PHYSICAL REVIEW D **93,** 043540 (2016)

A tale of two timescales: Mixing, mass generation, and phase transitions in the early universe

Keith R. Dienes,[1,2,*] Jeff Kost,[1,†] and Brooks Thomas[3,4,‡]

[1]*Department of Physics, University of Arizona, Tucson, Arizona 85721, USA*
[2]*Department of Physics, University of Maryland, College Park, Maryland 20742, USA*
[3]*Department of Physics, Reed College, Portland, Oregon 97202, USA*
[4]*Department of Physics, Colorado College, Colorado Springs, Colorado 80903, USA*

(Received 9 September 2015; published 24 February 2016)

Light scalar fields such as axions and string moduli can play an important role in early-universe cosmology. However, many factors can significantly impact their late-time cosmological abundances. For example, in cases where the potentials for these fields are generated dynamically—such as during cosmological mass-generating phase transitions—the duration of the time interval required for these potentials to fully develop can have significant repercussions. Likewise, in scenarios with multiple scalars, mixing amongst the fields can also give rise to an effective timescale that modifies the resulting late-time abundances. Previous studies have focused on the effects of either the first or the second timescale in isolation. In this paper, by contrast, we examine the new features that arise from the interplay between these two timescales when both mixing and time-dependent phase transitions are introduced together. First, we find that the effects of these timescales can conspire to alter not only the total late-time abundance of the system—often by many orders of magnitude—but also its distribution across the different fields. Second, we find that these effects can produce large parametric resonances which render the energy densities of the fields highly sensitive to the degree of mixing as well as the duration of the time interval over which the phase transition unfolds. Finally, we find that these effects can even give rise to a "reoverdamping" phenomenon which causes the total energy density of the system to behave in novel ways that differ from those exhibited by pure dark matter or vacuum energy. All of these features therefore give rise to new possibilities for early-universe phenomenology and cosmological evolution. They also highlight the importance of taking into account the time dependence associated with phase transitions in cosmological settings.

DOI: 10.1103/PhysRevD.93.043540

I. INTRODUCTION

Light scalar fields are a common feature in many cosmological scenarios and thus frequently play an important role in early-universe cosmology. Such scalars include, for example, the QCD axion [1–4] and other axionlike particles [5], supersymmetric partner particles such as sneutrinos and staus [6–9], string moduli such as the dilaton and other geometric moduli [10–12], Q-balls [13,14], quintessence fields [15] and chameleons [16–18], familons [19] and Majorons [20,21], certain degrees of freedom within inert Higgs-doublet models [22], additional scalars present in little-Higgs theories [23], and others [24–27]. Some of these fields may even serve as candidates for dark matter or as significant contributors to dark energy. Moreover, the number of such scalars can be relatively large; indeed, the recent Dynamical Dark Matter framework [28,29] posits a dark sector containing an entire ensemble of such fields which together conspire to produce a number of phenomenological, astrophysical, and cosmological effects which differ markedly from those arising from more traditional dark sectors. Likewise, string-inspired models typically involve significant numbers of moduli—and frequently a large number of light axions and axionlike fields as well [30–33].

When a scalar field is light, the reason is usually that a mass term for this field is forbidden by a symmetry of the full, high-temperature theory. Such a mass term must therefore be generated at a lower temperature scale through some dynamics which violates these symmetries—dynamics typically associated with a cosmological phase transition. For example, masses for axions or axionlike particles can be generated by nonperturbative instanton effects which become significant at temperatures near the confining temperature of the corresponding gauge group. Indeed, dynamical mass generation through cosmological phase transitions provides a natural mechanism for engineering mass-scale hierarchies.

However, dynamical mass generation can also have potentially significant ramifications for the cosmology of the particles involved. First, cosmological phase transitions are never instantaneous. In the case of a second-order phase transition, the transition between the high- and low-temperature phases is smooth but nevertheless time-dependent. Even in the case of a first-order transition, a

*dienes@email.arizona.edu
†jkost@email.arizona.edu
‡bthomas@ColoradoCollege.edu

rapid epoch of bubble nucleation and expansion interpolates between the two phases. As a result, scalar masses which are dynamically generated during either kind of phase transition do not "turn on" instantaneously everywhere in the universe; rather, these masses depend nontrivially on time and evolve continuously over the course of the phase transition towards whatever asymptotic values they will have at late times. It turns out that this time dependence—and the existence of an associated timescale or "width" over which the phase transition unfolds—can have a significant impact on the cosmology of light scalars. For example, in the case of an axion receiving its mass via nonperturbative instanton effects, the nontrivial consequences of this time-dependence on the resulting axion cosmological abundance were assessed in Ref. [34].

Another important consequence of dynamical mass generation is that it can give rise to mixing amongst scalars sharing the same quantum numbers. Indeed, if multiple such scalars couple to the same source of dynamical mass generation, there is no guarantee that the mass eigenstates of the theory will be the same before and after the corresponding phase transition. The phenomenological consequences of such mixing were investigated in Refs. [29,35,36] for an ensemble of light axionlike fields. Mixing can also be induced amongst light scalar fields at low energies through a variety of other dynamical processes, including the integration out of massive fields to which our light scalars are coupled. For example, the phenomenological implications of such mixings amongst axionlike fields have been studied in Ref. [37]. Regardless of the source, however, we find that dynamically induced mixing generally results in a nontrivial modification of the coupling structure of the fields involved and can thus have a significant impact on the resulting energy densities of the fields at late times.

The effects of mixing and of a nontrivial time dependence for scalar mass generation have historically been studied separately. In this paper, by contrast, we shall examine the effects that arise when the mass matrix for a set of multiple light scalar fields includes *both* mixing *and* a nontrivial time dependence. As we shall demonstrate, the interplay between the different timescales associated with these effects can, in the presence of a time-dependent background cosmology, give rise to a number of qualitatively new features which are not seen when these effects are each considered in isolation.

These new features generally concern the late-time cosmological abundances (energy densities) associated with the scalar fields in our coupled multiscalar system. For example, we shall demonstrate that the total late-time energy density of our system can be altered—often by many orders of magnitude–compared with traditional expectations. This includes situations in which the late-time energy density is enhanced to greater values as well as situations in which it is strongly suppressed. Second, we shall demonstrate that not only can the total energy density be altered, but so can the distribution of this total energy density across the different fields in our system—often in dramatic ways. Indeed, these effects can even completely change which field carries the largest abundance. Third, we shall demonstrate that under certain circumstances, the combined effects of our two timescales can give rise to large parametric resonances which render the energy densities of the fields highly sensitive to the degree of mixing as well as the duration of the time interval over which the phase transition unfolds. As a result of this heightened sensitivity, any effects which modify the mixing slightly or which change the rate at which our mass-generating phase transition unfolds can have a huge influence on the corresponding late-time cosmological abundances. Finally, we shall demonstrate that mixing in conjunction with a time-dependent phase transition can even give rise to a "reoverdamping" phenomenon which causes the field values and total energy density of the system to behave in novel ways which differ from those normally associated with pure dark matter or vacuum energy.

Taken together, we see that all of these features give rise to new possibilities for the phenomenology of the early universe and cosmological evolution. In particular, many new possibilities for model-building emerge. However, our results can also serve as a warning: in the presence of multiple mixed scalar fields, it is essential to treat mass-generating phase transitions rigorously, with the proper time dependence included. Indeed, approximations in which such phase transitions are treated as instantaneous can produce late-time energy densities which differ from their true values by many orders of magnitude. This warning is particularly relevant in the case of axions, where the mass-generating phase transition is nothing but the instanton-induced QCD phase transition. This phase transition unfolds over a calculable nonzero timescale, and we shall demonstrate explicitly that ignoring this timescale (by setting it to be either zero or effectively infinite) can lead to highly inaccurate late-time axion abundances.

