UC Santa Cruz
UC Santa Cruz Previously Published Works

Title
Cosmological implications of the KOTO excess

Permalink

Journal

ISSN

Authors
Altmannshofer, Wolfgang
Lehmann, Benjamin V
Profumo, Stefano

Publication Date

DOI

Cosmological implications of the KOTO excess

Wolfgang Altmannshofer,[1, *] Benjamin V. Lehmann,[1, †] and Stefano Profumo[1, ‡]

[1]*Department of Physics, University of California Santa Cruz,*

1156 High St., Santa Cruz, CA 95064, USA and

Santa Cruz Institute for Particle Physics,

1156 High St., Santa Cruz, CA 95064, USA

The KOTO experiment has reported an excess of $K_L \to \pi^0 \nu \bar{\nu}$ events above the standard model prediction, in tension with the Grossman–Nir bound. The GN bound heavily constrains new physics interpretations of an excess in this channel, but another possibility is that the observed events originate from a different process entirely: a decay of the form $K_L \to \pi^0 X$, where X denotes one or more new invisible species. We introduce a class of models to study this scenario with two light scalars playing the role of X, and we examine the possibility that the lighter of the two new states may also account for cosmological dark matter. We show that this species can be produced thermally in the presence of additional interactions apart from those needed to account for the KOTO excess. Conversely, in the minimal version of the model, dark matter must be produced non-thermally. In this case, avoiding overproduction imposes constraints on the structure of the low-energy theory. Moreover, this requirement carries significant implications for the scale of reheating in the early universe, generically preferring a low but observationally-permitted reheating temperature of $\mathcal{O}(10\,\mathrm{MeV})$. We discuss astrophysical and terrestrial signatures that will allow further tests of this paradigm in the coming years.

---

* waltmann@ucsc.edu

† blehmann@ucsc.edu

‡ profumo@ucsc.edu

CONTENTS

I. INTRODUCTION

The rare kaon decays $K^+ \to \pi^+ \nu\bar\nu$ and $K_L \to \pi^0 \nu\bar\nu$ are widely recognized as very sensitive probes of new physics (NP). In the Standard Model (SM), the branching ratios of these decays are strongly suppressed, and can be precisely predicted [12, 14] to be

$$\text{BR}(K^+ \to \pi^+ \nu\bar\nu)_{\text{SM}} = (8.4 \pm 1.0) \times 10^{-11} , \tag{1}$$

$$\text{BR}(K_L \to \pi^0 \nu\bar\nu)_{\text{SM}} = (3.4 \pm 0.6) \times 10^{-11} . \tag{2}$$

On the experimental side, several $K^+ \to \pi^+ \nu\bar\nu$ candidate events have been observed by the E787/E949 experiment [2, 5, 6] and the NA62 experiment [18], but a discovery of $K^+ \to \pi^+ \nu\bar\nu$ has still to be established. The current best limit on the branching ratio is from a preliminary analysis of NA62 data and reads [47]

$$\text{BR}(K^+ \to \pi^+ \nu\bar\nu)_{\text{exp}} < 2.44 \times 10^{-10} \quad (95\% \text{ C.L.}), \tag{3}$$

not far above the SM prediction. The NA62 experiment aims to measure the SM branching ratio with $\mathcal{O}(10\%)$ uncertainty. In the case of $K_L \to \pi^0 \nu\bar\nu$, the current most stringent bound on the branching ratio comes from the KOTO experiment [3], and is still two orders of magnitude above the SM prediction:

$$\text{BR}(K_L \to \pi^0 \nu\bar\nu)_{\text{exp}} < 3.0 \times 10^{-9} \quad (90\% \text{ C.L.}). \tag{4}$$

Interestingly, in the latest status update by KOTO [49], 4 events are seen in the signal box, with an expected number of 0.05 ± 0.01 SM $K_L \to \pi^0 \nu\bar\nu$ events and 0.05 ± 0.02 background events. One of the events has been identified as likely background. If the remaining events are interpreted as signal, one finds a branching ratio of $\text{BR}(K_L \to \pi^0 \nu\bar\nu) \sim 2 \times 10^{-9}$ [40]. A branching ratio of this size would be a spectacular discovery. Not only does it imply NP, it also violates the Grossman-Nir (GN) bound [31], $\text{BR}(K_L \to \pi^0 \nu\bar\nu) \lesssim 4.3 \times \text{BR}(K^+ \to \pi^+ \nu\bar\nu) \lesssim 10^{-9}$, when combined with the NA62 constraint in eq. (3). The GN bound is very robust in models where the $K \to \pi \nu\bar\nu$ decays are modified by heavy new physics well above the kaon mass. However, in the presence of light new physics, the GN bound can be violated and the observed events at KOTO may find an explanation [19, 21, 23, 25, 30, 36–38, 40–42, 52].

Here we focus on a new physics scenario first discussed in [45]. Two new light scalars S and P, neutral under the SM gauge interactions, are introduced such that K_L can decay
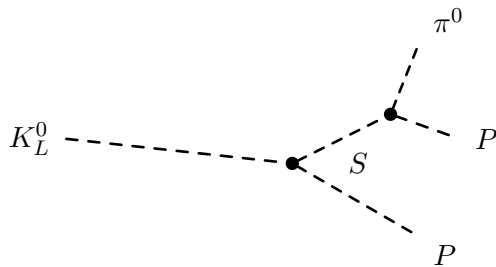
FIG. 1. Decay chain accounting for the KOTO signal in our scenario.

into a pair of the new particles, $K_L \to SP$. If the decay $S \to \pi^0 P$ is allowed and P is stable on the relevant experimental scales, then the decay chain $K_L \to SP \to \pi^0 PP$ can mimic the $K_L \to \pi^0 \nu \bar{\nu}$ signature (see fig. 1). The corresponding chain of two-body decays does not exist for the charged kaon. A possible decay $K^+ \to \pi^+ SP$ is suppressed by three-body phase space or may be forbidden entirely by kinematics.

If P is absolutely stable, it is also a candidate for cosmological dark matter. In the minimal setup that can provide a NP explanation of the KOTO events, P couples to the SM very weakly, implying that annihilation cross sections into SM states are too small for production by freeze-out. We therefore investigate alternative scenarios for cosmological production, and interpret overproduction of P as a cosmological constraint on the structure of the low-energy theory. We show that P is readily produced non-thermally if the scale of reheating is low, close to but safely above the current observational bound. We also show that this class of models can account for the KOTO excess without requiring a low reheating temperature, but only in the presence of additional interactions. We investigate prospects for testing this model with future experiments and with additional data from KOTO, and show that much of the parameter space will be probed in the near future.

This paper is organized as follows: in section II, we present the model and discuss how it can explain the KOTO events. In section III, we evaluate astrophysical and terrestrial constraints on the parameter space of our model. In section IV, we consider cosmological production of P, and relate the production of P to the scale of reheating. We discuss the implications of our results in section V and conclude in section VI.

II. MODEL

We start with very simple kinematical considerations concerning the masses of the two scalars S and P. Figure 2 shows the plane of the two scalar masses m_S and m_P. As described in the introduction, we are interested in regions of parameter space where the decay $K_L \to \pi^0 PP$, which mimics $K_L \to \pi^0 \nu\bar{\nu}$, can be realized as a sequence of the two-body decay $K_L \to SP$ followed by $S \to \pi^0 P$. For m_S too large, the decay $K_L \to SP$ is kinematically forbidden, while for m_S too small, the $S \to \pi^0 P$ decay is not open, excluding the dark gray regions in the plot. In the light gray region one faces potential constraints from the charged kaon decay $K^+ \to \pi^+ SP$ that is generically expected in the models discussed below. In the white region, however, this decay is kinematically forbidden, while $K_L \to \pi^0 \nu\bar{\nu}$ remains open.

The plot also indicates two other interesting kinematical boundaries. If $m_P < m_{\pi^0}/2$, the exotic pion decay $\pi^0 \to PP$ is possible which, as we will discuss in section IV, can impact cosmological production considerably. If $m_S > 3m_P$, the decay $S \to 3P$ can be allowed, thus modifying the lifetime of S, which is a crucial parameter for beam dump constraints. Note that low P masses may be subject to constraints from supernova cooling, which we will discuss further in section III A. A weaker lower bound on the P mass also follows from assuming a particular thermal history, a point to which we shall return in section V.

In the following sections, we will discuss four benchmark parameter points covering the most interesting regimes:

$$
\begin{aligned}
\text{BM1:} \quad & m_S = 400\,\text{MeV}, \quad m_P = 10\,\text{MeV}, \\
\text{BM2:} \quad & m_S = 350\,\text{MeV}, \quad m_P = 100\,\text{MeV}, \\
\text{BM3:} \quad & m_S = 300\,\text{MeV}, \quad m_P = 125\,\text{MeV}, \\
\text{BM4:} \quad & m_S = 200\,\text{MeV}, \quad m_P = 10\,\text{MeV}.
\end{aligned}
\tag{5}
$$

Next we discuss in detail the interactions of S and P with SM quarks. We first focus on non-renormalizable effective couplings and identify viable regions of parameter space. Then we comment on simplified UV models that map onto the effective couplings.
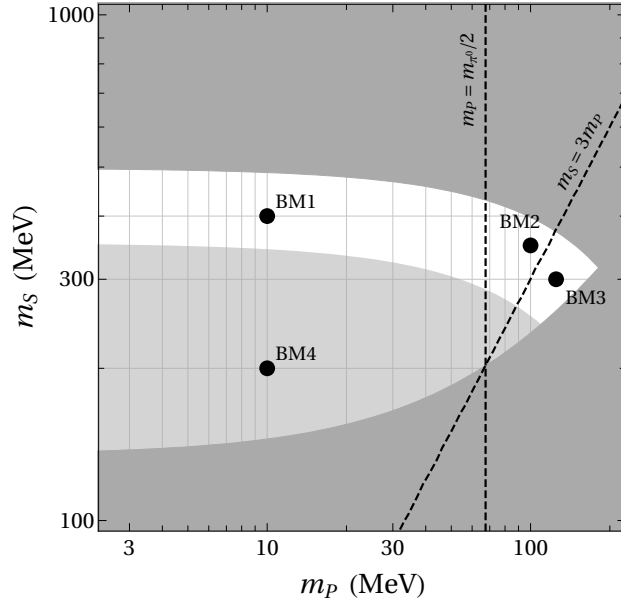
FIG. 2. The plane of the scalar masses m_S vs. m_P. In the dark gray region the $K_L \to \pi^0 PP$ decay cannot be realized as a sequence of 2-body decays. In the light gray region the $K^+ \to \pi^+ SP$ decay is open. The black dots indicate four benchmark scenarios that we consider later (eq. (5)).

A.   Effective interactions of the scalars and meson decay rates

We assume that the scalars S and P interact with SM particles via the effective couplings

$$
\mathcal{L}_{\text{int}} \supset iSP \left(\frac{g_{dd}^{SP}}{\Lambda_{\text{NP}}} (\bar{d}d) + \frac{\tilde{g}_{dd}^{SP}}{\Lambda_{\text{NP}}} (\bar{d}i\gamma_5 d) + \frac{g_{ss}^{SP}}{\Lambda_{\text{NP}}} (\bar{s}s) + \frac{\tilde{g}_{ss}^{SP}}{\Lambda_{\text{NP}}} (\bar{s}i\gamma_5 s) \right)
$$
$$
+ iSP \left(\frac{g_{sd}^{SP}}{\Lambda_{\text{NP}}} (\bar{s}d) + \frac{\tilde{g}_{sd}^{SP}}{\Lambda_{\text{NP}}} (\bar{s}i\gamma_5 d) + \text{h.c.} \right). \quad (6)
$$

The factors of i in the above Lagrangian are reminiscent of considering S to be a CP-even scalar and P to be a CP-odd pseudoscalar, a notational pattern that we will retain when matching onto low-energy QCD later on. The coefficients g_{dd}^{SP}, g_{ss}^{SP}, $\tilde{g}_{dd}^{SP}$, and $\tilde{g}_{ss}^{SP}$ are purely imaginary (by hermiticity of the Lagrangian) while the g_{sd}^{SP} and $\tilde{g}_{sd}^{SP}$ coefficients can have an arbitrary complex phase. There could also be interactions involving b quarks, but as long as they are not considerably larger than the interactions with the light quarks, their impact on phenomenology will be negligible.

In the following, we will also entertain the possibility of additional interactions involving

P^2 and S^2, of the form

$$\mathcal{L}_{\text{int}} \supset P^2 \left(\frac{g_{dd}^{P^2}}{\Lambda_{\text{NP}}}(\bar{d}d) + \frac{\tilde{g}_{dd}^{P^2}}{\Lambda_{\text{NP}}}(\bar{d}i\gamma_5 d) + \frac{g_{ss}^{P^2}}{\Lambda_{\text{NP}}}(\bar{s}s) + \frac{\tilde{g}_{ss}^{P^2}}{\Lambda_{\text{NP}}}(\bar{s}i\gamma_5 s) \right)$$
$$+ P^2 \left(\frac{g_{sd}^{P^2}}{\Lambda_{\text{NP}}}(\bar{s}d) + \frac{\tilde{g}_{sd}^{P^2}}{\Lambda_{\text{NP}}}(\bar{s}i\gamma_5 d) + \text{h.c.} \right). \quad (7)$$

While the interactions in eq. (7) are not directly relevant for the KOTO signal, they do have important implications for other meson decays and in particular for the dark matter phenomenology as we will discuss in section IV below.

The decays relevant for an enhanced KOTO signal, $K_L \to SP$ and $S \to \pi^0 P$ are induced by the couplings $\text{Re}(\tilde{g}_{sd}^{SP})$ and $\text{Im}(\tilde{g}_{dd}^{SP})$, respectively. For the corresponding decay rates we find

$$\Gamma(K_L \to SP) = \frac{1}{8\pi} \frac{f_K^2 m_{K_L}^3}{m_s^2} \left(\frac{\text{Re}(\tilde{g}_{sd}^{SP})}{\Lambda_{\text{NP}}} \right)^2 \eta_{\text{QCD}} \sqrt{\lambda\left(1, m_S^2/m_{K_L}^2, m_P^2/m_{K_L}^2\right)}, \quad (8)$$

$$\Gamma(S \to \pi^0 P) = \frac{1}{128\pi} \frac{f_\pi^2 m_{\pi^0}^4}{m_S m_d^2} \left(\frac{\text{Im}(\tilde{g}_{dd}^{SP})}{\Lambda_{\text{NP}}} \right)^2 \eta_{\text{QCD}} \sqrt{\lambda\left(1, m_{\pi^0}^2/m_S^2, m_P^2/m_S^2\right)}, \quad (9)$$

with the phase space function $\lambda(a, b, c) = a^2 + b^2 + c^2 - 2(ab + ac + bc)$. The down and strange quark masses in the above expressions should be interpreted as the $\overline{\text{MS}}$ masses at a renormalization scale of $\mu = 2$ GeV. Leading-log QCD corrections are then taken into account through the factor η_{QCD}

$$\eta_{\text{QCD}} = \left(\frac{\alpha_s(m_t)}{\alpha_s(M)} \right)^{8/7} \left(\frac{\alpha_s(m_b)}{\alpha_s(m_t)} \right)^{24/23} \left(\frac{\alpha_s(2 \text{ GeV})}{\alpha_s(m_b)} \right)^{24/25}, \quad (10)$$

where M is the scale of new physics that is responsible for the effective interactions of S and P with the SM quarks. Because of $SU(2)_L$ invariance we expect $M \sim \sqrt{\Lambda_{\text{NP}} v}$, where $v = 246$ GeV is the vacuum expectation value of the SM Higgs. Note that including the η_{QCD} factor is equivalent to evaluating the down and strange masses in eqs. (8) and (9) at the scale M.

The coupling $|g_{sd}^{SP}|$ can lead to the decay $K^+ \to \pi^+ SP$, if kinematically allowed. The differential 3-body decay rate of $K^+ \to \pi^+ SP$ is given by

$$\frac{d\Gamma(K^+ \to \pi^+ SP)}{dq^2} = \frac{1}{256\pi^3} \frac{m_{K^+}^3}{m_s^2} \left(\frac{|g_{sd}^{SP}|}{\Lambda_{\text{NP}}} \right)^2 \eta_{\text{QCD}} \left(1 - \frac{m_{\pi^+}^2}{m_{K^+}^2} \right)^2$$
$$\times \sqrt{\lambda\left(1, m_S^2/q^2, m_P^2/q^2\right)} \sqrt{\lambda\left(1, m_{\pi^+}^2/m_{K^+}^2, q^2/m_{K^+}^2\right)}, \quad (11)$$

where we estimated the relevant scalar form factor as $\langle \pi^+ | \bar{s} d | K^+ \rangle \simeq (m_{K^+}^2 - m_{\pi^+}^2)/m_s$ and q^2 is the invariant mass of the SP system, with $(m_P + m_S)^2 < q^2 < (m_{K^+} - m_{\pi^+})^2$.

