi}\rangle)$. This is what is known as the Higgs mechanism [14, 15]. No new particles appear, informally speaking people usually describe this mechanism by saying that the gauge bosons have eaten up the would-be Goldstone bosons.

Now we have all the elements needed if we want to construct a Lagrangian starting from symmetry arguments. Different models will be obtained by considering different symmetry groups and different representations of the fields under these symmetry groups. Once the Lagrangian of the theory is established, it is just a matter of calculations to derive the physical quantities. From the Lagrangian, the Feynman rules can be derived and from those the values of the scattering amplitudes follow straightforwardly (although the calculations might be very complicated).

The SM is, from this point of view, just one of the many theories that can be constructed. The value of the SM comes from the fact that it describes nature very precisely.

1.1.3 The SM Lagrangian

Given the great success of the SM, we now turn our attention to its description. We will describe the building of the SM Lagrangian by employing the construction described in the previous section.

Since we are using a construction based on symmetry, the first thing we need to specify is the symmetry group:

$$\mathcal{G}_{\text{SM}} = \text{SU}(3)_c \times \text{SU}(2)_L \times \text{U}(1)_Y. \tag{1.11}$$

We then need to choose our field content and decide in which representation of the group the fields are. In the case of the SM we have the ones reported in table 1.1.

	q_L	u_R	d_R	l_L	e_R	H
$\mathrm{SU}(3)_C$	3	3	3	1	1	1
$\mathrm{SU}(2)_L$	2	1	1	2	1	2
$\mathrm{U}(1)_Y$	1/6	2/3	$-1/3$	$-1/2$	-1	1/2
spin	1/2	1/2	1/2	1/2	1/2	0

Table 1.1: The representation of each field is specified. The spin representation is also made explicit, since it enters the Lagrangian construction. For what concerns $U(1)_Y$ the value of the hypercharge is reported.

Firstly, we construct a Lagrangian that is globally invariant under $\mathcal{G}_{\mathrm{SM}}$. In order to do so, we just need to build invariants using the fields. In table 1.1 we have all the information we need, we specified here the value of the spin as well, since it enters the construction. Each fermion type will actually be composed of three different generations, for instance $u_R = (u_R, c_R, t_R)$, but this does not influence the construction of the Lagrangian.

Hence, our globally invariant Lagrangian will be (we also need to make sure that it is Lorentz invariant):

$$\begin{aligned}\mathcal{L} = & \, i\bar{q}_L\gamma^\mu\partial_\mu q_L + i\bar{u}_R\gamma^\mu\partial_\mu u_R + i\bar{d}_R\gamma^\mu\partial_\mu d_R + i\bar{l}_L\gamma^\mu\partial_\mu l_L + i\bar{e}_R\gamma^\mu\partial_\mu e_R + \\ & + \partial_\mu H(\partial^\mu H)^\dagger + \mu^2|H|^2 - \lambda|H|^4 + \\ & - (y_u\bar{q}_L u_R\tilde{H} + y_d\bar{q}_L d_R H + y_l\bar{l}_L e_R H + h.c).\end{aligned} \tag{1.12}$$

where $\tilde{H} \equiv i\sigma_2 H^*$, σ_2 being a Pauli matrix. These are all the terms admitted at the renormalizable level, which for us means that there are no terms of dimension higher than 4. Of course, some index contraction is understood in order to make each term symmetric under the global group.

From the discussion in the previous section it should be clear that now we want to gauge $\mathcal{G}_{\mathrm{SM}}$. In doing so, we will make the derivatives covariant and add the Yang-Mills terms. For what concerns the field content, this means that we are adding several gauge fields. The gauging of each group leads to a number of gauge bosons equal to the dimensionality of the group, which for SU(N) is $N^2 - 1$. This implies that making $\mathrm{SU}(3)_c$ local leads to 8 gluons A^a_μ, from $\mathrm{SU}(2)_L$ we get 3 W^i_μ and from $\mathrm{U}(1)_Y$ we finally get the last boson B_μ. Up until now, we have kept generic group indices, but now we have specified the groups, so we have $i = 1, 2, 3$ and $a = 1, , 8$. We can generically denote the covariant derivative as

$$D_\mu = \partial_\mu - ig_s T^a_R A^a_\mu - igT^i_R W^i_\mu - ig' Y B_\mu, \tag{1.13}$$

here of course R indicates the representation of the field to which the derivative is applied and it is what distinguishes the various fields. For singlets we have $T_1 = 0$ and the field does not interact with the related gauge boson.

There will be also three Yang-Mills term, one for each symmetry group:

$$\mathcal{L}_{\mathrm{YM}} = -\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu,a} - \frac{1}{4}W^i_{\mu\nu}W^{\mu\nu,i} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \tag{1.14}$$

intuitively $G_{\mu\nu}$, $W_{\mu\nu}$ and $B_{\mu\nu}$ are the fields strength related to $\mathrm{SU}(3)_c$, $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ respectively. We now have all the pieces of the SM Lagrangian in the unbroken phase:

$$\begin{aligned}\mathcal{L}_{\mathrm{SM}} = & \, i\bar{q}_L\gamma^\mu D_\mu q_L + i\bar{u}_R\gamma^\mu D_\mu u_R + i\bar{d}_R\gamma^\mu D_\mu d_R + i\bar{l}_L\gamma^\mu D_\mu l_L + i\bar{e}_R\gamma^\mu D_\mu e_R + \\ & + D_\mu H (D^\mu H)^\dagger + \mu^2|H|^2 - \lambda|H|^4 + \\ & - (y_u \bar{q}_L u_R \tilde{H} + y_d \bar{q}_L d_R H + y_l \bar{l}_L e_R H + h.c.) + \\ & - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu,a} - \frac{1}{4} W^i_{\mu\nu} W^{\mu\nu,i} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}.\end{aligned} \tag{1.15}$$

This Lagrangian can tell us many things, but for now all the fermions are massless and the photon is nowhere to be found. A spontaneous breaking will occur: $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \to \mathrm{U}(1)_{em}$, leading to several particles acquiring a mass. This breaking is spontaneous and can be traced back to the Higgs potential $V_H = \mu^2|H|^2 - \lambda|H|^4$. The parameters here can be either positive or negative, but we fixed the sign in order to reflect the values derived experimentally. By minimizing we find that $-\mu^2|H|^2 + \lambda|H|^4$ has a minimum for $|H| = \sqrt{\frac{\mu^2}{2\lambda}}$, which is real if $\mu^2 > 0$. This vacuum expectation value (VEV) breaks the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry.

We parametrize the Higgs doublet in one possible way:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} h_2 + i h_1 \\ h_4 - i h_3 \end{pmatrix} \quad \xrightarrow{\langle |H| \rangle = v} \quad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ h + v \end{pmatrix}, \tag{1.16}$$

so, after the symmetry breaking, we only retain one scalar field h which is a singlet under both $\mathrm{SU}(3)_c$ and $\mathrm{U}(1)_{em}$ and is obtained by expanding around the vacuum.

The expectation value is now responsible for giving a mass to both the fermions and three of the gauge bosons. The mass of the fermions is easily derived from the third line of eq. (1.15). This becomes:

$$\begin{aligned} & -(y_u \bar{q}_L u_R \tilde{H} + y_d \bar{q}_L d_R H + y_l \bar{l}_L e_R H + h.c.) \quad \xrightarrow{\langle |H| \rangle = v} \\ & - \left(m_u \bar{u}_L u_R \left(1 + \frac{h}{v}\right) + m_d \bar{d}_L d_R \left(1 + \frac{h}{v}\right) + m_l \bar{e}_L e_R \left(1 + \frac{h}{v}\right) + h.c. \right). \end{aligned} \tag{1.17}$$

Here, since each fermion has three generations, the parameters $(m_u, m_d, m_l) = \frac{v}{\sqrt{2}}(y_u, y_d, y_l)$ are actually 3×3 matrices which can be diagonalized. The consequences of this are not trivial and we defer the discussion to section 1.2.3, where we discuss the hierarchy of the fermion masses in the SM.

The mass of the gauge bosons comes instead from the term $(D_\mu H)(D^\mu H)^\dagger$. This term gives also rise to interactions between the gauge bosons and the Higgs particle. After carrying out the calculations we get:

$$(D_\mu H)(D^\mu H)^\dagger \xrightarrow{\langle |H| \rangle = v} \quad \frac{1}{2}\partial_\mu h \partial^\mu h + \left(m_W^2 W^\mu_- W^+_\mu + \frac{1}{2} m_Z^2 Z_\mu Z^\mu \right) \left(1 + 2\frac{h}{v} + \frac{h^2}{v^2}\right), \tag{1.18}$$

here we have defined some new gauge bosons as combinations of the previous ones. As a matter of fact, given the symmetry breaking pattern, we expect one gauge boson to remain

massless, this is the photon. However, in the breaking $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \to \mathrm{U}(1)_{em}$, the unbroken subgroup is a combination of $\mathrm{U}(1)_Y$ and $\mathrm{SU}(2)_L$, so the massless gauge boson will be associated to a combination of B_μ and W^i_μ. After working out the algebra we get:

$$W^\pm_\mu = \frac{W^1_\mu \mp i W^2_\mu}{\sqrt{2}}, \qquad Z_\mu = \frac{g W^3_\mu - g' B_\mu}{\sqrt{g^2 + g'^2}}, \qquad A_\mu = \frac{g' W^3_\mu + g B_\mu}{\sqrt{g^2 + g'^2}} \tag{1.19}$$

$$m^2_W = \frac{g^2 v^2}{4}, \qquad m^2_Z = \frac{(g^2 + g'^2) v^2}{4}, \qquad m^2_A = 0 \tag{1.20}$$

At this point it would be convenient to rewrite the covariant derivatives in eq. (1.15) in terms of these gauge bosons. This inevitably introduces many interaction terms among gauge bosons and between gauge bosons and fermions. However, this is just a change of basis and a matter of rewriting interactions.

Although the final Lagrangian may look complicated and involve several terms, the construction was rather straightforward:

- Choose a symmetry group and the fields belonging to the theory. Construct a Lagrangian globally invariant under the group.
- Promote the symmetry from global to local. This will introduce a number of gauge bosons corresponding to the dimensionality of the gauged group. Make the derivatives covariant and add kinetic terms for the new gauge fields.
- Check if there is spontaneous symmetry breaking. A number of gauge bosons equal to the number of spontaneously broken directions will acquire a mass. The field acquiring the VEV will have to be expanded around its vacuum, thus changing the interactions.

Of course some subtleties have been overlooked, but we will focus more on them in the next part, where we will discuss some issues of the SM and the way these problems may be solved.

1.2 Problems of the standard model

The SM is in great agreement with experimental evidences and yet there are several clues that tell us we need to go beyond. These warning signs may be hints related to something unexpected happening in the SM or even strong evidences that the SM is not enough. The issue we will care about the most is the absence of a dark matter candidate. Since this is the main issue we are trying to address in this thesis, we dedicated chapter 2 to a deeper discussion. Briefly stated, the problem is that there are very strong indications suggesting the presence of non-visible matter and yet the SM does not provide a suitable candidate. Many models have been proposed in order to address this issue.

This section will be dedicated to an outline of different problems of the SM. Notice that these problems are not necessarily unrelated to the presence of DM. It is indeed possible to have a good DM candidate whose properties help solve other SM issues. So, in the following we will try to underline if any of the proposed solutions provide a possible DM candidate.

1.2.1 Naturalness

The first issue we are going to mention is the one that has the community most divided: naturalness. Naturalness has been a guiding principle in the development of particle models. The naturalness principle states that the couplings of a theory have a "natural" size. Dimensionless couplings should be of order 1, while couplings of dimension D should be of the order of the largest mass scale of reference to the power of D. The only way out is if we recover a symmetry by sending the coupling to 0, in that case the coupling can be arbitrarily small. We need to stress here that the naturalness principle is not a theorem, so theories do not need to comply to this characterization in order to be valid. That being said, when developing a model, discussing the naturalness of such a model is recommended.

Why is naturalness bad for the SM? The gauge couplings and the leading couplings between the Higgs and the fermions are satisfactorily natural. However, the parameter μ^2 is not natural, for that we have $\mu^2 = m_h^2/2 \sim (100\,\text{GeV})^2$. This would need to be compared to the largest mass scale of the theory, which in this case is the Planck mass $m_{\text{Pl}} \simeq 1.2 \times 10^{19}\,\text{GeV}$, so about 17 orders of magnitude larger. So, if a physicists adheres to the naturalness principle, this is an issue. Many solutions have been proposed, mainly introducing new physics at an energy much lower than the Plank mass, so that we no longer need to compare m_h with m_{Pl}, but rather with an energy scale of the order of the TeV.

Other parameters of the Standard Model have some unnaturalness, mainly the couplings of first and second generation fermions to the Higgs, this also creates a hierarchy between the fermion masses. This topic is briefly discussed in section 1.2.3.

1.2.2 Hierarchy (Higgs)

A problem related to naturalness, but of a different nature, is the hierarchy of the Higgs mass. The biggest difference with respect to naturalness is that while the lack of naturalness does not forbid a model, the hierarchy problem deeply affects the plausibility of a theory.

This issue in the SM is related to the loop corrections of the Higgs mass. At tree level we know that we will have $m_h = \sqrt{2\mu^2} \simeq 125\,\text{GeV}$, which is consistent with the experimentally measured value. But we know from renormalization theory that we need to consider also loop correction to the mass, these contributions come from top-loops, self-interactions and gauge loops as depicted in fig. 1.1. After carrying out the calculations we get

$$m_h^{2(1L)} \simeq 2\mu^2 + \frac{\Lambda^2}{32\pi^2}\left(-y_t^2 + +6\lambda + \frac{9}{4}g^2 + \frac{3}{4}g'^2\right), \tag{1.21}$$

where Λ is a hard cut-off we impose to the momentum circulating in the loop. This energy scale Λ is supposed to be the scale at which the SM description breaks down. If the SM was indeed the final theory (except for gravity), this description would break down at $\Lambda = m_{\text{Pl}}$. This implies that the corrections to the mass might be several orders of magnitude larger than the tree-level. It is in principle possible for the higher order contributions to cancel each other. This is certainly a possibility, although it would require cancellations of more than 20 digits, so that would be quite the coincidence.

Many solutions have been proposed: introduce a mechanism to protect the Higgs mass (SUSY [16]), make the Higgs a non-elementary particle (composite models [17]), dilute the Planck scale through extra dimensions ([18], [19]). SUSY [20] and composite Higgs models

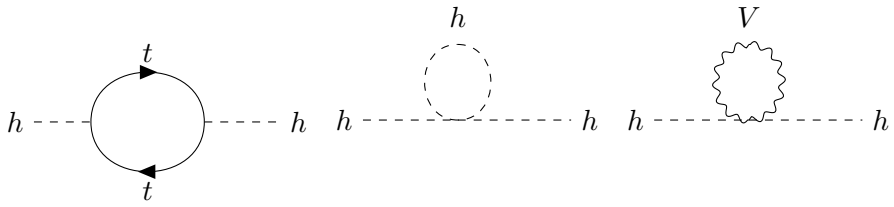

Figure 1.1: Diagrams contributing to the 1-loop Higgs mass. Sub-leading contrbutions come from loops with other fermions beside the top.

[21] in particular may introduce additional particles which are also good DM candidates. A composite Higgs model providing a good DM candidate is studied in detail by us in chapter 3.

1.2.3 Hierarchy (fermion masses)

The hierarchy problem enters also in the fermion masses. In the building of the SM we overlooked how to derive the fermion masses, so let us focus on that here.

After spontaneous breaking of $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$, we ended up with the following mass term for the fermions:

$$-(m^u_{ij}\bar{u}^i_L u^j_R + m^d_{ij}\bar{d}^i_L d^j_R + m^l_{ij}\bar{e}^i_L e^j_R + h.c.),$$

where we specified now the indices $i, j = 1, 2, 3$ which run over the fermion generations. There are a priori several free parameters here. However, things simplify by considering that, except for this part of the Lagrangian, everything is globally symmetric under $\mathrm{U}(3)^5 \times \mathrm{U}(1)$. The reason for this is that we can rotate all the 5 fermion types and the Higgs without changing the Lagrangian. This allows some freedom in redefining and diagonalizing the mass matrices, we can reduce to the case where:

$$m_u \to m_u^{\mathrm{diag}}, \quad m_d \to m_d^{\mathrm{diag}} V^\dagger_{\mathrm{CKM}}, \quad m_l \to m_l^{\mathrm{diag}}, \tag{1.22}$$

so that we have 9 mass parameters and then the parameters entering the CKM matrix. The information encoded in the CKM matrix is related to the mixing between families and CP violation.

Where does the hierarchy enter? We have now 9 mass parameters which a priori should be around the same order of magnitude, since they are basically the product of an order 1 coupling and the VEV of the SM. It actually turns out that these masses go from $m_t \simeq 170\,\mathrm{GeV}$ to $m_e \simeq 5 \times 10^{-4}\,\mathrm{GeV}$, spanning 6 orders of magnitudes. And this is without considering the neutrinos. Solutions to this problem can be found in the so called $3-3-1$ models [22, 23]. These models may also include the presence of dark matter candidates [24].

1.2.4 Neutrino masses

In the previous subsection we treated the mass generation in more detail and still three particles remain massless: the neutrinos. This is in tension with experiments, since we know that neutrinos oscillate [25, 26], but this is possible only if they have a mass. Experimentally, we know not only that neutrino have masses, but also the relative mass differences between

them [27] and a bound on the sum of their masses: $\sum_i m_{\nu_i} < 0.12\,\text{eV}$ [28]. The SM Lagrangian lacks a term capable of giving mass to the neutrinos.

A possible way out is introducing higher dimensional operators. In the construction of the SM Lagrangian renormalizability was a requirement, but we might let this requirement go. Already at dimension 5 it is possible to introduce the Weinberg operator [29], which can generate the mass of neutrinos:

$$\mathcal{L}_W = \frac{y_{ij}}{\Lambda}(\tilde{H} l_L{}^i)(H \bar{l}_L{}^j), \tag{1.23}$$

here Λ is again a new physical scale. In this case it would correspond to the mass scale where this operator is generated. We know that the masses of the neutrinos are much lower than the masses of the other leptons and a nice feature of introducing a dimension 5 operator is that it gives a natural suppression to the mass of the neutrinos of order v/Λ.

A possible way of generating this operator is through the see-saw mechanism [30]. In this case, a right-handed neutrino with a large mass M_R is introduced. The generated mass for the SM neutrinos is of order v^2/M_R.

See-saw and Weinberg operators are the standard way in which this problem is solved, but it needs to be mentioned that scotogenic models are popular solutions that can accommodate both a viable DM candidate and the generation of neutrino masses [31].

1.2.5 Strong CP

Yet another issue of the SM is related to the size of a parameter. When we constructed the SM Lagrangian, we said that we wanted to put all the renormalizable gauge invariant terms. Well, we actually omitted one:

$$\mathcal{L}_{CP} = -\frac{\theta_{CP}}{32\pi^2}\epsilon^{\mu\nu\rho\sigma} G_{\mu\nu} G_{\rho\sigma}, \tag{1.24}$$

where $\epsilon^{\mu\nu\rho\sigma}$ is the anti-symmetric Levi-Civita symbol. We point out that the θ_{CP} we are mentioning here is the physical one, so that the ambiguity due to the possible phases of the quark masses has already been taken away. This Lagrangian term break CP, but there is already CP breaking in CKM matrix of the SM, so there is no reason why this term should not be present as well.

This term might actually be present, but θ_{CP} should be in this case very small, as we have not seen any effect of its presence yet. A bound on this parameter comes from measuring the neutron electric dipole moment [32] and so we know that $\theta_{CP} < 10^{-9}$. This small value clearly should be justified, a popular solution to this problem is the presence of a new particle: the axion [33, 34]. Both the axion itself or particles with similar properties (the so-called axion-like particles) can be good DM candidates.

1.2.6 Other issues of the SM

This is not a complete list of the SM problems, many others are present and we briefly mention the most interesting ones.

Other experimental problems of the SM include the necessity to integrate gravity in the description. For now the Standard Model and gravity do not talk to each other. Another problem is the lack of an explanation of the baryon anti-baryon asymmetry. Here, some

dark matter candidates (asymmetric dark matter) have been proposed as having origin from this asymmetry [35]. Finally, another quantity does not have the expected magnitude: the vacuum energy of the SM. This is also known as the cosmological constant problem: we know from cosmology the expected value of this parameter and we also know that in the SM this quantity is originated by the vacuum of the theory. Unfortunately, the value derived from cosmology and the one calculated in the SM differ by around a hundred orders of magnitude.

All these issues deserve study and discussion, but our main object of study is dark matter. This is why we focused our attention on SM problems that can be solved by introducing a dark matter candidate. The next chapter will be fully dedicated to the discussion of dark matter: why do we believe it exists, which properties should it have and how do we find it?

Chapter 2

Dark matter and how to find it

We have stated that the lack of a dark matter candidate in the SM is a problem that should be addressed. But are we sure there is dark matter in the universe? And which properties should a good dark matter candidate have?

These are all legitimate questions that we will try to answer in this chapter. Firstly, we will list all the evidences in support of the existence of DM. It will be clear that the amount of evidences is stunning. Then, since we have so many evidences for the presence of dark matter, we also have some information about the properties a good DM candidate should have. Many solutions have been proposed and we will outline the best motivated ones. Finally, we would like to check whether a specific theory of dark matter is consistent with experiments, so we will talk about what people are doing in order to detect and study dark matter.

2.1 Evidences of the presence of dark matter

An important first step when addressing a problem is understanding if the problem is there at all. The evidences for the presence of dark matter are so numerous and so different between each other that it is hard not to believe in its existence. However, alternatives have also been proposed [36]. The focus of these solutions is on the laws used to describe the universe, rather than on its content. The idea is that Newtonian mechanics breaks down at very large scales and needs to be modified, hence the name modified Newtonian dynamics (MOND). The problem is that these solutions may be consistent with some of the evidences, but not with all of them. Nevertheless, these theories have their own interest. Discussing them is outside the scope of this thesis, so in the following we will rather talk about why we do believe there is indeed dark matter out there.

2.1.1 First hints and galaxy rotation curves

The fact that there was something weird going on with the galaxies started to appear clear from the beginning of the 20th century. In the same years the idea that there might be unseen matter in the universe started to become popular. The first links between the strange galaxy behavior and the idea that there might be non-luminous matter in the universe are due to Kapteyn (1922, [37]), Oort (1932, [38]) and Zwicky (1933, [39]). Most notably, Zwicky's paper is usually considered the reference paper for the beginning of the dark matter era.

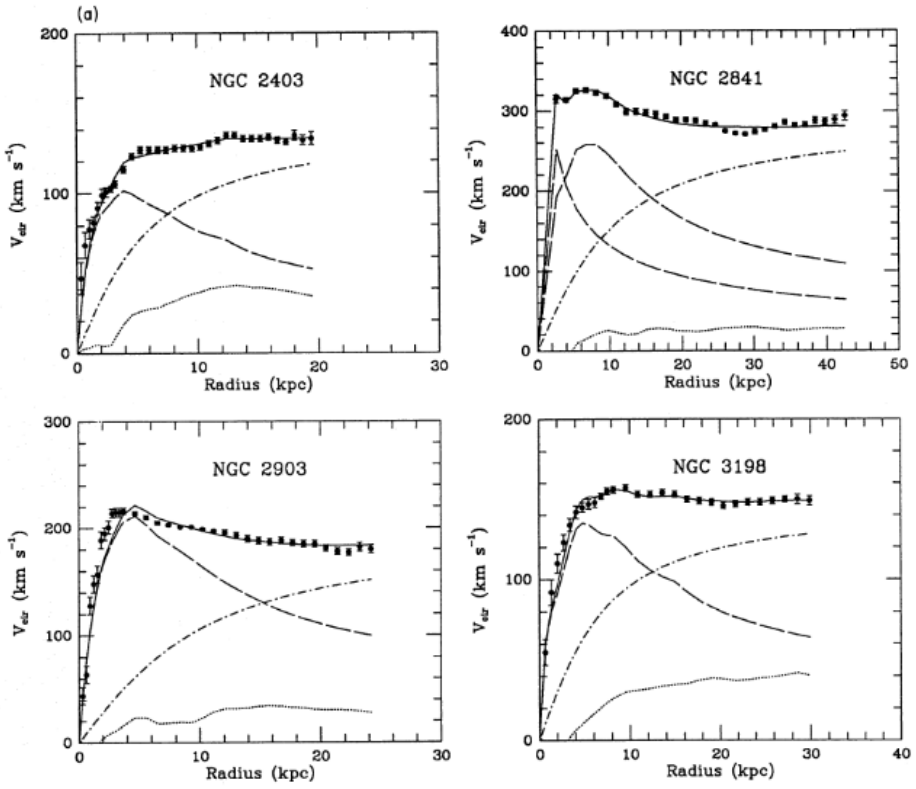

Figure 2.1: Velocity profile of some sample galaxies. The dotted and dashed lines refer to the gas and visible component respectively. The dot-dashed curve is the reconstructed dark matter profile. Plots from Ref. [40].

What Zwicky did was using the virial theorem in order to relate the kinetic energy and the potential energy of galaxies. After carrying out the calculation, the velocity profile at a distance r from the center of the galaxy was

$$v(r) \propto \sqrt{\frac{m(r)}{r}}, \tag{2.1}$$

here $m(r)$ indicates the mass enclosed within a radius r. The luminous matter is distributed with a certain average density $\bar{\rho}$ within a radius R. This implies two regimes:

$$v(r) \propto \sqrt{\frac{m(r)}{r}} \overset{r \leq R}{=} \sqrt{\frac{4}{3}\frac{\bar{\rho} r^3}{r}} \propto r, \qquad v(r) \propto \sqrt{\frac{m(r)}{r}} \overset{r \geq R}{=} \sqrt{\frac{M}{r}} \propto r^{-1/2}. \tag{2.2}$$

However, this was in disagreement with the experimental results. We report in fig. 2.1 the measured profile for some galaxies as an illustrative example. The behavior for large values of r is not the one expected from virialization. The velocity does not seem to scale as the inverse of the square root of r for large r, but seems rather constant. A constant velocity implies a mass linearly proportional to r, which in turn implies $\rho(r) \propto r^{-2}$.

So already at this point some information about the predicted density profile was available.

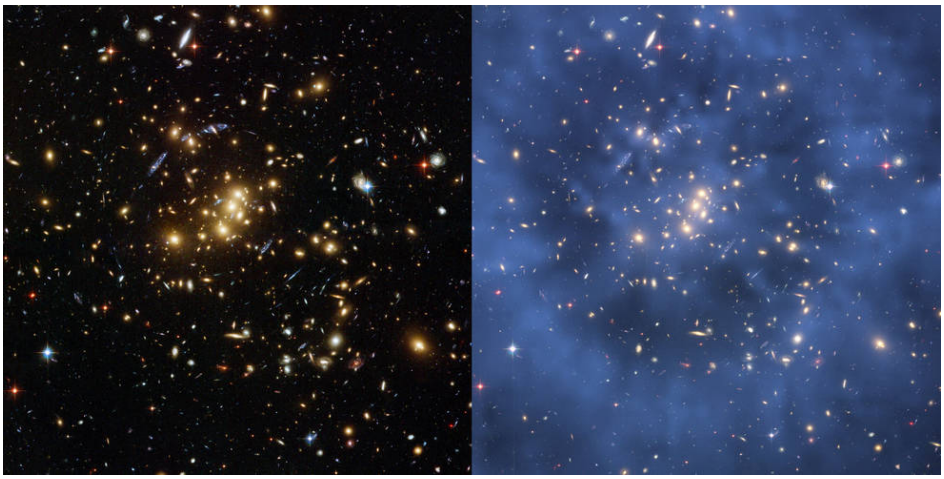

Figure 2.2: *On the left:* View from Hubble of the galaxy cluster Cl 0024+17 (ZwCl 0024+1652). Images of objects behind the cluster are distorted as the light is bent by gravtiational lensing. *On the right:* the reconstructed dark matter mass profile is presented in blue. **Credits:** NASA, ESA, M.J. Jee and H. Ford (Johns Hopkins University).

And not only that. Zwicky was able to relate the mass of the luminous and non-luminous part and conclude that the dark matter should be more than the visible matter.

2.1.2 Gravitational lensing

From general relativity we know that even light interacts gravitationally and can be influenced by mass distributions. Because of this, lensing has become a very powerful tool to describe the universe. Light from distant stars arrives on Earth and can be observed. However, the path of the ray of light on its way to Earth is distorted by the massive objects found along the way. Depending on the mass of the objects encountered and the relative position of the ray of light and the objects, the intensity of this effect can vary. So we go from microlensing, for the weakest perturbation, to strong lensing for total distortion of the reconstructed images. The result we get from observing the light arriving on Earth is that the amount of distortion is not compatible with interaction with visible matter only. In fig. 2.2 we see both the images distorted by the gravitational effect (on the left) and the reconstructed dark matter profile (on the right). Lensing can be used then to trace the dark matter mass distribution in the galaxies [41].

2.1.3 Bullet cluster

Lensing can be used to describe the gravitational potential of galactic objects. This was used in order to study the gravitational potential of two galactic clusters before and after the collision [42]. An illustrative reconstruction after collision is shown in fig. 2.3: the galaxy clusters seem to be made up of two components behaving very differently. One component is the visible one, this interacts strongly and it is deeply affected and slowed down by the cluster collision. The other component, which could be dark matter, proceeds almost unaltered after the collision. This is consistent with the fact that we expect dark matter to interact very

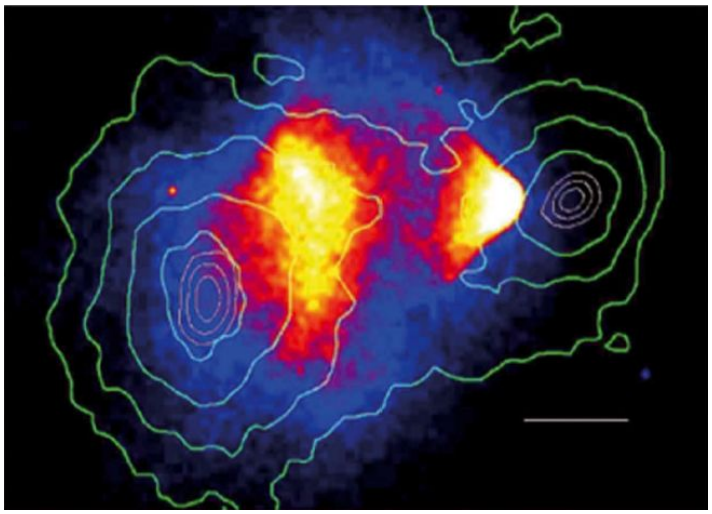

Figure 2.3: Bullet cluster 1E 0657-558 and measured properties. Iso-potential lines reconstructed from lensing are shown in green. The abundance of visible matter is shown through a temperature map [43].

weakly. The potential reconstructed through lensing after the collision could not be explained by considering visible matter alone. This evidence is particularly important as it is completely consistent with a dark matter interpretation, but not with a MOND explanation.

2.1.4 Large scale structure

From perturbation theory we know how structures in the universe should develop. However, the way the universe looks now, especially the large scale structure (LSS), is not the predicted one. The easiest way to understand this is considering the expected time for the beginning of structure formation. Visible matter interacts significantly with light, this means that collapse into structure should happen only after matter-radiation equality. The study of structure formation after that epoch tells us that there has not been enough time for the present structures to develop. A possible solution is given by dark matter. Since it does not interact with light, DM would start collapsing before visible matter, thus allowing for the current LSS [44]. Not only that, but if indeed structure formation is affected by dark matter, we also know that dark matter should be mainly composed by a cold (non-relativistic) component. Warm and hot (relativistic) components, if present, should be sub-dominant.

The prediction for the LSS with dark matter is in good agreement to what is experimentally found. Two well-known problems are present if we consider the small scale structure [45]. One is the fact that few dwarf galaxies are seen, while many are predicted, this is the "missing satellite problem". The other issue is that the predicted density profile should have a peak in the center of the galaxy according to simulations, while a smoother profile is favored by DM observations.

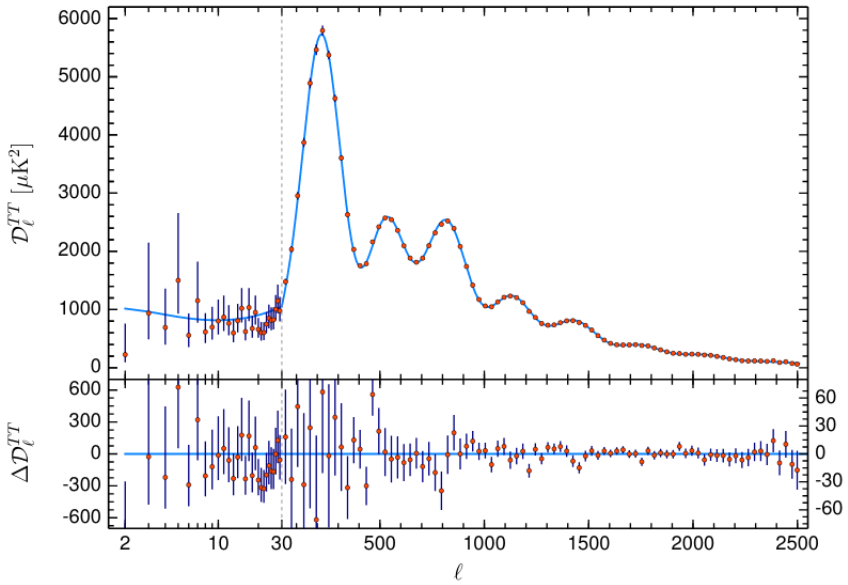

Figure 2.4: Temperature power spectrum from Planck 2018. The error bars show the 1σ uncertainties. A different behavior is displayed above and below $l = 30$.

2.1.5 Cosmological evidences

Finally, we turn to the largest scale accessible: cosmological scales. We can infer information about the history of the universe from the cosmic microwave background (CMB) observations. What we get from measurements is a spectrum made of peaks and valleys as shown in fig. 2.4. By analyzing their widths and heights, cosmologists understand the evolution of the universe. In particular, similarly to what we said about the large scale structure, we can derive information about the interaction of matter with photons. In the radiation dominated epoch all visible matter oscillates together with photons and this should be seen in the CMB spectrum. The conclusion is that if we want to reproduce this spectrum, then there must be non-baryonic matter. Cosmological observations are even capable of deriving how much dark matter should be present in the universe [28]:

$$\Omega_c h^2 = 0.1198 \pm 0.012. \tag{2.3}$$

This value comes from a best fit of the CMB observations and lensing results. We will use this experimental value throughout this work to constrain the parameter space of our models.

2.2 Dark matter candidates

From the previous section we understood that dark matter does seem to be present in the universe. Physicists have been developing dark matter models for tens of years now and the number of available candidates is increasing every day. The history of dark matter model building has actually seen several shifts in the focus. Weakly interacting massive particles (WIMP) have been a favored candidate for a long time, but recently they have been losing

some interest in favor of other theories. No matter which is your preferred dark matter model, it will have to adhere to some requirements and respect experimental limits.

2.2.1 Required properties

Since we have so many evidences for the presence of dark matter, we also know some things about the properties it should have. These requirements must be taken into account when developing a dark matter model.

- **Relic abundance**

  From cosmological observations of the CMB, baryon acoustic oscillations and lensing we can derive the abundance of dark matter in the universe. We already reported a best fit value in eq. (2.3). Since this number is known to very good precision, reproducing this result is usually the first step to see whether a dark matter model works. This requirement can be somewhat loosened since there might be more dark matter components, so we might propose a candidate which does not reproduce all the abundance, but just a part. Not producing more dark matter than what is observed remains a strict requirement, as that would lead to over-closing the universe.

- **Lifetime**

  We know how much dark matter there is today in the universe, this means that it did not decay up until now. This tells us something about the stability of the dark matter. The exact time at which the dark matter is produced is model dependent, but we are talking about the early universe in any case. Thus, we require the dark matter to either be completely stable or at least to have a lifetime comparable to the current time of the universe:

  $$\tau_{\mathrm{DM}} \geq t_{\mathrm{U}} \sim H_0^{-1}. \tag{2.4}$$

  Here H_0 is the Hubble constant and its inverse can be used to derive an order of magnitude estimate for how long the universe has existed.

- **Dark**

  Dark matter is indeed dark. By this we mean that it should interact at most very weakly with visible matter and light. In particular, it should not be in equilibrium with photons in the radiation dominated epoch.

- **Collisionless**

  This is related to the previous property. Dark matter should interact mainly gravitationally, so in a very weak way with other matter. Hence, when dark matter collides with other matter in the universe it should not be affected by the collision. This is clearly highlighted by the bullet cluster. Observations on the bullet cluster in particular allow to set a limit on the self-interactions:

  $$\sigma_{\text{DM-DM}} < 1.8\,\mathrm{b}\,\mathrm{GeV}^{-1}. \tag{2.5}$$

- **Cold (or even a little bit warm)**

  From structure formation we know that cold dark matter is favored. The size of the objects formed is set by the free streaming length, which is the distance traveled between creation and radiation-matter equality:

$$\lambda_{\mathrm{FS}} = \int_{t_i}^{t_{\mathrm{eq}}} \frac{v(t)}{a(t)}\,\mathrm{d}t, \tag{2.6}$$

  here $v(t)$ and $a(t)$ are the velocity and universe expansion parameter at time t. Hot dark matter would lead to top-down structure formation, while cold dark matter would imply the favored bottom-up formation. Warm dark matter is not completely excluded by this result.

We know also other properties, for instance we know that dark matter behaves like a fluid and this puts strong constraints on primordial black holes as a candidate. We see that quantum properties of DM are not relevant at galactic scale: this sets lower bounds for the mass of the DM particle. More precisely, we have that for a fermion $m_{\mathrm{DM}} \gtrsim 100\,\mathrm{eV}$ and for a boson $m_{\mathrm{DM}} \gtrsim 1 \times 10^{-22}\,\mathrm{eV}$.

2.2.2 WIMPs and their thermal history

Many proposed candidates are consistent with these requirements. We now turn to describe the thermal history of WIMPs. The two candidates we studied in detail in this thesis belong indeed to this class. It is important to understand why we focused on them and this will be clear after we talk about the "WIMP miracle".

First of all, we add a new particle χ to the standard model. We want to be as general as possible here, so we try to treat everything independently from the specific properties of χ. In the early universe everything is extremely hot and the particle χ is in equilibrium with the SM, so we can trace its abundance. This equilibrium is maintained by processes of the type:

$$\chi\chi \leftrightarrow \mathrm{SM}\,\mathrm{SM}. \tag{2.7}$$

These interactions will however become inefficient as the temperature drops, as a consequence of the expansion of the universe. In order to track the density after decoupling, we need to understand when the interactions become inefficient. This can be accurately described by Boltzmann equations [46].

The object we want to study is the phase space distribution $f_\chi(x^\mu, p^\mu)$ of the DM particle. Its evolution is given by the Liouville equation:

$$\hat{L}[f] = \hat{C}[f], \tag{2.8}$$

the LHS gives the evolution of f from the Liouville operator $\hat{L} = \frac{\mathrm{d}}{\mathrm{d}t}$. This is a total derivative which can be decomposed in the various contributions and these contributions can be simplified by assuming a specific cosmological model, in our case we assume a FRW universe. The RHS includes instead the collisional operator $\hat{C}$. This encodes information about all the possible interactions of the particle χ. We now integrate over the phase space Π_χ of the DM

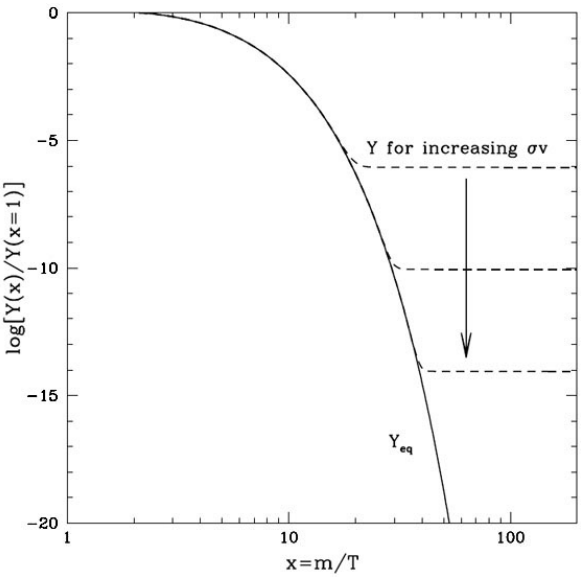

Figure 2.5: Time evolution of the relic density $Y(x)$. Different values at current time are obtained by varying the thermally averaged cross section. Image from [46].

particle both the LHS and the RHS and we obtain the number density, which is by definition $n_\chi = \int f_\chi(x^\mu, p^\mu)\mathrm{d}\Pi_\chi$. We now have:

$$\frac{\mathrm{d}n_\chi}{\mathrm{d}t} + 3Hn_\chi = \int \hat{C}[f_\chi]\mathrm{d}\Pi_\chi. \tag{2.9}$$

The RHS can be simplified under the assumption of kinetic and chemical equilibrium. This assumption is related to the relative rates of different processes involving χ. If chemical equilibrium failed, we would have to solve a system of coupled Boltzmann equations. In this simple picture, the final result is

$$\frac{\mathrm{d}n_\chi}{\mathrm{d}t} + 3Hn_\chi = -\langle\sigma v\rangle(n_\chi^2 - n_{\chi,eq}^2), \tag{2.10}$$

where $\langle\sigma v\rangle$ is the thermally averaged cross section for the scattering $\chi\chi \leftrightarrow \mathrm{SM\,SM}$ and $n_{\chi,eq}$ is the number density if there were no interactions. Its explicit value is:

$$n_{\chi,eq} = \frac{T}{2\pi^2} g_\chi m_\chi^2 K_2\left(\frac{m_\chi}{T}\right), \tag{2.11}$$

this only depends on a few properties of the DM candidate, namely its degrees of freedom g_χ and its mass m_χ. $K_2(x)$ is the modified Bessel function of the second kind.

The expression actually simplifies even more if we consider, as it is customary in this study, the new quantities $Y \equiv n/s$ and $x \equiv m_\chi/T$. Here s is the entropy of the universe and x can be used to track the time evolution in place of time or temperature (as x increases, time increases and temperature decreases). The final equation, which is the one used for practical purposes, is:

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = -\sqrt{\frac{\pi g_*}{45G}}\frac{m}{x^2}\langle\sigma v\rangle(Y^2 - Y_{eq}^2). \tag{2.12}$$

This equation can be solved for varying values of the cross section and we would get the behavior in fig. 2.5. From this figure we can derive some useful insight. First of all, the relic abundance present today is strongly related to the decoupling time, that is the time where the interactions with the SM stop being effective and the density is not following the equilibrium density anymore. This also explains why larger cross sections lead to a lower abundance today. If the cross section is very large, then interactions are effective for a longer time and the decoupling happens later. Since here the DM is non-relativistic, $Y_{eq}(x) \propto \exp(-x)$ and it is suppressed at later times.

The time of freeze-out is the one at which interactions are no longer effective. This is quantitatively stated by defining the freeze-out time as the one at which the rate of interactions $\Gamma(x) = n(x)\langle\sigma v\rangle_x$ is equal to the Hubble parameter $H(x)$, which in turn describes the dilution of the universe as a function of x. We have

$$n(x_f)\langle\sigma v\rangle_{x_f} = H(x_f), \tag{2.13}$$

this usually implies $x_f = 25 + O(\text{logarithmic corrections})$. We then track the evolution of $Y(x)$ from x_f to today to derive the current abundance:

$$Y_\infty = \left(\frac{1}{Y(x_f)} + \int_{x_f}^{+\infty} \sqrt{\frac{\pi g_*}{45G}}\frac{m_\chi}{x^2}\langle\sigma v\rangle_x \mathrm{d}x\right)^{-1}. \tag{2.14}$$

Here it is clear that the final result will depend largely on the cross section. Finally, the abundance today is related to Y_∞ just by some constants:

$$\Omega_\chi h^2 = \frac{\rho_\chi}{\rho_c}h^2 = \frac{Y_\infty m_\chi}{3.5\,\mathrm{eV}}. \tag{2.15}$$

The mass of the particle appears in both the numerator and the denominator, so the explicit dependence from this parameter is not strong. Here $\rho_\chi \equiv n_\chi/m_\chi$, ρ_c is the critical density and $H = H_0/100$.

Then, the question naturally arises: how large should the cross section be in order to reproduce the dark matter abundance today? The magic number that is usually mentioned is $\langle\sigma v\rangle \simeq 1\,\mathrm{pb} \simeq 3\times10^{-26}\,\mathrm{cm^3 s^{-1}}$.

An order of magnitude estimate for an s-wave cross section would give:

$$\langle\sigma v\rangle = k\frac{\alpha'^2}{m_\chi^2}, \tag{2.16}$$

here α' is the coupling strength relevant for the interaction and k is an order 1 constant. If we say that $\alpha' \sim \alpha_w$, then we derive that $m_\chi \sim 100\,\mathrm{GeV} - 1\,\mathrm{TeV}$. This means that the correct relic density is reproduced for couplings of the order of the weak couplings and masses of the order of v: this is the "WIMP miracle".

The previous discussion was rather straightforward, but it relied on some simplifying assumption and currently many models where some of these assumption fail have been and are being studied [47, 48, 49]. However, the fact that the relic calculation needs to be changed if an assumption fails is nothing new and it was already noted in Ref. [50]. Here we briefly discuss each of the three exceptions, focusing the most on co-annihilations, as they will be relevant in our models.

- **Forbidden channel**

The channels considered in the averaged cross section of eq. (2.12) are the ones allowed when χ is at rest. This is usually not a problem, since at freeze-out we have $T \simeq m_\chi/25$ and so the kinetic energy is much lower than the rest mass. However, it might be possible to open up channels with this small kinetic energy. The requirement in order to open up the channel $\chi\chi \to \sum_{j=1}^{N} \phi_j$ is

$$2m_\chi < \sum_{j=1}^{N} m(\phi_j), \quad 2\sqrt{m_\chi^2 + p_\chi^2} > \sum_{j=1}^{N} m(\phi_j). \tag{2.17}$$

If this is the case, we need to add neglected contributions to eq. (2.12).

- **Near threshold resonance**

The interaction between dark matter and standard model can be mediated by a particle S of finite particle width Γ_S. In this case, the scattering cross section gets the Breit-Wigner form:

$$\langle \sigma v \rangle \propto \frac{\Gamma(S \to \chi\chi)\Gamma(S \to \text{SM SM})}{(s - M_S^2)^2 + M_S^2\Gamma_S^2}. \tag{2.18}$$

s here is the Mandelstam variable corresponding to the scattering energy in the center of mass (CM) frame, so $s \gtrsim 4m_\chi^2$. If we have that $M_S \simeq 2m_\chi$, then we have a resonance as the first term in the denominator becomes very small and the scattering cross section can be increased substantially. This situation is particularly important for models where the Higgs particle can be the mediator and we are in a situation where $m_\chi \simeq m_h/2$.

- **Coannihilations**

This is the case that is most interesting to us, as it will be central in chapter 4.

It is often the case that a DM model does not introduce a single new particle χ, but also other companion particles called dark matter partners. We then denote the class of new particles with χ_i and the DM candidate will be the lightest stable one, which we call χ_1. In this case not only the processes of the type $\chi_i\chi_i \leftrightarrow$ SM SM maintain the equilibrium between the SM and a particle χ_i, but also scatterings of the type $\chi_i\chi_j \leftrightarrow$ SM SM.

The evolution of the densities n_i is hence described by a system of coupled Boltzmann equations:

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} + 3Hn_i = -\sum_j^N \langle \sigma_{ij} v \rangle \left(n_i n_j - n_i^{eq} n_j^{eq}\right). \tag{2.19}$$

Things simplify considerably if we consider that we do not actually care about the evolution of every single density. The heavier particles will decay into the DM candidate χ_1, this means that we can safely talk about the total density only: $n_{\mathrm{DM}} = \sum_i n_i$.

Now the N equations reduce to just one:

$$\frac{\mathrm{d}n_{\mathrm{DM}}}{\mathrm{d}t} + 3Hn_{\mathrm{DM}} = -\langle\sigma_{\mathrm{eff}}v\rangle\,(n_{\mathrm{DM}}^2 - n_{eq}^2), \tag{2.20}$$

where the effective cross section is defined as:

$$\sigma_{\mathrm{eff}} \equiv \sum_{ij}\sigma_{ij}\frac{g_i g_j}{g_{\mathrm{eff}}^2}(1+\Delta_i)^{\frac{3}{2}}(1+\Delta_j)^{\frac{3}{2}}e^{-x(\Delta_i+\Delta_j)}, \qquad \text{with} \qquad \Delta_i \equiv \frac{m_i - m_1}{m_1}. \tag{2.21}$$

We can see that the weight with which each cross section enters is defined by the relative mass splitting. In particular, if we have $m_i + m_j - 2m_1 \geq T$, the contribution of the cross section σ_{ij} is suppressed.

In chapter 4 we have only two particles: χ_1 of mass m_1 and χ_2 of mass $m_1 + \Delta m$. The equations become:

$$\sigma_{\mathrm{eff}} = \frac{1}{(1+w)^2}\sigma_{11} + \frac{2w}{(1+w)^2}\sigma_{12} + \frac{w^2}{(1+w)^2}\sigma_{22} \qquad \text{with} \qquad w \equiv \frac{g_2}{g_1}\left(1+\frac{\Delta m}{m_1}\right)^{\frac{3}{2}} e^{-x\frac{\Delta m}{m_1}}. \tag{2.22}$$

2.2.3 Other candidates

Even though the WIMP paradigm looks very appealing, it is experiencing a slight decrease in popularity. One of the reasons for this is that, due to the scale of interactions and masses, it should not be difficult to see a WIMP particle. However, there is no clear evidence for a WIMP particle in any of the detection experiments (which we will discuss in section 2.3). This means that the "reasonable" parameter space for WIMPs is getting more and more constrained. There are anyway several non-WIMP candidates and modifications of the WIMP paradigm. In the following, we will describe in a little more detail the most popular and well-motivated alternative DM candidates.

Axions

Axions are a very intriguing candidate, as they allow to solve the strong CP problem and provide a good DM particle. The strong CP problem was discussed by us in section 1.2.5.

The idea is to deal with the smallness of θ_{CP} by saying that it corresponds to the vacuum of a new field $a(x)$, deriving from the breaking of a $\mathrm{U}(1)_{\mathrm{PQ}}$ symmetry. This idea came from Peccei and Quinn [33], who also give the name to the broken symmetry. With the introduction of this field the Lagrangian term becomes

$$\mathcal{L}_{CP} = -\frac{1}{32\pi^2}\frac{a(x)}{f_a}\epsilon^{\mu\nu\rho\sigma}G_{\mu\nu}G_{\rho\sigma}, \tag{2.23}$$

here f_a is the axion decay constant.

The mass of the axion depends on the specific model, but up to corrections it is predicted to be [51]:

$$m_a = 6 \times 10^{-6}\,\text{eV}\left(\frac{10^{12}\,\text{GeV}}{f_a}\right), \tag{2.24}$$

this is very useful if you want to relate limits on f_a to bounds on m_a.

The axion is expected to be able to decay into a pair of photons, the coupling with the photons depends on the specific axion model. This decay means that the axion is not stable, so we need to guarantee that its decay time is larger than the time scale of the universe. This implies an upper bound on the mass of the axion of the order of the eV. Stronger constraints come from astrophysical observation of supernovae and star cooling, in this case the upper bound on the mass become of the order of the meV [52].

We also would like to provide a good DM candidate, but with these constraints on the mass it is not possible to reproduce the relic density with a thermal evolution such as the one of the WIMPs. Anyway, there are non-thermal axion production scenarios which are able to reproduce the current DM relic density. The axion can for instance oscillate around the initial value of θ_{CP} and then stabilize when $m_a(T) = H(T)$ [53].

Sterile neutrinos

Another candidate, which is interesting if we want to solve more SM problems at once, are right-handed neutrinos. As we outlined in section 1.2.4, we have no way of giving mass to the neutrinos in the SM, but if we added heavy sterile neutrinos we could give them mass via the seesaw mechanism.

In order to make the neutrino masses as light as experimentally observed, we would need $M_R \sim 10^{12}\,\text{GeV}$. This large mass is not compatible with the reproduction of the correct DM relic density as we would need $M_R \sim \text{keV}$. This problem can be solved by having some hierarchy in the couplings giving mass to the neutrinos. As a matter of fact, we do not need all three masses to be very large and it is possible to have one which is much lower than the others. This would help provide a good DM candidate, at the price of introducing some hierarchy in the right handed neutrinos, similarly to what is actually already present for all the other fermions.

Primordial black holes

Primordial Black Holes are another popular candidate. The reason for their popularity is that they do not require the addition of other fields to the SM picture. The idea is that massive black holes were produced in the early universe and now they are still present. The problem with this scenario is that there is a plethora of constraints, making most of the parameter space non-viable [54]. Many of these constraints are actually disputed in the community and there is not a unanimous agreement on many of them.

2.3 Looking for WIMPs

At this point we have evidences for the existence of dark matter and we have models that provide DM candidates, more specifically our preferred candidate are WIMPs. How do we

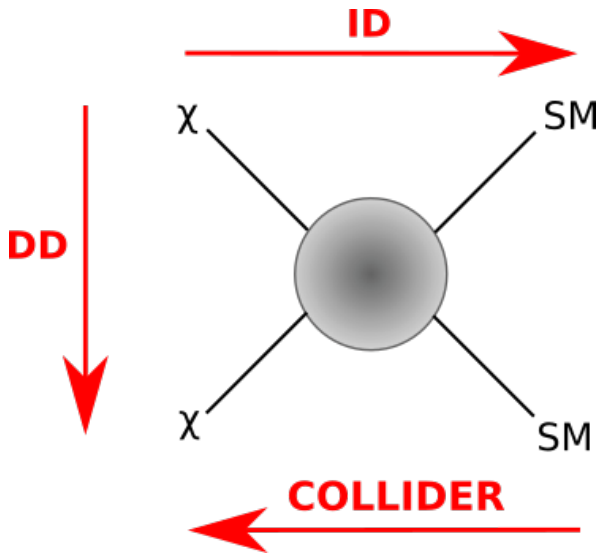

Figure 2.6: Scheme of the three main searches for DM. χ is a generic DM candidate, not necessarily a fermion.

tell if a model is good and how do we look for dark matter? There are three main ways in which this can be done and they are depicted in fig. 2.6. First of all, we can either rely on detecting dark matter already present in the universe or we can try to produce it in the laboratory. This latter case is the one of collider experiments, which will be the search most relevant to us. If instead we want to detect the dark matter already present, we can look for the byproducts of DM-DM annihilation arriving on Earth. Here we talk about indirect detection (ID), since we do not see DM, but only effects of DM. Otherwise, we can look for dark matter crossing the Earth and interacting with some isolated materials here, this is called direct detection (DD).

