WIMP Dark Matter and Unitarity-Conserving Inflation via a Gauge Singlet Scalar

Felix Kahlhoefer\textsuperscript{a} and John McDonald\textsuperscript{b}
\textsuperscript{a}DESY, Notkestrasse 85, D-22607 Hamburg, Germany
\textsuperscript{b}Lancaster University, Lancaster LA1 4YB, United Kingdom

E-mail: felix.kahlhoefer@desy.de, j.mcdonald@lancaster.ac.uk

Abstract. A gauge singlet scalar with non-minimal coupling to gravity can drive inflation and later freeze out to become cold dark matter. We explore this idea by revisiting inflation in the singlet direction (S-inflation) and Higgs Portal Dark Matter in light of the Higgs discovery, limits from LUX and observations by Planck. We show that large regions of parameter space remain viable, so that successful inflation is possible and the dark matter relic abundance can be reproduced. Moreover, the scalar singlet can stabilise the electroweak vacuum and overcome the problem of unitarity-violation during inflation encountered by Higgs Inflation, provided the singlet is a real scalar. The 2-$\sigma$ Planck upper bound on $n_s$ imposes that the singlet mass is below 2 TeV, so that almost the entire allowed parameter range can be probed by XENON1T.

Keywords: Dark matter theory, Cosmology of theories beyond the SM, Inflation
1 Introduction

There is strong evidence in support of the idea that the Universe underwent a period of primordial inflation. In particular, the observation of adiabatic density perturbations with a spectral index which deviates from unity by a few percent [1] is consistent with the generic prediction of scalar field inflation models. However, the identity of the scalar field responsible for inflation remains unknown. Another unsolved problem of similar importance for cosmology is the nature of dark matter (DM). While it is possible to explain DM by the addition of a new particle, there is presently no experimental evidence for its existence or its identity.

The most studied proposal is that DM is a thermal relic weakly-interacting massive particle (WIMP). WIMPs typically have annihilation cross-sections comparable to the value required to reproduce the observed density of DM, the so-called “WIMP miracle”. Nevertheless, the non-observation of any new weak-scale particles at the LHC beyond the Standard Model (SM) places strong constraints on many models for WIMPs, such as in supersymmetric extensions of the SM. The absence of new particles may indeed indicate that any extension of the SM to include WIMP DM should be rather minimal. In the present work we therefore focus on a particularly simple extension of the SM, namely an additional gauge singlet scalar, which is arguably one of the most minimal models of DM [2–4].

A similar issue arises from recent constraints on inflation. In fact, the non-observation of non-Gaussianity by Planck [5] suggests that the inflation model should also be minimal, in the sense of being due to a single scalar field. The absence of evidence for new physics then raises the question of whether the inflaton scalar can be part of the SM or a minimal extension of the SM. The former possibility is realized by Higgs Inflation [6], which is a version of the non-minimally coupled scalar field inflation model of Salopek, Bond and Bardeen (SBB) [7] with the scalar field identified with the Higgs boson. A good example for the latter option are gauge singlet scalar extensions of the SM, because the DM particle can also provide a well-motivated candidate for the scalar of the SBB model. In other words, in these models the same scalar particle drives inflation and later freezes out to become cold DM.

The resulting gauge singlet inflation model was first considered in [8], where it was called S-inflation (see also [9]). All non-minimally coupled scalar field inflation models based on the SBB model are identical at the classical level but differ once quantum corrections to the inflaton potential are included. These result in characteristic deviations of the spectral index from its classical value, which have been extensively studied in both Higgs Inflation [6, 11–15] and S-inflation [16].

Since the original studies were performed, the mass of the Higgs boson [17] and the Planck results for the inflation observables [1] have become known. In addition, direct DM detection experiments, such as LUX [18], have imposed stronger bounds on gauge singlet scalar DM [19–25]. This new data has important implications for these models, in particular for S-inflation, which can be tested in Higgs physics and DM searches. The main objective of the present paper is to compare the S-inflation model with the latest results from CMB observations and direct DM detection experiments.

We will demonstrate that — in spite of its simplicity — the model still has a large viable parameter space, where the predictions for inflation are consistent with all current constraints and the observed DM relic abundance can be reproduced. In addition, we observe that this model can solve the potential problem that the electroweak vacuum may be metastable, because the singlet gives a positive contribution to the running of the quartic Higgs coupling. Intriguingly, the relevant parameter range can be almost completely tested by XENON1T.

Another important aspect of our study is perturbative unitarity-violation, which may be a significant problem for Higgs Inflation. Since Higgs boson scattering via graviton exchange violates unitarity at high energies [26, 27], one might be worried that the theory is either incomplete or that perturbation theory breaks down so that unitarity is only conserved non-perturbatively [28–31]. In both cases there can be important modification of the inflaton potential due to new physics or strong-coupling effects. Indeed, in conventional Higgs Inflation, the unitarity-violation scale is of the same magnitude as the Higgs field during inflation [14, 32], placing in doubt the predictions of the model or even its viability.

In contrast, we will show that S-inflation has sufficient freedom to evade this problem, provided that the DM scalar is specifically a real singlet. By choosing suitable values for the non-minimal couplings at the Planck scale, it is possible for the unitarity-violation scale to be much larger than the inflaton field throughout inflation, so that the predictions of the model are robust. Therefore, in addition to providing a minimal candidate for WIMP DM, the extension of the SM by a non-minimally coupled real gauge singlet scalar can also account for inflation while having a consistent scale of unitarity-violation.

The paper is organized as follows. In section 2 we review the real gauge singlet scalar ex-

---

1The case of singlet DM added to Higgs Inflation was considered in [10].
tension of the SM and the S-inflation model. We estimate the predictions of the model for the spectral index $n_s$ and discuss the effect of constraints from inflation on the model parameter space. Section 3 considers the DM phenomenology of the model and the implications from DM searches. In section 4 we discuss how to connect these two aspects via renormalisation group evolution and which constraints follow from electroweak vacuum stability and perturbativity. The scale of unitarity-violation during inflation and the consistency of S-inflation are discussed in section 5. Finally, we present our results in section 6 and our conclusions in section 7. Additional details are provided in the Appendix.

2 The S-inflation model

S-inflation is a version of the non-minimally coupled inflation model of [7] in which the scalar field is identified with the gauge singlet scalar responsible for thermal relic cold DM. In the present work, we focus on the case of a real singlet scalar $s$. In the Jordan frame, which is the standard frame for interpreting measurements and calculating radiative corrections, the action for this model is

$$S_J = \int \sqrt{-g} d^4x \left[ \mathcal{L}_{\text{SM}} + (\partial_\mu H)^\dagger (\partial^\mu H) + \frac{1}{2} \partial_\mu s \partial^\mu s - \frac{m_P^2}{2} R - \xi s H^\dagger H R - \frac{1}{2} \xi s^2 R - V(s, H^\dagger H) \right],$$

where $\mathcal{L}_{\text{SM}}$ is the SM Lagrangian density minus the purely Higgs doublet terms, $m_P$ is the reduced Planck mass and

$$V(s^2, H^\dagger H) = \lambda_h \left[ \left( H^\dagger H - \frac{v^2}{2} \right)^2 + \frac{1}{2} \lambda_h s^2 H^\dagger H + \frac{1}{4} \lambda_s s^4 + \frac{1}{2} \xi s^2 H^\dagger H + \frac{1}{4} \lambda_s h^2 s^2 \right],$$

with $v = 246\text{GeV}$ the vacuum expectation value of the Higgs field. Writing $H = (h + v, 0)/\sqrt{2}$ with a real scalar $h$ we obtain

$$V(s^2, h) = V(h) + \frac{1}{2} m_s^2 s^2 + \frac{1}{4} \lambda_s s^4 + \frac{1}{2} \lambda_h v^2 s^2 + \frac{1}{4} \lambda_h h^2 s^2,$$

where we have introduced the physical singlet mass $m_s^2 = m_{s_0}^2 + \lambda_h v^2/2$.