Similarly to the $K_L \to SP$ decay, the interactions in eq. (6) also lead to the exotic eta decay $\eta \to SP$, which has been identified as a possible source of the scalar S at beam dump experiments [37]. Neglecting η–η' mixing, we find

$$\Gamma(\eta \to SP) = \frac{3}{512\pi} \frac{f_\eta^2 m_\eta^3}{m_s^2} \left(\frac{2 \operatorname{Im}(\tilde{g}_{ss}^{SP}) - \operatorname{Im}(\tilde{g}_{dd}^{SP})}{\Lambda_{\mathrm{NP}}} \right)^2 \eta_{\mathrm{QCD}} \sqrt{\lambda\left(1, m_S^2/m_\eta^2, m_P^2/m_\eta^2\right)}. \tag{12}$$

For completeness, we also provide the expression for the decay $K_S \to SP$:

$$\Gamma(K_S \to SP) = \frac{1}{32\pi} \frac{f_K^2 m_{K_S}^3}{m_s^2} \left(\frac{\operatorname{Im}(\tilde{g}_{sd}^{SP})}{\Lambda_{\mathrm{NP}}} \right)^2 \eta_{\mathrm{QCD}} \sqrt{\lambda\left(1, m_S^2/m_{K_S}^2, m_P^2/m_{K_S}^2\right)}. \tag{13}$$

In the presence of the P^2 interactions in eq. (7), there are additional exotic meson decays, $\pi^0 \to PP$, $\eta \to PP$, $K_{L/S} \to PP$, and $K^+ \to \pi^+ PP$, with the following decay rates:

$$\Gamma(\pi^0 \to PP) = \frac{1}{64\pi} \frac{f_\pi^2 m_{\pi^0}^3}{m_d^2} \left(\frac{\operatorname{Re}(\tilde{g}_{dd}^{P^2})}{\Lambda_{\mathrm{NP}}} \right)^2 \eta_{\mathrm{QCD}} \sqrt{1 - \frac{4m_P^2}{m_{\pi^0}^2}}, \tag{14}$$

$$\Gamma(\eta \to PP) = \frac{3}{256\pi} \frac{f_\eta^2 m_\eta^3}{m_s^2} \left(\frac{2 \operatorname{Re}(\tilde{g}_{ss}^{P^2}) - \operatorname{Re}(\tilde{g}_{dd}^{P^2})}{\Lambda_{\mathrm{NP}}} \right)^2 \eta_{\mathrm{QCD}} \sqrt{1 - \frac{4m_P^2}{m_\eta^2}}, \tag{15}$$

$$\Gamma(K_L \to PP) = \frac{1}{4\pi} \frac{f_K^2 m_{K_L}^3}{m_s^2} \left(\frac{\operatorname{Im}(\tilde{g}_{sd}^{P^2})}{\Lambda_{\mathrm{NP}}} \right)^2 \eta_{\mathrm{QCD}} \sqrt{1 - \frac{4m_P^2}{m_{K_L}^2}}, \tag{16}$$

$$\Gamma(K_S \to PP) = \frac{1}{4\pi} \frac{f_K^2 m_{K_S}^3}{m_s^2} \left(\frac{\operatorname{Re}(\tilde{g}_{sd}^{P^2})}{\Lambda_{\mathrm{NP}}} \right)^2 \eta_{\mathrm{QCD}} \sqrt{1 - \frac{4m_P^2}{m_{K_S}^2}}, \tag{17}$$

$$\frac{\mathrm{d}\Gamma(K^+ \to \pi^+ PP)}{\mathrm{d}q^2} = \frac{1}{128\pi^3} \frac{m_{K^+}^3}{m_s^2} \left(\frac{|g_{sd}^{P^2}|}{\Lambda_{\mathrm{NP}}} \right)^2 \eta_{\mathrm{QCD}} \left(1 - \frac{m_{\pi^+}^2}{m_{K^+}^2} \right)^2$$
$$\times \sqrt{1 - \frac{4m_P^2}{q^2}} \sqrt{\lambda\left(1, m_{\pi^+}^2/m_{K^+}^2, q^2/m_{K^+}^2\right)}, \tag{18}$$

In the $K^+ \to \pi^+ PP$ decay width, q^2 denotes the PP invariant mass, which lies in the range $4m_P^2 < q^2 < (m_{K^+} - m_{\pi^+})^2$.

The interactions of S and P with quarks that we have introduced preserve a Z_2 symmetry under which S and P are odd, while all SM particles are even. We assume that the Z_2 symmetry is also respected by the scalar potential, such that P is an absolutely stable dark matter candidate. Among the allowed Z_2 symmetric terms in the scalar potential, the SP^3 interaction

$$\mathcal{L}_{\mathrm{int}} \supset \lambda_{SP^3} SP^3, \tag{19}$$

will turn out to be relevant. When kinematically allowed, this interaction leads to the decay $S \to 3P$ with rate

$$\Gamma(S \to 3P) = \frac{3}{256\pi^3}\lambda_{SP^3}^2 m_S \; f(m_P/m_S), \tag{20}$$

where f is the three-body phase space integral,

$$f(y) = 2 \int_{4y^2}^{(1-y)^2} \mathrm{d}x \; \sqrt{\lambda\left(1, x, y^2\right)\lambda\left(1, y^2/x, y^2/x\right)}, \tag{21}$$

which is normalized to 1 in the limit $y \to 0$. The $S \to 3P$ rate will modify the lifetime of S and can therefore have a crucial impact on possible constraints from beam dump experiments.

B. Events at the KOTO experiment

The model introduced in the previous section will lead to $K_L \to \pi^0 PP$ events at the KOTO experiment. We now identify the regions of parameter space in which this decay can mimic the KOTO signal.

The number of events that can be expected to be detected at KOTO can be written as

$$N = \frac{\mathrm{BR}(K_L \to SP) \times \mathrm{BR}(S \to \pi^0 P)}{\mathrm{BR}(K_L \to \pi^0\nu\bar{\nu})_{\mathrm{SM}}} \times R \times N_{\mathrm{SM}}, \tag{22}$$

where $\mathrm{BR}(K_L \to \pi^0\nu\bar{\nu})_{\mathrm{SM}} = (3.4 \pm 0.6) \times 10^{-11}$ is the SM prediction for the $K_L \to \pi^0\nu\bar{\nu}$ branching ratio [12, 14], $N_{\mathrm{SM}} = 0.05 \pm 0.01$ is the expected number of SM signal events at KOTO [49], and

$$R = \frac{A(K_L \to SP \to \pi^0 PP)}{A(K_L \to \pi^0\nu\bar{\nu})} \tag{23}$$

is the ratio of acceptances of the considered model signal and the SM signal at the KOTO detector. As has been pointed out before [30, 37, 40], an exotic contribution to the KOTO signal (in our case $K_L \to SP \to \pi^0 PP$) can have a considerably different acceptance. We determine the acceptance ratio R using a Monte Carlo simulation. Details are provided in appendix A. The result is given in fig. 3, which shows R as a function of the S lifetime for our four benchmark points (eq. (5)). For prompt decays, $\tau_S \to 0$, we find $\{R_{\mathrm{BM1}}, R_{\mathrm{BM2}}, R_{\mathrm{BM3}}, R_{\mathrm{BM4}}\} \simeq \{102\%, 51\%, 10\%, 73\%\}$. Once the lifetime of S becomes comparable to the size of the KOTO detector, $\tau_S \sim 1\,\mathrm{m}$, R starts to decrease as more and more S leave the detector before decaying.
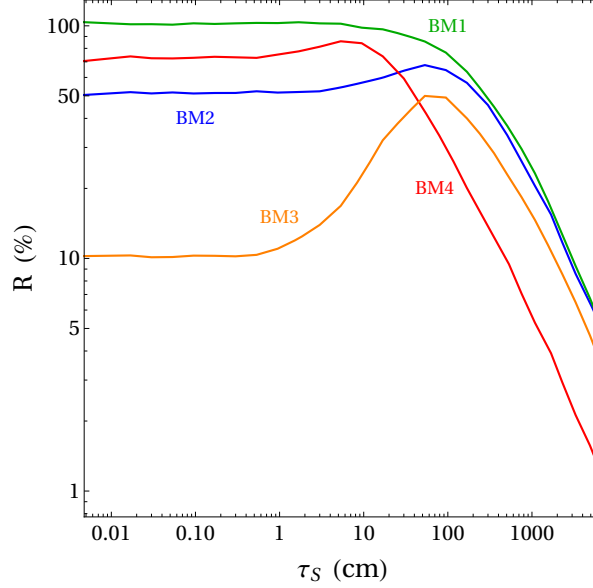
FIG. 3. The acceptance ratio R of the $K_L \to SP \to \pi^0 PP$ signal over the SM $K_L \to \pi^0 \nu \bar\nu$ signal at KOTO as a function of the S lifetime τ_S for the four benchmark scenarios.

In our setup, the lifetime of S is determined by the $S \to \pi^0 P$ and $S \to 3P$ decays. In the four benchmark cases for the scalar masses defined above we find

$$\left\{ \Gamma(S \to \pi^0 P)_{\mathrm{BM1}}, \Gamma(S \to \pi^0 P)_{\mathrm{BM2}}, \Gamma(S \to \pi^0 P)_{\mathrm{BM3}}, \Gamma(S \to \pi^0 P)_{\mathrm{BM4}} \right\} \simeq$$
$$\left\{ \frac{1}{3.3\,\mathrm{cm}}, \frac{1}{3.4\,\mathrm{cm}}, \frac{1}{4.4\,\mathrm{cm}}, \frac{1}{2.7\,\mathrm{cm}} \right\} \times \left(\frac{10^6\ \mathrm{GeV}}{\Lambda_{dd}} \right)^2 \left(\frac{\alpha_s(10^4\ \mathrm{GeV})}{\alpha_s(M)} \right)^{8/7}, \quad (24)$$

$$\left\{ \Gamma(S \to 3P)_{\mathrm{BM1}}, \Gamma(S \to 3P)_{\mathrm{BM2}}, \Gamma(S \to 3P)_{\mathrm{BM4}} \right\} \simeq$$
$$\left\{ \frac{1}{2.0\,\mathrm{cm}}, \frac{1}{49\,\mathrm{cm}}, \frac{1}{4.3\,\mathrm{cm}} \right\} \times \left(\frac{\lambda_{SP3}}{10^{-5}} \right)^2, \quad (25)$$

where in the $S \to \pi^0 P$ decay width we have defined $\Lambda_{dd} = \Lambda_{\mathrm{NP}} / \mathrm{Im}(\tilde{g}_{dd}^{SP})$. Note that $S \to 3P$ is not kinematically allowed in benchmark BM3. The $S \to \pi^0 P$ branching ratio is given by $\mathrm{BR}(S \to \pi^0 P) = \Gamma(S \to \pi^0 P) / [\Gamma(S \to \pi^0 P) + \Gamma(S \to 3P)]$.

Finally, we find the following $K_L \to SP$ branching ratios

$$\left\{ \mathrm{BR}(K_L \to SP)_{\mathrm{BM1}}, \mathrm{BR}(K_L \to SP)_{\mathrm{BM2}}, \mathrm{BR}(K_L \to SP)_{\mathrm{BM3}}, \mathrm{BR}(K_L \to SP)_{\mathrm{BM4}} \right\} \simeq$$
$$\left\{ 1.7,\ 1.8,\ 2.3,\ 4.0 \right\} \times 10^{-9} \times \left(\frac{10^{12}\ \mathrm{GeV}}{\Lambda_{sd}} \right)^2 \left(\frac{\alpha_s(10^4\,\mathrm{GeV})}{\alpha_s(M)} \right)^{8/7}, \quad (26)$$

where we have defined $\Lambda_{sd} = \Lambda_{\mathrm{NP}}/\operatorname{Re}(\tilde{g}_{sd}^{SP})$.

Figures 4 and 5 show the number of expected events in the Λ_{sd}–Λ_{dd} plane for our benchmark cases in the absence of the $S \to 3P$ decay (fig. 4) and in the presence of the $S \to 3P$ decay induced by a coupling $\lambda_{SP3} = 10^{-5}$ (fig. 5). Along the solid green lines one expects 3 events, in the dark green regions one expects 2–4 events, and in the light green regions one expects 1–5 events. In the gray regions labeled "$K_L \to \pi^0$ inv.", the number of predicted events exceeds the limit from KOTO (see eq. (4)). The right vertical axis shows the lifetime of S corresponding to Λ_{dd}. In fig. 5, the lifetime is approximately constant for $\Lambda_{dd} > 10^7\,\mathrm{GeV}$, as in this region of parameter space, the lifetime is set by the $S \to 3P$ decay width.

For S lifetimes of $\tau_S \gtrsim 1\,\mathrm{m}$, existing beam dump constraints apply (see section III) as indicated in fig. 4 by the dashed contours. A proposed upgrade of the SeaQuest experiment might probe S lifetimes as low as $\tau_S \gtrsim 5\,\mathrm{cm}$. In the scenarios shown in fig. 5 with $\lambda_{SP3} = 10^{-5}$, the S lifetimes are short enough throughout the parameter space that existing beam dump constraints are avoided.

In fig. 5 we also show additional constraints from other meson decays. The known K_L branching fractions add up to a value compatible with 1 with very high precision. Any additional K_L branching ratio, in particular $K_L \to SP$, is thus bounded above as $\mathrm{BR}(K_L \to SP) < 6.3 \times 10^{-4}$ [28]. In fig. 5 the gray regions left of the dashed vertical lines denoted "$K_L \to$ inv." are excluded by this constraint. Note that this gives an absolute lower bound $\Lambda_{sd} \gtrsim \mathrm{few} \times 10^9\,\mathrm{GeV}$.