2.3.1 Collider experiments

Let us first focus on collider searches. Here much of the control is in the hands of the experimentalists. Differently from the other searches, where we need to look at the DM already present in the universe, here we are the ones producing dark matter. Because of this, this is the only search strategy that is not affected by ρ_χ, the DM density profile today. However, that does not imply that the life of collider experimentalists is easy. First of all, DM needs to be produced and this requires very high energies. This translates in the necessity of large, highly technological accelerator stages. Then, we still need to detect DM after we have produced it. Here the detectors actually reconstruct (almost) everything, but definitely not the DM. Let us discuss these two stages in a little more detail.

Production stage

In order to produce DM, we need to reach center of mass energies of the order of $s \gtrsim 4m_{\text{DM}}^2$. Since the most looked for candidates are WIMPs, this means that hopefully we should reach energies of the order of some TeV. Of course, this is far from easy and the reason is twofold.

On one hand, we need very large structures, whether we are considering a linear accelerator or a circular accelerator. On the other hand, we need top notch technology in order to control everything: precise magnetic fields, effective cooling etc. So building an accelerator is a massive task that needs careful consideration.

Depending on what we are looking for, we have two main possibilities for the particles we want to accelerate and then collide.

- Lepton collider

  When talking about lepton collider we mean collisions between electrons and positrons. These particles have no substructure as far as we know, which makes the description of the interaction much clearer with respect to hadron collisions as we will see. A downside lies in the synchrotron emission. Due to their low mass, electrons reach very high velocities quite easily, but electrons in a circular path also lose energy proportionally to their velocity, so that once the electrons are very fast they start losing energy significantly. This sets a limit on how energetic electrons can be in a circular collider and we would not get energies above 500 GeV. So, if we want to get higher energies, we need to work with linear colliders. Given a linear accelerator of length L, we will have that the energy of the electron is linearly proportional to L.

  One of the most important circular colliders of the past was LEP. LEP was intended for the study of electroweak physics [55] and in order to possibly find the Higgs boson [56]. New linear colliders are expected to be built in the future, namely ILC (see [57] for the prospects for physics) and CLIC [58].

- Hadron collider

  Hadron colliders work in a fairly different framework. Here we collide protons with other protons or anti-protons. Since protons are much heavier than electrons, they lose much less energy due to synchrotron radiation. This means that circular accelerator are a viable option and we can reach energies of several TeV. Since the energy is proportional to the radius R of the structure, this clearly implies very large accelerators. The price we have to pay is that protons are not fundamental objects and they have a substructure: this signifies that when we study the collision we have to take into account the parton distribution function. Moreover, since protons interact strongly with each other, the collision is affected by a severe pile-up problem.

  Despite the difficulties of the analysis, hadron colliders have proven to be very powerful. The hadron collider currently working and the reference point for these analyses is the Large Hadron Collider. LHC is notably the responsible for the Higgs discovery [59, 60]. LHC also sets numerous constraints on dark matter models as we will see later on. Any physicist dealing with BSM today needs to pay attention to LHC. This collider should work up until 2036, providing the community with experimental results for still a couple of decades. LHC is limited to a center of mass energy of 14 TeV, but an even larger circular collider (FCC) is expected to be built. FCC could reach energies of up to 100 TeV, strengthening the prospects for new physics discoveries[61].

Detection stage

Regardless of how you produce the particles in the collision, you still need to see what you have created. There is a vacuum zone before reaching the first stage of detectors, hence the types of particle actually present in the detector is limited. Let us discuss the several detection stages and their role.

In the first part of the detector we have the tracker. Here electromagnetic fields deflect charged particles ($e^{\pm}, \mu^{\pm}, \pi^{\pm}, p$) in order to study their charge and momentum.

The second stage is given by the electromagnetic calorimeter, here $e^{\pm}$ and photons are absorbed, giving us information about their energy.

A similar stage is the hadronic calorimeter, here instead the hadrons are absorbed ($p, n, \pi^{\pm}$) and the energy deposited is measured. Even further away, there is the muon chamber, where muons are captured. Clearly the stronger these particles interact with matter, the sooner they will be detected.

But there are also particles escaping and the only particles escaping in the SM are the neutrinos. When we reconstruct the energy, we find that something is missing and the SM explanation is that there are neutrinos. But if DM is produced, then it should escape as well. So, also DM leaves a missing energy trace, in principle not distinguishable from the neutrino one.

We would like then to mention that further detection stages are possible. Indeed, if DM escapes and it is not absolutely stable, then it may decay after exiting the muon chamber. In order to see if this actually happens, we need a further stage, a proposed solution is MATHUSLA for LHC [62].

Analysis stage

After producing and detecting the particles, many of their properties are measured, such as their transverse momenta and the missing energy. Let us briefly overview the physical quantities of interest for LHC physics.

First of all, a necessary geometric disclaimer: collisions happen inside a cylindrical beam and the detectors surround this. This means that only the transverse (perpendicular to the beam line) properties are measured. Conventionally, the beam line is identified with the z axis.

Every particle will be described by a 4-momentum $p^{\mu} = (p_0, \vec{p})$. We can decompose the 3-momentum as:

$$\begin{cases} p_x = |\vec{p}| \sin\theta \cos\phi \\ p_y = |\vec{p}| \sin\theta \sin\phi \\ p_z = |\vec{p}| \cos\theta \end{cases}$$

The only thing we measure is then $\vec{p}_T = p_x \hat{x} + p_y \hat{y}$. Its modulus will be $|\vec{p}_T| = \sqrt{p_x^2 + p_y^2} = |\vec{p}| \sin\theta$. Of course, we can measure this only for the detected particles. But, given the momentum conservation, we can introduce a new quantity which is the missing transverse momentum $\vec{\not{p}}_T$:

$$\vec{\not{p}}_T + \sum_i \vec{p}_{T,i} = 0, \tag{2.25}$$

here the sum is over all seen particles. As stated earlier, both neutrinos and DM will result in missing transverse momentum. The modulus of this vector will be the missing transverse energy $\not{E}_T \equiv |\vec{\not{p}}_T|$.

Cuts at LHC can be performed based on geometrical properties, such as the alignment with the z axis. It is then useful to introduce some new parameters: the rapidity y and the pseudo-rapidity η,

$$y = \frac{1}{2} \log \left(\frac{p_0 + p_z}{p_0 - p_z} \right), \tag{2.26}$$

$$\eta = \frac{1}{2} \log \left(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z} \right) = -\log \left(\tan \frac{\theta}{2} \right). \tag{2.27}$$

η in particular is the quantity on which the cut is usually imposed in the searches.

But what are we looking for? If we are looking for dark matter, a first necessary requirement is the presence of missing energy. However, that is not enough, the reason for this is that there are too many events in the SM with just this characteristic. If we want to reduce the background, we need to look for something else on top of missing energy. The most popular additional tag is the presence of initial state radiation. This can either be an hadronic jet (monojet), or a Z-W boson (mono-V), or a photon (mono-γ) or even a Higgs boson (mono-h). Monojet is possibly the most popular one, but depending on the specific model, it may be that another one of these handles is the most relevant. Even requiring the presence of initial state radiation is not enough. One of the goals of LHC analyses is finding the best regions where to look for. Identifying the regions where we can find the best signal to background ratio is far from trivial. Help can possibly come from machine learning and data analysis, as we explain in section 5.2. Identifying the best regions means imposing a cut on some measured properties, usually the transverse momentum of the jets, the pseudo-rapidity and most importantly the missing energy. Since we are looking for DM, we expect missing energies of the order of the hundreds of GeV.

Instead of trying to reduce the background of common signals, something else can be done: look for uncommon signals. This direction has become quite popular in recent years. Several dark matter models predict signatures that are expected to be quite rare in the SM, these are especially due to long-lived particles (LLP). We can have for instance displaced vertices, where a particle decays inside the detector or we can have R-hadrons, where a colored particle arrives at the detector and then hadronizes, as it is the case for our model in chapter 4. For a review on the possibility of detecting LLP at LHC and some useful software for carrying out simulations, see Ref. [3].

2.3.2 Direct detection

Dark matter is flowing through the Earth all the time, but since it interacts so weakly we do not notice that. If we want to be able to observe these interactions, we need to highly isolate the detector in order to reduce all possible noise due to other interactions. This is what is done in direct detection experiments, such as XENON1T [63] and LUX [64].

In the standard picture, we are going to describe how the flux of DM coming from space scatters off the nuclei imprinting some energy. This is measured as a recoil and can be an indication of the presence of DM. However, very recently an excess in the scatter off electrons

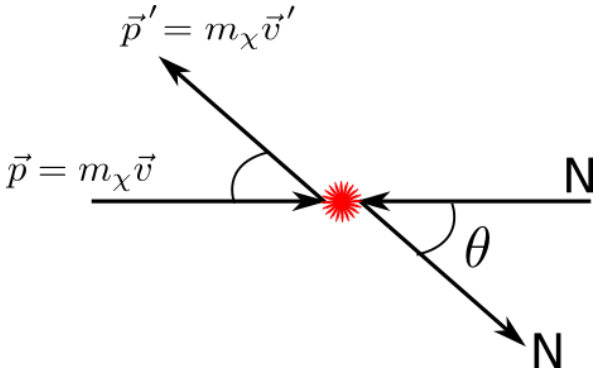

Figure 2.7: Interaction between a nucleon N and a DM particle χ.

has been measured around 2 keV [65]. The theoretical community is currently proposing models in order to explain that and the experimental community is checking whether some effect has been neglected in the analysis.

For what concerns WIMPs, since their masses are of the order of hundreds of GeV, scatterings with nuclei are most relevant as we turn to describe now. The simple non-relativistic interaction is represented in fig. 2.7. This process is non-relativistic, as we are talking about WIMPs exchanging an energy of order of the keV.

What we want to evaluate is the recoil energy, which is the quantity relevant in order to understand the prospect of finding DM with this search:

$$E_R = \frac{|\vec{q}|^2}{2M_N} = \frac{|\vec{p} - \vec{p}'|^2}{2M_N} = \frac{\mu^2 v^2 (1 - \cos\theta)}{M_N} \quad \rightarrow \quad E_R^{\max} = 2\frac{\mu^2 v^2}{M_N}, \tag{2.28}$$

here μ is the reduced mass of the system and v is the velocity of the incoming flux in the center of mass frame. Since there is a lower cut on E_R in the measured events, this means that there will be a lower bound on the velocity required in order to detect the event.

What is actually measured is the number of events, so we care about the rate of events at varying energy. The differential rate at a certain energy will be given by the product of the scattering cross section and the dark matter flux at that energy. Proceeding as in Ref. [66], we derive:

$$\frac{\mathrm{d}R}{\mathrm{d}E_R} = N_N \int \frac{\mathrm{d}\sigma}{\mathrm{d}E_R} v \, \mathrm{d}n_{DM} = N_N \frac{\rho_\chi}{m_\chi} \int_{v > v_{\min}} \frac{\mathrm{d}\sigma}{\mathrm{d}E_R} v f(\vec{v}) \, \mathrm{d}^3 v. \tag{2.29}$$

Here ρ_χ and $f(\vec{v})$ are the local DM density and the velocity profile of DM. Notice that these quantities have different predicted values based on the model used to describe the DM distribution. For the purpose of correctly evaluating the results coming from direct detection, it is important to consider the errors coming from model dependence.

The other relevant part of this expression is given by the cross section. Clearly, the DM is not interacting with isolated nucleons, but rather with some nuclei composed of both neutrons and protons. This means that we also need information coming from nuclear physics. Given the size of the energy exchange the scattering is non-relativistic, this implies that the particle scattering can be approximated with an effective interaction between the DM particle and the quarks and gluons of the nucleon. All possible non-relativistic interactions of this type

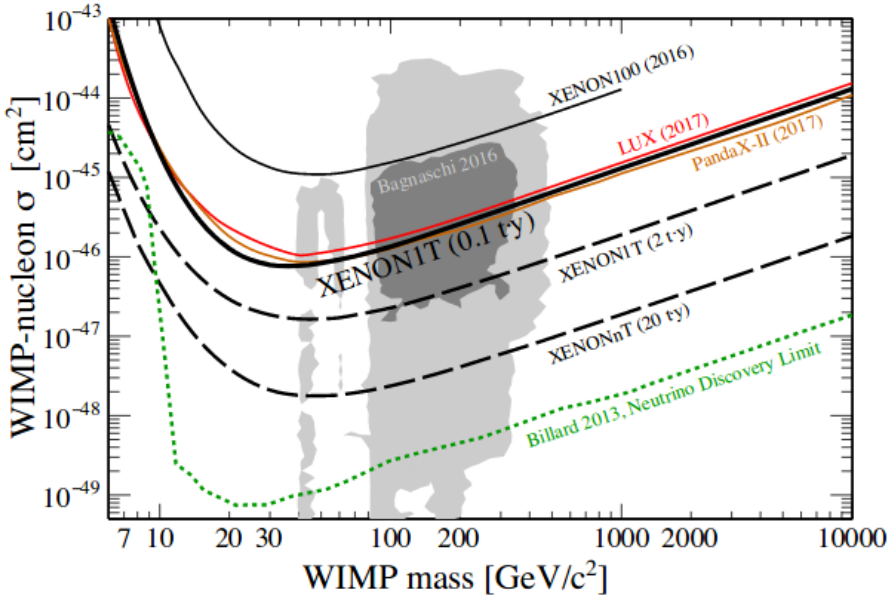

Figure 2.8: Current limits on spin independent cross sections from XENON1T [63], Lux [64], Panda X-II [68]. The SUSY parameter space considered by [69] is presented as a grey area. Image from [63].

are summarized in Ref. [66]. Once the cross section has been decomposed in the various components, we have (for N=neutron, proton):

$$\sigma_{\rm NR} = \sum_i \sum_N c_i(N) O_i^{(N)}, \tag{2.30}$$

where i sums over all the possible operators. When we consider how the nucleons are embedded inside the nucleus T, the form factors will appear. These form factors are described and summarized in Ref. [67]. Using their expression we have:

$$\overline{|\mathcal{M}_T|}^2 = \frac{m_T^2}{m_{\rm N}^2} \sum_{i,j,{\rm N},{\rm N}'} c_i^{\rm N} c_j^{{\rm N}'} F_{i,j}^{{\rm N}{\rm N}'}, \tag{2.31}$$

where $F_{i,j}^{{\rm N}{\rm N}'}$ are the form factors we are interested in.

Once the differential rate is known, a limit on the cross section can be established. An example of such limits can be seen in fig. 2.8.

A phenomenologist usually does not deal with how the bound is derived exactly, but they still need to compare their predictions with the experimental results. In this case, it is useful to reduce the scattering cross section to the non-relativistic limit. A fundamental step to address is understanding whether the cross section in the non-relativistic limit is spin-dependent or spin-independent. Very different bounds apply to these two cases and most notably the bounds on spin-dependent cross sections are around 4 orders of magnitude weaker. Bounds on spin-independent cross section on the other hand are very strong and they are starting to sensibly limit the available parameter space of WIMPs.

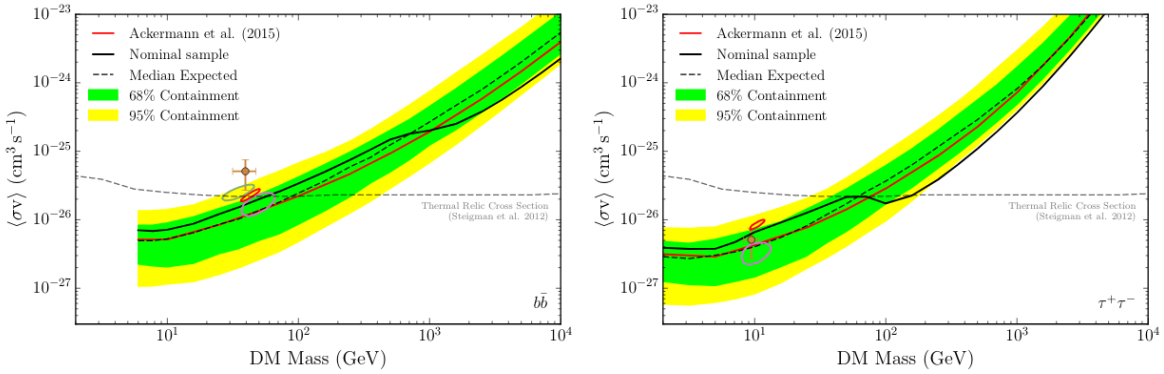

Figure 2.9: Bounds on the DM annihilation cross section from γ-ray observations in two specific channels. Results from FERMI-LAT[70]. *On the left:* limits on the $b\,\bar{b}$ channel. *On the right:* limits on the $\tau^+\,\tau^-$ channel.

2.3.3 Indirect detection

Dark matter that does not flow through Earth can still be detected, even if indirectly. We said that the DM self-interactions have an upper bound and we know that DM should be stable or very long-lived. This does not however completely forbid annihilations into SM model or possibly even decays. Of course, these processes are supposed to be very rare, but at the same time DM is very abundant: annihilations should proceed proportionally to ρ_χ^2 and decays proportionally to ρ_χ. It is worth mentioning that the annihilation rates are proportional to the scattering rates today, this says that if the scatterings are not $s-$wave, they are further suppressed by a factor $v^2 \sim 10^{-6}$.

What can arrive on Earth are the byproducts of annihilations, but since these particles have to travel all the way to Earth, only stable components will survive. We have some freedom here. First of all, we can decide what to look for, since there are several stable particles: γ, $e^\pm$, p, $\bar{p}$, ν and $\bar{\nu}$. Each of these particles have their advantages and disadvantages as we will explain in a bit. But we can also decide where to look and here there are two effects that need to be balanced. On one hand, we want to look at parts of the sky where ρ_χ is large, so that we might have more events. On the other hand, we are talking about astrophysics, so there are a lot of uncertainties in the modeling of the background. This means that sources with abundance of dark matter, such as the centers of the galactic halos, may also carry too much background to be good observation points.

Let us discuss the observed particles in a little more detail.

- Photons

  Photons have the advantage of not interacting sensibly during their path. This implies that we can track their path from production to detection without large uncertainties. As a matter of fact, most of the needed information about the DM flux will be encoded in the J-factor:

$$\mathbf{J} = \int\int \rho_\chi^2(l,\Omega)\mathrm{d}l\mathrm{d}\Omega. \tag{2.32}$$

Crucial here is the model used for the DM profile, some examples of J-factors for different dark matter distributions can be found in Ref. [71].

A disadvantage of photons is that we know dark matter to be dark, hence the production of photons should be heavily suppressed. Nevertheless, relevant bounds are derived from γ observations. Among the most constraining bounds we have the ones coming from FERMI-LAT [70], which we also report in fig. 2.9.

- Neutrinos

  Neutrinos have the same advantage of photons: since they interact very weakly, they arrive on Earth basically unaltered from production. The downside is that this also implies that they are very difficult to detect. The focus in these kind of searches is isolating the detector in order to kill the background. Relevant searches for neutrinos are carried out by ICECUBE [72].

- Antiparticles

  We know that there is more matter than antimatter in the universe, this implies that the detection of antimatter excesses would be a clear indication of unknown physics. Antiparticles considered for this purpose can either be positrons or anti-protons. Unfortunately, each of these observed candidate has its downside. Positrons lose most of their energy due to their electromagnetic interactions and found excesses can be easily explained with the background sources. Anti-protons do not lose as much energy as positrons, but their production and propagation are affected by large uncertainties. Measurements in this direction are carried out by AMS-02 (see [73] for e^+ and [74] for $\bar{p}$).

Chapter 3

A model of theoretical interest: composite 2HDM with singlets

3.1 Motivation

There are many available DM models, but which ones are more interesting? The answer to this question clearly depends on what you are interested in. We know that the SM has several problems, so from the point of view of a theorist, an interesting model could be one that solves more than one problem at once. Indeed, from the point of view of minimality, it would make sense for a single extension of the SM to solve more than one issue. This is the direction we try to pursue here. The problems we will try address are the presence of DM, obviously, the naturalness and hierarchy of the Higgs mass.

The composite Higgs (CH) paradigm [75, 76, 77, 78] can in principle solve or at least mitigate all of these problems, see e.g. the reviews [79, 80, 81]. The hierarchy of the Higgs mass is not necessarily solved in these models, but the issue is surely ameliorated, since quadratic corrections should not be present above the compositeness scale. CH models rely on the concept of spontaneous symmetry breaking we described in section 1.1.2. Here, a new global symmetry $\mathcal{G}$ is introduced at high energies. This symmetry is then spontaneously broken giving rise to one or more particles, among these particles we will surely find the Higgs boson, but possibly we can also find stable DM candidates. The Higgs boson of the SM would then not be an elementary particle, but a pseudo Nambu-Goldstone boson (pNGB). This breaking occurs at a scale f, so a new energy scale is introduced, possibly fixing naturalness, if we can have $v \lesssim f$. This also means that loop-corrections to the Higgs mass must be cut at a lower energy, helping with the Higgs mass hierarchy. We will go into the details of the construction later, but for now we need to clarify how to find the Higgs boson and DM after the breaking. As stated in section 1.1.2, the number of particles appearing after the breaking will depend on the broken and unbroken group, but not only that. Depending on the breaking pattern, we can also establish the representation of the new pNGBs. What we require is (at least) the presence of the Higgs doublet and of a DM candidate, which should be a singlet of $\mathrm{SU}(3)_c \times \mathrm{U}(1)_{em}$. As a plus, since the Higgs particle and the DM arise from the same mechanism, they should have similar masses. This means that we will be dealing with a WIMP.

We will study the model at low-energy, meaning after the breaking of the symmetry, when

the pNGBs are already present. We will construct our model step-by-step. There are several models already studied in the literature for different combinations of $\mathcal{G}$ and $\mathcal{H}$. The most minimal CH model including a DM candidate among the pNGBs is based on the SO(6)/SO(5) coset. In this case, in addition to the Higgs doublet, a scalar particle appears. This is probably the most popular among the studied CH models (see e.g. Refs. [21, 82, 83, 84, 85]). Non-minimal models, where more and more particles appear, have also been widely considered $\mathrm{SO}(6) \to \mathrm{SO}(4) \times \mathrm{SO}(2)$ [83], $\mathrm{SO}(7) \to G_2$ [86], $\mathrm{SO}(7) \to \mathrm{SO}(6)$ [87, 88, 89], $\mathrm{SO}(7) \to \mathrm{SO}(5)$ [90], $\mathrm{SU}(4) \times \mathrm{SU}(4) \to \mathrm{SU}(4)$ [91, 92], $\mathrm{SU}(5) \to \mathrm{SO}(5)$ [93], and $\mathrm{SU}(6) \to \mathrm{SO}(6)$ [94] (see also [95, 96]). So our minimal requirement of having an Higgs doublet and a DM candidate still leaves us with considerable freedom on the choice of $\mathcal{G}$ and $\mathcal{H}$. Is there any reason to prefer one model to another? In choosing these groups there are mainly two different motivations. It might be the case that we are interested in a particular UV completion of the model and based on the UV completion we can derive $\mathcal{G}$ and $\mathcal{H}$. Or we can choose certain groups because we are interested in the phenomenology that arises from the fields generated, possibly leading to interesting effects. In the model we construct, we will follow the latter point of view.

We will construct and study a CH model based on $\mathrm{SO}(7) \to \mathrm{SO}(5) \times \mathrm{SO}(2)$. As we said, we are studying this because of the field content. So what is this model providing us with? We will have two Higgs doublets and two real scalars, the lightest of which is stable and will be our dark matter candidate. As will be explained in our discussion, the presence of the second doublet will be crucial to relax the constraints from electroweak precision tests (EWPTs) and DD. The second scalar does not add much to the first singlet in the standard scenario, except of course for the introduction of new interactions. However, we will see that the presence of the second singlet will open the possibility of a different DM production mechanism. This choice of groups also avoids having to deal with Wess-Zumino-Witten anomalies, which could in principle endanger the stability of DM. This model can be seen as an extension of the composite two-Higgs doublet model (2HDM) based on the coset $\mathrm{SO}(6) \to \mathrm{SO}(4) \times \mathrm{SO}(2)$, studied in detail in Refs. [97, 83, 98, 99]. We do share some features with that model, but in our model there is the addition of the two scalars. Or from another point of view we can see our model as a doubling of $\mathrm{SO}(6) \to \mathrm{SO}(5)$, since we have both a copy of the doublet and a copy of the singlet. Here we still share some features, such as the DM candidate being a scalar, but the presence of the second doublet, as we mentioned, influences our parameter space.

For the purpose of building the low-energy effective theory we will be using the Callan-Coleman-Wess-Zumino (CCWZ) construction [100, 101] and naive dimensional analysis (NDA) [102, 103]. NDA will be central to establish the size of many interactions.

The main parameter of interest for composite models is f, the breaking scale: the EWPTs put severe constraints on its size (usually requiring it to be large). In order to address hierarchy and naturalness, we would actually prefer a small value of f, so in these models there is a sort of tension, which translates in the level of fine tuning. If we also want to have DM, this means more constraints on the parameter space. This is because in addition to the tests of the SM, we also need to be consistent with the relic abundance and with the fact that DM has not been detected yet. One of the nice features of the model is that it actually allows a good DM candidate in a consistent composite Higgs model without requiring high fine tuning. This is thanks to the field content of the model, in particular the presence of the second Higgs doublet. The addition of the other scalar instead introduces the possibility of LLP. Here, this particle is not relevant for its detectability, but rather because it allows us

to open up another part of the parameter space, though here we will have larger fine tuning. As far as we know, this is the first time a LLP pseudo Nambu-Goldstone boson has been studied in a CH model.

The chapter is organized as follows: in section 3.2, we build the effective Lagrangian of the model based on symmetry arguments, while in section 3.3 we introduce the particles, describing their masses and interactions. We then move to describe the DM phenomenology: first, we study the usual thermal WIMP candidate in section 3.4; then we describe the possibility of having a non-thermal contribution in section 3.5. For both scenarios the relic density is reproduced and we study the experimental constraints.

3.2 Lagrangian construction

The starting point in describing the model will be deriving the Lagrangian, we will do this for the coset $\mathrm{SO}(7)/\mathrm{SO}(5) \times \mathrm{SO}(2)$. Just by fixing the coset, we can derive the field content and the CCWZ Lagrangian. By fixing the fermionic representation and using symmetry arguments, we will be able to derive also the fermionic interactions and the pNGB potential.

3.2.1 Coset and the pseudo-NGB

The global symmetry $\mathcal{G} = \mathrm{SO}(7)$ is respected at an energy scale above $\Lambda = m_* \sim$ (few) TeV, but it is spontaneously broken to a subgroup $\mathcal{H} = \mathrm{SO}(5) \times \mathrm{SO}(2)$ at a scale $f \sim m_*/g_*$. Similarly to what happens in chiral QCD, the spontaneous breaking is due to a condensate of the strong dynamics. This strong sector is the one symmetric under $\mathcal{G}$ and with g_* we indicate a typical strong coupling of the composite sector.

We can already see the representations of the pNGB that will be produced. Since we have the generators of the group (whose explicit expression we report in section 3.A), using these is the easiest way to find out their representation, but more general group theory arguments could be used. So we know that we will have 10 new particles transforming as a $(\mathbf{5}, \mathbf{2})$ of $\mathcal{H}$.

The global symmetry $\mathcal{G}$ of the strong sector will be also explicitly broken by the gauging of the SM and the interactions which generate the Yukawa couplings. This means that there will be a potential for these new particles and they will possibly acquire a mass (hence the reason why they are pseudo NGB).

The first thing we need to do is understand where the spontaneous breaking of $\mathcal{G}_{\mathrm{EW}} = \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ is coming from. The EW gauge group $\mathcal{G}_{\mathrm{EW}}$ should be somewhere in the model and in particular it should be embeddable in $\mathcal{H}'$. This new group in general does not coincide with the subgroup individuated by the vacuum of the theory, $\mathcal{H}$. This well known mechanism of vacuum misalignment between these two subgroups of $\mathcal{G}$, shown schematically in fig. 3.1, is responsible for the spontaneous breaking of $\mathcal{G}_{\mathrm{EW}}$. The VEVs here are physical, since $\mathcal{G}$ will also be explicitly broken: it is not possible to rotate $\mathcal{H}$ into $\mathcal{H}'$! Since the vacuum identified by $\mathcal{H}'$ is the one related to the gauging of $\mathcal{G}_{\mathrm{EW}}$, we will call this the *gauge* vacuum. While the vacuum identified by $\mathcal{H}$ will be the *physical* vacuum, as it is more closely related to the generation of the fields, though it is in fact not more physical than the other. In the physical basis what we will have is that the particles do not in fact acquire a VEV, the analogous in the SM is considering a basis where H is defined in such a way that the Higgs field is already considered around its vacuum and does not take a VEV.

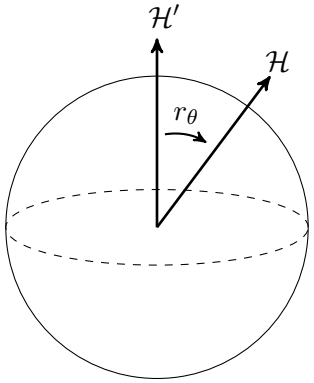

Figure 3.1: The effect of gauging $\mathcal{H}' \subset \mathcal{G}$ is a breaking of the global symmetry due to vacuum misalignment.

In general, we can then consider two basis of generators: one, $\{T_\theta\}$, related to the breaking $\mathcal{G} \to \mathcal{H}$, and a second one, $\{T\}$, related to $\mathcal{G} \to \mathcal{H}'$. This misalignment between the two groups is originated by the VEVs of the theory. But in general, think for instance of a 2HDM, we know that more than one field can acquire a VEV. So the rotation angle can actually be a vector of rotation angles, one for each particle acquiring a VEV. The two vacuum structures are related by a rotation matrix r_θ and, if we assume that the generators are normalized as $\mathrm{Tr}\{T^A T^B\} = \delta_{AB}$, we have:

$$T_\theta = r_\theta \, T \, r_\theta^{-1} \, . \tag{3.1}$$

We then introduce the Goldstone matrices in the two basis as:

$$U \equiv U(\Pi) = e^{i\frac{\sqrt{2}}{f}\Pi} \quad , \quad U_\theta \equiv U(\Pi_\theta) = e^{i\frac{\sqrt{2}}{f}\Pi_\theta} \, , \tag{3.2}$$

where we defined $\Pi \equiv \Pi^I \hat{T}^I$ and $\Pi_\theta \equiv \Pi^I \hat{T}^I_\theta$. Here $\hat{T}^I$ and $\hat{T}^I_\theta$ are the broken generators in the gauge and physical basis respectively, Π^I being the respective pNGB fields.

The rotation r_θ can be obtained by considering the Goldstone matrix in the *gauge* basis $\{T\}$, and setting the NGBs at the corresponding VEVs, i.e.:

$$r_\theta \equiv U(\langle \Pi \rangle) \, . \tag{3.3}$$

The various generators will be embedded in a 7×7 matrix as follows:

$$\{T_L, T_R, T_5, T_2, \hat{T}_1, \hat{T}_2\} \sim \left(\begin{array}{cc|cc} & & & \\ & T_{L,R} & T_5 & \hat{T}_1 \quad \hat{T}_2 \\ & & & \\ \hdashline & \bar{T}_5 & 0 & \\ \hline & \hat{T}_1 & & \\ & \hat{T}_2 & & T_2 \end{array} \right) . \tag{3.4}$$

The most important information coming from this expression is the position of the broken generators, along which the NGBs will appear. The explicit expressions for T and $\hat{T}$ can

be found in section 3.A, but the treatment is standard, as we just extend the representation of SO(5) and SO(6). Notice that the unbroken subgroup SO(5)′× SO(2)′ is represented in a block-diagonal form in this basis. We indicate with the first letters of the alphabet the indices of a generic SO(7) transformation, $a, b, c, \dots = 1, \dots, 7$. Because of its block-diagonal form, instead, an SO(5)′×SO(2)′ transformation will have $\bar{a}, \bar{b}, \dots$ indices, where $\bar{a} = \{i, \mu\}$, with i and μ being SO(5)′ fiveplet and SO(2)′ doublet indices, respectively. In the following, the position of the indices (upper, lower) will be based on practical convenience and have no physical meaning. Repeated indices still indicate a summation and we still use the first index to indicate the row index and the second one to indicate the column.

As already stated, the generators in the *physical* vacuum ($\mathcal{G}/\mathcal{H}$) basis are related to these by $T_\theta = r_\theta\, T\, r_\theta^{-1}$. A nice feature of this approach is that we can expand U_θ in the fields to extract the interactions, so this part will be perturbative. But we can still maintain the exact expression for the VEVs.

Everything in the end will be misaligned, but it is much easier to consider the aligned case, if we want to discuss how the symmetries affect the model. So, for now let us consider $r_\theta = \mathbf{1}$. The NGB matrix will take then the form:

$$\Pi = \hat{T}^I \Pi^I \equiv -\frac{i}{\sqrt{2}} \begin{pmatrix} 0_{5\times 5} & \Phi_1 & \Phi_2 \\ -\Phi_1^T & 0 & 0 \\ -\Phi_2^T & 0 & 0 \end{pmatrix} . \tag{3.5}$$

Φ_1 and Φ_2 are two five-plets of SO(5)′. We usually work with the SM, so we would rather decompose them under the custodial symmetry $\mathrm{SO}(4)_c \subset \mathrm{SO}(5)'$. They are then $(\mathbf{5}, \mathbf{2}) = 2 \times \mathbf{4} + 2 \times \mathbf{1}$:

$$\Phi_1 = (\phi_1, \eta)^T , \qquad \Phi_2 = (\phi_2, \kappa)^T . \tag{3.6}$$

The two fourplets ϕ_1, ϕ_2 describe the two Higgs doublets of the theory. In the limit in which there is no mixing between the two we have:

$$\phi_1 = \begin{pmatrix} G_1 \\ G_2 \\ G_3 \\ h \end{pmatrix} , \qquad \phi_2 = \begin{pmatrix} \frac{-i}{\sqrt{2}}(H_+ - H_-) \\ \frac{1}{\sqrt{2}}(H_+ + H_-) \\ H_0 \\ A_0 \end{pmatrix} , \tag{3.7}$$

where h is the SM-like Higgs. G_i are the would-be longitudinal polarizations of the EW gauge bosons, the ones that will be eaten in order to give mass to Z_μ and $W_\mu^\pm$. Then H_0 and A_0 are the neutral components of ϕ_2, the first one will be CP-even and the second one CP-odd. $H^\pm$ are instead the charged components. There are two singlets η and κ: the lightest of the two singlets, η, will be our DM candidate. This is however the limit in which the second doublet does not take a VEV and hence there is no need to distinguish between mass eigenstates and particles acquiring a VEV. We will see that the second doublet does take a VEV and the mass eigenstates (h, H_0) will be obtained by diagonalizing the mass matrix of (h_1, h_2), which will be our fields acquiring a VEV.

As noticed in Ref. [97], it is very convenient for the feasibility of the model to introduce a discrete symmetry C_2 of the pNGBs. The transformation under this symmetry can be represented as

$$C_2 = \mathrm{diag}(1, 1, 1, 1, 1, 1, -1) , \tag{3.8}$$

Field	$\mathrm{SO}(4)_c$	C_2	P_7
ϕ_1	**4**	$+$	$+$
ϕ_2	**4**	$-$	$+$
η	**1**	$+$	$-$
κ	**1**	$-$	$-$

Table 3.1: Representations and quantum numbers of the NGB fields under the custodial symmetry $\mathrm{SO}(4)_c$ and the accidental symmetries C_2 and P_7. Given these representations ϕ_1 could be identified with the SM Higgs doublet, in the limit in which ϕ_2 does not take a VEV.

acting as $U \to C_2 U C_2$ on the Goldstone bosons matrix. This implies for our fiveplets $(\Phi_1, \Phi_2) \to (\Phi_1, -\Phi_2)$. The role of this symmetry is distinguishing the two doublets, in particular if this symmetry is at least approximate it forbids the second doublet from acquiring a sizable VEV. Without this symmetry, a VEV in the second doublet could easily break custodial symmetry. This is usually very bad, since there are very strong limits on this breaking. This C_2 symmetry will be however broken in our model due to the interactions with fermions. Another useful symmetry is P_7. This can be represented by a transformation

$$P_7 = \mathrm{diag}(1, 1, 1, 1, -1, 1, 1) \ , \tag{3.9}$$

under which $(\eta, \kappa) \to -(\eta, \kappa)$. Clearly, we have that this symmetry stabilizes the dark matter candidate. In the following, we use P_7 as a symmetry of the theory. We will see that this symmetry will be unbroken even in the low-energy regime.

As we stated in the introductory part, with our choice of groups there is no Wess-Zumino-Witten term [104, 105] since the fifth de Rham cohomology group of $\mathrm{SO}(7)/\mathrm{SO}(5) \times \mathrm{SO}(2)$ vanishes [106, 107][1]. This implies that the P_7 parity remains unbroken by the strong dynamics to all orders in the chiral expansion. We summarize our field content, their representation under $\mathrm{SO}(4)'$, their parities in table 3.1.

Lagrangian terms which break explicitly the global symmetry $\mathcal{G}$ could in principle break these parities too. We will see that this is the case for C_2, since the interaction of the top quark with the composite sector will not be symmetric under C_2. As a direct consequence, also the potential terms generated by fermionic operators will not respect this symmetry. Instead P_7 remains a symmetry of the model in both the high and low energy description.

We know that at the end of the day our model will be symmetric under both P_7 and CP, this means that we know how the vacuum structure will look in general. In the physical basis identified by $\mathcal{H}$ the vacuum is exact and retains its trigonometric expression. Then, we can talk about the two angles $\theta_1 = \langle h_1 \rangle / f$ and $\theta_2 = \langle h_2 \rangle / f$, where the expectation values are the

[1] We thank Joe Davighi for sharing this result with us.

ones found instead in the gauge basis. The misalignment matrix of eq. (3.3) is then:

$$r_\theta = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & c_{\theta_2} & 0 & 0 & 0 & s_{\theta_2} \\ 0 & 0 & 0 & c_{\theta_1} & 0 & s_{\theta_1} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -s_{\theta_1} & 0 & c_{\theta_1} & 0 \\ 0 & 0 & -s_{\theta_2} & 0 & 0 & 0 & c_{\theta_2} \end{pmatrix} , \tag{3.10}$$

where $s(c)_{\theta_{1,2}} = \sin(\cos)\theta_{1,2}$.

3.2.2 CCWZ Lagrangian

Since we want to construct a symmetric Lagrangian, the possible allowed interactions are limited. How we build these invariant terms is described by the Callan-Coleman-Wess-Zumino construction [100, 101]. We will not present the complete CCWZ construction here, but just outline the main ingredients which are needed. In the case where $\mathcal{H}$ is compact, we know how the Goldstone matrix will transform under $\mathcal{G}$: $U \to gUh^\dagger(g)$. For the purpose of constructing invariants, it is then useful to consider the object $\bar{A}_\mu$, which is the dressed gauge field:

$$\bar{A}_\mu = iU^{-1}(\partial_\mu - iA_\mu)U \equiv d^i_\mu \hat{T}^i + e^a_\mu T^a . \tag{3.11}$$

From the transformation of U, the transformations of the new objects d_μ and e_μ follow. For what concerns us, e_μ can be used to write covariant derivatives and d_μ describes derivative interactions. This last term is the one that interests us the most and will be used to write the derivative Lagrangians.

The CCWZ description allows to describe the derivative interactions between pNGBs, in particular, it allows to expand the interactions in terms of how many derivatives there are. So of course at leading order there will be two derivatives:

$$\mathcal{L}^{(2)}_\Pi \equiv \frac{f^2}{4} \operatorname{Tr}\left[d^{(\theta)}_\mu d^{(\theta)\mu} \right] , \tag{3.12}$$

where $d^{(\theta)}_\mu \equiv i \sum_I \operatorname{Tr}[U_\theta^{-1} D_\mu U_\theta \hat{T}^I_\theta] \hat{T}^I_\theta$ in our basis. The Goldstones are derived from this interaction term by expanding U_θ, which we remind contains all 10 NGBs. Also the SM gauge bosons are present in this d_μ, since we have a covariant derivative. After expanding, many pieces of the Lagrangian are found. We find the kinetic terms for the Nambu-Goldstone fields and the mass terms for Z_μ and $W^\pm_\mu$. Of course, since we derive parts of the SM Lagrangian from a different perspective, we will have to be consistent with what we know about the SM. We also derive the interactions among different NGBs and the interactions between them and the SM gauge bosons. Let us check the SM quantities we need to reproduce. The mass of the W boson is given by:

$$m^2_W = \frac{g^2 f^2}{4} (\sin^2\theta_1 + \sin^2\theta_2) . \tag{3.13}$$

Then, as usual, in 2HDM we define:

$$\begin{aligned} \sin\theta_1 &\equiv \sqrt{\xi}\cos\beta , \\ \sin\theta_2 &\equiv \sqrt{\xi}\sin\beta , \end{aligned} \tag{3.14}$$

so that the vacuum of the theory is related to the W mass in the usual way: $m_W^2 = g^2 v^2/4$. Where from the previous definition we have:

$$\xi \equiv \frac{v^2}{f^2} = \sin^2\theta_1 + \sin^2\theta_2 \ , \qquad \tan\beta = \frac{\sin\theta_2}{\sin\theta_1} \ , \tag{3.15}$$

with $v = 246\,\text{GeV}$ being the SM VEV.

With these definitions the Z mass becomes:

$$m_Z^2 = \frac{v^2(g^2+g'^2)}{4}\left[1 - \frac{\xi}{4}\left(1-\cos 4\beta\right)\right] \ , \tag{3.16}$$

leading to a tree-level positive contribution to the $\hat{T}$ parameter:

$$(\Delta\hat{T})_{\text{2HDM}} \approx \frac{\xi}{4}(1-\cos 4\beta) \approx 2\xi\beta^2 + \mathcal{O}(\xi\beta^4) \ . \tag{3.17}$$

The $\hat{T}$ parameter is related to the breaking of the custodial symmetry and its SM prediction is 0.

This contribution to $\hat{T}$ is due to the fact that $\langle h_2\rangle$ explicitly breaks $\text{SO}(4)_c$ to $\text{SO}(2)$. This is not what happens for instance in [97], a model which we have said share several properties with ours. The two models differ because of the different CP structure, which in our case implies that we have custodial breaking even if we assume CP to be preserved. The breaking of custodial symmetry does not forbid the model completely, but rather tells us the expected size of some parameters. In our case, we will have that in order not to have very high fine tuning (meaning ξ very small), we would need $\beta \lesssim 0.1$. Such a contribution might even help to improve the fit of electroweak precision observables (see section 3.B for more details). Of course, we should justify why our parameter β has that size. As shown in section 3.3.1, small values of β are predicted by this model. Interestingly enough, the positive contribution to $\hat{T}$ given by eq. (3.17) can help relax the EWPT limits on ξ. This allows us to have rather large values of ξ. We find for instance that for $\beta \approx 0.1$, a fine tuning up to $\xi \approx 0.08$ is compatible with both EWPT and Higgs data (see section 3.B).

3.2.3 Partial compositeness

The previous discussion followed just by the choice of the coset. We are still missing many pieces, since our fermions are massless for now. We know that fermions can couple to H in the SM and acquire a mass after the spontaneous breaking of $\mathcal{G}_{\text{EW}}$. We need here then to couple all our NGBs to the fermions in order to allow a similar mechanism to take place.

To couple the SM fermions to the Higgs field and generate their masses, we use the partial compositeness paradigm: the idea is that quarks are linearly coupled to fermionic operators $\mathcal{O}_{L,R}$ belonging to the strong sector [108]. This is defined as partial compositeness, because in the end the physical particles will be a mixture of elementary and composite fermions. Depending on the size of the coupling, a larger fraction of compositeness will be allowed. This will imply that the "most composite" fermion will be the top quark. We have now to choose the representation of these fermionic operators and this is the second choice we will make beyond the choice of the coset. In the following, we assume that the operators coupled to the top and bottom quarks transform in the fundamental representation of SO(7).

Other choices are possible and would lead to different results (see e.g. Ref. [88] for other representations).

Since we want to reproduce the SM, one of the quantities we will have to reproduce are the SM charges. For this purpose we will need to enlarge the group $\mathcal{G}$ with a spectator group $\mathrm{U}(1)_X$. This group will have no physical consequences on the description, besides the correct reproduction of the SM quantum numbers. We fix the charge of the top quark to be $X = 2/3$. In this way, if we define the hypercharge to be $Y = T^3_R + X$, we have the correct SM result.

The fermionic operators are in a **7** representation, which decomposes under $\mathrm{SO}(5)'\times\mathrm{SO}(2)'$ and $\mathcal{G}_{\mathrm{EW}} = \mathrm{SU}(2)_L\times\mathrm{U}(1)_Y$ as:

$$\mathbf{7}_{\frac{2}{3}} = (\mathbf{5},\mathbf{1})_{\frac{2}{3}} \oplus (\mathbf{1},\mathbf{2})_{\frac{2}{3}} = \mathbf{2}_{\frac{7}{6}} \oplus \mathbf{2}_{\frac{1}{6}} \oplus \mathbf{1}_{\frac{2}{3}} \oplus \mathbf{1}_{\frac{2}{3}} \oplus \mathbf{1}_{\frac{2}{3}} \,. \tag{3.18}$$

We need to couple the left-handed quark doublet (which is in a $\mathbf{2}_{1/6}$ representation) and the right handed top quark (which instead is a $\mathbf{1}_{2/3}$) to these operators. We see that there is only one way to couple the left-handed doublet q_L, but there are, it seems, three ways to couple the right-handed quark t_R. But since the SM fermions are assumed to be even under both C_2 and P_7, we will see that one of these three coupling is actually forbidden by P_7, which we recall is never broken, not even in the low energy. Once we correctly couple the quarks, our Lagrangian will look like this:

$$\mathcal{L}^f_{\mathrm{int}} = \bar{q}^\alpha_L \, \mathcal{Y}^\alpha_{L,a}{}^T \, \mathcal{O}^a_L + \bar{t}_R \, \mathcal{Y}^T_{R,a} \, \mathcal{O}^a_R + \mathrm{h.c.}\,, \tag{3.19}$$

where $a = 1,\dots,7$ is an SO(7) index, $\alpha = 1,2$ is the flavor index of the quark doublet and $\mathcal{Y}_{L,R}$ are the spurions. The spurions are a practical way of ensuring the correct symmetries are respected, the idea is that instead of transforming the fields, we transform the couplings $\mathcal{Y}_{L,R}$. These fields will hence be called spurions (as they are not real fields) and they will have transformations under $\mathcal{G}$. The way they transform is established by requiring that their interactions with $\mathcal{O}_{L,R}$ is invariant under the symmetry $\mathcal{G}$. Compatibly with P_7 (which basically forbids the 5th component of $\mathcal{Y}_R$), we would have still two (complex) ways to couple the right handed top quark with the fermionic operator. This means that there are 4 parameters for the coupling. Actually, three of these parameter are nonphysical and can be rotated away, thanks to the residual elementary symmetries $\mathrm{U}(2)^{el}_L$ and $\mathrm{U}(1)^{el}_R$ and we remain with just one real parameter. Finally, we have that the most general VEVs for the spurions are:

$$\mathcal{Y}_L = \frac{y_L}{\sqrt{2}}\begin{pmatrix} 0 & 0 & i & 1 & 0 & 0 & 0 \\ i & -1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T, \qquad \mathcal{Y}_R = y_R \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \cos\theta_t & i\sin\theta_t \end{pmatrix}^T, \tag{3.20}$$

with y_L and y_R real. Only one parameter remains in the coupling of the right handed quark, this is θ_t, which we will see will be central in the phenomenology. This parameter basically regulates how much the fermion will couple to the 6th component of $\mathcal{O}_R$, rather than the 7th. The breaking of C_2 will be in some sense proportional to the value of $\cos\theta_t$. Since the spurions acquire a VEV, it is important to notice that there are two energy regimes. We have the strong sector, before the breaking, which is symmetric under the spurionic symmetries. These symmetries can be broken after the spurions acquire a VEV. These symmetries might even be present at the EW scale, in which case we call them residual symmetries. In our construction of the potential we will assume that at least the spurionic symmetries are respected by the theory.

In the treatment of invariance under $\mathcal{G}$ it is convenient to dress up the spurions with the NGB matrix. The "dressed" spurions will be defined as:

$$\bar{\mathcal{Y}}_L^\alpha \equiv \left(r_\theta^{-1} U_\theta^\dagger \, \mathcal{Y}_L\right)^\alpha , \tag{3.21a}$$

$$\bar{\mathcal{Y}}_R \equiv r_\theta^{-1} U_\theta^\dagger \, \mathcal{Y}_R \, . \tag{3.21b}$$

The reason we do this is that the dressing takes an object transforming with an index a of $\mathcal{G}$ and returns an object transforming with an index $\bar{a}$ of $\mathcal{H}$. Now we can construct invariants under $\mathcal{H}$, which is something that we know how to do, and we are sure that those terms will also be invariant under $\mathcal{G}$. In the literature this dressing procedure is usually carried out in the *gauge* basis, but we can see that our dressing is consistent with the standard one, i.e. $\bar{\mathcal{Y}} = U^\dagger \mathcal{Y}$. We can check this by going to the basis of the VEVs, where $\langle U_\theta \rangle = \mathbf{1}$ and $\langle U \rangle = r_\theta$.

The derivation of the low-energy effective Lagrangian follows the standard procedure, described for example in Ref. [97]. We will use our dressed spurions and the SM gauge fields to write operators invariant under $\mathcal{H}$. We know that our terms should actually be invariant under $\mathcal{G}$, but this is guaranteed by the dressing. The spurions, similarly to the term d_μ of the CCWZ Lagrangian, contain the matrix U_θ, so that by expanding it we derive the interactions among the pNGBs.

For the purpose of constructing invariants, we can work in the case $\theta_{1,2} = 0$, where the dressed spurions transform as a $(\mathbf{5},\mathbf{1}) \oplus (\mathbf{1},\mathbf{2})$ under SO(5)×SO(2). These two components will be called:

$$\left(\bar{\mathcal{Y}}_5\right)^i \ , \ \left(\bar{\mathcal{Y}}_2\right)^\mu \ , \tag{3.22}$$

with $i = 1, \ldots, 5$, while $\mu = 1, 2$. We then need to contract these components with each other in order to build invariants. This can be achieved by using δ_{ij}, $\delta_{\mu\nu}$. Other possibilities, such as using $\epsilon_{\mu\nu}$, are forbidden due to the fact that we want to respect the C_2-spurionic. This symmetry will be broken in the low energy, but should be respected at least in the strong sector. Because of chirality, only left-right combinations are allowed. Now it might seem that there are two possible invariant, one constructed by contracting a LH $\left(\bar{\mathcal{Y}}_5\right)^i$ with a RH one, the other by contracting a LH $\left(\bar{\mathcal{Y}}_2\right)^\mu$ with one of opposite chirality. That is not actually the case. These two invariants are not independent, due to the singlet one can obtain by combining two **7** with a δ.

The leading order operator generating the top mass is thus given by:

$$\mathcal{L}_t = c_t \frac{m_*}{g_*^2} \, \bar{q}_L^\alpha \left(\bar{\mathcal{Y}}_{L,2}^\alpha\right)_\mu^\dagger \left(\bar{\mathcal{Y}}_{R,2}\right)^\mu t_R \, . \tag{3.23}$$

As we were saying in the introduction, NDA is a guide for us here and it explains why the coefficient m_*/g_*^2 is present. The free coefficient c_t is expected to be $\mathcal{O}(1)$, since the other dependencies from the processes generating the term have been included with m_*/g_*^2. From this effective Lagrangian one obtains both the top mass and (after expansion of U_θ) the interactions between the top quark and all the pNGBs. We know from EWPTs that β should be smaller than 1, then let us expand around $\beta = 0$. The top quark Yukawa coupling is given by:

$$Y_t \approx c_t \frac{y_L y_R}{g_*} (\sqrt{1-\xi} \cos\theta_t + \beta \sin\theta_t) \ . \tag{3.24}$$

Note that the factor in parenthesis approaches β for $\theta_t \to \pi/2$. This would lead to a suppression, which could generate problems, since we need to reproduce the large top Yukawa coupling. This suppression can be compensated by a slightly larger value of c_t or a large value of $y_L y_R/g_*$.

What about the other quarks? They can be coupled with similar Lagrangian terms, but their physics is not too relevant since the couplings will be much smaller. Let us at least consider the bottom quark. We have $Y_b \ll Y_t$, so its effect will not be as relevant, especially for the potential contributions. We set $\theta_b = 0$ for simplicity, since we have checked that different values do not lead to different results.

We can now see explicitly that C_2 is broken in the low energy. The Lagrangian in eq. (3.23) leads to the interaction term $ihA_0\bar{t}\gamma^5 t$: this explicitly violates the residual C_2. Since C_2 is not respected at low energy, the second doublet is allowed to take a VEV (whose size will depend on how much the residual symmetry is broken).

By looking at other terms generated, we also see that if we do not want further contributions to CP breaking, η and κ will need to have opposite CP-parities. We assume that η is CP-even and κ is CP-odd.

We know that CH models can have issues with flavor changing neutral currents (FCNCs). We can avoid issues if the embedding of all the other SM fermions is fixed carefully in order to align the Yukawa interactions. Since the fermions are in the fundamental representation, we would have two independent strong sector invariants; however, the spurionic C_2 forbids one of these. This is already good, since we do not have the possibility of different generations coupling to the two invariants in different proportions. Nevertheless, to avoid FCNCs, different generations should have the same embedding, including the same choice of $\theta_u = \theta_c = \theta_t$, $\theta_d = \theta_s = \theta_b$, if we want to align the Yukawa interactions and recover the SM. We are then imposing minimal flavor violation. Operators from the strong dynamics could also induce flavor violation, but in our analysis we are focusing our attention on the low energy phenomenology, so we do not consider these contributions.