This paper is devoted to a general study of the new features discussed above. Because these features are common to many systems of scalars which exhibit mixing in conjunction with a time-dependent mass-generating phase transition, in this paper we shall work within the context of a general model of scalars ϕ_i whose identities remain unspecified. We shall likewise make no assumptions about the nature of the mass-generating phase transition except that it unfolds over a certain timescale. Moreover, quite remarkably, we shall find that a simple toy model consisting of only *two* scalar fields is sufficient to illustrate all of the features outlined above. This paper will therefore focus on an analysis of this two-component toy model, and we shall defer a study of more complex scenarios to future work [38].

This paper is organized as follows. In Sec. II, we introduce the two-component toy model which forms the basis for all subsequent discussions in this paper. As outlined above, this toy model contains both a mixing between our two fields as well as a time-dependent mass-generating phase transition. Along the way we also introduce all needed definitions, conventions, and notation. We also discuss some of the basic properties of this model. Then, in Sec. III, we study the behavior of the total late-time energy density of our two-component system, while in Sec. IV we study how this total late-time energy density is distributed between the two fields of our model. In Sec. V we discuss the parametric resonance which appears in the energy density, while in Sec. VI we turn our attention to the "reoverdamping" phenomenon.

Although our analysis up to this point is completely general, our results have many immediate phenomenological applications. In Sec. VII, we therefore focus on a special case by considering the effects on the standard QCD axion that emerge when a second axion is incorporated into the theory. In Sec. VIII we then sketch some of the additional features that emerge when more than two fields are considered. These features will ultimately be explored more fully in Ref. [38]. Finally, in Sec. IX, we conclude with a discussion of our main results and possible avenues for extension and generalization.

This paper also contains two appendices. Although most of the results of this paper are obtained through numerical analysis, Appendix A provides exact analytical results in certain tractable special cases. Appendix B then discusses an alternative approach towards analyzing some of the features presented in this paper.

II. A TOY MODEL

In this section we delineate the toy model which shall be the basis of our analysis in this paper. This section will also serve to establish our notation and conventions.

A. Motivation: A one-component warm-up

As a prelude to the presentation of our model, we begin with a short discussion of the one-component case. Along the way we will also review some basic facts about scalar fields in a Friedmann-Robertson-Walker (FRW) cosmology and provide motivation for the particular construction of our eventual multicomponent toy model.

Towards this end, let us consider a single scalar field ϕ of mass m evolving in a flat FRW universe. If all spatial variations in ϕ are assumed to be negligible, such a field evolves according to the standard equation of motion

$$\frac{\partial^2\phi}{\partial t^2} + 3H(t)\frac{\partial\phi}{\partial t} + m^2\phi = 0 \tag{2.1}$$

where the Hubble parameter $H(t)$ scales as $3H(t) \approx \kappa/t$ with $\kappa = 2$ (respectively $\kappa = 3/2$) during a matter- (radiation-)dominated epoch. As a result, the field behaves as a damped oscillator with a time-dependent damping ratio $\zeta(t) \equiv 3H(t)/2m$. Since $H(t)$ falls with time, a field which is initially overdamped will inevitably become underdamped and experience decaying oscillations.

For simplicity, the energy-momentum tensor for such a field can be modeled as a perfect fluid with energy density ρ and pressure p given by

$$\rho = \frac{1}{2}[(\partial\phi/\partial t)^2 + m^2\phi^2],$$
$$p = \frac{1}{2}[(\partial\phi/\partial t)^2 - m^2\phi^2]. \tag{2.2}$$

If ϕ is initially in an overdamped phase with $\partial\phi/\partial t = 0$, then $\rho = -p$: the energy density ρ associated with such a field in this phase behaves as vacuum energy. By contrast, after ϕ transitions to an underdamped phase, the field eventually experiences oscillations which are approximately virialized: in this phase $p = 0$ and the corresponding energy density ρ can be associated with massive matter. Of course, there is also an intermediate time interval *during* the transition from the overdamped to underdamped phase within which the behavior of the field exhibits transient features that eventually dissipate.

The equation of motion (2.1) can be simplified by defining a dimensionless time variable $\tau \equiv mt$. The solutions can then be expressed analytically in terms of Bessel functions of the first and second kind. The exact evolution of the corresponding energy density ρ is shown in Fig. 1, where τ_ζ denotes the critical damping time at which $\zeta = 1$. The different phases of the system are clearly evident in Fig. 1. For early times $\tau \ll \tau_\zeta$, the system is overdamped and the energy density ρ is essentially constant: $\rho \sim \rho_0$. For late times $\tau \gg \tau_\zeta$, by contrast, the system is underdamped and the energy density ρ scales as $\rho \sim \tau^{-\kappa}$. Finally, during the intermediate times $\tau \sim \tau_\zeta$, the energy density ρ exhibits a corresponding transition between the two limiting behaviors above, punctuated by small transient oscillations. Although the specific dynamics of the field during this transitional period does not affect the late-time asymptotic scaling of ρ with time, these transients are nevertheless important in determining the overall scale of the late-time energy density.

In this example, we have taken the mass m to be a constant, nonzero for all time. However, in many cosmological situations, masses are generated by phase transitions. In such cases, the masses of such fields can be time-dependent. A well-motivated example of this is case of the QCD axion: the axion potential is flat (i.e., $m = 0$) at temperatures $T \gg \Lambda_{\rm QCD}$, but this flat potential is modified at lower temperatures by instanton effects which generate an effective mass.

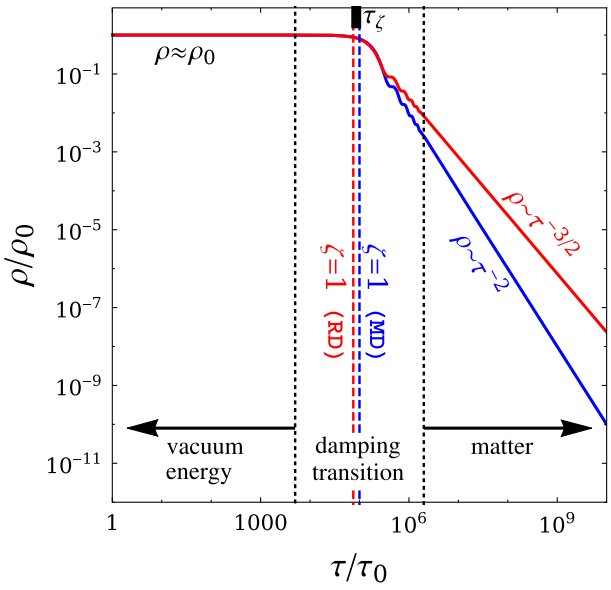

FIG. 1. The energy density ρ, normalized to its initial value ρ_0, for a solution to Eq. (2.1) in a radiation-dominated (RD) universe, i.e., $\kappa = 3/2$ (red), or a matter-dominated (MD) universe, i.e., $\kappa = 2$ (blue). In the overdamped phase, the energy density remains nearly constant at $\rho = -p > 0$, behaving as vacuum energy. By contrast, once the critical damping $\zeta = 1$ threshold is crossed at τ_ζ (a critical time which is slightly different for RD and MD universes), the energy density begins to dissipate, asymptotically exhibiting the simple power-law scaling behavior $\rho \sim \tau^{-\kappa}$ expected for matter.