The other meson decay constraints shown in fig. 5 are less robust as they depend on couplings that are in principle unrelated. If we assume that the coupling g_{sd}^{P2} (corresponding to $(\bar{s}d)P^2$) is of the same order as the coupling $\tilde{g}_{sd}^{SP}$ (corresponding to $(\bar{s}i\gamma_5 d)SP$), we find relevant constraints from the searches for $K^+ \to \pi^+ \nu\bar{\nu}$. To evaluate the constraints we compare the predicted $K^+ \to \pi^+ PP$ branching ratio with the bound from NA62 given in eq. (3). We correct for the different signal acceptances of $K^+ \to \pi^+ PP$ compared to $K^+ \to \pi^+ \nu\bar{\nu}$ that arise due to kinematical cuts on the missing mass and the charged pion momentum. For the three P masses relevant to our benchmarks, we find the bounds $\mathrm{BR}(K^+ \to \pi^+ PP) < 2.7 \times 10^{-10}$ for $m_P = 10\,\mathrm{MeV}$, $\mathrm{BR}(K^+ \to \pi^+ PP) < 3.5 \times 10^{-10}$ for $m_P = 100\,\mathrm{MeV}$, and $\mathrm{BR}(K^+ \to \pi^+ PP) < 2.4 \times 10^{-9}$ for $m_P = 125\,\mathrm{MeV}$. Setting $\Lambda_{\mathrm{NP}}/|g_{sd}^{P2}| = \Lambda_{\mathrm{NP}}/\operatorname{Re}(\tilde{g}_{sd}^{SP}) = \Lambda_{sd}$, we find that in fig. 5, the regions left of the dotted
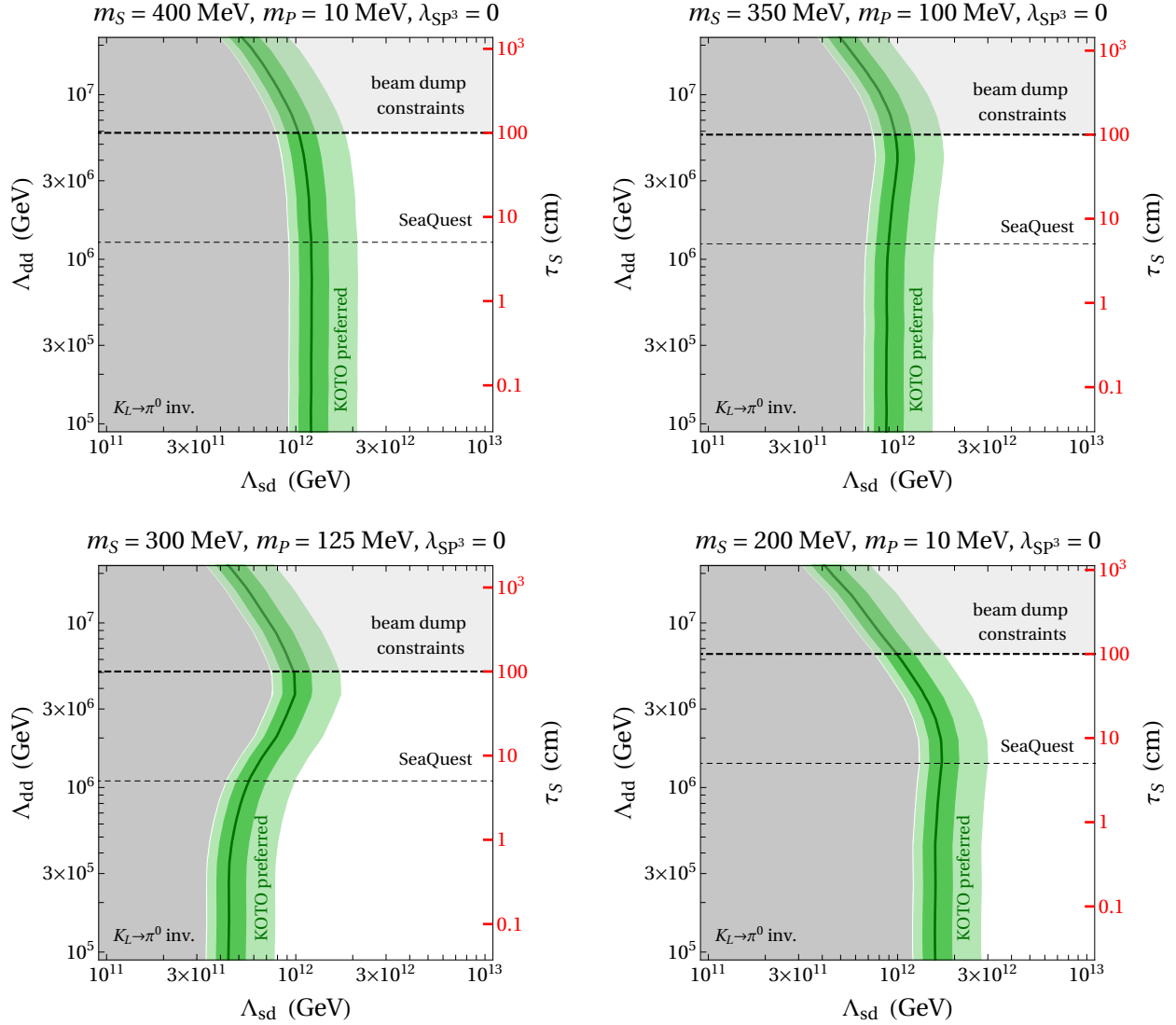
FIG. 4. Number of expected $K_L \to SP \to \pi PP$ events at KOTO in the Λ_{sd}–Λ_{dd} plane for four benchmark points of the S and P masses. The SP^3 coupling is set to zero. The right vertical axis indicates the S lifetime. One expects 3 events along the solid dark green line, 2–4 events in the dark green region, and 1–5 events in the light green region. In the gray regions labeled "$K_L \to \pi^0$ inv.", the number of predicted events exceeds the limit from KOTO. The dashed lines show constraints from existing beam dump experiments and the potential reach of the SeaQuest upgrade. The

vertical lines are excluded.

If we assume that the coupling $\tilde{g}_{dd}^{P^2}$ (corresponding to $(\bar{d} i \gamma_5 d) P^2$) is of the same order as the coupling $\tilde{g}_{dd}^{SP}$ (corresponding to $(\bar{d} i \gamma_5 d) SP$), we find relevant constraints from the invisible branching fraction of the neutral pion, $\mathrm{BR}(\pi^0 \to \mathrm{inv.}) < 4.4 \times 10^{-9}$ [47]. Setting
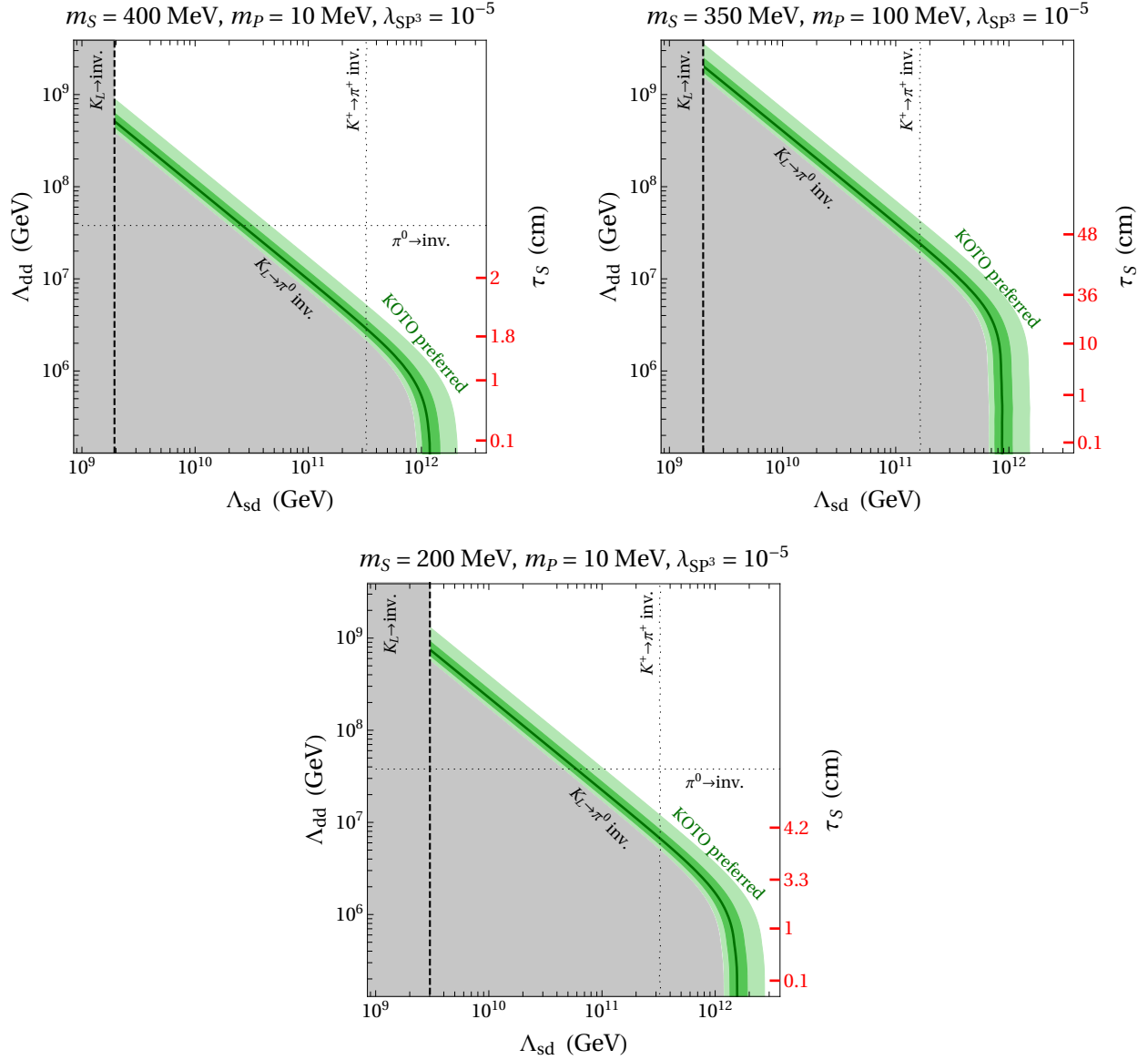
FIG. 5. Number of expected $K_L \to SP \to \pi PP$ events at KOTO in the Λ_{sd}–Λ_{dd} plane for three benchmark points of the S and P masses. The SP^3 coupling is set to $\lambda_{SP^3} = 10^{-5}$. The right vertical axis indicates the S lifetime, which is approximately constant for $\Lambda_{dd} > 10^7\,\mathrm{GeV}$. One expects 3 events along the solid dark green line, 2–4 events in the dark green region, and 1–5 events in the light green region. The gray regions are excluded by the KOTO limit on $K_L \to \pi^0$ inv. or the bound on the invisible K_L branching ratio. The dotted lines show the generic location of other constraints that depend on additional model parameters. Benchmark BM3 is not shown, as the $S \to 3P$ decay is kinematically forbidden.

$\Lambda_{\rm NP}/\operatorname{Re}(\tilde{g}_{dd}^{P2}) = \Lambda_{\rm NP}/\operatorname{Im}(\tilde{g}_{dd}^{SP}) = \Lambda_{dd}$ in the benchmarks BM1 and BM4, the regions below the dotted horizontal lines are excluded. For benchmarks BM2 and BM3, the P mass is too large for the $\pi^0 \to PP$ decay, so the couplings are therefore completely unconstrained by $\mathrm{BR}(\pi^0 \to \mathrm{inv.})$.

C. Simplified UV models

The higher dimensional interactions in eq. (6) that lead to the exotic meson decays can be UV completed by simplified models in various ways. In this section, we discuss briefly two possibilities: (1) vector-like quarks and (2) an inert Higgs doublet.

1. Vector-like quark model

We introduce two sets of heavy vector-like quarks D and Q which have quantum numbers of the right-handed down quark singlets, $D = (\mathbf{3}, \mathbf{1})_{-\frac{1}{3}}$, and of the left-handed quark doublets $Q = (\mathbf{3}, \mathbf{2})_{\frac{1}{6}}$, respectively. These quantum number assignments admit the following terms in the Lagrangian:

$$\begin{aligned}
\mathcal{L} \supset \ & m_Q \bar{Q}_L Q_R + m_D \bar{D}_L D_R + Y_{QD}(\bar{Q}_L D_R)h + Y_{DQ}(\bar{D}_L Q_R)h^c \ + \ \mathrm{h.c.} \\
& + X_{Dd}(\bar{D}_L d_R)S + X_{Ds}(\bar{D}_L s_R)S + Z_{Qd}(\bar{Q}_R d_L)iP + Z_{Ds}(\bar{Q}_R s_L)iP \ + \ \mathrm{h.c.} \ .
\end{aligned} \tag{27}$$

The first line contains the masses m_Q and m_D for the vector-like quarks, as well as interactions with the SM Higgs doublet h. The masses m_Q, m_D and the couplings Y_{QD}, Y_{DQ} are in general complex parameters. However, not all of their phases are observable. Using the freedom to re-phase the vector-like quark fields, we will choose real m_Q, m_D and Y_{QD} without loss of generality. The second line in eq. (27) contains couplings of the SM down and strange quarks with S and the vector-like quark D as well as with P and the vectorlike quark Q. The couplings X_{Dd}, X_{Ds}, Z_{Qd}, and Z_{Qs} contain physical phases.

Note that the above Lagrangian is invariant under a Z_2 symmetry under which all SM particles are even, while the vector-like quarks as well as S and P are odd. Thus P remains an absolutely stable dark matter candidate. In addition to the couplings shown, the model could also contain Z_2 invariant couplings involving S and Q or P and D. However, such
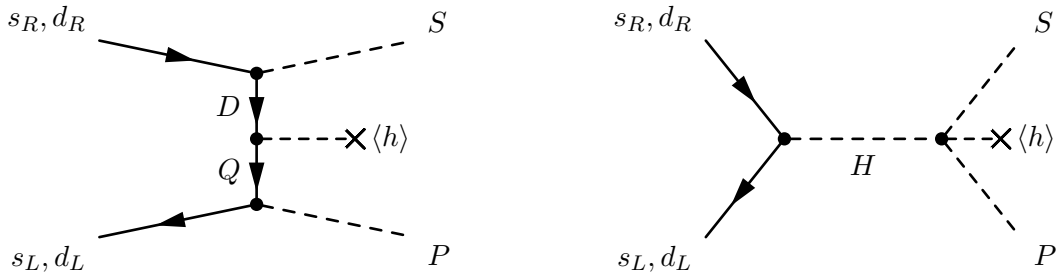
FIG. 6. Feynman diagrams that show the matching of the vector-like quark model (left) and the inert Higgs model (right) onto the effective $SPqq'$ interactions in eq. (6).

couplings are not required to generate the desired low energy interactions and we will neglect them in the following.

Integrating out the vector-like quarks at tree level (see fig. 6, left diagram), and matching onto the effective Lagrangian of eq. (6), we find

$$
\frac{g_{dd}^{SP}}{\Lambda_{\mathrm{NP}}} = \frac{-iY_{QD}v}{\sqrt{2}m_Q m_D}\,\mathrm{Im}(X_{Dd}Z_{Qd}^*)\,, \qquad \frac{\tilde{g}_{dd}^{SP}}{\Lambda_{\mathrm{NP}}} = \frac{iY_{QD}v}{\sqrt{2}m_Q m_D}\,\mathrm{Re}(X_{Dd}Z_{Qd}^*)\,, \qquad (28)
$$

$$
\frac{g_{ss}^{SP}}{\Lambda_{\mathrm{NP}}} = \frac{-iY_{QD}v}{\sqrt{2}m_Q m_D}\,\mathrm{Im}(X_{Ds}Z_{Qs}^*)\,, \qquad \frac{\tilde{g}_{ss}^{SP}}{\Lambda_{\mathrm{NP}}} = \frac{iY_{QD}v}{\sqrt{2}m_Q m_D}\,\mathrm{Re}(X_{Ds}Z_{Qs}^*)\,,
$$

$$
\frac{g_{sd}^{SP}}{\Lambda_{\mathrm{NP}}} = \frac{Y_{QD}v}{\sqrt{2}m_Q m_D}\frac{1}{2}(Z_{Qs}X_{Dd}^* - X_{Ds}Z_{Qd}^*)\,, \qquad \frac{\tilde{g}_{sd}^{SP}}{\Lambda_{\mathrm{NP}}} = \frac{Y_{QD}v}{\sqrt{2}m_Q m_D}\frac{i}{2}(Z_{Qs}X_{Dd}^* + X_{Ds}Z_{Qd}^*)\,.
$$

As required by $SU(2)_L$ invariance, the effective interactions $g_{ij}^{SP}/\Lambda_{\mathrm{NP}}$ and $\tilde{g}_{ij}^{SP}/\Lambda_{\mathrm{NP}}$ are proportional to the SM Higgs vev $v \simeq 246\,\mathrm{GeV}$. If all couplings X_{ij}, Y_{ij}, Z_{ij} are of $\mathcal{O}(1)$, we can expect vector-like quark masses $m_{Q,D} \sim \sqrt{\Lambda_{\mathrm{NP}}v} \sim 10^6\,\mathrm{GeV}$. The couplings above are not all independent but obey the relation

$$
|\tilde{g}_{sd}^{SP}|^2 - |g_{sd}^{SP}|^2 + 2i\,\mathrm{Re}(g_{sd}^{SP}\tilde{g}_{sd}^{SP*}) = \tilde{g}_{dd}^{SP}\tilde{g}_{ss}^{SP*} - g_{dd}^{SP}g_{ss}^{SP*} + i(\tilde{g}_{dd}^{SP}g_{ss}^{SP*} + \tilde{g}_{ss}^{SP*}g_{dd}^{SP})\,. \qquad (29)
$$

One therefore expects that the flavor changing couplings are of the order of the geometric mean of the flavor conserving couplings.

The vector-like quarks also give 1-loop contributions to kaon mixing. We checked explicitly that those contributions scale as $v^2/(m_Q^2 m_D^2)$ and are completely negligible.

2. Inert Higgs doublet model

In a second scenario, we introduce an inert Higgs doublet H with mass m_H, which couples to down and strange quarks, the SM Higgs, and the scalars S and P through the following

interactions:

$$\mathcal{L} \supset m_H^2 H^\dagger H + \lambda_{SP}(H^\dagger h + h^\dagger H)SP$$
$$+ Y_{dd}(\bar{d}_L d_R)H + Y_{ds}(\bar{d}_L s_R)H + Y_{sd}(\bar{s}_L d_R)H + \text{h.c.} . \quad (30)$$

As in the vector-like quark scenario, this inert Higgs Lagrangian is invariant under a Z_2 symmetry: S and P are odd, while all other particles are even. Additional Z_2 symmetric quartic couplings of the inert Higgs involving e.g. S^2 or P^2 are also possible but are not required to generate the low energy interactions in eq. (6), and we neglect them in the following.