In order to calculate the observables predicted by inflation, we perform a conformal transformation to the Einstein frame, where the non-minimal coupling to gravity disappears. In the case that $s \neq 0$ and $h = 0$, this transformation is defined by

$$\dot{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}, \quad \Omega^2 = 1 + \frac{\xi s^2}{m_P^2}.$$

The transformation yields

$$S_E = \int \sqrt{-\dot{g}} d^4x \left[ \mathcal{L}_{\text{SM}} + \frac{1}{2} \left( \frac{1}{\Omega^2} + \frac{6 \xi_s^2 s^2}{m_P^2 \Omega^4} \right) \dot{g}^{\mu\nu} \partial_\mu s \partial_\nu s - \frac{m_P^2}{2} \ddot{R} - \frac{V(s, 0)}{\Omega^4} \right],$$

where $\ddot{R}$ is the Ricci scalar with respect to $\dot{g}_{\mu\nu}$. We can then rescale the field using

$$\frac{d\chi_s}{ds} = \sqrt{\frac{\Omega^2 + 6 \xi_s^2 s^2/m_P^2}{\Omega^4}},$$

- 3 –
which gives
\[ S_E = \int \sqrt{-g} \, d^4x \left( \hat{\mathcal{L}}_{SM} - \frac{m_P^2}{2} \hat{R} + \frac{1}{2} \hat{g}^{\mu\nu} \partial_\mu \chi_s \partial_\nu \chi_s - U(\chi_s, 0) \right), \tag{2.7} \]
with
\[ U(\chi_s, 0) = \frac{\lambda_s s^4(\chi_s)}{4 \Omega^4}. \tag{2.8} \]
The relationship between \( s \) and \( \chi_s \) is determined by the solution to eq. (2.6). In particular, for \( s \gg m_P/\sqrt{\xi_s} \), the Einstein frame potential is
\[ U(\chi_s, 0) = \frac{\lambda_s m_P^4}{4 \xi_s^2} \left( 1 + \exp \left( -\frac{2\chi_s}{\sqrt{6} m_P} \right) \right)^2. \tag{2.9} \]
This is sufficiently flat at large \( \chi_s \) to support slow-roll inflation.

An analogous expression is obtained for the potential along the \( h \)-direction. In both cases, the Einstein frame potential is proportional to \( \lambda \phi/\xi_s^2 \), where \( \phi = s \) or \( h \). Therefore the minimum of the potential at large \( s \) and \( h \) will be very close to \( h = 0 \) and inflation will naturally occur along the \( s \)-direction if \( \lambda_s/\xi_s^2 \ll \lambda_h/\xi_h^2 \), which is true for example if \( \xi_s \gg \xi_h \) and \( \lambda_s \sim \lambda_h \).

In the following, inflation is always considered to be in the direction of \( s \) with \( h = 0 \). The conventional analysis of inflation can then be performed in the Einstein frame. After inflation, the Jordan and Einstein frames will be indistinguishable since \( \xi_s^2 \ll m_P^2 \) and so \( \Omega \to 1 \). Therefore the curvature perturbation spectrum calculated in the Einstein frame becomes equal to that observed in the physical Jordan frame at late times.

The classical (tree-level) predictions for the spectral index and tensor-to-scalar ratio are [33]
\[ n_s^{\text{tree}} \approx 1 - \frac{2}{N} - \frac{3}{2 N^2} + \mathcal{O}\left( \frac{1}{N^3} \right) = 0.965, \tag{2.10} \]
\[ r^{\text{tree}} \approx \frac{12}{N^2} + \mathcal{O}\left( \frac{1}{\xi_s N^2} \right) = 3.6 \times 10^{-3}, \tag{2.11} \]
while the field during inflation is
\[ s_{\text{N}}^2 \approx 4 m_P^2 \tilde{N}/(3 \xi_s). \tag{2.12} \]
In the equations above \( \tilde{N} \) is the number of e-foldings as defined in the Einstein frame, which differs from that in the Jordan frame by \( \tilde{N} \approx N + \ln(1/\sqrt{N}) \) [8], and we have used \( \tilde{N} = 58 \).\(^2\) The classical predictions are in good agreement with the most recent Planck values, \( n_s = 0.9677 \pm 0.0060 \) (68\% confidence level (CL), Planck TT + lowP + lensing) and \( r_{0.002} < 0.11 \) (95\% CL, Planck TT + lowP + lensing) [1].

The classical predictions for S-inflation are the same as those of any model based on the SBB model. Differences between S-inflation and other models do however arise from

\(^2\)Reheating in S-inflation occurs via stochastic resonance to Higgs bosons through the coupling \( \lambda_{hs} \). It was shown in [16] that this process is very efficient and makes quite precise predictions for the reheating temperature and the number of e-foldings of inflation, with \( 57 \leq \tilde{N} \leq 60 \) at the WMAP pivot scale. This in turn allows for quite precise predictions of the inflation observables.
quantum corrections to the effective potential. To include these corrections, we calculate
the RG evolution of the various couplings as a function of the renormalisation scale $\mu$ (see
section 4 for details). We can then obtain the renormalisation group (RG)-improved effective
potential for $s$ in the Jordan frame by replacing the couplings in eq. (2.3) by the running
couplings and setting $\mu$ equal to the value of the field. For $h = 0$ (and neglecting the singlet
mass term, $m_s \sim 1 \text{ TeV} \ll s$), this approach yields

$$V_{\text{RG}}(s^2, 0) = \frac{\lambda_s(s) s^4}{4}.$$  (2.13)

The RG-improved potential can then be transformed into the Einstein frame in order to
calculate the observables predicted by inflation.\(^3\) The inflationary parameters are calculated
using the methods discussed in [16]. In particular, the Einstein parameters are given by

$$\tilde{\epsilon} = \frac{m_P^2}{2} \left( \frac{1}{U} \frac{dU}{d\chi_s^2} \right)^2,$$
$$\tilde{\eta} = \frac{m_P^2}{U} \frac{d^2U}{d\chi_s^2},$$
$$\tilde{\xi}^2 = \frac{m_P^4}{U^2} \frac{dU}{d\chi_s} \frac{d^3U}{d\chi_s^2}.$$  (2.14)

3 Singlet scalar as dark matter

Let us now turn to the phenomenology of the singlet scalar in the present Universe and
at energies well below the scale of inflation [19, 24, 35]. Most importantly, the assumed
$\mathbb{Z}_2$ symmetry ensures the stability of the scalar, so that it can potentially account for the
observed abundance of DM [2, 3]. If the mass of the singlet is comparable to the electroweak
scale, the singlet is a typical WIMP, which obtains its relic abundance from thermal freeze-out.
Indeed, at low energies, where the effects of the non-minimal coupling to gravity are
negligible, our model becomes identical to what is often referred to as Higgs Portal Dark
Matter [21, 22, 36], because all interactions of the singlet with SM particles are mediated
by the Higgs. In this section we review the constraints on these models and determine the
parameter space allowed by the most recent experimental results. In the process, we point
out several discrepancies in the literature and resolve the resulting confusion.

3.1 Relic abundance

The calculation of the relic abundance of singlet scalars is discussed in detail in [19]. Three
kinds of processes are relevant for the annihilation of singlets into SM states: annihilation
into SM fermions, annihilation into SM gauge bosons and annihilation into two Higgs
particles. The first kind dominates as long as $m_s < m_W$, while for larger masses the second
kind gives the largest contribution. Notably, all of these processes can proceed via an $s$-channel
Higgs boson, leading to a resonant enhancement of the annihilation cross-section and a corresponding suppression of the DM relic abundance for $m_s \sim m_h/2$.\(^4\)

\(^3\)In [34] it was proposed to use of the Einstein frame for the computation of quantum corrections. The
Jordan frame analysis is, however, easier to implement correctly, being a straightforward extension of the
Standard Model analysis.