3.2.4 Pseudo-NGB potential

In the previous parts we have said that our NGBs are actually pseudo-NGBs, meaning that they are massive. If they have a mass, they should have a potential. This potential is generated by the explicit breaking of $\mathcal{G}$ due to both the gauging of the EW subgroup of $\mathcal{G}$ and to the mixing between SM fermions and the composite operators. We can guess the size of this potential by employing NDA again. The generated potential can be written as (see also e.g. Ref. [103] for a justification of each term):

$$V(\Pi) \sim m_*^2 f^2 \left(\frac{g_*^2}{16\pi^2}\right)^L \left(\frac{g_{\rm SM}}{g_*}\right)^{\mu_G} \left(\frac{y}{g_*}\right)^{\mu_F} \hat{V}\left(\frac{\Pi}{f}\right) \tag{3.25}$$

where $\hat{V}$ is a dimension-less function of the NGBs. These potential will be generated by loops of strong sector fermions, the number of such loops will be L and each additional loop will give a suppression of order $g_*^2/16\pi^2$. Since the breaking derives from interactions with SM fields (either the gauge bosons or the top), we need also to count SM insertions with μ_G and μ_F for the gauge and fermionic contributions respectively.

Our construction is analogous to the one of Ref. [97].

Gauge contributions

Let us start by considering the gauge contributions.

It is convenient to introduce a set of spurions in order to describe the role of SM gauge bosons:

$$\mathcal{G}' \equiv \sum_{A=1}^{21} \mathcal{G}'_A \, T^A = g' \, T_3^R \ , \qquad \mathcal{G}^\alpha \equiv \sum_{A=1}^{21} \mathcal{G}^\alpha_A \, T^A = g \, T_L^\alpha \ , \tag{3.26}$$

transforming under $g \in \mathcal{G}$ as $\mathcal{G}^X \to g\mathcal{G}^X g^\dagger$. Since they are in the adjoint representation, they need to be dressed as follows:

$$\bar{\mathcal{G}}^X \equiv r_\theta^{-1} U_\theta^\dagger \mathcal{G}^X \, U_\theta r_\theta \, . \tag{3.27}$$

Their components $\bar{\mathcal{G}}^X_A = \mathrm{Tr}\left[\bar{\mathcal{G}}^X T^A\right]$ transform under $\mathrm{SO}(5) \times \mathrm{SO}(2)$ in this way:

$$\mathbf{21} = (\mathbf{10}, \mathbf{1}) \oplus (\mathbf{5}, \mathbf{2}) \oplus (\mathbf{1}, \mathbf{1}) \, . \tag{3.28}$$

Each of these three terms corresponds to a transformation under some of the generators. The three components are associated respectively to $\{T_L, T_R, T_5\}$, $\{\hat{T}_1, \hat{T}_2\}$ and T_2. This tells us which indices need to be considered for the contraction. Let us rewrite the three components as:

$$\left(\bar{\mathcal{G}}^X_{10}\right)^I \ , \ \left(\bar{\mathcal{G}}^X_{\hat{T}}\right)^i_\mu \ , \ \bar{\mathcal{G}}^X_2 \, , \tag{3.29}$$

with $I = 1, \dots, 10$ being an index in the adjoint, while $i = 1, \dots, 5$ and $\mu = 1, 2$ are the indices associated to SO(5) and SO(2), respectively. We now know how to contract the indices (still compatibly with the spurionic C_2 and P_7):

$$\begin{aligned} \mathcal{I}^{(1)}_{g'} &\equiv -\left(\bar{\mathcal{G}}'_{10}\right)^I \left(\bar{\mathcal{G}}'_{10}\right)_I \ , & \mathcal{I}^{(2)}_{g'} &\equiv -\bar{\mathcal{G}}'_2 \, \bar{\mathcal{G}}'_2 \ , \\ \mathcal{I}^{(1)}_{g} &\equiv -\left(\bar{\mathcal{G}}^\alpha_{10}\right)^I \left(\bar{\mathcal{G}}^\alpha_{10}\right)_I \ , & \mathcal{I}^{(2)}_{g} &\equiv -\bar{\mathcal{G}}^\alpha_2 \, \bar{\mathcal{G}}^\alpha_2 \ . \end{aligned} \tag{3.30}$$

Then by plugging these expressions in eq. (3.25) we get the potential generated by the gauging:

$$V_{\text{gauge}} = \frac{m_*^4}{16\pi^2} \sum_i \sum_{\tilde{g}=g,g'} \frac{1}{g_*^2} \, c^{(i)}_{\tilde{g}} \, \mathcal{I}^{(i)}_{\tilde{g}} \, , \tag{3.31}$$

where $c^{(i)}_{g,g'}$ should be $\mathcal{O}(1)$ coefficients due to naturalness.

Fermionic contribution

We will see that the main source of explicit breaking of the Goldstone symmetry is due to the coupling of the composite sector with the top quark. This is kind of intuitive, considering that among the Yukawa couplings y_t is the largest one and it also larger than g and g'. The Lagrangian describing the interaction between the elementary fermions and the strong sector is the one in eq. (3.19). This term is the one leading to the breaking of $\mathcal{G}$ and to the generation of the potential. The spurions we will use to construct the invariant terms of the potential are the same used for this Lagrangian, so the ones in eq. (3.22). These invariant terms will no longer contain elementary quarks, so rather than caring about chirality, as we

did when constructing eq. (3.19), we are now limited by requiring gauge invariance. Hence we can construct the objects:

$$\bar{\Delta}_L^{\bar{a}\bar{b}} \equiv \bar{\mathcal{Y}}_L^{*\,\alpha,\bar{a}} \bar{\mathcal{Y}}_L^{\alpha,\bar{b}}\,, \tag{3.32a}$$

$$\bar{\Delta}_R^{\bar{a}\bar{b}} \equiv \bar{\mathcal{Y}}_R^{*\,\bar{a}} \bar{\mathcal{Y}}_R^{\bar{b}}\,. \tag{3.32b}$$

The lowest order invariants constructed from these are:

$$\mathcal{I}^{(1)}_{(1,0)} = \bar{\Delta}_L^{ii}\,, \quad \mathcal{I}^{(1)}_{(0,1)} = \bar{\Delta}_R^{ii}\,, \tag{3.33}$$

they are generated at 1-loop, with the minimum amount of SM insertions, so they are $\mathcal{O}(y^2)$. At the next order in the SM coupling, which is $\mathcal{O}(y^4)$, the non-vanishing invariants are:

$$\begin{aligned}
&\mathcal{I}^{(1)}_{(2,0)} \equiv \bar{\Delta}_L^{ij}\,\bar{\Delta}_L^{ji}\,, \quad \mathcal{I}^{(1)}_{(1,1)} \equiv \bar{\Delta}_L^{ij}\,\bar{\Delta}_R^{ji}\,, \quad \mathcal{I}^{(1)}_{(0,2)} \equiv \bar{\Delta}_R^{ij}\,\bar{\Delta}_R^{ji}\,,\\
&\mathcal{I}^{(2)}_{(2,0)} \equiv \bar{\Delta}_L^{ii}\,\bar{\Delta}_L^{jj}\,, \quad \mathcal{I}^{(2)}_{(1,1)} \equiv \bar{\Delta}_L^{ii}\,\bar{\Delta}_R^{jj}\,, \quad \mathcal{I}^{(2)}_{(0,2)} \equiv \bar{\Delta}_R^{ij}\,\bar{\Delta}_R^{ij}\,,\\
&\mathcal{I}^{(3)}_{(0,2)} \equiv \Im\left[\bar{\Delta}_R^{\bar{a}i}\,\bar{\Delta}_R^{\bar{a}i}\right]\,,
\end{aligned} \tag{3.34}$$

here only the operators indicated with $^{(1)}$ are generated at one loop, all the other ones are generated at two loops [97], and are thus accompanied by a further factor of $g_*^2/(4\pi)^2$.

Similarly to what we did for the gauge contributions, we can now plug these expressions in eq. (3.25) and get the fermionic contribution to the potential:

$$V_{\text{fermion}} = N_c\,\frac{m_*^4}{16\pi^2}\sum_{n_L,n_R,i}\frac{1}{g_*^{2(n_L+n_R)}}\,c^{(i)}_{(n_L,n_R)}\,\mathcal{I}^{(i)}_{(n_L,n_R)}\,, \tag{3.35}$$

where $\mathcal{I}^{(i)}_{(n_L,n_R)}$ is an invariant formed with $n_{L,R}$ powers of $\bar{\Delta}_{L,R}$, and $c^{(i)}_{(n_L,n_R)}$ are $\mathcal{O}(1)$ coefficients. The strong sector fermions in the loops are colored, so there is a factor $N_c = 3$ accounting for the number of colors. Since we assumed there is no further CP breaking coming from the effective Lagrangian of the composite Higgs model, we take $c^{(3)}_{(0,2)} = 0$, given that $\mathcal{I}^{(3)}_{(0,2)}$ breaks CP.

We now have 8 free coefficients from the fermionic operators and 4 from the gauge contribution. However, they are not all equally important. From the calculations we derived that $c^{(1)}_{(1,0)}$, $c^{(1)}_{(0,1)}$, $c^{(1)}_{(2,0)}$, $c^{(1)}_{(1,1)}$, $c^{(1)}_{(0,2)}$, $c^{(2)}_{(1,1)}$ and $c^{(2)}_{(0,2)}$ are the ones relevant for the phenomenology. All the other coefficients are set equal to 1 for simplicity. These coefficients, which from now on will be called c_i, are also not arbitrary. We want to address naturalness, so we require them to be order 1. In order to make this "order 1" criterion more quantitative, we define three possible ranges of variation of these coefficients, depending on how close they are to unity:

- *strictly* natural coefficients: $0.2 \leq |c_i| \leq 5$;
- *loosely* natural coefficients: $0.1 \leq |c_i| \leq 10$;
- unnatural coefficients: $|c_i| < 0.1$ or $|c_i| > 10$.

3.3 NGB dynamics

Now we have our field content and the Lagrangians describing all the interactions, so we dedicate this section to a discussion of their properties (VEVs, masses, effective couplings).

3.3.1 Vacuum structure

Up until now, we have talked about two general angles θ_1 and θ_2, but now we can calculate their values explicitly. Actually, we cheated a little bit in the previous part, since in truth not until after the calculation of the potential we could have known which symmetries were effectively respected in the low-energy regime. So, in principle we could have had even more VEVs, but from the VEV analysis we can do now, we know that there are actually just two, as we have always claimed. For the purpose of carrying out this vacuum structure analysis we set to zero the pNGB fields in the potential $V_{\rm tot} = V_{\rm gauge} + V_{\rm fermion}$. Then, similarly to what we did for the SM, we minimize the potential to find the vacuum expectation values. This yields a value for $\langle h_1 \rangle$ and $\langle h_2 \rangle$. In practice, it is more practical to deal with ξ and β, rather than the two angles, since they have more physical meaning. We thus need to use eq. (3.14) and after the minimization of the potential we would find that these two parameters are a function of the various coefficients. What we do is fix ξ and β to values of our interest, this will be our benchmark, then two coefficients c_i are fixed so that we are always in our benchmark. Namely, we solve for $c^{(1)}_{(1,0)}$ and $c^{(1)}_{(1,1)}$ and check that the values of these coefficients lie within the desired naturalness range. An approximate expression for ξ, obtained at leading order in $y_{L,R}/g_* \ll 1$, is:

$$\xi = \sin^2\theta_1 + \sin^2\theta_2 \approx \frac{2N_c y_L^4 c^{(1)}_{(2,0)} + g_*^2\left(N_c y_L^2 c^{(1)}_{(1,0)} - 3g^2 c^{(1)}_g - g'^2 c^{(1)}_{g'}\right)}{N_c y_L^4 c^{(1)}_{(2,0)}} \,. \tag{3.36}$$

Here we can see that the tuning in the coefficients appears. We will see that we require $\xi \lesssim 0.1$, this means that some cancellation must occur in the numerator of this expression. The amount of tuning that needs to be imposed is of order $\Delta \sim \xi^{-1}$. For what concerns the other VEV parameter β, we know that it should be small. This was already discussed in Ref. [97], where a hierarchy $\theta_2 \ll \theta_1$, i.e. $\beta \ll 1$, can be obtained in CH models. That is the case for our model in some portions of the parameter space. By minimizing the potential, we get:

$$\tan\beta = \frac{\sin\theta_2}{\sin\theta_1} \approx \frac{N_c\, c^{(1)}_{(1,1)} y_L^2 y_R^2 \sin 2\theta_t}{2g_*^2(g'^2 c^{(1)}_{g'} + 2N_c\, y_R^2 c^{(1)}_{(0,1)} \cos 2\theta_t)} \,. \tag{3.37}$$

We see that indeed $\tan\beta$ is proportional to $\sin 2\theta_t$, implying that a suppression is present for $\theta_t \approx \pi/2$. This portion of the parameter space will be of primary importance for us, since here we can reproduce the DM relic density in the standard freeze-out paradigm. Since the gauge contribution in the denominator should be small, we can see that also having $y_L \ll g_*$ leads to smaller values of β. The other region of parameter space interesting for us is when $\theta_t \approx \pi/4$. In this case we still manage to have $\tan\beta \sim 0.1$, although here we are not helped by the natural suppression from $\sin 2\theta_t$.

3.3.2 Spectrum

We have seen in the previous part that β is usually suppressed and small. Even if that did not happen naturally, we would still need small values of β. This justifies treating the masses and interactions of the particles by expanding around small values of this angle. We then look at the potential in order to derive the spectrum of our theory, meaning the masses of our pNGBs.

The mass matrix is not diagonal, since a mixing between h_1 and h_2 is present. We can rotate with an angle $\alpha \approx \beta$ and recover the expression for the masses:

$$\begin{pmatrix} h \\ H_0 \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \tag{3.38}$$

Remember that two coefficients c_i have been fixed, in this way we can talk about ξ and β. We then express the physical Higgs mass at leading order in these small parameters:

$$\begin{aligned} m_h^2 &\approx \frac{N_c f^2 \xi}{16\pi^2}\left(2y_L^4 c^{(1)}_{(2,0)} + y_R^4 c^{(1)}_{(0,2)}(3+4\cos 2\theta_t + \cos 4\theta_t)\right) \\ &\approx \frac{N_c g_*^2}{8\pi^2} m_t^2 \left(2\frac{y_L^2}{y_R^2} c^{(1)}_{(2,0)} + \frac{y_R^2}{y_L^2} c^{(1)}_{(0,2)}(3+4\cos 2\theta_t + \cos 4\theta_t)\right) , \end{aligned} \tag{3.39}$$

as mentioned in the previous section, not all coefficients are equally important and for this expression we reported only the relevant ones. We then substituted the expression for the top-Yukawa (cf. eq. (3.24)) in order to be able to compare the Higgs mass and the top-mass, since we know that there is less than a factor 2 between their values. We can now discuss the role of the strong sector coupling g_*. In principle, this coupling can be as high as 4π, but we see from the expression of the Higgs mass that such a large value would force us to tune the coefficients in order to recover the Higgs mass. On the contrary, a small value of g_*, i.e. light top partners, would help us in this sense. In the light of this, we fix $g_* = 3$ for our calculations. Similarly to what we did for ξ and β, also here we fix the coefficient $c^{(1)}_{(2,0)}$ in such a way that the Higgs mass is always the right one. Of course, we will always check that $c^{(1)}_{(2,0)}$ is order 1.

Now that ξ, β, and m_h have been fixed, only a few free coefficients remain. The masses of the other pNGBs, H_0, A_0, $H_\pm$, η and κ can now be written as functions of these. To the leading order in ξ, we have:

$$m_{H_0}^2 \approx -N_c \frac{y_R^2}{8\pi^2}\left(c^{(1)}_{(0,1)} + \frac{y_L^2}{8\pi^2} c^{(2)}_{(1,1)}\right) m_*^2 \cos(2\theta_t) , \tag{3.40a}$$

$$m_{A_0}^2 \approx m_{H_0}^2 , \tag{3.40b}$$

$$m_{H_\pm}^2 \approx m_{H_0}^2 - \frac{m_h^2}{2} , \tag{3.40c}$$

$$m_\eta^2 \approx N_c \frac{y_R^2}{8\pi^2}\left(c^{(1)}_{(0,1)} + \frac{y_L^2}{8\pi^2} c^{(2)}_{(1,1)}\right) m_*^2 \cos^2\theta_t , \tag{3.40d}$$

$$m_\kappa^2 \approx N_c \frac{y_R^2}{8\pi^2}\left(c^{(1)}_{(0,1)} + \frac{y_L^2}{8\pi^2} c^{(2)}_{(1,1)}\right) m_*^2 \sin^2\theta_t . \tag{3.40e}$$

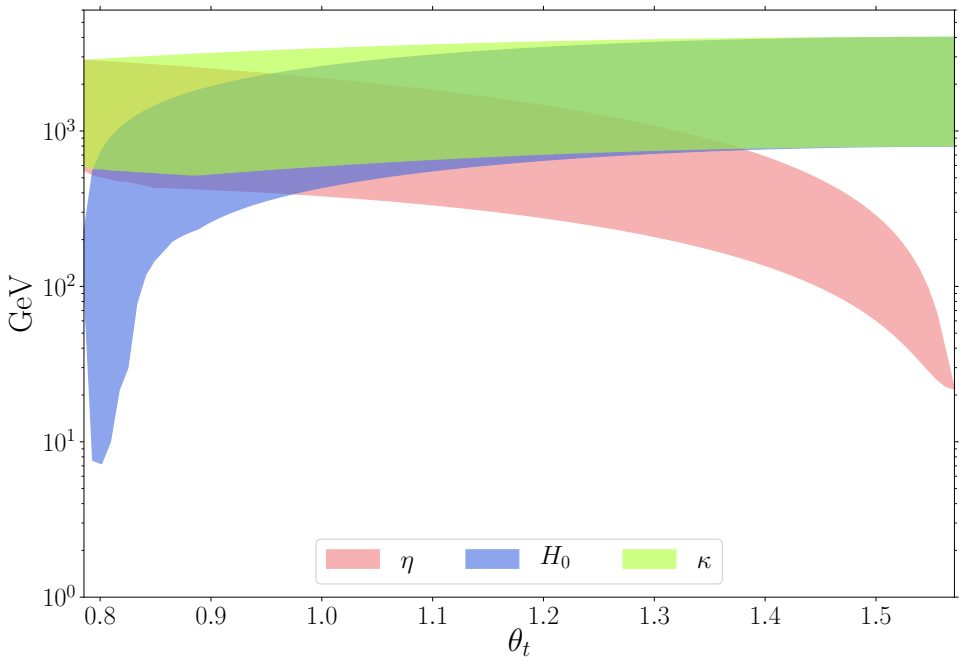

Figure 3.2: Mass spectrum of the pNGBs for $\xi = 0.061$, $\beta = 0.1$ and $g_* = 3$. The bands are obtained by varying the remaining coefficients ($c^{(1)}_{(0,1)}$, $c^{(1)}_{(0,2)}$, $c^{(2)}_{(0,2)}$, $c^{(2)}_{(1,1)}$) in the strictly natural range: $|c_i| \in [0.2, 5]$.

Here there would be a gauge contribution at leading order to the masses, but due to the way we fixed some of our coefficients, this contribution cancels at the first order in ξ. We would however recover the dependency from the gauge contribution at the next to leading order in ξ. As we can see, H_0, A_0 and $H_\pm$ are close in mass. We want all masses to be positive, this means that we should have $\pi/4 \leq \theta_t \leq \pi/2$ and $c^{(1)}_{(0,1)}, c^{(2)}_{(1,1)} > 0$. We calculated the various masses for different values of the remaining free coefficients, we show the result in fig. 3.2 for $\xi = 0.061$ and $\beta = 0.1$, which will be our benchmark point.

Since the mass of η is proportional to $\cos^2\theta_t$, we need to consider the next order in ξ, at least when $\theta_t \approx \pi/2$. We then have for m_η:

$$m_\eta^2 \approx N_c\, c^{(1)}_{(0,1)} \frac{y_R^2}{8\pi^2} m_*^2 \cos^2\theta_t + \xi\, \frac{m_h^2}{2}\,, \tag{3.41}$$

this approximate expression holds in particular in the case in which the order 1 coefficients do not differ too much in value. From fig. 3.2, we notice two extreme regions of the parameter space. It will turn out that these two regimes are also the ones interesting for DM phenomenology, albeit for two completely different reasons. On the right of the plot we have $\theta_t \lesssim \pi/2$, where η is by far the lightest pNGB (other than the SM Higgs), with a mass $m_\eta \sim \mathcal{O}(100\,\text{GeV})$, at least a factor 10 below the masses of the other pNGBs. Since the other pNGBs are so much more massive, they do not participate in a relevant way to the

phenomenology. The other region is on the left of the plot, for $\theta_t \gtrsim \pi/4$. Here we have that η and κ very close in mass and have masses of $\mathcal{O}(1\,\text{TeV})$. This region could in principle be interesting: if κ had a long enough lifetime, it could freeze-out in the early universe and then decay into η. η is our DM candidate, this implies that the relic would be generated not (only) by the freeze-out of η, but also by the late time-decay of κ. We denote this latter contribution as a "non-thermal" contribution to the relic density, where we took the naming from Ref. [109], since a similar effect appears in SUSY.

We will dedicate the next two sections (3.4, 3.5) to a discussion of both regimes.

3.3.3 pNGB interactions

In the construction of our model we encountered three pieces contributing to the Lagrangian of the model: the CCWZ Lagrangian, the one from partial compositeness and the pNGB potential. Since each of these pieces leads to different types of interactions, we consider them separately.

Interactions from CCWZ

The CCWZ Lagrangian in eq. (3.12) contains both the SM EW gauge fields $(Z_\mu, W^\pm_\mu)$ and the pNGBs. So we have all their interactions and self-interactions:

$$\begin{aligned}
\mathcal{L}^{(2)}_\Pi \supset & \mathcal{L}_{\text{kin}} + \left(m_W^2 W^+_\mu W^{-\mu} + \frac{m_Z^2}{2} Z_\mu Z^\mu\right)\left(1 + \frac{2g_V}{v}h + \frac{b_h}{v^2}h^2 + \frac{\lambda^{(V)}_\eta}{2v^2}\eta^2 + \frac{\lambda^{(V)}_\kappa}{2v^2}\kappa^2 + \dots\right) \\
& - \frac{2}{v}\left(g_{H_0 W} m_W^2 W^+_\mu W^{-\mu} + \frac{g_{H_0 Z} m_Z^2}{2} Z_\mu Z^\mu\right) H_0 \\
& - \frac{k_{\text{der}}}{4v^2}\Big[\left(\Phi_1^2(\partial_\mu \Phi_1)^2 - (\Phi_1 \partial_\mu \Phi_1)^2\right) + \left(\Phi_2^2(\partial_\mu \Phi_2)^2 - (\Phi_2 \partial_\mu \Phi_2)^2\right) + \\
& \qquad 2\left(\Phi_1\Phi_2(\partial_\mu \Phi_1 \partial^\mu \Phi_2) - (\Phi_1 \partial_\mu \Phi_2)(\Phi_2 \partial_\mu \Phi_1)\right) + \left(\Phi_1 \overset{\leftrightarrow}{\partial}_\mu \Phi_2\right)^2\Big] \\
& - \frac{m_W m_Z}{2v} g_{H_+ V} W^-_\mu Z^\mu H_+ + \text{h.c.} + \dots
\end{aligned} \tag{3.42}$$

where $\mathcal{L}_{\text{kin}}$ contains the pNGB kinetic terms and $k_{\text{der}} = 2\xi/3$, its value being fixed by the choice of coset. Given the symmetries of the CCWZ Lagrangian, it is convenient to talk about $\Phi_{1,2}$: the two pNGB five-plets of *physical* fields introduced in eq. (3.6). Other couplings which break custodial symmetry could be present, but they would become negligible once the limits from EWPTs are taken into account. We also omitted other interactions with two gauge bosons and two pNGBs, since such terms would have no consequences in our discussion.

People familiar with CH models may notice the lack of interactions with three pNGBs and two derivatives, once the explicit expression for the fields are used. We do not have terms like $h\partial_\mu\eta\partial^\mu\eta$, but these kind of interactions are indeed usually present in these models and their relevance has been repeatedly stressed (see e.g. Refs. [21, 82, 83, 88]). These terms would usually arise from Π^4 terms, once the pNGBs take a VEV. However, we are discussing everything in the *physical* basis (as stated in section 3.2.1), this means that the fields do not have a finite VEV and therefore these terms are not generated. Indeed, let us start from a situation in which the two pictures are equivalent. If we set to zero the gauge and Yukawa

couplings, the *physical* basis and the *gauge* basis coincide, since the vacua are degenerate and this implies that we can rotate the basis in a way that has no physical consequences. In this situation, the two-derivative NGBs interactions start at $\mathcal{O}(\Pi^4)$ in both vacua as it is the case for us in eq. (3.42). If instead the gauge and Yukawa couplings are present, this selects SO(5)$'$ as the vacuum and we can no longer rotate the basis. However, in our basis fields do not take a VEV and certainly the CCWZ Lagrangian does not depend on these couplings, so the derivative interactions are not affected and therefore cubic ones are not generated. Of course, it must be the case that the physics is the same in both basis.

If we want to change the basis to the one where cubic terms are present (i.e. recover the fields around the *gauge* vacuum SO(5)$'$ and allowing then a VEV for the pNGBs), then this can be easily obtained via a non-linear field redefinition ([97, 83]). Let us consider the fields $\tilde{h}$ and $\tilde{\eta}$ as the physical fields in the *gauge* description, then in the limit of $\theta_2 = 0$ we can transform in such a way:

$$h \;\to\; \tilde{h} + \theta_1 \frac{\tilde{\eta}^2}{3f} + \mathcal{O}(\theta_1^2)\,, \tag{3.43}$$

$$\eta \;\to\; \tilde{\eta} - \theta_1 \frac{\tilde{\eta}\tilde{h}}{3f} + \mathcal{O}(\theta_1^2)\,. \tag{3.44}$$

We see that with such a transformation we would recover the cubic derivative interactions from the kinetic terms, as well as non-derivative interactions from the pNGBs mass terms. Under these transformations the physical observables remain invariant.

Interactions with fermions and from the potential

While the interactions in the CCWZ Lagrangian are all derivative and thus involve also the SM gauge bosons, the terms from the potential and the fermionic Lagrangian are all scalar terms. In particular, from the partial compositeness Lagrangian we recover the interactions between the quarks and the pNGBs, including the Higgs field of the SM, whose coupling to the quarks could be in principle modified. We know that in the SM the production of the Higgs particle is boosted by gluon fusion processes, similarly to that we will also have the generation of loop-induced couplings between the gluons and some pNGBs. This will be relevant for the discussion of LHC limits. Finally, from the potential we derive all the scalar interaction between pNGBs. We summarize all these terms:

$$\mathcal{L}_q \supset -\frac{m_q}{v}\,\bar{q}q\left(k_q h + k_{H_0 q} H_0 - \frac{g_q}{2v^2}\eta^2 + \frac{g_{\kappa q}}{2v^2}\kappa^2\right) - \frac{g_{\eta\kappa q}}{v^2}\, m_q \eta\,\kappa\,\bar{q}\gamma^5 q \tag{3.45a}$$

$$\mathcal{L}_g^{t-\text{loop}} \supset \frac{g_{gh}}{v} h G^a_{\mu\nu} G_a^{\mu\nu} + \frac{g_{gH_0}}{v} H_0 G^a_{\mu\nu} G_a^{\mu\nu} + \frac{g_{g\eta}}{v^2}\eta^2 G^a_{\mu\nu} G_a^{\mu\nu} \tag{3.45b}$$

$$\begin{aligned} V \supset & -\frac{g_{\eta h}}{2}\, v\,\eta^2 h - \frac{g_{\eta H_0}}{2}\, v\,\eta^2 H_0 - \frac{g_{\kappa h}}{2}\, v\,\kappa^2 h - \frac{g_{\kappa H_0}}{2}\, v\,\kappa^2 H_0 \\ & - \frac{\lambda_{\eta h}}{4}\,\eta^2 h^2 - \frac{\lambda_{\eta H_0}}{4}\,\eta^2 H_0^2 + \frac{\lambda_{\eta A_0}}{4}\,\eta^2 A_0^2 + \frac{\lambda_{\eta H_+}}{2}\,\eta^2 H_+ H_- \\ & - \frac{g_{A_0 h}}{2}\, v\, h\, A_0^2 - \frac{g_{A_0 H_0}}{2}\, v\, H_0 A_0^2 - g_{H_+ h} v h\, H_+ H_- - g_{H_+ H_0} v H_0 H_+ H_- \\ & + \frac{m_h^2}{2v^2}\,\lambda v\, h^3 - \frac{\lambda_{H_0}}{6}\, v\, H_0^3 + \frac{g_{H_0}}{2}\, v\, h\, H_0^2 - \frac{g_{H_0 hh}}{2}\, v\, h^2 H_0 \end{aligned}$$

$$-\frac{\lambda_{\kappa h}}{4}\,\kappa^2 h^2-\frac{\lambda_{\kappa H_0}}{4}\,\kappa^2 H_0^2+\frac{\lambda_{\kappa A_0}}{4}\,\kappa^2 A_0^2+\frac{\lambda_{\kappa H_+}}{2}\,\kappa^2 H_+ H_-. \tag{3.45c}$$

There are several effective couplings, their values as functions of the free coefficients c_i, ξ, β, θ_q are reported in section 3.C. We will always check that they are less than 4π for perturbative reasons: this often forces $c^{(1)}_{(0,1)}$ to be smaller than 1.

3.4 Thermal dark matter scenario

Let us start from the standard WIMP paradigm. This is realized in the part of the parameter space where $\theta_t \lesssim \pi/2$. Here η, our stable DM candidate, is the lightest pNGB and the dark matter abundance is generated via the usual thermal freeze-out. As we already hinted earlier, the fact that $\theta_t \approx \pi/2$ has several physical implications: from fig. 3.2 we see that η is much lighter than the other resonances, this means that they will not have a relevant role in setting the relic abundance, except if θ_t starts being considerably less than $\pi/2$, as we will see. We know that β should be smaller than 1, and this is easily the case for such values of θ_t: no further tuning or unnaturally small value of the potential coefficients are needed, as can be seen from eq. (3.37). Finally, by looking at the expression for the top mass in eq. (3.24), we understand that there is a suppression in a term giving rise to the Yukawa coupling. This could be compensated by having either a large c_t or $y_L y_R \gtrsim g_*$: the first option could imply an unnaturally large coefficient, so we choose the second option and take $y_L = 2$, $y_R = 3$ and $g_* = 3$. This choice is consistent with perturbativity and with the phenomenological requirement of having the left-handed top less composite than the right-handed one. We see that in any case the couplings are strong and of the same order, so we are making a reasonable assumption. Furthermore, we have checked that some other possible choices do not change the results qualitatively.

The other opposite region of the parameter space, namely when $\theta_t \approx \pi/4$, could be potentially interesting because of co-annihilations with other pNGBs (cf. fig. 3.2). In this case, we would still be in the picture where the relic density is set by the standard freeze-out mechanism, as discussed in section 2.2.2. However, we verified that in this region, with the mentioned constraints on the coefficients, it is not possible to reproduce the observed relic density in this paradigm. Instead, in this region the non-thermal DM production mechanism due to the late time decay of κ can take place. Since the properties and the limits on this candidate change considerably with respect to the completely thermal one, an in-depth discussion will be carried out in section 3.5.

3.4.1 Relic density

For the purpose of calculating the relic density, we need to consider the annihilations into SM particles. In particular, the main channels will be the ones into a couple of SM EW gauge bosons, Higgses and top quarks. Other channels, such as the ones with other quarks, will be subleading. Anyway, in our computations we included all the contributions: their explicit expressions can be found in section 3.D. The relic density profile as a function of the DM mass is shown in fig. 3.3. We have here two areas with different colors: as discussed in section 3.2.4, we want the coefficients to be order 1, but this still leaves us rather unrestrained. We then introduced the concept of loosely natural and strictly natural coefficients, these concepts

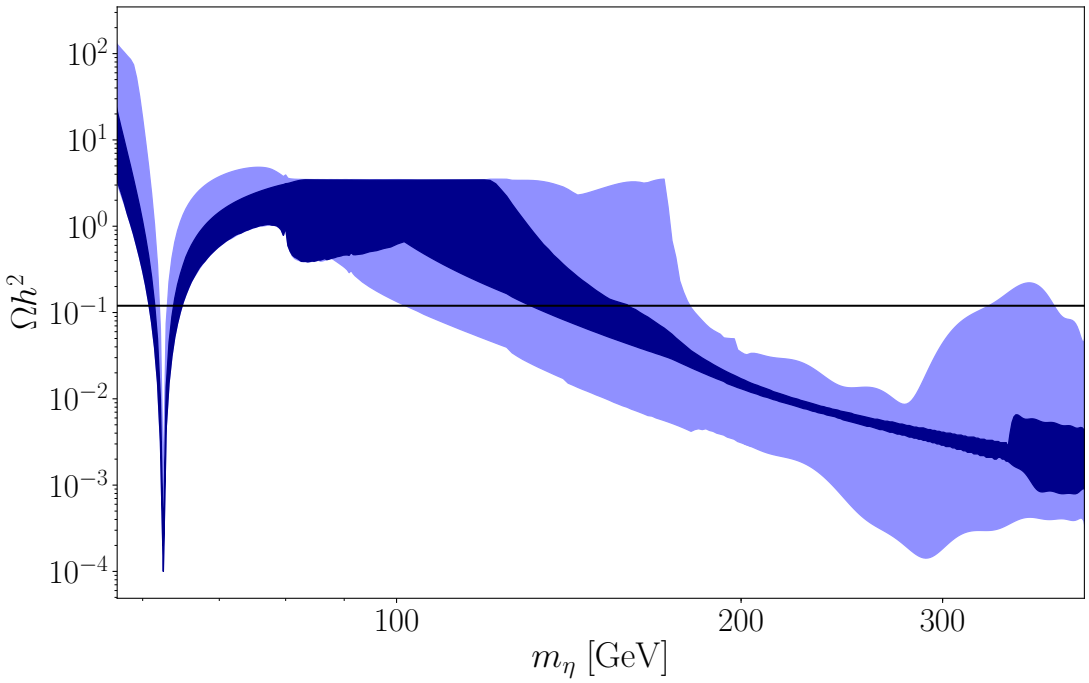

Figure 3.3: Dark matter relic density as a function of the DM mass m_η; we have considered here our benchmark point $f = 1\,\text{TeV}$ ($\xi = 0.061$) and $\beta = 0.1$. The dark (light) blue region is obtained by varying the free coefficients c_i in the strictly (loosely) natural range (see section 3.2.4). We see that the correct relic density ($\Omega h^2 = 0.1198$ [28]), indicated by the black line, can be reproduced for natural values of these coefficients.

are represented in the picture with the lighter and darker areas respectively. It should be understood that in the darker region all the coefficients are strictly natural and the relic here could also be reproduced if some coefficients were loosely natural. Conversely, the lighter region cannot be reproduced unless at least one coefficient is outside of the range $[0.2, 5]$.

It is useful in order to interpret the plot in fig. 3.3 to remember the dependence of the relic density from the cross section (cf. eqs. (2.14) and (2.15)): $\Omega h^2 \propto 1/\langle\sigma v\rangle$. So an upper bound on the relic density corresponds to a lower bound on the cross section. This is the reason why a plateau appears for $m_\eta \lesssim m_h$. By varying the coefficients we regulate the channels coming from the pNGB potential, but the CCWZ terms do not depend on these. So in this region the cross section is dominated by annihilations into SM gauge bosons, even varying the coefficients we cannot decrease the cross section for this annihilation channel. At larger masses we see that the value of the coefficients gains importance: new annihilation channels open up and even terms coming from the potential become relevant. Usually we expect the relic density to decrease with increasing mass, given the scaling of the cross section with energy, even if we play with the coefficients. However, it is possible that a cancellation in the main contributions to the effective cross section occurs: this is what can be grasped from the

plot of fig. 3.3 for $m_\eta \approx 400\,\text{GeV}$. Here η starts getting heavier and θ_t start getting smaller than $\pi/2$: this implies that the exchange of H_0 is no longer negligible. But now we have a channel with h in s-channel and another one with H_0: these two contributions can have opposite signs and give destructive interference, leading to a suppression of the cross section and a bump in the relic density.

We have included everything in the calculations, but of course not all couplings have the same importance. Among the effective couplings, the one that plays the most important role is definitely $g_{\eta h}$: this coupling describes the interaction between two h's and one η and comes from a scalar interaction in the pNGB potential. This effective coupling enters different channels and can either be positive or negative in our parameter space.

The relic density result depends strongly on the VEV parameters ξ and β. For the purpose of setting a meaningful benchmark, we look at the limits coming from EWPTs (which are discussed in section 3.B). So we will present the results for $\xi = 0.061$ (corresponding to $f = 1\,\text{TeV}$) and $\beta = 0.1$. This point is allowed by EWPTs, but we still do not know whether it respects also the constraint from DM detection. At least, we know that we can reproduce the correct relic density. If we look at fig. 3.3, we can see that there are three good mass ranges: $m_\eta \approx m_h/2$, $m_\eta \approx 150\,\text{GeV}$ and $m_\eta \approx 400\,\text{GeV}$. The latest one is present because of the aforementioned cancellation in the $s-$channel, but we will see that it is excluded by DD in our benchmark point and the fine tuning will have to be increased considerably in order to save this mass range. But let us now discuss the experimental constraints so that we may understand why it is excluded.

3.4.2 LHC searches

We first discuss the LHC searches. We start by considering $m_\eta \leq 62.5\,\text{GeV}$, here the Higgs can decay into two DM particles, $h \to \eta\eta$. We start from here since the invisible Higgs decay is widely studied experimentally and there are strong constraints coming from ATLAS and CMS [110, 111]. We notice that since we modify the couplings of the Higgs particle to the SM, also the decay widths for the processes visible at LHC are modified. This implies that we will need to rescale them appropriately when dealing with the branching ratio. However, in our model no major departure from the SM result is expected, since the corrections depends on the values of ξ and β and these are small due to the limits from electroweak precision tests. We take as an upper bound $BR_{\text{inv}} < 19\%$ at 95% CL [110]. It is interesting to look at what will happen to this constraint in the future, since it is a channel of deep interest. The HL-LHC aims at reaching a 95% CL exclusion sensitivity of 1.9%, which implies a much stronger bound on our model. This process can also be investigated at future electron-positron colliders, since the signal is much clearer here, they plan to reach sub-percent bound on the BR (see, e.g., Ref. [112] for a recent review of Higgs boson measurements at future colliders).

The invisible Higgs decay width for our model is:

$$\Gamma_{h\to\eta\eta} = \frac{g_{\eta h}^2}{32\pi m_h} v^2 \sqrt{1 - \frac{4m_\eta^2}{m_h^2}}\,, \tag{3.46}$$

which is proportional to $g_{\eta h}^2$. In the previous section we mentioned that this coefficient will be the most important one and this one of the several cases where it will play a central role. Since $g_{\eta h}$ comes from the pNGB potential it is a function of the coefficients c_i. By varying

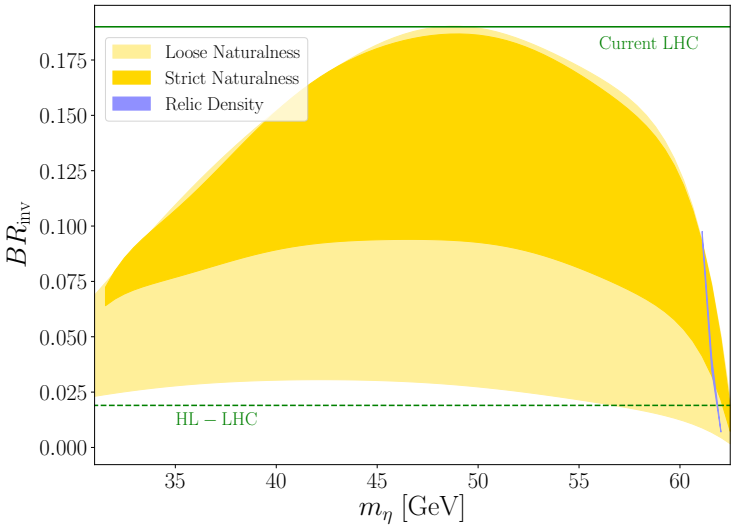

Figure 3.4: Branching ratio of Higgs into DM as a function of the DM mass. The dark (light) yellow region corresponds to strictly (loosely) natural coefficients c_i. We present both the current bound, $BR_{\rm inv} < 0.19$ at 95% CL [110] as a green solid line and the HL-LHC prospect $BR_{\rm inv} < 0.019$ as a green dashed line. On the right side of the plot the narrow region where the correct relic density is obtained is shown. This is currently unprobed by the invisible Higgs searches.

the coefficients c_i, we then get different values for this width and in general for the invisible branching ratio. Its range of variation is depicted in fig. 3.4, again with a darker (lighter) color for strictly (loosely) natural coefficients. As can be seen, at the moment we are unconstrained in the part of the parameter space that also reproduces the DM relic abundance. However, it is worth pointing out that HL-LHC will be able to test most of the available parameter space with this search.

The invisible Higgs limit is relevant only for $m_\eta \leq 62.5\,\text{GeV}$, while for a generic mass the bounds from searches looking at $\not{E}_T$ + tag apply. For our model the missing energy trace could be produced by direct production of η particles or by the decay of other pNGB into η. In this case, the relevant tags are an energetic jet, so the usual monojet signature (MJ) or two well-separated jets, so vector boson fusion signature (VBF). This latter signature is typical of models involving Higgs-like particles. For both cases, it will be important to consider the effective coupling of gluons to the massive scalars, cf. eq. (3.45b), since this gives the main contribution to the production cross section.

We implemented the model in FeynRules [113, 114] and generated events with MadGraph5 [115]. Since h is the lightest particle which can be in the s-channel, it will be the most relevant one. However, we have also included processes involving H_0, as they could give important corrections to the MJ and VBF processes. After performing our simulations, we obtain that for $m_\eta > 50\,\text{GeV}$ both MJ and VBF do not put any further constraint on the model. Our predicted production cross sections are always at least an order of magnitude below the experimental limits ([116, 117]). This result has been obtained by varying ξ, β and

the c_i into reasonable ranges.

Concluding, DM searches at LHC do not put important constraints on the parameter space of our model at the moment, though BR_{inv} could test the lowest masses in the future.

3.4.3 Direct detection

In CH models the strongest limits usually come from DD, we will see that this holds also for us. For DD we need to look at how the DM candidate η interacts with quarks. In our model this interaction is possible either via the contact interaction generated by partial compositeness, which will be proportional to the coupling g_q (see eq. (3.45a)), or through an exchange of h and H_0. These are all tree-level processes, so we expect our model to be probed by DD. In order to study these constraints for our model, we need to evaluate the spin-independent DM-nucleon cross section. Given that we want to compare our prediction with the experimental limits on DM-nucleon scattering cross section, the easiest thing to do is parametrize the interaction Lagrangian as:

$$\mathcal{L}_{DD}^{(\text{eff})} = \sum_q a_q m_q \eta^2 \bar{q} q \,, \tag{3.47}$$

where for our model:

$$a_q = \frac{1}{2}\left[\frac{g_q}{v^2} - \left(k_q \frac{g_{\eta h}}{m_h^2} - k_{H_0 q}\frac{g_{\eta H_0}}{m_{H_0}^2}\right)\right]. \tag{3.48}$$

In our case, since we have imposed minimal flavor violation, we must have $a_t = a_c = a_u$ and $a_b = a_s = a_d$. In the expression in eq. (3.48) we can see that different terms are summed, one of them in particular contains the coupling $g_{\eta h}$, which can either be positive or negative. This implies that cancellations are possible. Of course, it is not certain that the values of $g_{\eta h}$ that cancel the DD amplitude are compatible with the values that provide the correct relic density. But that will actually be the case for our benchmark. The allowed parameter space will then be set by requiring that this amplitude is small. The first term is necessarily negative, since g_q is negative (see eq. (3.57a)): this implies that to have a partial cancellation in the amplitude, also the term in parenthesis will need to be negative.

From the effective Lagrangian above we derive the spin-independent DM-nucleon cross section (cf. the treatment in section 2.3.2 and Ref. [66]). At present, the strongest constraint on it comes from the XENON1T experiment [63]. The coupling g_q comes from partial compositeness, so it does not depends on the coefficients c_i, but both $g_{\eta h}$ and $g_{\eta H_0}$ have origin in the pNGB potential and thus depend on the coefficients c_i. Since these coefficients leave us with some freedom, it will be possible for the cancellation to actually occur. As can be seen in fig. 3.5, DD will impose strong constraints on all the three mass regions where we get the correct relic density. The large mass range is completely excluded by DD for our benchmark values of ξ and β.

Similarly to what we did for LHC, also here we consider the future prospects for direct detection. We make the comparison in this case with the predicted values of the XENONnT experiment [118]. The available parameter space will be sensibly reduced by taking into account these future limits.

3.4.4 Indirect detection

Finally, we turn our attention to indirect detection. For this search we mainly focus on limits from dwarf spheroidal galaxies (dSphs) given by Fermi-LAT. Among the channels reported in Ref. [70], the most constraining one for our model is given by DM-annihilation into $b\bar{b}$. In the region $m_\eta \approx m_h/2$, the process $\eta\eta \to b\bar{b}$ is the only one that needs to be considered and it is also the dominant one in setting the relic density. For higher DM masses, instead, processes leading to the production of other particles, which will later decay into bottoms are important. We have considered the branching ratios into bottoms in this case, but as a conservative estimate, we can assume that these intermediate states completely decay into $b\bar{b}$. The final result will not change appreciably with this assumption and we see that in general ID does not put any further constraint on our model.

Anyway, we also know that bounds from AMS-02 are worth exploring, these are derived from the observation of anti-proton excesses. As discussed for instance in Ref. [88], the systematic uncertainties for this search are still under debate and we limit ourselves to the bounds from observations of γ-rays. A comprehensive treatment of ID is beyond the scope of our study.

3.4.5 Discussion

We summarize the main results for our benchmark point $f = 1\,\text{TeV}$ ($\xi = 0.061$) and $\beta = 0.1$ in fig. 3.5.

As understood from fig. 3.3, three regions are available based on the requirement of having the correct relic density: $m_\eta \approx m_h/2$, $m_\eta \approx 150\,\text{GeV}$ and $m_\eta \approx 400\,\text{GeV}$. The three plots refer to each of these regions. However, good parameter points for the relic density could be experimentally excluded. In the blue region the correct DM abundance at 3σ, $\Omega h^2 = 0.1198 \pm 0.0036$ [28], is obtained. By looking at the orange and purple hatched areas, we can see instead what is excluded by DD and ID respectively. In these plots, just to be more clear in the results presented, we only show the regions obtained when the coefficients are loosely natural. We can see that there is a gray region: this part of the parameter space is unnatural, as it would require coefficients outside the range $[0.1, 10]$.

The bound we considered in less detail, as it is the one least of interest for our study, is the one from ID. However, we see that current indirect detection limits from dSphs do not exclude regions where the correct relic is reproduced. They are in any case close to this parameter space, so a future stronger bound could be able to test this model, especially in the low mass range.

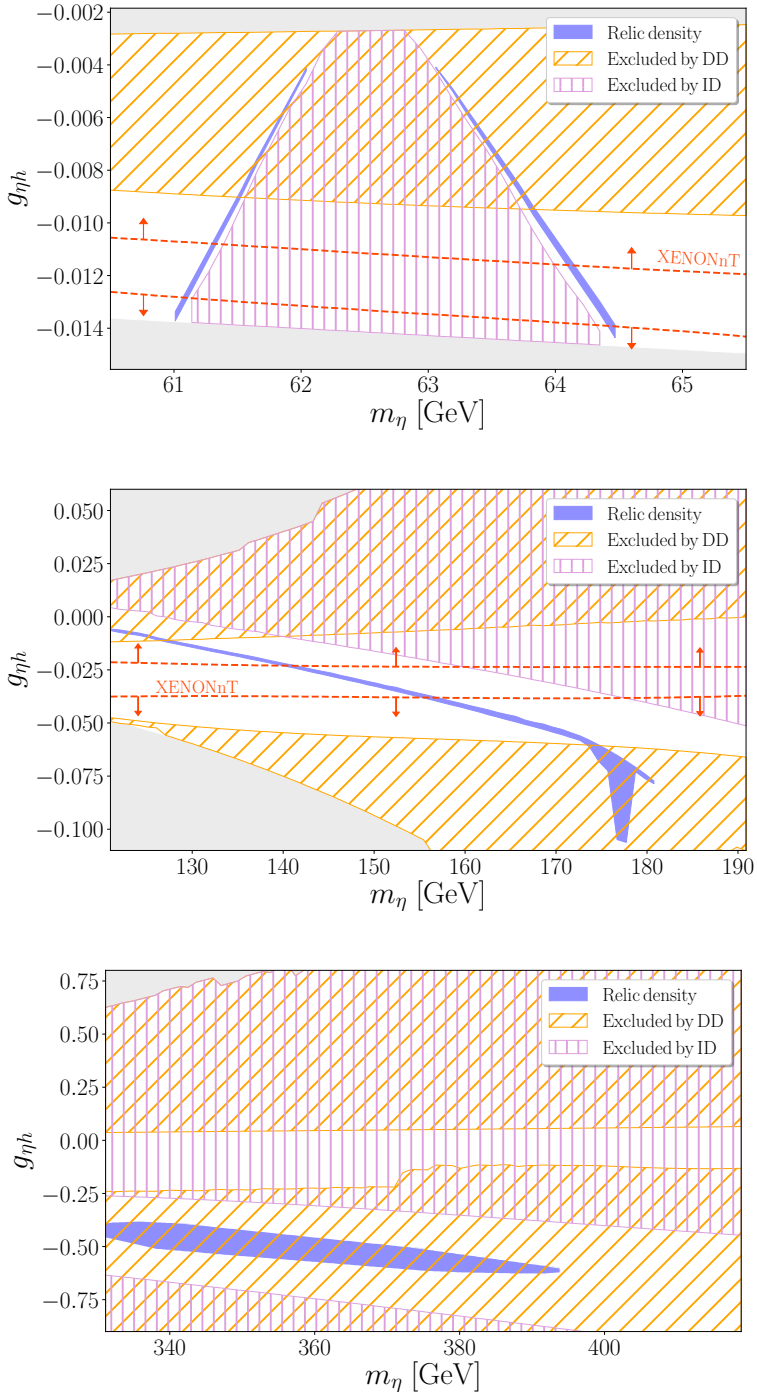

Figure 3.5: We plot the study of the DM phenomenology for the three different mass ranges for $\xi = 0.061$, $\beta = 0.1$. The results are given as functions of m_η and the coupling $g_{\eta h}$, for loosely natural coefficients c_i. The blue region corresponds to the 3σ-relic density contour. The regions of parameter space excluded by direct and indirect detection are depicted with the orange and purple hatched regions. A gray region is also present where no result is depicted: this is the part of the parameter space that cannot be accessed with loosely natural coefficients.

Low mass range ($m_\eta \approx m_h/2$)

The correct relic density is reproduced for masses just below and just above the on-shell Higgs production threshold of 62.5 GeV. This is the well-known Higgs resonance, so the allowed mass range will be a very narrow band around the resonance value. Such a specific value for m_η should be justified somehow, in absence of such a justification, we would need to introduce further tuning in the model parameters. The results from the Higgs invisible BR and from ID do not exclude any parts of the region where the correct relic density is obtained. DD probes part of the parameter space and sets an upper bound on $g_{\eta h}$. Even if ID does not exclude any points in the parameter space at the current state, the cross section in our model is only a factor $1.1 - 1.2$ below the bound. This suggests that upgraded searches are expected to completely probe this region of the parameter space. It is also worth mentioning that predicted precise measurements of the invisible branching ratio of the Higgs will tackle most of the parameter space where the correct relic density is reproduced. However, in the near future DD and especially ID seem more promising directions.

We conclude by mentioning that in general this case looks like the standard Higgs portal, but it is not. For a comprehensive study of the differences between a non-linear Higgs portal and the standard one see Ref. [119].

Intermediate mass range ($m_\eta \approx 150$ GeV)

The most interesting region for us will be the one for masses larger than around 100 GeV. Here direct detection is the search that at the end of the day is constraining our parameter space. Simulations tell us that limits from LHC experiments are not relevant and considering that we are far from the experimental bounds, we do not expect this search to be able to probe this model in the near future. The benchmark values of ξ and β we considered are of course permitted by EWPTs and constraints on the Higgs couplings. In this mass range, differently from the previous one, we are overestimating the bound from ID, and yet this search is not limiting our model. Indeed, as we anticipated at the beginning of this discussion, upper and lower bounds on m_η are set by considering the limits from DD (cf. fig. 3.5). In the plots we are always in the loosely natural range, so let us say a few words for the case in which these coefficients are "more natural". If we were to limit to coefficients in the range $[0.2, 5]$, then the masses below 135 GeV or above 160 GeV would be excluded. The feature at $m_\eta \approx 180$ GeV, namely an enlargement in the available parameter space, is because of a cancellation in the cross section for $\eta\eta \to t\bar{t}$, due again to the interference between the channels with h and H_0 in the $s-$ channel (cf. eq. (3.59c)). It was not obvious a priori that the correct relic density would lie in the same region where the DD amplitude cancels, but that was indeed the case.

Here ID and LHC are not putting any limits and DD leaves a significant region of the parameter space available. These values of m_η are natural, if $\theta_t \lesssim \pi/2$. These observations motivate a deeper study of this case, so we start looking at different values of ξ.

Decreasing a bit the value of ξ has the effect of enlarging the allowed mass range, so that it is interesting to see which masses are available with varying fine tuning: this is shown in fig. 3.6. We observe that the constraints from DM phenomenology in this model allow low values of the breaking scale f. We can reach up to $f \approx 750$ GeV for strictly natural coefficients and even below 600 GeV for loosely natural ones. Indeed, the bounds from DM phenomenology can be weaker than the ones from EWPTs and Higgs precision measurements.