Integrating out the inert Higgs at tree level (see fig. 6, right diagram), and matching onto the effective Lagrangian of eq. (6), we find

$$\frac{g_{dd}^{SP}}{\Lambda_{\text{NP}}} = \frac{i\lambda_{SP}v}{\sqrt{2}m_H^2}\,\text{Re}(Y_{dd})\,, \qquad \frac{\tilde{g}_{dd}^{SP}}{\Lambda_{\text{NP}}} = \frac{i\lambda_{SP}v}{\sqrt{2}m_H^2}\,\text{Im}(Y_{dd})\,, \quad (31)$$

$$\frac{g_{ds}^{SP}}{\Lambda_{\text{NP}}} = \frac{\lambda_{SP}v}{\sqrt{2}m_H^2}\frac{i}{2}(Y_{ds} + Y_{sd}^*)\,, \qquad \frac{\tilde{g}_{ds}^{SP}}{\Lambda_{\text{NP}}} = \frac{\lambda_{SP}v}{\sqrt{2}m_H^2}\frac{1}{2}(Y_{ds} - Y_{sd}^*)\,. \quad (32)$$

In addition, integrating out the inert Higgs gives 4-fermion contact interactions of the type $(\bar{d}_L s_R)(\bar{d}_R s_L)$ that modify kaon oscillations. We find the following contributions to the kaon mixing matrix element:

$$M_{12} = \frac{m_{K^0}^3 f_K^2}{4 m_s^2 m_H^2}\eta_{\text{QCD}}B_4 Y_{sd}Y_{ds}^*\,, \quad (33)$$

where $B_4 \simeq 0.78$ [15] (see also [17, 26]) and η_{QCD} is the QCD correction factor given in eq. (10), with $M = m_H$. Modifications to the mixing matrix alter the neutral kaon oscillation frequency ΔM_K and the observable ϵ_K that measures CP violation in kaon mixing. The above contribution to M_{12} modifies these two quantities as

$$\Delta M_K = \Delta M_K^{\text{SM}} + 2\,\text{Re}(M_{12})\,, \qquad \epsilon_K = \epsilon_K^{\text{SM}} + \frac{\text{Im}(M_{12})}{\sqrt{2}\Delta M_K}\,. \quad (34)$$

Taking into account the SM predictions ΔM_K^{SM} and ϵ_K^{SM} from [11, 13], and the corresponding experimental values from [50], we find the bounds

$$\text{Re}(Y_{sd}Y_{ds}^*) < 7.3 \times 10^{-9} \times \left(\frac{m_H}{1\text{ TeV}}\right)^2 \left(\frac{\alpha_s(m_H)}{\alpha_s(1\text{ TeV})}\right)^{8/7}\,, \quad (35)$$

$$\text{Im}(Y_{sd}Y_{ds}^*) < 4.5 \times 10^{-12} \times \left(\frac{m_H}{1\text{ TeV}}\right)^2 \left(\frac{\alpha_s(m_H)}{\alpha_s(1\text{ TeV})}\right)^{8/7}\,. \quad (36)$$

Assuming $|Y_{ds}| \simeq |Y_{sd}|$ and $\mathcal{O}(1)$ CP violating phases, the kaon mixing bounds are compatible with $\Lambda_{sd} \gtrsim 3 \times 10^9$ GeV. Also, note that the bounds are entirely avoided if either of Y_{sd} or Y_{ds} is set to zero.

III. ASTROPHYSICAL AND TERRESTRIAL CONSTRAINTS

We now consider extant astrophysical and terrestrial constraints that may apply to our model.

First, anticipating our treatment of P as a dark matter candidate, we note that direct detection, indirect detection, and self-interaction constraints are not relevant for our model in its minimal configuration (see eq. (6)). If our P is the cosmological dark matter, but the SM is only coupled to the current SP, then direct detection is only sensitive to the inelastic scattering process $P + \text{SM} \to S + \text{SM}$, which is kinematically forbidden unless the dark matter is boosted. Similarly, indirect detection and self-interaction processes require two vertices, and thus the cross sections are suppressed by Λ_{NP}^4.

Extensions of our minimal model containing couplings to P^2 (see eq. (7)) may be subject to these constraints due to the presence of additional interactions. However, we first treat constraints from supernova cooling and beam dump experiments, which apply directly to the minimal model.

A. Supernova constraints

Supernova cooling provides powerful constraints on new weakly-coupled light particles. Evaluating these bounds properly requires a detailed analysis that lies beyond the scope of this work, but we can perform an order-of-magnitude estimate to determine the regions of our parameter space that are likely to be subject to such constraints.

In the case of axions, the cross section for axion production $NN \to NNa$ is constrained by SN1987A to lie in the range [46]

$$3 \times 10^{-20} \lesssim \frac{\sigma}{\text{GeV}^{-2}} \lesssim 10^{-13}. \tag{37}$$

Below the lower limit, axions are not produced in sufficient numbers to affect the cooling process. Above the upper limit, the produced axions are trapped within the supernova environment, and are unable to cool the system more effectively than neutrinos. Many details of the calculation for axions should be modified in our case, but we will make a crude estimate of the constraints by requiring our production cross section to lie in the same range.

Since P is stabilized by a Z_2 symmetry, it can only be produced in pairs, or in association with S. The process $NN \to NNPP$ is mediated at the loop level in the minimal model,

involving two insertions of the effective interaction vertex. Since $T_{\rm SN} \simeq 30\,{\rm MeV}$ [46], we estimate the cross section as

$$\sigma_{NN \to NNPP} \sim \frac{1}{16\pi^2} \frac{T_{\rm SN}^2}{\Lambda_{dd}^4} \simeq 6 \times 10^{-34}\,{\rm GeV}^{-2} \left(\frac{T_{\rm SN}}{30\,{\rm MeV}} \right)^2 \left(\frac{10^7\,{\rm GeV}}{\Lambda_{dd}} \right)^4, \qquad (38)$$

lying below the constrained range of cross sections, even neglecting exponential suppression when $m_P \gtrsim T_{\rm SN}$. In the case of SP production, $NN \to NNSP$, since $m_S \gg T_{\rm SN}$, we estimate the cross section as

$$\sigma_{NN \to NNSP} \sim \frac{1}{4\pi \Lambda_{dd}^2} e^{-(m_S + m_P)/T_{\rm SN}}$$

$$\simeq 7 \times 10^{-21}\,{\rm GeV}^{-2} \exp\left[\frac{35}{3} \left(1 - \frac{m_S + m_P}{350\,{\rm MeV}} \frac{30\,{\rm MeV}}{T_{\rm SN}} \right) \right] \left(\frac{10^7\,{\rm GeV}}{\Lambda_{dd}} \right)^2. \qquad (39)$$

While parts of our parameter space are thus expected to be unconstrained by supernova limits, it is important to note that if m_P is small, or if $\Lambda_{dd} \lesssim 10^6\,{\rm GeV}$, the estimated production cross section enters the prohibited range. In particular, if $\Lambda_{dd} = 10^6\,{\rm GeV}$, then avoiding the bound requires $m_S + m_P \gtrsim 450\,{\rm MeV}$, favoring the larger P masses in fig. 2. However, in this naive projection of supernova constraints, our model remains viable in a wide region of the parameter space.

B.   Beam dump constraints

In minimal form, our model of the KOTO excess is potentially subject to constraints from long-lived particle searches: the partial decay width of $S \to \pi^0 P$ is bounded from below by the observed KOTO event rate, so in the absence of additional interactions, the S lifetime can be $\mathcal{O}({\rm m})$ or larger. Such lifetimes are probed very effectively by beam-dump experiments with $\mathcal{O}(100\,{\rm m})$ baseline lengths. In such an experiment, a proton beam is directed at a target, potentially producing a large number of S particles. The S particles travel unimpeded through shielding and earth over a distance L_B, reaching an instrumented decay volume with length L_D. The $S \to \pi^0 P$ events within the decay volume can be typically detected with an $\mathcal{O}(1)$ efficiency $\mathcal{E}$. Thus, the strength of the constraints is mainly determined by two factors: (1) how many S particles are produced, and (2) what fraction of these undergo $S \to \pi^0 P$ within the decay volume.

First we estimate the number of S particles produced. There are at least two channels to consider: direct production from nucleon-nucleon scattering, and secondary production

from kaon and other meson decays. Observe, however, that the fraction of proton-proton collisions that produce an SP pair is of order $(s/\Lambda_{\mathrm{NP}})^2/\alpha_S^2$, which is much smaller than the branching ratios $\mathrm{BR}(K_L \to SP)$ and $\mathrm{BR}(K_S \to SP)$ implied by our interpretation of the KOTO excess. We also checked that the number of S from eta decays $\eta \to SP$ is small compared to those coming from the kaon decays in our scenarios.

Given N_p protons on target, we expect that of order $N_K \sim 10^{-2} N_p$ kaons are produced [30], and this is sufficient for kaon decays to dominate production. However, of these kaons, most will be stopped or scattered away from the axis of the beam before they decay. The dynamics of kaon energy loss and deflection in materials are complicated, but the nuclear interaction length for relativistic kaons in most materials is $L_{\mathrm{nuc}} \sim \mathcal{O}(10\,\mathrm{cm})$ [50], so we will assume that any kaons traveling this far before decaying are sufficiently slowed down or deflected such that only a negligible fraction of the S particles are directed towards the detector. Thus, the number of S particles produced and directed towards the detector is of order

$$N_S \sim \frac{1}{2} \sum_{X=L,S} 10^{-2} N_p \frac{\Gamma(K_X \to SP)}{\Gamma_{K_X}} \left[1 - \exp\left(-\frac{\Gamma_{K_X} L_{\mathrm{nuc}}}{\gamma_{K_X}} \right) \right], \tag{40}$$

where γ is the boost factor. Now, accounting for the fraction of S particles which decay in the decay volume, and accounting for the efficiency of the detector, the number of events is given by

$$N_E \sim \frac{1}{2} \sum_{X=L,S} 10^{-2} N_p \, \mathrm{BR}(K_X \to SP) \, \mathrm{BR}(S \to \pi^0 P) \mathcal{E}$$
$$\times \underbrace{\left[1 - \exp\left(-\frac{\Gamma_{K_X} L_{\mathrm{nuc}}}{\gamma_{K_X}} \right) \right]}_{\text{avoid kaon deflection}} \underbrace{\exp\left(-\frac{\Gamma_S L_B}{\gamma_S} \right)}_{\text{reach decay volume}} \underbrace{\left[1 - \exp\left(-\frac{\Gamma_S L_D}{\gamma_S} \right) \right]}_{\text{decay in decay volume}}. \tag{41}$$

In the minimal scenario, with no additional interactions, $\mathrm{BR}(S \to \pi^0 P) = 1$.

We now estimate the event counts in the CHARM [7] and NuCal [10] experiments. CHARM conducted a search for decays of axion-like particles with 2.4×10^{18} protons incident on a copper target at $400\,\mathrm{GeV}$, a baseline length of $480\,\mathrm{m}$, and a $35\,\mathrm{m}$-long decay volume. The detector efficiency is approximately 0.5. No candidate events were observed. NuCal conducted a similar search, with 1.7×10^{18} protons incident on an iron target at $70\,\mathrm{GeV}$, a baseline length of $64\,\mathrm{m}$, and a $23\,\mathrm{m}$-long decay volume. One candidate event was observed with an expected standard model background of 0.3. To estimate the event counts

that would be produced by our model, we set $\gamma_{K_X} = \gamma_S = 10$ for CHARM and reduce these proportionally for NuCal's lower beam energy.

Assuming $\mathrm{BR}(S \to \pi^0 P) = 1$, the resulting event count is shown as a function of the S lifetime in fig. 7. The minimum expected number of events at long S lifetime is large unless $\tau_S \gtrsim 10^5$ m, and lifetimes as large as 10^9 m may be excluded. This potentially rules out a significant portion of our parameter space, as indicated in fig. 4. On the other hand, the event rate cuts off sharply for $\tau_S \lesssim 1$ m, and there is indeed a region of our parameter space where $\tau_S \sim 1$ cm. These constraints can be relaxed if the coupling of the SP^3 interaction in our model is non-zero, which can shorten the S lifetime significantly if m_P is small (see fig. 5). The presence of this additional interaction greatly extends the parameter space consistent with the null results at CHARM and NuCal.

Looking towards future prospects, most proposed beam-dump experiments are competitive in the same regime of S lifetimes. However, it has been suggested [8] that the SeaQuest experiment [4] may be modified to serve as a short-baseline beam dump experiment, with the instrumented area starting only ~ 5 m from the beam target. Such an experiment would have sensitivity to lifetimes as short as 5 cm, and could probe most of the parameter space in which the minimal model can account for the KOTO excess. However, if the SP^3 coupling is unconstrained, the S lifetime can be shortened by many orders of magnitude, potentially evading even these experiments.

C. Direct dark matter detection

Direct detection of P can occur in the extended model via the interactions in eq. (7). While the interaction terms containing $(\bar{q} i \gamma_5 q) P^2$ give rise to suppressed velocity-dependent cross sections off of nucleons, the operators $(\bar{q} q) P^2$ with $q = d, s$ produce potentially detectable scattering off of nucleons. We define the integrated nucleon form factors

$$B_q^N \equiv \langle N | \bar{q} q | N \rangle = \frac{m_N}{m_q} f_q^N, \tag{42}$$

where f_q^N are the form factors for nucleon N of quark q [39]. The direct detection cross section can be cast as

$$\sigma = \sum_{q=d,s} \left(\frac{2 m_N}{m_P + m_N} \frac{g_{qq}^{P^2}}{\Lambda_{\mathrm{NP}}} B_q^N \right)^2 \approx \frac{4}{\Lambda_{\mathrm{NP}}^2} \left[(B_d^N)^2 (g_{dd}^{P^2})^2 + (B_s^N)^2 (g_{ss}^{P^2})^2 \right]. \tag{43}$$
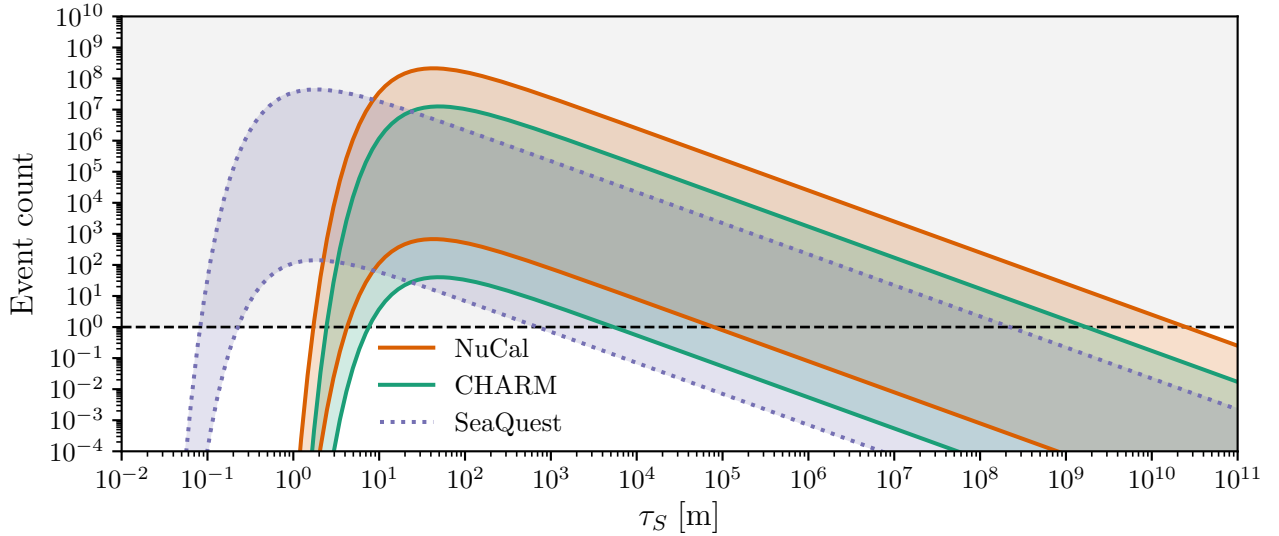
FIG. 7. Estimated event counts at CHARM and NuCal and prospective event counts at SeaQuest as a function of the S lifetime. The top curve fixes $\Gamma(K_L \to SP)$ to saturate the experimental bound on the invisible K_L width. The bottom curve fixes $\Gamma(K_L \to SP)$ such that $\mathrm{BR}(K_L \to SP)$ is equal to the ratio inferred from the KOTO excess, i.e., $\Gamma(K_L \to SP)$ is the smallest width for which this model can account for the excess. Both curves assume that $\Gamma(K_L \to SP) = \Gamma(K_S \to SP)$ and that $\mathrm{BR}(S \to \pi^0 P) = 1$. Under these conditions, $1\,\mathrm{m} \lesssim \tau_S \lesssim 10^5\,\mathrm{m}$ is ruled out. SeaQuest may eventually probe lifetimes as short as $\tau_S \sim 5\,\mathrm{cm}$.