\(^4\)The process $ss \rightarrow hh$ also receives a contribution from $t$-channel singlet exchange, which gives a relevant
contribution if $\lambda_{hs}$ is large compared to $\lambda_h$. 
For the present work we calculate the singlet abundance using micrOMEGAs_3 [37], which numerically solves the Boltzmann equation while calculating the Higgs width in a self-consistent way. It is then straightforward to numerically find the coupling $\lambda_{hs}$ that gives $\Omega_s h^2 = 0.1197$, in order to reproduce the value of the DM density $\Omega_{DM} h^2 = 0.1197 \pm 0.0022$ determined by Planck (TT + lowP, 68% CL) [38]. For example, we find $\lambda_{hs} \approx 0.08$ for $m_s = 300$ GeV and $\lambda_{hs} \approx 0.30$ for $m_s = 1000$ GeV. These values agree with the ones found in [19, 23], but disagree with [20–22, 24] by a factor of 2 after accounting for the different conventions.\(^5\)

3.2 Direct detection constraints

The strongest constraints on $\lambda_{hs}$ stem from DM direct detection experiments, since the singlet-Higgs coupling induces spin-independent interactions between the DM particle and nuclei. The scattering cross-section at zero momentum transfer is given by [19]

$$\sigma_{SI} = \frac{\lambda_{hs}^2 f_N^2}{4\pi} \frac{\mu_r^2 m_n^2}{m_h^4 m_s^2}, \quad (3.1)$$

where $m_n$ is the neutron mass, $\mu_r = (m_s m_n)/(m_s + m_n)$ is the reduced mass and $f_N$ is the effective Higgs-nucleon coupling.\(^6\) In terms of the light-quark matrix elements $f_{Tq}^N$, the effective coupling can be written as

$$f_N = \left[ \frac{2}{9} + \frac{7}{9} \sum_{q=u,d,s} f_{Tq}^N \right]. \quad (3.2)$$

The values of $f_{Tq}^N$ can either be determined phenomenologically from baryon masses and meson-baryon scattering data or computed within lattice QCD. A comparison of the different methods was recently performed in [19] and we adopt their result of $f_N = 0.30$ for the effective coupling.

The scattering cross-section given above can be directly compared to the bound obtained from the LUX experiment [18]. Indeed, as shown in figure 1, LUX is typically sensitive to the same range of values for $\lambda_{hs}$ as what is implied by the relic density constraint. Specifically, the LUX bound excludes the mass ranges $5.7$ GeV $< m_s < 52.6$ GeV and $64.5$ GeV $< m_s < 92.8$ GeV.

3.3 Invisible Higgs decays

Direct detection experiments cannot constrain singlet scalars with a mass of a few GeV or less, since such particles would deposit too little energy in the detector to be observable. This parameter region can however be efficiently constrained by considering how the Higgs-singlet coupling $\lambda_{hs}$ would modify the branching ratios of the SM Higgs boson. The partial decay width for $h \rightarrow ss$ is given by\(^7\)

$$\Gamma(h \rightarrow ss) = \frac{\lambda_{hs}^2 v^2}{32\pi m_h} \sqrt{1 - \frac{4 m_s^2}{m_h^2}}. \quad (3.3)$$

\(^5\)The Higgs-singlet coupling is called $\lambda_{S}$ in [19], $\lambda_{HS}$ in [20], $\lambda_{SS}$ in [21, 22], $\lambda_{DM}$ in [23] and $a_2$ in [24]. The respective conventions are captured by $\lambda_{hs} = \lambda_{S}$, $\lambda_{HS}/2 = \lambda_{SS}/2 = \lambda_{DM}/2 = 2 a_2$.

\(^6\)Note that [24] uses an approximate expression valid for $m_s \gg m_n$, such that $\mu_r^2 \approx m_n^2$.

\(^7\)This equation agrees with [4, 19–22, 35] but disagrees with [23].
This theoretical prediction can be compared to the experimental bound on invisible Higgs decays from the LHC. Direct searches for invisible Higgs decays in the vector boson fusion channel give $\text{BR}(h \to \text{inv}) \lesssim 0.29$ [39]. A somewhat stronger bound can be obtained from the observation that in our model there are no additional contributions to the Higgs production cross-section and no modifications of the partial decay widths of the Higgs boson into SM final states. Therefore the presence of an invisible decay channel leads to an overall reduction of the signal strength in visible channels. A global fit of all observed decay channels (combined with the bounds on invisible Higgs decays) then gives $\text{BR}(h \to \text{inv}) \lesssim 0.26$ [40].

Crucially, the bound from invisible Higgs decays becomes independent of the singlet mass for $m_s \ll m_h/2$. Invisible Higgs decays will therefore provide the strongest constraints for small singlet masses. Indeed, this constraint rules out the entire mass region where direct detection experiments lose sensitivity (see figure 1). As a result, only two mass regions remain viable: a low-mass region $52.6 \text{ GeV} < m_s < 64.5 \text{ GeV}$ and a high-mass region $m_s \gtrsim 93 \text{ GeV}$.

### 3.4 Other constraints

It has been pointed out recently [24] that bounds on $\gamma$-ray lines from Fermi-LAT [41] rule out the parameter region where $m_s$ is slightly above $m_h/2$. To be safe from this constraint, we will focus on the mass range $52.6 \text{ GeV} < m_s < 62.4 \text{ GeV}$, which we shall refer to as the low-mass region. For the high-mass region, on the other hand, there are no strong constraints from indirect detection. Moreover, collider searches for singlet scalars with $m_s > m_h/2$ are extremely challenging [20–23, 35] and consequently, there are no relevant bounds from the LHC for the high-mass region [42].

The most significant improvements in sensitivity in the near future are expected to come from direct detection experiments. Indeed, XENON1T [43] is expected to improve upon current LUX constraints on the DM scattering cross section by a factor of about 50 and will therefore be able to probe the high-mass region up to $m_s \approx 4 \text{ TeV}$. As we will show, in S-inflation singlet masses larger than about $2 \text{ TeV}$ are excluded by the Planck 2-\(\sigma\) upper bound on $n_s$ and perturbativity. XENON1T will therefore be able to probe the entire high-mass region relevant for singlet inflation. Similarly, XENON1T can also further constrain the low-mass region and potentially probe singlet masses in the range $53 \text{ GeV} < m_s < 57 \text{ GeV}$.
4 Renormalisation group evolution and theoretical constraints

In order to connect the inflationary observables of our model to the measured SM parameters and the DM phenomenology discussed in the previous section, we need to calculate the evolution of all couplings under the RG equations \[8, 10, 12, 14, 44–47\]. Existing analyses have considered the RG equations for the SM at two-loop order and the contributions of the singlet sector and non-minimal coupling at one-loop order (see also the Appendix) \([8, 10]\).

To examine the issue of vacuum stability, we improve the accuracy of our analysis further by incorporating the three-loop RG equations for the SM gauge couplings \[48\] and the leading order three-loop corrections to the RG equations for \(\lambda_h\) and \(y_t\) \[49\].

When considering large field values for either \(s\) or \(h\), the RG equations are modified, because there is a suppression of scalar propagators. This suppression is captured by inserting a factor

\[
c_\phi = \frac{1 + \frac{\xi_s \phi^2}{m^2}}{1 + (6\xi_s + 1)\frac{\phi^2}{m^2}}
\]

with \(\phi = s (\phi = h)\) for each \(s (h)\) propagating in a loop \([8]\). The changes in the RG equations for large values of the Higgs field have been discussed in detail in \([47]\). The modifications resulting from large singlet field values can be found in \([8]\) and are reviewed in the Appendix.

Note that, when considering S-inflation, such that \(s \gg h\), we can set the suppression factor \(c_h = 1\).

We determine the values of the SM parameters at \(\mu = m_t\) following \([46]\). Using the most recent values from the Particle Data Group \([50]\)

\[
m_t = (173.2 \pm 0.9) \text{ GeV}, \quad m_H = (125.09 \pm 0.24) \text{ GeV}, \quad \alpha_S(m_Z) = 0.1185 \pm 0.0006, \quad (4.2)
\]

we obtain at \(\mu = m_t\)

\[
y_t = 0.936 \pm 0.005, \quad \lambda_h = 0.1260 \pm 0.0014, \quad g_S = 1.164 \pm 0.003. \quad (4.3)
\]

Unless explicitly stated otherwise, we will use the central values for all calculations below. For given couplings at \(\mu = m_t\), we then use the public code \textsc{RGErun 2.0.7} \([51]\) to calculate the couplings at higher scales.