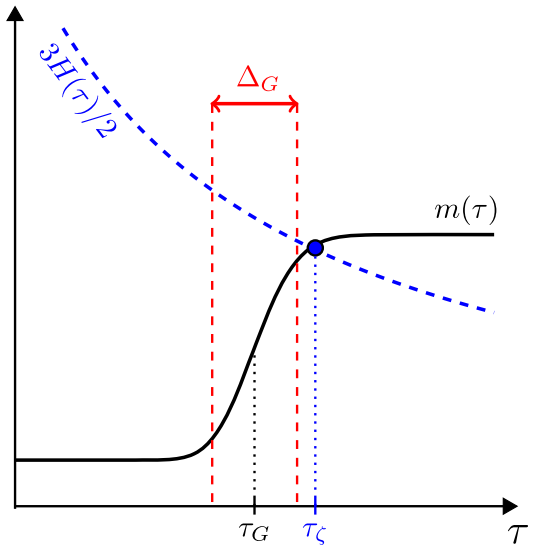

FIG. 2. Two time-dependent mass scales: a growing mass function $m(\tau)$ and a falling Hubble friction scale $3H(\tau)/2$. Their intersection determines the transition time τ_ζ. We assume that the growth of $m(\tau)$ takes place during an interval of approximate duration Δ_G centered at τ_G. As we shall see, the introduction of a nonzero Δ_G can have many dramatic effects on the late-time energy densities of those fields which (either directly or indirectly) are sensitive to this change in mass. For example, one new possibility is that the transition between overdamped and underdamped phases can occur *during* the mass-generating phase transition.

Time-dependent masses $m(\tau)$ can significantly modify the evolution of such scalar fields. In this paper, we are primarily interested in cases in which at least a portion of the contributions to the scalar masses are generated as the result of phase transitions. We shall therefore generally consider cases in which $m(\tau)$ has one (smaller) value at early times, a second (larger) value at asymptotically late times, and a smooth time-dependent transition between the two. Indeed, we shall let τ_G represent the "central" time at which this mass-generating phase transition occurs, and imagine that this transition unfolds over a time interval of duration or width Δ_G. We then obtain the situation illustrated schematically in Fig. 2.

The introduction of a nonzero time interval Δ_G during which our phase transition unfolds can have a significant effect on the time-evolution of ϕ. When $\Delta_G = 0$, there are essentially two regimes which one can consider, each with its own distinctive phenomenology: $\tau_\zeta < \tau_G$ and $\tau_\zeta > \tau_G$. However, a nonzero width Δ_G opens up a third possibility, with $\tau_\zeta \sim \tau_G$. Indeed, more generally, the introduction of a nonzero Δ_G introduces a new timescale into the problem, and we shall see that this can have dramatic effects. We shall explore the phenomenology of this region as part of our larger general study. We emphasize, however, that phenomenon of having $\tau_\zeta \sim \tau_G$ is completely different from having τ_G during the "damping transition" region of Fig. 1. Indeed, while Δ_G may be considered to represent a width around τ_G, the "damping region" of Fig. 1 can instead be considered a width around τ_ζ, a width which exists—as shown in Fig. 1—*even if the mass is constant*.

B. Defining our toy model

The introduction of a nonzero time interval Δ_G for our phase transition is just one feature we wish to study in this paper. The other concerns the possibility of *mixing* between different field components ϕ_i, $i = 1, \ldots, N$. The phenomenological effects stemming from each can certainly be studied in isolation. However, as we shall find, these two features can conspire to produce a number of remarkable effects that transcend what is possible with either alone. For this reason, we shall study the implications of both features together. The effects of either feature in isolation can then be extracted through the limiting cases in which the effects due to the other feature are gradually turned off.

In this paper, we shall explore these effects within the context of a simple toy model involving only two components, ϕ_0 and ϕ_1. As we shall see, our toy model is simple enough to be tractable, yet rich enough to incorporate all of the phenomena of interest.

Our model consists of two real scalar fields, ϕ_0 and ϕ_1. If we again assume that the spatial variations in these fields

are negligible, their equations of motion in a flat FRW universe take the form

$$\frac{\partial^2\phi_i}{\partial t^2}+3H(t)\frac{\partial\phi_i}{\partial t}+\sum_j\mathcal{M}^2_{ij}\phi_j=0 \tag{2.3}$$

where $\mathcal{M}^2$ is the corresponding squared-mass matrix. At times $t\ll t_G$, long before mass generation occurs, we shall take $\mathcal{M}^2$ constant and diagonal. In fact, for simplicity, we shall further assume that ϕ_0 is massless at such early times. Thus, at early times, we shall assume

$$\mathcal{M}^2\underset{t\ll t_G}{\longrightarrow}\begin{bmatrix}0&0\\0&M^2\end{bmatrix} \tag{2.4}$$

where $M\neq 0$ is a general unfixed mass parameter. By contrast, long after the phase transition has occurred, we assume that new components $\bar{m}^2_{ij}$ will have been generated in the squared-mass matrix:

$$\mathcal{M}^2\underset{t\gg t_G}{\longrightarrow}\begin{bmatrix}0&0\\0&M^2\end{bmatrix}+\begin{bmatrix}\bar{m}^2_{00}&\bar{m}^2_{01}\\\bar{m}^2_{01}&\bar{m}^2_{11}\end{bmatrix}. \tag{2.5}$$

In order to connect these two asymptotic extremes, we shall let $m^2_{ij}(t)$ denote time-dependent elements of the mass matrix which interpolate between zero at early times and $\bar{m}_{ij}$ at late times. We can then write

$$\mathcal{M}^2(t)=\begin{bmatrix}0&0\\0&M^2\end{bmatrix}+\begin{bmatrix}m^2_{00}(t)&m^2_{01}(t)\\m^2_{01}(t)&m^2_{11}(t)\end{bmatrix}. \tag{2.6}$$

Having specified the mass matrix, we can now introduce a dimensionless time variable $\tau\equiv Mt$, as in the single-component case. Our equations of motion then take the form

$$\ddot{\phi}_i+3H(\tau)\dot{\phi}_i+\sum_j\mathcal{M}^2_{ij}(\tau)\phi_j=0 \tag{2.7}$$

where the dots indicate $\partial/\partial\tau$ and where our mass matrix is also dimensionless and takes the form

$$\mathcal{M}^2(\tau)=\begin{bmatrix}m^2_{00}(\tau)&m^2_{01}(\tau)\\m^2_{01}(\tau)&1+m^2_{11}(\tau)\end{bmatrix} \tag{2.8}$$

with each dimensionless m_{ij} now understood to be a fraction of M. *We shall adopt these conventions throughout the rest of this paper.*

In general, there is no reason to expect that the mass generation occurs in the same basis as in Eq. (2.4). We shall therefore allow for the possibility that $m^2_{01}\neq 0$—i.e., the possibility that our phase transition induces a mixing between our primordial field components ϕ_0 and ϕ_1. However, we shall nevertheless make the simplifying assumption that the time dependence of each component is identical, allowing us to focus on those effects that come from a single common timescale for mass generation. Indeed, since the mass matrix is nothing but the curvature matrix associated with the potential $V(\phi_0,\phi_1)$ induced by the phase transition, this assumption is tantamount to assuming a single time dependence for the potential as a whole. As a result, we can write each of the individual mass components in a factorized form

$$m_{ij}(\tau)=\bar{m}_{ij}\cdot h(\tau), \tag{2.9}$$

where $h(\tau)$ is a smooth function of time which describes the time development of the phase transition.

Our next step in specifying our toy model is to choose a suitable function $h(\tau)$. As indicated above, we require that $h(\tau)\to 0$ as $\tau\to 0$ and $h(\tau)\to 1$ as $\tau\to\infty$. However, because one of our main interests in this paper concerns the timescale associated with the mass-generating phase transition, we would also like $h(\tau)$ to incorporate a dimensionless parameter σ which controls how abruptly the phase transition occurs. The limit $\sigma\to 0$ might then correspond to a phase transition which is effectively instantaneous, while nonzero values of σ correspond to phase transitions which occur increasingly slowly. It is also desirable, regardless of the width of the transition, that the midpoint at τ_G be a fixed point of reference. We therefore include in our construction the requirement that $h(\tau_G)=1/2$ for all σ. Indeed, this may be taken as a definition of τ_G.