Using the central values $B_d^p \approx 6.77$ and $B_s^p \approx 0.50$, it is clear that the dominant effect is scattering off of d quarks if $g_{ss}^{P2} \simeq g_{dd}^{P2}$. The scattering cross section off of protons is then

$$\sigma_p \approx 7 \times 10^{-38}\,\mathrm{cm}^2 (g_{dd}^{P2})^2 \left(\frac{10^6\,\mathrm{GeV}}{\Lambda_{\mathrm{NP}}} \right)^2, \tag{44}$$

i.e., close to $0.1\,\mathrm{pb}$. Cross sections of this order are above the expected neutrino background, and are within reach of future planned experimental sensitivity [22]. We will return to direct detection prospects in section V.

IV. COSMOLOGICAL PRODUCTION

We now turn to the question of cosmological production of the dark matter candidate P: which scenarios allow P to be produced with the observed dark matter density?

The standard thermal freeze-out paradigm is not viable in our minimal scenario. Esti-

mating the freeze-out temperature by $n_P \sigma(PP \to \mathrm{SM}) \sim H(T)$, we have

$$T_{\mathrm{FO}} \sim \frac{4\pi \Lambda_{\mathrm{NP}}^2}{m_{\mathrm{Pl}}} \sim \left(\frac{\Lambda_{\mathrm{NP}}}{10^{12}\,\mathrm{MeV}} \right)^2 10^3\,\mathrm{MeV}, \tag{45}$$

where Λ_{NP} is the scale of new physics in question—for practical purposes, the lesser of Λ_{sd} and Λ_{dd}. For typical values of Λ_{NP} consistent with the KOTO excess, $T_{\mathrm{FO}} \gg m_P$, so P freezes out as a hot relic, with relic abundance

$$\Omega_P h^2 \sim \frac{m_P}{\mathrm{keV}} \left(\frac{g_\star|_{T_{\mathrm{FO}}}}{100} \right) \sim 0.1 \left(\frac{m_P}{80\,\mathrm{eV}} \right). \tag{46}$$

Thus, for the masses and couplings considered in this work, P is generically overproduced in the freeze-out scenario. If the P mass were small enough to be produced with the right relic abundance, then P would be ruled out as a dark matter candidate because of structure formation constraints on relativistic relics.

Departing from the minimal scenario outlined above opens up the possibility that an *additional* effective interaction with SM species keeps P in thermal equilibrium, and that the P relic abundance is set by thermal decoupling (freeze-out). Since generally thermal decoupling happens at temperatures $T \sim m_P/25$, in order to avoid possible constraints from BBN, one can assume that the effective interaction only involves SM neutrinos:

$$\mathcal{L} \supset \frac{1}{\Lambda_{\nu\nu}} \bar{\nu}\nu PP. \tag{47}$$

For the effective dimension five operator in the equation above, we find that the zero-velocity thermally averaged product of the pair-annihilation cross section and relative velocity is

$$\lim_{v \to 0} \langle \sigma v \rangle = \frac{1}{4\pi} \frac{1}{\Lambda_{\nu\nu}^2}. \tag{48}$$

A standard treatment of the relic abundance for the pair-annihilation cross section above indicates that P would be produced in the right amount if $\Lambda_{\nu\nu} \simeq 7$ TeV. This is several orders of magnitude above current limits for dark matter interactions with SM neutrinos, independent of flavor [9]. Thus, if P were in equilibrium at high temperatures, an effective interaction with SM neutrinos—which, incidentally, can be quite naturally embedded in the UV completions described above—could suppress the P abundance to an acceptable relic density in agreement with observations.

In the absence of the additional neutrino portal described in the paragraph above, the only alternative is production via freeze-in [32]. Here the dark species is produced out of

equilibrium by some standard model species, and the abundance increases until cosmological expansion halts production. It is thus possible to avoid overproduction of dark matter with extremely small couplings. Note that while other mechanisms might allow for additional production of P, the freeze-in contribution is unavoidable in the range of temperatures where our effective theory is valid.

Typically, freeze-in is applied to a UV-complete theory, where the dark matter production rate can be computed starting at very high temperatures. In the context of a renormalizable model, it can be shown that dark matter is produced primarily at lower temperatures, so the details of the UV physics are unimportant. Thus, freeze-in can be used to consistently calculate the non-thermal relic abundance, even though a formal dependence on initial conditions remains. Note that this is in contrast to the freeze-out paradigm, where equilibrium with the standard model bath erases any non-trivial initial conditions in the dark sector.

However, in our scenario, the dark matter is produced through non-renormalizable interactions, and the standard freeze-in mechanism cannot be directly applied: our effective theory cannot be applied at scales above some $\mathcal{O}(\Lambda_{\mathrm{NP}})$ cutoff. At first, this does not seem to be a problem: in standard freeze-in, production is IR-dominated, and we can apply our effective theory in this regime. But for higher-dimension operators, production is no longer IR-dominated, and it is no longer possible to self-consistently estimate the relic abundance unless an initial condition is fixed at a temperature where the effective theory is valid.

Naively, one can place a lower bound on the relic abundance by fixing the dark matter abundance to zero at $T \sim \Lambda_{\mathrm{NP}}$ and computing the amount of dark matter produced at lower temperatures, where the effective theory is valid. However, as we shall see in the following section, this still leads to overproduction of P. Thus, in our model, it would seem that dark matter is overproduced in the freeze-in scenario, even with the most favorable initial conditions.

There is, however, a significant loophole in this argument: setting the dark matter abundance to zero at $T \sim \Lambda_{\mathrm{NP}}$ is in fact not the most favorable initial condition. If reheating occurs at a temperature $T_{\mathrm{rh}} \ll \Lambda_{\mathrm{NP}}$, then the dark matter abundance should be set to zero at this lower temperature, allowing for a much lower relic abundance. There is nothing particularly unnatural about this scenario: in general, freeze-in production of dark matter depends on the reheating temperature. This dependence is weak if the reheating scale happens to be much higher than any scale in the theory, but the convenience of this arrangement

does not constitute evidence for it. Moreover, if $T_{\rm rh} \ll \Lambda_{\rm NP}$, then our effective theory can be used to self-consistently compute the dark matter relic abundance independently of any UV completion. This paradigm is known as UV freeze-in [24].

A.   Computing the yield

First, we briefly review the computation of the dark matter relic abundance in the standard freeze-in paradigm. The basic technology of UV freeze-in is identical to that of standard freeze-in, but the initial condition is fixed at the reheating temperature $T_{\rm rh}$, which becomes an important free parameter of the theory. In certain scenarios, the dark matter yield is quite sensitive to temperatures near $T_{\rm rh}$, and decreasing $T_{\rm rh}$ can significantly reduce the relic abundance.

The starting point is the Boltzmann equation,

$$\dot{n}_\chi + 3H n_\chi = \sum_{I,F} [N_\chi(F) - N_\chi(I)] \int \mathrm{d}^{n_I}\Pi_I \, \mathrm{d}^{n_F}\Pi_F (2\pi)^4 \delta^4 \left(p_I - p_F\right) |\mathcal{M}_{I \to F}|^2 \prod_{i \in I} f_i. \quad (49)$$

Here n_χ denotes the number density of a dark species χ, I and F index initial and final states, $N_\chi(S)$ denotes the number of χ particles in the state S, $\mathrm{d}\Pi_i = g_i \, \mathrm{d}^3\mathbf{p}_i/(2\pi)^3 2E_i$, $|\mathcal{M}_{I \to F}|^2$ is the spin-averaged squared matrix element, and f_k is the phase space density of the species k. We assume Maxwell-Boltzmann statistics, and by conservation of comoving entropy density, we rewrite the left-hand side of eq. (49) as $\dot{n}_\chi + 3H n_\chi = S\dot{Y}_\chi$, where $S = (2\pi^2/45)g_{\star S}T^3$ is the entropy density and $Y_\chi \equiv n_\chi/S$. In turn, since $\dot{T} \approx -HT$, we have $S\dot{Y}_\chi \approx xHSY'_\chi(x)$, where $x = \mu/T$ for any fixed mass μ.

In freeze-in, one assumes that the phase space density of the dark species is always small, so that any initial state with $N_\chi(I) > 0$ makes a negligible contribution in eq. (49). If all of the initial-state species are now in equilibrium, the phase space densities f_i can be replaced with equilibrium distributions $e^{-E_i/T}$. Now eq. (49) reads

$$Y'_\chi(x) = \frac{1}{xHS} \sum_{I \not\ni \chi, F} N_\chi(F) \int \mathrm{d}^{n_I}\Pi_I \, \mathrm{d}^{n_F}\Pi_F (2\pi)^4 \delta^4 \left(p_I - p_F\right) |\mathcal{M}_{I \to F}|^2 \exp\left(-xE_I/\mu\right). \quad (50)$$

We will be interested in two types of processes: $1 \to 2$ decays and $2 \to 2$ scattering. In the $1 \to 2$ case, with a process $i \to \chi f$, we set $\mu = m_i$, i.e., $x = m_i/T$. We recognize the decay width $\Gamma_{i \to \chi f}$ in eq. (50), which becomes

$$Y'_\chi(x) = \frac{1}{2\pi^2} \frac{g_i m_i^3}{x^2 HS} N_\chi(F)\Gamma_{i \to \chi f} K_1\left(x\right), \quad (51)$$

where K_1 is a modified Bessel function of the second kind, and now $N_\chi(F)$ is either 1 or 2, depending on whether $f = \chi$. Substituting $H = 1.66 g_\star^{1/2} x^{-2} m_i^2 m_{\mathrm{Pl}}^{-1}$, the total yield can now be computed by performing a 1-dimensional integration of eq. (51), as

$$Y_\chi(\infty) = \frac{45 N_\chi(F) g_i m_{\mathrm{Pl}} \Gamma_{i \to \chi f}}{1.66 \times 4\pi^4 m_i^2} \int_{x_{\mathrm{min}}}^{\infty} \mathrm{d}x \, \frac{x^3 K_1(x)}{g_\star^{1/2} g_{\star S}}. \tag{52}$$

In particular, suppose that $f = \chi$, $m_\chi \ll m_i$, and $|\mathcal{M}_{i \to \chi\chi}|^2 = \lambda^2$. If production mainly takes place during an epoch when $g_\star$ and $g_{\star S}$ are not changing rapidly, then we can estimate the yield as

$$Y_\chi(\infty) \simeq \frac{135 N_\chi(F) g_i m_{\mathrm{Pl}} \lambda^2}{1.66 \times 8(2\pi)^4 g_\star^{1/2} g_{\star S} m_i^3} \begin{cases} 1 & x_{\mathrm{min}} \ll 1 \\ \frac{1}{3}\sqrt{\frac{2}{\pi}} x_{\mathrm{min}}^{5/2} \exp(-x_{\mathrm{min}}) & x_{\mathrm{min}} \gg 1. \end{cases} \tag{53}$$

Similarly, if the abundance of χ is set by $2 \to 2$ processes of the form $ij \to \chi f$, then the integrals over the final-state phase space produce the cross section $\sigma_{ij \to \chi f}$, and eq. (50) becomes

$$Y_\chi'(x) = \frac{N_\chi(F) g_i g_j}{x H S} \int \frac{\mathrm{d}^3 \mathbf{p}_i}{(2\pi)^3} \frac{\mathrm{d}^3 \mathbf{p}_j}{(2\pi)^3} \, \sigma v \exp(-x E_i/\mu) \exp(-x E_j/\mu). \tag{54}$$

This remaining integrals can be reduced to a single 1d integral, following e.g. [29]. Integrating in x, the yield is then

$$Y_\chi(\infty) = \frac{\mu N_\chi(F) g_i g_j}{2(2\pi)^4} \int_{x_{\mathrm{min}}}^{\infty} \frac{\mathrm{d}x}{x^2 H S} \int_{s_{\mathrm{min}}}^{\infty} \mathrm{d}s \, \sigma(s) \, r_- r_+ \times$$
$$\times \left\{ \frac{m_+ m_-}{s} \left(\frac{\mu}{x} + \sqrt{s} \right) \exp(-x\sqrt{s}/\mu) + \frac{r_- r_+}{\sqrt{s}} K_1(x\sqrt{s}/\mu) \right\}, \tag{55}$$

where $m_\pm = |m_i \pm m_j|$, $r_\pm = (s - m_\pm^2)^{1/2}$, and $s_{\mathrm{min}} = \min(m_i + m_j, m_\chi + m_f)^2$. As in the $1 \to 2$ case, we can estimate the yield analytically for a process $ii \to \chi\chi$ when $m_i \ll m_\chi$ and the evolution of $g_\star$ and $g_{\star S}$ is negligible. If $|\mathcal{M}_{ii \to \chi\chi}|^2 = \lambda^2$, then the result is

$$Y_\chi(\infty) \simeq \frac{45 N_\chi(F) g_i^2 m_{\mathrm{Pl}} \lambda^2}{1.66 \times 512\pi^5 g_\star^{1/2} g_{\star S} m_i} \begin{cases} (3\pi/8) m_i/m_\chi & x_{\mathrm{min}} \ll 1 \\ x_{\mathrm{min}} \exp(-2 x_{\mathrm{min}} m_\chi/m_i) & x_{\mathrm{min}} \gg 1, \end{cases} \tag{56}$$

where $x_{\mathrm{min}} = m_i/T_{\mathrm{max}}$. The analogous expression for $m_\chi \ll m_i$ is obtained by interchanging m_i and m_χ and taking $\mu = m_\chi$ (i.e. $x_{\mathrm{min}} = m_\chi/T_{\mathrm{max}}$). However, in our model, $2 \to 2$ processes are driven by effective 4-point vertices suppressed by a scale Λ_{NP}, so we should

instead set $|\mathcal{M}_{ii\to\chi\chi}|^2 = s/\Lambda_{\mathrm{NP}}^2$. In this case, the result is

$$
Y_\chi(\infty) \simeq \frac{45 N_\chi(F) g_i^2 m_{\mathrm{Pl}} m_\chi^2}{1.66 \times 128\pi^4 g_\star^{1/2} g_{\star S} m_i \Lambda_{\mathrm{NP}}^2}
\begin{cases}
\frac{8}{\pi} (m_\chi/m_i)^{-2} x_{\mathrm{min}}^{-1} & x_{\mathrm{min}} \ll 1 \\[2mm]
x_{\mathrm{min}} \exp\left(-2 x_{\mathrm{min}} m_\chi/m_i\right) & x_{\mathrm{min}} \gg 1.
\end{cases}
\tag{57}
$$

This demonstrates a key difference between standard freeze-in and UV freeze-in: a naive extrapolation of the production rate to arbitrarily high temperatures (small x_{min}) diverges. Of course, one should not expect to accurately compute the production rate in the effective theory at $T \gg \Lambda_{\mathrm{NP}}$. But even so, if $\Lambda_{\mathrm{NP}} \gg T_{\mathrm{max}} \gg \max\{m_\chi, m_i\}$, then production can dominated by $2 \to 2$ processes, whereas $1 \to 2$ decays typically dominate in standard freeze-in. In our case, m_χ and m_i are MeV-scale, while $\Lambda_{\mathrm{NP}} \gtrsim 10^6\,\mathrm{GeV}$. Thus, production by $2 \to 2$ processes at high temperatures is potentially very significant.

Using the approximate forms of the yield derived above together with the dark matter abundance today, $Y_\chi(\infty) \approx 2 \times 10^{-6} (m_\chi/\mathrm{MeV})$, we can estimate the ranges of parameters which account for all of dark matter—or, at least, those which do not overclose the universe. If dark matter in our model is produced dominantly by quark annihilation via an interaction of the form $\Lambda_{dd}^{-1} d(i\gamma_5)\bar{d} S P$, then the only important parameters are Λ_{dd} and x_{min}. Note that if this is the only interaction at work, there is no contribution from decays.