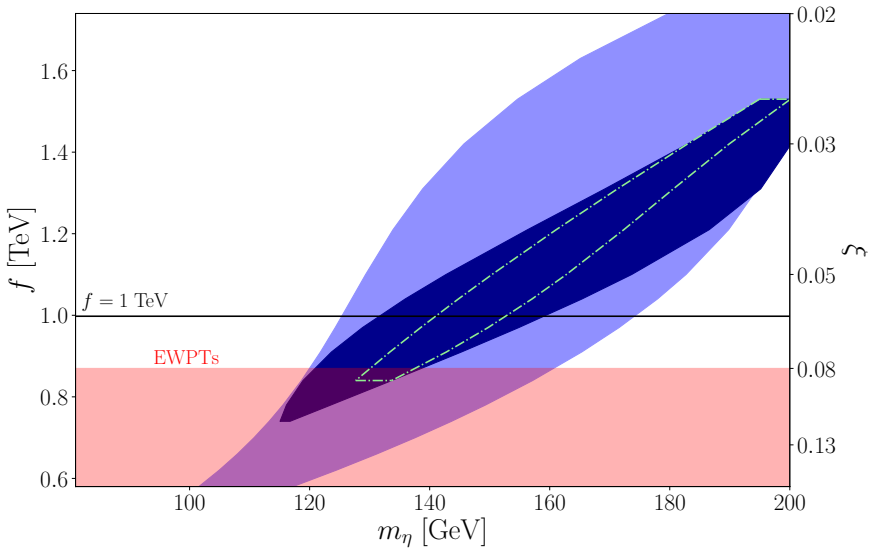

Figure 3.6: All the points in the dark or light blue region are available parameter points: the correct relic density is reproduced and we respect the experimental constraints from DD. As usual, the dark and light region refer to coefficients that are strictly and loosely natural respectively. We can see how different ranges of m_η become available as we change f. The red region is excluded by our combined fit of EWPTs and Higgs couplings (see section 3.B) and sets a lower bound on the breaking scale f. The black horizontal line corresponds to our benchmark point $f = 1$ TeV ($\xi = 0.061$). We see that, considering the expected future limits from XENONnT, part of the parameter space where the coefficients are strictly natural would still be available: this is represented by the green, dashed line.

The relevant lower bound on f are set by these analyses. We can compare our result with other similar non-minimal composite DM models in the literature, we see that significantly larger values of f are usually found to be necessary (see e.g. [83, 87]). The reason why we can have viable DM phenomenology with a fine tuning which is lower than in other models is two-fold: on the one hand, a small positive contribution to $\hat{T}$ from the second doublet can help relax the constraint and allow a lower fine tuning; in addition, because of the richness of the model, we can have a cancellation in the DD amplitude and, luckily, this cancellation occurs exactly for values of $g_{\eta h}$ which are compatible with the reproduction of the correct relic density. We can even see from fig. 3.6 that lowering too much the value of f would lead to smaller available mass ranges: this is because varying ξ we are changing the part of the parameter space where the DD amplitude can cancel.

We can also think about changing β, but the consequences are less straightforward. β controls in some sense the role of subleading terms and increasing its value would make the results more dependent on the specific values of the coefficients c_i. This can in principle influence the results, however, we checked that the values of β allowed by EWPTs are too small to make this effect relevant.

Large mass range ($m_\eta \approx 400$ GeV)

The last region of interest is for $m_\eta \approx 400\,\text{GeV}$. The existence of this region is due to a cancellation between the terms with the exchanges of h and H_0 in the cross section determining the relic density. This makes this region interesting, as its very presence is a consequence of the richness of our model. Unfortunately, the correct relic abundance can only be reproduced in the loosely natural range of the c_i coefficients. Furthermore, our benchmark point is already excluded by current DD constraints. We could try different, allowed values of ξ and β and see what happens. Changing ξ shifts the part of the parameter space excluded by DD and in order to evade these limits one would need $\xi \lesssim 0.01$. We do not study this region further given the high fine tuning required.

3.5 Non-thermal dark matter production

Now we move to the other part of the parameter space, the one for which $\theta_t \gtrsim \pi/4$. Here, given the large number of pseudo-Goldstones and their sprectrum, another possibility is present. It could be possible indeed that the production of the DM candidate η is caused by the late time decay of another pNGB particle.

By looking at fig. 3.2, we see that the particle which could possibly decay is κ, given that it can be very close in mass to η. Being close in mass suppresses the available phase space and guarantees a late time decay.

In this scenario, the η relic density receives two contributions. η will still freeze-out in the usual way, so this thermal term will have to be taken into account. The second term will come from the freeze-out of κ, provided it is sufficiently long-lived. Once κ and η have frozen out, it will be possible for κ to decay into η, thus increasing the DM abundance. In the decay, for each particle κ we have the generation of one particle η, so the number-abundance of κ is completely converted into the one of η. The total DM relic density is thus given by (see e.g. Ref. [120]):

$$\Omega_{\text{DM}} h^2 = \Omega_\eta h^2 + \frac{m_\eta}{m_\kappa} \Omega_\kappa h^2 \,. \tag{3.49}$$

In the previous part, the one where we discussed the thermal scenario, we stated that for the relic calculation we only cared about the scattering into SM particles, since the one into the other pNGB was not allowed at freeze-out energies. Instead, here we have that H_0, A_0, H_+ are all lighter than η (cf. fig. 3.2), so they also have to be included as final states of η- and κ-annihilations when calculating the relic density. As in the standard co-annihilation paradigm discussed in section 2.2.2, κ will decay into η (though at a later time and not during freeze-out), so we only care about the sum of the two densities and hence conversion processes are not relevant. The cross sections for the processes leading to the freeze-out of κ can be easily computed by looking at the cross sections for the freeze-out of η and changing the effective coupling and the masses appropriately. The Lagrangians given in eq. (3.42) and eqs. (3.45a) to (3.45c) can guide the conversion of η cross sections into κ cross sections.

By looking at the partial compositeness Lagrangian, we understand that κ can decay into an η and a quark-antiquark couple. The decay width for such a process is:

$$\Gamma_{\kappa\to\eta q\bar{q}} = \frac{3}{32\pi^3 m_\kappa}\frac{m_q^2}{v^4}|g_{\eta\kappa q}|^2\int_{m_\eta}^{\frac{m_\kappa^2+m_\eta^2-4m_q^2}{2m_\kappa}} dq_0\sqrt{q_0^2-m_\eta^2}\,(m_\kappa^2+m_\eta^2-2m_\kappa q_0)$$
$$\sqrt{1-\frac{4m_q^2}{m_\kappa^2+m_\eta^2-2m_\kappa q_0}}\,. \tag{3.50}$$

We see that the decay width is proportional to m_q^2, this means that the decay into heavier particles will be much faster with respect to lighter ones. This is important, because this decay must occur in a specific time window. For this non-thermal contribution to be present, we need the decay to occur later than the freeze-out time, otherwise we would just have thermal terms. Furthermore, we need to respect the strong constraints from Big Bang nucleosynthesis (BBN) and a decay at a later time would imply an entropy injection which will not be consistent with BBN history. This latter constraint is actually not restrictive for our parameter space, since our decays usually occur earlier than BBN time, even being cautious and setting the bound $1/\Gamma < 1\,\mathrm{s} < t_{\mathrm{BBN}}$. So let us instead check that we decay later than the freeze-out time t_F:

$$\frac{t_F}{1\,\mathrm{s}} \approx \frac{1.5^2}{\sqrt{g_{*,EU}}}\left(\frac{1\,\mathrm{MeV}}{m_\eta}\right)^2 x_F^2\,, \tag{3.51}$$

where as always $g_{*,EU}\approx 100$ is the effective number of relativistic species in the early universe and the freeze-out time is approximated by $x_F \approx 25$.

We need to ensure that our DM partner κ decays at a later time, but we also know that $\Gamma_{\kappa\to\eta q\bar{q}} \propto m_q^2$, so if $\Delta m_{\kappa,\eta} \equiv m_\kappa - m_\eta > 2m_t$, the decay into $t\bar{t}$ is allowed and extremely quick. If we were to allow the decay into a $t\bar{t}$ couple, then we would need to highly fine tune the coefficients in order to suppress $g_{\eta\kappa t}$. A more natural approach is requiring that η and κ are so close in mass that the decay into top-antitop is not allowed. Of course, also in this case there is some degree of fine tuning, but less so than by killing the effective coupling. On the other hand, in the case where $\Delta m_{\kappa,\eta} < 2m_t$, the two contributions to the relic density are of the same order, i.e. $\Omega_\eta h^2 \approx \Omega_\kappa h^2$. This is because η and κ undergo similar processes and they also share very similar masses. A very nice feature of the thermal case was that relic density at low fine tuning was reproduced exactly where DD cancellations could occur, unfortunately that is not the case here. In order to reproduce the relic density and avoid DD constraint, an higher fine tuning (and hence a smaller ξ) is required. We see that the phenomenology of the model has drastically changed near $\theta_t = \pi/4$.

We show in fig. 3.7 the results in the m_η-$g_{\eta h}$ plane, as we did for the thermal case, even if here $g_{\eta h}$ loses part of its importance. We choose a benchmark point that is not excluded at the moment, so we have $\xi = 0.01$, $\beta = 0.2$, $y_L = y_R = 1$, $g_* = 3$ and loosely natural coefficients. Here, differently from the thermal case, there are no good points for strictly natural coefficients. We see that DD excludes a portion of the parameter space for the benchmark, as a matter of fact, it is the search forcing us to lower the value of ξ. Conversely, current bounds from ID and LHC are not relevant here. The allowed mass-splitting range is $20\,\mathrm{GeV} \lesssim \Delta m_{\kappa,\eta} \lesssim 50\,\mathrm{GeV}$, meaning that either we justify such a small mass splitting or

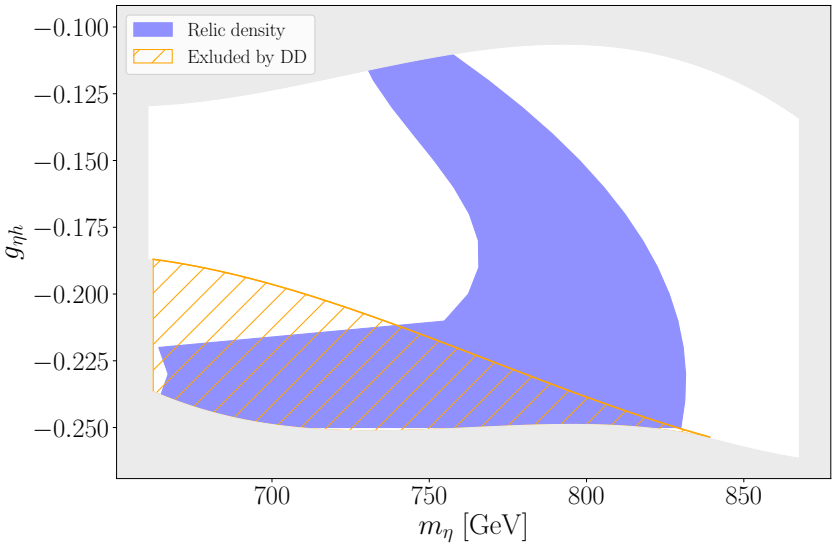

Figure 3.7: Relic density from non-thermal DM production for the benchmark point $\xi = 0.01$, $\beta = 0.2$, $y_L = y_R = 1$, $g_* = 3$ and loosely natural coefficients c_i. The blue region is the 3σ DM relic density contour. Areas excluded by DD are represented by the orange hatched region, while regions of the parameter space that cannot be accessed with natural c_is are denoted by the gray region. No further constraint on the model is relevant in this part of the parameter space.

we need some degree of fine tuning. The masses of η and κ are close in the region $\theta_t \gtrsim \pi/4$, but that is not enough to completely justify these mass splittings. Higher fine tuning will be required for larger masses of η and κ, since we require mass splittings of this order even if the involved particles get heavier.

In the end, in this part of the parameter space we have a new intriguing effect given by the non-thermal production mechanism. Unfortunately, for our model the available parameter space is in a fine tuned region. However, this is not a general result: other models may have this non-thermal production via pNGB decay, but it might be the case that this effect is present in a region of low fine tuning. Enlarging our coset has provided us with a greater level of complexity with respect to the minimal case, but this also enriched the spectrum and delivered interesting results for DM phenomenology.

Appendices

3.A Generators

We write here our generators explicitly for the breaking pattern $\mathrm{SO}(7) \to \mathrm{SO}(5)' \times \mathrm{SO}(2)'$. Proceeding similarly to other SO(N) models, we define them as:

$$(T_L^\alpha)_{ab} = -\frac{i}{2}\left[\epsilon_{\alpha\beta\gamma}\,\delta_{\beta a}\,\delta_{\gamma b} + (\delta_{\alpha a}\,\delta_{4b} - \delta_{\alpha b}\,\delta_{4a})\right] \tag{3.52a}$$

$$(T_R^\alpha)_{ab} = -\frac{i}{2}\left[\epsilon_{\alpha\beta\gamma}\,\delta_{\beta a}\,\delta_{\gamma b} - (\delta_{\alpha a}\,\delta_{4b} - \delta_{\alpha b}\,\delta_{4a})\right] \tag{3.52b}$$

$$(T_5^\omega)_{ab} = -\frac{i}{\sqrt{2}}\left[\delta_{\omega a}\,\delta_{5b} - \delta_{5a}\,\delta_{\omega b}\right] \tag{3.52c}$$

$$(T_2)_{ab} = -\frac{i}{\sqrt{2}}\left[\delta_{6a}\,\delta_{7b} - \delta_{7a}\,\delta_{6b}\right] \tag{3.52d}$$

$$\left(\hat{T}_1^i\right)_{ab} = -\frac{i}{\sqrt{2}}\left[\delta_{ia}\,\delta_{6b} - \delta_{6a}\,\delta_{ib}\right] \tag{3.52e}$$

$$\left(\hat{T}_2^i\right)_{ab} = -\frac{i}{\sqrt{2}}\left[\delta_{ia}\,\delta_{7b} - \delta_{7a}\,\delta_{ib}\right] \tag{3.52f}$$

where $\alpha, \beta, \gamma = 1, 2, 3$, $a, b = 1, \ldots, 7$, $\omega = 1, \ldots, 4$ and $i = 1, \ldots, 5$. $T_{L,R}$ are the generators of $\mathrm{SO}(4)' \subset \mathrm{SO}(5)'$, T_5 are the four remaining generators of $\mathrm{SO}(5)'$ that are not in $\mathrm{SO}(4)'$ and T_2 is the generator of $\mathrm{SO}(2)'$. We generally denote with a "hat" the broken generators, which in this case are the ten $\hat{T}_{1,2}$.

We can write down some commutation relations for our generators. These are very useful when we have to decompose terms in the CCWZ Lagrangian:

$$\begin{aligned}
&[\hat{T}_1^i, \hat{T}_1^j] = \left(t^A\right)_{ij} T^A = [\hat{T}_2^i, \hat{T}_2^j] \quad , \quad [\hat{T}_1^i, \hat{T}_2^j] = -\frac{i}{\sqrt{2}}\,\delta_{ij} T_2 \quad , \quad [T^A, T_2] = 0 \quad , \\
&[\hat{T}_{1,2}^i, T^A] = \left(t^A\right)_{ij} \hat{T}_{1,2}^j \quad , \quad [\hat{T}_1^i, T_2] = \frac{i}{\sqrt{2}} \hat{T}_2^i \quad , \quad [\hat{T}_2^i, T_2] = -\frac{i}{\sqrt{2}} \hat{T}_1^i \,,
\end{aligned} \tag{3.53}$$

here we have defined $T^A \equiv \{T_L, T_R, T_5\}$ and t^A as the upper 5×5 block of T^A. Other useful

relations are:

$$[T^\alpha_{L,R}, T^\beta_{L,R}] = i\epsilon_{\alpha\beta\gamma} T^\gamma_{L,R} \quad , \quad [T^\alpha_L, T^4_5] = -\frac{i}{2} T^\alpha_5 \quad , \quad [T^\alpha_R, T^4_5] = \frac{i}{2} T^\alpha_5 \quad ,$$

$$[T^\alpha_L, T^\beta_5] = \frac{i}{2}\left(\delta_{\alpha\beta} T^4_5 + \epsilon_{\alpha\beta\gamma} T^\gamma_5\right) \quad , \quad [T^\alpha_R, T^\beta_5] = \frac{i}{2}\left(-\delta_{\alpha\beta} T^4_5 + \epsilon_{\alpha\beta\gamma} T^\gamma_5\right) \quad , \tag{3.54}$$

$$[T^{\omega_1}_5, T^{\omega_2}_5] = \frac{i}{2}\epsilon_{\omega_1\omega_2\omega_3}\left(T^{\omega_3}_L + T^{\omega_3}_R\right) \quad , \quad [T^\alpha_L, T^\beta_R] = 0\,.$$

3.B Higgs couplings fit and EWPTs

In CH models it is important to consider the limits coming from EWPTs and from Higgs physics.

First, we consider the limits coming from Higgs measurements. Here, we use the global Higgs coupling analysis published by ATLAS with 80 fb^{-1} of luminosity [121]. In our model the couplings of the Higgs particle to the EW gauge bosons and to the quarks are modified and we will consider the limits on them in the κ framework [122]. In particular, the most model independent modified couplings we can look at are $\{\kappa_{Zg}, \lambda_{tg}, \lambda_{Zg}, \lambda_{WZ}, \lambda_{\gamma Z}, \lambda_{\tau Z}, \lambda_{bZ}\}$. We can express these modifications in terms of our couplings, once we make some reasonable assumption for some subleading couplings. In particular, we assume the same modification for different particles: $k_c = k_t$ and $k_b = k_s = k_\tau = k_\mu$. The effective couplings of interest for these limits are $g_{W,Z}$ and $k_{t,b}$. These were introduced in the Lagrangian terms in eqs. (3.45a) and (3.42), while their values in our model are reported in section 3.C and they only depend on $\xi, \beta, \theta_t, \theta_b$. We have set $\theta_b = 0$ in our phenomenological treatment and we will do the same here.

Now that we have an expression of these modified couplings in terms of ξ, β, θ_t, we can set limits by looking at the experimental limits on these modifications reported in [121]. These modifications are not independent of each other, so it will be necessary to consider their cross-correlations. In this way we obtain the plot on the left of fig. 3.8.

For the EWPT, we use the combined limits on $S = 4s^2_W \hat{S}/\alpha$ and $T = \hat{T}/\alpha$ from GFitter [123]. Except from the contribution to $\hat{T}$ deriving from the small custodial breaking of our model, the contributions to $\hat{S}$ and $\hat{T}$ are those common to every CH model [124]. So we have:

$$(\Delta\hat{S})_{\rm IR} \approx \frac{g^2}{192\pi^2}\xi \log\frac{\Lambda^2}{m_h^2} \ , \quad (\Delta\hat{T})_{\rm IR} \approx -\frac{3g^2}{64\pi^2}\tan^2\theta_W \xi \log\frac{\Lambda^2}{m_h^2} \ , \tag{3.55a}$$

$$(\Delta\hat{S})_{\rm UV} \approx \frac{m_W^2}{\Lambda^2} \ , \quad (\Delta\hat{T})_{\rm UV} \approx 0 \ , \tag{3.55b}$$

$$(\Delta\hat{S})_{\rm 2HDM} = 0 \ , \quad (\Delta\hat{T})_{\rm 2HDM} = \frac{\xi}{4}\left(1 - \cos 4\beta\right) \ , \tag{3.55c}$$

where for us $\Lambda = m_* = g_* f$ and we take $g_* = 3$, as it is our benchmark value in the thermal case. With these contributions we get the central plot of fig. 3.8.

Let us briefly remind the origin of each of these terms. In the SM, the coupling of the gauge bosons to the Higgs field are such that the loop contributions for $\hat{S}$ and $\hat{T}$ are not present. This is due to the cancellation between the terms with the Higgs in the loop and the terms with the would-be Goldstones. However, in CH models the couplings of h to the gauge

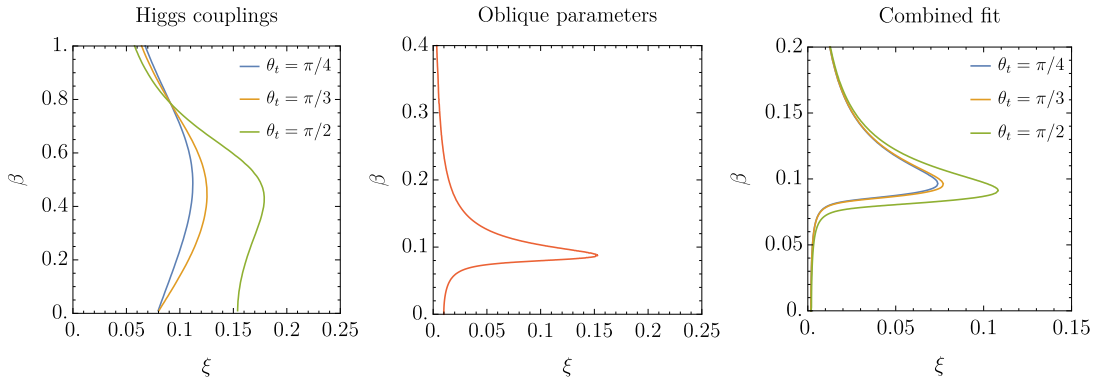

Figure 3.8: *Left*: 95%CL exclusion limits as a function of ξ and β from the Higgs data for different values of θ_t. *Center:* 95%CL exclusion limits from EWPT. *Right*: Higgs + EWPT combined 95% exclusion limits for different values of θ_t.

bosons are modified and these loops do give a contribution, which is the one we denoted with IR. The terms we denoted with UV come instead from the mixing of the SM gauge bosons and the vector resonances of the composite sector. The parameter $\hat{T}$ is protected at tree-level from these contributions, while that is not the case for $\hat{S}$. Finally, in our case the CP structure of the 2HDM allows the breaking of custodial symmetry, giving a positive contribution to $\hat{T}$.

By combining the result of Higgs and EWPT we get the final exclusion plot on the right of fig. 3.8. This combined fit has been obtained by performing a combined χ^2 analysis. We see that a value of $\beta \simeq 0.1$ is selected and this is mainly due to EWPT. In order to understand why this specific value is selected, let us look at the different contributions to the oblique parameters. We represented this result in fig. 3.9.

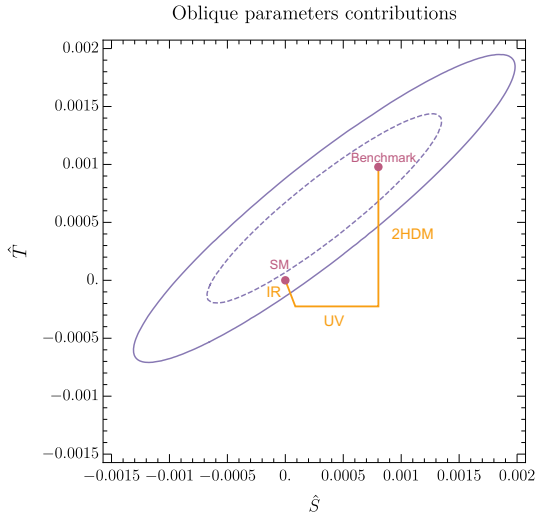

Figure 3.9: *Left*: 95% (68%) exclusion limits on the $\hat{S}$ and $\hat{T}$ parameter are represented by the solid (dashed) purple line. The SM result $(0, 0)$ is more than 1σ from the best fit value. Our benchmark point ($\xi = 0.061$, $\beta = 0.1$) is portrayed as well and falls within the 1σ band.

We see that we can relax the EWPT thanks to the small breaking of custodial symmetry specific to our model. In the plot we can see indeed that if this term was not present, we would be outside the best fit by more than two σs, while for $\beta = 0.1$ we fall inside the 1σ band.

3.C Expression of the effective couplings

The expressions for the effective couplings can be rather involved, but they can be simplified by introducing a new set of coefficients. This new set incorporates in their definition the relevant coupling pre-factors:

$$\begin{aligned}
&\tilde{c}_y^{(1)} \equiv N_c\, c_{(1,0)}^{(1)}\, y_L^2 && \tilde{c}_y^{(2)} \equiv N_c\, c_{(0,1)}^{(1)}\, y_R^2 && \tilde{c}_y^{(3)} \equiv N_c\, c_{(2,0)}^{(1)}\, y_L^4/g_*^2 \\
&\tilde{c}_y^{(4)} \equiv N_c\, c_{(1,1)}^{(1)}\, y_L^2 y_R^2/g_*^2 && \tilde{c}_y^{(5)} \equiv N_c\, c_{(0,2)}^{(1)}\, y_R^4/g_*^2 && \tilde{c}_y^{(6)} \equiv N_c\, c_{(2,0)}^{(2)}\, y_L^4/(4\pi)^2 \\
&\tilde{c}_y^{(7)} \equiv N_c\, c_{(1,1)}^{(2)}\, y_L^2 y_R^2/(4\pi)^2 && \tilde{c}_y^{(8)} \equiv N_c\, c_{(0,2)}^{(2)}\, y_R^4/(4\pi)^2 && \tilde{c}_y^{(9)} \equiv N_c\, c_{(0,2)}^{(3)}\, y_R^4/(4\pi)^2
\end{aligned}$$

$$\begin{aligned}
\tilde{c}_g^{(1)} &\equiv c_{g'}^{(1)}\, g'^2 \quad \tilde{c}_g^{(2)} \equiv c_{g'}^{(2)}\, g'^2 \\
\tilde{c}_g^{(3)} &\equiv c_g^{(1)}\, g^2 \quad \tilde{c}_g^{(4)} \equiv c_{g'}^{(2)}\, g^2
\end{aligned}$$

In the following, we provide the list of all the effective couplings used in the computations with their expressions in term of this new coefficients. We listed the effective couplings in different sections, depending on the particles involved in the interaction.

3.C.1 Interactions between NGBs and gauge bosons

The first set of couplings is generated by CCWZ and involves the massive gauge bosons:

$$g_V \approx \sqrt{1-\xi}, \qquad b_h \approx 1-2\xi, \tag{3.56a}$$

$$\lambda_\eta^{(V)} \approx 2\xi, \qquad \lambda_\kappa^{(V)} \approx -\xi\beta^2, \tag{3.56b}$$

$$g_{H_0 W} \approx -\frac{\beta\xi}{2}, \qquad g_{H_0 Z} \approx \frac{3\beta\xi}{2}, \qquad g_{H_+ V} \approx \xi\beta. \tag{3.56c}$$

g_V is particularly important here, since it enters the constraint from Higgs physics. As mentioned in the text, LHC production of the Higgs particle is dominated by gluon fusion. Such a coupling to gluons is generated by top-loops, in our model other pNGBs are coupled to the gluons in the same way:

$$g_{gh} = -i\,\frac{\alpha_S}{8\pi}\,\tau_h\left[1+(1-\tau_h)f(\tau_h)\right], \qquad \tau_h = \frac{4m_t^2}{m_h^2}, \tag{3.56d}$$

$$g_{gH_0} = -i\,k_{H_0 t}\frac{\alpha_S}{8\pi}\tau_{H_0}[1+(1-\tau_{H_0})f(\tau_{H_0})], \qquad \tau_{H_0} = \frac{4m_t^2}{m_{H_0}^2}, \tag{3.56e}$$

$$g_{g\eta} = -i\,g_t\frac{\alpha_S}{8\pi}\,\tau_\eta[1+(1-\tau_\eta)f(\tau_\eta)], \qquad \tau_\eta = \frac{m_t^2}{m_\eta^2}. \tag{3.56f}$$

where $f(\tau_X)$ is the usual function appearing for gluon effective couplings (see for instance eq. (1.198) of Ref.[125]).

Notice that the coupling of η and H_0 to quarks has a factor ξ, so also the coupling to gluons will receive this suppression. The gluon fusion production of η and H_0 will not be relevant at LHC, but it has been included in our simulations.

3.C.2 Interactions between NGBs and fermions

Interactions with quarks are generated by the partial compositeness Lagrangian. In writing this Lagrangian we had to choose the embedding of the fermions, different choices would lead to different expressions. For our model, with fermions in the fundamental representation, we have:

$$k_q \approx 1 - \frac{7}{6}\xi - \frac{\xi}{3}\frac{\cos(3\beta + \alpha_q\theta_q)}{\cos(\beta - \alpha_q\theta_q)}\,, \tag{3.57a}$$

$$g_q \approx -2\xi\frac{\cos\beta\cos\theta_q}{\cos(\beta - \alpha_q\theta_q)}\,, \tag{3.57b}$$

$$k_{H_0 q} \approx \frac{2\xi\sin(4\beta) + (-6+\xi)\sin(2\beta - 2\alpha_q\theta_q) + 4\xi\sin(2\beta + 2\alpha_q\theta_q)}{12\cos^2(\beta - \alpha_q\theta_q)}\,, \tag{3.57c}$$

$$g_{\kappa q} \approx -2\alpha_q\xi\frac{\sin\beta\sin\theta_q}{\cos(\beta - \alpha_q\theta_q)}\,, \tag{3.57d}$$

$$g_{\eta\kappa q} \approx -i\alpha_q\xi\tan(\beta - \alpha_q\theta_q)\,. \tag{3.57e}$$

with $\alpha_q = 1$ (-1) for quarks with charge $2/3$ $(-1/3)$.

3.C.3 Interactions among NGBs

Since the potential has been generated using an expansion of the NGB matrix, we list here only the relevant orders in ξ and β:

$$\begin{aligned} g_{\eta h} \approx & -\frac{g_*^2}{8\pi^2}\cos^2\theta_t\left[2\tilde{c}_y^{(5)} - \tilde{c}_y^{(7)} + 2\tilde{c}_y^{(8)} + 2\cos(2\theta_t)(\tilde{c}_y^{(5)} + \tilde{c}_y^{(8)})\right] \\ & + \frac{g_*^2\beta}{4\pi^2}\cot\theta_t\cos(2\theta_t)(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)}) \\ & - \frac{g_*^2\beta^2}{8\pi^2}\cot^2\theta_t\left[\tilde{c}_y^{(5)} - 2\tilde{c}_y^{(8)} - 2\cos(2\theta_t)(\tilde{c}_y^{(2)} + \tilde{c}_y^{(5)} + 2\tilde{c}_y^{(7)} - \tilde{c}_y^{(8)}) + \tilde{c}_y^{(5)}\cos(4\theta_t)\right] \end{aligned} \tag{3.58a}$$

$$\begin{aligned} \lambda_{\eta h} \approx & \frac{g_*^2}{24\pi^2}\cos^2\theta_t\left[2\tilde{c}_y^{(2)} - 6\tilde{c}_y^{(5)} + 7\tilde{c}_y^{(7)} - 6\tilde{c}_y^{(8)} - 6\cos(2\theta_t)(\tilde{c}_y^{(5)} + \tilde{c}_y^{(8)})\right] \\ & + \frac{g_*^2\beta}{4\pi^2}\cot\theta_t\cos(2\theta_t)(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)}) \\ & - \frac{g_*^2\beta^2}{24\pi^2}\cot^2\theta_t\left[\tilde{c}_y^{(2)} + 3\tilde{c}_y^{(5)} + 2\tilde{c}_y^{(7)} - 6\tilde{c}_y^{(8)} - \cos(2\theta_t)(7\tilde{c}_y^{(2)} + 6\tilde{c}_y^{(5)} + 14\tilde{c}_y^{(7)} - 6\tilde{c}_y^{(8)})\right. \end{aligned}$$

$$+ 3\tilde{c}_y^{(5)} \cos(4\theta_t)\Big] \tag{3.58b}$$

$$\begin{aligned}\lambda \approx 1 - \beta^2 & \\ - \frac{g_*^2\beta^2}{16\pi^2}\frac{v^2}{m_h^2}\Big[&\tilde{c}_g^{(1)} + \tilde{c}_g^{(2)} + \tilde{c}_g^{(3)} + \tilde{c}_g^{(4)} - 16\tilde{c}_y^{(2)} + \tilde{c}_y^{(5)} - 32\tilde{c}_y^{(7)} + 2\tilde{c}_y^{(8)} \\ &- \cos(2\theta_t)(14\tilde{c}_y^{(2)} - 2\tilde{c}_y^{(5)} + 29\tilde{c}_y^{(7)} - 2\tilde{c}_y^{(8)}) \\ &+ 8\csc^2\theta_t(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)}) + \tilde{c}_y^{(5)}\cos(4\theta_t)\Big]\end{aligned} \tag{3.58c}$$

$$\begin{aligned}g_{H_0} \approx \frac{m_h^2}{v^2}\left(1-\beta^2\right) & \\ - \frac{g_*^2}{8\pi^2}\big[&\tilde{c}_g^{(1)} + \tilde{c}_g^{(2)} + \tilde{c}_g^{(3)} + \tilde{c}_g^{(4)} + \tilde{c}_y^{(5)} + 2\tilde{c}_y^{(8)} + \cos(2\theta_t)(2\tilde{c}_y^{(5)} - \tilde{c}_y^{(7)} + 2\tilde{c}_y^{(8)}) \\ &+ \tilde{c}_y^{(5)}\cos(4\theta_t)\big] \\ + \frac{g_*^2\beta}{4\pi^2}\csc\theta_t[\cos\theta_t + \cos(3\theta_t)]&(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)}) \\ + \frac{g_*^2\beta^2}{16\pi^2}\Big[&15(\tilde{c}_g^{(1)} + \tilde{c}_g^{(2)} + \tilde{c}_g^{(3)} + \tilde{c}_g^{(4)} + \tilde{c}_y^{(5)} + 2\tilde{c}_y^{(8)}) \\ &+ \cos(2\theta_t)(14\tilde{c}_y^{(2)} + 6\tilde{c}_y^{(5)} + 25\tilde{c}_y^{(}6\tilde{c}_y^{(8)}) + 15\tilde{c}_y^{(5)}\cos(4\theta_t)\Big]\end{aligned} \tag{3.58d}$$

$$\begin{aligned}g_{H_0hh} \approx \frac{g_*^2\beta}{8\pi^2}\Big[&3\left[\tilde{c}_g^{(1)} + \tilde{c}_g^{(2)} + \tilde{c}_g^{(3)} + \tilde{c}_g^{(4)} + \tilde{c}_y^{(5)} + 2\tilde{c}_y^{(8)}\right] + \cos(2\theta_t)(4\tilde{c}_y^{(2)} + 6\tilde{c}_y^{(5)} + 5\tilde{c}_y^{(7)} + 6\tilde{c}_y^{(8)}) \\ &+ 3\tilde{c}_y^{(5)}\cos(4\theta_t)\Big] \\ - \frac{g_*^2\beta^2}{2\pi^2}\cot(2\theta_t)&[\cos\theta_t + \cos(3\theta_t)](\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)})\end{aligned} \tag{3.58e}$$

$$\begin{aligned}\lambda_{\eta H_0} \approx - \frac{g_*^2}{8\pi^2}\cos^2\theta_t&\left[2\tilde{c}_y^{(5)} - \tilde{c}_y^{(7)} - 2\tilde{c}_y^{(8)} - 2\cos(2\theta_t)(\tilde{c}_y^{(5)} - \tilde{c}_y^{(8)})\right] \\ + \frac{g_*^2\beta^2}{24\pi^2}\Big[&\tilde{c}_y^{(2)} - 3\tilde{c}_y^{(5)} + 2\tilde{c}_y^{(7)} - 6\tilde{c}_y^{(8)} - \cos(2\theta_t)\left(7\tilde{c}_y^{(2)} + 6\tilde{c}_y^{(5)} + 14\tilde{c}_y^{(7)} + 6\tilde{c}_y^{(8)}\right) \\ &- 3\tilde{c}_y^{(5)}\cos(4\theta_t)\Big]\end{aligned} \tag{3.58f}$$

$$\begin{aligned}\lambda_{H_0} \approx - \frac{3g_*^2\beta}{8\pi^2}&\left[\tilde{c}_g^{(1)} + \tilde{c}_g^{(2)} + \tilde{c}_g^{(3)} + \tilde{c}_g^{(4)} + \tilde{c}_y^{(5)} + 2\tilde{c}_y^{(8)} - \cos(2\theta_t)(2\tilde{c}_y^{(5)} - \tilde{c}_y^{(7)} + 2\tilde{c}_y^{(8)}) + \tilde{c}_y^{(5)}\cos(4\theta_t)\right] \\ + \frac{3g_*^2\beta^2}{2\pi^2}&\tan\theta_t\cos(2\theta_t)(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)})\end{aligned} \tag{3.58g}$$

$$g_{\eta H_0} \approx \frac{g_*^2\beta}{8\pi^2}\left[\tilde{c}_y^{(5)} + 2\tilde{c}_y^{(8)} + 2\cos(2\theta_t)(\tilde{c}_y^{(2)} + \tilde{c}_y^{(5)} + 2\tilde{c}_y^{(7)} + \tilde{c}_y^{(8)}) + \tilde{c}_y^{(5)}\cos(4\theta_t)\right]$$
$$- \frac{g_*^2\beta^2}{4\pi^2}\tan\theta_t\cos(2\theta_t)(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)}) \tag{3.58h}$$

$$g_{A_0 h} \approx \frac{g_*^2}{8\pi^2}\left[-\frac{8\pi^2}{g_*^2}\frac{m_h^2}{v^2} + \tilde{c}_y^{(5)} + \tilde{c}_y^{(8)} + \cos(2\theta_t)(2\tilde{c}_y^{(5)} - \tilde{c}_y^{(7)} + 2\tilde{c}_y^{(8)}) + \cos(4\theta_t)(\tilde{c}_y^{(5)} + \tilde{c}_y^{(8)})\right]$$
$$- \frac{g_*^2\beta}{4\pi^2}\csc\theta_t\sec\theta_t(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)})\cos^2(2\theta_t)$$
$$- \frac{g_*^2\beta^2}{16\pi^2}\Big[-\frac{16\pi^2}{g_*^2}\frac{m_h^2}{v^2} + 7\tilde{c}_g^{(1)} + 7\tilde{c}_g^{(2)} + 7\tilde{c}_g^{(3)} + 7\tilde{c}_g^{(4)} + 16\tilde{c}_y^{(2)} + 7\tilde{c}_y^{(5)} + 32\tilde{c}_y^{(7)} + 14\tilde{c}_y^{(8)}$$
$$+ \cos(2\theta_t)(14\tilde{c}_y^{(2)} + 6\tilde{c}_y^{(5)} + 25\tilde{c}_y^{(7)} + 6\tilde{c}_y^{(8)})$$
$$- 16\csc^2(2\theta_t)(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)}) + 7\tilde{c}_y^{(5)}\cos(4\theta_t)\Big] \tag{3.58i}$$

$$g_{A_0 H_0} \approx -\frac{g_*^2\beta}{8\pi^2}\left[\tilde{c}_g^{(1)} + \tilde{c}_g^{(2)} + \tilde{c}_g^{(3)} + \tilde{c}_g^{(4)} + \tilde{c}_y^{(5)} + 2\tilde{c}_y^{(8)} - \cos(2\theta_t)(2\tilde{c}_y^{(5)} - \tilde{c}_y^{(7)} + 2\tilde{c}_y^{(8)}) + \tilde{c}_y^{(5)}\cos(4\theta_t)\right]$$
$$+ \frac{g_*^2\beta^2}{2\pi^2}\tan\theta_t\cos(2\theta_t)(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)}) \tag{3.58j}$$

$$g_{\kappa h} \approx \frac{g_*^2}{8\pi^2}\sin^2\theta_t\left[\tilde{c}_y^{(7)} - 4(\tilde{c}_y^{(5)} - \tilde{c}_y^{(8)})\cos^2\theta_t\right]$$
$$- \frac{g_*^2\beta^2}{8\pi^2}\left[\tilde{c}_y^{(5)} + 2\tilde{c}_y^{(8)} - 2\cos(2\theta_t)(\tilde{c}_y^{(2)} + \tilde{c}_y^{(5)} + 2\tilde{c}_y^{(7)} + \tilde{c}_y^{(8)}) + \tilde{c}_y^{(5)}\cos(4\theta_t)\right] \tag{3.58k}$$

$$g_{\kappa H_0} \approx -\frac{g_*^2\beta}{2\pi^2}\sin^2\theta_t(\tilde{c}_y^{(8)} - \tilde{c}_y^{(5)}\cos(2\theta_t)) + \frac{g_*^2\beta^2}{4\pi^2}\tan\theta_t\cos(2\theta_t)(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)}) \tag{3.58l}$$

$$\lambda_{\kappa h} \approx \frac{g_*^2}{8\pi^2}\sin^2\theta_t(\tilde{c}_y^{(7)} - 4\cos^2\theta_t(\tilde{c}_y^{(5)} - \tilde{c}_y^{(8)}))$$
$$+ \frac{g_*^2\beta^2}{24\pi^2}\Big[\tilde{c}_y^{(2)} - 3\tilde{c}_y^{(5)} + 2\tilde{c}_y^{(7)} - 6\tilde{c}_y^{(8)} + \cos(2\theta_t)(5\tilde{c}_y^{(2)} + 6\tilde{c}_y^{(5)} + 10\tilde{c}_y^{(7)} + 6\tilde{c}_y^{(8)})$$
$$- 3\tilde{c}_y^{(5)}\cos(4\theta_t)\Big] \tag{3.58m}$$

$$\lambda_{\kappa H_0} \approx -\frac{g_*^2}{24\pi^2}\left[2\left(2\tilde{c}_y^{(7)} + \tilde{c}_y^{(2)}\right)\cos^2\theta_t - \sin^2\theta_t\left(4\tilde{c}_y^{(2)} + 11\tilde{c}_y^{(7)}\right) + 12\sin^4\theta_t\left(\tilde{c}_y^{(5)} + \tilde{c}_y^{(8)}\right)\right]$$
$$+ \frac{g_*^2\beta}{24\pi^2}[3 - 5\cos(2\theta_t)]\csc\theta_t\sec\theta_t\cos(2\theta_t)(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)})$$
$$- \frac{g_*^2\beta^2}{24\pi^2}\Big[13\tilde{c}_y^{(2)} - 3\tilde{c}_y^{(5)} + 26\tilde{c}_y^{(7)} - 6\tilde{c}_y^{(8)} - \cos(2\theta_t)(13\tilde{c}_y^{(2)} - 3\tilde{c}_y^{(5)} + 26\tilde{c}_y^{(7)} - 6\tilde{c}_y^{(8)})$$

$$+ 2(\csc^2\theta_t - 4\sec^2\theta_t)(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)})\Big] \tag{3.58n}$$

$$\lambda_{\eta A_0} \approx \frac{g_*^2}{8\pi^2}\cos^2\theta_t\left[2\tilde{c}_y^{(5)} - \tilde{c}_y^{(7)} - 2\tilde{c}_y^{(8)} - 2\cos(2\theta_t)(\tilde{c}_y^{(5)} - \tilde{c}_y^{(8)})\right]$$
$$- \frac{\beta^2 g_*^2}{24\pi^2}\Big[\tilde{c}_y^{(2)} - 3\tilde{c}_y^{(5)} + 2\tilde{c}_y^{(7)} - 6\tilde{c}_y^{(8)} - \cos(2\theta_t)(7\tilde{c}_y^{(2)} + 6\tilde{c}_y^{(5)} + 14\tilde{c}_y^{(7)} + 6\tilde{c}_y^{(8)})$$
$$-3\tilde{c}_y^{(5)}\cos(4\theta_t)\Big] \tag{3.58o}$$

$$\lambda_{\eta H_+} \approx \frac{g_*^2}{8\pi^2}\cos^2\theta_t\left[2\tilde{c}_y^{(5)} - \tilde{c}_y^{(7)} - 2\tilde{c}_y^{(8)} - 2\cos(2\theta_t)(\tilde{c}_y^{(5)} - \tilde{c}_y^{(8)})\right]$$
$$- \frac{\beta^2 g_*^2}{24\pi^2}\Big[\tilde{c}_y^{(2)} - 3\tilde{c}_y^{(5)} + 2\tilde{c}_y^{(7)} - 6\tilde{c}_y^{(8)} - \cos(2\theta_t)(7\tilde{c}_y^{(2)} + 6\tilde{c}_y^{(5)} + 14\tilde{c}_y^{(7)} + 6\tilde{c}_y^{(8)})$$
$$-3\tilde{c}_y^{(5)}\cos(4\theta_t)\Big] \tag{3.58p}$$

$$g_{H_+h} \approx -\frac{g_*^2}{16\pi^2}\left[\tilde{c}_y^{(5)} + 2\tilde{c}_y^{(6)} - 2\tilde{c}_y^{(7)} - \tilde{c}_y^{(8)} - \cos(4\theta_t)(\tilde{c}_y^{(5)} - \tilde{c}_y^{(8)}) - 2\tilde{c}_g^{(3)} - 2\tilde{c}_g^{(4)}\right]$$
$$- \frac{\beta^2 g_*^2}{4\pi^2}\left[\tilde{c}_y^{(5)} + 2\tilde{c}_y^{(8)} + \tilde{c}_y^{(5)}\cos(4\theta_t) + \tilde{c}_g^{(1)} + \tilde{c}_g^{(2)} + \tilde{c}_g^{(3)} + \tilde{c}_g^{(4)}\right] \tag{3.58q}$$

$$g_{H_+H_0} \approx -\frac{\beta g_*^2}{8\pi^2}\cos^2\theta_t\Big[\tilde{c}_y^{(5)} + 2\tilde{c}_y^{(8)} - \cos(2\theta_t)(2\tilde{c}_y^{(5)} - \tilde{c}_y^{(7)} + 2\tilde{c}_y^{(8)}) + \tilde{c}_y^{(5)}\cos(4\theta_t)$$
$$+ \tilde{c}_g^{(1)} + \tilde{c}_g^{(2)} + \tilde{c}_g^{(3)} + \tilde{c}_g^{(4)}\Big]$$
$$+ \frac{\beta^2 g_*^2}{2\pi^2}(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)})\cos(2\theta_t)\tan\theta_t \tag{3.58r}$$

$$\lambda_{\kappa A_0} \approx \frac{g_*^2}{24\pi^2}\left[2(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)})\cos^2\theta_t - \sin^2\theta_t[4\tilde{c}_y^{(2)} + 11\tilde{c}_y^{(7)} - 12\sin^2\theta_t(\tilde{c}_y^{(5)} + \tilde{c}_y^{(8)})]\right]$$
$$+ \frac{\beta g_*^2}{24\pi^2}(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)})\cos(2\theta_t)[5\cos(2\theta_t) - 3]\csc\theta_t\sec\theta_t$$
$$+ \frac{\beta^2 g_*^2}{24\pi^2}\Big[13\tilde{c}_y^{(2)} - 3\tilde{c}_y^{(5)} + 26\tilde{c}_y^{(7)} - 6\tilde{c}_y^{(8)} - \cos(2\theta_t)(13\tilde{c}_y^{(2)} - 6\tilde{c}_y^{(5)} + 26\tilde{c}_y^{(7)} - 6\tilde{c}_y^{(8)})$$
$$- 3\tilde{c}_y^{(5)}\cos(4\theta_t) + 2(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)})(\csc^2\theta_t - 4\sec^2\theta_t)\Big] \tag{3.58s}$$

$$\lambda_{\kappa H_+} \approx \frac{g_*^2}{24\pi^2}\left[2(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)})\cos^2\theta_t - \sin^2\theta_t[4\tilde{c}_y^{(2)} + 11\tilde{c}_y^{(7)} - 12\sin^2\theta_t(\tilde{c}_y^{(5)} + \tilde{c}_y^{(8)})]\right]$$
$$+ \frac{\beta g_*^2}{24\pi^2}(\tilde{c}_y^{(2)} + 2\tilde{c}_y^{(7)})\cos(2\theta_t)[5\cos(2\theta_t) - 3]\csc\theta_t\sec\theta_t$$

$$+\frac{\beta^2 g_*^2}{24\pi^2}\Big[13\tilde{c}_y^{(2)}-3\tilde{c}_y^{(5)}+26\tilde{c}_y^{(7)}-6\tilde{c}_y^{(8)}-\cos(2\theta_t)(13\tilde{c}_y^{(2)}-6\tilde{c}_y^{(5)}+26\tilde{c}_y^{(7)}-6\tilde{c}_y^{(8)})$$

$$-3\tilde{c}_y^{(5)}\cos(4\theta_t)+2(\tilde{c}_y^{(2)}+2\tilde{c}_y^{(7)})(\csc^2\theta_t-4\sec^2\theta_t)\Big] \tag{3.58t}$$

$$k_{\rm der}=\frac{2\xi}{3} \tag{3.58u}$$

3.D Relic density

We consider here only the processes involved in the thermal freeze-out scenario, so only the ones involving the annihilations of η into SM particles. The final states are then h, W, Z, t and b. However, as noticed for instance in Ref. [126], below the W, Z threshold, the channels $\eta\eta\to WW^*$, ZZ^* can give a sizable contribution. By using the effective couplings in section 3.C, we have for the thermally-averaged cross sections :

$$\langle\sigma v_{\rm rel}\rangle_{\eta\eta\to hh}=\frac{1}{64\pi m_\eta^2}\left|\lambda_{\eta h}+\frac{3g_{\eta h}\lambda\, m_h^2}{4m_\eta^2-m_h^2+im_h\Gamma_h}-\frac{4g_{\eta H_0}g_{H_0hh}\,v^2}{4m_\eta^2-m_{H_0}^2+im_{H_0}\Gamma_{H_0}}-\frac{2g_{\eta h}^2v^2}{m_h^2-2m_\eta^2}\right.$$

$$\left.+\frac{k_{\rm der}(5m_\eta^2-m_h^2)}{v^2}\right|^2\sqrt{1-\frac{m_h^2}{m_\eta^2}} \tag{3.59a}$$

$$\langle\sigma v_{\rm rel}\rangle_{\eta\eta\to VV}=\frac{\alpha_V}{32\pi m_\eta^2}\frac{m_V^4}{v^4}\left|\lambda_\eta^{(V)}+\frac{2g_{\eta h}g_V\,v^2}{4m_\eta^2-m_h^2+im_h\Gamma_h}-\frac{2g_{\eta H_0}g_{H_0V}\,v^2}{4m_\eta^2-m_{H_0}^2}\right|^2$$

$$\left[2+\left(\frac{2m_\eta^2-m_V^2}{m_V^2}\right)^2\right]\sqrt{1-\frac{m_V^2}{m_\eta^2}} \tag{3.59b}$$

$$\langle\sigma v_{\rm rel}\rangle_{\eta\eta\to q\bar{q}}=\frac{3}{4\pi}\frac{m_q^2}{v^4}\left|g_q+\frac{g_{\eta h}k_q\,v^2}{4m_\eta^2-m_h^2+im_h\Gamma_h}-\frac{g_{\eta H_0}k_{H_0q}\,v^2}{4m_\eta^2-m_{H_0}^2}\right|^2\left(1-\frac{m_q^2}{m_\eta^2}\right)^{\frac{3}{2}}, \tag{3.59c}$$

with $\alpha_V=1\,(1/2)$ for W (Z).

When $m_\eta\lesssim m_V$ we have also the process $\eta\eta\to VV^*$ to discuss. In this case we have:

$$\langle\sigma v_{\rm rel}\rangle_{\eta\eta\to VV^*}=\sum_f\frac{k_{(V)}^2N_c^{(f)}}{1536\pi^3m_\eta^2}\frac{m_V^4}{v^4}\left(\lambda_\eta^{(V)}+\frac{2g_{\eta h}g_V\,v^2}{4m_\eta^2-m_h^2}\right)^2F(\varepsilon_V,\zeta_f)\,, \tag{3.60}$$

where $N_c^{(f)}$ is the number of colors of the final state f to which V^* is going. Here there would be also a channel with H_0, but for such a low mass of m_η, we have that H_0 is very heavy and plays no role. The coefficient in front is:

$$F(\varepsilon_V,\zeta_f)=\int_{\varepsilon_V}^{1+\frac{\varepsilon_V^2}{4}-\zeta_f^2}dy\,\frac{\sqrt{y^2-\varepsilon_V^2}}{(1-y)^2}\,\frac{1}{\varepsilon_V^2}\sqrt{1-\frac{4\zeta_f^2}{4-4y+\varepsilon_V^2}}$$

$$\left\{ \left(\tau_{(V)}^2 + \chi_{(V)}^2\right) \left[4y^2 - 12\varepsilon_V^2 y + 8\varepsilon_V^2 + 3\varepsilon_V^4\right] \right.$$
$$+ \frac{2\zeta_f^2}{4 - 4y + \varepsilon_V^2} \left[\tau_{(V)}^2 \left(4y^2 - 12\varepsilon_V^2 y + 8\varepsilon_V^2 + 3\varepsilon_V^4\right) \right.$$
$$\left. \left. + 2\chi_{(V)}^2 \left(2y^2 + 12\varepsilon_V^2 y - 14\varepsilon_V^2 - 3\varepsilon_V^4\right) \right] \right\}. \tag{3.61}$$

Here we have defined some quantities that depend on whether we are considering the W^* or Z^* channel:

$$k_{(V)} = \begin{cases} \frac{g}{2\sqrt{2}} \\ \frac{\sqrt{g^2+g'^2}}{2} \end{cases} \quad , \quad \tau_{(V)} = \begin{cases} 1 \\ c_V \end{cases} \quad , \quad \chi_{(V)} = \begin{cases} 1 & , \quad V = W \\ c_A & , \quad V = Z \end{cases} \quad , \tag{3.62}$$

$\varepsilon_V \equiv m_V/m_\eta$ and $\zeta_f \equiv (m_{f_1} + m_{f_2})/(2m_\eta)$, f_1 and f_2 being the final states of V^* off-shell decay. c_V and c_A are the vector and axial current couplings, as such they depend on the final states.

The relic density is computed then from the effective cross section as in the usual WIMP paradigm (see eqs. (2.14) and (2.15)). The same calculation, with the effective annihilation cross section of κ and including other final states, would yield the abundance of κ after its freeze-out for the non-thermal case.

Chapter 4

A model of experimental interest: chromo-electric dark matter

4.1 Motivation

In the previous chapter we saw why and how to build a dark matter model based on theoretical arguments. Here we go in a complementary direction. If on one hand a model can be justified by the problems it solves, on the other hand a model can be interesting if we know how to probe it experimentally. We have already stated several times how powerful the detection experiments are and yet there is no clear trace of dark matter.

Why are we not seeing dark matter at LHC? Maybe it is not there, in which case there is little we can do. Maybe it is there, but there is too much SM to make any significant conclusion. If this is the case, then maybe it would make sense to look elsewhere. In section 2.3.1 we explained how these searches are performed and in particular we talked about the mono-X signal. This signal is great, since it is common to many DM models, but not so great because it has a considerable SM background and also it cannot distinguish easily the different DM models. The route we are taking now is looking at LHC signatures that are model-dependent, but well-justified and with low SM background. This implies that these signatures will not be common to all DM models, but for those models they will be a distinctive feature. Instead of looking at where many models are hard to find, we look at where a few models are easy to find.