In contrast to the remaining couplings, we fix the non-minimal couplings \(\xi_h\) and \(\xi_s\) at \(\mu = m_P\). In order to obtain the correct amplitude of the scalar power spectrum, we require

\[
\frac{U}{\bar{\epsilon}} = (0.00271 m_P)^4, \quad (4.4)
\]

where \(U\) and \(\bar{\epsilon}\) are the potential and the first slow-roll parameter in the Einstein frame at the beginning of inflation, as defined in eq. (2.8) and eq. (2.14) respectively. Imposing equation (4.4) allows us to determine \(\xi_s\) at the scale of inflation once all other parameters have been fixed. Note that, since the value of \(s\) at the beginning of inflation also depends on \(\xi_s\), equation (4.4) can only be solved numerically. We then iteratively determine the required value of \(\xi_s\) at the electroweak scale such that RG evolution yields the desired value at the scale of inflation.

The coupling \(\xi_h\) plays a very limited role for the phenomenology of our model because we do not consider the case of large Higgs field values for inflation. As a result our predictions

\[8\] We thank Kyle Allison for sharing his numerical implementation of these equations.
for the inflationary observables show only a very mild dependence on $\xi_h$, so that $\xi_h$ can essentially be chosen arbitrarily. Nevertheless, it is not possible to simply set this parameter to zero, since radiative corrections induce a mixing between $\xi_s$ and $\xi_h$. Moreover, we will see below that the value of $\xi_h$ plays an important role for determining whether our model violates unitarity below the scale of inflation. As with $\xi_s$, we fix $\xi_h$ at $m_P$ and then determine iteratively the required value of $\xi_h$ at the electroweak scale.\(^9\)

4.1 Metastability

We now discuss various theoretical constraints related to the RG evolution of the parameters in our model. It is a well-known fact that for the central values of the measured SM parameters, the electroweak vacuum becomes metastable at high scales, because the quartic Higgs coupling $\lambda_h$ runs to negative values (see e.g. [46]). This metastability is not in any obvious way a problem, as the lifetime of the electroweak vacuum is well above the age of the Universe [52] (note, however, that this estimate may potentially be spoiled by effects from Planck-scale higher-dimensional operators [53–55]). However, one may speculate that a stable electroweak vacuum is necessary for a consistent theory, for example if the vacuum energy relative to the absolute minimum is a physical energy density leading to inflation. It is therefore an interesting aspect of singlet extensions of the SM that scalar singlets give a positive contribution to the running of $\lambda_h$ [56, 57]:

$$\beta_{\lambda_h} = \beta_{\lambda_h}^{SM} + \frac{1}{32\pi^2} \xi_s^2 \lambda_h^2. \quad (4.5)$$

In fact, it was shown in [58] that, for the case of a minimally-coupled singlet, $\lambda_h s$ can be chosen such that the electroweak vacuum remains stable all the way up to the Planck scale and at the same time (for appropriate choices of the singlet mass $m_s$) the singlet obtains a thermal relic density compatible with the observed DM abundance.\(^10\)

In order to study electroweak vacuum stability, we need to consider the potential in the $h$-direction with $s = 0$.\(^11\) Vacuum stability then requires that $\lambda_h(\mu) > 0$ for $\mu$ up to $m_P$. In the present study, we consider both the case where $\lambda_h s$ is sufficiently large to stabilise the electroweak vacuum and the case where $\lambda_h s$ only increases the lifetime of the metastable vacuum, but does not render it completely stable. We focus throughout on the case where $\lambda_h s$ is positive.

4.2 Examples

Figure 2 shows an example for the evolution of scalar couplings (left) and the non-minimal couplings (right) under the RG equations discussed above. Solid lines correspond to the case $s \gg h$, which is relevant for inflation, while dotted lines correspond to $h \gg s$, which is relevant for vacuum stability. We fix the scalar couplings at the weak scale, choosing $\lambda_h s = 0.25$ and $\lambda_s = 0.01$, such that the observed relic abundance can be reproduced for $m_s = 835$ GeV. We consider $\lambda_s \ll \lambda_h s$, in which case the value of $\xi_s$ necessary to obtain the correct amplitude of the scalar power spectrum is reduced. For our choice, we find $\xi_s \sim 10^4$. Note, however, that

---

\(^9\)The running of $\xi_h$ between $m_P$ and the scale of inflation is completely negligible, since the relevant diagrams are strongly suppressed for large values of $s$.

\(^10\)Note that if the singlet mixes with the Higgs, there will be additional threshold effects at $\mu = m_s$ from integrating out the singlet [59]. In the setup we consider, however, this effect is not important [60].

\(^11\)A more detailed study of the potential in general directions with both $s \neq 0$ and $h \neq 0$ (along the lines of [61]) is beyond the scope of the present work.
while $\lambda_{hs}$ exhibits only moderate running, $\lambda_s$ grows significantly with the renormalisation scale $\mu$, because its $\beta$-function contains a term proportional to $\lambda_{hs}^2$. Choosing even smaller values of $\lambda_s$ at the weak scale will therefore not significantly reduce its value at the scale of inflation nor the corresponding value of $\xi_s$.

An important observation from figure 2 is that $\lambda_h$ does not run negative and hence the electroweak vacuum remains stable all the way up to the Planck scale. The additional contribution from the singlet scalar is sufficient to ensure $\lambda_h > 10^{-3}$ for all renormalisation scales up to $m_P$. For field values $s \gtrsim 10^{15}$ GeV $\gg h$, the singlet propagator is suppressed, leading to a visible kink in the running of $\lambda_h$. We show the propagator suppression factor $c_s$ in the right panel of figure 2. One can clearly see how this suppression factor affects the running of $\xi_h$, which becomes nearly constant for $s \gtrsim 10^{15}$ GeV. In this particular example, we have chosen $\xi_h(m_P) = 1000$ (for $s \gg h$). This choice, together with $\lambda_s \ll 1$, implies that the running of $\xi_s(\mu)$ from $m_t$ to $m_P$ is negligible.

In the low-mass region, we are interested in much smaller values of $\lambda_{hs}$, typically below $10^{-2}$. A particular example is shown in figure 3 (left) for the representative choice $\lambda_{hs}(m_t) = 0.002$ and $\lambda_s = 0.0005$, which yields the observed relic abundance for $m_s \approx 57$ GeV. We observe that if $\lambda_s$ and $\lambda_{hs}$ are both small at the electroweak scale, these couplings exhibit only very little running up to the scale of inflation. For the same reason, it is impossible to influence the running of the Higgs couplings sufficiently to prevent $\lambda_{hs}$ from running negative at around $10^{11}$ GeV.

If $\lambda_{hs}$ is small, we can obtain the correct scalar power spectrum amplitude with a much smaller value of $\xi_s$ during inflation. For the specific case considered in figure 3, we find $\xi_s \sim 10^3$. For these values of $\xi_s$ and $\lambda_{hs}$, the loop-induced corrections to $\xi_h$ are very small and hence this coupling changes only very slightly under RG evolution.

4.3 Perturbativity

In order for our calculation of the running couplings and the radiative corrections to the potential to be reliable, we must require that all couplings remain perturbative up to the scale of inflation, which is typically $10^{17}$–$10^{18}$ GeV. This requirement is easily satisfied for the SM couplings, but needs to be checked explicitly for the couplings of the singlet, which can grow significantly with increasing renormalisation scale $\mu$. We follow [58, 62] and use the requirement of perturbative unitarity to impose an upper bound on the scalar couplings.
Figure 3. Running of the scalar couplings $\lambda_h$, $\lambda_s$ and $\lambda_{hs}$ (left) and of the non-minimal couplings $\xi_h$ and $\xi_s$ (right) as a function of the renormalisation scale $\mu$ for a typical parameter point in the low-mass region. In the right panel, we also show the suppression factor $c_s$, which modifies the running at large field value. Note that $\lambda_h$ runs negative for $\mu \gtrsim 10^{11}$ GeV.

This procedure gives

$$\lambda_s < \frac{4\pi}{3} \quad \text{and} \quad \lambda_{hs} < 8\pi.$$  \hspace{1cm} (4.6)

As we will see below, the non-minimal coupling $\xi_s$ can be much larger than unity without invalidating a perturbative calculation. Nevertheless, if $\xi_s$ and $\xi_h$ are both very large, processes involving both couplings may violate perturbative unitarity, implying that there may be new physics or strong coupling below the scale of inflation. We will now discuss this issue in more detail.