First, suppose that $x_{\mathrm{min}} \ll 1$. Then the scale Λ_{dd} must satisfy

$$
\Lambda_{dd} \gtrsim \left(\frac{g_\star|_{T_{\mathrm{rh}}}}{100}\right)^{-3/4} \left(\frac{T_{\mathrm{rh}}}{\mathrm{GeV}}\right)^{1/2} 3 \times 10^{10}\,\mathrm{GeV}.
\tag{58}
$$

Per the analysis in section II B, this is too large to account for the KOTO excess—and this estimate accounts for only one production channel! In particular, if $T_{\mathrm{rh}} > \Lambda_{dd}$, dark matter is dramatically overproduced. At the very least, one requires $T_{\mathrm{rh}} \lesssim 100\,\mathrm{MeV}$, where the approximations made for this estimate are no longer trustworthy. However, suppose instead that reheating indeed takes place near the MeV scale, so that $x_{\mathrm{min}} \gg 1$. Then the situation is quite different: neglecting the difference between m_S and m_P, we have

$$
\Lambda_{dd} \gtrsim \left(\frac{g_\star|_{T_{\mathrm{rh}}}}{10}\right)^{-3/4} \left(\frac{T_{\mathrm{rh}}}{10\,\mathrm{MeV}}\right) \exp\left[-\left(\frac{m_S}{T_{\mathrm{rh}}} - 30\right)\right] 300\,\mathrm{GeV}.
\tag{59}
$$

This bound poses no obstacle to accounting for the KOTO excess. When combined, these two estimates naively suggest that our model can account for all of dark matter if reheating takes place between $100\,\mathrm{MeV}$ and $10\,\mathrm{MeV}$. While the scale of reheating is often assumed to be much higher, the strongest observational lower bound on the reheating temperature is in

fact only $T_{\rm rh} \gtrsim 5\,{\rm MeV}$ [20, 33]. There is no particularly strong motivation for a very high reheating temperature, and certainly nothing inconsistent about reheating taking place at $10\,{\rm MeV}$.

However, production at such low temperatures introduces a new complication: our simplistic estimates above have presumed not only that production is dominated by $2 \to 2$ processes, but also that the initial state consists of free quarks. If $T_{\rm rh} < 100\,{\rm MeV}$, then quarks are confined into hadrons during the entire production period. One must then modify the effective couplings to account for hadronic scattering, and since the initial and final states are all (pseudo)scalars, the matrix elements no longer carry any s-dependence. Additionally, since single hadrons can now decay to S and P, hadronic decays can dominate the relic abundance, and must be included in the calculation of the yield.

In the following section, we treat these issues in detail and calculate the relic density numerically.

B. Determining the reheating temperature

Our estimates in the previous section suggest that P can be produced non-thermally, and can account for all of dark matter, if the initial temperature of the SM bath is between $100\,{\rm MeV}$ and $10\,{\rm MeV}$. We now refine our estimate of the yield to account for confinement and hadronic decays, and then numerically compute the yield to establish the required reheating temperature in our model.

At $T \lesssim 200\,{\rm MeV}$, quarks are confined into hadrons, and the effective interactions of the hadrons with S and P are well described by chiral perturbation theory (chiPT). The effective couplings of hadrons to S and P are built from a combination of the new physics scales and QCD parameters. Since the couplings in the quark-level effective Lagrangian are proportional to $\Lambda_{\rm NP}^{-1}$, and the hadron-level $1 \to 2$ coupling must have mass dimension 1, the latter must be of order $\Lambda_{\rm chiPT}^2/\Lambda_{\rm NP}$, where $\Lambda_{\rm chiPT}$ is some scale associated with low-energy QCD. Similarly, in the $2 \to 2$ case, the hadron-level coupling should have the form $\Lambda'_{\rm chiPT}/\Lambda_{\rm NP}$. As we will see momentarily, $\Lambda_{\rm chiPT}^{(\prime)}$ is a combination of two constants, $f_\pi \approx 92\,{\rm MeV}$ and $B_0 \approx 2666\,{\rm MeV}$. To determine the couplings explicitly, we match our effective quark-level Lagrangian onto the chiPT Lagrangian following [27, 44]. Our application of this method to light scalars is also similar to the treatment in section 3.1 of [52].

The interactions of QCD degrees of freedom with our light scalars can be written as the couplings of quarks to external currents s and p, respectively a scalar and pseudoscalar. These take the form

$$\mathcal{L}_{\text{QCD}}[s,p] = -\bar{\mathbf{q}}\left(s(x) - i\gamma_5 p(x)\right)\mathbf{q}. \tag{60}$$

Interactions of hadrons with these currents enter the chiPT Lagrangian via the current $\chi = 2B_0(s + ip)$. At lowest order, we have

$$\mathcal{L}_2 \supset \frac{f_\pi^2}{4}\,\text{tr}\left(\chi U^\dagger + U\chi^\dagger\right), \qquad U = \exp\left(\frac{i\sqrt{2}}{f_\pi}\,\Phi\right), \tag{61}$$

where Φ is the PNGB matrix [see e.g. 44]. Now consider a quark-level interaction of the form

$$\mathcal{L} \supset \frac{1}{2}\bar{q}_i\left(g_{ij}^{\mathcal{O}_S} - i\tilde{g}_{ij}^{\mathcal{O}_S}\gamma_5\right)q_j\mathcal{O}_S + \frac{i}{2}\bar{q}_i\left(g_{ij}^{\mathcal{O}_P} - i\tilde{g}_{ij}^{\mathcal{O}_P}\gamma_5\right)q_j\mathcal{O}_P + \text{h.c.} \tag{62}$$

where $\mathcal{O}_S$ is a scalar (CP-even) and $\mathcal{O}_P$ is a pseudoscalar (CP-odd). We can then identify

$$s_{ij} = -\frac{1}{2}\left(g_{ij}^{\mathcal{O}_S} + g_{ji}^{\mathcal{O}_S*}\right)\mathcal{O}_S - \frac{1}{2}\left(g_{ij}^{\mathcal{O}_P} - g_{ji}^{\mathcal{O}_P*}\right)\mathcal{O}_P, \tag{63}$$

$$p_{ij} = -\frac{i}{2}\left(\tilde{g}_{ij}^{\mathcal{O}_S} - \tilde{g}_{ji}^{\mathcal{O}_S*}\right)\mathcal{O}_S - \frac{i}{2}\left(\tilde{g}_{ij}^{\mathcal{O}_P} + \tilde{g}_{ji}^{\mathcal{O}_P*}\right)\mathcal{O}_P. \tag{64}$$

Substituting these expressions into eq. (61) with $\mathcal{O}_S = S^2$, P^2, and $\mathcal{O}_P = SP$ gives the interactions of S and P with the PNGBs. For instance, the interactions of S and P with π^0 are specified by

$$\mathcal{L}_2 \supset B_0 f_\pi \pi^0\left(SP\,\text{Im}\,g_{dd}^{SP} - S^2\,\text{Im}\,\tilde{g}_{dd}^{S^2} - P^2\,\text{Im}\,\tilde{g}_{dd}^{P^2}\right)$$
$$- \frac{1}{2}B_0(\pi^0)^2\left(SP\,\text{Re}\,\tilde{g}_{dd}^{SP} - S^2\,\text{Re}\,\tilde{g}_{dd}^{S^2} - P^2\,\text{Re}\,\tilde{g}_{dd}^{P^2}\right) + \cdots, \tag{65}$$

where the ellipsis denotes a series of higher-dimensional operators. We include all terms up to second order in the PNGB fields in our analysis, and the form of the hadron-level Lagrangian is as expected from dimensional analysis. Note that it is essential to consider complex-valued g_{ij} and $\tilde{g}_{ij}$, without which some interactions will vanish.

We can now determine the reheating temperature required to produce the observed dark matter density as a function of our model parameters. First, using the normalization factors as they appear in eq. (65), we can now estimate the relative significance of decays and scattering, starting with eqs. (54) and (56). Assuming that all dimensionless couplings are $\mathcal{O}(1)$, we set the coupling λ for 3-point vertices equal to $B_0 f_\pi/\Lambda_{\text{NP}}$, and we set the coupling

for 4-point vertices to $B_0/\Lambda_{\rm NP}$. In this regime, we typically have $m_i \gg \max\{m_P, T_{\rm rh}\}$, and in this limit,

$$\frac{Y_P^{1\to2}(\infty)}{Y_P^{2\to2}(\infty)} \simeq 32 \left(\frac{f_\pi}{m_i}\right)^2 \begin{cases} 1 & m_P \ll T_{\rm rh} \ll m_i \\ \frac{3\pi}{8}(T_{\rm rh}/m_i)\exp\left(2m_i/T_{\rm rh}\right) & T_{\rm rh} \ll m_P \ll m_i. \end{cases} \qquad (66)$$

Our parameter space includes $1\,{\rm MeV} \lesssim m_P \lesssim 200\,{\rm MeV}$, so the ratio above can be large or $\mathcal{O}(1)$ depending on the choice of the P mass, but it is never small. Note, however, that increasing m_P can also close certain decay channels. In particular, if there exist interactions allowing the decay $\pi^0 \to PP$, this channel naively dominates production at low temperatures, but is closed for $2m_P > m_{\pi^0}$.

Since decays dominate in most of the parameter space, we can make a first estimate of the yield by considering only production via $K_L \to SP$, the same decay process which is necessary to account for the KOTO excess. Neglecting the distinction between m_S and m_P, the yield is

$$Y_P^{K_L\to SP}(\infty) \simeq \frac{45}{1.66 \times 4(2\pi)^{9/2}g_\star^{1/2}g_{\star S}} \left(\frac{Bf_\pi}{m_K\Lambda_{sd}}\right)^2 \frac{m_{\rm Pl}}{T_{\rm rh}}\exp\left(-2m_S/T_{\rm rh}\right), \qquad (67)$$

and the resulting upper bound on Λ_{sd} is

$$\Lambda_{sd} \gtrsim \left(\frac{g_\star|_{T_{\rm rh}}}{10}\right)^{-3/4}\left(\frac{T_{\rm rh}}{15\,{\rm MeV}}\right)^{1/2}\exp\left[-\left(\frac{m_S}{T_{\rm rh}}-20\right)\right] 5 \times 10^6\,{\rm GeV}. \qquad (68)$$

For the typical parameter values selected above, this upper bound is towards the lower edge of our parameter space of interest for the KOTO excess. Thus, although hadronic decays significantly enhance production relative to the prediction of eq. (59), this channel on its own does not pose an obstacle to accounting for the KOTO excess.

However, in general, it is necessary to numerically evaluate the yield to determine the extent of the viable parameter space—and, in particular, to identify the reheating temperature that produces the observed relic density at each parameter point. The resulting reheating temperatures are shown in fig. 8, and are of order $10\,{\rm MeV}$ throughout the parameter space of interest. The required reheating temperature is mainly controlled by the smaller of Λ_{sd} and Λ_{dd}, with a slight bias towards Λ_{sd}, since production by η decays is suppressed compared to production by K^0 decays due to their relative masses. Note that all couplings except for g_{sd}^{SP} and g_{dd}^{SP} are neglected in fig. 8, so, in particular, $\pi^0 \to PP$ does not contribute to the

relic density even when $2m_P < m_{\pi^0}$. If we suppose that all of the couplings in the effective theory are of similar order, the viable parameter space can change significantly.

We can estimate this effect by taking $g_{q_1 q_2}^{S^2} = g_{q_1 q_2}^{SP} = g_{q_1 q_2}^{P^2}$ and setting $g_{sd}^{\mathcal{O}} = \left(g_{ss}^{\mathcal{O}} g_{dd}^{\mathcal{O}}\right)^{1/2}$ to fix $g_{ss}^{\mathcal{O}}$. The resulting reheating temperatures are shown in fig. 9. With these choices for the couplings, our two benchmark points with $m_P = 10\,\text{MeV}$ are incompatible with freeze-in as a production mechanism, since the required reheating temperature is below observational bounds throughout the relevant parameter space. This is due to the open $\pi^0 \to PP$ decay, which is kinematically closed for the other two benchmark points with $m_P = 100\,\text{MeV}$ and $m_P = 125\,\text{MeV}$. For these points, the required reheating temperature is again of order $10\,\text{MeV}$ throughout the relevant parameter space. At the top-left of the corresponding panel of fig. 9, the required reheating temperature *decreases* with increasing Λ_{dd}. This is just because of our assumption that $g_{sd}^{\mathcal{O}}$ is the geometric mean of $g_{dd}^{\mathcal{O}}$ and $g_{ss}^{\mathcal{O}}$: increasing Λ_{dd} corresponds to decreasing $g_{dd}^{\mathcal{O}}$, so if $g_{sd}^{\mathcal{O}}$ is held fixed, then $g_{ss}^{\mathcal{O}}$ must increase to compensate. This increases the relic density, forcing a lower reheating temperature.

Finally, we note that the reheating temperatures shown in figs. 8 and 9 are potentially imprecise, and should be viewed as lower bounds. Our calculation of the yield assumes that all of the initial-state species are thermalized, but the mesons freeze out at temperatures of the same order considered here. In particular, π^0, K^0, and η freeze out at $3\,\text{MeV}$, $10.5\,\text{MeV}$, and $11.6\,\text{MeV}$, respectively. In a scenario with a high reheating temperature, this concern would be less significant: the mesons would have a thermal distribution at early times, so as long as dark matter production is not dominated by temperatures well below the mesons' freeze-out temperatures, the effect should be small. However, we are speculating that the reheating temperature itself is lower than e.g. the kaon freeze-out temperature in parts of our parameter space, in which case the kaons may never be populated with anything resembling a thermal distribution. It is thus possible that eq. (50) overestimates the dark matter relic abundance.

This does not have a significant effect on our qualitative results: we can safely predict that DM is overproduced if $T_{\text{rh}} \gtrsim 15\,\text{MeV}$, in which case all of the relevant mesons are thermalized, so this is an upper bound on T_{rh}. Likewise, we can see that DM would be underproduced for T_{rh} below a particular value even if the mesons have their equilibrium number densities.[1] This lower threshold is $\mathcal{O}(7\,\text{MeV})$ if $\pi^0 \to PP$ is forbidden, and

---

[1] Since production is dominated by decays, the dark matter relic abundance is mainly determined only
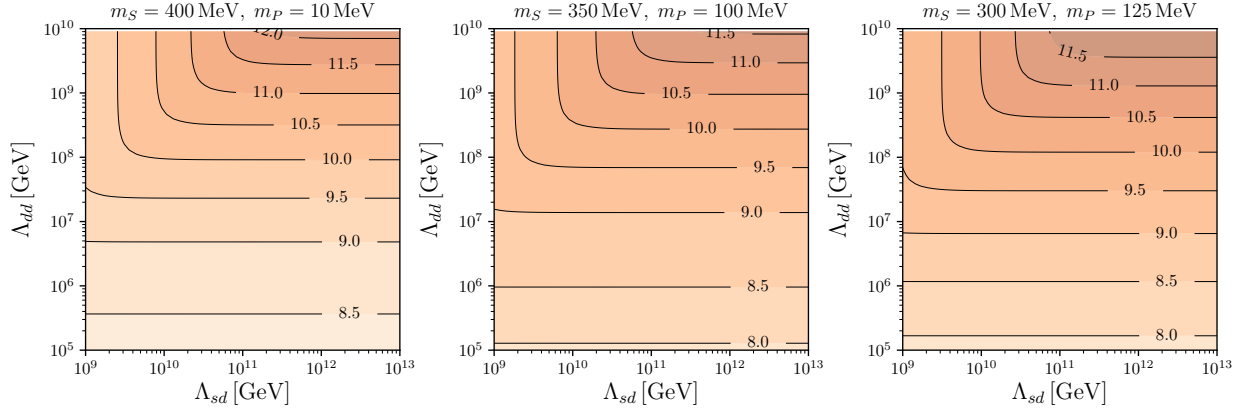
FIG. 8. Reheating temperature in MeV to produce the observed DM relic density, including all production channels with no DM in the initial state. The couplings g_{sd}^{SP} and g_{dd}^{SP} are taken to be purely imaginary, while all other couplings are set to zero, corresponding to the minimal scenario to account for the KOTO excess. In the leftmost panel (BM1), all decay channels are open. In the middle panel (BM2), $S \to 3P$ is kinematically closed, so there are no number-changing interactions in the dark sector: S decays via $S \to \pi^0 P$. In the rightmost panel (BM3), $S \to 3P$ and $\pi^0 \to PP$ are both closed, so there is no contribution to the relic density from π^0 decays.

$\mathcal{O}(2\,\mathrm{MeV})$ if it is not. The only qualitative importance of out-of-equilibrium effects is that it may be possible to construct a cosmologically-viable model in which dark matter is not overproduced even if $\pi^0 \to PP$ is open. However, such a model would depend on the details of reheating, and this analysis lies beyond the scope of this work.

V.   DISCUSSION

In the foregoing sections, we have introduced a model to account for the KOTO excess and explored the cosmological effects. We now discuss the implications of our results and future experimental prospects.