In this context we work with long lived particles (LLP). SM particles do not have a lifetime comparable to the detector timescale. This means that such a process would have very low SM background (mainly due to detector defects). Fortunately, models that allow for LLPs are not made *ad hoc* to be observable, but are well justified. In many SM extensions it is the case that not only DM appears, but also some accompanying particle: all these particles belong to the dark sector. If indeed the DM candidate and one of the partners are close in mass and/or very weakly coupled, it is possible to have the decay of the partner into DM inside the detector [35, 127, 128, 129, 130, 131, 132]. Well justified models where this arises organically are SUSY and the aforementioned composite Higgs models (see e.g. [133, 134, 135, 136, 137, 138, 139]). The model we will study in detail in this section adds just one partner on top of DM and this partner has the peculiarity of being colored.

Having a colored partner adds on top of the LLP also the possibility of strong signals at

LHC, this is because the production of the partner is driven by QCD interactions. This type of models has indeed gathered interest in the literature [140, 141, 142, 143].

Since QCD is the dominant interaction driving much of the physics, many results we will derive are somewhat model independent. The specifics will depend on the representation of the colored partner under $\mathrm{SU}(3)_c$ [140]. However, many features for collider phenomenology and most importantly the LLP signatures are indeed determined by the interaction between the dark sector and the SM.

In the model we considered here the partner is in an octet (adjoint) representation of $\mathrm{SU}(3)_c$ and it interacts with the SM via a dimension-5 effective operator (the validity of effective theories for DM searches has been widely discussed in the literature, see e.g. Refs. [144, 145, 146, 147, 148, 149, 150, 151, 152]). The LLP signature for our model will then be the hadronization in bound states like ordinary quarks and gluons. We will see that due to the suppressed coupling the produced bound state can travel a macroscopic distance inside the detector and leave an observable track. In a supersymmetric context this is a well-known possibility and such bound states, originally introduced in Ref. [153], are called R-hadrons. We use here the same terminology, although our considerations do not assume any underlying supersymmetry. For more recent papers about R-hadrons, see e.g. Refs. [154, 155, 156, 157, 158].

The chapter is organized as follows: first we introduce the model and discuss some of its features in section 4.2; we proceed by considering the usual LHC constraints derived from monojet and the distinctive ones from R-hadron searches in section 4.3. We will put the focus on the interplay between them.

4.2 Lagrangian and DM phenomenology

4.2.1 Model

We want to couple our dark matter candidate to the colored sector of the standard model. In order to allow such a coupling, colored particles within the dark sector are required [159, 160, 161, 162, 163, 164, 143, 165, 166]. In this work, we consider an extension of the minimal scenario containing just the DM particle χ_1, as it is accompanied by a slightly heavier partner χ_2. We denote the masses of these particles by m_1 and $m_2 \equiv m_1 + \Delta m$, respectively. Both χ_1 and χ_2 are Majorana fermions in our model.

Scalar or fermionic partners in the fundamental representation of $\mathrm{SU}(3)_c$ can couple the DM to quarks at the renormalizable level [167, 168, 143]. If, instead, we want to consider a coupling to gluons, the lowest dimensional operator has $D = 5$ and involves a colored partner χ_2 in the adjoint representation of $\mathrm{SU}(3)_c$. The free Lagrangian for the dark sector is:

$$\mathcal{L}_0 = \frac{1}{2}\bar{\chi}_1 \left(i\gamma^\mu \partial_\mu - m_1\right) \chi_1 + \frac{1}{2}\bar{\chi}_2^a \left(i\gamma^\mu D_\mu - m_2\right) \chi_2^a \,, \tag{4.1}$$

where $a = 1, , 8$ is the index of the adjoint representation.

The coupling to gluons can be obtained with effective operators mimicking the (chromo-) electric and (chromo-)magnetic dipole moments, in the following way:

$$\mathcal{L}_{\mathrm{int}} = \frac{i}{2m_1}\bar{\chi}_2^a \sigma^{\mu\nu} \left(\mu_\chi - i d_\chi \gamma^5\right) \chi_1 G^a_{\mu\nu} \,, \tag{4.2}$$

where $\sigma^{\mu\nu} = (i/2)[\gamma^\mu, \gamma^\nu]$ and $G^a_{\mu\nu}$ is the canonically normalized field strength of the gluons.

The two operators in eq. (4.2) give rise to similar phenomenology and no interference effect between the two arises in the observables we will study. Therefore, we will only study the operator with d_χ in the rest of the chapter. Since this operator is analogous to the electric moment, we will call the model chromo-electric DM.

Effective operators describing dipole moments typically arise after integrating out heavy particles of the underlying ultraviolet theory at loop-level. This would imply a further suppression of our chromo-electric operator. Depending on the amount of suppression, it might be possible that higher order operators become relevant. We did not pursue a complete theoretical analysis of the origin of the interaction in eq. (4.2) and of the role of higher-order operators, as this was not the focus of our work.

The interactions of the dark sector with the SM particles are then described by the parameters $\{m_1, \Delta m, d_\chi\}$. Due to its UV-origin, we have that $d_\chi \ll 1$. This interaction term could be written as an effective term suppressed by $1/\Lambda$, with Λ being the scale of some underlying new physics. It is then natural to identify $d_\chi \sim m_1/\Lambda$. In order for the effective theory to be a good description, this has to be small. For the values of d_χ considered in our analysis, the energy scales of the processes of interest are always well below the operator scale Λ, thus ensuring we are in the regime where the effective field theory is valid.

We devised this model in order to discuss the role of LLPs at LHC, so let us see if this particle is indeed long-lived. The simplest process leading to the decay of χ_2 is $\chi_2 \to \chi_1 g$, whose width is:

$$\Gamma_{\chi_2} = \frac{d_\chi^2}{\pi}\frac{\Delta m^3}{m_1^2}\left(\frac{1+\frac{\Delta m}{2m_1}}{1+\frac{\Delta m}{m_1}}\right)^3 \simeq \frac{d_\chi^2}{\pi}\frac{\Delta m^3}{m_1^2}. \tag{4.3}$$

Since d_χ is expected to be small, we expect χ_2 to be a long-lived particle with lifetime on the detector timescale. More on this in the following, but by looking at the lifetime, the model already looks promising, if we want late time decays.

4.2.2 Relic density

We first discuss the DM phenomenology and as usual our first concern is reproducing the observed dark matter abundance $\Omega h^2 = 0.1198$ [28]. Such a relic density is determined by processes of the form $\sigma(\chi_i\chi_j \to \mathrm{SM\,SM})$. In order to give a sense of the size of these interactions, we reported the order of the contributions to the cross section in table 4.1. More complete expressions, depending on the ratio $\Delta m/m_1$ and the ratio between the quark masses and the DM mass, have been used in the calculations. The complete expressions can be found in section 4.B. If we want to determine the relic density predicted by this model, two modifications to the standard procedure have to be taken into account. First, if the mass splitting between χ_1 and χ_2 is small compared to their masses, co-annihilations must be included as we discussed in section 2.2.2. Second, due to the color charge of χ_2, Sommerfeld enhancement modifies the value of $\sigma(\chi_2\chi_2 \to \mathrm{SM\,SM})$. We discuss how these modifications are taken into account.

As far as the co-annihilations are concerned, the effective cross-section which determines the dark matter relic abundance is (cf eq. (2.22)):

$$\langle\sigma v\rangle_{\mathrm{eff}} = \frac{1}{(1+w)^2}\left(\langle\sigma v\rangle_{11} + 2w\langle\sigma v\rangle_{12} + w^2\langle\sigma v\rangle_{22}\right), \tag{4.4}$$

$SMSM$ / ij	$q\,\bar{q}$	$g\,g$
11		$\frac{2\,d_\chi^4}{\pi}\,\frac{1}{m_1^2}$
12	$\frac{d_\chi^2\,g_s^2}{96\pi}\,\frac{1}{m_1^2}v^2$	$\frac{3\,d_\chi^2\,g_s^2}{16\pi}\,\frac{1}{m_1^2}$
22	$\frac{3\,g_s^4}{256\pi}\,\frac{1}{m_1^2}$	$\frac{27\,g_s^4}{512\pi}\,\frac{1}{m_1^2}+\mathcal{O}(d_\chi^2)$

Table 4.1: Different contributions to the effective cross-section $\langle\sigma v\rangle_{\chi_i\chi_j\to \mathrm{SM\ SM}}$. The QCD coupling is denoted by g_s, while v is the relative velocity in the $\chi_i\chi_j$ center-of-mass frame.

where $w \equiv g_2/g_1(1+\Delta m/m_1)^{3/2}e^{-x\Delta m/m_1}$, $x \equiv m_1/T$, $\langle\sigma v\rangle_{ij} \equiv \langle\sigma v\rangle_{\chi_i\chi_j\to SM\,SM}$ and g_i is the number of degrees of freedom of χ_i. We already discussed how to derive the relic density from the effective cross-section and so we will use eqs. (2.14) and (2.15). In the following, we take $g_* = 106.75$ as a reference.

Another effect we have not discussed up until now is Sommerfeld enhancement. This plays an important role in the determination of the relic abundance [169, 170, 171]: Sommerfeld enhancement is a non-perturbative effect due to the exchange of soft gluons between the colored particles in the initial state. What happens is that the a potential between the particles is generated and this deforms the wave-function. Depending on whether the potential is attractive or repulsive, the two particles are more or less likely to be close to each other. This effect is then relevant for the self-annihilation of χ_2, since here we can have a significant potential between the two particles. A model independent discussion of this effect can be found in Refs. [140, 142]. These analyses assume that the relic DM density is dominated by QCD, remaining agnostic about the particular phenomenology deriving from the new BSM coupling. In our case this simplification makes completely sense, since we know that we require $d_\chi \ll 1$. Another model where the DM co-annihilations contribute negligibly to the relic density has been analyzed in Ref. [172] in the more general context of *sterile co-annihilations*.

When the final state is characterized by a single representation Q, the Sommerfeld-corrected cross-section is $\sigma_{\mathrm{Somm}} = S\left(C_2(Q)\alpha_s/\beta\right)\sigma_{\mathrm{Pert}}$, where $S(x)$ is the non-perturbative correction depending on the final representation (through the quadratic Casimir $C_2(Q)$). This also depends on the velocity of the particles through β: the Sommerfeld effect is stronger for slow particles. If, on the other hand, we have more than one possible final state representation, we need to consider the decomposition $R\otimes R' = \oplus_Q Q$, where R and R' are the initial state representations (in our case $R = R' = 8$) and Q are instead the possible representations of the final state.s Each Q gives a contribution to the total cross-section and has its own value of $C_2(Q)$. After group decomposition, the final result is given by eqs. (2.24, 2.25) of Ref. [140].

As a result, the profile yielding the correct relic density, with and without the inclusion of such a non-perturbative effect, can be found in fig. 4.1. In this plot, we only considered

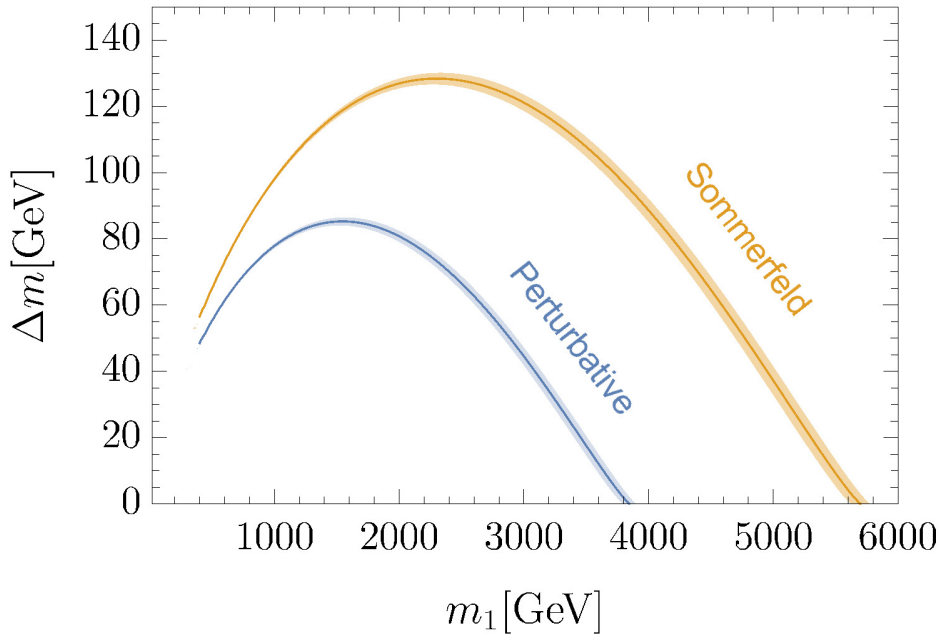

Figure 4.1: Contours corresponding to the measured relic abundance $\Omega h^2 = 0.1198 \pm 0.0012\,(1\sigma)$, together with its 3-$\sigma$ bands, in the case the QCD processes are the dominant ones. The perturbative result and the result including Sommerfeld enhancement are both shown. Note that part of the parameter space is already excluded by LHC searches, as explained in section 4.3.3 and shown in fig. 4.6.

the dominant contributions from QCD self-annihilations, not including the sub-leading contributions of processes proportional to d_χ, which we know will be negligible in the part of the parameter space where the effective theory is valid.

While the relic density is dominated by QCD processes, we see from eq. (4.3) that the decay length instead depends quadratically on d_χ. This is what we meant in the beginning when we said that the LLP signature does not depend on QCD and is instead model-dependent. Since we know that d_χ must be small, we conclude that the decay lengths can be be macroscopic. This is a central feature we will use in our analysis. From the plot we derive that Δm can be seen as a function of the mass of the DM candidate m_1, if we require the reproduction of the correct DM relic density. We use then the Sommerfeld corrected curve in fig. 4.1 in order to be able to express $\Delta m(m_1)$, a relation we will use implicitly. By doing this identification, we automatically derive the correct relic density for all the parameter points (m_1, d_χ) we will consider. Since we fixed Δm, the decay length now only the other two parameters (m_1, d_χ). The numerical results for the decay length can be found in fig. 4.2, where we show contours of the proper decay length of χ_2. By looking at the magnitudes, we understand that macroscopic decay lengths are admitted, provided that d_χ is small enough.

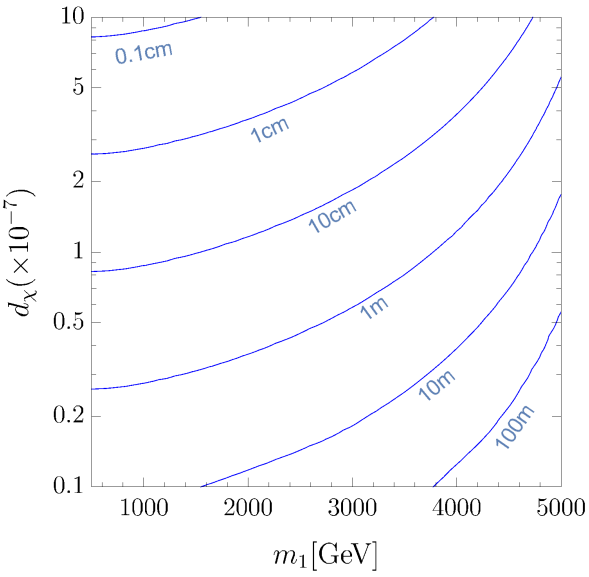

Figure 4.2: Proper decay lengths for the colored partner χ_2 in a parameter space where $\Gamma_{\chi_2}^{-1}$ is in the cm − m range. As explained in the text, the mass splitting Δm is fixed by requiring the correct relic density. Small values of d_χ and large values of m_1 give origin to larger values of decay length, as could already be understood by eq. (4.3).

4.2.3 Departure from chemical equilibrium

We discussed how co-annihilations modify the standard picture of freeze-out in section 2.2.2. But in our description we implicitly assumed chemical equilibrium between the particles. An assumption, we said, that is usually justified by the relative rate of number changing interactions and interactions keeping the kinetic equilibrium. Under particular conditions, however, it is possible that this assumption might not be valid [143]: this can happen, for instance, when the relic abundance is dominated by a SM coupling, which is exactly our case. Here the coupling characterizing the BSM physics (d_χ for our model) remains unconstrained. This implies that this parameter can be arbitrarily small and we can still reproduce the correct DM relic density. But if d_χ is too small, the assumption of chemical equilibrium can fail.

We are interested in the relative rates of some interactions. In particular, we want to consider the ratios $\Gamma_{\chi_i\chi_j}/H$, where $\Gamma_{\chi_i\chi_j}$ represents the rate of any process involving both χ_i and χ_j. This can be the scattering $\chi_2\,\chi_2 \to \chi_1\,\chi_1$, the decay $\chi_2 \to \chi_1\,g$, the conversions $\chi_2\,g \to \chi_1\,g$ and $\chi_2\,q \to \chi_1\,q$. Of course, we always need to take into account also the inverse reactions.

The rates of decay and conversion are proportional to d_χ^2, while the one for the scattering is proportional to d_χ^4. The scattering is thus expected to be the most sensitive to d_χ and the one with the smallest rate, since d_χ must be much smaller than 1.

When the largest of these rates $\Gamma_{\chi_i\chi_j}^{(\max)}$ is such that $\Gamma_{\chi_i\chi_j}^{(\max)}/H \lesssim 1$, the assumption of chemical equilibrium is no longer valid. This prevents us from using the effective cross section of eq. (2.22) to study the relic abundance. However, the failure of chemical equilibrium does

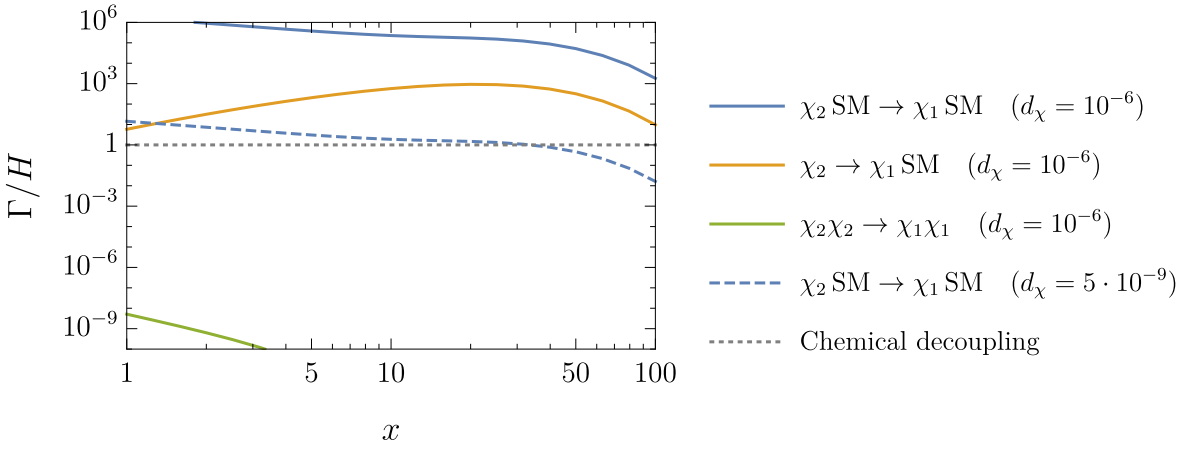

Figure 4.3: Interaction rates for the case $m_1 = 1\,\mathrm{TeV}$ and $d_\chi = 10^{-6}$. Different choices of m_1 give similar results; the scattering with gluons is always the leading contribution for maintaining the chemical equilibrium. The leading process is also shown for the threshold value of d_χ below which chemical equilibrium no longer holds.

not invalidate the model. The simplified picture we are using stops being valid and the set of coupled Boltzmann equations must be solved. This means that in order to describe this situation we would need to solve a system of differential equations (which is possible, but much harder). Nevertheless, before getting worried about deriving numerical solutions to these equations, let us look at the rates of these processes.

Since the rate corresponding to $\langle\sigma v\rangle_{\chi_2\, g \leftrightarrow \chi_1\, g} \propto g_s^2\, d_\chi^2$ turns out to be the largest one, scatterings with gluons are ultimately responsible for maintaining chemical equilibrium.

We want to use the simplified picture to describe freeze-out, this means that chemical equilibrium should at least hold before and during freeze-out. To understand if the chemical equilibrium hypothesis is valid, we look at the ratio between the scattering rate and the Hubble constant: $\Gamma_{\chi_2\,\chi_1}/H \equiv H^{-1}\, n \langle\sigma v\rangle_{\chi_2\, g \leftrightarrow \chi_1\, g}$. We look how this ratio varies in the region where freeze-out is supposed to happen, meaning $20 \lesssim x_F \lesssim 30$.

From the plot of the decay lengths in fig. 4.2 we see that in order to obtain LLPs we need $d_\chi \lesssim 1 \times 10^{-6}$. We check if for such a value chemical equilibrium holds in fig. 4.3. Here we show the ratios Γ/H for the three processes maintaining the chemical equilibrium. We see that the leading process, in the region of possible freeze-out, has an interaction rate of $\sim 10^4$ for $d_\chi = 10^{-6}$. Since this interaction rate scales as d_χ^2, it is easy to rescale and find the critical value of d_χ:

$$\frac{\Gamma_{\chi_2\,\chi_1}}{H} \lesssim 1 \quad \Leftrightarrow \quad d_\chi \lesssim 10^{-8}\,. \tag{4.5}$$

In the plot in fig. 4.3 the ratio Γ/H for the leading process is showed also for $d_\chi = 5 \times 10^{-9}$, which is the correct critical value if $m_1 = 1\,\mathrm{TeV}$. We want to be cautious and take into account also the mass dependency of this critical value, so we will consider $d_\chi \gtrsim 10^{-8}$ as our threshold. For larger values of d_χ we will not have to deal with the failure of chemical equilibrium. We will see that the physics we are interested in happens in a regime where the chemical equilibrium actually holds.

4.3 LHC searches

We have arrived at the central part of our analysis. This model reproduces the correct relic abundance and has the potential for LLPs. Here we want to assess the limits on the parameter space coming from both R-hadrons and monojet. Other LHC bounds may apply, such as the dijet-resonance bounds coming from the production and breaking of a bound state of two χ_2 particles, similar to a gluinonium in SUSY. This latter constraint, DD and ID have been considered, but put rather weak bounds on the model, so we comment on them in section 4.A.

4.3.1 R-hadron constraints

Let us start from the unusual signature. Colored particles can hadronize and χ_2, our DM partner, is charged under $\mathrm{SU}(3)_c$. When the particle reaches the detector (and it can do this because of its lifetime), it interacts with the SM particles constituting it. Here it can form particles analogous to the R-hadrons in SUSY. These particles then proceed through the detector and, if their lifetime allows it, they can travel the whole detector leaving an ionization track. This means that they can be detected.

For the purpose of setting bounds on the model, we need to recast the results of SUSY searches. We consider Ref. [154], which searches for R-hadrons at $\sqrt{s} = 13\,\mathrm{TeV}$ with $3.2\ \mathrm{fb}^{-1}$ of data. In this ATLAS search what we are interested in are the limits on the gluinos, since they are in the octet representation just like our χ_2.

We are not only interested in what is happening now, but also in the future prospects. This leads us to consider an approximate high-luminosity (HL) projection of these limits to $\mathcal{L} = 3000\,\mathrm{fb}^{-1}$, using the procedure outlined in Ref. [139]. What we need to consider for the projection are the background counts in Table 3 of Ref. [154] for the gluino search. These will be straightforwardly rescaled with the future luminosity. We make the raw assumption that the same efficiencies of Table 3 will hold also in the future. Increasing the luminosity so much implies that we will be limited by systematics, rather than by limited statistics. We then finish by carrying out our LHC simulations.

In order to simulate the pair production of χ_2 particles at parton level we have used MadGraph5 [115], where the model has been implemented using FeynRules [113], and apply the R-hadronisation routine from Pythia 8.230 [173].

We have said that the DM partner has the potential of reaching the first layer of the detector, but the probability of doing so depends on the decay length. We also need the R-hadron to travel through the whole calorimeter, if we want to measure the ionizing track. For the purpose of evaluating this probability, we rely on the assumption that the lifetime of the resultant R-hadron is at least as long as the unhadronized χ_2 lifetime. Then we have:

$$\mathcal{P}(\ell > \ell_{\mathrm{calo}}) = \exp\left(-\frac{\ell_{\mathrm{calo}}}{\ell_T}\right), \tag{4.6}$$

where $\ell_{\mathrm{calo}} = 3.6\,\mathrm{m}$ is the transverse distance to the edge of the calorimeter and we defined $\ell_T = p_2^T/(m_2\,\Gamma_{\chi_2})$. This probability will be used to rescale the cross section of the process and derive the effective cross-section for the process of yielding at least one R-hadron. Note that each event generated will have a different probability, since l_T depends on the p_T of the events. Not all χ_2 will lead to the production of such a bound-state, following Ref. [154] we assume that 90% of the χ_2 form charged R-hadrons.

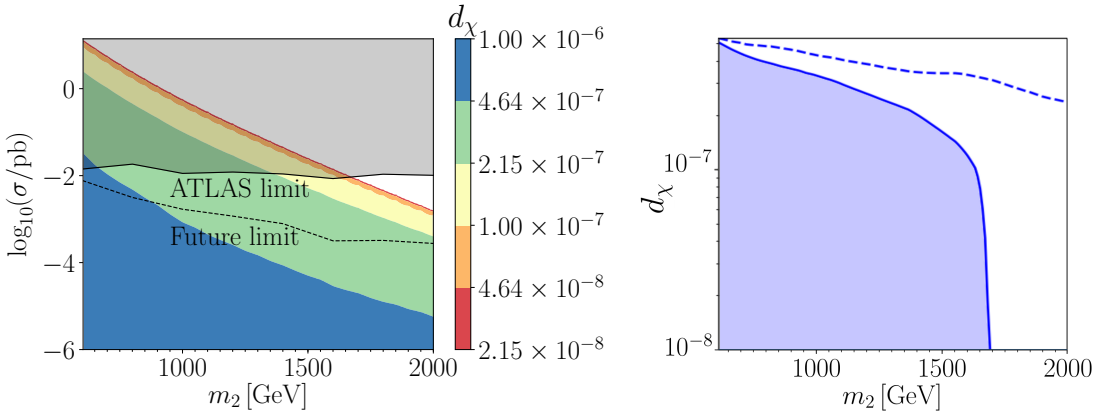

Figure 4.4: *Left:* Values of the effective cross-section for the prodcution of R-hadrons for different values of d_χ and m_2. The effective cross-section has been obtained by taking into account both the probability of hadronizing and the probability of reaching the edge of the calorimeter. The current and future bounds from ATLAS are also showed. *Right:* Lower limit on d_χ from current (solid) and projected high-luminosity (dashed) R-hadron constraints as a function m_2. In these plots we used m_2 as variable, since it is the one entering directly the calculations. We remember that by imposing the correct relic abundance m_1 and m_2 are not independent.

On the left of fig. 4.4 we show how this effective cross section depends on m_2 and d_χ. We also show here the cross-section bound coming from ATLAS searches. On the right, we can see the limit in the (m_2, d_χ) parameter space.

4.3.2 Monojet

While the R-hadron signature is peculiar to specific DM models, a general signature for DM searches is the one of monojet as we explained in section 2.3.1. Here what is measured is a general large missing energy $\not{E}_T$. The invisible particle we are looking for can be traced if it recoils against an energetic jet, whose energy is measured [174, 175, 116, 176].

For the chromo-electric model the production processes leading to the monojet signature are of the form $pp \to \chi_i \chi_l j$ with $i, l = 1, 2$. Since $d_\chi \ll 1$, the production channel $pp \to \chi_2 \chi_2 j$ involving QCD will be the leading one, similarly to what happened for relic production. All other terms are proportional to powers of d_χ and we know that this parameter is supposed to be smaller than g_s. As we did previously, we apply the relic constraint in fig. 4.1.

We consider the monojet results from ATLAS [116], which searches for events with large missing energy and at least one high-energy jet. This search is performed with $\sqrt{s} = 13\,\mathrm{TeV}$ and $\mathcal{L} = 36.1\,\mathrm{fb}^{-1}$. The cuts for this search are as follows: $\not{E}_T > 250\,\mathrm{GeV}$, leading-$p_T > 250\,\mathrm{GeV}$, $|\eta|_{\mathrm{leading-jet}} < 2.4$. A maximum of four jets with $p_T > 30\,\mathrm{GeV}$ and $|\eta| < 2.8$ are allowed, and the condition $\Delta\phi(\mathrm{jet}, \boldsymbol{p}_T^{\mathrm{miss}}) > 0.4$ must be satisfied for any jet. The ATLAS analysis then establishes 10 signal regions. These signal regions are identified by different lower cuts on $\not{E}_T$. The lowest cut requires $\not{E}_T > 250\,\mathrm{GeV}$ and corresponds to region IM1; while the last region IM10 requires $\not{E}_T > 1000\,\mathrm{GeV}$. Less events and less background are present if we increase the cut on $\not{E}_T$, so it is not obvious which region is the best for our

model. For the purpose of investigating this we need to set up our monojet simulations.

We simulate events at parton level using MADGRAPH5 [115], then apply the same cuts as Ref. [116]. Opposed to what we wanted before, we now require the DM partner to decay promptly. What we need is that the colored partner should decay before reaching the inner part of the detector, this corresponds to a length $\ell_{\text{beam}} = 2.5$ cm. Otherwise, if it enters the detector material, it could form an R-hadron as we discussed in the previous subsection. This hadronization happens on timescales of the order of $\Lambda_{\text{QCD}}^{-1} \sim 10^{-24}\,\text{s}$. Here contrarily to what we did before, we need to consider the probability that the transverse decay length ℓ_T of χ_2 is less than l_{beam} [142]:

$$\mathcal{P}(\ell_T < \ell_{\text{beam}}) = 1 - \exp\left(-\frac{\ell_{\text{beam}}}{\ell_T}\right), \tag{4.7}$$

where $\ell_T = p_2^T(i)/(m_2\,\Gamma_{\chi_2})$ is the transverse distance traveled by χ_2 in an event i. As we did for the R-hadron, also here we consider each event with its probability of decaying promptly. When we say that the particles reaching the detector hadronize, we are being conservative, as not all χ_2 will do that. But by imposing that, we ensure that we are not applying our analysis to cases which are not under our control.

We want to see what will happen to monojet bounds in future searches, so we need to make a projection for the bounds at higher luminosities. We are looking for a bound on N_{NP}, the number of new physics events. For both current and future luminosities, we apply a χ^2 analysis and consider the 95% CL with unit efficiency and acceptance, according to [140]:

$$\chi^2 = \frac{[N_{\text{obs}} - (N_{\text{SM}} + N_{\text{NP}})]^2}{N_{\text{NP}} + N_{\text{SM}} + \sigma_{\text{SM}}^2}, \tag{4.8}$$

where the error on the SM background is assumed to be normally distributed.

As we were saying before setting up our simulations, the best signal region must be found for our model. For that purpose, we consider the different signal regions from Ref. [116] and see how our results vary for different $\not{E}_T$ cuts. For a given value of m_1, we calculate the ratio between our MADGRAPH5 simulated cross section and the bound from the ATLAS paper and find that the strongest bound comes from IM9 (which requires $\not{E}_T > 900\,\text{GeV}$). This can be seen in fig. 4.5 where we set $m_1 = 860\,\text{GeV}$. This value of the mass was chosen (a posteriori) for the plot, as it was the one leading to a ratio of the cross sections equal to 1 and will hence be the lower bound. It should be noted that changing the mass varies both the value of the cross section and the kinematic distribution of the particles. For each value of m_1 a simulation must be performed and the results from fig. 4.5 cannot be recast into a bound on the mass.

For this region IM9, the number of predicted and observed SM events is:

$$N_{\text{SM}} = 464 \pm 34\,, \qquad N_{\text{obs}} = 468\,. \tag{4.9}$$

Then from the χ^2 analysis we have that the cross section for a new physics (NP) model predicting new events here has to satisfy the constraint $\sigma_{\text{NP}} < 2.3\,\text{fb}$ for $\mathcal{L} = 36.1\,\text{fb}^{-1}$. Using this value and the procedure outlined earlier in this section, we find a lower bound on the mass of the DM of 860 GeV for $d_\chi \gtrsim 3 \times 10^{-7}$. Full results are shown as the blue lines in fig. 4.6. This lower bound on d_χ comes from the requirement that χ_2 must not reach the

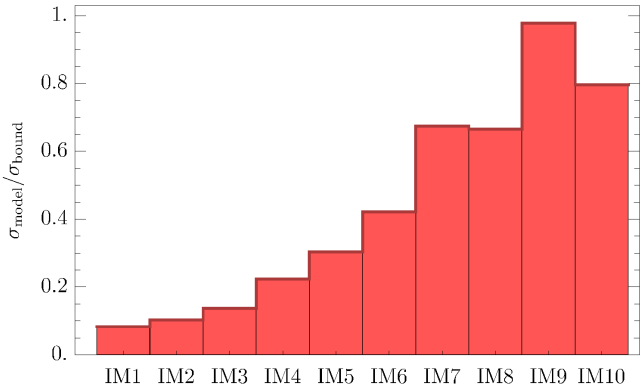

Figure 4.5: Ratio between the cross section from our model and the bound from Ref. [116] in the case of $m_1 = 860\,\text{GeV}$. This ratio is shown for the different inclusive regions considered in the ATLAS paper.

inner detector. For smaller values of d_χ the lifetime of χ_2 increases, leading to the possibility of reaching the detector and hadronizing.

We project the monojet bound from Ref. [116] to higher luminosity, in doing so we need to consider the statistical and systematic uncertainties separately. The relative statistical error scales with the inverse square root of the number of events (and hence of the luminosity). Increasing the luminosity will decrease this type of error. On the other hand, it is not possible to know exactly how systematics will scale at higher luminosities. We can decrease these errors with improved technology and computations, but using the LHC for such a long time means wearing out the detectors, thus increasing systematics. For this reason, we parametrize it in general as $\delta^{\rm sys}(\mathcal{L}_2) \equiv r\,\delta^{\rm sys}(\mathcal{L}_1)$. Knowing the upper bound on the cross-section at a luminosity $\mathcal{L}_1$, we can then estimate the corresponding upper bound at a different luminosity $\mathcal{L}_2$ as:

$$\sigma_{NP}^{(\mathcal{L}_2)}(r) \leq \sigma_{NP}^{(\mathcal{L}_1)} \sqrt{r^2 + \left(\frac{\mathcal{L}_1}{\mathcal{L}_2} - r^2\right) \frac{N_1}{\delta N_1^2}}\,. \tag{4.10}$$

We carry out this HL projection at 3000 fb^{-1}. The values we will report refer to an optimistic scenario where systematic uncertainties have been cut to half with respect to the current values. The limit on m_1 from monojet is then $m_1 > 1020\,\text{GeV}$. The results including also the dependency on d_χ are shown as the dashed blue curve in fig. 4.6.

Needless to say, we also considered other scenarios. In the case the systematics will be unchanged with respect to their current value we get the bound $m_1 > 900\,\text{GeV}$. Notice that this limit is not much stronger than the current one. As it is known, monojet searches are dominated by systematics uncertainties: increasing only the luminosity does not improve the current situation much. In the optimal and unrealistic case in which systematics will be completely negligible, we would get $m_1 > 1250\,\text{GeV}$.

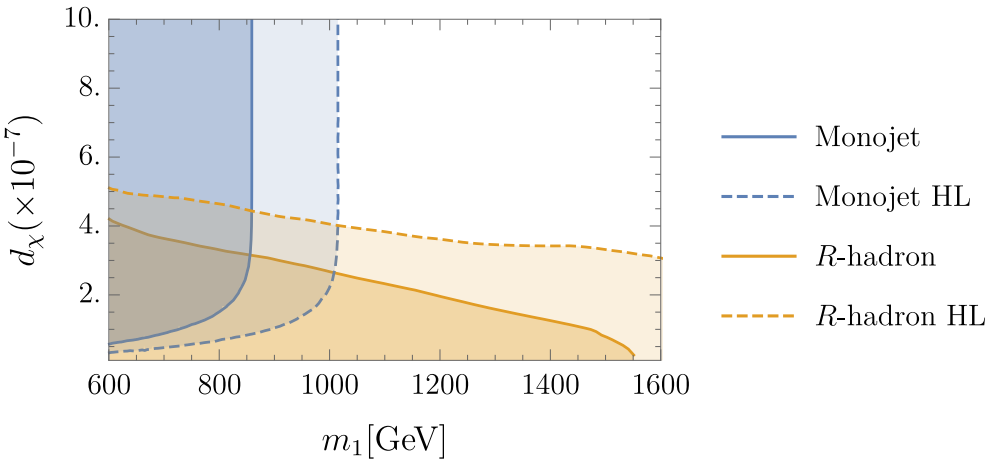

Figure 4.6: Current (solid) and foreseen future (dashed) status of the parameter space as excluded by monojet and R-hadrons searches, in blue and orange, respectively. The future bounds make assumptions on the role of systematics, as explained in the text. For every point represented the correct relic density is reproduced by fixing Δm accordingly.

4.3.3 Comparison between different searches

The LHC bounds relevant for this model are mainly monojet and R-hadrons. Since the different LHC signals are most effective in different regions of parameter space, it is important to understand the interplay between them and their complementarity. A first relevant feature is that the monojet analysis is insensitive to the value of the BSM coupling d_χ in most of the parameter space. This is because the production at LHC is QCD dominated. However, for low enough values of d_χ, we cannot assure the prompt decay of χ_2 and the bound on the mass becomes rapidly weaker. Since particles start being too long-lived, a new signal starts becoming relevant: R-hadrons. The search for these states becomes the one putting relevant constraints, as can be inferred from fig. 4.6. In this plot both bounds are shown, together with the future projections. We remind that for the monojet we assumed systematics are reduced by 50% and for the R-hadron we considered unchanged systematics.

These two searches are affected by different errors, meaning that increased luminosity has a distinct effect on each of them. Bluntly put, R-hadron searches are limited by statistics, so that increasing luminosity improves the situation significantly, while the same does not hold for monojets. This suggests that a high-luminosity projection might show us which of these searches has the best prospects of becoming interesting. These projections are shown in fig. 4.6, where the role of higher luminosity in probing a larger parameter space is made explicit.

If we only care about a bound on the mass today, we can clearly see two complementary regimes. For $d_\chi \gtrsim 3 \times 10^{-7}$ the strongest bound comes from monojet searches and tells us that $m_1 > 860\,\text{GeV}$. This bound starts becoming weaker for smaller couplings. Here instead the R-hadron search gives the relevant result and the bound on the mass depends on the specific value of d_χ. We can see that the bound on the mass becomes much stronger in this part of the parameter space, limiting masses of up to 1.5 TeV.

Appendices

4.A Estimate of other constraints

There are of course other constraints on the model, coming from both direct and indirect detection. Furthermore, another possible LHC signal is given by the disintegration of a bound state into two jets. Let us briefly comment on the reasons they are subleading

For DD we are interested in the scatterings $\chi_1 q \to \chi_1 q$. These are generated at 1-loop, with two insertions of g_s and two insertions of d_χ. We have that:

$$\sigma_{\rm DD} \propto d_\chi^2 g_s^2 \frac{m_p^2}{m_1^2}, \tag{4.11}$$

here the proportionality constant will depends on the parton distribution function of the proton, loop factors and the sum of different diagrams. The DD will impose an upper bound on d_χ, even overestimating the strength of this search, we derive that values of $d_\chi \leq 1 \times 10^{-4}$ are always allowed. Thus, the DD search does not operate in the same parameter space interesting for the comparison of LHC searches.

We also checked the ID limits by applying bounds from [177] to our model. These bounds are derived by looking at the anti-protons fluxes in cosmic rays. From the upper bound on the annihilation cross section $\sigma(\chi_1\chi_1 \to gg)$ we can set a bound on m_1 and d_χ. Also in this case, the limit on d_χ is an upper bound, since the cross section will be proportional to d_χ^4. The bounds found were not relevant for us, since they were in a region where $d_\chi \sim 1$, this means that we are in a region where our effective operator description is not accurate. Just to mention some numbers, the upper limit on d_χ was found to be $d_\chi \leq 0.2$ for $m_1 = 1\,\text{TeV}$ and $d_\chi \leq 1$ for $m_1 = 5\,\text{TeV}$. We have stated before that large uncertainties affect this detection channel, so we do not expect these constraints to be precise. Anyway, the parameter space probed is far from the one interesting for LHC searches, so we are satisfied with this result.

It is also possible, especially if the particles are slow, that a couple of χ_2 forms a bound state. This bound state can later decay into a couple of jets, giving a distinctive signature. This kind of signal is rather model-independent and it falls under the category of signals controlled by QCD. This kind of bound states are predicted in SUSY and take the name of gluinonium. This means that we can use analyses and bounds carried out on gluinonium and use them for our model. In this spirit, we use the production rate of gluinonia as calculated by Ref. [178] as the value for the production rate of $\chi_2\chi_2$ bound states. We use model-independent constraints from ATLAS [179] and put a constraint on the dijet resonance production cross-section. As we did for R-hadrons and monojet, we investigate the future prospects, so we make projections for a luminosity of 3000 fb^{-1} using methods analogous to

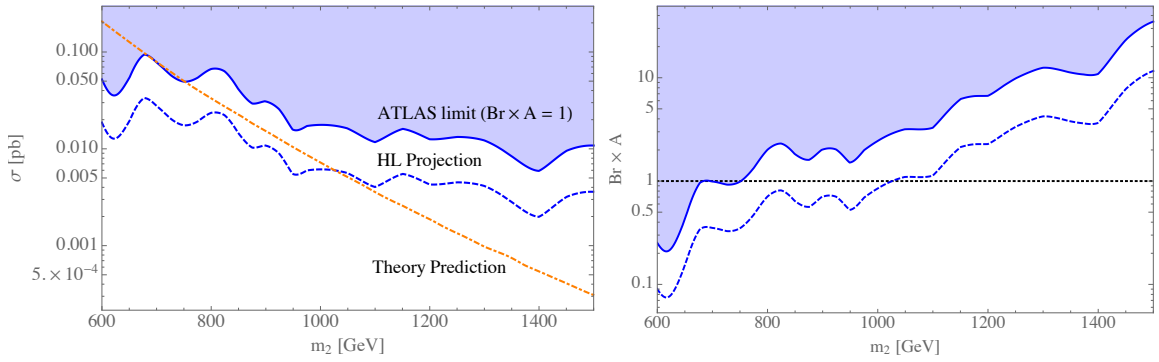

Figure 4.7: The excluded region (shaded) and projected future limit (dashed) are presented as a function of m_2. *Left:* Limits on the production cross-section obtained by setting $Br \times A = 1$. The theoretical prediction for the dijet production cross-section (dot-dashed) used to impose the limits is also showed. *Right*: The limits in the plot on the left have been recasted as a limit on $Br \times A$. $Br \times A$ cannot be greater than 1, as highlighted in the plot.

those used in previous sections. For the purpose of evaluating the bounds, we have worked under the assumption that the fitting function for the dijet mass distribution used in [179] is still valid at higher luminosities. For what concerns systematic uncertainties, here we adopted the same criterion we used for R-hadrons and assumed they are unchanged at higher luminosities.

These constraints are shown in fig. 4.7 (left), along with the theoretical production rate from Ref. [178]. The limits assume that both the branching ratio to dijets (Br) and the acceptance (A) are 1. This overestimated the bound, since actually $Br \times A < 1$. We can then recast the bound as a lower limit on $Br \times A$, as we did in fig. 4.7 (right).

Since $Br \times A$ cannot be greater than 1, all the parameter points for which the bound is greater than this value are certainly not limited by dijets. This critical value is reached for $m_2 \sim 650\,\text{GeV}$. Below this mass, dijet constraints rule out the model only if $Br \times A$ is larger than 0.1-1 (the precise value depending on m_2).

This constraint is conservative for two reasons. First, the production rate is calculated for a CM energy of 14 TeV, while LHC works at 13 TeV, thus in reality the cross section will be smaller. Second, we have taken the dijet constraints assuming a narrow Gaussian width. A broader width weakens the constraints, as seen in Fig. 5 of Ref. [179]. In conclusion, at the current status dijet searches do not put further constraints on the model with respect to monojet, but depending on the systematics reduction for monojet, they may become a more interesting channel in the future.

4.B Analytical expressions for the differential cross sections

We present here the analytical expressions for the differential cross sections used in the calculations. We simplified the expressions by using the quantities $E_1 = \sqrt{m_1^2 + p^2}$ and $E_2 = \sqrt{m_2^2 + p^2}$, which are the energies of the incoming particles χ_1 and χ_2. m_q is the mass of the outgoing quarks, we report the values for just one flavor. The expressions for the

cross sections are the ones derived in the center of mass frame (CM), so that p and θ must be interpreted as the momentum of the incoming particles in the CM frame and the angle between incoming and outcoming momenta in the CM frame.

4.B.1 $\chi_1\chi_1 \to gg$

The self-annihilation of the DM particles into gluons proceeds via t and u channel exchanges of χ_2.

$$\begin{aligned}\left.\frac{d\sigma v}{d\Omega}\right|_{\text{CM}} =& \frac{d_\chi^4 E_1^2}{2\pi^2 m_1^4 \left(-2E_1 p\cos\theta + E_1^2 + m_2^2 + p^2\right)^2 \left(2E_1 p\cos\theta + E_1^2 + m_2^2 + p^2\right)^2} \\ & \Big[4E_1^6 \left(m_2^2 + 2p^2\sin^2\theta\right) \\ & + 2E_1^4 \left(2m_2^2 p^2(3 - 4\cos(2\theta)) + 4m_2^4 + p^4\sin^2\theta(-(7\cos(2\theta)+1))\right) \\ & + E_1^2 \left(m_2^2 p^4(7\cos(4\theta) - 12\cos(2\theta) + 1)\right) \\ & + E_1^2 \left(4m_2^4 p^2(1 - 5\cos(2\theta)) + 4m_2^6 + p^6\sin(2\theta)\sin(4\theta)\right) \\ & - 4p^2 \left(m_2^2 + p^2\right)^2 \left(p^2\sin^4\theta - m_2^2\right) \Big]\end{aligned} \tag{4.12}$$

4.B.2 $\chi_1\chi_2 \to gg$

The co-annihilation into gluons proceeds via s-channel gluon interaction, t and u exchange of χ_2 and via the quartic $\chi_1\chi_2 gg$ coupling.

$$\begin{aligned}\left.\frac{d\sigma v}{d\Omega}\right|_{\text{CM}} =& \frac{3 g_s^2 d_\chi^2}{512\pi^2 E_1 E_2 m_1^2 \left(E_2^2 - p^2\cos^2\theta\right)^2} \\ & \Big[-8E_1^2 p^2 \left(E_2^2(1 - 2\cos(2\theta)) + p^2\cos^2\theta\right) \\ & + 8E_1 E_2 p^2 \left(E_2^2\left(4 - 5\cos^2\theta\right) + p^2\cos^2\theta\left(4 - 3\cos^2\theta\right)\right) \\ & + 8E_1^3 E_2 \left(E_2^2 - p^2\cos^2\theta\right) \\ & + E_2^2 p^2 \left(4m_1 m_2(1 - \cos(2\theta)) + p^2(3\cos(4\theta) - 11)\right) \\ & + p^4 m_1 m_2(\cos(4\theta) - 1) + p^6(\cos(4\theta) + 7)\cos^2\theta \Big]\end{aligned} \tag{4.13}$$

4.B.3 $\chi_1\chi_2 \to q\bar{q}$

The co-annihilation into quarks proceeds via s-channel gluon interaction.

$$\begin{aligned}\left.\frac{d\sigma v}{d\Omega}\right|_{\text{CM}} =& \frac{g_s^2 d_\chi^2 \sqrt{(E_1 + E_2)^2 - 4m_q^2}}{64\pi^2 E_1 E_2 (E_1 + E_2)^3 m_1^2} \\ & \Big[p^2 \left(2m_q^2\cos(2\theta) - (E_1 + E_2)^2\cos^2\theta\right)(E_1 E_2 - m_1 m_2)\left((E_1 + E_2)^2 + 2m_q^2\right) \Big]\end{aligned} \tag{4.14}$$

4.B.4 $\chi_2\chi_2 \to gg$

The annihilation of the partners into gluons proceeds via five different channels: t and u exchange of χ_1, whose amplitude will be proportional to d_χ^2, s-channel gluon interaction, and t and u exchange of χ_2, whose amplitude will be proportional to g_s^2.

Hence the total cross section will be the sum of three terms proportional to g_s^4, $g_s^2 d_\chi^2$ and d_χ^4.

$$\begin{aligned}
\left.\frac{d\sigma v}{d\Omega}\right|_{\rm CM} =& -\frac{9g_s^4}{16384\pi^2 E_2^4\left(E_2^2-p^2\cos^2\theta\right)^2} \\
&\Big[E_2^2p^4(5\cos(4\theta)-12\cos(2\theta)+31)+4E_2^4p^2(5\cos(2\theta)-7) \\
&-24E_2^6+p^6(\cos(4\theta)-4\cos(2\theta)+11)\cos^2\theta\Big] \\
&+\frac{3g_s^2d_\chi^2}{512\pi^2m_1^2(E_2^2-p^2\cos^2\theta)(-2E_2p\cos\theta+E_2^2+m_1^2+p^2)(2E_2p\cos\theta+E_2^2+m_1^2+p^2)} \\
&\Big[4E_2^4(m_1m_2+2p^2\sin^2\theta)+2m_1p^2(-2m_1^2m_2\cos(2\theta)+p^2m_2(\cos(4\theta)+1) \\
&+m_1p^2\sin^2\theta(3\cos(2\theta)+1))-2E_2^2(4m_1p^2m_2\cos(2\theta)-2m_1^3m_2-4m_1^2p^2\sin^2\theta \\
&+p^4\sin^2\theta(5\cos(2\theta)+3))+p^6\sin^2\theta(6\cos(2\theta)-\cos(4\theta)+3)\Big] \\
&+\frac{d_\chi^4E_2^2}{64\pi^2m_1^4(-2E_2p\cos\theta+E_2^2+m_1^2+p^2)^2(2E_2p\cos\theta+E_2^2+m_1^2+p^2)^2} \\
&\Big[2E_2^6(8m_1^2+9p^2\sin^2\theta)+E_2^4(m_1^2p^2(27-43\cos(2\theta))+32m_1^4+2p^4\sin^2\theta(3-14\cos(2\theta))) \\
&+E_2^2(m_1^2p^4(21\cos(4\theta)-34\cos(2\theta)-3) \\
&-m_1^4p^2(59\cos(2\theta)+5)+16m_1^6+2p^6\sin^2\theta(4\cos(4\theta)-6(\cos(2\theta))-3)) \\
&+p^2(m_1^2+p^2)^2(m_1^2(7\cos(2\theta)+9)+2p^2\sin^2\theta(4\cos(2\theta)+3))\Big]
\end{aligned} \tag{4.15}$$

4.B.5 $\chi_2\chi_2 \to q\bar{q}$

The annihilation of the partners into quarks proceeds via purely QCD s-channel gluon interactions.

$$\left.\frac{d\sigma v}{d\Omega}\right|_{\rm CM} = \frac{3g_s^4\sqrt{E_2^2-m_q^2}\left(E_2^2\left(m_q^2+p^2\cos^2\theta-p^2\right)+2E_2^4-m_q^2p^2\cos^2\theta\right)}{2048\pi^2E_2^7} \tag{4.16}$$

Chapter 5

A data-driven approach: finding dark matter with LHC data

5.1 Preamble: DarkMachines

Before delving into the discussion of what we did and why, let us briefly mention how the particle physics community is using machine learning (ML) at LHC. Machine learning has seen an exponential increase in popularity and it is starting to be used in any field. Particle physics is here no exception.

The end goal here is using an efficient ML algorithm on the LHC data in order to guide the searches. Of course, we do not actually know if there is some beyond the standard model physics at LHC yet, so if we were to study the LHC data we would have no way of checking whether the algorithm works well or not. The idea is then to simulate what might happen at LHC: a dataset containing both SM events and new physics event is provided, with known labels. Efforts in this direction have been put forward by both DarkMachines (section 23 of [5]) and LHCOlympics [180]. We focus on the former, since it is the one we are directly involved in and it is more focused on the topic of dark matter.

In Ref. [5] several datasets are provided at a luminosity of $\mathcal{L} = 10\,\text{fb}^{-1}$. First of all, a dataset for the SM events is presented, comprised of 26 processes. Then, three BSM signals are considered:

$$\begin{aligned} pp &\to \tilde{g}\,\tilde{g}\,, \quad \tilde{g} \to t\bar{t}\tilde{\chi}_0 \\ pp &\to \tilde{t}\,\tilde{t}\,, \quad \tilde{t} \to t\tilde{\chi}_0 \\ pp &\to Z' \to t\bar{t} \end{aligned}$$

where $\tilde{g}, \tilde{t}, \chi_0$ are the gluino, stop and lightest neutralino of SUSY, Z' is instead the massive mediator of a leptophobic model [181]. For each signal several values for the mass parameters are considered, for a total of 16 possible BSM signals. These, in combination with the SM background, can be used to train and validate ML algorithms and the DarkMachines simulations serve as a benchmark. An ideal algorithm should be robust against the type of BSM signal. We will later use this dataset to study the properties of our proposed anomaly detector.

5.2 Motivation

Now that we have a benchmark dataset provided by DarkMachines, we want to study it someway. To this end one might be tempted to go and apply their favorite ML algorithm to the dataset. ML usually provides the best results and can tackle any data analysis task (outlier detection [182, 183], classification [184, 185], clustering [186, 187]). That is a legitimate approach and considering how large the dataset is, a well constructed architecture will provide good results for the data analysis task of interest. However, a ML algorithm is not something easy to control and may require a long time for computation. In particular, we are here worried about the "black box" problem [188]: we will deal with complex datasets, so we want to try and understand what is going on and in which direction we should go.

To this end, we take a step back and consider density based approaches. The advantage of such approaches is that the results will be much clearer with respect to ML: some probabilities are evaluated on the datasets provided and by comparing their ratios we can assert some properties of the datasets. In our treatment we will not say anything new about the way the probability estimate is used for the specific task, meaning that we will rely on the established literature. Our interest will be rather be on the first step: we introduce here a new, simple idea on how to estimate the probability given a sample. The density estimator we will build and study is based on Markov chains. Markov processes are everywhere in statistical physics and have already found several applications in data analysis [189, 190]. The most popular approach to density estimation is kernel density estimation (KDE), which is fast, reliable and easy to understand [191]. We will see that, though we start from a very different idea, our estimator can be related to the well-known KDE.