5 Unitarity-violation during inflation

In this section we estimate the scale of perturbative unitarity-violation as a function of the background inflaton field. Note that by “unitarity-violation scale” we mean the scale at which perturbation theory in scalar scattering breaks down, so this may in fact indicate the onset of unitarity-conserving scattering in a strongly-coupled regime [28, 31]. We will consider unitarity-violation in the scattering of scalar particles corresponding to perturbations about the background field. In the case of a real scalar $s$ and the fields of the Higgs doublet $H$, there are two distinct scattering processes we need to consider: (i) $\delta s h_1 \leftrightarrow \delta s h_1$ and (ii) $h_1 h_2 \leftrightarrow h_1 h_2$, where $\delta s$ is the perturbation about the background $s$ field and $h_1$ and $h_2$ are two of the Higgs doublet scalars. Scattering with the other scalars in $H$ is equivalent to these two processes. We will use dimensional analysis to estimate the scale of unitarity-violation by determining the leading-order processes in the Einstein frame which result in unitarity-violating scattering.

It will be sufficient to consider the Einstein frame Lagrangian for two real scalar fields $\phi_i$, where — using the notation of [30] — $\phi_i$ stands for either $s$ or a component of the Higgs doublet. Unitarity-violation requires that there are two different scalars in the scattering process, since in the case of a single scalar there is a cancellation between $s$-, $t$- and $u$-channel amplitudes [63].

Since unitarity-violating scattering in the Jordan frame is due to graviton exchange via the non-minimal coupling to $R$, we can set $V = 0$. The Einstein frame action for two real
scalars is then of the form

\[ S_E = \int d^4 x \sqrt{-\tilde{g}} \left[ \mathcal{L}_{ii} + \sum_{i<j} \mathcal{L}_{ij} - \frac{1}{2} m_P^2 \tilde{R} \right], \]  

(5.1)

where

\[ \mathcal{L}_{ii} = \frac{1}{2} \left( \frac{\Omega^2 + 6 \xi_i \phi_i^2}{\Omega^4} \right) \tilde{g}^{\mu\nu} \partial_\mu \phi_i \partial_\nu \phi_i \quad \text{and} \quad \mathcal{L}_{ij} = \frac{6 \xi_i \xi_j \phi_i \phi_j \tilde{g}^{\mu\nu} \partial_\mu \phi_i \partial_\nu \phi_j}{m_P^2 \Omega^4} \]  

and

\[ \Omega^2 = 1 + \frac{\xi_j \phi_j^2}{m_P^2}, \]  

(5.3)

The interaction terms proportional to \(\xi_i, \xi_j\) are responsible for the dominant unitarity-violation in scattering cross-sections calculated in the Einstein frame. These interactions are the Einstein frame analogue of scalar scattering via graviton exchange in the Jordan frame due to the non-minimal coupling. To obtain the scale of unitarity-violation in terms of the physical energy defined in the Jordan frame, we first canonically normalize the fields in the Einstein frame, then estimate the magnitude of the scattering matrix element and finally transform the unitarity-violation scale in the Einstein frame back to that in the Jordan frame.

In the following we will denote the inflaton by \(\phi_1 (\equiv s)\), which we expand about the background field, i.e. \(\phi_1 = \bar{\phi}_1 + \delta \phi_1\). The Higgs doublet scalars are denoted by \(\phi_2 (\equiv h_1)\) and \(\phi_3 (\equiv h_2)\). The corresponding canonically normalized scattering fields in the Einstein frame are then defined to be \(\varphi_1, \varphi_2\) and \(\varphi_3\). Once we have determined the interactions of the canonically normalised fields, we use dimensional analysis to estimate the scale of tree-level unitarity-violation. For this purpose we introduce appropriate factors of \(\tilde{E}\) to make the coefficient of the interaction terms in \(\mathcal{L}\) dimensionless. Energy scales \(\tilde{E}\) in the Einstein frame are related to the ones in the Jordan frame via \(\tilde{E} = E/\Omega\), where during inflation \(\Omega^2 \approx \xi_1 \phi_1^2 / m_P^2 \approx N \gg 1\). Unitarity conservation implies that the matrix element for any \(2 \leftrightarrow 2\) scattering process should be smaller than \(O(1)\), so we can determine the scale of unitarity-violation (denoted by \(\Lambda\)) by determining the value of \(\tilde{E}\) that saturates this bound. This was demonstrated explicitly in [8], by comparing the dimensional estimate with the exact value from the full scattering amplitude.

In the following, we consider three regimes for \(\bar{\phi}_1\), each of which leads to a different form of the Lagrangian and the scattering amplitudes:

- **Regime A**: Large field values. In this regime we have \(\Omega > 1\) (implying that \(\xi_1 \bar{\phi}_1^2 / m_P^2 > 1\) and \(6 \xi_1^2 \bar{\phi}_1^2 / m_P^2 > 1\).

- **Regime B**: Medium field values. In this regime we have approximately \(\Omega \approx 1\) (implying that \(\xi_1 \bar{\phi}_1^2 / m_P^2 < 1\), but still \(6 \xi_1^2 \bar{\phi}_1^2 / m_P^2 > 1\).

- **Regime C**: Small field values. Finally we consider \(6 \xi_1^2 \bar{\phi}_1^2 / m_P^2 < 1\), which in particular implies \(\Omega \approx 1\).
5.1 Regime A: Large field values

In this case the canonically normalized fields are $\varphi_1 = \sqrt{6} m_P \delta \phi_1/\phi_1$ and $\varphi_{2,3} = \phi_{2,3}/\Omega$. The interaction leading to unitarity-violation in $\delta s_1$ scattering is

$$\tilde{\mathcal{L}} \supset \frac{6 \xi_1 \xi_2}{m_P \Omega^4} (\delta \phi_1 + \delta \phi_1) \phi_2 \tilde{g}^{\mu \nu} \partial_\mu \delta \phi_1 \partial_\nu \phi_2.$$  

(5.4)

This results in a 3-point and a 4-point interaction. After rescaling to canonically normalized fields, the 3-point interaction is

$$\tilde{\mathcal{L}} \supset \frac{\sqrt{6} \xi_2}{m_P} \phi_2 \tilde{g}^{\mu \nu} \partial_\mu \delta \phi_1 \partial_\nu \phi_2.$$  

(5.5)

This interaction can mediate $\varphi_1 \varphi_2 \leftrightarrow \varphi_1 \varphi_2$ scattering at energy $\tilde{E}$ via $\varphi_2$ exchange, with a matrix element given dimensionally by $|\mathcal{M}| \sim \tilde{E}^2 \xi_2^2 / m_P^2$. Unitarity is violated once $|\mathcal{M}| \sim 1$, therefore the unitarity-violation scale in the Einstein frame is

$$\tilde{\Lambda}_{12}^{(3)} \sim \frac{m_P}{\xi_2},$$  

(5.6)

where the superscript (3) denotes unitarity-violation due to the 3-point interaction. In the Jordan frame $\Lambda_{12}^{(3)} = \Omega \tilde{\Lambda}_{12}^{(3)}$, where $\Omega \approx \sqrt{\xi_1} \phi_1/m_P$, therefore

$$\Lambda_{12}^{(3)} \sim \frac{\sqrt{\xi_1}}{\xi_2} \phi_1.$$  

(5.7)

Similarly, the 4-point interaction has an Einstein frame matrix element given by $|\mathcal{M}| \sim \xi_2 \tilde{E}^2 / m_P^2$, therefore the scale of unitarity-violation in the Jordan frame is

$$\Lambda_{12}^{(4)} \sim \sqrt{\frac{\xi_1}{\xi_2}} \phi_1.$$  

(5.8)

For $\xi_2 > 1$, this is larger than $\Lambda_{12}^{(3)}$, therefore $\Lambda_{12}^{(3)}$ is the dominant scale of unitarity-violation. In general these estimates of the unitarity-violation scales are valid if the scalars can be considered massless, which will be true if $\phi_1 < \Lambda_{12}^{(3)}$, i.e. for $\sqrt{\xi_1} > \xi_2$.