Beyond these constraints, the choice of $h(\tau)$ is completely arbitrary, and many functions may be chosen. For concreteness, however, we shall take

$$h(\tau)=\frac{1}{2}\left\{1+\mathrm{erf}\left[\frac{1}{\sigma}\log\left(\frac{\tau}{\tau_G}\right)\right]\right\}, \tag{2.10}$$

where the error function $\mathrm{erf}(z)$ is given by

$$\mathrm{erf}(z)\equiv\frac{2}{\sqrt{\pi}}\int_0^z e^{-x^2}dx. \tag{2.11}$$

As illustrated in Fig. 3, this function satisfies all of our requirements. Of course, many other choices for $h(\tau)$ are possible. However, none of the qualitative results of this paper will ultimately depend on the specific choice for $h(\tau)$. Thus, any smooth, monotonic function $h(\tau)$ satisfying the above constraints will lead to similar results.

Corresponding to each nonzero value of the parameter σ there exists a nonzero timescale Δ_G over which the phase transition occurs. In general, we may define Δ_G in terms of the slope of the $h(\tau)$ function at its midpoint τ_G:

$$h(\tau_G+\delta\tau)\approx\frac{1}{2}+\frac{\delta\tau}{\Delta_G}\quad\text{for }\delta\tau\ll\Delta_G. \tag{2.12}$$

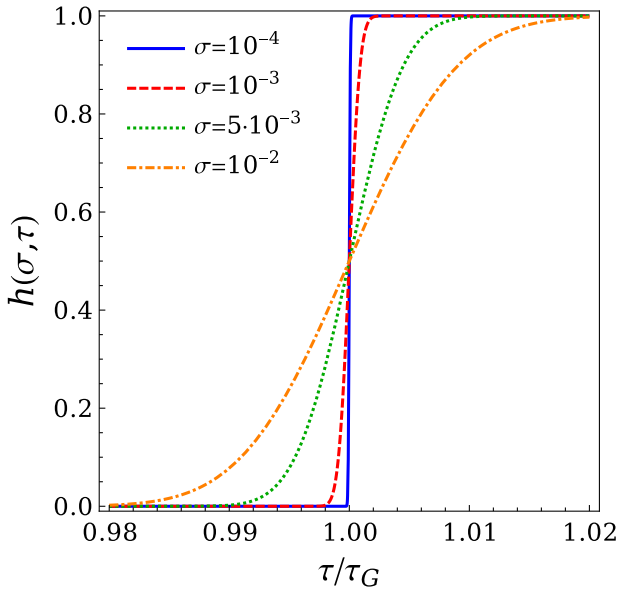

FIG. 3. The general form of the h-function in Eq. (2.10), plotted for several choices of the parameter σ. In all cases we see that $h(\tau)$ rises smoothly from $h=0$ to $h=1$ and crosses $h=1/2$ at $\tau=\tau_G$. Large values of σ correspond to more gradual transitions, while for $\sigma\to 0$ we find that h approaches the Heaviside step function $\Theta(\tau-\tau_G)$.

Adopting this definition for Δ_G, we then find for our $h(\tau)$ function that

$$\Delta_G \equiv \sqrt{\pi}\sigma\tau_G. \tag{2.13}$$

We thus see that $\Delta_G\to 0$ as $\sigma\to 0$, as expected. Indeed, this limit corresponds to the case of an instantaneous phase transition, with the $h(\tau)$ taking the form of the Heaviside step function $\Theta(\tau-\tau_G)$.

In this paper, we shall consider the effects that arise when σ is nonzero. There is, however, a critical value σ^* above which the behavior of $h(\tau)$ changes in an important way. To see this, let us first consider the limit in which $\sigma\to\infty$. In this limit, we find that $h(\tau)$ is essentially constant and τ_G-independent: $h(\tau)\approx\frac{1}{2}\Theta(\tau)$ for all τ. This limit may therefore be interpreted as one in which our original phase transition at τ_G effectively disappears and is replaced by a new, infinitely sharp "phase transition" at $\tau=0$. Note that this latter transition is nothing but an artifact of our boundary condition that $h=0$ at $\tau=0$. As a result, dialing σ from 0 to ∞ has the effect of slowing our original phase transition near τ_G while simultaneously building up an artificial phase transition near $\tau=0$. Indeed, if σ grows too large, the most rapid changes in the mass parameters of our system will no longer be associated with our original phase transition near $\tau=\tau_G$, but with the artificial one near $\tau=0$.

In this paper, we wish to maintain the notion that increasing the value of σ corresponds to slowing the time development of our mass matrix. Even more importantly, we also wish to ensure that the time period exhibiting the most rapid time development of our mass matrix is still associated with our original phase transition near τ_G and not the artifact near $\tau=0$. Therefore, in this paper, we shall always restrict our attention to values $0\le\sigma\le\sigma^*$ where σ^* is that value of σ for which $\max_\tau(dh/d\tau)$ is minimized. For the $h(\tau)$ function given in Eq. (2.10), we find that $\sigma^*=\sqrt{2}$ for all τ_G. As a result, in what follows we shall only consider values of Δ_G in the range

$$0\le\Delta_G\le\sqrt{2\pi}\tau_G. \tag{2.14}$$

In order to gain insight into the values of σ that we might expect for a well-motivated phase transition, let us consider the case of the instanton-induced phase transition during which a mass is generated for the QCD axion. Detailed lattice studies [39,40] have yielded approximate expressions for the mass of the QCD axion as a function of temperature T. However, the cosmological time-temperature relation appropriate for a radiation-dominated epoch is given by

$$T(t)=\left(\frac{5}{2g_*}\right)^{1/4}\left(\frac{3M_P}{\pi t}\right)^{1/2}, \tag{2.15}$$

where M_P is the Planck mass and where g_* is the temperature-dependent effective number of relativistic degrees of freedom. Use of this relation then allows us to determine the axion mass as a function of cosmological time t, with the results shown in Fig. 4. For comparison, our h-function with $\sigma\approx 0.47$ is also shown in Fig. 4. It is clear that these two functions match quite well over the entire range of times shown.

The only remaining ingredient to be specified as part of our toy model is a set of initial conditions to be imposed at some early time τ_0. For the differential equations in Eq. (2.7), this means specifying initial values for our two fields $\phi_{0,1}$ and their first derivatives $\dot\phi_{0,1}$. While many possibilities exist, one particularly natural choice is to consider a mere displacement for the ϕ_0 field:

$$\begin{pmatrix}\phi_0\\ \phi_1\end{pmatrix}_{\tau=\tau_0}=\begin{pmatrix}A_0\\ 0\end{pmatrix},\qquad \begin{pmatrix}\dot\phi_0\\ \dot\phi_1\end{pmatrix}_{\tau=\tau_0}=\begin{pmatrix}0\\ 0\end{pmatrix}. \tag{2.16}$$

There are three special properties associated with this initial configuration which make this choice especially appealing. First, as long as $\tau_0\ll\tau_G$, this configuration does not introduce any initial energy into the system. Thus, all of the energy that our system accrues will be solely that injected through the phase transition. Second, as long as $\tau_0\ll\tau_G$, the specific choice of the initial time τ_0 will be irrelevant. In other words, given our other assumptions, the system is essentially time-independent for all $\tau\ll\tau_G$ and thus completely insensitive to variations of τ_0. As a result,