In section 5.3 we describe how the estimator is built and provide a general algorithm for density estimation. We will show by means of simulations that the estimator is consistent and performs well. The task of interest to us for LHC is outlier detection. A background independent approach is described in section 5.4, where we build a local outlier detector with our density estimates and we perform a comparative study on general scope datasets. We finally perform a mock analysis of the dataset provided by DarkMachines in section 5.5, here we perform anomaly detection for the case in which the background is known. In section 5.A a discussion on the error analysis for two variants of our estimator is presented, with more technical details in section 5.B.

The code for the Markov chain density estimator (MCDE) with an example of usage is publicly available[1].

5.3 Algorithm for Density Estimation

5.3.1 Construction of the Markov chain

Consider the sample points $\{\boldsymbol{x}_j : \boldsymbol{x}_j \in \mathbb{R}^D\}_{j=1}^N$, which are N independent and identically distributed D-dimensional realizations of a random vector $\boldsymbol{x}$ with unknown probability density function (PDF) $f(\boldsymbol{x})$:

$$\mathcal{S} \equiv \{\boldsymbol{x}_j\}_{j=1}^N \overset{\text{iid}}{\sim} f\,. \tag{5.1}$$

[1] https://github.com/de-simone/MarkovChainDensityEstimator

Throughout the discussion we assume $N > 2$ and we consider a distance metric $d(\cdot, \cdot)$ on $\mathbb{R}^D$, which for all our purposes will be the Euclidean metric. Let us consider a stochastic process starting from any $\boldsymbol{x}_j \in \mathcal{S}$ and that moves from one point of the sample $\mathcal{S}$ to another, with defined probabilities of jumping specified by the initial and final point. This process can be described as a discrete-time Markov chain (MC) having the finite set $\mathcal{S}$ as state space.

For the construction of our estimator, we need to specify this jumping probability. In order to retain information about the density of points (which is strictly related to probability), we construct the $N \times N$ distance matrix d_{mn}. For each pair of points $\boldsymbol{x}_m, \boldsymbol{x}_n \in \mathcal{S}$, we will have $d_{mn} \equiv d(\boldsymbol{x}_m, \boldsymbol{x}_n)$. We need to transform this distance matrix in a way that is consistent with the reconstruction of the correct probability, this means that we will need to transform in such a way that it will be more likely to jump to nearby points, rather than to far away ones. This implies that we will have to consider a regular decreasing function $g(d)$ of the distances. However, one probability remains undefined: the probability of remaining in the point. While intuitively we can say that it should be likely to jump to nearby points, it is not clear how likely it should be to remain in the starting point. To describe this freedom, we introduce the parameter b, which we will call the movement bias. The reason for this name is that for $b = 0$ the MC will have the maximal probability of staying in the point, while for $b = 1$ we will force the MC to always change state at every step of the chain. We can thus define the $N \times N$ symmetric matrix of weights W_{mn}, as a function g of the distance between points $W_{mn} \equiv g(d_{mn})\,(1 - b\,\delta_{mn})$, where δ_{ij} is the Kronecker delta and b is the movement bias. For our purposes we will have $b \in [0, 1]$, even though values outside this range could also be allowed a priori.

Some properties are required for the function $g(d)$: they are monotonically decreasing, continuous, differentiable, with bounded first derivative. As it will be clear in the next paragraphs, this function can be related to the kernel function $K(u)$ of KDE. Since we are talking about the probabilities of jumping, we need to require that the sum of all probabilities is equal to 1. This implies a normalization for the $N \times N$ transition matrix Q_{mn}, which provides the probability to move from $\boldsymbol{x}_m$ to $\boldsymbol{x}_n$:

$$Q_{mn} \equiv \frac{W_{mn}}{\sum_{k=1}^{N} W_{mk}} = \frac{g(d_{mn})(1 - b\,\delta_{mn})}{\sum_{k=1}^{N} g(d_{mk})(1 - b\,\delta_{mk})}\,, \tag{5.2}$$

where we normalized the weights in W_{mn} by requiring that the sum of the elements in each row is 1. Because of this row-dependent normalization, Q_{mn} is not symmetric. The transition probability matrix defined by eq. (5.2) is now a right stochastic matrix with the following properties: (1) all entries are non-negative ($Q_{mn} \geq 0$); (2) each row sums to 1 ($\sum_{n=1}^{N} Q_{mn} = 1$). Since $g(d)$ is a decreasing function, it is more likely to move to closer points, according to the distance metric d.

Given that all states in $\mathcal{S}$ communicate (i.e. $Q_{mn} > 0$ for $m \neq n$), there is only one communication class and the MC is then irreducible. Furthermore, the state space $\mathcal{S}$ is finite, which implies that the irreducible MC is also positive recurrent. Given this latter property, we also know that it has a unique stationary distribution $\boldsymbol{\pi}$ such that $\boldsymbol{\pi}$ is invariant under Q, i.e. $\boldsymbol{\pi} Q = \boldsymbol{\pi}$. So, $\boldsymbol{\pi}$ is a left eigenvector of Q with eigenvalue 1, or equivalently, $\boldsymbol{\pi}$ is the principal right eigenvector of Q^T (Perron-Frobenius eigenvector) [192]. The reason we are interested in the stationary distribution is that it encodes the information on the proportion of time the MC spends on a given point in the long run. This means that the component π_j is the probability to find the chain at point $\boldsymbol{x}_j$ after infinite iterations. We will try to

relate this probability to the underlying PDF from which the sample is extracted. It would be nice to have an easy way to get this stationary distribution. Given that we start from the weight matrix W_{mn}, we can easily derive an explicit expression for a left eigenvector of Q_{mn}. Looking at eq. (5.2), it is easy to see that $\pi_m = \sum_n W_{mn}$ satisfies the eigenvector equation $\boldsymbol{\pi} Q = \boldsymbol{\pi}$ for any value of b, though its components do not sum to 1.

From the Markov chain point of view, the stationary distribution $\boldsymbol{\pi}$ can be viewed as a mapping from the state space $\mathcal{S}$ of the MC to $\mathbb{R}^N$: for a given sample, it provides an N-dimensional row vector whose entries sum to 1. The density estimator we describe in this chapter is built upon the observation that the MC constructed above is spending more time in higher density regions. Therefore, our estimate for the probability density at point $\boldsymbol{x}_j$ is directly related to the j-th component of the principal eigenvector

$$\hat{f}(\boldsymbol{x}_j) \propto \pi_j \,. \tag{5.3}$$

Obviously, this is just the starting idea and for now we have no guarantee that it should actually work.

5.3.2 Link with KDE

It is easy to see, though definitely not obvious a priori, that the stationary distribution of a Markov chain with movement bias $b = 0$ is proportional to the KDE estimate of the probability in the points of the sample. This is because we have $\sum_{i=1}^{N} K(d(\boldsymbol{x}_i, \boldsymbol{x}_m)/h) = \sum_{i=1}^{N} g(d_{im}) \propto \pi_m$, where $K(u)$ are the usual kernel functions of KDE. Here h is the bandwidth parameter, which in this discussion is supposed to be fixed, and since $K(d/h)$ satisfies the requirements we imposed for $g(d)$, it is a possible choice for this function. From now on, we will identify our function $g(d)$ with the kernel function $K(d/h)$ as we will only consider these. This means that the KDE estimate in the points of the sample can be recovered with our density estimator (up to a multiplying coefficient) by setting the parameter b to 0. However, independently of the value of b, it is always possible to establish a link between our density estimator and KDE. We remind that the KDE estimate in a generic point $\boldsymbol{x}$ can be written as

$$\hat{f}_{\mathrm{KDE}}(\boldsymbol{x}) = \frac{1}{N} \frac{1}{\det(H)} \sum_{i=1}^{N} K(H^{-1}(\boldsymbol{x} - \boldsymbol{x}_i)), \tag{5.4}$$

here H is a $D \times D$ matrix carrying information about the bandwidth in each direction. In order to easily link this with our estimators let us write $K(H^{-1}(\boldsymbol{x}, \boldsymbol{x}_i)) = K(d(\boldsymbol{x}, \boldsymbol{x}_i)/h)$ and $\det(H) = h^D$. With this definition the different weight H can give to each direction has been translated in the fact that in evaluating the distance we can give a different weight to each direction.

For our estimator with a generic b we have

$$\pi_i \propto \frac{1}{N} \sum_k W_{ik} = \frac{1}{N h^D} \left[\sum_k K\left(\frac{d(\boldsymbol{x}_k, \boldsymbol{x}_i)}{h} \right) (1 - b\delta_{ik}) \right] = \hat{f}_{\mathrm{KDE}}(\boldsymbol{x}_i) - b \frac{K(0)}{N h^D}. \tag{5.5}$$

The symmetry of the kernel around 0 is portrayed here by the fact that $d(\boldsymbol{x_k}, \boldsymbol{x}_i) = d(\boldsymbol{x_i}, \boldsymbol{x}_k)$. As stated in eq. (5.3), our PDF estimate will be proportional to this π_i, which is now related

to the KDE estimate. In particular, by fixing b and the proportionality constant by hand, we could recover both the usual KDE estimator and the leave-one-out density estimator on the points of the sample. However, as explained in the next section, this is not what we do. Indeed, we fix the proportionality constant by normalizing the PDF estimate, which is something that we can do independently from KDE. The way we fix the proportionality constant is central for the results we will derive and it is noteworthy that fixing the constant in this way can lead to lowering the bias in the main population with respect to KDE. The link with KDE will be employed by us to discuss the error analysis in section 5.A and will also be used for faster computations. Since optimized software for KDE is already available, we will rely for the PDF estimate and optimization on scikit-learn [193].

5.3.3 Density estimate

As mentioned in the previous section, for $b = 0$ we can recover the KDE estimate, where normalization of the PDF is automatically guaranteed and it is possible to evaluate the probability in every point $\boldsymbol{x} \in \mathbb{R}^D$, since we have analytic expressions. However, for $b > 0$ we need to establish how to generalize to every point of $\mathbb{R}^D$ and fix the proportionality constant. We will see that even in the case $b = 0$ we do not necessarily recover KDE, depending on the way we derive an estimate from our vector $\boldsymbol{\pi}$.

The starting point for our PDF estimate will be the vector $\boldsymbol{\pi}$ found in section 5.3.1. This vector maps a sample $\mathcal{S}$ of size N to a vector of probabilities in $\mathbb{R}^N$, so it does not have the correct domain, since it should be defined on $\mathbb{R}^D$.

It will be then necessary to extend the previous result to the correct domain. This means that, starting from N points in $\mathbb{R}^D$ and having found the corresponding probabilities in $\boldsymbol{\pi}$, we will construct a continuous function going from $\mathbb{R}^D$ to $\mathbb{R}^+$. The new function will be called $q(\boldsymbol{x})$. This extension can be carried out in several ways: first of all we can perform an interpolation, but we can also take advantage of the KDE estimate and extend it to points not belonging to the sample. We have discussed an extension of the KDE estimate and the linear interpolation in sections 5.B.1 and 5.B.2 respectively. For what concerns the interpolation, we have several options, but we cannot avoid some error and the more points we have, that is the bigger N is, the more accurate the interpolation will be. Due to the fact that data are sparser at higher dimensions, it will be more difficult to correctly interpolate for large D and the final error will be larger.

The function we have now, $q(\boldsymbol{x})$, is defined on the correct domain, but it is not normalized. Having a normalized function is not a mandatory requirement for density estimates, but it gives us a way to fix the proportionality constant and it would anyway be necessary if we wanted to calculate expected values. What we have is a correctly normalized vector $\boldsymbol{\pi}$, whose components were the probabilities of ending up in one of the points of the sample, but now we are dealing with the underlying distribution. This means that for the normalization we require the integral over the domain of $f(\boldsymbol{x})$ to be 1. This can be easily obtained by defining the new function $\tilde{q}(\boldsymbol{x})$:

$$\tilde{q}(\boldsymbol{x}) \equiv \frac{q(\boldsymbol{x})}{\int q(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}} \quad \longrightarrow \quad \int \tilde{q}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1. \tag{5.6}$$

The integration can be carried out in several ways, but at higher dimensions the more practical solution is a Monte-Carlo integration. Another source of error will come from the integration

error, this will be smaller the more points are used for the Monte-Carlo, but it will be larger for higher dimensionalities. The legitimate doubt one might have now is whether the combination of interpolation and integration error is such to prevent a good performance. We will see that for some practical applications, e.g. the outlier detector constructed in 5.4.2, these types of error will not be relevant for the final performance. These errors can instead be relevant for the density estimate, but we will also see in section 5.3.4 that at least for $D \leq 6$ we can still do better than KDE.

In the following we focus on the case where $q(\boldsymbol{x})$ is obtained through interpolation, rather than by an extension of KDE. We present an estimator that is constructed relying on an extension of KDE in section 5.B. There are several reasons why we focus on the case where we interpolate $\boldsymbol{\pi}$. First of all, this is the estimator we will be using in practice for our numerical experiments. Secondly, it is possible to optimize this estimator in a way completely independent of KDE as we turn to describe next. Indeed, deriving $q(\boldsymbol{x})$ from an interpolation starting from $\boldsymbol{\pi}$ is the most agnostic thing we can do and the only option we had if we did not realize there was a link with KDE.

Similarly to what is done with KDE, we will work with a set of functions depending on some parameters to be fixed. This parameter is basically the bandwidth of KDE and so we will call the the set of functions K_h. This allows the estimator to adapt to very different situations, while still retaining its non-parametric nature. We have discussed the different ways we can optimize in section 5.B.3, if the sample is large enough a good way to optimize is use results from KDE. But we also have a way independent of KDE, which we describe now.

We have now a set of functions K_h and following from them we have a set of probability estimates $\tilde{q}_h(\boldsymbol{x})$. The idea here is to minimize a loss function constructed only with our estimates $\tilde{q}_h(\boldsymbol{x})$. The most popular and easiest loss function we can build for this purpose is the negative log-likelihood:

$$\text{Loss}(h) = -\sum_{j=1}^{N} \log \tilde{q}_h(\boldsymbol{x}_j) \quad \longrightarrow \quad h^* = \arg\min_h \text{Loss}(h)\,, \tag{5.7}$$

where h^* will be the optimal value for the free parameter. The intuition behind this optimization is that among all the possible normalized PDFs, the ones closer to the correct $f(\boldsymbol{x})$ will maximize the probability in the drawn points. This does not work for KDE, since the loss function has a global minimum for $h = 0$ and intuitively it implies that our optimization procedure will work better for b close to 1.

Finally, the estimate of the PDF will be given by

$$\hat{f}(\boldsymbol{x}) = \tilde{q}_{h^*}(\boldsymbol{x})\,. \tag{5.8}$$

A summary of the procedure used to find $\hat{f}(\boldsymbol{x})$ is presented in algorithm 1. It is up to the user to define a range of values of h where to look for the optimal value: the user is supposed to provide to the algorithm a minimal value h_{min} and a maximal value h_{max}. At step 4 of the algorithm the left eigenvector of Q is found, in our software we used the link with KDE, but numerical methods independent of KDE are possible [194].

Input : Sample $\mathcal{S}$ of size $N \times D$,
b, a range of values for the parameter h: $h \in [h_{min}, h_{max}]$
Output: Probability estimate $\hat{f}(\boldsymbol{x})$ defined on $\mathbb{R}^D$

1 choose a family of functions K_h;
2 $h \leftarrow h_{min}$;
3 **while** $h \leq h_{max}$ **do**
4 $\quad \boldsymbol{\pi}_h \leftarrow$ eq. (5.5) ;
5 $\quad$ interpolate $\boldsymbol{\pi}_h$ to get $q_h(\boldsymbol{x})$ defined on $\mathbb{R}^D$;
6 $\quad$ normalize $q_h(\boldsymbol{x})$ to get $\tilde{q}_h(\boldsymbol{x})$ where $\int \tilde{q}_h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 1$;
7 $\quad$ calculate $\mathrm{Loss}(h) = -\sum_j^N \log \tilde{q}_h(\boldsymbol{x}_j)$;
8 $\quad$ increase h;
9 $h^* \leftarrow \arg\min \mathrm{Loss}(h)$;
10 $\hat{f}(\boldsymbol{x}) \leftarrow \tilde{q}_{h^*}(\boldsymbol{x})$;

Algorithm 1: MCDE with optimization. In line 1, $K_h(d/h)$ needs to respect the requirements mentioned in section 5.3. In our case we have considered the set of functions as depending on the parameter h only.

There are many possible choices for the set of functions K_h, but for our purposes we have found that the Gaussian kernel $K_h(d/h) \propto \exp(-d^2/(2h^2))$ works best. Similarly to KDE, our final estimate does not depend sensibly on the choice of the kernel, but rather on the choice of h. However, it might be the case that for some particular distributions $f(\boldsymbol{x})$ and/or purposes a different set of functions works best. In particular, one might prefer to work with compact kernels. For our set of functions $K_h(d/h)$ small values of h correspond to large transition probabilities to nearby points, meaning that local variations will be important in order to estimate the probability. On the contrary, large values of the parameter h correspond to higher probabilities for the MC to jump to distant points, hence even far away points have an impact on the probability estimate. This should sound familiar to people working with density estimation, as the role of h in our estimator is analogous to the role of the bandwidth for KDE.

5.3.4 Numerical performance

From now on we will work with $b = 1$ and the extension $\boldsymbol{\pi} \to q(\boldsymbol{x})$ is carried out with an interpolation. As a kernel function we considered the family of Gaussian kernels. This setup has to be understood when we will talk about MCDE.

The first thing we want to assess is the consistency for increasing number of points N at different dimension D. The consistency has also been shown to follow from the error analysis of section 5.A, but in this section we focus more on when we use the estimator in practice. Once we have established that our estimator is consistent, we want to see how well it works compared to others, in our case we use for comparison the well known KDE. As a matter of fact, we have seen that our estimator can be seen in some sense as a modification of KDE.

We want to start by working in the cases we have under control, so we work with $D \leq 6$, because at low-dimensionality we can perform a precise linear interpolation and we can normalize the PDF by performing a Monte Carlo integration without incurring in a large integration error.

First, we pre-process the data, starting from our sample points $\mathcal{S}$, we will get $\mathcal{S}_w$, where $\mu_{\mathcal{S}_w} = 0^D$ and $\sigma_{\mathcal{S}_w\mathcal{S}_w} = \mathbb{I}_D$. The probability estimate will be performed on this new set $\mathcal{S}_w$, since in this way we have no issue with non-diagonal cross correlations and all the directions will have the same weight in the final result. This means that we can safely use the Euclidean metric when we calculate distances. We have checked that the pre-processing improves the final result for both our estimator and KDE. Pre-processing of the data is in anyway a common first step for many machine learning applications.

The procedure to derive $\hat{f}(\boldsymbol{x})$ with our estimator has already been outlined in the previous subsection and summarized in algorithm 1, including our optimization procedure. For what concerns KDE we will scan over a range of values of bandwidth and for each value do a 5-fold cross validation with the log-likelihood as scoring function. Since we do know the real $f(\boldsymbol{x})$, we were able to check explicitly that our procedure finds optimal or near-optimal values for both MCDE and KDE. Other optimization procedures have been considered for KDE, namely a 10-fold cross-validation, the leave-one-out cross validation and the rule of thumbs from Scott and Silverman [195]. The other optimization procedures lead to analogous results, we chose the 5-fold cross validation since for our simulations it gave the most stable results. As mentioned in section 5.B.3, using the optimal bandwidth from the KDE optimizaton can work well for us, provided the sample is large enough. We then report two plots: one where the optimization of MCDE has been performed with eq. (5.7) and this is fig. 5.1, the other where we took the optimal h of KDE and used it for MCDE, this instead is in fig. 5.2. In the second case it will be easier to see the intrinsic error of the estimator, while in the first one the result will also depend on how good our optimization procedure is.

Using these two approaches we will derive the results for a different number of sample points N, dimensionality D and probability distribution $f(\boldsymbol{x})$.

As error we will consider the mean square error (MSE), which we can calculate, since we know the real PDF. Given a sample $\mathcal{S}$ of size N, the MSE averaged over the sample can be computed as:

$$\mathrm{MSE}(\mathcal{S}) = \frac{1}{N}\sum_{j=1}^{N}[f(\boldsymbol{x}_j) - \hat{f}(\boldsymbol{x}_j)]^2. \tag{5.9}$$

This is exactly the error whose properties we discuss in section 5.A. Other error metrics are of course possible, we considered here the MSE, since it is the error usually considered for consistency. For our estimator it is natural to average the MSE over the sample points, since we started from the vector $\boldsymbol{\pi}$ defined over the sample $\mathcal{S}$. Integrating the MSE and hence derive the MISE would imply giving more relevance on the way we go from $\boldsymbol{\pi}$ to $q(\boldsymbol{x})$, which is not the focus here. Our procedure to assess the performance of our estimator (MCDE) compared to KDE, in a setup where the underlying distribution of data is known, is as follows:

1. We fix the dimensionality D, the number of points N and the univariate PDF $f(x)$. For $D > 1$ we will multiply the PDF along each direction, so that $f(\boldsymbol{x}) = f(x^1, \dots, x^D) = f(x^1)\cdots f(x^D)$.

2. Parameter selection. We generate a training set of N points from the distribution and take the optimal values of bandwidth with the mentioned procedures. Since we want to fix these parameters for all following simulations, we want to be cautious. We can

do this step more than once and take the average of the values we find each time for the bandwidths, so that the final result will be more stable.

3. Testing. We generate a test set of N points from the distribution. We estimate $f(\boldsymbol{x})$ with both MCDE and KDE, with fixed values of bandwidth found in step 2. Using fixed values for the bandwidth avoids the possible overfitting we could encounter by deriving the optimal values each time we generate a sample and makes the results easier to interpret. Then we calculate the MSE averaged over the sample for both KDE and MCDE. This step will be repeated R times, because we want to estimate the statistical uncertainty on the errors and performances.

4. We compute the mean errors and performances:

$$\langle \text{MSE} \rangle = \frac{1}{R} \sum_{i}^{R} \text{MSE}(\mathcal{S}_i) \,, \tag{5.10}$$

$$\mathcal{P} = \frac{\langle \text{MSE} \rangle_{\text{KDE}}}{\langle \text{MSE} \rangle_{\text{MCDE}}} \,. \tag{5.11}$$

Note that step 2. and step 3. can be done multiple times in order to take the mean value of the optimal parameters and of the errors respectively. In our case we repeated step 2. two times and we took $R = 16$ and $R = 9$ for figs. 5.1 and 5.2 respectively. The bandwidth values used for optimization are evenly spaced in the logarithmic interval $[10^0, 10^2]/\sqrt{N}$. Dividing by $\sqrt{N}$ we ensure that the optimal value always lies within the range defined for any N and any D.

In the plots we will show only the statistical error. We have roughly estimated the systematic errors coming from sub-optimal selection of h and they are around the same order of the statistical error showed. In the plots we show on the left the case for a unimodal distribution: $f(x) = \chi^2$ with 5 degrees of freedom. On the right, the case for a multimodal distribution given by a mixture of normal PDFs: $f(x) = \mathcal{N}(0, 2)+\mathcal{N}(8, 3)$, where $\mathcal{N}(\mu, \sigma^2)$ is a Gaussian with mean μ and variance σ^2. The way we consider multidimensional distribution is the one explained in step 1 of the procedure. Of course, many other distributions are possible and could be considered, but we focus on these for simplicity as they have properties that are interesting for our study. The χ^2 distribution is fairly smooth and peaked, but shows asymmetry around the maximum. The multimodal is asymmetric and also has several peaks, which usually are difficult to identify.

In the plots in the upper part of figs. 5.1 and 5.2 we show the convergence when increasing the number of points. For a better representation, we have chosen to show the relative increase compared to when $N = 100$, so what you see is ($\langle$MSE$\rangle$ for N)/($\langle$MSE$\rangle$ for $N = 100$). Looking at the upper row the convergence for increasing N is clear. Here the contributions to the error come from both the intrinsic error of the estimator and the normalization error. Both these types of error are supposed to decrease with increasing size of the sample. Higher dimensionalities are the ones that suffer the most from these errors, so it is not surprising that the convergence is steeper for $D = 6$ rather than for $D = 3$, since not only the intrinsic error of the estimator is decreasing, but also these other contributions.

The behavior of the performance, in the lower row of figs. 5.1 and 5.2, is far less regular, this is because here also the uncertainties on the MSE of KDE contribute to the total error.

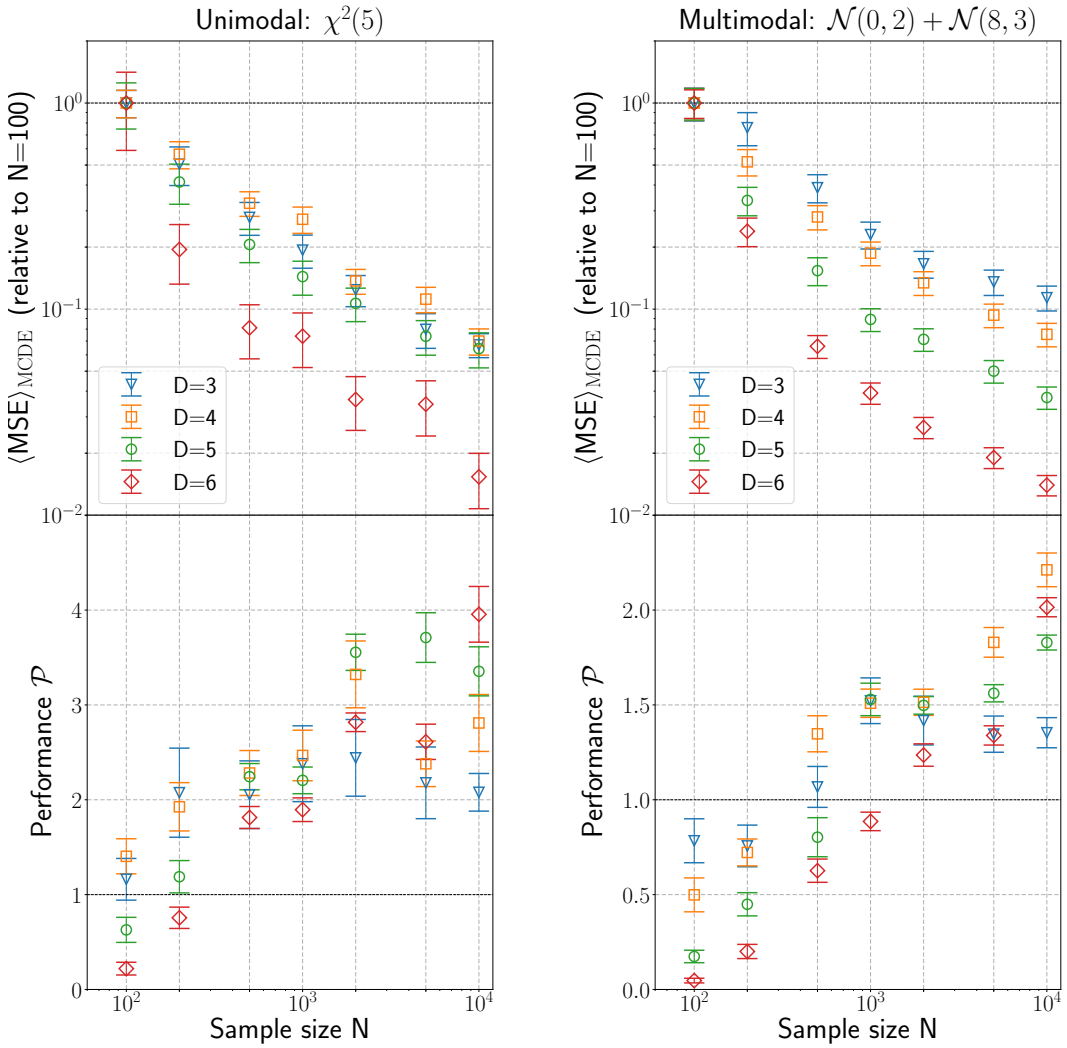

Figure 5.1: The average MSE (5.10) and performance ratio (5.11) of our estimator, for different sample sizes. Different bandwidth optimization have been used for the two estimators: for MCDE we used the procedure outlined in algorithm 1, for KDE we performed a 5-fold cross-validation. The error bars indicate the 2σ statistical uncertainty evaluated with $R = 16$ repetitions. The $\langle \mathrm{MSE} \rangle$ for a value of N is scaled by dividing by the $\langle \mathrm{MSE} \rangle$ obtained for $N = 100$.

Here systematics are of the order of twice the statistical error. Anyway, compatibly with the errors, an enhancing in the performance is present. Here we need to distinguish again between the intrinsic errors of MCDE and the normalization error. Our estimator works better than KDE at higher dimensionality, but this gain is hidden for small samples due to the big uncertainty from normalization, whereas KDE always returns a correctly normalized function. In order for the improvement to become apparent, it is necessary to reduce the error coming from the normalization and hence increase the sample size. The sparser the data, the more points will be needed. For instance for the multimodal at $D = 3$ having 500 points is enough to already perform better than KDE, but at $D = 6$ at least 2000 points are

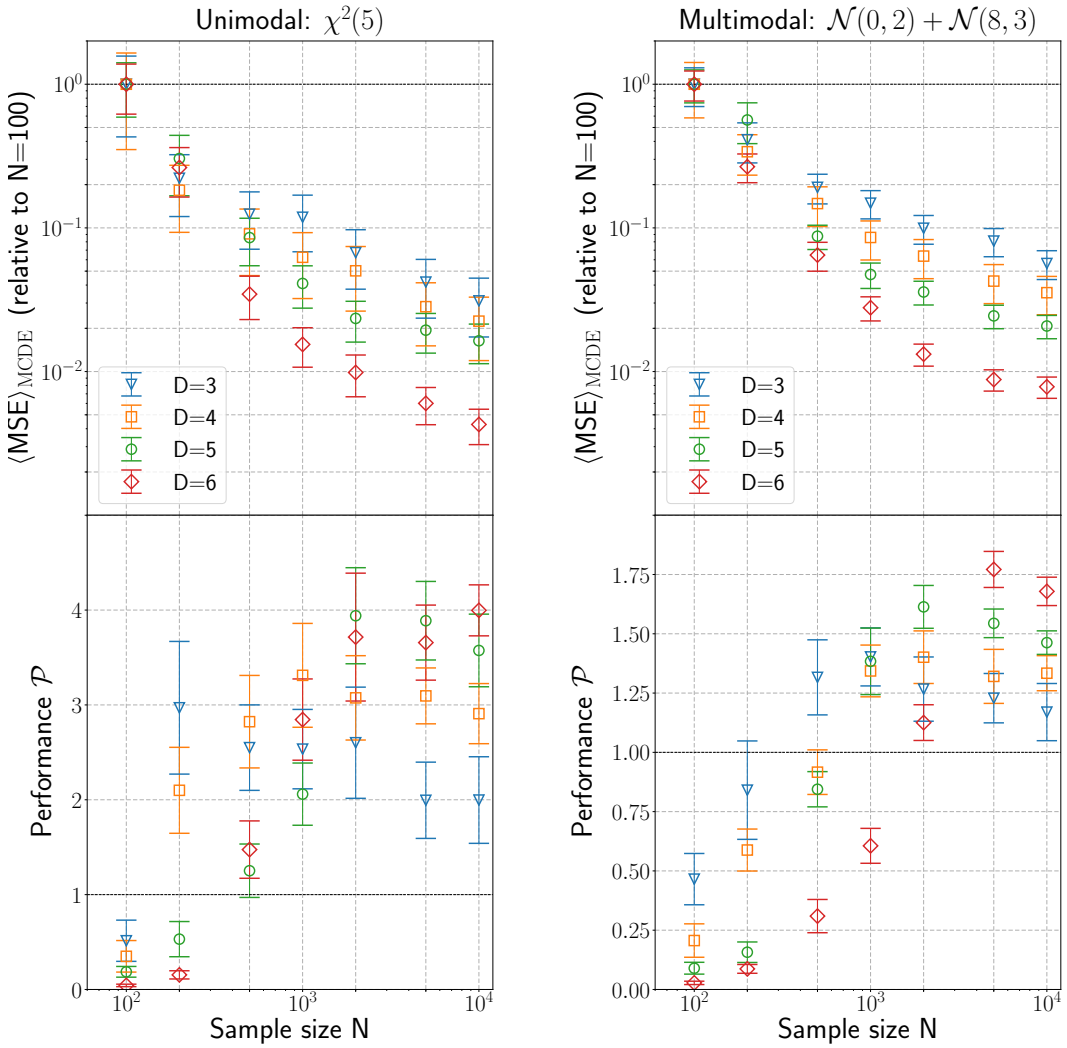

Figure 5.2: The average MSE (5.10) and performance ratio (5.11) of our estimator, for different sample sizes. For both MCDE and KDE the bandwidth used is the same, chosen by performing 5-fold cross-validation on KDE. The error bars indicate the 2σ statistical uncertainty evaluated with $R = 9$ repetitions. The $\langle \mathrm{MSE} \rangle$ for a value of N is scaled by dividing by the $\langle \mathrm{MSE} \rangle$ obtained for $N = 100$.

necessary.

We can see that the two different optimization procedures give consistent results, but there are some differences. First of all, the plot of fig. 5.1 has a less regular behavior, especially in the lower row. This is because the performance is not only influenced by the relative performance of the estimators, but also by the goodness of the optimization procedure. Moreover, the optimization from 5-fold cross validation is very stable, while the one from the log-likelihood of $\tilde{q}(\boldsymbol{x})$ less so. This is because 5-fold cross validation splits the data in 5 sub-samples, so there are 5 scores for each sample, while with our optimization for each sample we only have one score. We know that the optimization procedure of fig. 5.2 should provide good results for

our estimator only for large samples. Indeed, this can be seen in the plots. The convergence in the upper row of the plot in fig. 5.2 seems faster than the one in fig. 5.1, but this is due to the poor performance for small N, not to a greater improvement at large N.

In our procedure we were able to avoid overfitting by performing the probability estimate on samples different from the ones used for the parameter estimation. In real life situations usually only one sample dataset is available and it must be used for both the parameter selection and the testing. It is then natural to wonder whether our algorithm works well when the parameters are estimated from the same set used for testing. We have explicitly checked that in this situation we can still perform better than KDE, provided we have enough points, and that our overfitting is manageable or at least it does not compromise the performance. The case where one sample is used for both parameter optimization and testing is the one of a local outlier detection and we will see that our density estimator can work well for this data analysis task in section 5.4.2.

5.4 Application to local outlier detection

While estimating the probability distribution of data can be useful per se, it is often a tool used for a specific task. As a matter of fact, many data analysis tasks can be density-based and as such require the probability of the data points as an input [196, 197, 198].

In the following, we will focus on a density-based approach to outlier detection, where the probabilities are estimated with MCDE. Of course, our density estimator can be used in principle also for other tasks that take PDFs as input [199, 200].

5.4.1 Description of the method

The aim of outlier detection (also known as anomaly detection) is finding, given a dataset, the elements which are most different from the main population. The task of finding anomalies is very important in real life, for instance it is fundamental in order to detect frauds, but it has applications also in high energy physics, where beyond the standard model signals are expected to constitute a minority population [201]. We will see an example of this in section 5.5.

In this section, we will focus on the case where only one dataset is given, so that outliers should be found based on local properties. Our aim is giving a score to every point of the dataset in order to establish which points are more likely to be anomalies. The way we establish this score for a specific point in a local density-based approach is by looking at how the density in that point compares to the density in the nearby points. An outlier will most likely be distant from the main population and hence have low probability compared to its neighbors.

As mentioned in the introduction, we are interested in new ways to evaluate the density, not in new ways to use the density. For practical purposes we will use an anomaly score similar to the one of [202]. Given a set of points and an estimate of the PDF $\hat{f}$, we can give a score to a specific point $\boldsymbol{x}_i$ by considering its k first nearest neighbors $(\boldsymbol{x}_i^1, \dots, \boldsymbol{x}_i^k)$ and

averaging the density over these points:

$$S(\boldsymbol{x}_i) = \frac{\langle \hat{f}(\boldsymbol{x}_i) \rangle_k}{\hat{f}(\boldsymbol{x}_i)}, \quad \text{where} \quad \langle \hat{f}(\boldsymbol{x}_i) \rangle_k = \frac{1}{k} \sum_{l=1}^{k} \hat{f}(\boldsymbol{x}_i^l) \,. \tag{5.12}$$

The greater the score, the more likely the point will be an anomaly. This score is an indicator of how much our point is declustered from nearby points: the score will be high if the neighbors are in an higher probability region compared to the point. Notice that we are dealing with a local outlier detector: the anomaly score depends on the number of nearest neighbor we consider and the final performance will depend on local properties.

The general procedure by which we do anomaly detection will be the following:

1. Choose a dataset and evaluate $\boldsymbol{\pi}_h$ for various values of h;
2. Normalize the interpolated PDF and choose h^* as the one that minimizes the negative log-likelihood loss function;
3. Build a score as in eq. eq. (5.12) for every point of the sample at varying/fixed k;
4. If possible, consider some performance metric.

Having established how the score is given, it should be clear that we are not worried about the errors coming from interpolation and integration. In order to give a score, we are only interested in the ratio between PDFs, so the overall rescaling with the normalization cancels out and we can work directly with $\boldsymbol{\pi}_{h^*}$. Of course, we still need to establish the optimal h^* and for that we do need to work with normalized PDFs. However, the interpolation/integration error may only result in a sub-optimal selection of h^* and this will not compromise substantially the performance of the detector, as will be clear in the following discussions. Actually, it is important to mention that this optimization procedure is the one that yields the best result. One might wonder why do we bother normalizing the PDFs if other (faster) optimization procedures would still give us a value for the optimal bandwidth. As explained in section 5.B.3, these other optimization procedures rely on KDE results and how well this optimization adapts to MCDE depends on the size of the sample. Especially in section 5.4.2 we will work with high dimensional samples, here the optimal bandwidth for MCDE derived by minimizing the negative log-likelihood will be different from the one derived with an optimization procedure relying on KDE and ultimately lead to a better performance.

Though we do not do it here, there is another reason why it might be worth normalizing the PDF. In the ratios in eq. (5.12) we use the same density estimator for both the numerator and the denominator and here the threshold value $S(\boldsymbol{x}) > 1$ tells us that a point might be interesting. If we use two different density estimators, then in order to retain the same threshold on $S(\boldsymbol{x})$ we need the numerator and denominator to be comparable, hence with the same normalization. Normalizing the PDFs is useful for comparison also in the case where the background is known, as we have in section 5.5.

5.4.2 Numerical experiments

Synthetic datasets

We will first try to establish if our density estimator can be used for anomaly detection in situations under our control. We will generate some datasets in low-dimension ($D = 2, 4, 6$) for different background and anomaly distributions. In this case, we will use a linear interpolator and Monte Carlo integration in step 2 of the procedure, but as mentioned in the introduction, once we have fixed h to its optimal value there will be no need to normalize our estimated PDF. As performance metric we will use the area under the receiving operator characteristics curve (AUC) for different values of k. Other metrics are of course possible, for instance the AUC for the precision-recall curve. In our numerical examples it is almost always the case that the two performances follow the same trend, but we opted for the AUC of the receiving operator characteristic (ROC) as it is easier to interpret and more commonly quoted in the literature. There are also metrics which require a threshold value for the score (F-score, precision, recall), but in anomaly detection setting a threshold value is non-trivial. The reason for this is that the correct threshold depends strongly on the specific dataset, in particular on the imbalance ratio, that is the ratio between the points in the majority and minority population.

For our synthetic data we consider three possible datasets, because we want to encompass different possibilities for the background and signal distributions. All these datasets consider the case where the anomaly is localized, meaning that it is concentrated around some values described by the underlying PDF. We will have a one-dimensional background distribution $f_{\text{in}}(x)$ and a signal distribution $f_{\text{out}}(x)$, we go from 1D to generic D as in the previous part by multiplying along each direction. Once we have fixed the distributions and the dimensionality, the true distribution will be given by $c\, f_{\text{in},D}(\boldsymbol{x}) + (1-c)\, f_{\text{out},D}(\boldsymbol{x})$, where $c/(1-c)$ is the imbalance ratio and $c \in [0,1]$. In order for the score defined in eq. (5.12) to be useful for localized outliers, the neighborhood of each outlier point must include inlier points. For this to be the case it is sufficient (but not necessary) that when we take k neighbors we have $k > (1-c)N_{tot}$. It is also possible that the outliers are not confined within a small region, in which case this procedure works the same way, but there is not a defined minimum value for k. That would be for instance the case when the outliers are uniformly distributed in the domain of the problem.

We will consider each dataset at different dimensions, but we fix the number of points and the imbalance ratio for each dataset.

Datasets:

1. Dataset 1: $N_{in} = 450$, $N_{out} = 50$. $f_{\text{in}} = \mathcal{N}(4, 0.5)$, $f_{\text{out}} = \text{Log-Laplace}(2, 1)$
2. Dataset 2: $N_{in} = 180$, $N_{out} = 20$. $f_{\text{in}} = \text{Exp}(1)$, $f_{\text{out}} = \mathcal{N}(5, 1)$
3. Dataset 3: $N_{in} = 950$, $N_{out} = 50$. $f_{\text{in}} = \text{Gamma}(2)$, $f_{\text{out}} = \text{Gamma}(12)$

The result is presented in fig. 5.3. A clear trend in the dimensionality D and the number of neighbors is present.

We were talking about the fact that the outliers in this example are localized. In the plots a shaded gray area is present, showing the values of k for which we know the performance

might not be satisfactory. Blatantly put, if there are N_{out} outliers, we would need N_{out} neighbors in order to be able to see some inlier points, at least in the worst case scenario. It can be seen that in the gray area the performance is very irregular exactly because of this. On the other hand, once k is sufficiently large, the performance increases with increasing number of neighbors. This is the case only for localized outliers.

The problem of the curse of dimensionality implies that data become sparser at higher dimensions, but this also means that the outliers will be more distant from the main population and easier to identify. As a matter of fact, at higher dimensionality the AUC performance is usually larger with respect to lower dimensionality (for the same value of k). Stated a bit differently, we can see that if we want to get a satisfying performance a smaller value of k is needed at higher dimensions.

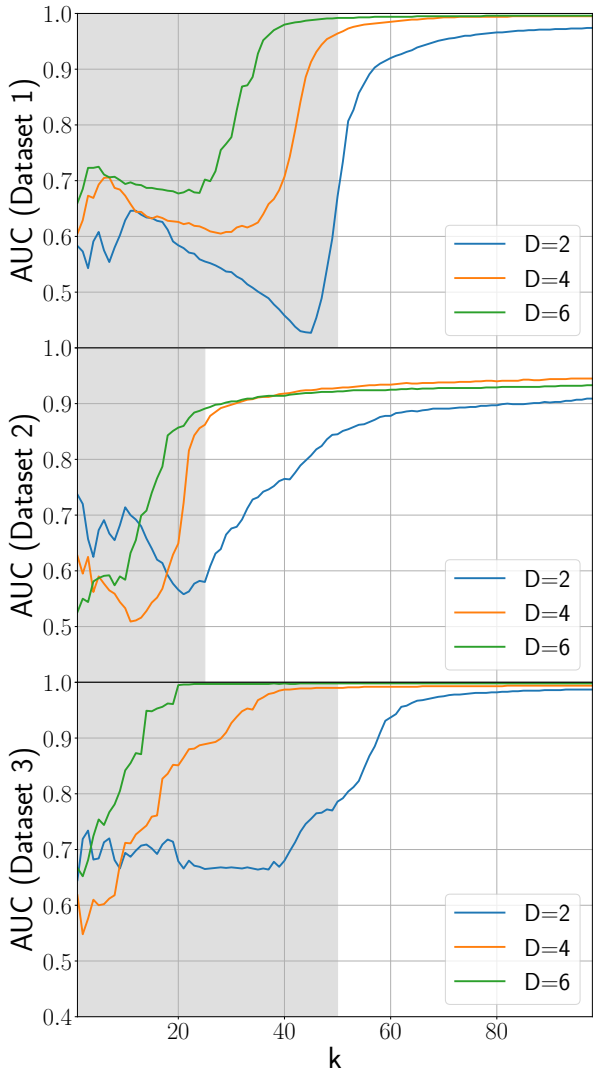

Figure 5.3: AUC performances for our three datasets at different dimensions, for varying number of first neighbors on the x axis. A shaded gray area is present to indicate values of k for which we do not expect a good performance.

Dataset	Method	$k = 5$	$k = 10$	$k = 20$
Breast cancer [203]	MCDE	0.95	0.96	0.96
$N_{in} = 347$, $N_{out} = 10$	RDOS	0.97	0.98	0.98
$D = 30$	LOF	0.93	0.98	0.98
Pen-local [204]	MCDE	0.98	0.98	0.98
$N_{in} = 6704$, $N_{out} = 10$	RDOS	0.98	0.99	0.99
$D = 16$	LOF	0.90	0.99	0.99
Pen-global [205]	MCDE	0.70	0.86	0.95
$N_{in} = 629$, $N_{out} = 90$	RDOS	0.52	0.73	0.98
$D = 16$	LOF	0.51	0.62	0.88
Satellite [206]	MCDE	0.79	0.80	0.92
$N_{in} = 4950$, $N_{out} = 75$	RDOS	0.73	0.77	0.71
$D = 36$	LOF	0.73	0.75	0.68

Table 5.1: The AUC performance for different datasets is presented for benchmark values of k. k can be interpreted as a measure of locality. The values for RDOS and LOF are taken from the plots of Ref. [202].

Real datasets

Since we want to be general for the practical applications, we cannot limit ourselves to the low dimensional cases. This means that the previously employed linear interpolation might be no longer feasible, as the computing time scales exponentially with the dimensionality. For the practical applications at $D > 6$ we will use a nearest neighbor interpolator, which is less precise, but whose computation time is short even at high dimensions. In order to normalize, we will still use the Monte Carlo integration, though we are aware that at high dimensions, if we do not introduce a very high number of points, the calculation will not be accurate. Fortunately, we will see that the performances are still competitive in comparison with other methods. The reason for this is that the optimization still proceeds smoothly, as our negative log-likelihood loss function presents a clear minimum with this interpolation method as well.

In the anomaly detection task there is a well-known issue: there are very few points belonging to the minority class and this creates problems in establishing a metric for the performance, since it might be important not to get false negatives. This is not the case for instance for a classifier with balanced classes. For outlier detectors it is not easy to compare algorithms, since there is no agreement on a performance metric. One other issue deriving from the low population is that it is not possible to split the sample in a training, validation and test set. This is because the test set would not have enough points to be meaningful. This means that the choice of some hyperparameters cannot be pursued by a validation procedure. Of course, we can still fix h for MCDE as we have done before, but in a agnostic approach we have no way of fixing other parameters, e.g. the number of first neighbors.

The work in Ref. [202] was developed in this spirit. In this work there is a comparison between different algorithms with a free parameter k. This parameter k is the number of neighbors to be considered and it is a measure of locality. We compare their results with ours in table 5.1. In this discussion we consider only the two algorithms which work best, namely RDOS [202] and LOF [207]. The value of k is not fixed, so it is important to compare the

Dataset	Method	$k=\sqrt{N}$ $c=0.9$	$k=\sqrt{N}$ $c=0.95$	$k=\sqrt{N}$ $c=0.98$	$k=3\sqrt{N}$ $c=0.9$	$k=3\sqrt{N}$ $c=0.95$	$k=3\sqrt{N}$ $c=0.98$
Ionosphere [208, 209]	MCDE	0.93	0.95	0.98	0.93	0.95	0.98
$N_{in}=225\,, D=34$	LOF	0.91	0.92	0.98	0.94	0.95	0.97
Heart [208]	MCDE	0.63	0.74	0.84	0.70	0.74	0.82
$N_{in}=150\,, D=13$	LOF	0.69	0.74	0.80	0.73	0.73	0.84
Iris [208]	MCDE	0.74	0.84	0.78	0.83	0.86	0.78
$N_{in}=100\,, D=4$	LOF	0.85	0.89	0.83	0.88	0.82	0.68
Parkinson [210]	MCDE	0.42	0.50	0.58	0.43	0.51	0.57
$N_{in}=147\,, D=22$	LOF	0.42	0.52	0.62	0.44	0.52	0.56
Transfusion [211]	MCDE	0.61	0.74	0.77	0.61	0.73	0.74
$N_{in}=570\,, D=4$	LOF	0.60	0.70	0.76	0.69	0.76	0.73
Vehicle [208, 212]	MCDE	0.78	0.85	0.87	0.77	0.83	0.85
$N_{in}=647\,, D=18$	LOF	0.85	0.90	0.91	0.76	0.81	0.85
CMC [208]	MCDE	0.43	0.48	0.46	0.40	0.45	0.43
$N_{in}=1140\,, D=9$	LOF	0.45	0.49	0.46	0.40	0.45	0.43
Yeast [208, 213]	MCDE	0.47	0.48	0.53	0.48	0.51	0.56
$N_{in}=1240\,, D=8$	LOF	0.46	0.47	0.50	0.51	0.56	0.58
PC1 [214]	MCDE	-	0.70	0.72	-	0.71	0.72
$N_{in}=1032\,, D=21$	LOF	-	0.67	0.70	-	0.69	0.69

Table 5.2: AUC for eight benchmark datasets. We have considered two reasonable choices of k and three different imbalance regime ($c = N_{\rm in}/N$). See Ref. [215] for a description of each dataset.

algorithms for different values. In the original paper values between 1 and 32 were considered, so we focused our attention on three benchmark values $k = 5, 10, 20$.

In the first two datasets we are competitive, being at most 0.02 below the best results. Considering the last two datasets we either perform comparably or considerably better, with up to a 0.21 improvement on the second best result. For all datasets our estimator works satisfyingly well for $k = 5$, indicating that our algorithm correctly encapsulates local deviations.

Of course the performance of the outlier detector depends on many factors, but the most relevant one is the imbalance ratio. It is then interesting to consider some benchmark datasets known in the literature, used for instance in [215], and see what happens by changing the number of outliers. For comparison we have used LOF, as it was easy to implement and the algorithm is based on a free parameter k, so that there can be a meaningful comparison with our algorithm. Isolation forests (IF) have also been considered, but what we lack there is a free parameter whose meaning is similar to the one of k. Furthermore, IF are based on a random selection, so the performance depends on this initialization and has a fairly large uncertainty. Qualitatively speaking, taking into account the uncertainties, IF is comparable with the two algorithms mentioned. Though the comparison is not very relevant, because of the issue with the meaning and selection of the parameter mentioned earlier. In any case, neither LOF or MCDE would be overshadowed, as there are several cases where IF loses even taking into account the uncertainties.

Having established our setup, we can look at table 5.2. Three values of c, which we remind is N_{in}/N, are present for comparison. Since we are completely agnostic, we look at $k = \sqrt{N}$, because of the usual rule of thumb. From the analysis of the synthetic datasets we also know that, if we have some reason to believe the outliers are localized, large values of k should show the saturated performance, so we also report the performance for $k = 3\sqrt{N}$. By looking at how many times we win, it seems our outlier detector performs better for larger imbalance ratios.

As stated in the introduction, our goal is not to introduce an algorithm that works better than any other. First of all, it is well known that it is not possible to have an algorithm that performs better in any possible setup [216]. Secondly, we want an easy to interpret algorithm, so we do not want to compare our method with deep neural networks, as they are not the rivals we are interested in. What we wanted to show in this section is that a straightforward application of density estimation (implemented with MCDE) to outlier detection is competitive with other easy to understand approaches to this task. In this sense we are satisfied with the result, as our approach is the best one in a considerable amount of configurations.

5.5 Anomaly detection at LHC

It has been a while now since we last mentioned LHC data, but the time to discuss them has finally arrived. We need of course some data to study and as mentioned in section 5.1 we will be using the ones from DarkMachines. Our aim is not to carry out a precise analysis and deliver the best possible anomaly score, but rather to gather some insight from the data. However, there are also ML algorithms based on density estimation that carry out more accurate analyses [198].

As highlighted for instance in Ref. [217], different outlier detectors applied at the same LHC dataset can find different outliers. In their paper they show that "shallow" outlier detectors (such as ours) can give a different insight on the data with respect to "deep" outlier detectors. We carry out a simple analysis in this spirit: we will study three different signals with a different level of "outlierness", meaning that the relative position of the signal points with respect to the main background will depend on the specific signal. This will be better clarified when we present the data.

We will work with a subsample of luminosity $\mathcal{L} = 0.25\,\text{fb}^{-1}$ and we will work in the channel where $n_{\text{jets}} \geq 2$, $n_{\text{bottom}} \geq 1$ and $n_{\text{lepton}} \geq 1$. This allows to kill the main background generated by the njets SM process, but it is otherwise a pretty agnostic cut on the data. In order to increase the significance of our analysis, we also boost the signal by a factor of 5. In this channel we will then have the numbers of events represented in table 5.3. We see that we will have 53475 events for the background and around 100 events for the signal, meaning that the imbalance ratio will be more than 100. Of course, this imbalance ratio is specific to the cuts we are considering: different cuts could lead to a lower imbalance ratio.

We want to compare now the different signals, so we need to study them in similar ways. For this purpose, we will consider the same features in the analysis: $\not{E}_T$ and p_1, the transverse momentum of the leading jet. The reason we choose these features is that they have a discriminant role for all the three signals considered. The datasets will be used as follows for each signal: both the SM and the BSM dataset will be split in 3 parts. One part of

bkg	#events	#cut
njets	4929478	125
wjets	2634155	5048
gamjets	1981720	592
zjets	938239	1427
ttbar	135304	18740
wtop	79422	9663
...	...	...
total	10941897	53475

signal	#events	#cut
gluino (1TeV)	251	76
stop (400GeV)	1560	243
Z' (2TeV)	483	36

Table 5.3: Number of events before and after the cut for the SM on the left and for our three signals on the right. The name of the SM channels are consistent with the definitions in [5]

the SM is used solely for the evaluation of the background density. Then, one part of SM and one part of BSM will constitute a dataset on which we can evaluate the PDF of signal plus background. On this same dataset we can perform anomaly detection by assigning a score to the points: this dataset simulates a possible dataset coming from LHC and containing BSM physics. We can now move our analysis in two directions: we can either consider the ROC performance on this dataset and calculate the AUC or perform a cut. With this second procedure we cut our parameter space in order to identify the most anomalous region. Since we are trying to guide the cuts at LHC, the cuts performed must be good also for a different dataset coming from the same signal. In order to assess this, we will consider the significance increase in an independent dataset which we call y_{test}, built with a part of SM and a part of BSM, just like the previous one. This analysis is summarized in fig. 5.4. In order for the comparison between different signals to be meaningful, the background points in the three datasets (the one for estimating the background, the one for estimating the background plus signal and y_{test}) are the same and only the signal points are different.