In the case of Higgs scattering $\varphi_2 \varphi_3 \leftrightarrow \varphi_2 \varphi_3$, there is only the 4-point interaction following from

$$\tilde{\mathcal{L}} \supset \frac{6 \xi_2 \xi_3}{m_P^2 \Omega^4} \phi_2 \phi_3 \tilde{g}^{\mu \nu} \partial_\mu \phi_2 \partial_\nu \phi_3.$$  

(5.9)

As the canonically normalized Higgs fields are in general given by $\varphi_{2,3} = \phi_{2,3}/\Omega$, the unitarity-violation scale in the Einstein frame is generally $\tilde{\Lambda}_{23} \sim m_P / \sqrt{\xi_2 \xi_3} \equiv m_P / \xi_2$ (since $\xi_2 = \xi_3$ if both scalars are part of the Higgs doublet). On translating the energy to the Jordan frame, the unitarity-violation scale becomes

$$\Lambda_{23} \sim \frac{\sqrt{\xi_1}}{\xi_2} \phi_1,$$  

(5.10)

which is the same expression as for $\Lambda_{12}^{(3)}$. 

– 13 –
5.2 Regime B: Medium field values

In this case the canonically normalized fields are $\varphi_1 = \sqrt{6} \xi_1 \phi_1 / m_p$ and $\varphi_2 = \phi_2$. Since $\Omega = 1$, the energies are the same in the Einstein and Jordan frames. Using the same procedure as before, we find

$$\Lambda_{12}^{(3)} \sim \frac{m_p}{\xi_2}, \quad (5.11)$$

and

$$\Lambda_{12}^{(4)} \sim \sqrt{\frac{\xi_1}{\xi_2}} \phi_1. \quad (5.12)$$

For Higgs scattering we obtain

$$\Lambda_{23} \sim \frac{m_p}{\xi_2}. \quad (5.13)$$

5.3 Regime C: Small field values

In this case the Einstein and Jordan frames are completely equivalent. Therefore

$$\Lambda_{12}^{(3)} \sim \frac{m_p^2}{6 \xi_1 \xi_2 + 1}, \quad (5.14)$$

and

$$\Lambda_{12}^{(4)} \sim \frac{m_p}{\sqrt{\xi_1 \xi_2}}. \quad (5.15)$$

For Higgs scattering we obtain

$$\Lambda_{23} \sim \frac{m_p}{\xi_2}. \quad (5.16)$$

5.4 Discussion

In summary, for $\xi_s > \xi_h$ the smallest (and so dominant) scale of unitarity-violation in each regime is given by:

A : $\Lambda_{sh}^{(3)} \sim \Lambda_h \sim \frac{\sqrt{6} \xi_h}{\xi_s}$

B : $\Lambda_{sh}^{(3)} \sim \Lambda_h \sim \frac{m_p}{\sqrt{\xi_s}}$

C : $\Lambda_{sh}^{(4)} \sim \frac{m_p}{\sqrt{\xi_s} \xi_h}$, \quad (5.17)

where $\Lambda_{sh}^{(3)} \equiv \Lambda_{12}^{(3)}, \Lambda_h \equiv \Lambda_{23}, \xi_s \equiv \xi_1$ and $\xi_h \equiv \xi_2$.

In the case of Higgs Inflation, the scale of unitarity-violation is obtained as above but with $\xi_s$ set equal to $\xi_h$, since the inflaton is now a component of $H$. During inflation $\Lambda \approx \phi_1 / \sqrt{\xi_h}$, with $\xi_h \sim 10^{15}$. Since this energy scale is less than $\phi_1$, the gauge bosons become massive and only the physical Higgs scalar takes part in scattering. Since unitarity-violation requires more than one massless non-minimally coupled scalar, there is no unitarity-violation at energies less than $\phi_1$. Unitarity-violation therefore occurs at $\Lambda \approx m_W(\phi_1) \approx \phi_1$, i.e. the unitarity-violation scale is essentially equal to the Higgs field value during inflation [14, 32].

As a result, either the new physics associated with unitarising the theory or strong coupling effects are expected to significantly modify the effective potential during inflation. It is uncertain in this case whether inflation is even possible and its predictions are unclear.
In S-inflation, on the other hand, it is possible to ensure that $\Lambda \gg \sigma$ provided $\xi_h$ is sufficiently small compared to $\xi_s$ at the scale of inflation. This is illustrated in figure 4 for $\xi_h(m_P) = 1000$ (left) and $\xi_h(m_P) = 1$ (right), taking $\lambda_{hs}(m_t) = 0.25$, $\lambda_s(m_t) = 0.01$ and $\xi_s(m_P) \approx 10^2$ as above. Both panels show the scale of unitarity-violation $\Lambda$ as a function of the field value $\sigma$. For $\xi_h(m_P) = 1000$ we observe that $\Lambda < \sigma$ for $\sigma \gtrsim 10^{16}$ GeV, which is significantly smaller than the field value at the beginning of inflation. For $\xi_h(m_P) = 1$, on the other hand, $\Lambda$ always remains larger than $\sigma$. In this case it is reasonable to assume that new physics, in the form of additional particles with mass of order $\Lambda$ (or strong coupling effects\textsuperscript{12}), will have only a small effect on the effective potential at the scale $\mu = \sigma$.

The right panel of figure 4 exhibits another new feature: We find $\Lambda \rightarrow \infty$ for $\sigma \sim 10^{15}$ GeV. The reason is that, as already observed in figure 2, $\xi_h(m_P)$ exhibits a strong running for $\mu < 10^{15}$ GeV. Consequently, if we fix $\xi_h$ to a rather small value at the Planck scale, e.g. $\xi_h(m_P) = 1$, $\xi_h(\mu)$ will run negative at lower scales. During the transition, $\xi_h(\mu)$ will be very small and hence the scale of unitarity-violation can be very large.

It should be emphasized that the advantage of S-inflation with respect to unitarity-violation is only obtained if the singlet is a real scalar. In the case of a complex singlet, the real and imaginary parts of the scalar both have the same non-minimal coupling $\xi_s$. Therefore we would have $\xi_1 = \xi_2$ in the above analysis and the unitarity-violation scale would become the same as in Higgs Inflation. Therefore the requirement that unitarity is not violated during inflation predicts that the DM scalar is a real singlet scalar.

6 Results

In this section we combine the experimental and theoretical constraints discussed above and present the viable parameter space for our model. Out of the five free parameters ($\lambda_s$, $\lambda_{hs}$, $m_s$, $\xi_s$ and $\xi_h$) we can eliminate $\lambda_{hs}$ (or $m_s$) by requiring the model to yield the observed DM abundance (see figure 1) and $\xi_s$ by imposing the correct amplitude of the scalar power

\textsuperscript{12}In unitarisation by strong coupling, $\Lambda$ is automatically field dependent and equal to the scale at which the potential is expected to change. In unitarisation by new particles, on the other hand, $\Lambda$ is only an upper bound on the masses of the new particles. Moreover the masses need not be field-dependent in order to unitarise the theory. Therefore strong coupling is more naturally compatible with the scale of inflation.
Figure 5. Predictions for inflation in the high-mass region (500 GeV < m_s < 2500 GeV) as a function of m_s and λ_s for fixed ξ_h(m_P) = 100. Shown are the deviations from the tree-level predictions n_s^{tree} = 0.965 (left) as well as the value of ξ_s at the beginning of inflation (right). For each value of m_s the coupling λ_h s has been fixed by the relic density requirement, as shown on the top of each panel. The grey shaded region indicates the parameter region where couplings become non-perturbative below the scale of inflation and the purple shaded region indicates the parameter space where λ_h runs negative below the scale of inflation, leading to a metastable vacuum. We furthermore show the parameter region excluded by the upper bound on n_s from Planck at 95% CL (shaded in blue). The green line in the right panel indicates the value of ξ_s where unitarity is violated at the scale of inflation.

spectrum. In the following we always ensure that these two basic requirements are satisfied, and then consider additional constraints in terms of the remaining parameters λ_s, m_s (or λ_h s) and ξ_h. We begin with a detailed discussion of the high-mass region and then turn to the low-mass region.