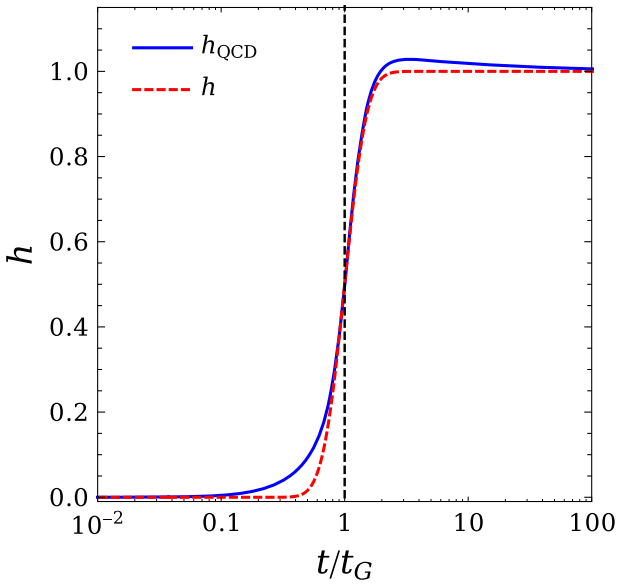

FIG. 4. The function $h_{\rm QCD}$ describing the instanton-induced time-dependent mass of the axion (solid blue line), plotted as a function of t/t_G. This curve is generated for $\Lambda_{\rm QCD} = 200$ MeV, with t_G defined as that location where the curve crosses $h = 1/2$. For comparison purposes, our h-function in Eq. (2.10) with $\sigma \approx 0.47$ is superimposed (dashed red line). We see that our h-function matches $h_{\rm QCD}$ surprisingly well throughout the entire range of values of plotted, and fits particularly well near the "center" of the phase transition at $t = t_G$.

we shall never need to refer to τ_0 again, understanding implicitly throughout this paper that τ_0 has always been chosen to be sufficiently early that our results are independent of τ_0. Finally, the choice of A_0 in Eq. (2.16) merely serves to set an overall mass scale for our system. However, since we shall always be considering *ratios* of field values or energy densities in this paper, our results will also be insensitive to A_0.

This completes the specification of our toy model. It contains only five free parameters: two (namely τ_G and $\Delta_G \sim \sigma$) which describe the time development of the phase transition, and three (namely m_{00}^2, m_{11}^2, and m_{01}^2) which describe the effects of the phase transition and the mixing it induces. Despite its simplicity, however, we shall find that this toy model is not only of sufficient generality to accommodate a wide variety of physical systems but also of sufficient richness to give rise to a number of surprising phenomena.

C. Preliminary analysis: Constraints on mixing

In subsequent sections, we shall analyze the behavior of our fields and of their corresponding energy densities within different regions of our five-dimensional parameter space. However, even before proceeding, there are a number of preliminary observations that hold more generally and which serve to significantly constrain the allowed parameter space.

These constraints all ultimately stem from the observation that not every choice of the masses m_{ij}^2 can be made independently. Requiring that the eigenvalues of $\mathcal{M}^2$ be real is tantamount to demanding that $\mathcal{M}^2$ be Hermitian. However we must further demand that these eigenvalues be non-negative, which requires that $\mathcal{M}^2$ be positive-semidefinite. This requires that all m_{ij}^2 be real and satisfy the three constraints

$$\begin{cases} \bullet & m_{00}^2 \geq 0 \\ \bullet & m_{11}^2 \geq -1 \\ \bullet & m_{00}^2(1+m_{11}^2) \geq m_{01}^4. \end{cases} \tag{2.17}$$

These constraints provide a set of intrinsic limits on the magnitude of the mixing that may occur within our toy model. Indeed, these constraints imply that

$$|m_{01}^2| \leq [m_{01}^2]_{\rm max} \equiv \sqrt{m_{00}^2(1+m_{11}^2)}. \tag{2.18}$$

However, for many purposes it will prove useful to introduce the variables

$$\begin{aligned} m_{\rm sum}^2 &\equiv \mathcal{M}_{11}^2 + \mathcal{M}_{00}^2 = 1 + m_{00}^2 + m_{11}^2 \\ \Delta m^2 &\equiv \mathcal{M}_{11}^2 - \mathcal{M}_{00}^2 = 1 - m_{00}^2 + m_{11}^2 \end{aligned} \tag{2.19}$$

in terms of which our squared-mass matrix $\mathcal{M}^2$ in Eq. (2.8) takes the more symmetric form

$$\mathcal{M}^2 = \begin{bmatrix} \frac{1}{2}(m_{\rm sum}^2 - \Delta m^2) & m_{01}^2 \\ m_{01}^2 & \frac{1}{2}(m_{\rm sum}^2 + \Delta m^2) \end{bmatrix}. \tag{2.20}$$

Note while the individual m_{ij}^2 quantities each carry the same time dependence [proportional to $h(\tau)^2$], the same is no longer true for $m_{\rm sum}^2$ and Δm^2. In terms of these variables, the constraints in Eq. (2.17) then take the form

$$\begin{cases} \bullet & m_{\rm sum}^2 \geq 0 \\ \bullet & |\Delta m^2| \leq m_{\rm sum}^2 \\ \bullet & |m_{01}^2| \leq \frac{1}{2}\sqrt{(m_{\rm sum}^2)^2 - (\Delta m^2)^2}, \end{cases} \tag{2.21}$$

whereupon we find from Eq. (2.18) that

$$[m_{01}^2]_{\rm max} = \frac{1}{2}\sqrt{(m_{\rm sum}^2)^2 - (\Delta m^2)^2}. \tag{2.22}$$

Finally, perhaps the most useful way to parametrize the mixing in our toy model is in terms of a rotation angle θ which relates the mass eigenstates ϕ_{λ_0} and ϕ_{λ_1} at any instant of time to the original mass eigenstates prior to the onset of the phase transition:

$$\begin{pmatrix}\phi_{\lambda_0}\\ \phi_{\lambda_1}\end{pmatrix}=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\phi_0\\ \phi_1\end{pmatrix}, \tag{2.23}$$

from which it follows that

$$\tan(2\theta)=\frac{2m_{01}^2}{\Delta m^2}. \tag{2.24}$$

Indeed, while θ generally populates the range $-\pi\le\theta\le\pi$ in Eq. (2.23), we see from Eq. (2.24) that it is sufficient to focus on $-\pi/2\le\theta\le\pi/2$ for the purpose of calculating mixing angles, since these are invariant under the mapping $(\phi_{\lambda_0},\phi_{\lambda_1})\to-(\phi_{\lambda_0},\phi_{\lambda_1})$. The same will also be true for calculating energy densities—our main interest in this paper—because these energy densities will depend only quadratically on the fields in Eq. (2.23). However, we further observe that in the $\{\phi_{\lambda_0},\phi_{\lambda_1}\}$ basis introduced in Eq. (2.23), the equations of motion in Eq. (2.7) now take the form

$$\begin{aligned}&\ddot{\phi}_{\lambda_0}+3H\dot{\phi}_{\lambda_0}+(\lambda_0^2-\dot{\theta}^2)\phi_{\lambda_0}\\&\qquad=-2\dot{\theta}\dot{\phi}_{\lambda_1}-(\ddot{\theta}+3H\dot{\theta})\phi_{\lambda_1}\\&\ddot{\phi}_{\lambda_1}+3H\dot{\phi}_{\lambda_1}+(\lambda_1^2-\dot{\theta}^2)\phi_{\lambda_1}\\&\qquad=+2\dot{\theta}\dot{\phi}_{\lambda_0}+(\ddot{\theta}+3H\dot{\theta})\phi_{\lambda_0}.\end{aligned} \tag{2.25}$$

These equations of motion exhibit an invariance under the simultaneous correlated transformations $(\phi_{\lambda_0},\phi_{\lambda_1})\to(\phi_{\lambda_0},-\phi_{\lambda_1})$ and $\theta\to-\theta$. Since this is also an invariance of our initial conditions in Eq. (2.16), our toy model will exhibit this invariance at all points during its time evolution. Note that this invariance is also tantamount to the simultaneous correlated transformations $(\phi_0,\phi_1)\to(\phi_0,-\phi_1)$ and $m_{01}^2\to-m_{01}^2$. We can henceforth truncate our attention to $m_{01}^2\ge0$ and $0\le\theta\le\pi/2$ without loss of generality, and we shall do so throughout the rest of this paper. The region $0\le\theta<\pi/4$ then corresponds to $\Delta m^2>0$, while the region $\pi/4<\theta\le\pi/2$ corresponds to $\Delta m^2<0$.