We need to assign a score to each point of the second dataset and intuitively it will still be the ratio between two PDF estimates. Differently from the previous section, we now have the possibility of evaluating two different probability distributions: one containing only the SM and one containing the SM plus possibly signal. We then define the score on the points of the second dataset as:

$$S(\boldsymbol{x}) = \frac{\langle \hat{f}_{s+b}(\boldsymbol{x}) \rangle}{\langle \hat{f}_b(\boldsymbol{x}) \rangle}, \tag{5.13}$$

where $\hat{f}_b(\boldsymbol{x})$ and $\hat{f}_{s+b}(\boldsymbol{x})$ are the probability estimates on the background only (derived from the first dataset) and background plus signal (derived from the second dataset itself). The brakets indicate some averaging: this is useful to avoid false positives, related to the presence of isolated background points. We will perform this averaging in two different ways: we will average over the k nearest neighbor or we will average over the points inside a ball of fixed radius r. The score of truly isolated points will not be smoothed by the second procedure. The two averaging procedures contain a parameter that needs to be fixed in such a way that is robust against different datasets, in one case it will be the number of nearest neighbors, in the other it will be the ratio between the ball radius and the standard deviation of the dataset.

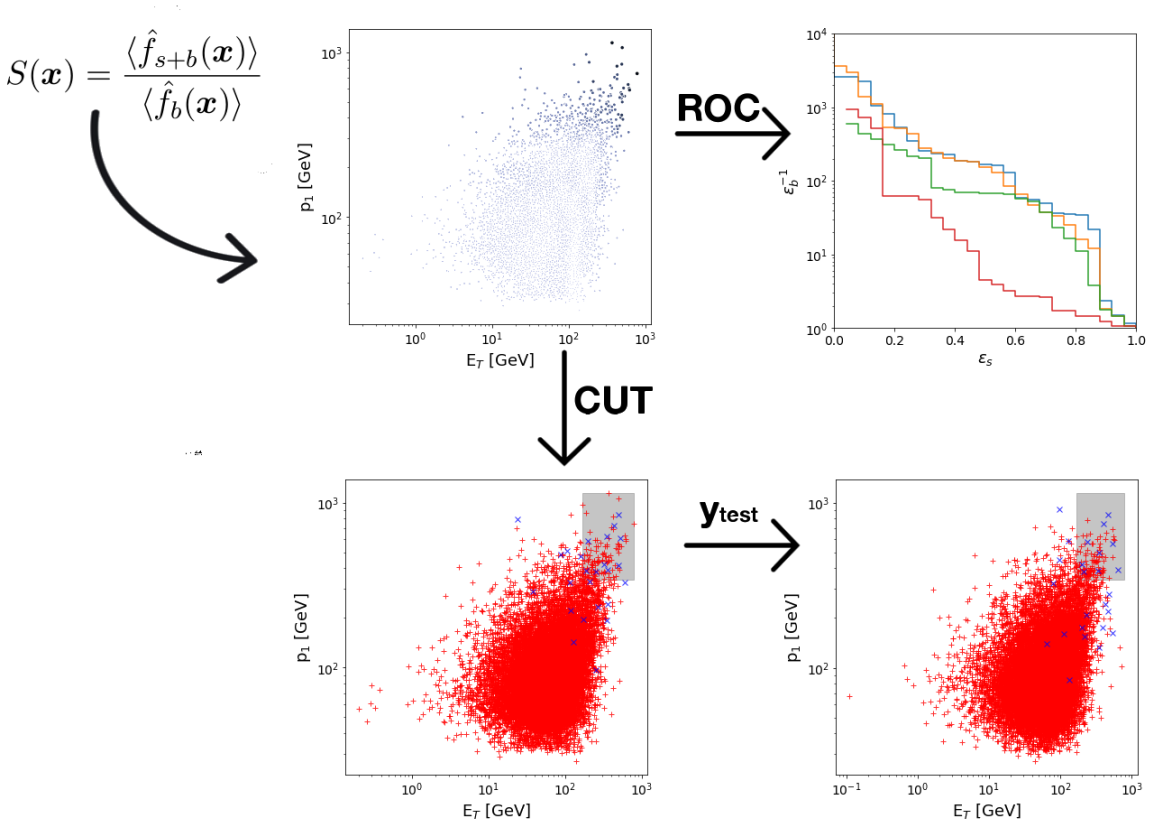

Figure 5.4: Setup of the analysis. A score is assigned to each point of the dataset containing both SM and BSM events. The value of the score is represented by the size and darkness of the points on the top left. Here we perform a ROC analysis and consider the significance increase after the cut on a independent dataset y_{test}. The shaded region indicates the cut found in the second dataset and applied to the third independent dataset y_{test}.

Now comes the interesting part. Nobody is telling us to use the same density estimator for both the background and the background plus signal. Furthermore, we know that KDE and MCDE are similar, but they share some important differences. KDE puts a kernel in each point of the sample, so it is great to evaluate local overdensities. MCDE on the other hand lowers the probability in the tails, if we set $b = 1$. Then, by combining them differently, we will have four outlier detectors looking in different parts of the parameter space. We will call these outlier detectors depending on which estimator is used for the signal plus background and which one is used for the background in eq. (5.13). We will have then K/M, K/K, M/M and M/K, where intuitively K stands for KDE and M stands for MCDE. We already ordered them based on the expected behavior: K/M will be the one looking the most into the tail, while M/K will be the one looking the most into the main population. Of course we do not expect this one to work well, since it is much harder to find outliers in the main population. Furthermore, we know that our density estimator works well for outlier detection, but the improved performance is evident at higher dimensions, so for $D = 2$ we do not expect that

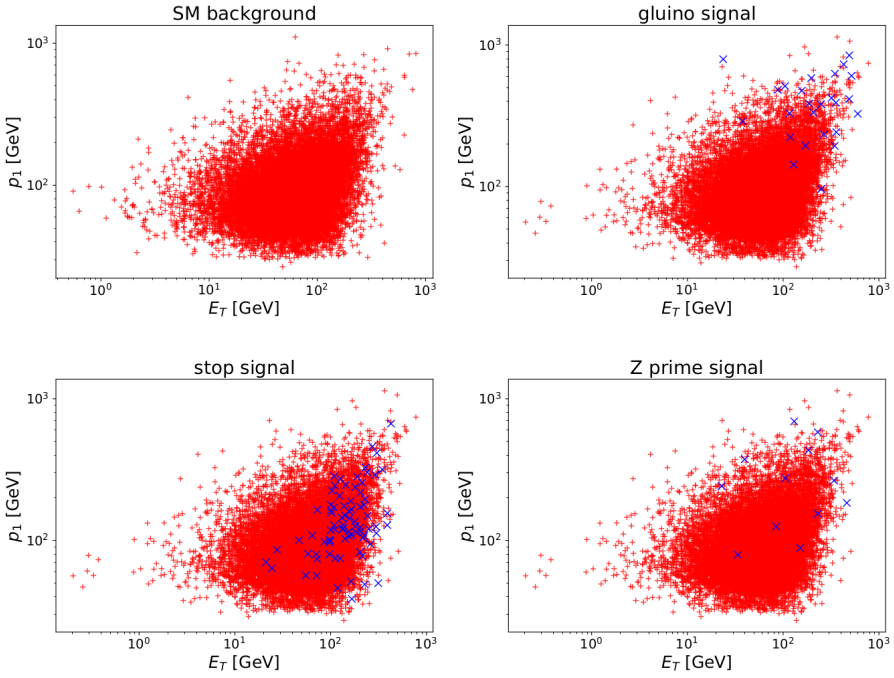

Figure 5.5: Datasets used for anomaly detection: with red + we indicate SM events and with blue × we indicate BSM. The analysis of the datasets containing both SM and signal is performed separately and they have the same SM background.

to be very relevant in the final performance.

Let us look at the datasets we will use: one for the SM background and three from background plus signal, we can see them in fig. 5.5. Here in red we can see the SM events and in blue the BSM events.

We can see what we meant by different degree of "outlierness". Indeed, the signal coming from the stop is the most abundant one, but that does not mean it will be the easiest to study. Many outliers are actually very close to the main population. On the contrary, the gluino signal is rather concentrated in a tail of the distribution and the Z' signal has very few points, distributed rather uniformly in the main population and the tails.

First, let us perform the ROC analysis by considering the AUC as performance metric: once we assign a score as in eq. (5.13), we can directly derive the AUC performance. We do this by averaging in the two ways explained and we report the result in fig. 5.6.

First of all, the two different averaging procedures lead to slightly different results. For what concerns the averaging over k nearest neighbor, we see that $k = \sqrt{N}$ would be a good choice for the parameter. However, the smoothing is such that the differences between the outlier detectors are not so obvious. Anyway, a hierarchy between them is clear and it is the same hierarchy outlined earlier, based on the degree of outlierness. The stop signal is the hardest to identify. Looking at the lower row we see the averaging over balls of fixed radius r, here the hierarchy and the differences between the anomaly detectors are much clearer. In particular, for the gluino signal and apparently for the Z' we can see that K/M outperforms

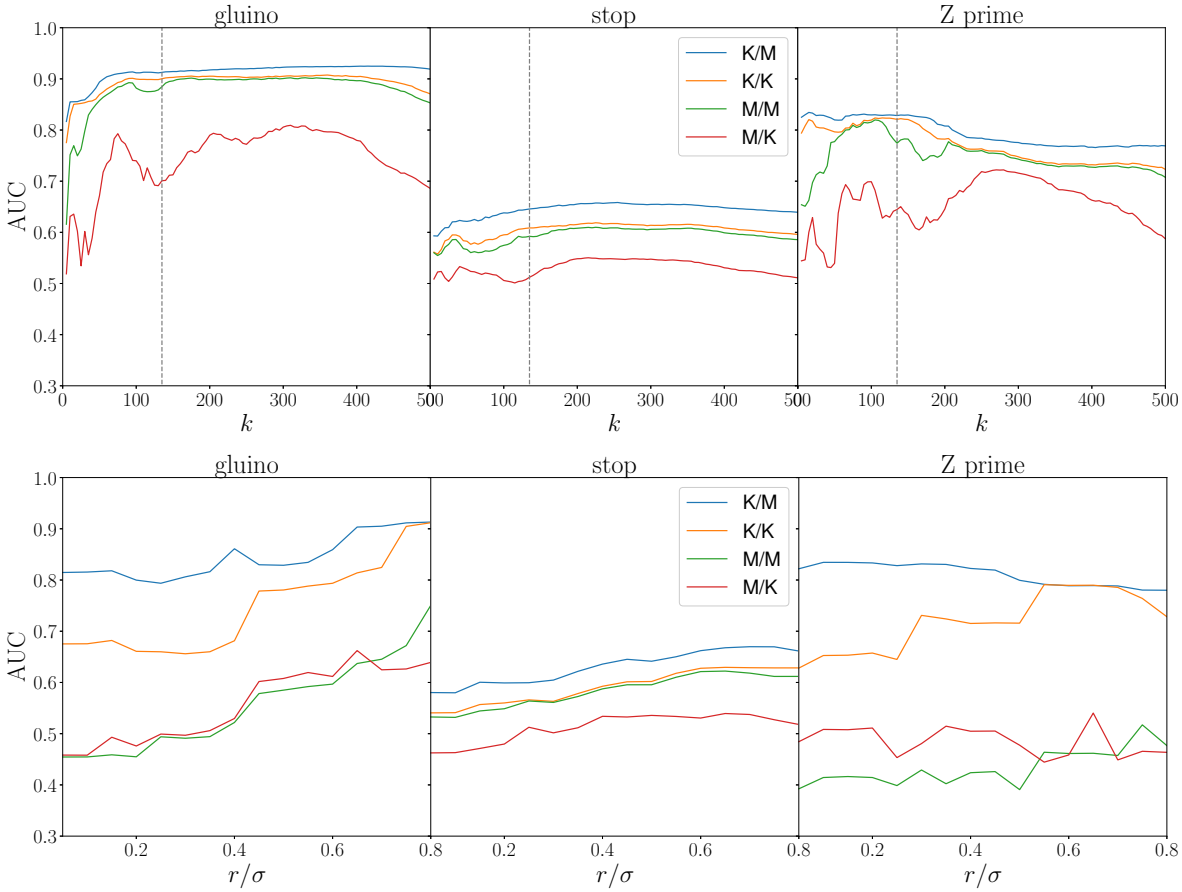

Figure 5.6: ROC analysis with the two averaging procedures for the datasets containg SM plus signal. *On top:* averaging over k nearest neighbors. The value $k = \sqrt{N}$ has been highlighted by a dashed gray line. *In the bottom:* averaging over a ball of radius r.

the other combinations considerably, with the K/K still performing much better than the others. The situation is different for the stop signal. Here the outliers are not really "out" of the main population, so focusing the anomaly detection in the tails does not give a sensible improvement. Even more so, the fact that the outliers here are different from the outliers in the other signals can be inferred from the fact that K/K and M/M are performing very similarly, indicating that the outliers are in fact not in the tails.

The study of the AUC was already very informative, but let us look at a more meaningful discussion about cuts. By considering the score assigned to the points of the dataset, we look at where the top 0.2% of point is and we cut the parameter space in order to identify that region. We now have cuts on the missing energy and the leading jet momentum, we look at how many points from an independent sample (with the same underlying PDF) are inside the cut region. Quantitatively, we consider y_{test} which has N_s signal points and N_b SM points, then we cut deriving N_s^{cut} and N_b^{cut}, meaning the number of outliers and inliers after the cut.

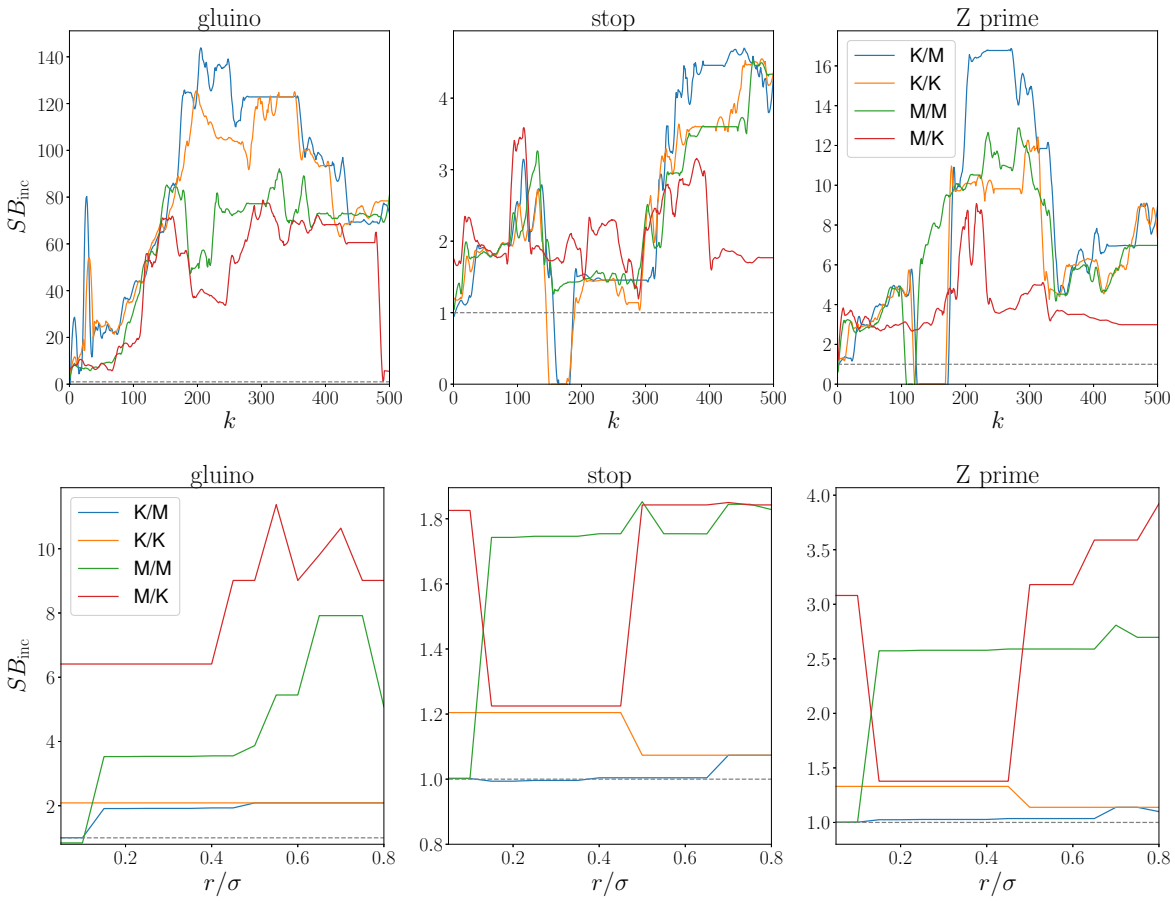

Figure 5.7: Analysis of the signal over background increase after the cut with the two averaging procedures for the three $y_{\rm test}$. A dashed gray line is present for a performance equal to 1: points below this mean that we are making the situation worse. *On top:* averaging over k nearest neighbors. *In the bottom:* averaging over a ball of radius r.

Our performance will be the signal over background increase $SB_{\rm inc}$:

$$\frac{S}{B} = \frac{N_s}{N_b}\,, \quad \left(\frac{S}{B}\right)^{cut} = \frac{N_s^{cut}}{N_b^{cut}}\,, \quad SB_{\rm inc} \equiv \left(\frac{S}{B}\right)^{cut}\left(\frac{S}{B}\right)^{-1} = \frac{N_s^{cut}}{N_s}\frac{N_b}{N_b^{cut}} \tag{5.14}$$

We plot the results in fig. 5.7.

We can see that the results are here instead very sensitive to the averaging parameters. In particular, the averaging over balls of fixed radius does not perform as well as the averaging over k nearest neighbors. This was in some sense expected: the smoothing here is milder, so isolated points can still constitute a considerable source of false positives: the reason why the cut does not work is that the cut is too naive and tends to include areas of the parameter space that are too large. A more cleverly devised cut could improve the situation considerably. With the averaging over k nearest neighbor the result is much better, but it

ultimately depends on the averaging parameter k. Unfortunately, for this performance it does not seem possible to choose a value of k that would yield good results for all datasets. A result from the ROC analysis is still reproduced: the hierarchy between the different outlier detectors and their role in looking at different regions of the parameter space. A direction worth following in the future will be looking at what happens at higher dimensionalities, when the difference between MCDE and KDE is emphasized.

Appendices

5.A Error analysis

In the main part of this chapter we talked about how our density estimator works in practice. There, we showed that under some error metric (defined in eq. (5.10)), our estimator is consistent. We discuss in this first appendix the error analysis of two variations of MCDE in 1D: one obtained from a linear interpolation and the other from an extension of KDE.

We remind that for KDE the necessary requirements for consistency of the estimator are that as $N \to \infty$, we also have that $h \to 0$ and $Nh \to \infty$. In our case a further requirement needs to be satisfied as we also need $V(N)/Nh \to 0$, where $V(N)$ is the volume of the convex-hull of the sample points, which in 1D is simply the difference between the largest sample point and the smallest sample point. The new condition on the volume can be satisfied as long as $V(N)$ goes to infinity slower than N, a condition satisfied if for instance $f(x)$ admits finite first and second moments.

The link between our estimator and KDE discussed in section 5.3.2 can be very useful to prove the consistency of our estimator. In this way we can use results from KDE and apply them to our estimator. After the eigenvector equation, we know that our estimator in the points of the sample can be written as

$$\hat{f}(x_i) \propto \pi_i \propto \frac{1}{N}\sum_k W_{ik} = \frac{1}{Nh}\left[\sum_k K\left(\frac{x_i - x_k}{h}\right) - bK(0)\right] = \hat{f}_{\text{KDE}}(x_i) - b\frac{K(0)}{Nh}, \quad (5.15)$$

where $K(x)$ is a kernel function such that $\int K(x)\,\mathrm{d}x = h$ and $\int x\,K(x)\,\mathrm{d}x = 0$. This is a pointwise estimator and we want to extend its domain to the real axis. We will do it in two different ways and get two different density estimators: $\hat{f}_1(x)$ and $\hat{f}_2(x)$. First, we consider extending the KDE estimate to points not belonging to the sample:

$$\hat{f}_1(x) = C_1\left(\hat{f}_{\text{KDE}}(x) - b\frac{K(0)}{Nh}\right)\mathbb{I}_{\mathbb{D}_1}(x), \qquad \text{with } \mathbb{D}_1 = \{x | \hat{f}_{\text{KDE}}(x) \geq b\frac{K(0)}{Nh}\}, \quad (5.16)$$

here $\mathbb{I}(x)$ is the indicator function and $\mathbb{D}_1$ is the part of the real axis where $\hat{f}_{\text{KDE}}(x) - bK(0)/Nh \geq 0$. Notice that $\mathbb{D}_1$ is not compact in general as it can show discontinuities. C_1 is a constant to be fixed and we decide to fix it by requiring the normalization of $\hat{f}_1(x)$. With these definitions we will have $\hat{f}_1(x) \geq 0, \forall x \in \mathbb{R}$ and $\int_{\mathbb{R}} \hat{f}_1(x)\,\mathrm{d}x = 1$.

The second estimator we build is constructed on a linear interpolation of the pointwise estimate, we call this interpolation $f_{\text{lin}}(x)$. This linear interpolation can be made explicit.

Let us consider the sample points $\{x_i\}$ and let us re-order the sample as $\{x_{(i)}\}$, defined in such a way that $x_{(1)} < x_{(2)} \cdots < x_{(N)}$. We now have for a point $x \in [x_{(i)}, x_{(i+1)}]$ and with $i = 1, , N-1$:

$$f_{\text{lin}}(x) = \hat{f}_{\text{KDE}}(x_{(i)}) + \frac{x - x_{(i)}}{x_{(i+1)} - x_{(i)}} \left(\hat{f}_{\text{KDE}}(x_{(i+1)}) - \hat{f}_{\text{KDE}}(x_{(i)}) \right). \tag{5.17}$$

Clearly it is true that $f_{\text{lin}}(x_{(i)}) = \hat{f}_{\text{KDE}}(x_{(i)})$, since the estimate in the points of the sample does not depend on the interpolation. In this case our estimator will be written as:

$$\hat{f}_2(x) = C_2 \left(f_{\text{lin}}(x) - b\frac{K(0)}{Nh} \right) \mathbb{I}_{\mathbb{D}_2}(x), \qquad \text{with } \mathbb{D}_2 = [x_{(1)}, x_{(N)}]. \tag{5.18}$$

Here $\mathbb{D}_2$ is the convex-hull of the sample points, which is nothing else that the range of the sample. In this case $\mathbb{D}_2$ is compact and at least for non-compact kernels we have $\hat{f}_2(x) > 0, \forall x \in \mathbb{D}_2$. Like before, we fix C_2 by requiring the normalization of the density estimate and with our definitions it holds that $\hat{f}_2(x) \geq 0, \forall x \in \mathbb{R}$ and $\int_{\mathbb{R}} \hat{f}_2(x)\, \mathrm{d}x = 1$.

We keep here the technicalities to a minimum, more details can be found in section 5.B. As we were saying, we will use results from the KDE error analysis, these results hold in the case that $f''(x)$ is continuous and there are no boundary bias terms, so our analysis must be considered under these same conditions. We will have then that $C_1, C_2 \leq 1 + O\left((x_{(N)} - x_{(1)})/Nh\right) + O\left(1/N\right)$. This result holds for both estimators, so in the following we will omit the index on the normalization constant C and keep the discussion general.

All the points of the sample belong to both $\mathbb{D}_1$ and $\mathbb{D}_2$, so there is no need to distinguish between the two when talking about the mean square error averaged over the sample points, which is our error metric. We now calculate bias and variance by leaving C unspecified. This is the part where we will use the KDE results (see e.g. Ref. [191] for these):

$$\begin{aligned}
\text{Bias}[\hat{f}](x_i) &= \mathbb{E}[\hat{f}](x_i) - f(x_i) = C\mathbb{E}[\hat{f}_{\text{KDE}}](x_i) - \frac{C}{Nh} bK(0) - f(x_i) = \\
&= C\left[f(x_i) \int K(w)\mathrm{d}w - hf'(x_i) \int wK(w)\mathrm{d}w + \frac{1}{2}h^2\sigma^2 f''(x_i) \right] + \\
&\quad - f(x_i) - b\frac{CK(0)}{Nh} + O(h^4) = \\
&= (C-1)f(x_i) + \frac{C}{2}h^2\sigma^2 f''(x_i) - b\frac{CK(0)}{Nh} + O(h^4).
\end{aligned} \tag{5.19}$$

Here σ^2 is the variance of $K(x)$: $\sigma^2 \equiv \int x^2 K(x)\mathrm{d}x$. The variance of $\hat{f}$ is even simpler than the bias since:

$$\text{Var}[\hat{f}](x_i) = C^2 \text{Var}[\hat{f}_{\text{KDE}}](x_i) = C^2 \frac{R(K)}{Nh} f(x_i) - C^2 \frac{f^2(x_i)}{N}. \tag{5.20}$$

where we defined the L^2-norm of a generic function ϕ as

$$R(\phi) \equiv \int \phi^2(x)\, dx\,. \tag{5.21}$$

Putting everything together we have the MSE on the points of the sample x_i:

$$\text{MSE}(x_i) = \left((C-1)f(x_i) + \frac{C}{2}h^2\sigma^2 f''(x_i) - b\frac{CK(0)}{Nh}\right)^2 + C^2\frac{R(K)}{Nh}f(x_i) - C^2\frac{f^2(x_i)}{N} + h.o. \tag{5.22}$$

Notice that it is in principle possible to reduce the bias with an appropriate choice of C, notably the bias in the main population is lowered for a value of C slightly larger than 1. We will see in sections 5.B.1 and 5.B.2 that this is our case, since both C_1 and C_2 are equal to 1 plus some small positive contribution.

Since we want to discuss asymptotic properties of our estimators, we now look at the behavior for $N \to \infty$. We have specified earlier that $C \leq 1 + O\left((x_{(N)} - x_{(1)})/Nh\right) + O\left(1/N\right)$ and this gives us enough information for our analysis. If we want to discuss the asymptotic behavior we need to specify the dependence of $x_{(N)} - x_{(1)}$ from N. $x_{(1)}$ and $x_{(N)}$ are points of the sample, so they will be independent of h. Formally, this is a problem of order statistics, more details can be found in section 5.B. We parametrize this contribution as $x_{(N)} - x_{(1)} = V(N)$, where some upper bounds can be found if $f(x)$ satisfies some regularity conditions. Keeping only the lowest orders in $1/N$ and h, we have that the asymptotic (averaged) mean square error (AMSE) is

$$\text{AMSE} \leq \frac{1}{N}\sum_i \left[\left(K(0)\frac{V(N)}{Nh}f(x_i) + \frac{1}{2}h^2\sigma^2 f''(x_i)\right)^2 + \frac{R(K)}{Nh}f(x_i)\right] \tag{5.23}$$

We know that the requirements needed for KDE in order to be consistent are $N \to \infty$, $h \to 0$ and $Nh \to \infty$. In this case a further requirement is necessary: $V(N)/Nh \to 0$. We can see that all these requirements can be satisfied at the same time as long as $V(N)$ does not grow faster than N. Indeed, under the condition that $f(x)$ is smooth enough to have a well-defined mean and variance, we know that $V(N)$ cannot grow faster than $\sqrt{N}$ [218]. This is sufficient to ensure that AMSE goes to 0 for $N \to \infty$ and thus our estimator is consistent under this error metric.

Our expression looks very similar to the one for KDE, except for the volume term. It is instructive to see what happens if the volume term is indeed not there, in that case we recover the usual expressions:

$$\begin{aligned} h^* &= \left(\frac{R(K)}{\sigma^4}\right)^{1/5} N^{-1/5}\left(\frac{\sum_i f(x_i)}{\sum_i (f''(x_i))^2}\right)^{1/5} \\ \text{AMSE}^* &= \frac{5}{4}\left(\sigma R(K)\right)^{4/5}\langle f''^2\rangle^{1/5}\langle f\rangle^{4/5}N^{-4/5}, \end{aligned} \tag{5.24}$$

where $\langle\cdot\rangle$ indicates the average over the sample points. This is very useful in order to understand whether the volume term adds sub-leading terms or slows down the convergence.

Under the condition that $f(x)$ admits mean and variance, we derive that the convergence of the AMSE* to 0 is always guaranteed and it is always faster than $N^{-2/3}$. This rate of convergence is slightly worse than the optimal rate of convergence of KDE, which is $N^{-4/5}$. By looking at the first two terms of eq. (5.23), we can see that the condition for the volume term to be subleading is that the volume should not grow faster than $N^{2/5}$. If this is the behavior of that term, then the AMSE is the same as the one of KDE and AMSE* converges

as $N^{-4/5}$. A family of functions for which this is guaranteed are the functions admitting a moment generating function (see Lemma 5.1 of [219] and the discussion in section 5.B). In this case we have that $V(N)$ cannot grow faster than $\log N$, so the volume term is subleading and we recover the asymptotic properties of KDE.

In our error analysis we relied on the link with KDE in order to see the consistency of the estimator. This consistency has been shown pointwise, meaning that our error metric was evaluated only on the points of the sample. An analysis of global properties, such as the convergence of MISE to 0, clearly needs a different discussion for each of the two estimators, but let us briefly comment on that. For $\hat{f}_1(x)$ the $\text{MSE}(x_i)$ can be extended to every point x of the domain $\mathbb{D}_1$. So the MISE analysis would follow straightforwardly from $\text{MSE}(x)$. However, the same does not hold for $\hat{f}_2(x)$, since the value in a generic $x \in \mathbb{D}_2$ depends on the linear interpolation in eq. (5.17), making the link with KDE hard to implement in the MISE analysis. As a first raw approximation we could consider the trapezoidal integration of $\text{MSE}(x_i)$, since the error is known pointwise. In this approximation the MISE analysis would be similar to the AMSE analysis and consistency would be guaranteed under the same conditions.

Let us conclude with a discussion on the consistency of these interpolators at higher dimensionalities, which is actually the case we reported in section 5.3.4. Practical information about the behavior at higher D and finite N can be found in section 5.B.3. The extension of eq. (5.23) with a generalized bandwidth matrix H is pretty straightforward:

$$\text{AMSE} \leq \frac{1}{N} \sum_i \left[\left(K(0) \frac{V(N)}{N h^D} f(x_i) + \frac{1}{2} h^2 \Sigma(x_i) \right)^2 + \frac{R(K)}{N h^D} f(x_i) \right]. \tag{5.25}$$

Here we have defined $\Sigma(\boldsymbol{x}_i) \equiv \text{Tr}(A^T A \nabla^2 f(\boldsymbol{x}_i))$ and $A = H/h$. It is clear that also in this case the rate of convergence of the AMSE is determined by the behavior of $V(N)$, which is now the convex hull at a generic dimension D and it is harder to treat with respect to the 1D case. As for the 1-dimensional case, the requirement for consistency is that $V(N)$ does not grow faster than N, since in this case the requirements $N \to \infty$, $h \to 0$ and $Nh \to \infty$ can be satisfied at the same time and they are sufficient to assure consistency. If we also have that $V(N)$ does not grow faster than $N^{2/D+4}$, then the volume term is sub-leading and we recover the asymptotic properties of KDE.

5.B Comparison between the two estimators

In section 5.A we performed the error analysis for our density estimators. This section is dedicated to the discussion of more technical details, in particular we will try to understand the asymptotic behavior of C_1 and C_2. We conclude this appendix with some practical discussion about how to use these estimators, focusing on the relative advantages of each of the two. The kernels considered in this discussion are the exponential, the normal and several compact ones (Epanechnikov, uniform, triangular, cosine). Our software is implemented using scikit-learn and hence these kernels are the ones present in our software.

5.B.1 KDE extension

Let us remember our first estimator:

$$\hat{f}_1(x) = C_1 \left(\hat{f}_{\text{KDE}}(x) - b\frac{K(0)}{Nh} \right) \mathbb{I}_{\mathbb{D}_1}(x), \qquad \text{with } \mathbb{D}_1 = \{x | \hat{f}_{\text{KDE}}(x) \geq b\frac{K(0)}{Nh}\}, \tag{5.26}$$

in this case we extended the pointwise estimator to the whole domain by using the KDE estimate. What we mean by this is that in all the points x such that $\hat{f}_{\text{KDE}}(x) \geq bK(0)/Nh$, the estimate will be proportional to $\hat{f}_{\text{KDE}}(x) - bK(0)/Nh$. These points define our set $\mathbb{D}_1$ and outside of this set the estimate is set to 0. Clearly this procedure makes this estimator more similar to KDE with respect to $\hat{f}_2(x)$, in which case we perform a linear interpolation. This estimator not only becomes similar to KDE for $N \to \infty$, but it also reduces to the KDE estimator for $b = 0$, differently to what happens to $\hat{f}_2$.

For the purpose of studying C_1, it is useful to introduce a new set of points $\mathbb{D}'_2 = [x_a, x_b]$, where we have $x_a = \min x \in \mathbb{D}_1$ and $x_b = \max x \in \mathbb{D}_1$. We also introduced the set of points $\mathbb{D}_2 = [x_{(1)}, x_{(N)}]$ in section 5.A, when we discussed the estimator $\hat{f}_2$ and this set of points corresponds to the sample range. The two sets $\mathbb{D}_2$ and $\mathbb{D}'_2$ have similar names, because they look very similar and more often than not they almost coincide. In order to understand this we have that surely $x_{(1)}, x_{(N)} \in \mathbb{D}_1$, but we also know that the kernels are decreasing over a typical length scale of h, so we can state that $x_a = x_{(1)} - O(h)$ and $x_b = x_{(N)} + O(h)$. For compact kernels we can even be more precise and say that $x_{(1)} - h \leq x_a \leq x_{(1)}$ and $x_{(N)} \leq x_b \leq x_{(N)} + h$. Notice that $\mathbb{D}_2$ and $\mathbb{D}'_2$ are both compact.

With this, we are now ready to discuss the constant C_1. We remind that we fix this by requiring the normalization of the estimate:

$$C_1 \int_{\mathbb{D}_1} \hat{f}_1(x) = 1 \quad \to \quad C_1 \left(\int_{\mathbb{D}_1} \hat{f}_{\text{KDE}}(x)\mathrm{d}x - b\frac{K(0)}{Nh} V(\mathbb{D}_1) \right) = 1 \tag{5.27}$$

We know that the KDE estimate is normalized over the real axis: $\int_{\mathbb{R}} \hat{f}_{\text{KDE}}(x)\mathrm{d}x = 1$, so we can recast some terms in function of this equality. We will also use the fact that $\mathbb{D}_1 \subseteq \mathbb{D}'_2$ by definition. If $\mathbb{D}_1$ was already compact, the two sets would be equal. We then rewrite the integral term in eq. (5.27) as:

$$\int_{\mathbb{D}_1} \hat{f}_{\text{KDE}}(x)\mathrm{d}x = 1 - \int_{\overline{\mathbb{D}}_1} \hat{f}_{\text{KDE}}(x)\mathrm{d}x = 1 - \int_{\overline{\mathbb{D}}'_2} \hat{f}_{\text{KDE}}(x)\mathrm{d}x - \int_{\mathbb{D}'_2 - \mathbb{D}_1} \hat{f}_{\text{KDE}}(x)\mathrm{d}x, \tag{5.28}$$

here we denote with an overline the complementary of a set with respect to $\mathbb{R}$: $\overline{\mathbb{D}} \equiv \mathbb{R} - \mathbb{D}$. Now we have all the terms:

$$C_1^{-1} = 1 - \int_{\overline{\mathbb{D}}'_2} \hat{f}_{\text{KDE}}(x)\mathrm{d}x - \int_{\mathbb{D}'_2 - \mathbb{D}_1} \hat{f}_{\text{KDE}}(x)\mathrm{d}x - b\frac{K(0)}{Nh} V(\mathbb{D}_1). \tag{5.29}$$

Notice that there are three contributions that shift C_1^{-1} from 1 and these contributions are all negative. They are small, so in considering C_1 we can expand them and they will become three small positive contributions making C_1 slightly larger than 1. We now turn to describe each contribution separately.

- Boundary term: $\int_{\overline{\mathbb{D}}_2'} \hat{f}_{\mathrm{KDE}}(x)\mathrm{d}x$.

  This term is of order $1/N$. This can be seen by expressing the kernel explicitly:

$$\int_{\overline{\mathbb{D}}_2'} \hat{f}_{\mathrm{KDE}}(x)\mathrm{d}x = \int_{-\infty}^{x_a} \hat{f}_{\mathrm{KDE}}(x)\mathrm{d}x + \int_{x_b}^{+\infty} \hat{f}_{\mathrm{KDE}}(x)\mathrm{d}x = \\ = \hat{F}_{\mathrm{KDE}}(x_a) + 1 - \hat{F}_{\mathrm{KDE}}(x_b) \leq h(\hat{f}_{\mathrm{KDE}}(x_a) + \hat{f}_{\mathrm{KDE}}(x_b)) = 2\frac{K(0)}{N}. \tag{5.30}$$

  The inequality in the second line is the least obvious passage. That inequality holds immediately for a compact kernel, while an explicit computation is needed for the exponential and the Gaussian. From the calculation we actually get an equality for the exponential kernel and an inequality for the Gaussian one.

- Discontinuity in the domain: $\int_{\mathbb{D}_2'-\mathbb{D}_1} \hat{f}_{\mathrm{KDE}}(x)\mathrm{d}x$.

  As easily understood, $\mathbb{D}_1$ can be non-compact and the normalization constant C_1 receives contributions, because we are setting the estimate to 0 in parts of $\mathbb{R}$ where $\hat{f}_{\mathrm{KDE}}$ is not 0. However, by definition, the KDE estimate in this case is small, since we have that $\hat{f}_{\mathrm{KDE}} \leq bK(0)/Nh$. This terms will then be small and we can set an upper bound explicitly:

$$\int_{\mathbb{D}_2'-\mathbb{D}_1} \hat{f}_{\mathrm{KDE}}(x)\mathrm{d}x \leq b\frac{K(0)}{Nh}\int_{\mathbb{D}_2'-\mathbb{D}_1} \mathrm{d}x = b\frac{K(0)}{Nh}(V(\mathbb{D}_2') - V(\mathbb{D}_1)). \tag{5.31}$$

  The difference between the two volumes increases the more discontinuities we have in $\mathbb{D}_1$. This clearly depends on the specific sample and also on the value of h, since more discontinuities will be present the smaller h is. All these non trivial dependencies are taken into account if we combine this term with the next one.

- Volume term: $b\frac{K(0)}{Nh}V(\mathbb{D}_1)$.

  We can add this term to the previous one and get:

$$\int_{\mathbb{D}_2'-\mathbb{D}_1} \hat{f}_{\mathrm{KDE}}(x)\mathrm{d}x + b\frac{K(0)}{Nh}V(\mathbb{D}_1) \leq b\frac{K(0)}{Nh}(V(\mathbb{D}_2') - V(\mathbb{D}_1)) + b\frac{K(0)}{Nh}V(\mathbb{D}_1) = \frac{K(0)}{Nh}V(\mathbb{D}_2') \tag{5.32}$$

  This is the mentioned term of order $V(N)/Nh$. We have that: $V(\mathbb{D}_2') = V(\mathbb{D}_2) + O(h) = x_{(N)} - x_{(1)} + O(h)$. So, in the end we have:

$$\int_{\mathbb{D}_2'-\mathbb{D}_1} \hat{f}_{\mathrm{KDE}}(x)\mathrm{d}x + b\frac{K(0)}{Nh}V(\mathbb{D}_1) \leq b\frac{K(0)}{Nh}V(\mathbb{D}_2') = b\frac{K(0)}{Nh}(x_{(N)} - x_{(1)}) + O\left(\frac{1}{N}\right). \tag{5.33}$$

This last contribution $K(0)(x_{(N)} - x_{(1)})/Nh$ will be the leading one and hence the one relevant for the error analysis in section 5.A. The time has come to talk about $x_{(1)}$ and $x_{(N)}$, as they regulate the volume of $\mathbb{D}_2'$.

These two points are the smallest and largest point of the sample. Because of this, they depend on the $f(x)$ we are estimating and on N, but not on h. From order statistics we know the expressions for their expected values as a function of N and $f(x)$ [220]:

$$\langle x_{(1)} \rangle = -\int_{\mathbb{R}} x \frac{d(1-F(x))^N}{dx}\, dx = N \int_{\mathbb{R}} x f(x)(1-F(x))^{N-1}\, dx, \tag{5.34}$$

$$\langle x_{(N)} \rangle = \int_{\mathbb{R}} x \frac{dF^N(x)}{dx}\, dx = N \int_{\mathbb{R}} x f(x) F^{N-1}(x)\, dx. \tag{5.35}$$

This clearly depends on the specific distribution $f(x)$ and even knowing $f(x)$ there are usually no analytic solutions for general N.

We use two bounds on $x_{(1)}$ and $x_{(N)}$ deriving from properties of $f(x)$. A first loose bound was already mentioned and comes from the requirement that $f(x)$ admits a mean μ and a variance σ^2 [218]. In this case, we know that $\langle x_{(1)} \rangle \geq \mu - \sigma\sqrt{N}$ and $\langle x_{(N)} \rangle \leq \mu + \sigma\sqrt{N}$. This requirement is very loose, by being more strict and asking that the function $f(x)$ admits a moment generating function (a requirement equivalent to admitting a bilateral Laplace transform or admitting moments of every order), we can also derive that $x_{(N)}$ cannot grow faster with N than $\log N$. This can be understood by looking at Lemma 5.1 of [219]. In particular, in the proof we find the inequality:

$$\langle x_{(N)} \rangle \leq \inf_{s>0} \frac{\log N + \log(m(s))}{s}, \tag{5.36}$$

here $m(s)$ is the moment generating function of the unknown distribution. $m(s)$ is a strictly increasing function of s that takes values between 1 and $+\infty$. This means that for any N we can find $\bar{s}$ such that $m(\bar{s}) = N$. In particular, since $m(s)$ is increasing with s, we have that increasing N will also increase $\bar{s}$. This last observation tells us that increasing N, the denominator of eq. (5.36) will also increase. Putting everything together and considering $s = \bar{s}$, we conclude that $x_{(N)} - x_{(1)}$ cannot grow faster than $\log N$ for functions admitting a moment generating function.

Let us summarize our results. C_1^{-1} is equal to 1 except for three negative contributions. A boundary term of order $1/N$ and a term which is less than $V(N)/Nh$ in absolute value. Depending on the properties of $f(x)$ we can set some upper bounds on $V(N)$: if $f(x)$ admits first and second central moments, then $V(N)$ cannot increase faster than $N^{1/2}$. If $f(x)$ admits all moments, then we know that it cannot increase faster than $\log N$.

5.B.2 Linear interpolation

Let us now recall the second estimator, the one obtained from a linear interpolation:

$$\hat{f}_2(x) = C_2 \left(f_{\text{lin}}(x) - b\frac{K(0)}{Nh} \right) \mathbb{I}_{\mathbb{D}_2}(x), \qquad \text{with } \mathbb{D}_2 = [x_{(1)}, x_{(N)}]. \tag{5.37}$$

A multidimensional generalization of this estimator is the one we used for numerical experiments. The two estimators $\hat{f}_1$ and $\hat{f}_2$ work differently and hence give different results. Already the first estimator was consistent, but this other estimator is more "stable". The reasons for this will be made explicit in section 5.B.3, where a comparison between the two will be presented.

Also in this case we fix the constant C_2 by requiring the normalization of $\hat{f}_2(x)$:

$$C_2 \int_{\mathbb{D}_2} \hat{f}_2(x) = 1 \quad \rightarrow \quad C_2 \left(\int_{\mathbb{D}_2} f_{\text{lin}}(x) \mathrm{d}x - b \frac{K(0)}{Nh} V(\mathbb{D}_2) \right) = 1 \tag{5.38}$$

First, let us point out that the integral of the linear interpolation is nothing other than the integral calculated with the trapezoidal rule over a non-uniform grid. This grid will have nodes at $x_{(k)}$ and spacings $h_k \equiv x_{(k+1)} - x_{(k)}$. Deriving an expectation value for these spacings is again a non-trivial order statistics problem. Let us use here the properties of KDE to discuss the integral term, as we did for C_1:

$$\int_{\mathbb{D}_2} f_{\text{lin}}(x) \mathrm{d}x = \int_{\mathbb{D}_2} \left(f_{\text{lin}}(x) - \hat{f}_{\text{KDE}}(x) \right) \mathrm{d}x + \int_{\mathbb{D}_2} \hat{f}_{\text{KDE}}(x) \mathrm{d}x = 1 + \delta_{\text{trapz}} - \int_{\overline{\mathbb{D}}_2} \hat{f}_{\text{KDE}}(x) \mathrm{d}x, \tag{5.39}$$

where we have defined $\delta_{\text{trapz}} \equiv \int_{\mathbb{D}_2} \left(f_{\text{lin}}(x) - \hat{f}_{\text{KDE}}(x) \right) \mathrm{d}x$. This term can be a priori either positive or negative, which already constitutes a difference with respect to C_1.

Let us made the contributions to C_2 explicit:

$$C_2^{-1} = 1 - \int_{\overline{\mathbb{D}}_2} \hat{f}_{\text{KDE}}(x) \mathrm{d}x - b \frac{K(0)}{Nh} V(\mathbb{D}_2) + \delta_{\text{trapz}}. \tag{5.40}$$

This expression is very similar to the one derived for C_1: there are a boundary term and a term proportional to $V(\mathbb{D}_2)$, which in the previous part came from the combination of two different contributions. The difference lies in the addition of the third term δ_{trapz}. We will see that this contribution is usually positive, thus partially compensating the other two negative terms.

In the treatment of the contributions, in order to be as general as possible, we will consider two portions of $\mathbb{D}_2$: one where we have $h_k \leq h$ asymptotically and the other where we have $h_k > h$.

We now turn to describe each contribution separately.

- Boundary term: $\int_{\overline{\mathbb{D}}_2} \hat{f}_{\text{KDE}}(x) \mathrm{d}x = \int_{-\infty}^{x_{(1)}} \hat{f}_{\text{KDE}}(x) \mathrm{d}x + \int_{x_{(N)}}^{+\infty} \hat{f}_{\text{KDE}}(x) \mathrm{d}x$.

  This term is of order $1/N$, under the assumptions we are considering. Let us express this more explicitly:

$$\begin{aligned} \int_{-\infty}^{x_{(1)}} \hat{f}_{\text{KDE}}(x) \mathrm{d}x + \int_{x_{(N)}}^{+\infty} \hat{f}_{\text{KDE}}(x) \mathrm{d}x = \hat{F}_{\text{KDE}}(x_{(1)}) + 1 - \hat{F}_{\text{KDE}}(x_N) \leq \\ \leq h(\hat{f}_{\text{KDE}}(x_{(1)}) + f_{\text{KDE}}(x_{(N)})). \end{aligned} \tag{5.41}$$

  The inequality comes from the same reasoning carried out in the previous section. If we have $h_{N-1} \gtrsim h$, then the largest sample point is isolated and we have $\hat{f}_{\text{KDE}}(x_{(N)}) =$

$(Nh)^{-1}\sum K(x_{(N)}-x_i) \simeq (Nh)^{-1}K(0)$ and we recover the order $1/N$. This approximated equality becomes an equality for compact kernels and $h < h_{N-1}$. The same reasoning holds for $x_{(1)}$. Terms of order h could be present, if $f(x_{(1)})$ or $f(x_{(N)})$ would not go to 0 for increasing N. However, this would imply that KDE has problems with boundary bias, a condition we need to avoid in order to carry out the error analysis. Ways to treat this situation and improve the convergence are discussed in section 5.B.3, where we refer to standard solutions to the issue.

- Volume term: $b\frac{K(0)}{Nh}V(\mathbb{D}_2)$.

  The discussion of this term is identical to the one carried out in the previous section and the same upper bounds on its behavior hold.

- $\delta_{\text{trapz}} : \int_{\mathbb{D}_2}\left(f_{\text{lin}}(x) - \hat{f}_{\text{KDE}}(x)\right)\mathrm{d}x$

  Here it is very useful to distinguish between the two regimes of h_k, since we want to be general. We will show that this term is (asymptotically) less than the volume term with $b = 1$ in absolute value. In order to easily compare, let us split the volume contribution and this contribution over the intervals between each sample point:

$$\delta_{\text{volume}} = \frac{K(0)}{Nh}V(\mathbb{D}_2) = \frac{K(0)}{Nh}\sum_{k=1}^{N-1} h_k = \sum_{k=1}^{N-1} h_k \delta_{\text{volume}}(k) \tag{5.42}$$

$$\begin{aligned}\delta_{\text{trapz}} &= \sum_{k=1}^{N-1}\frac{h_k}{Nh}\sum_i\left[K\left(\frac{x_{(k)}+h_k-x_i}{h}\right) + K\left(\frac{x_{(k)}-x_i}{h}\right) - \frac{\int_{x_{(k)}}^{x_{(k)}+h_k} K(\frac{x-x_i}{h})\mathrm{d}x}{h_k}\right] = \\ &= \sum_{k=1}^{N-1} h_k \delta_{\text{trapz}}(k)\end{aligned} \tag{5.43}$$

The term in eq. (5.43) has been derived by considering the trapezoidal integration error explicitly. Notice that this integration error is not really an error for our estimator, meaning that it is present because of the way our estimator is defined.

In the first regime we have $h_k/h < 1$, so we can expand around small h_k getting (this is nothing else that the error for the trapezoidal rule):

$$\delta_{\text{trapz}}(k) = \frac{h_k^2}{12Nh^3}\sum_i K''\left(\frac{x_{(k)}-x_i}{h}\right). \tag{5.44}$$

Now, by comparing interval by interval the volume term and the trapezoidal term, we would like to show that:

$$\delta_{\text{volume}}(k) = \frac{K(0)}{Nh} \geq \frac{h_k^2}{12Nh^3}\sum_i K''\left(\frac{x_{(k)}-x_i}{h}\right). \tag{5.45}$$

In order to show this, we first consider that only a finite number of terms is contributing to the trapezoidal term, namely all x_i such that $x_{(k)} - x_i \lesssim h$. The number of such points is of the order of h/h_k. Secondly, we have that for all the kernels we considered $K''((x_{(k)} - x_i)/h) \leq |K''(0)| \leq \gamma K(0)$. Then putting all the inequalities together we have:

$$\frac{h_k^2}{12Nh^3} \sum_i K'' \left(\frac{x_{(k)} - x_i}{h} \right) \lesssim \frac{h_k^2}{12Nh^3} \gamma K(0) \frac{h}{h_k} = \frac{\gamma}{12} \frac{h_k}{h} \frac{K(0)}{Nh}. \tag{5.46}$$

This term is indeed smaller than $K(0)/(Nh)$ for $h_k \leq h$ for all the kernels considered, an explicit value of γ is reported in table 5.4. Intuitively and from numerical simulations, we gather that the contribution to δ_{trapz} in this regime will be much smaller than the ones when $h_k > h$. This is also due to the fact that here these term can be either positive or negative, so there will be cancellations.

In the second regime we have isolated points, this means that we have $f_{x_{(k)}} \simeq f_{x_{(k+1)}} \simeq K(0)/Nh$ (this equality is exact for compact kernels). In this case we cannot expand around small h_k, but it is easy to evaluate δ_{trapz} explicitly:

$$\delta_{\text{trapz}}(k) = \frac{1}{h_k} \left(\int_{x_{(k)}}^{x_{(k)}+h_k} \left(f_{\text{lin}}(x) - \hat{f}_{\text{KDE}}(x) \right) \mathrm{d}x \right) \simeq \frac{K(0)}{Nh} - \frac{K(0)}{Nh_k}. \tag{5.47}$$

This is equal in absolute value to the volume term in the asymptotic limit and smaller in the other cases, but it is always positive. Since this is in general larger than the previous contribution, we can safely state that in general δ_{trapz} is positive.

The description of δ_{trapz} we carried out would break down if there were strong discontinuities in the spacings h_k. These can be due to two effects: discontinuities in $f(x)$, which can be avoided by requiring a continuous $f(x)$. Or they can be due to statistical fluctuations, these are however not supposed to be present in the asymptotic limit $N \to \infty$.

For the purpose of considering every term contributing to $\delta_{\text{trapz}}(k)$, we need to make a further comment for what concerns compact kernels. In the case $h_k \leq h$ we have three sub-cases (let us assume $x_i < x_{(k)}$ for simplicity): $x_i + h < x_{(k)}$ and in this case there is no contribution to $\delta_{\text{trapz}}(k)$. Then, we can have $x_i + h > x_{(k)} + h_k$ and the contribution is the same described before. Finally, we have a contribution if $x_{(k)} < x_i + h < x_{(k)} + h_k$. In this case the number of points contributing will be order 1 and the scaling of the contribution will depend on the specific kernel. More explicitly we will have contributions of the form:

$$\delta_{\text{trapz}}(k) = \frac{1}{Nh} \sum_i \left(K \left(\frac{x_{(k)} - x_i}{h} \right) - \frac{1}{h_k} \int_{x_{(k)}-x_i}^{h} K \left(\frac{x}{h} \right) \mathrm{d}x \right) \leq \frac{C}{Nh_k} \left(\frac{h_k}{h} \right)^{\alpha} a^{\beta}, \tag{5.48}$$

here C is an order 1 term with subleading dependencies on h_k/h and a. a is defined as $a \equiv (x_i + h - x_{(k)})/h_k$ and so $a \in [0, 1]$. α and β are kernel dependent and we report

Kernel	$K(u)$	support	α	β	γ
Gaussian	$\exp(-u^2/2)$	$\mathbb{R}$	-	-	1
Exponential	$\exp(-\lvert u\rvert)$	$\mathbb{R}$	-	-	1
Uniform	1	$\lvert u\rvert \leq 1$	1	0	0
Triangular/Linear	$1-\lvert u\rvert$	$\lvert u\rvert \leq 1$	2	1	0
Epanechnikov	$1-u^2$	$\lvert u\rvert \leq 1$	2	1	2
Cosine	$\cos(\pi u/2)$	$\lvert u\rvert \leq 1$	2	1	$\pi^2/4$
Biweight	$(1-u^2)^2$	$\lvert u\rvert \leq 1$	3	2	4
Triweight	$(1-u^2)^3$	$\lvert u\rvert \leq 1$	4	3	6

Table 5.4: Values of $\gamma \equiv |K''(0)|/K(0)$, α, β. α and β are defined in eq. (5.48) and apply only to compact kernels. We reported the expression for the kernel functions as implemented in KDE. We usually expressed the kernel as a function of the distance, in which case we just need to take $d = h|u|$ in 1D. Biweight and triweight have not been implemented in the software we provide, but they would give no issues with the performance as can be understood from the values of α, β, γ.

their values in table 5.4. For all compact kernels (except the uniform), we have $\alpha \geq 2$ and $\beta \geq 1$, so this additional term gives no problems. For the uniform kernel we have $\alpha = 1$, but we also have positive contributions for $a < 1/2$ and negative for $a > 1/2$. So this term is not problematic, if there are no strong discontinuities in h_k, a condition we had already imposed to evaluate the trapezoid contribution.