6.1 The high-mass region

Let us for the moment fix ξ_h(m_P) = 100 and study how the predictions depend on λ_s and m_s (or, alternatively, λ_h s). The left panel of figure 5 shows the predicted value of n_s compared to the tree-level estimate n_s^{tree} = 0.965. We find that in our model n_s is always slightly larger than the tree-level estimate, but the differences are typically ∆n_s < 0.01. Only for λ_h s > 0.5, corresponding to m_s > 2 TeV, do the differences grow so large that the model can be excluded by the Planck 2-σ bound, n_s < 0.98. In the same parameter region we find the largest differences between the tree-level predictions of r and α (see section 2) and the value predicted in our model. However, we find these deviations to be negligibly small. In particular our model predicts r < 0.01 everywhere, i.e. the tensor-to-scalar ratio would be very difficult to observe in the near future. Figure 5 also shows the parameter region excluded by the requirements that all couplings remain perturbative up to the scale of inflation (shaded in grey). This constraint requires λ_s \lesssim 0.3 for small values of λ_h s and becomes more severe with increasing λ_h s.

Finally, we also show the parameter region where λ_h s is too small to prevent λ_h from running to negative values in the h-direction (shaded in purple). While this is not fatal for

\[\text{Next generation CMB satellites, such as PIXIE [64] and LiteBIRD [65], plan to measure } r \text{ to an accuracy of } \delta r < 0.001. \text{ This would be sufficient to detect the tensor-to-scalar ratio in our model.}\]
the model (the electroweak vacuum remains metastable with a lifetime that is longer than the one predicted for the SM alone), this constraint may be physically significant depending upon the interpretation of the energy of the metastable state. We note, however, that the bound from metastability depends very sensitively on the assumed values of the SM parameters at the electroweak scale. Indeed, it is still possible within experimental uncertainties (at 95% CL) that the electroweak vacuum is completely stable even in the absence of any new physics [46].

For the currently preferred values of the SM parameters, we find the interesting parameter region to be $0.2 \lesssim \lambda_{hs} \lesssim 0.6$ corresponding roughly to $700 \text{ GeV} \lesssim m_s \lesssim 2 \text{ TeV}$. Very significantly, this entire range of masses and couplings can potentially be probed by XENON1T.

To study the predictions of inflation — and in particular the scale of unitarity-violation — in more detail, we show in the right panel of figure 5 the value of $\xi_s$ (at the scale of inflation) required by the scalar power spectrum amplitude. We typically find values around $10^4$, although values as large as $10^5$ become necessary as $\lambda_s$ comes close to the perturbative bound. Since we have fixed $\xi_h(m_P) = 100$ in this plot, such large values of $\xi_s$ imply that $\sqrt{\xi_s}/\xi_h > 1$, which in turn means that the scale of unitarity-violation is larger than the field value $\bar{s}$ at the beginning of inflation. Conversely, if both $\lambda_s$ and $\lambda_{hs}$ are small, $\xi_s$ can be significantly below $10^4$, such that unitarity is violated below the scale of inflation. The parameters for which the scale of unitarity-violation is equal to the scale of inflation is indicated by a green line.

Let us now turn to the dependence of our results on the value of $\xi_h(m_P)$. For this purpose, we fix $\lambda_s = 0.01$ and consider the effect of varying $\xi_h(m_P)$ in the range $0 \leq \xi_h(m_P) \leq 1000$. We find that neither the constraint from Planck, nor the bounds from metastability and perturbativity depend strongly on $\xi_h(m_P)$. Nevertheless, as discussed in section 5, $\xi_h(m_P)$ does play a crucial role for the scale of unitarity-violation. We therefore show in the left panel of figure 6 the scale of unitarity-violation at the beginning of inflation divided by the field value $\bar{s}$ at the beginning of inflation. This ratio is to be larger than unity in order to avoid unitarity-violation. As indicated by the green line, this requirement implies $\xi_h(m_P) \lesssim 150$ in the parameter region of interest.

As discussed in section 4, small values of $\xi_h(m_P)$ imply that $\xi_h(\mu)$ will run to negative values for $\mu \rightarrow m_t$. To conclude our discussion of the high-mass region we therefore show in the right panel of figure 6 the magnitude of $\xi_h(m_t)$ as a function of $m_s$ and $\xi_h(m_P)$. The thick black line indicates the transition between $\xi_h(m_t) > 0$ and $\xi_h(m_t) < 0$. By comparing this plot with the one to the left, we conclude that within the parameter region that avoids unitarity-violation $\xi_h(m_t)$ necessarily becomes negative.

6.2 The low-mass region

We study the predictions for the low-mass region in figure 7. The top row shows $\Delta n_s$ and $\xi_s$ at the beginning of inflation as a function of $\lambda_s$ and $\lambda_{hs}$ for $\xi_h(m_P) = 1$. These plots are analogous to the ones for the high-mass region in figure 5. The crucial observation is that, unless $\lambda_s \ll \lambda_{hs}$, radiative corrections to the inflationary potential are completely negligible, because any contribution proportional to $\lambda_s$ is suppressed by powers of $c_s \ll 1$ during inflation. Consequently, in most of the low-mass region, $n_s$ is identical to its tree-level value. We furthermore find that, as expected, the tensor-to-scalar ratio and the running of the spectral index are both unobservably small. Since radiative corrections play such a small
role, the value of $\xi_s$ required from inflation depends almost exclusively on $\lambda_s$. As a result, it is easily possible to have $\xi_s < 1000$ during inflation for $\lambda_s < 10^{-3}$ and $\xi_s < 100$ for $\lambda_s < 10^{-5}$.

Since $\xi_s$ can be much smaller in the low-mass region than in the high-mass region, it is natural to also choose a very small value for $\xi_h$. In fact, for typical values in the low-mass region $\xi_s \sim 10^3$ and $\lambda_s \sim \lambda_{hs} \sim 10^{-3}$, the loop-induced contribution to $\xi_h$ is $\Delta \xi_h < 10^{-2}$, so that it is technically natural to have $\xi_h \ll 1$. One then obtains $\sqrt{\xi_s/\xi_h} \gg 1$ and hence the scale of unitarity-violation is well above the scale of inflation. It is therefore possible without difficulty to solve the issue of unitarity-violation in the low-mass region.

This conclusion is illustrated in the bottom row of figure 7, which should be compared to figure 6 from the high-mass case, except that we keep $\xi_h(m_P) = 1$ fixed and vary $\lambda_s$ instead. The bottom-left plot clearly shows that (for our choice of $\xi_h$) the scale of unitarity-violation is always well above the field value at the beginning of inflation, so that the problem of unitarity-violation can easily be solved in the low-mass region. In addition, if the scalar couplings are sufficiently small, $\xi_h$ will have negligible running from the Planck scale down to the weak scale. It is therefore easily possible to set e.g. $\xi_h(m_P) = 1$ and still have $\xi_h(m_t) > 0$, as illustrated in the bottom-right panel of figure 7. It is not possible, however, to ensure at the same time that $\lambda_h$ remains positive for large field values of $h$. In other words, the electroweak vacuum is always metastable in the low-mass region (for the preferred SM parameters).

In the high-mass region we found that the coupling $\lambda_{hs}$ is bounded from below by the desire to stabilise the electroweak vacuum and from above by constraints from Planck and the requirement of perturbativity. In the low-mass region, on the other hand, we obtain an upper bound on $\lambda_{hs}$ from LUX and a lower bound on $\lambda_{hs}$ from the relic density requirement. Compared to the high-mass region, the allowed range of couplings in the low-mass region is much larger and therefore much harder to probe in direct detection experiments. If indeed $m_s$ is very close to $m_h/2$, and $\lambda_s$, $\lambda_{hs} < 10^{-3}$, it will be a great challenge to test the model predictions with cosmological or particle physics measurements.
Figure 7. Top row: Predictions for inflation in the low-mass region (52.5 GeV < m_s < 62.5 GeV) as a function of the couplings λ hs and λ s for fixed ξ_h (m_P) = 1. Shown are the deviations from the tree-level predictions n_s^{\text{tree}} = 0.965 (left) and the value of ξ_s at the beginning of inflation (right). Bottom row: The scale of unitarity-violation Λ compared to the field value at the beginning of inflation s_{\inf} (left) and the value of ξ_h at the electroweak scale (right). In all panels the grey shaded region indicates the parameter region where couplings become non-perturbative below the scale of inflation and the light blue shaded region represents the 95% CL bound from LUX. We do not show the metastability bound, since it covers the entire low-mass region.