It then follows from Eq. (2.24) that for any values of $m_{\rm sum}^2$ and Δm^2, the allowed mixing angles θ are those for which

$$|\tan(2\theta)|\le|\tan(2\theta_{\rm max})|\equiv\sqrt{\left(\frac{m_{\rm sum}^2}{\Delta m^2}\right)^2-1} \tag{2.26}$$

where $\theta_{\rm max}$ is defined as the value of θ which maximizes $|\tan(2\theta)|$. Thus $\theta\le\theta_{\rm max}$ for $\theta_{\rm max}\le\pi/4$, while $\theta\ge\theta_{\rm max}$ for $\theta_{\rm max}\ge\pi/4$. These angles are illustrated in Fig. 5. Note that for $\Delta m^2=\pm m_{\rm sum}^2$ (corresponding to $m_{00}^2=0$ or $m_{11}^2=-1$, respectively), the corresponding values of θ are restricted to 0 or $\pi/2$. Indeed, any other values would lead to a squared-mass matrix $\mathcal{M}^2$ with at least one negative eigenvalue, corresponding to a tachyonic mode. Finally, for $\Delta m^2=0$, we have $\theta=\pi/4$ for all $m_{01}^2>0$, while the angle θ is in principle undetermined for $m_{01}^2=0$.

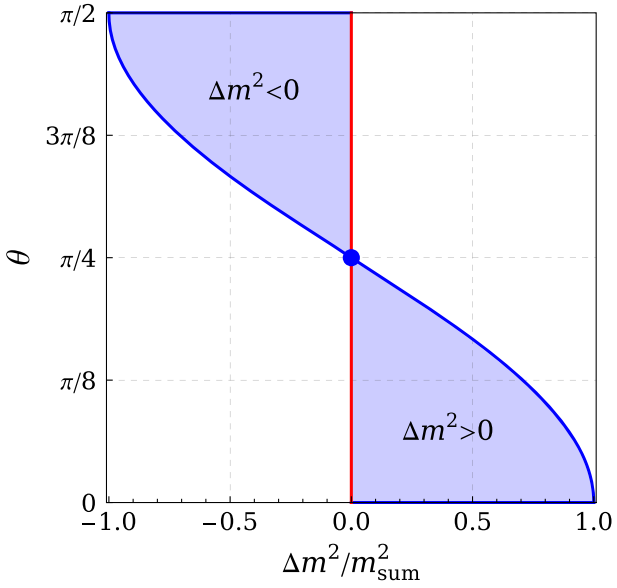

FIG. 5. Allowed ranges for the mixing angle θ (blue shaded region), plotted as a function of $\Delta m^2/m_{\rm sum}^2$. Note that θ is restricted to 0 or $\pi/2$ when $\Delta m^2=\pm m_{\rm sum}^2$—i.e., when $m_{00}^2=0$ or $m_{11}^2=-1$, respectively. By contrast, when $\Delta m^2=0$, the allowed range for θ depends on the value of m_{01}^2: for $m_{01}^2\neq0$, the only allowed θ-value is $\theta=\pi/4$ (blue dot), while for $m_{01}^2=0$ we find that $\tan(2\theta)$ is indeterminate and any value for θ is allowed (red line), depending on how relevant limits are taken.

We conclude this section with some further definitions that shall also prove useful in what follows. Quite often, we shall be evaluating a quantity X which is a function of the time τ. For example, such quantities might include the mass parameters $m_{ij}(\tau)$, the mixing angle $\theta(\tau)$, the field values $\phi_\lambda(\tau)$, or the total energy density $\rho(\tau)$. For any such quantity $X(\tau)$, we shall define

$$\bar{X}\equiv\lim_{\tau\to\infty}X(\tau). \tag{2.27}$$

In other words, $\bar{X}$ shall denote the asymptotic *late-time* value of $X(\tau)$, where in practical terms the notion of a "late" time can be taken as referring to a time at which both fields ϕ_λ have reached the asymptotic underdamped regime, with corresponding energy densities ρ_λ exhibiting the virialized damped scaling behavior $\rho_\lambda\sim\tau^{-\kappa}$ shown in Fig. 1. Finally, we shall also find that a crucial measure for the degree of mixing that might be present in a given system is not the absolute value of the mixing parameter m_{01}^2 (or θ), but rather the value of this parameter *as a fraction of the total degree of mixing that would have been allowed for that system*, given the constraints discussed above. The same is often true for the splitting parameter Δm^2. Towards this end, we define the *mixing saturation*

$$\xi \equiv \frac{m_{01}^2}{[m_{01}^2]_{\rm max}} = \frac{|\tan(2\theta)|}{|\tan(2\theta_{\rm max})|} = \frac{m_{01}^2}{\sqrt{m_{00}^2(1+m_{11}^2)}} \tag{2.28}$$

as well as the analogous *splitting saturation*

$$\eta \equiv \frac{\Delta m^2}{[\Delta m^2]_{\rm max}} = \frac{\Delta m^2}{m_{\rm sum}^2} = \frac{1-m_{00}^2+m_{11}^2}{1+m_{00}^2+m_{11}^2}. \tag{2.29}$$

Note that in terms of these variables[1] we have

$$\tan(2\theta) = \frac{\xi}{\eta}\sqrt{1-\eta^2}. \tag{2.30}$$

We emphasize, however, that a maximally saturated mixing configuration with $\xi = 1$ does *not* necessarily imply that the mixing itself is maximal or even large on an absolute scale. For this reason, we shall usually specify the degree of mixing in a given configuration by quoting both ξ and $\theta_{\rm max}$. Indeed, for many quantities, it will be ξ rather than θ or $\theta_{\rm max}$ which characterizes the behavior of interest and which allows us to compare across systems with different values of θ or $\theta_{\rm max}$.

Note that for $\Delta m^2 = 0$, ambiguities can arise when defining ξ. In particular, the middle expression in Eq. (2.28) is formally indeterminate in such situations. In this paper, for $\Delta m^2 = 0$, we shall therefore define ξ through the final expression in Eq. (2.28). Thus ξ can vary even though θ may be fixed at $\pi/4$. Moreover, again following the final expression in Eq. (2.28), in this paper we shall define

$$X|_{\substack{\Delta m^2=0\\ \xi=0}} \equiv \lim_{m_{01}^2\to 0}\left(X|_{\Delta m^2=0}\right) \tag{2.31}$$

for any quantity X. Thus the $\xi \to 0$ limit will always be a smooth one with $\theta = \pi/4$, even for $\Delta m^2 = 0$.

D. Temporal properties of the model: Basic features

Given the definition of our toy model, it is now relatively straightforward to study the corresponding dynamics. Indeed, this can be done numerically if not analytically, and certain features are entirely as expected and relatively easy to understand. However, many other features are surprising and will play a significant role in what follows. In the remainder of this section, therefore, we shall discuss general features of the time evolution of this model. In particular, we shall focus on the time dependence of the mass eigenvalues, mixing angles, and mass eigenstates—quantities upon which our future results will rest.