Let us summarize our results. C_2^{-1} is equal to 1 except for three contributions: two always negative, one usually positive. The volume term and the trapezoid term are both bounded in absolute value by something of the order of $V(N)/Nh$. Depending on the properties of $f(x)$, we can set some upper bounds on $V(N)$: if $f(x)$ admits first and second central moments, then $V(N)$ cannot increase faster than $N^{1/2}$. If $f(x)$ admits all moments, then we know that it cannot increase faster than $\log N$. There is then a boundary term which is of order $1/N$ under the assumptions needed to carry out our error analysis.

5.B.3 Practical tips

The previous part of this section was dedicated to a comparison between the asymptotic properties of the two estimators, while here we will talk about when N is actually finite. The previous discussion was fairly theoretical, while now we turn to more practical distinctions between the two. We also conclude by mentioning how to treat problematic boundary terms, an issue afflicting our estimators and KDE alike.

Large N, small N

Asymptotic properties for $N \to \infty$ are important to show consistency, but in practice we always work with finite N. In the following, we will often talk about properties that hold for large or small N. These concepts of large and small N are loose, by large N we mean a situation where the domain is densely populated and we are then close to the asymptotic description. On the contrary, small N indicates that the population is rather sparse. It

should be clear that these concepts depend on the dimensionality and the distribution we are studying: $N = 1000$ is large, if we are studying a unimodal distribution in 1D, but it is rather small if we want to study a multimodal distribution in 6D.

Optimization

Under the conditions specified in section 5.A, we have that our estimators and KDE share the same asymptotic properties. Knowing this and looking at eq. (5.15), we see that for large N our estimators and KDE will be similar to each other, more specifically we expect the optimal bandwidths of all these estimators to be comparable. This helps us for optimization. As a matter of fact, since the optimization of KDE is widely discussed, we can rely on finding the optimal bandwidth for KDE and then use that for our estimators. This is true whether we establish the bandwidth from a rule of thumb or from a data-based optimization. Indeed, this is what we used in fig. 5.8, where we used a 5-fold cross validation to determine the optimal bandwidth of KDE and then used the same value for our estimators. For these plots, we have checked that a 10-fold cross validation or the leave-one-out cross validation would have yield similar results. Of course, using the same bandwidth for our estimators is an approximation, which holds better, the larger N is.

However, we do not need to introduce KDE in order to optimize. For both $\hat{f}_1$ and $\hat{f}_2$ we can minimize the negative log-likelihood before rescaling with C_1 and C_2. This is equivalent to considering:

$$\text{Loss}(h) = -\sum_i \log\left(\hat{f}_{\text{KDE}}(x_i) - \frac{K(0)}{Nh}\right). \tag{5.49}$$

From simulations we know that this loss function shows a clear minimum. We can realize that this loss function works, since it is equal to the leave-one-out log-likelihood function, apart from a term independent of h. Optimization based on this loss function is very fast with the software we provide, since there is an handle that can be used if we do not want to normalize.

The most immediate choice for a loss function would be the negative log-likelihood of the estimate after the rescaling with C_1 or C_2. However, from the theory of KDE we know that the negative log-likelihood of the KDE estimate does not work well as a loss function, this is because there is a trivial minimum for $h = 0$, where the estimate becomes a series of delta functions. Similarly, this loss function does not work for $\hat{f}_1$, even for $b = 1$. The situation is different for $\hat{f}_2$, whose negative log-likelihood

$$\text{Loss}(h) = -\sum_i \log(\hat{f}_2(x_i)). \tag{5.50}$$

shows a single non-trivial minimum for $b = 1$. A minimum is usually found also for $0 \leq b < 1$. This loss function is specific to our estimator and it is also the one used in the main text, both for the linear interpolator and the nearest neighbor interpolator. Since it is in some sense "new", let us comment a little more on it. As all loss functions, there are situations for which it does not provide the best possible value of bandwidth. In our simulations, we have found that a very good value of the bandwidth is found for $D \geq 2$ for any kernel, while non-optimal values are found at $D = 1$ for kernels that are not the Gaussian one. The optimization works better for large values of b, although a minimum in the loss function is usually found

even for $b = 0$. In the main text we worked in the situation $b = 1$ and Gaussian kernel, so everything was perfectly under control. The value of h found here is in general different from the one derived from eq. (5.49), since C_2 depends on h and hence enters in the value of the loss function. Since eq. (5.50) requires a longer computation with respect to eq. (5.49), we need to have some advantage in doing so. Indeed, it is clear that eq. (5.50) encodes more information, as eq. (5.49) is calculated before rescaling with the proportionality constant and does not even distinguish $\hat{f}_1$ and $\hat{f}_2$. C_1, just like C_2, depends on h, so we might think that $\text{Loss}(h) = -\sum_i \log(\hat{f}_1(x_i))$ should work well for optimizing $\hat{f}_1$. As we stated earlier, that is usually not the case and this can be understood by looking at the h dependent contributions to C_1 and C_2. C_1 receives a positive contribution proportional to $1/Nh$, while the dependence of C_2 is less trivial, since there are two terms of similar order compensating each other. This implies that C_1 (and in turn $\hat{f}_1$) has a strong dependence on h, while the same is not true for C_2. This is what ultimately leads to the stability of the loss function in eq. (5.50) and to the fact that the performance of $\hat{f}_1$ depends on the bandwidth much more than the one of $\hat{f}_2$.

This last optimization procedure gives reliable results and it is the one that relies the least on KDE results, thus being the most self consistent. It has unfortunately the disadvantage of being lengthy to compute, since we are normalizing for each value of h. The loss function in eq. (5.50) is the one we used throughout the main text of this chapter.

Numerical comparison between the estimators

We will now consider the same setup of the main text, meaning two distributions (one unimodal, the other multimodal) and several sample sizes and dimensionalities. But here, rather than the convergence, we will focus on the relative behavior of KDE and our two estimators. The results are reported in fig. 5.8. For what concerns optimization, we have used the same bandwidth for each estimator, this bandwidth was derived with a 5-fold cross validation where the bandwidths spanned the linearly spaced range $[0.05, 1]$. The reason we use the same bandwidth is to highlight the different behaviors of the estimators, rather than the effects of optimization. We report the mean square error averaged over the sample points as a measure of the error at different dimensionalities and for different sample sizes. The mean values and the error bars are derived by making $R = 9$ repetitions in the same way we did for section 5.3.4. Let us look at what we can say.

First of all, there is no clear winner, meaning that, depending on the size of the sample and the dimensionality, we get different hierarchies between the estimators. However, some common trend is exhibited.

$\hat{f}_2$ is the most regular one: bad for small samples and increasingly good as we enlarge the sample size. The errors for small samples are due to the poor reconstruction of the interpolator, changing the bandwidth does not improve the result much. $\hat{f}_2$ is less sensitive to the h selection with respect to $\hat{f}_1$, this also explains part of its regularity.

$\hat{f}_1$ on the other hand is far less regular, this can be understood because there can be a large positive contribution to C_1 as discussed in section 5.B.1. This contribution is inversely proportional to Nh, so it makes things worse for small h and small N. This implies that the estimator is well-behaved at large N, but not at small N. From the plots it is clear that the estimator is most irregular for $D \geq 5$. It is anyway possible to check a posteriori if we are in the well-behaved case or not by looking at C_1. When we perform the density estimate, we can ask the software to return the value of C_1 to us, so we can see how much

it differs from 1 and how worried we should be. Differently from the case of $\hat{f}_2$, here the value of h enters critically in the goodness of the estimate. This strong dependence on h is also perceived by looking at the error bars: the same value of bandwidth can be perfect for a sample and not-so-good for a different sample, even if D and N are the same. Changing (usually enlarging) the bandwidth at small N could improve significantly the final result, as this is most sensitive to the bandwidth. Of course, some different way of optimizing should be introduced. As a practical tip, the optimization procedure that would lead to the largest h is the safest choice for optimizing this estimator. By safest we mean that it is the one that avoids the irregular behavior the most, but it is not necessarily the one that minimizes the error.

Boundary term

A well known problem of KDE is the boundary bias and our estimators can suffer from the same problem. Fortunately, many of the solutions that have been studied for KDE can be used for our estimators as well. The boundary bias problem arises when we have a point such that $f(x)$ is finite in that point, but it is identical to 0 in a neighborhood of that point. This is the case for instance for an exponential distribution or even a uniform distribution. There are two cases of interest for us. In the first case, a single boundary point is present, then we can always reduce to the case where $f(x)$ has support on $[0, +\infty)$. In the second case we have two boundary points and the function has support on a compact interval, which can be sent to the unit interval $[0, 1]$.

If there is a single boundary, then a simple solution is doubling the points by considering $x \to -x$ [221]. A generalized reflection, which further improves the reduction of the bias when $f'(0) \neq 0$, is proposed in Ref. [222].

A general solution which applies to both cases is transforming the variable x in such a way that whichever support $f(x)$ has, the new $y = g(x)$ will take values on $\mathbb{R}$ [223, 224]. For a discussion on the consistency of the estimator with this change of variable see for instance Ref. [224].

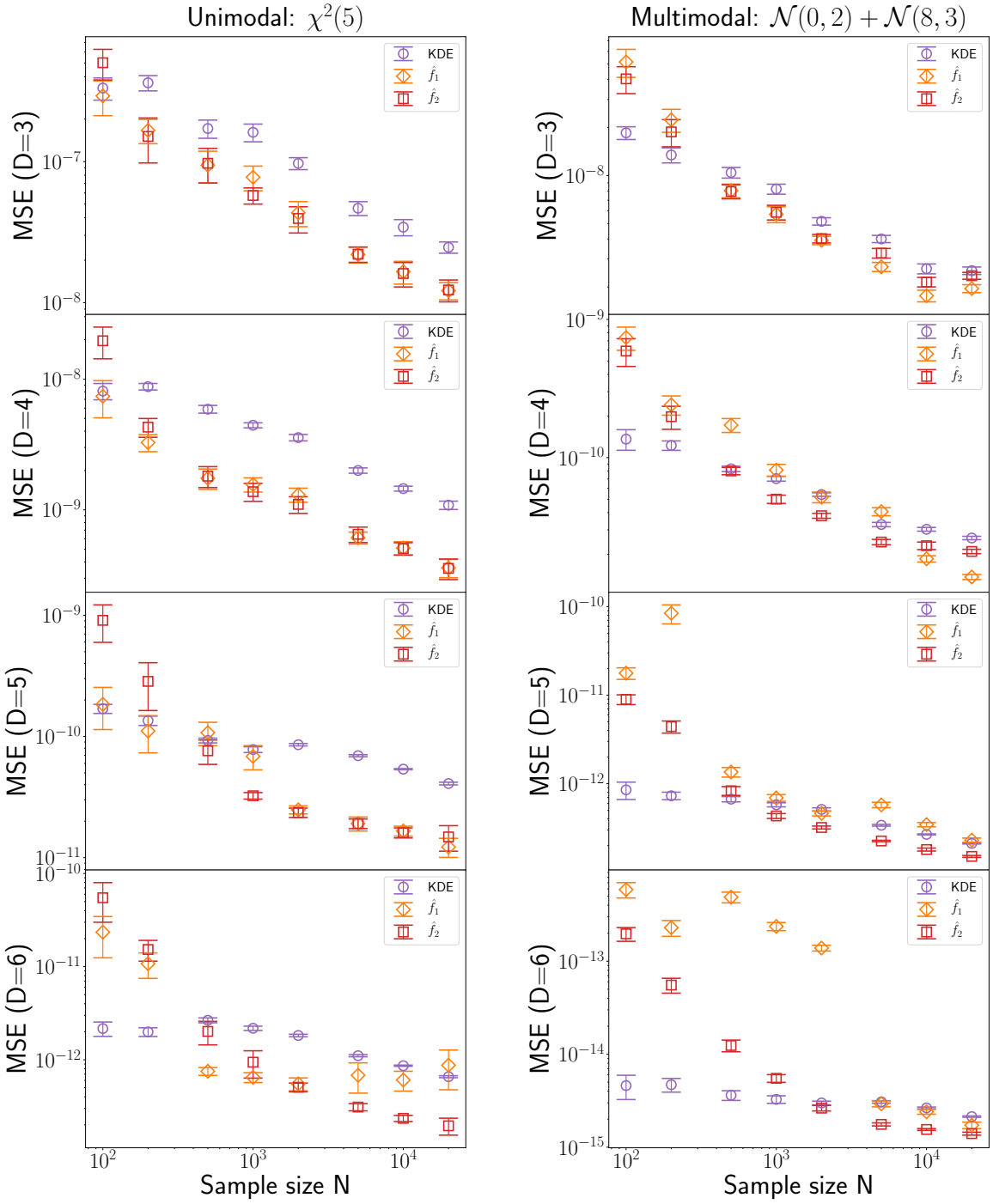

Figure 5.8: MSE for KDE and our two estimators for different values of D and with increasing sample size. For all estimators the bandwidth used is the same, chosen by performing 5-fold cross-validation on KDE. The error bars indicate the 2σ statistical uncertainty evaluated with $R = 9$ repetitions. *On the left:* Unimodal case: $\chi^2(5)$. *On the right:* Multimodal case: $\mathcal{N}(0,2)+\mathcal{N}(8,3)$

Chapter 6

Conclusions and outlook

Dark matter is out there and addressing this problem is a task that falls on the shoulders of both theorists and experimentalists. This work lies somehow in between the two approaches and shows the possibility of moving in either direction.

We built the SM Lagrangian starting from symmetry arguments and similarly we derived our composite Higgs model based on symmetry only. This model constitutes an extension of the SM providing a DM candidate satisfying all experimental requirements. We described this model as of "theoretical" interest, this is because of the well-motivated construction and the opportunity of addressing more than one SM problem at once. This gives the model a higher degree of realism, but unfortunately it does not tell us much about the direction we should follow for DM experiments, especially at LHC.

We then moved in the opposite direction: we know what experiments are looking (and are going to look) for DM. The model we devised aimed at describing the role of unusual signals at LHC, in our case a signal analogous to R-hadron, which is part of the broad category of LLP signals. In this case, the model was experimentally justified: this is an effective field theory, so it does not want to construct a more complete description, but exemplifies some properties of it. The advantage of this approach is that it is telling us where to look with experiments at LHC, which was our focus.

So apparently there would be two directions: either motivate your model with an elegant theoretical construction or build a model that is useful in order to guide experiments. However, we proposed a third way: a pragmatic, data-driven approach that uses the data that are already present. This approach is in a certain sense also agnostic with respect to the specific BSM model, since the aim is finding new physics regardless of its nature. Of course, understanding where to look and how to find new physics would be impossible without a physical input: we still need to use our models to simulate the BSM signal and pursue our study.

In conclusion, hopefully we have showed that an astroparticle phenomenologist can take advantage of both experimental and theoretical knowledge in order to tackle the problem of dark matter. This already leaves us with the freedom to decide on which of the two aspects we want to focus our attention. But the world is evolving and machine learning is starting to play a relevant role in high energy physics. This means that now we have the opportunity of combining our theoretical models, our knowledge of the experiments and the computer science progress for the purpose of finding what dark matter really is.

Bibliography

[1] A. Davoli, A. De Simone, D. Marzocca and A. Morandini, *Composite 2HDM with singlets: a viable dark matter scenario*, *JHEP* **10** (2019) 196, [1905.13244].

[2] A. Davoli, A. De Simone, T. Jacques and A. Morandini, *LHC Phenomenology of Dark Matter with a Color-Octet Partner*, *JHEP* **07** (2018) 054, [1803.02861].

[3] J. Alimena, J. Beacham, M. Borsato, Y. Cheng, X. C. Vidal, G. Cottin et al., *Searching for long-lived particles beyond the standard model at the large hadron collider*, *Journal of Physics G: Nuclear and Particle Physics* **47** (sep, 2020) 090501.

[4] A. D. Simone and A. Morandini, *Nonparametric density estimation from markov chains*, 2020.

[5] G. Brooijmans et al., *Les Houches 2019 Physics at TeV Colliders: New Physics Working Group Report*, in *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*, 2, 2020. 2002.12220.

[6] H. Georgi, *Lie Algebras In Particle Physics: from Isospin To Unified Theories*. Frontiers in Physics. Avalon Publishing, 1999.

[7] B. Hall and B. Hall, *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Graduate Texts in Mathematics. Springer, 2003.

[8] É. Cartan, *La théorie des groupes finis et continus et l'analysis situs*. 1952.

[9] E. Cartan, *Sur la structure des groupes de transformations finis et continus*. Thèses présentées a la Faculté des Sciences de Paris pour obtenir le grade de docteur ès sciences mathématiques. Nony, 1894.

[10] S. Weinberg, *The Quantum Theory of Fields: Volume 2, Modern Applications*. Cambridge University Press, 1996.

[11] M. Schwartz, *Quantum Field Theory and the Standard Model*. Quantum Field Theory and the Standard Model. Cambridge University Press, 2014.

[12] J. Goldstone, *Field Theories with Superconductor Solutions*, *Nuovo Cim.* **19** (1961) 154–164.

[13] J. Goldstone, A. Salam and S. Weinberg, *Broken symmetries*, *Phys. Rev.* **127** (Aug, 1962) 965–970.

[14] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508–509.

[15] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) 321–323.

[16] S. P. Martin, *A Supersymmetry primer*, vol. 21, pp. 1–153. 2010. hep-ph/9709356. 10.1142/9789812839657_0001.

[17] K. Agashe, R. Contino and A. Pomarol, *The Minimal composite Higgs model*, *Nucl. Phys. B* **719** (2005) 165–187, [hep-ph/0412089].

[18] N. Arkani-Hamed, S. Dimopoulos and G. Dvali, *The Hierarchy problem and new dimensions at a millimeter*, *Phys. Lett. B* **429** (1998) 263–272, [hep-ph/9803315].

[19] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, *Phys. Rev. Lett.* **83** (1999) 3370–3373, [hep-ph/9905221].

[20] G. Jungman, M. Kamionkowski and K. Griest, *Supersymmetric dark matter*, *Phys. Rept.* **267** (1996) 195–373, [hep-ph/9506380].

[21] M. Frigerio, A. Pomarol, F. Riva and A. Urbano, *Composite Scalar Dark Matter*, *JHEP* **07** (2012) 015, [1204.2808].

[22] M. Singer, J. W. F. Valle and J. Schechter, *Canonical neutral-current predictions from the weak-electromagnetic gauge group su(3) x u(1)*, *Phys. Rev. D* **22** (Aug, 1980) 738–743.

[23] F. Pisano and V. Pleitez, *An SU(3) x U(1) model for electroweak interactions*, *Phys. Rev. D* **46** (1992) 410–417, [hep-ph/9206242].

[24] J. Mizukoshi, C. de S.Pires, F. Queiroz and P. Rodrigues da Silva, *WIMPs in a 3-3-1 model with heavy Sterile neutrinos*, *Phys. Rev. D* **83** (2011) 065024, [1010.4097].

[25] Super-Kamiokande Collaboration collaboration, Y. Fukuda, T. Hayakawa, E. Ichihara, K. Inoue, K. Ishihara, H. Ishino et al., *Evidence for oscillation of atmospheric neutrinos*, *Phys. Rev. Lett.* **81** (Aug, 1998) 1562–1567.

[26] SNO Collaboration collaboration, Q. R. Ahmad, R. C. Allen, T. C. Andersen, J. D.Anglin, J. C. Barton, E. W. Beier et al., *Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory*, *Phys. Rev. Lett.* **89** (Jun, 2002) 011301.

[27] M. Gonzalez-Garcia, M. Maltoni, J. Salvado and T. Schwetz, *Global fit to three neutrino mixing: critical look at present precision*, *JHEP* **12** (2012) 123, [1209.3023].

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

[29] S. Weinberg, *Baryon and Lepton Nonconserving Processes*, *Phys. Rev. Lett.* **43** (1979) 1566–1570.

[30] M. Gell-Mann, P. Ramond and R. Slansky, *Complex Spinors and Unified Theories*, *Conf. Proc. C* **790927** (1979) 315–321, [1306.4669].

[31] C. Boehm, Y. Farzan, T. Hambye, S. Palomares-Ruiz and S. Pascoli, *Is it possible to explain neutrino masses with scalar dark matter?*, *Phys. Rev. D* **77** (2008) 043516, [hep-ph/0612228].

[32] nEDM collaboration, C. Abel et al., *Measurement of the permanent electric dipole moment of the neutron*, *Phys. Rev. Lett.* **124** (2020) 081803, [2001.11966].

[33] R. Peccei and H. R. Quinn, *Constraints Imposed by CP Conservation in the Presence of Instantons*, *Phys. Rev. D* **16** (1977) 1791–1797.

[34] T. Donnelly, S. Freedman, R. Lytel, R. Peccei and M. Schwartz, *Do Axions Exist?*, *Phys. Rev. D* **18** (1978) 1607.

[35] D. E. Kaplan, M. A. Luty and K. M. Zurek, *Asymmetric Dark Matter*, *Phys. Rev. D* **79** (2009) 115016, [0901.4117].

[36] M. Milgrom, *A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis.*, *Astrophysical Journal* **270** (July, 1983) 365–370.

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

[38] J. H. Oort, *The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems*, *Bulletin of the Astronomical Institutes of the Netherlands* **6** (Aug., 1932) 249.

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

[40] K. G. Begeman, A. H. Broeils and R. H. Sanders, *Extended rotation curves of spiral galaxies: dark haloes and modified dynamics*, *Monthly Notices of the Royal Astronomical Society* **249** (04, 1991) 523–537, [https://academic.oup.com/mnras/article-pdf/249/3/523/18160929/mnras249-0523.pdf].

[41] N. Kaiser and G. Squires, *Mapping the dark matter with weak gravitational lensing*, *Astrophys. J.* **404** (1993) 441–450.

[42] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones et al., *A direct empirical proof of the existence of dark matter*, *Astrophys. J. Lett.* **648** (2006) L109–L113, [astro-ph/0608407].

[43] S. W. Randall, M. Markevitch, D. Clowe, A. H. Gonzalez and M. Bradač, *Constraints on the self-interaction cross section of dark matter from numerical simulations of the merging galaxy cluster 1e 0657-56*, *The Astrophysical Journal* **679** (jun, 2008) 1173–1180.

[44] G. R. Blumenthal, S. Faber, J. R. Primack and M. J. Rees, *Formation of Galaxies and Large Scale Structure with Cold Dark Matter*, *Nature* **311** (1984) 517–525.

[45] D. H. Weinberg, J. S. Bullock, F. Governato, R. Kuzio de Naray and A. H. G. Peter, *Cold dark matter: controversies on small scales*, *Proc. Nat. Acad. Sci.* **112** (2015) 12249–12255, [1306.0913].

[46] E. Kolb, *The Early Universe*. CRC Press, 2018.

[47] J. McDonald, *WIMP Densities in Decaying Particle Dominated Cosmology*, *Phys. Rev. D* **43** (1991) 1063–1068.

[48] L. J. Hall, K. Jedamzik, J. March-Russell and S. M. West, *Freeze-In Production of FIMP Dark Matter*, *JHEP* **03** (2010) 080, [0911.1120].

[49] M. Cirelli, Y. Gouttenoire, K. Petraki and F. Sala, *Homeopathic Dark Matter, or how diluted heavy substances produce high energy cosmic rays*, *JCAP* **02** (2019) 014, [1811.03608].

[50] K. Griest and D. Seckel, *Three exceptions in the calculation of relic abundances*, *Phys. Rev. D* **43** (1991) 3191–3203.

[51] J. E. Kim, *Light Pseudoscalars, Particle Physics and Cosmology*, *Phys. Rept.* **150** (1987) 1–177.

[52] G. G. Raffelt, *Astrophysical axion bounds*, *Lect. Notes Phys.* **741** (2008) 51–71, [hep-ph/0611350].

[53] G. Ballesteros, J. Redondo, A. Ringwald and C. Tamarit, *Standard Model—axion—seesaw—Higgs portal inflation. Five problems of particle physics and cosmology solved in one stroke*, *JCAP* **08** (2017) 001, [1610.01639].

[54] B. Carr, K. Kohri, Y. Sendouda and J. Yokoyama, *Constraints on Primordial Black Holes*, 2002.12778.

[55] ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group collaboration, S. Schael et al., *Precision electroweak measurements on the Z resonance*, *Phys. Rept.* **427** (2006) 257–454, [hep-ex/0509008].

[56] LEP Working Group for Higgs boson searches, ALEPH, DELPHI, L3, OPAL collaboration, R. Barate et al., *Search for the standard model Higgs boson at LEP*, *Phys. Lett. B* **565** (2003) 61–75, [hep-ex/0306033].

[57] *The International Linear Collider Technical Design Report - Volume 2: Physics*, 1306.6352.

[58] *Physics and Detectors at CLIC: CLIC Conceptual Design Report*, 1202.5940.

[59] ATLAS collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1–29, [1207.7214].

[60] CMS collaboration, S. Chatrchyan et al., *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30–61, [1207.7235].

[61] FCC collaboration, A. Abada et al., *FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2*, *Eur. Phys. J. ST* **228** (2019) 261–623.

[62] D. Curtin et al., *Long-Lived Particles at the Energy Frontier: The MATHUSLA Physics Case*, *Rept. Prog. Phys.* **82** (2019) 116201, [1806.07396].

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

[64] LUX collaboration, D. Akerib et al., *First results from the LUX dark matter experiment at the Sanford Underground Research Facility*, *Phys. Rev. Lett.* **112** (2014) 091303, [1310.8214].

[65] XENON collaboration, E. Aprile et al., *Observation of Excess Electronic Recoil Events in XENON1T*, 2006.09721.

[66] M. Cirelli, E. Del Nobile and P. Panci, *Tools for model-independent bounds in direct dark matter searches*, *JCAP* **10** (2013) 019, [1307.5955].

[67] A. Fitzpatrick, W. Haxton, E. Katz, N. Lubbers and Y. Xu, *The Effective Field Theory of Dark Matter Direct Detection*, *JCAP* **02** (2013) 004, [1203.3542].

[68] PandaX collaboration, A. Tan et al., *Dark Matter Search Results from the Commissioning Run of PandaX-II*, *Phys. Rev. D* **93** (2016) 122009, [1602.06563].

[69] E. Bagnaschi et al., *Likelihood Analysis of Supersymmetric SU(5) GUTs*, *Eur. Phys. J. C* **77** (2017) 104, [1610.10084].

[70] Fermi-LAT collaboration, M. Ackermann et al., *Searching for Dark Matter Annihilation from Milky Way Dwarf Spheroidal Galaxies with Six Years of Fermi Large Area Telescope Data*, *Phys. Rev. Lett.* **115** (2015) 231301, [1503.02641].

[71] N. Evans, J. Sanders and A. Geringer-Sameth, *Simple J-Factors and D-Factors for Indirect Dark Matter Detection*, *Phys. Rev. D* **93** (2016) 103512, [1604.05599].

[72] IceCube collaboration, M. Aartsen et al., *Search for dark matter annihilations in the Sun with the 79-string IceCube detector*, *Phys. Rev. Lett.* **110** (2013) 131302, [1212.4097].

[73] AMS collaboration, M. Aguilar et al., *First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–350 GeV*, *Phys. Rev. Lett.* **110** (2013) 141102.

[74] AMS collaboration, M. Aguilar et al., *Antiproton Flux, Antiproton-to-Proton Flux Ratio, and Properties of Elementary Particle Fluxes in Primary Cosmic Rays Measured with the Alpha Magnetic Spectrometer on the International Space Station*, *Phys. Rev. Lett.* **117** (2016) 091103.

[75] D. B. Kaplan and H. Georgi, *SU(2) x U(1) Breaking by Vacuum Misalignment*, *Phys. Lett.* **136B** (1984) 183–186.

[76] H. Georgi and D. B. Kaplan, *Composite Higgs and Custodial SU(2)*, *Phys. Lett.* **145B** (1984) 216–220.

[77] D. B. Kaplan, H. Georgi and S. Dimopoulos, *Composite Higgs Scalars*, *Phys. Lett.* **136B** (1984) 187–190.

[78] M. J. Dugan, H. Georgi and D. B. Kaplan, *Anatomy of a Composite Higgs Model*, *Nucl. Phys.* **B254** (1985) 299–326.

[79] R. Contino, *The Higgs as a Composite Nambu-Goldstone Boson*, in *Physics of the large and the small, TASI 09, proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, Colorado, USA, 1-26 June 2009*, pp. 235–306, 2011. 1005.4269. DOI.

[80] B. Bellazzini, C. Csáki and J. Serra, *Composite Higgses*, *Eur. Phys. J.* **C74** (2014) 2766, [1401.2457].

[81] G. Panico and A. Wulzer, *The Composite Nambu-Goldstone Higgs*, *Lect. Notes Phys.* **913** (2016) pp.1–316, [1506.01961].

[82] D. Marzocca and A. Urbano, *Composite Dark Matter and LHC Interplay*, *JHEP* **07** (2014) 107, [1404.7419].

[83] N. Fonseca, R. Zukanovich Funchal, A. Lessa and L. Lopez-Honorez, *Dark Matter Constraints on Composite Higgs Models*, *JHEP* **06** (2015) 154, [1501.05957].

[84] S. Bruggisser, F. Riva and A. Urbano, *Strongly Interacting Light Dark Matter*, *SciPost Phys.* **3** (2017) 017, [1607.02474].

[85] B. Gripaios, A. Pomarol, F. Riva and J. Serra, *Beyond the Minimal Composite Higgs Model*, *JHEP* **04** (2009) 070, [0902.1483].

[86] G. Ballesteros, A. Carmona and M. Chala, *Exceptional Composite Dark Matter*, *Eur. Phys. J.* **C77** (2017) 468, [1704.07388].

[87] R. Balkin, M. Ruhdorfer, E. Salvioni and A. Weiler, *Charged Composite Scalar Dark Matter*, *JHEP* **11** (2017) 094, [1707.07685].

[88] R. Balkin, M. Ruhdorfer, E. Salvioni and A. Weiler, *Dark matter shifts away from direct detection*, *JCAP* **1811** (2018) 050, [1809.09106].

[89] L. Da Rold and A. N. Rossia, *The Minimal Simple Composite Higgs Model*, 1904.02560.

[90] M. Chala, R. Gröber and M. Spannowsky, *Searches for vector-like quarks at future colliders and implications for composite Higgs models with dark matter*, *JHEP* **03** (2018) 040, [1801.06537].

[91] T. Ma and G. Cacciapaglia, *Fundamental Composite 2HDM: SU(N) with 4 flavours*, *JHEP* **03** (2016) 211, [1508.07014].

[92] Y. Wu, T. Ma, B. Zhang and G. Cacciapaglia, *Composite Dark Matter and Higgs*, *JHEP* **11** (2017) 058, [1703.06903].

[93] R. Balkin, G. Perez and A. Weiler, *Little composite dark matter*, *Eur. Phys. J.* **C78** (2018) 104, [1707.09980].

[94] G. Cacciapaglia, H. Cai, A. Deandrea and A. Kushwaha, *Fundamental Composite Higgs model in SU(6)/SO(6)*, 1904.09301.

[95] T. Alanne, D. Buarque Franzosi, M. T. Frandsen and M. Rosenlyst, *Dark matter in (partially) composite Higgs models*, *JHEP* **12** (2018) 088, [1808.07515].

[96] A. Carmona and M. Chala, *Composite Dark Sectors*, *JHEP* **06** (2015) 105, [1504.00332].

[97] J. Mrazek, A. Pomarol, R. Rattazzi, M. Redi, J. Serra and A. Wulzer, *The Other Natural Two Higgs Doublet Model*, *Nucl. Phys.* **B853** (2011) 1–48, [1105.5403].

[98] S. De Curtis, L. Delle Rose, S. Moretti and K. Yagyu, *Supersymmetry versus Compositeness: 2HDMs tell the story*, *Phys. Lett.* **B786** (2018) 189–194, [1803.01865].

[99] S. De Curtis, L. Delle Rose, S. Moretti and K. Yagyu, *A Concrete Composite 2-Higgs Doublet Model*, *JHEP* **12** (2018) 051, [1810.06465].

[100] S. R. Coleman, J. Wess and B. Zumino, *Structure of phenomenological Lagrangians. 1.*, *Phys. Rev.* **177** (1969) 2239–2247.

[101] C. G. Callan, Jr., S. R. Coleman, J. Wess and B. Zumino, *Structure of phenomenological Lagrangians. 2.*, *Phys. Rev.* **177** (1969) 2247–2250.

[102] A. Manohar and H. Georgi, *Chiral Quarks and the Nonrelativistic Quark Model*, *Nucl. Phys.* **B234** (1984) 189–212.

[103] G. Panico and A. Wulzer, *The Discrete Composite Higgs Model*, *JHEP* **09** (2011) 135, [1106.2719].

[104] J. Wess and B. Zumino, *Consequences of anomalous Ward identities*, *Phys. Lett.* **37B** (1971) 95–97.

[105] E. Witten, *Global Aspects of Current Algebra*, *Nucl. Phys.* **B223** (1983) 422–432.

[106] J. Davighi and B. Gripaios, *Homological classification of topological terms in sigma models on homogeneous spaces*, *JHEP* **09** (2018) 155, [1803.07585].

[107] J. Davighi and B. Gripaios, *Topological terms in Composite Higgs Models*, *JHEP* **11** (2018) 169, [1808.04154].

[108] D. B. Kaplan, *Flavor at SSC energies: A New mechanism for dynamically generated fermion masses*, *Nucl. Phys.* **B365** (1991) 259–278.

[109] T. Moroi, M. Nagai and M. Takimoto, *Non-thermal production of wino dark matter via the decay of long-lived particles*, *Journal of High Energy Physics* **2013** (03, 2013) .

[110] CMS collaboration, A. M. Sirunyan et al., *Search for invisible decays of a Higgs boson produced through vector boson fusion in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, 1809.05937.

[111] ATLAS collaboration, M. Aaboud et al., *Combination of searches for invisible Higgs boson decays with the ATLAS experiment*, *Submitted to: Phys. Rev. Lett.* (2019) , [1904.05105].

[112] J. de Blas et al., *Higgs Boson Studies at Future Particle Colliders*, 1905.03764.

[113] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr and B. Fuks, *FeynRules 2.0 - A complete toolbox for tree-level phenomenology*, *Comput. Phys. Commun.* **185** (2014) 2250–2300, [1310.1921].

[114] N. D. Christensen and C. Duhr, *FeynRules - Feynman rules made easy*, *Comput. Phys. Commun.* **180** (2009) 1614–1641, [0806.4194].

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

[116] ATLAS collaboration, M. Aaboud et al., *Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector*, 1711.03301.

[117] ATLAS collaboration, M. Aaboud et al., *Search for invisible Higgs boson decays in vector boson fusion at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *Submitted to: Phys. Lett.* (2018) , [1809.06682].

[118] XENON collaboration, E. Aprile et al., *Physics reach of the XENON1T dark matter experiment*, *JCAP* **1604** (2016) 027, [1512.07501].

[119] I. Brivio, M. Gavela, L. Merlo, K. Mimasu, J. No, R. del Rey et al., *Non-linear Higgs portal to Dark Matter*, *JHEP* **04** (2016) 141, [1511.01099].

[120] M. Fairbairn and J. Zupan, *Dark matter with a late decaying dark partner*, *JCAP* **0907** (2009) 001, [0810.4147].

[121] ATLAS Collaboration collaboration, G. Aad, B. Abbott, D. C. Abbott, O. Abdinov, A. Abed Abud, K. Abeling et al., *Combined measurements of higgs boson production and decay using up to* 80 fb^{-1} *of proton-proton collision data at* $\sqrt{s} = 13$ TeV *collected with the atlas experiment*, *Phys. Rev. D* **101** (Jan, 2020) 012002.

[122] LHC Higgs Cross Section Working Group collaboration, J. R. Andersen et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, 1307.1347.

[123] J. Haller, A. Hoecker, R. Kogler, K. Mönig, T. Peiffer and J. Stelzer, *Update of the global electroweak fit and constraints on two-Higgs-doublet models*, *Eur. Phys. J.* **C78** (2018) 675, [1803.01853].

[124] R. Barbieri, A. Pomarol, R. Rattazzi and A. Strumia, *Electroweak symmetry breaking after LEP-1 and LEP-2*, *Nucl. Phys.* **B703** (2004) 127–146, [hep-ph/0405040].

[125] T. Plehn, *Lectures on LHC Physics*, *Lect. Notes Phys.* **844** (2012) 1–193, [0910.4182].

[126] L. Lopez Honorez and C. E. Yaguna, *The inert doublet model of dark matter revisited*, *JHEP* **09** (2010) 046, [1003.3125].

[127] M. Baumgart, C. Cheung, J. T. Ruderman, L.-T. Wang and I. Yavin, *Non-Abelian Dark Sectors and Their Collider Signatures*, *JHEP* **04** (2009) 014, [0901.0283].

[128] K. R. Dienes, S. Su and B. Thomas, *Distinguishing Dynamical Dark Matter at the LHC*, *Phys. Rev.* **D86** (2012) 054008, [1204.4183].

[129] I.-W. Kim and K. M. Zurek, *Flavor and Collider Signatures of Asymmetric Dark Matter*, *Phys. Rev.* **D89** (2014) 035008, [1310.2617].

[130] R. T. Co, F. D'Eramo, L. J. Hall and D. Pappadopulo, *Freeze-In Dark Matter with Displaced Signatures at Colliders*, *JCAP* **1512** (2015) 024, [1506.07532].

[131] Y. Hochberg, E. Kuflik and H. Murayama, *SIMP Spectroscopy*, *JHEP* **05** (2016) 090, [1512.07917].

[132] A. Davoli, A. De Simone, T. Jacques and V. Sanz, *Displaced Vertices from Pseudo-Dirac Dark Matter*, *JHEP* **11** (2017) 025, [1706.08985].

[133] J. D. Wells, *Implications of supersymmetry breaking with a little hierarchy between gauginos and scalars*, in *11th International Conference on Supersymmetry and the Unification of Fundamental Interactions (SUSY 2003) Tucson, Arizona, June 5-10, 2003*, 2003. hep-ph/0306127.

[134] N. Arkani-Hamed and S. Dimopoulos, *Supersymmetric unification without low energy supersymmetry and signatures for fine-tuning at the LHC*, *JHEP* **06** (2005) 073, [hep-th/0405159].

[135] A. Arvanitaki, N. Craig, S. Dimopoulos and G. Villadoro, *Mini-Split*, *JHEP* **02** (2013) 126, [1210.0555].

[136] N. Arkani-Hamed, A. Gupta, D. E. Kaplan, N. Weiner and T. Zorawski, *Simply Unnatural Supersymmetry*, 1212.6971.

[137] Y. Cui and B. Shuve, *Probing Baryogenesis with Displaced Vertices at the LHC*, *JHEP* **02** (2015) 049, [1409.6729].

[138] C. Csaki, E. Kuflik, S. Lombardo, O. Slone and T. Volansky, *Phenomenology of a Long-Lived LSP with R-Parity Violation*, *JHEP* **08** (2015) 016, [1505.00784].

[139] J. Barnard, P. Cox, T. Gherghetta and A. Spray, *Long-Lived, Colour-Triplet Scalars from Unnaturalness*, *JHEP* **03** (2016) 003, [1510.06405].

[140] A. De Simone, G. F. Giudice and A. Strumia, *Benchmarks for Dark Matter Searches at the LHC*, *JHEP* **06** (2014) 081, [1402.6287].

[141] M. J. Baker et al., *The Coannihilation Codex*, *JHEP* **12** (2015) 120, [1510.03434].

[142] S. El Hedri, A. Kaminska, M. de Vries and J. Zurita, *Simplified Phenomenology for Colored Dark Sectors*, *JHEP* **04** (2017) 118, [1703.00452].

[143] M. Garny, J. Heisig, B. Lülf and S. Vogl, *Coannihilation without chemical equilibrium*, *Phys. Rev.* **D96** (2017) 103521, [1705.09292].

[144] G. Busoni, A. De Simone, E. Morgante and A. Riotto, *On the Validity of the Effective Field Theory for Dark Matter Searches at the LHC*, *Phys. Lett.* **B728** (2014) 412–421, [1307.2253].

[145] E. Morgante, *On the validity of the effective field theory for dark matter searches at the LHC*, *Nuovo Cim.* **C38** (2015) 32, [1409.6668].

[146] G. Busoni, *Limitation of EFT for DM interactions at the LHC*, *PoS* **DIS2014** (2014) 134, [1411.3600].

[147] G. Busoni, A. De Simone, T. Jacques, E. Morgante and A. Riotto, *On the Validity of the Effective Field Theory for Dark Matter Searches at the LHC Part III: Analysis for the t-channel*, *JCAP* **1409** (2014) 022, [1405.3101].

[148] G. Busoni, A. De Simone, J. Gramling, E. Morgante and A. Riotto, *On the Validity of the Effective Field Theory for Dark Matter Searches at the LHC, Part II: Complete Analysis for the s-channel*, *JCAP* **1406** (2014) 060, [1402.1275].

[149] N. F. Bell, Y. Cai, J. B. Dent, R. K. Leane and T. J. Weiler, *Dark matter at the LHC: Effective field theories and gauge invariance*, *Phys. Rev.* **D92** (2015) 053008, [1503.07874].

[150] A. De Simone and T. Jacques, *Simplified models vs. effective field theory approaches in dark matter searches*, *Eur. Phys. J.* **C76** (2016) 367, [1603.08002].

[151] S. Bruggisser, F. Riva and A. Urbano, *The Last Gasp of Dark Matter Effective Theory*, *JHEP* **11** (2016) 069, [1607.02475].

[152] F. Kahlhoefer, *Review of LHC Dark Matter Searches*, *Int. J. Mod. Phys.* **A32** (2017) 1730006, [1702.02430].

[153] G. R. Farrar and P. Fayet, *Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry*, *Phys. Lett.* **76B** (1978) 575–579.

[154] ATLAS collaboration, M. Aaboud et al., *Search for heavy long-lived charged R-hadrons with the ATLAS detector in 3.2 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13$ TeV*, *Phys. Lett.* **B760** (2016) 647–665, [1606.05129].

[155] ATLAS collaboration, M. Aaboud et al., *Search for metastable heavy charged particles with large ionization energy loss in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS experiment*, *Phys. Rev.* **D93** (2016) 112015, [1604.04520].

[156] CMS collaboration, C. Collaboration, *Search for stopped long-lived particles produced in pp collisions at $\sqrt{s} = 13$ TeV*, .

[157] A. D. Bond, G. Hiller, K. Kowalska and D. F. Litim, *Directions for model building from asymptotic safety*, *JHEP* **08** (2017) 004, [1702.01727].

[158] M. Garny, J. Heisig, M. Hufnagel and B. Lülf, *Top-philic dark matter within and beyond the WIMP paradigm*, 1802.00814.

[159] A. Pierce and J. Thaler, *Natural Dark Matter from an Unnatural Higgs Boson and New Colored Particles at the TeV Scale*, *JHEP* **08** (2007) 026, [hep-ph/0703056].

[160] K. Hamaguchi, S. P. Liew, T. Moroi and Y. Yamamoto, *Isospin-Violating Dark Matter with Colored Mediators*, *JHEP* **05** (2014) 086, [1403.0324].

[161] A. Ibarra, A. Pierce, N. R. Shah and S. Vogl, *Anatomy of Coannihilation with a Scalar Top Partner*, *Phys. Rev.* **D91** (2015) 095018, [1501.03164].

[162] J. Ellis, F. Luo and K. A. Olive, *Gluino Coannihilation Revisited*, *JHEP* **09** (2015) 127, [1503.07142].

[163] S. P. Liew and F. Luo, *Effects of QCD bound states on dark matter relic abundance*, *JHEP* **02** (2017) 091, [1611.08133].

[164] A. Mitridate, M. Redi, J. Smirnov and A. Strumia, *Cosmological Implications of Dark Matter Bound States*, *JCAP* **1705** (2017) 006, [1702.01141].

[165] V. De Luca, A. Mitridate, M. Redi, J. Smirnov and A. Strumia, *Colored Dark Matter*, 1801.01135.

[166] S. Biondini and M. Laine, *Thermal dark matter co-annihilating with a strongly interacting scalar*, 1801.05821.

[167] M. Garny, A. Ibarra, S. Rydbeck and S. Vogl, *Majorana Dark Matter with a Coloured Mediator: Collider vs Direct and Indirect Searches*, *JHEP* **06** (2014) 169, [1403.4634].

[168] F. Giacchino, A. Ibarra, L. Lopez Honorez, M. H. G. Tytgat and S. Wild, *Signatures from Scalar Dark Matter with a Vector-like Quark Mediator*, *JCAP* **1602** (2016) 002, [1511.04452].

[169] A. Sommerfeld, *Über die beugung und bremsung der elektronen*, *Annalen der Physik* **403** (1931) 257–330.

[170] J. L. Feng, M. Kaplinghat and H.-B. Yu, *Sommerfeld Enhancements for Thermal Relic Dark Matter*, *Phys. Rev.* **D82** (2010) 083525, [1005.4678].

[171] S. El Hedri, A. Kaminska and M. de Vries, *A Sommerfeld Toolbox for Colored Dark Sectors*, *Eur. Phys. J.* **C77** (2017) 622, [1612.02825].

[172] R. T. D'Agnolo, C. Mondino, J. T. Ruderman and P.-J. Wang, *Exponentially Light Dark Matter from Coannihilation*, 1803.02901.

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

[174] ATLAS collaboration, M. Aaboud et al., *Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, *Phys. Rev. D* **94** (2016) 032005, [1604.07773].

[175] CMS collaboration, D. Vannerom, *Search for dark matter with jets and missing transverse energy at 13 TeV*, *PoS* **LHCP2016** (2016) 218.

[176] CMS collaboration, A. M. Sirunyan et al., *Search for new phenomena in final states with two opposite-charge, same-flavor leptons, jets, and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV*, 1709.08908.

[177] A. Cuoco, J. Heisig, M. Korsmeier and M. Krämer, *Constraining heavy dark matter with cosmic-ray antiprotons*, 1711.05274.

[178] M. R. Kauth, J. H. Kuhn, P. Marquard and M. Steinhauser, *Gluinonia: Energy Levels, Production and Decay*, *Nucl. Phys.* **B831** (2010) 285–305, [0910.2612].

[179] ATLAS collaboration, M. Aaboud et al., *Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector*, *Phys. Rev.* **D96** (2017) 052004, [1703.09127].

[180] G. Kasieczka, B. Nachman and D. Shih, *R&D Dataset for LHC Olympics 2020 Anomaly Detection Challenge*, Apr., 2019. 10.5281/zenodo.3832254.

[181] R. M. Harris and S. Jain, *Cross Sections for Leptophobic Topcolor Z' Decaying to Top-Antitop*, *Eur. Phys. J. C* **72** (2012) 2072, [1112.4928].

[182] V. Hodge and J. Austin, *A survey of outlier detection methodologies*, *Artificial intelligence review* **22** (2004) 85–126.

[183] R. Chalapathy and S. Chawla, *Deep learning for anomaly detection: A survey*, *CoRR* **abs/1901.03407** (2019) , [1901.03407].

[184] E. Real, A. Aggarwal, Y. Huang and Q. V. Le, *Regularized evolution for image classifier architecture search*, in *Proceedings of the aaai conference on artificial intelligence*, vol. 33, pp. 4780–4789, 2019.

[185] S. Kotsiantis, I. Zaharakis and P. Pintelas, *Machine learning: A review of classification and combining techniques*, *Artificial Intelligence Review* **26** (11, 2006) 159–190.

[186] A. K. Jain, M. N. Murty and P. J. Flynn, *Data clustering: a review*, *ACM computing surveys (CSUR)* **31** (1999) 264–323.

[187] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui and J. Long, *A survey of clustering with deep learning: From the perspective of network architecture*, *IEEE Access* **6** (2018) 39501–39514.

[188] D. Castelvecchi, *Can we open the black box of ai?*, *Nature News* **538** (2016) 20.

[189] W. R. Gilks, S. Richardson and D. Spiegelhalter, *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC, 1995.

[190] A. N. Langville and C. D. Meyer, *Updating markov chains with an eye on google's pagerank*, *SIAM journal on matrix analysis and applications* **27** (2006) 968–987.

[191] D. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley Series in Probability and Statistics. Wiley, 2015.

[192] S. M. Ross, J. J. Kelly, R. J. Sullivan, W. J. Perry, D. Mercer, R. M. Davis et al., *Stochastic processes*, vol. 2. Wiley New York, 1996.

[193] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, *J. Machine Learning Res.* **12** (2011) 2825–2830, [1201.0490].

[194] W. J. Stewart, *Numerical Methods for Computing Stationary Distributions of Finite Irreducible Markov Chains*, pp. 81–111. Springer US, Boston, MA, 2000.

[195] B. Silverman, *Density Estimation for Statistics and Data Analysis Estimation Density*. Monographs on statistics and applied probability. Kluwer Academic Publishers, 1986.

[196] S. Hido, Y. Tsuboi, H. Kashima, M. Sugiyama and T. Kanamori, *Statistical outlier detection using direct density ratio estimation*, *Knowledge and information systems* **26** (2011) 309–336.

[197] L. J. Latecki, A. Lazarevic and D. Pokrajac, *Outlier detection with kernel density functions*, in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 61–75, Springer, 2007.

[198] B. Nachman and D. Shih, *Anomaly Detection with Density Estimation*, *Phys. Rev. D* **101** (2020) 075042, [2001.04990].

[199] H.-P. Kriegel, P. Kröger, J. Sander and A. Zimek, *Density-based clustering*, *WIREs Data Mining and Knowledge Discovery* **1** (2011) 231–240, [https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.30].

[200] N. Saito and R. R. Coifman, *Local discriminant bases and their applications*, *Journal of Mathematical Imaging and Vision* **5** (1995) 337–358.

[201] J. H. Collins, K. Howe and B. Nachman, *Anomaly Detection for Resonant New Physics with Machine Learning*, *Phys. Rev. Lett.* **121** (2018) 241803, [1805.02664].

[202] B. Tang and H. He, *A local density-based approach for local outlier detection*, 2016.

[203] M. Goldstein, *breast-cancer-unsupervised-ad.tab*, in *Unsupervised Anomaly Detection Benchmark*. Harvard Dataverse, 2015. DOI.

[204] M. Goldstein, *pen-local-unsupervised-ad.tab*, in *Unsupervised Anomaly Detection Benchmark*. Harvard Dataverse, 2015. DOI.

[205] M. Goldstein, *pen-global-unsupervised-ad.tab*, in *Unsupervised Anomaly Detection Benchmark*. Harvard Dataverse, 2015. DOI.

[206] M. Goldstein, *satellite-unsupervised-ad.tab*, in *Unsupervised Anomaly Detection Benchmark*. Harvard Dataverse, 2015. DOI.

[207] M. M. Breunig, H.-P. Kriegel, R. T. Ng and J. Sander, *Lof: identifying density-based local outliers*, in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pp. 93–104, 2000.

[208] D. Dua and C. Graff, *UCI machine learning repository*, 2017.

[209] V. G. Sigillito, S. P. Wing, L. V. Hutton and K. B. Baker, *Classification of radar returns from the ionosphere using neural networks*, *Johns Hopkins APL Technical Digest* **10** (1989) 262–266.

[210] M. Little, P. Mcsharry, S. Roberts, D. Costello and I. Moroz, *Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection*, *Biomedical engineering online* **6** (02, 2007) 23.

[211] I. Yeh, K.-J. Yang and T.-M. Ting, *Knowledge discovery on rfm model using bernoulli sequence*, *Expert Syst. Appl.* **36** (04, 2009) 5866–5871.

[212] J. Siebert, *Vehicle recognition using rule based methods*, project report, Turing Institute, Glasgow, 1987.

[213] K. Nakai and M. Kanehisa, *Expert system for predicting protein localization sites in gram-negative bacteria*, *Proteins* **11** (1991) 95—110.

[214] J. Sayyad Shirabad and T. Menzies, "The PROMISE Repository of Software Engineering Databases.." School of Information Technology and Engineering, University of Ottawa, Canada, 2005.

[215] V. Karia, W. Zhang, A. Naeim and R. Ramezani, *Gensample: A genetic algorithm for oversampling in imbalanced datasets*, *arXiv preprint arXiv:1910.10806* (2019) .

[216] D. H. Wolpert and W. G. Macready, *No free lunch theorems for optimization*, *IEEE transactions on evolutionary computation* **1** (1997) 67–82.

[217] M. C. Romao, N. Castro and R. Pedro, *Finding New Physics without learning about it: Anomaly Detection as a tool for Searches at Colliders*, 2006.05432.

[218] E. J. Gumbel, *The maxima of the mean largest value and of the range*, *Ann. Math. Statist.* **25** (03, 1954) 76–84.

[219] R. van Handel, *Probability in high dimension*, tech. rep., PRINCETON UNIV NJ, 2014.

[220] G. Casella and R. Berger, *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning, 2002.

[221] D. B. H. Cline and J. D. Hart, *Kernel estimation of densities with discontinuities or discontinuous derivatives*, *Statistics: A Journal of Theoretical and Applied Statistics* **22** (1991) 69–84.

[222] R. J. Karunamuni and T. Alberts, *A generalized reflection method of boundary correction in kernel density estimation*, *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **33** (2005) 497–509.

[223] J. S. Marron and D. Ruppert, *Transformations to reduce boundary bias in kernel density estimation*, *Journal of the Royal Statistical Society. Series B (Methodological)* **56** (1994) 653–671.

[224] G. Geenens, *Probit transformation for kernel density estimation on the unit interval*, *Journal of the American Statistical Association* **109** (2014) 346–358.