If the KOTO excess is interpreted at face value, this suggests apparent violation of the GN bound. As has been discussed by several authors [19, 21, 23, 25, 30, 36–38, 40–42, 52], such a signal at KOTO can be mimicked by a decay of the form $K_L \to \pi^0 X$, where X denotes one or more invisible species. In contrast to most studies, we focus on a new physics

---

by the number density of the parent mesons, and is fairly insensitive to other details of the phase space distribution.
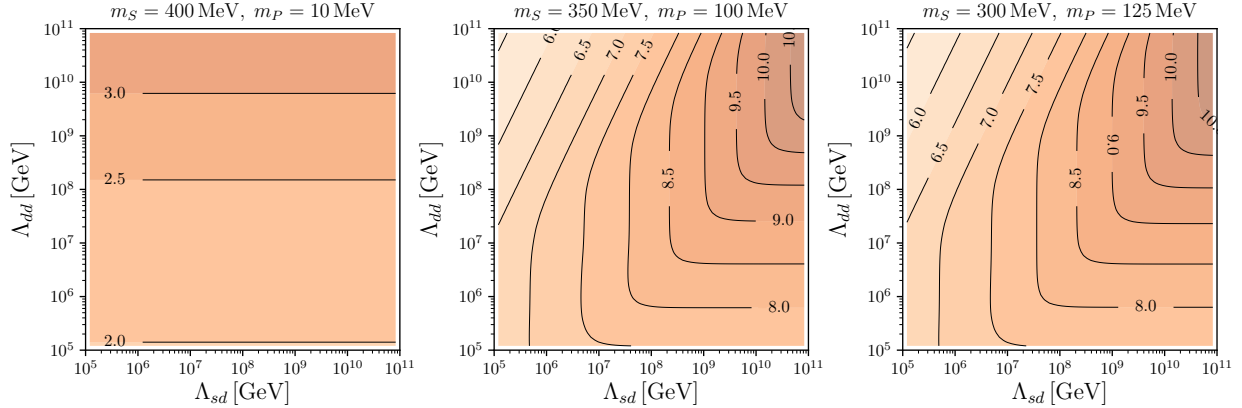
FIG. 9. Reheating temperature (in MeV) to produce the observed DM relic density, including all production channels with no DM in the initial state, as in fig. 8. Here it is assumed that S^2, P^2, and SP couple equally to light quark bilinears, and that $g_{sd}^{\mathcal{O}}$ is the geometric mean of $g_{ss}^{\mathcal{O}}$ and $g_{dd}^{\mathcal{O}}$. The real and imaginary parts of all couplings are taken to be equal. In the first panel (BM1), all decay channels are open, and production is dominated by π^0 decays. In the middle panel (BM2), $\pi^0 \to PP$ is closed, but $S \to 3P$ is still open. In the rightmost panel (BM3), both $\pi^0 \to PP$ and $S \to 3P$ are closed, so S decays only via $S \to \pi^0 P$. In the leftmost panel, since production is dominated by $\pi^0 \to PP$, the relic abundance is controlled exclusively by Λ_{dd}. In this case, the required reheating temperatures are observationally inviable throughout the parameter space. In the other two panels, production is dominated by K^0 and η decays, and their relative importance depends on Λ_{sd} and Λ_{dd}.

scenario where the decay $K_L \to \pi^0$ inv. is realized through a sequence of two-body decays $K_L \to SP \to \pi^0 PP$, where S and P are light neutral scalar particles. Similar scenarios were also studied in [37] where the light particles interact with the SM through a vector or scalar portal. Here we instead analyze a setup where S and P are coupled to the SM through effective operators at a characteristic new physics scale of $\Lambda_{\mathrm{NP}} \sim 10^6$–$10^9\,\mathrm{GeV}$. We have stabilized P with a Z_2 symmetry under which SM species are even and our new species are odd, and we have entertained the possibility of other interactions consistent with such Z_2 invariance, including an SP^3 term that could mediate the decay of $S \to 3P$. Our effective theory is readily UV-completed by e.g. very heavy vector-like quarks or a TeV-scale inert Higgs doublet. Such UV completions can realize a minimal case in which only interactions between SM quarks and SP are present at low energies, as well as more generic cases that

include interactions with S^2 and P^2.

If the KOTO excess persists, the GN bound heavily constrains new physics interpretations. A model of the type we consider, with new light scalars, is one of the simplest and most elegant solutions. Since the scale $\Lambda_{\rm NP} \sim 10^6$–$10^9\,\rm GeV$ indicated by the KOTO excess is so large, most other experiments are not substantially constraining (with the notable exception of beam dump experiments, to which we will return shortly). In particular, in our scenario, there is a large region of parameter space which can account for the KOTO excess while still unconstrained by other rare meson decays. However, it is important to consider astrophysical constraints. Supernova cooling limits can potentially rule out lower P masses: as discussed in section III A, supernova temperatures are high enough, at tens of MeV, to probe the lightest S and P masses that we consider in fig. 2. These constraints are most significant for $\Lambda_{dd} \lesssim 10^6\,\rm GeV$, and it is important to note that establishing firm constraints from supernova cooling requires a much more detailed analysis beyond the scope of this work. However, the simplistic expectation is that P masses of $\mathcal{O}(10\,\rm MeV)$ and below are disfavored, making our scenario easier to test.

Since the KOTO excess motivates the introduction of new feebly-coupled particles, it is natural to speculate that these new species might contribute to cosmological dark matter—and indeed, we have shown that S and P can constitute all of DM even in the most minimal scenarios needed to explain the KOTO signal. Nevertheless, this comes at a cost: in the absence of additional interactions, there is no mechanism to reduce the DM abundance, and cosmological reheating must take place at very late times, at a temperature of order $10\,\rm MeV$. This requirement should be interpreted as a cosmological constraint on our model and similar models accounting for the KOTO excess. The scale of the preferred reheating temperature originates mainly from the masses of the new scalars: since the DM abundance is exponentially suppressed in $m_{\rm DM}/T_{\rm rh}$, the required reheating temperature depends only logarithmically on the couplings and other scales of new physics.

Such a thermal history is necessary because the effective coupling lies in an intermediate regime: it is too small for freeze-out to deplete the DM abundance, but large enough that UV freeze-in generically overproduces DM. Thus, an additional feature is needed to prevent overproduction. The simplest mechanism to accomplish this, without any modification to the model, is to make a judicious choice of the reheating temperature. Since we are working with an effective theory, the DM relic density is inherently sensitive to the reheating

temperature—indeed, if $T_{\rm rh} \gtrsim \Lambda_{\rm NP}$, we cannot consistently calculate the relic density, but only bound it below. Thus, since $T_{\rm rh}$ is necessarily a parameter of our model, $T_{\rm rh} \sim 10\,{\rm MeV}$ is as natural as any other choice. As we have discussed, observational constraints are ineffective at temperatures above $\sim 5\,{\rm MeV}$.

We note that in principle low-temperature reheating might leave an imprint on early universe probes such as BBN and CMB. Unfortunately, such potential signals are highly model dependent. Specifically, low-reheating temperature scenarios have been shown in the literature to impart a significant effect on the synthesis of light elements, primarily via (i) modifications to the Hubble rate around BBN by changing the energy density of both relativistic and matter species; (ii) changing the momentum distribution of electron-flavor neutrinos, which directly enters charged current interactions, in turn governing the neutron-proton chemical equilibrium; and (iii) by entropy exchange that can affect the ratio of neutrino to photon temperature, which in turn is testable with CMB data.

Previous studies (see e.g. [35] and references therein) relied on simple assumptions such as a single massive matter species driving reheating, and decaying primarily into neutrinos [20], or electromagnetically-interacting species [33], or hadrons [34]. Generally, testable effects arise for $T_{\rm rh} \lesssim 5$ MeV, implying that no signal is expected for the scenario discussed here, where $T_{\rm rh} \gtrsim 10$ MeV. However, it is important to point out that the reheating scenario might include features that could manifest themselves when more stringent probes of CMB become available in the future [1]. For instance, the field driving reheating might actually be an *ensemble* of fields, with different masses; the S and P particles might be directly produced in the decay of the field(s) driving reheating, changing the predictions for $T_{\rm rh}$ made above; or new physics in the neutrino sector could make reheating temperatures in the 10s of MeV visible once constraints on $N_{\rm eff}$ significantly improve.

There are other mechanisms which prevent the overproduction of dark matter without requiring a particular temperature for reheating. One possibility is to add an interaction with the SM to restore freeze-out as a viable thermal history, as we discussed briefly in the context of a neutrino portal. This would be a heartening scenario: reheating can still take place at a very high temperature, and the coupling to leptons might allow for additional experimental probes. However, there are several other possibilities. In particular, it is possible that the DM abundance is depleted by additional interactions within the dark sector. This is not possible in our effective theory, but one can consider extensions which keep the DM in

thermal equilibrium long after decoupling from the SM bath, or which allow other number-changing processes at a sufficient rate to allow for freeze-out at high temperatures. We emphasize again that our results imply cosmological constraints on models of the KOTO excess: cosmology requires either a restricted range of reheating temperatures or additional features of the low-energy theory, regardless of what fraction of cosmological DM is composed of P.

Of course, one can also consider constraints which only apply if P makes up a significant fraction of DM. The simplest of these is the Lyman-α constraint on warm DM [51], which requires the P population to be non-relativistic at temperatures of $\mathcal{O}(\text{keV})$. If P is produced non-thermally via decays at $10\,\text{MeV}$, typical energies will be of order the masses of the parent states, i.e., $\mathcal{O}(100\,\text{MeV})$. Thus, in order for P to be non-relativistic when $T_\gamma \sim \text{keV}$, we require that $m_P \gtrsim 10\,\text{keV}$. This is a somewhat weaker bound than one expects from supernovae, but it is not subject to the complicated physics involved in such constraints.

The annihilation cross section into visible states is much too small ($\sim 10^{-50}\,\text{cm}^2$) for indirect detection to be viable, nor is there any significant self-interaction in the dark sector. However, the scattering cross section with nuclei could be as large as $\sim 0.1\,\text{pb}$, and thus potentially within reach of future, planned experimental sensitivity for sub-GeV direct dark matter searches. It is thus possible (albeit not guaranteed) that future experiments will probe such signatures associated with our model—particularly direct detection—but it is important to note that in the minimal scenario for the KOTO excess, these signatures are substantially suppressed even compared to the generic expectation. This is because the KOTO excess only requires SM interactions with the current SP, and not PP. Since any DM accounted for by our model is composed entirely of P, this means that any diagrams contributing to indirect detection must be suppressed by Λ_{NP}^{-4}. Moreover, at lowest order, direct detection is only sensitive to the inelastic scattering process $NP \to NS$, which is kinematically prohibited for non-relativistic DM. It is thus challenging to conclusively establish that P makes up cosmological DM through direct observational means.

However, it is potentially much easier to determine whether a model like ours accounts for the KOTO excess. If the excess persists at its present size, then as KOTO reaches its design sensitivity, hundreds of events will be observed. With a sample of this size, it is possible to distinguish our model from SM three-body decays kinematically in much of our parameter space, simply by measuring the pion's transverse momentum. In fig. 10, we show
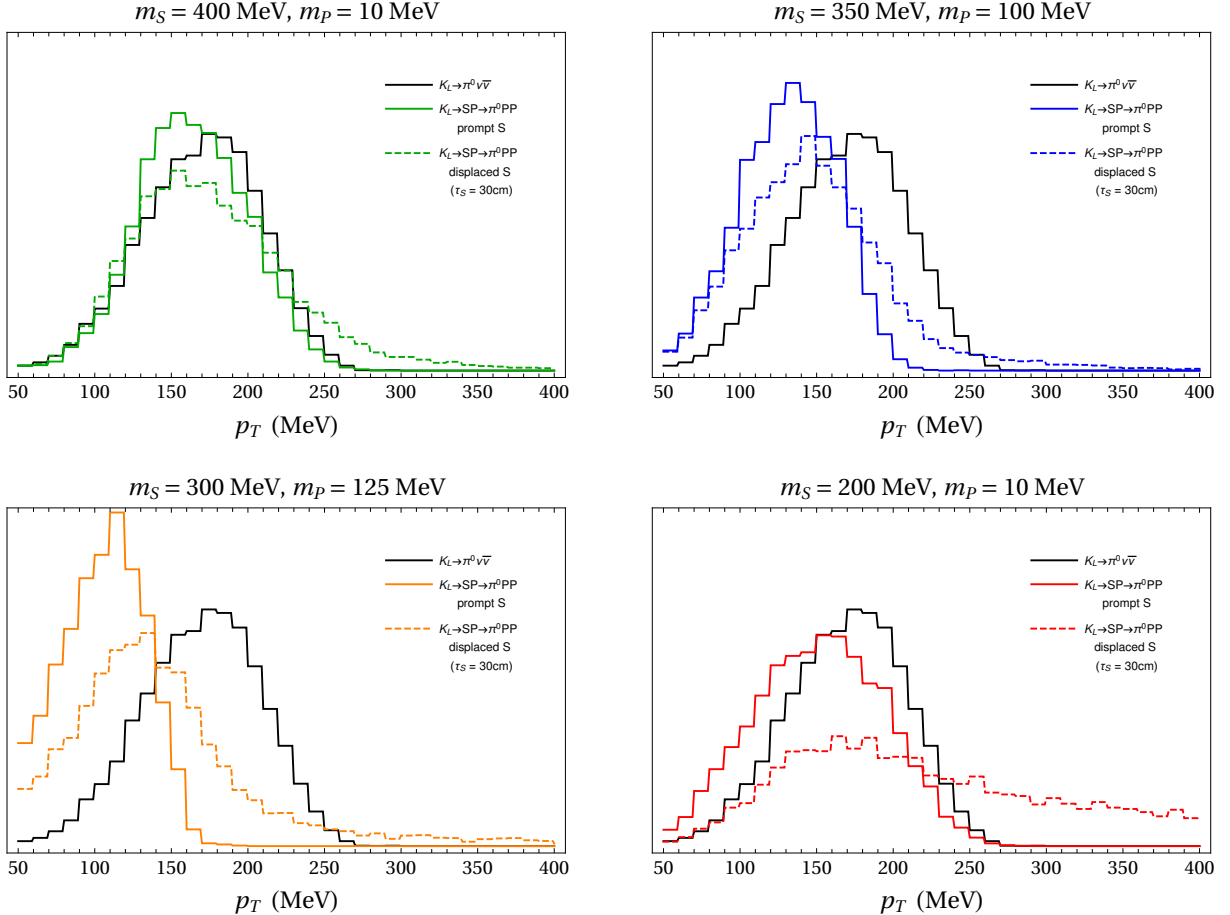
FIG. 10. Pion p_T distributions for the $K_L \rightarrow \pi^0 \nu \bar{\nu}$ decay and the $K_L \rightarrow SP \rightarrow \pi^0 PP$ decay in our benchmark points. The distributions are shown both for prompt S decays and S decays with a lifetime of $30\,\mathrm{cm}$.

the transverse momentum distributions expected at KOTO in the SM and in our model. By sampling from these distributions and applying the Kolmogorov–Smirnov test, we find that the p_T distribution in our model can be distinguished from the SM three-body decay at 5σ with $\mathcal{O}(100)$ events in much of our parameter space. Sensitivity is lost when m_P is small and $m_S \sim m_{K_L}$, and the distributions may also be too close to distinguish at smaller m_S if the S lifetime is shorter than $\mathcal{O}(10\,\mathrm{cm})$. Still, there are good prospects for making such a determination within the next several years, as KOTO continues to collect data.

There are also discovery prospects for S particles with meter- and centimeter-scale lifetimes at future beam-dump experiments. In particular, as discussed in section III B, the SeaQuest experiment can probe much shorter lifetimes than those to which CHARM and NuCal are sensitive. Backgrounds are relatively easy to control for experiments of this type,

and they remain sensitive even in our minimal scenario. The figure of merit is the S lifetime, which is at least $\mathcal{O}(\text{cm})$ in our minimal scenario. This can be reduced by enhancing the SP^3 interaction in our effective theory, but nonetheless, searches for long-lived particles promise to be a powerful probe of our scenario in the coming decade.

VI. CONCLUSIONS

Taken together, the anomalous KOTO events and the Grossman–Nir bound provide a strong hint for light new physics. In this work, we have introduced an effective theory that accounts for the excess in the $K_L \to \pi^0$ inv. channel with a metastable scalar S; a lighter, stable pseudoscalar P; and effective dimension-5 operators that mediate interactions between S, P and the d and s quarks. We provided two UV-complete models that would produce an effective theory consistent with our assumptions. We then investigated the implications of our effective theory for cosmology and vice versa. In particular, we have shown that cosmological overproduction of P places important constraints on the structure of the low-energy theory.