We begin by studying the mass eigenvalues in our model. At any moment in time, these masses are given by

$$\begin{aligned}\lambda_{0,1}^2 &= \frac{1}{2}m_{\rm sum}^2\left[1 \mp \sqrt{\eta^2 + (1-\eta^2)\xi^2}\right] \\ &= \frac{1}{2}m_{\rm sum}^2(1 \mp \eta\sec 2\theta).\end{aligned} \tag{2.32}$$

The late-time values of these masses are shown in Fig. 6, while the evolution of these masses from early to late times is shown in Fig. 7. Numerous features are immediately apparent. First, as predicted from Eq. (2.32), we see that the eigenvalues at early and late times ($\tau \ll \tau_G$ and $\tau \gg \tau_G$, respectively) are independent of the sign of Δm^2 (or η). [In this connection we recall from Fig. 5 that $\sec(2\theta) < 0$ when $\eta < 0$.] Thus, the early- and late-time values of $\lambda_{0,1}^2$ are identical in the left and right panels of Fig. 7. However, we see that the time evolution of the eigenvalues *between* these two endpoints is highly sensitive to the sign of $\Delta\bar{m}^2$. For $\Delta\bar{m}^2 > 0$, the eigenvalues evolve from initial to final values without any tendency towards level-crossing. For $\Delta\bar{m}^2 < 0$, by contrast, the eigenvalues initially head towards each other as if to experience a level-crossing. However, whether this level-crossing actually occurs depends on the value of the mixing. For $\bar{\xi} > 0$, the nonzero

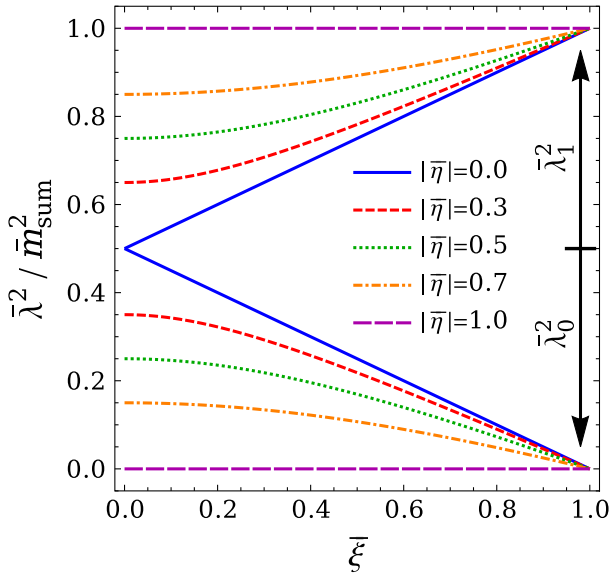

FIG. 6. Asymptotic (late-time) eigenvalues $\bar{\lambda}_i$ plotted as functions of the late-time splitting and mixing parameters $\bar{\eta}$ and $\bar{\xi}$. For all values of $\bar{\eta}$, we see that $\bar{\lambda}_0^2 \to 0$ and $\bar{\lambda}_1^2 \to \bar{m}_{\rm sum}^2$ as $\bar{\xi} \to 1$.

[1]We caution the reader that unlike the other variables we have thus far introduced, neither ξ nor η has a unique time dependence. In other words, knowledge of the value of either ξ or η at a given time does not fix the value of ξ or η at other times without knowledge of the values of the more primordial m_{ij}^2 variables from which ξ and η are derived. Thus, when we write a relation such as that in Eq. (2.30), we are illustrating the functional dependence of θ on ξ and η and asserting that θ does not depend on the more primordial variables except in these combinations. However, the time dependence of θ cannot be determined from such an expression and must be calculated directly in terms of the primordial variables themselves.

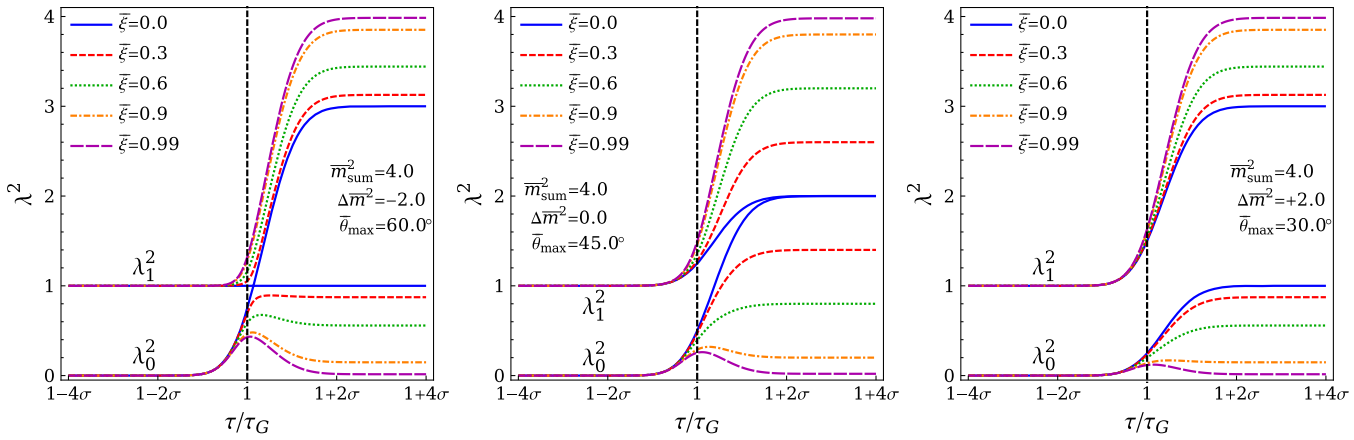

FIG. 7. Time evolution of the two mass eigenvalues $\lambda^2_{0,1}$ as the phase transition unfolds, plotted for different mixing saturations. Typical behaviors are shown for $\Delta\bar{m}^2 < 0$ (left panel), $\Delta\bar{m}^2 = 0$ (middle panel), and $\Delta\bar{m}^2 > 0$ (right panel). In all cases, the eigenvalues begin at $\{0, 1\}$, as expected. For $\Delta\bar{m}^2 > 0$, the eigenvalues slowly transition to their late-time values without any tendency towards level-crossing. For $\Delta\bar{m}^2 < 0$, by contrast, the eigenvalues meet and rebound off each other in the case of zero mixing ($\bar{\xi} = 0$) but such a meeting is thwarted by *level repulsion* for all nonzero mixing. In general, the strength of the level repulsion increases with degree of mixing $\bar{\xi}$. Despite these different features at intermediate times, we observe that the eigenvalues ultimately arrive at the same late-time values regardless of whether $\Delta\bar{m}^2$ is positive or negative, in accordance with Eq. (2.32). Finally, for $\Delta\bar{m}^2 = 0$, we see that there is no tendency towards level-repulsion, but for $\bar{\xi} = 0$ the two eigenvalues approach each other asymptotically.

mixing between mass eigenstates induces a *level repulsion* which ultimately prevents a direct level-crossing. As a result, the eigenvalues veer away from each other, and ultimately assume the same late-time values that they had for $\Delta\bar{m}^2 > 0$.

The case with $\Delta\bar{m}^2 < 0$ and $\bar{\xi} = 0$ is more subtle, and deserves special discussion. As $\bar{\xi} \to 0$, we see from the left panel of Fig. 7 that our eigenvalues λ_0 and λ_1 actually *meet* near τ_G before bouncing off each other. [Note, in this connection, that λ_0 is always defined as the *lighter* eigenvalue, consistent with Eq. (2.32).] In this sense, no actual level- *crossing* occurs, even for $\bar{\xi} = 0$; instead, each eigenvalue develops a "kink"—i.e., a discontinuous slope —at their meeting point near $\tau = \tau_G$.

Finally, for $\Delta\bar{m}^2 = 0$, we see from Fig. 7 that there is no tendency towards level-crossing. However, for $\bar{\xi} = 0$, the two eigenvalues approach each other asymptotically.

We can also study the mixing angle θ as a function of time. The result is shown in Fig. 8. As we see, for $\bar{\xi} > 0$ the mixing angle begins at zero, as expected, and rises to a nonzero late-time value $\bar{\theta}$; this corresponds to Δm^2 transitioning from its initial value $\Delta m^2 = m^2_{\rm sum} = 1$ to its final value $\Delta\bar{m}^2$. Once again, however, the behavior of the angle θ is highly sensitive to the sign of the late-time value $\Delta\bar{m}^2$. For $\Delta\bar{m}^2 > 0$, the angle θ remains below $\pi/4$ (i.e., within the lower right "lobe" of Fig. 5). For $\Delta\bar{m}^2 < 0$, by contrast, the angle θ eventually grows above $\pi/4$, moving from the lower right lobe of Fig. 5 to the upper left. The case with $\Delta\bar{m}^2 = 0$ will be discussed below.