7 Conclusions

The origin of inflation and the nature of DM are two of the fundamental questions of cosmology. In the present work, we have revisited the possibility that both issues are unified by having a common explanation in terms of a real gauge singlet scalar, which is one of the simplest possible extensions of the SM. Considering the most recent experimental constraints for this model from direct detection experiments, the LHC and Planck, we have shown that large regions of parameter space remain viable. Furthermore, we find that in parts of the parameter space the scalar singlet can stabilise the electroweak vacuum all the way up to the Planck scale, while at the same time avoiding the problem of unitarity-violation present in conventional models of Higgs inflation.

The scalar singlet can efficiently pair-annihilate into SM particles via the Higgs portal,
so that it is straight-forward in this model to reproduce the observed DM relic abundance via thermal freeze-out. We find two distinct mass regions where the model is consistent with experimental constraints from LUX, LHC searches for invisible Higgs decays and Fermi-LAT: the low-mass region, $53 \text{ GeV} \lesssim m_s \lesssim 62.4 \text{ GeV}$, where DM annihilation via Higgs exchange receives a resonant enhancement, and the high-mass region, $m_s \gtrsim 93 \text{ GeV}$, where a large number of annihilation channels are allowed.

In both mass regions it is possible without problems to fix the non-minimal couplings $\xi_s$ and $\xi_h$ in such a way that inflation proceeds in agreement with all present constraints. In particular, the tensor-to-scalar ratio and the running of the spectral index are expected to be unobservably small. On the other hand, radiative corrections to the spectral index typically lead to a value of $n_s$ slightly larger than the classical estimate, i.e. $n_s > 0.965$. This effect is largest for large values of $m_s$ and $\lambda_{hs}$ and current Planck constraints already require $m_s \lesssim 2 \text{ TeV}$. The entire high-mass region compatible with Planck constraints will therefore be tested by XENON1T, which can constrain gauge singlet scalar DM up to $m_s \sim 4 \text{ TeV}$.

In the high-mass region, the value of $\xi_s$ required to obtain a sufficiently flat potential during inflation is typically $\xi_s \sim 10^4$–$10^5$. In spite of such a large non-minimal coupling, it is possible to have unitarity-conservation during inflation, in the sense that the scale of unitarity-violation can be much larger than the inflaton field. The reason is that only $\xi_s$ needs to be large in order to reproduce the observed density perturbation, while the Higgs non-minimal coupling $\xi_h$ can be arbitrarily small. In the limit $\xi_h \rightarrow 0$ there will be only one non-minimally coupled scalar field and therefore no unitarity-violation, provided that the inflaton, and so the DM particle, is a real scalar.

We find that at large singlet field values $\bar{s}$ the scale of unitarity-violation is given by $\Lambda \sim \bar{s}\sqrt{\xi_s/\xi_h}$. If the non-minimal couplings satisfy $\xi_s(m_P) \gg \xi_h(m_P)$ at the Planck scale, it is possible for the unitarity-violation scale during inflation to be orders of magnitude larger than $\bar{s}$. Such a hierarchy of couplings is stable under radiative corrections and consistent with the assumption that inflation proceeds along the $s$-direction. Furthermore, in the low-mass region $\lambda_{hs}$ and $\lambda_s$ can be so small that $\xi_s \sim 10^2$–$10^3$ is sufficient to obtain a flat enough potential.

We conclude that it is possible for the inflaton potential in S-inflation to be safe from new physics or strong-coupling effects associated with the unitarity-violation scale. This contrasts with the case of Higgs Inflation, where unitarity is always violated at the scale of the inflaton.

Another interesting observation is that if the singlet mass and the coupling $\lambda_{hs}$ are sufficiently large (roughly $m_s \gtrsim 1 \text{ TeV}$ and $\lambda_{hs} \gtrsim 0.3$), the presence of the additional scalar singlet stabilises the electroweak vacuum, because the additional contribution to $\beta_{\lambda_h}$ prevents the quartic Higgs coupling from running to negative values. This observation becomes important if a metastable electroweak vacuum is physically disfavoured, for example if the potential energy relative to the absolute minimum defines an observable vacuum energy.

Given how tightly many models for DM are constrained by direct detection and LHC searches and how strong recent bounds on models for inflation have become, it is quite remarkable that one of the simplest models addressing both problems still has a large allowed parameter space. Nevertheless, the model is highly predictive. In particular, if the DM scalar is also the inflaton and unitarity is conserved during inflation, then DM is predicted to be a real scalar. Direct detection experiments will soon reach the sensitivity necessary to probe the entire parameter space relevant for phenomenology, with the exception of a small window in $m_s$ close to the Higgs resonance. The next few years will therefore likely tell us whether
indeed a singlet scalar extension of the SM can solve two of the central problems of particle physics and cosmology.

Acknowledgments

We would like to thank Rose Lerner for her contribution during the early stages of this project. We are grateful to Kyle Allison, Ido Ben-Dayan, Andreas Goudelis, Huayong Han, Thomas Konstandin, Kai Schmidt-Hoberg, Pat Scott and Christoph Weniger for helpful discussions, and to Guillermo Ballesteros for carefully reading the manuscript and providing a number of useful comments. The work of FK was supported by the German Science Foundation (DFG) under the Collaborative Research Center (SFB) 676 Particles, Strings and the Early Universe. The work of JMcD was partly supported by the Lancaster-Manchester-Sheffield Consortium for Fundamental Physics under STFC grant ST/J000418/1. FK would like to thank the Instituto de Fisica Teorica (IFT UAM-CSIC) in Madrid for its support via the Centro de Excelencia Severo Ochoa Program under Grant SEV-2012-0249, during the Program “Identification of Dark Matter with a Cross-Disciplinary Approach” where part of this work was carried out.

Appendix: RG Equations

The RG equations for the scalar couplings can be obtained using the techniques detailed in [66–68], as in [8]. The one-loop $\beta$-functions for the scalar couplings are

$$16\pi^2 \beta_{\lambda_h}^{(1)} = -6 y_t^4 + \frac{3}{8} \left(2 g^4 + (g^2 + g'^2)^2\right) + (-9 g^2 - 3 g'^2 + 12 y_t^2) \lambda_h + (18 c_h^2 + 6) \lambda_h^2 + \frac{1}{2} c_s^2 \lambda_{hs}^2 ,$$

$$16\pi^2 \beta_{\lambda_{hs}}^{(1)} = 4 c_h c_s \lambda_{hs}^2 + 6 (c_h^2 + 1) \lambda_h \lambda_{hs} - \frac{3}{2} (3 g^2 + g'^2) \lambda_{hs} + 6 y_t^2 \lambda_{hs} + 6 c_s^2 \lambda_s \lambda_{hs} ,$$

$$16\pi^2 \beta_{\lambda_s}^{(1)} = \frac{1}{2} (c_h^2 + 3) \lambda_{hs}^2 + 18 c_s^2 \lambda_s^2 .$$

The propagator suppression factors are given by

$$c_\phi = \frac{1 + \xi_\phi \phi^2}{1 + (6 \xi_\phi + 1) \frac{\xi_\phi \phi^2}{m_\phi^2}} ,$$

where $\phi$ is $s$ or $h$.

The RG equations for the non-minimal coupling can be derived following [69], as in [8] (see also [10]). One obtains

$$16\pi^2 \frac{d \xi_s}{d t} = (3 + c_h) \lambda_{hs} \left(\xi_h + \frac{1}{6}\right) + \left(\xi_s + \frac{1}{6}\right) 6 c_s \lambda_s ,$$

$$16\pi^2 \frac{d \xi_h}{d t} = \left(6 + 6 c_h\right) \lambda_h + 6 y_t^2 - \frac{3}{2} (3 g^2 + g'^2) \left(\xi_h + \frac{1}{6}\right) + \left(\xi_s + \frac{1}{6}\right) c_s \lambda_{hs} .$$
References