At face value, in our minimal scenario, P cannot account for either dark matter or the KOTO excess unless the reheating temperature is close to $10\,\text{MeV}$. While it is possible to escape this conclusion by augmenting the model, e.g. with couplings of P to neutrinos, a low reheating temperature is unavoidable in the model's simplest incarnation. However, unless P is very light, the required reheating temperature is compatible with current constraints from BBN and CMB, possibly even offering an observational handle on the model once CMB Stage IV experiments further probe the effective number of relativistic species.

Finally, we have discussed three experimental tests of our scenario. First, we have shown that portions of our parameter space are within reach of future dark matter direct detection experiments. Second, our metastable S may be discovered by upcoming long-lived particle searches, particularly the planned SeaQuest upgrade. Finally, if P is in our favored mass range, future KOTO data alone can discriminate between our decay chain and the SM three-body decay on the basis of the neutral pion p_T distribution. There are thus strong discovery prospects for P dark matter within the next decade.

Appendix A: KOTO simulation

In this appendix, we provide details of our calculation of the quantity R introduced in eq. (23). R is the acceptance of the $K_L \to SP \to \pi^0 PP$ signal relative to the SM $K_L \to \pi^0 \nu \bar{\nu}$ acceptance at KOTO. Our calculation is based on a Monte Carlo simulation following steps similar to the ones described in [37, 40].

The layout of the KOTO beamline and the KOTO detector is described e.g. in [43]. We start by generating K_L momenta, p_{K_L}, and K_L decay vertex locations, z_{K_L}, based on the distribution

$$f(p_{K_L}, z_{K_L}) \propto g(p_{K_L}) \times \exp\left(-\frac{(z_{K_L} - z_{\text{exit}})m_{K_L}}{\tau_{K_L} p_{K_L}}\right) , \tag{A1}$$

where $z_{\text{exit}} = 20\,\text{m}$ is the distance of the beam exit from the target and $g(p_{K_L})$ is the measured K_L momentum distribution at the beam exit from [43]. We include a small transverse component of the K_L momentum such that the beam profile at the beam exit is constant within an $8.5\,\text{cm} \times 8.5\,\text{cm}$ square and zero outside [43].

In the case of the SM decay, we generate pion momenta using the $K \to \pi$ form factor from [16]. In the case of the $K_L \to SP \to \pi^0 PP$ decay, we first generate momenta for S, based on the fixed energy of S in the K_L rest frame, $E_S = (m_{K_L}^2 + m_S^2 - m_P^2)/(2m_{K_L})$. We then decay S with a decay length distribution that is determined by the $S \to \pi^0 P$ and $S \to 3P$ partial widths. The pion momentum is generated based on the known pion energy in the S rest frame, $E_{\pi^0} = (m_S^2 + m_{\pi^0}^2 - m_P^2)/(2m_S)$.

Both in the SM case and the NP case, we let the pion decay promptly into two photons, each with energy $E_\gamma = m_{\pi^0}/2$ in the pion rest frame. We reject events with photons produced

less than $2.5\,\mathrm{m}$ after the front face of the front barrel (which starts $1.507\,\mathrm{m}$ after the beam exit), as they would be rejected by photon veto collar counters. All other photons are propagated to the calorimeter located $6.148\,\mathrm{m}$ after the front face of the front barrel [43]. The energy and location of the detected photons in the calorimeter is smeared using the parameters given in [48].

Based on the smeared energy and smeared location of the photons in the calorimeter, the transverse momentum and decay vertex location of the pion is inferred following the procedure described in [43]. If there is more than one solution for the vertex location in the decay volume, we pick the location further away from the calorimeter. We then perform the event selection as in [3], taking into account all cuts but timing and shape related cuts and the trigger related cut on the center of energy deposition. We use the updated signal region in the plane of the inferred pion transverse momentum and the pion decay vertex location from [49].

The results for R in our benchmark scenarios are shown in fig. 3 as function of the S lifetime.

---

[1] Kevork Abazajian et al. CMB-S4 Decadal Survey APC White Paper. *Bull. Am. Astron. Soc.*, 51(7):209, 2019. doi:10.2172/1556957.

[2] S Adler et al. Search for the decay $K^+ \to \pi^+ \nu\bar\nu$ in the momentum region p_π less than 195 MeV/c. *Phys. Lett. B*, 537:211–216, 2002. doi:10.1016/S0370-2693(02)01911-1.

[3] J.K. Ahn et al. Search for the $K_L \to \pi^0 \nu\bar\nu$ and $K_L \to \pi^0 X^0$ decays at the J-PARC KOTO experiment. *Phys. Rev. Lett.*, 122(2):021802, 2019. doi:10.1103/PhysRevLett.122.021802.

[4] C.A. Aidala et al. The SeaQuest Spectrometer at Fermilab. *Nucl. Instrum. Meth. A*, 930: 49–63, 2019. doi:10.1016/j.nima.2019.03.039.

[5] V.V. Anisimovsky et al. Improved measurement of the $K^+ \to \pi^+ \nu\bar\nu$ branching ratio. *Phys. Rev. Lett.*, 93:031801, 2004. doi:10.1103/PhysRevLett.93.031801.

[6] A.V. Artamonov et al. New measurement of the $K^+ \to \pi^+ \nu\bar\nu$ branching ratio. *Phys. Rev. Lett.*, 101:191802, 2008. doi:10.1103/PhysRevLett.101.191802.

[7] F. Bergsma et al. Search for Axion Like Particle Production in 400-GeV Proton - Copper Interactions. *Phys. Lett. B*, 157:458–462, 1985. doi:10.1016/0370-2693(85)90400-9.

[8] Asher Berlin, Stefania Gori, Philip Schuster, and Natalia Toro. Dark Sectors at the Fermilab SeaQuest Experiment. *Phys. Rev. D*, 98(3):035011, 2018. doi:10.1103/PhysRevD.98.035011.

[9] M. Blennow, E. Fernandez-Martinez, A. Olivares-Del Campo, S. Pascoli, S. Rosauro-Alcaraz, and A.V. Titov. Neutrino Portals to Dark Matter. *Eur. Phys. J. C*, 79(7):555, 2019. doi:10.1140/epjc/s10052-019-7060-5.

[10] J. Blumlein et al. Limits on neutral light scalar and pseudoscalar particles in a proton beam dump experiment. *Z. Phys. C*, 51:341–350, 1991. doi:10.1007/BF01548556.

[11] Joachim Brod and Martin Gorbahn. Next-to-Next-to-Leading-Order Charm-Quark Contribution to the CP Violation Parameter epsilon_K and Delta M_K. *Phys. Rev. Lett.*, 108:121801, 2012. doi:10.1103/PhysRevLett.108.121801.

[12] Joachim Brod, Martin Gorbahn, and Emmanuel Stamou. Two-Loop Electroweak Corrections for the $K \to \pi\nu\bar{\nu}$ Decays. *Phys. Rev.*, D83:034030, 2011. doi:10.1103/PhysRevD.83.034030.

[13] Joachim Brod, Martin Gorbahn, and Emmanuel Stamou. Standard-model prediction of ϵ_K with manifest CKM unitarity. 11 2019.

[14] Andrzej J. Buras, Dario Buttazzo, Jennifer Girrbach-Noe, and Robert Knegjens. $K^+ \to \pi^+\nu\bar{\nu}$ and $K_L \to \pi^0\nu\bar{\nu}$ in the Standard Model: status and perspectives. *JHEP*, 11:033, 2015. doi:10.1007/JHEP11(2015)033.

[15] N. Carrasco, P. Dimopoulos, R. Frezzotti, V. Lubicz, G. C Rossi, S. Simula, and C. Tarantino. ΔS=2 and ΔC=2 bag parameters in the standard model and beyond from N_f=2+1+1 twisted-mass lattice QCD. *Phys. Rev. D*, 92(3):034516, 2015. doi:10.1103/PhysRevD.92.034516.

[16] N. Carrasco, P. Lami, V. Lubicz, L. Riggio, S. Simula, and C. Tarantino. $K \to \pi$ semileptonic form factors with $N_f = 2 + 1 + 1$ twisted mass fermions. *Phys. Rev. D*, 93(11):114512, 2016. doi:10.1103/PhysRevD.93.114512.

[17] Benjamin J. Choi et al. Kaon BSM B-parameters using improved staggered fermions from $N_f = 2 + 1$ unquenched QCD. *Phys. Rev. D*, 93(1):014511, 2016. doi:10.1103/PhysRevD.93.014511.

[18] Eduardo Cortina Gil et al. First search for $K^+ \to \pi^+\nu\bar{\nu}$ using the decay-in-flight technique. *Phys. Lett. B*, 791:156–166, 2019. doi:10.1016/j.physletb.2019.01.067.

[19] Alakabha Datta, Saeed Kamali, and Danny Marfatia. Dark sector origin of the KOTO and MiniBooNE anomalies. 5 2020.

[20] P.F. de Salas, M. Lattanzi, G. Mangano, G. Miele, S. Pastor, and O. Pisanti. Bounds

on very low reheating scenarios after Planck. *Phys. Rev. D*, 92(12):123534, 2015. doi: 10.1103/PhysRevD.92.123534.

[21] P.S. Bhupal Dev, Rabindra N. Mohapatra, and Yongchao Zhang. Constraints on long-lived light scalars with flavor-changing couplings and the KOTO anomaly. *Phys. Rev. D*, 101(7): 075014, 2020. doi:10.1103/PhysRevD.101.075014.

[22] Bhaskar Dutta and Louis E. Strigari. Neutrino physics with dark matter detectors. *Ann. Rev. Nucl. Part. Sci.*, 69:137–161, 2019. doi:10.1146/annurev-nucl-101918-023450.

[23] Daniel Egana-Ugrinovic, Samuel Homiller, and Patrick Meade. Light Scalars and the KOTO Anomaly. *Phys. Rev. Lett.*, 124(19):191801, 2020. doi:10.1103/PhysRevLett.124.191801.

[24] Fatemeh Elahi, Christopher Kolda, and James Unwin. UltraViolet Freeze-in. *JHEP*, 03:048, 2015. doi:10.1007/JHEP03(2015)048.

[25] Kaori Fuyuto, Wei-Shu Hou, and Masaya Kohda. Loophole in $K \to \pi \nu \bar{\nu}$ Search and New Weak Leptonic Forces. *Phys. Rev. Lett.*, 114:171802, 2015. doi:10.1103/PhysRevLett.114.171802.

[26] Nicolas Garron, Renwick J. Hudspith, and Andrew T. Lytle. Neutral Kaon Mixing Beyond the Standard Model with $n_f = 2 + 1$ Chiral Fermions Part 1: Bare Matrix Elements and Physical Results. *JHEP*, 11:001, 2016. doi:10.1007/JHEP11(2016)001.

[27] J. Gasser and H. Leutwyler. Chiral Perturbation Theory: Expansions in the Mass of the Strange Quark. *Nucl. Phys. B*, 250:465–516, 1985. doi:10.1016/0550-3213(85)90492-4.

[28] S.N. Gninenko. Search for invisible decays of π^0, η, η', K_S and K_L: A probe of new physics and tests using the Bell-Steinberger relation. *Phys. Rev. D*, 91(1):015004, 2015. doi: 10.1103/PhysRevD.91.015004.

[29] Paolo Gondolo and Graciela Gelmini. Cosmic abundances of stable particles: Improved analysis. *Nucl. Phys.*, B360:145–179, 1991. doi:10.1016/0550-3213(91)90438-4.

[30] Stefania Gori, Gilad Perez, and Kohsaku Tobioka. KOTO vs. NA62 Dark Scalar Searches. 5 2020.

[31] Yuval Grossman and Yosef Nir. $K_L \to \pi^0 \nu \bar{\nu}$ beyond the standard model. *Phys. Lett. B*, 398: 163–168, 1997. doi:10.1016/S0370-2693(97)00210-4.

[32] Lawrence J. Hall, Karsten Jedamzik, John March-Russell, and Stephen M. West. Freeze-In Production of FIMP Dark Matter. *JHEP*, 03:080, 2010. doi:10.1007/JHEP03(2010)080.

[33] Steen Hannestad. What is the lowest possible reheating temperature? *Phys. Rev. D*, 70: 043506, 2004. doi:10.1103/PhysRevD.70.043506.

[34] Takuya Hasegawa, Nagisa Hiroshima, Kazunori Kohri, Rasmus S.L. Hansen, Thomas Tram, and Steen Hannestad. MeV-scale reheating temperature and thermalization of oscillating neutrinos by radiative and hadronic decays of massive particles. *JCAP*, 12:012, 2019. doi: 10.1088/1475-7516/2019/12/012.

[35] Takuya Hasegawa, Nagisa Hiroshima, Kazunori Kohri, Rasmus S.L. Hansen, Thomas Tram, and Steen Hannestad. MeV-scale reheating temperature and cosmological production of light sterile neutrinos. 3 2020.

[36] Xiao-Gang He, Xiao-Dong Ma, Jusak Tandean, and German Valencia. Breaking the Grossman-Nir Bound in Kaon Decays. *JHEP*, 04:057, 2020. doi:10.1007/JHEP04(2020)057.

[37] Matheus Hostert, Kunio Kaneta, and Maxim Pospelov. Pair production of dark particles in meson decays. 5 2020.

[38] Yongsoo Jho, Sung Mook Lee, Seong Chan Park, Yeji Park, and Po-Yan Tseng. Light gauge boson interpretation for $(g-2)_\mu$ and the $K_L \to \pi^0 +$ (invisible) anomaly at the J-PARC KOTO experiment. *JHEP*, 04:086, 2020. doi:10.1007/JHEP04(2020)086.

[39] Chris Kelso, Jason Kumar, Pearl Sandick, and Patrick Stengel. Charged mediators in dark matter scattering with nuclei and the strangeness content of nucleons. *Phys. Rev. D*, 91: 055028, 2015. doi:10.1103/PhysRevD.91.055028.

[40] Teppei Kitahara, Takemichi Okui, Gilad Perez, Yotam Soreq, and Kohsaku Tobioka. New physics implications of recent search for $K_L \to \pi^0 \nu \bar{\nu}$ at KOTO. *Phys. Rev. Lett.*, 124(7): 071801, 2020. doi:10.1103/PhysRevLett.124.071801.

[41] Yi Liao, Hao-Lin Wang, Chang-Yuan Yao, and Jian Zhang. An imprint of a new light particle at KOTO? 5 2020.

[42] Jia Liu, Navin McGinnis, Carlos E.M. Wagner, and Xiao-Ping Wang. A light scalar explanation of $(g-2)_\mu$ and the KOTO anomaly. *JHEP*, 04:197, 2020. doi:10.1007/JHEP04(2020)197.

[43] T. Masuda et al. Long-lived neutral-kaon flux measurement for the KOTO experiment. *PTEP*, 2016(1):013C03, 2016. doi:10.1093/ptep/ptv171.

[44] A. Pich. Chiral perturbation theory. *Rept. Prog. Phys.*, 58:563–610, 1995. doi:10.1088/0034-4885/58/6/001.

[45] M. Pospelov. Status and phenomenology of light bsm. talk at HC2NP 2019, 2019.

[46] G.G. Raffelt. *Stars as laboratories for fundamental physics: The astrophysics of neutrinos, axions, and other weakly interacting particles.* 5 1996. ISBN 978-0-226-70272-8.

[47] G. Ruggiero. New results on $k^+ \to \pi^+ \nu \bar{\nu}$ from the na62 experiment. talk at KAON2019, 2019.

[48] Kazufumi Sato, Eito Iwai, Koji Shiomi, Yasuyuki Sugiyama, Manabu Togawa, and Taku Yamanaka. Performance of KOTO CsI Calorimeter. *JPS Conf. Proc.*, 8:024007, 2015. doi: 10.7566/JPSCP.8.024007.

[49] S. Shinohara. Search for the rare decay $K_l \to \pi^0 \nu \bar{\nu}$ at j-parc koto experiment. talk at KAON2019, 2019.

[50] M. Tanabashi et al. Review of Particle Physics. *Phys. Rev. D*, 98(3):030001, 2018. doi: 10.1103/PhysRevD.98.030001.

[51] Matteo Viel, Julien Lesgourgues, Martin G. Haehnelt, Sabino Matarrese, and Antonio Riotto. Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-alpha forest. *Phys. Rev. D*, 71:063534, 2005. doi: 10.1103/PhysRevD.71.063534.

[52] Robert Ziegler, Jure Zupan, and Roman Zwicky. Three Exceptions to the Grossman-Nir Bound. 5 2020.