It is natural to define the time τ_θ at which θ crosses its midpoint value $\bar{\theta}/2$; in general, this is the time for which

$$h(\tau_\theta) = \frac{1}{\sqrt{1 + \bar{\lambda}_1^2 - \bar{\lambda}_0^2}}. \tag{2.33}$$

Likewise, for $\Delta\bar{m}^2 < 0$, we find that the transition between lobes at $\theta = \pi/4$ occurs at the time τ^* for which

$$h(\tau^*) = \frac{1}{\sqrt{1 - \Delta\bar{m}^2}}. \tag{2.34}$$

In general, this "lobe-crossing" time τ^* is distinct from τ_G and τ_θ. However, τ^* and τ_θ coincide when $\bar{\theta} = \pi/2$. As we see from Fig. 8, this happens only when $\Delta\bar{m}^2 < 0$ and $\bar{\xi} = 0$.

Once again, this case with $\Delta\bar{m}^2 < 0$ and $\bar{\xi} = 0$ deserves special mention. At early times we have $\Delta m^2 = m^2_{\rm sum} = 1$, and our mixing angle θ begins at zero. This angle then remains at zero all the way until the eigenvalue meeting time shown in the left panel of Fig. 7; note that this eigenvalue meeting time is indeed nothing but $\tau^* = \tau_\theta$. At this time, θ changes instantaneously to $\theta = \pi/2$, consistent with the fact that after τ^* we have $\Delta m^2 < 0$. Thus we see that θ behaves as a step function in this limit, with an effectively instantaneous field rotation. Note that this "instantaneous" behavior in the $\xi \to 0$ limit applies only for $\Delta\bar{m}^2 < 0$.

The situation with $\Delta\bar{m}^2 = 0$ also deserves special attention. As usual, we have $\theta = 0$ at early times. For all $\bar{\xi} \neq 0$, this mixing angle θ then transitions to $\bar{\theta} = \pi/4$ at late times. Moreover, as discussed above, it follows formally from Eq. (2.31) that $\bar{\theta} = \pi/4$ even for $\bar{\xi} = 0$.

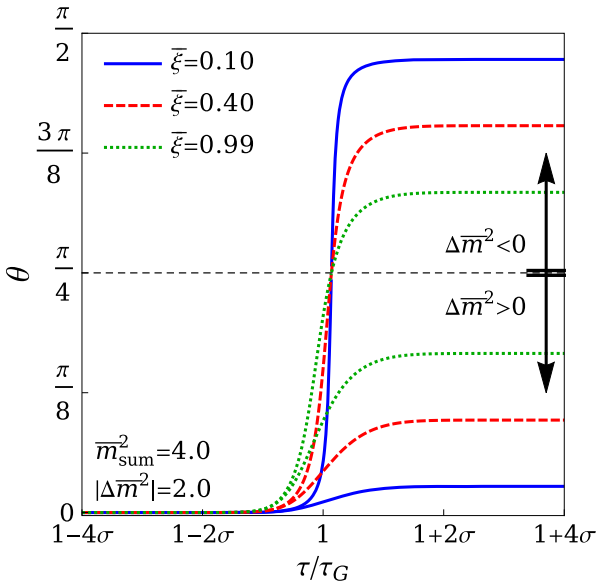

FIG. 8. The mixing angle θ, plotted as a function of time for $\Delta\bar{m}^2 = \pm 2$ and different values of $\bar{\xi}$. For early times $\tau \ll \tau_G$, we generally have $\Delta m^2 = m^2_{\rm sum} = 1$ and $\theta = 0$. The subsequent time dependence then depends critically on the sign of the late-time value $\Delta\bar{m}^2$. If $\Delta\bar{m}^2 < 0$, then Δm^2 eventually switches from positive to negative values; this occurs at $\tau = \tau^*$, where τ^* is defined in Eq. (2.34). At this point, the mixing angle θ passes from the lower "lobe" of Fig. 5 to the upper lobe through the only allowed transition point between the two lobes at $\Delta m^2 = 0$ and $\theta = \pi/4$. For $\Delta\bar{m}^2 > 0$, by contrast, the mixing angle stays entirely within the lower lobe of Fig. 5 and thus remains below $\pi/4$. Finally, for $\Delta m^2 = 0$, the angle θ moves to the central point between the two lobes, consistent with our expectation that $\bar{\theta} = \pi/4$. However, as discussed in the text, it takes infinite time to do this as $\bar{\xi} \to 0$.

However, there is an important subtlety in the latter case. As $\bar{\xi} \to 0$ with $\Delta\bar{m}^2 = 0$, it turns out that $h(\tau_\theta) = 1$. This in turn implies that $\tau_\theta \to \infty$. Thus, even though $\bar{\theta} = \pi/4$ in this case, we never actually reach the point at which the corresponding mixing occurs. Rather, our two original states ϕ_0 and ϕ_1 remain unmixed at all finite times. Indeed, in some sense, the transition to $\bar{\theta} = \pi/4$ occurs only when the two corresponding eigenvalues λ_0 and λ_1 actually meet. Note, in this connection, that even though the field rotation does not occur at any finite time, the mass eigenvalues $\lambda^2_{0,1}$ nevertheless experience their normal evolution in the neighborhood of τ_G, a result which follows directly from the time dependence of the original mass matrix and which is independent of mixing. In other words, in this special case, the two timescales τ_G and τ_θ are maximally separated.

Our discussion thus far has focused on the eigenvalue/mixing structure of the mass matrix as a function of time. However, the corresponding field values also generally behave as one might expect, at least as far as their grossest features are concerned. For times $\tau \ll \tau_\theta$, our two fields $\phi_{\lambda_{0,1}}$ are effectively uncoupled: ϕ_{λ_0} evolves independently of ϕ_{λ_1}, and indeed ϕ_{λ_1} remains vanishing. For times $\tau \sim \tau_\theta$, by contrast, the phase transition generates a nonzero, time-dependent mixing which couples the two fields together and thereby causes ϕ_{λ_1} to accrue a nonzero value as well. Finally, for times $\tau \gg \tau_\theta$, our mixing angle $\bar{\theta}$ is nonzero but essentially constant. This means that a similar decoupling exists during this period as well, except that our decoupled fields are now those linear combinations which are rotated relative to our original fields by $\bar{\theta}$. In general, each of these linear combinations will have a nonzero field value, and will evolve independently according to whether it is individually overdamped or underdamped.

The limiting case with $\Delta\bar{m}^2 < 0$ and $\bar{\xi} = 0$ is again worthy of special note. In this case, ϕ_{λ_0} retains its original amplitude until the eigenvalue meeting time $\tau^* = \tau_\theta$. The entire amplitude of ϕ_{λ_0} then transfers instantaneously to ϕ_{λ_1}, where it remains for all later times. (This follows from the observation that in this limit we have an instantaneous rotation of our fields from $\theta = 0$ to $\theta = \pi/2$.) This rapid amplitude transfer is consistent with the instantaneous infinite value of $\dot{\theta}$ at $\tau^* = \tau_\theta$. As a result of this amplitude transfer, quantities which depend on the field amplitudes (such as the associated energy densities) transfer instantaneously from ϕ_{λ_0} to ϕ_{λ_1} at this time. In other words, they "ride" smoothly across the eigenvalue collision in the left panel of Fig. 7, transitioning from ϕ_{λ_0} to ϕ_{λ_1}. Of course, this is completely as expected, reflecting nothing more than the fact that the energy density began in our original field ϕ_0 and remains there when $\bar{\xi} = 0$. All that has suddenly changed at the eigenvalue collision time is the identification between this field ϕ_0 and our mass-eigenstate fields $\phi_{\lambda_{0,1}}$.

One could, in principle, continue along this line of inquiry. For example, one could proceed to study the phase-space trajectories of our mass-eigenstate fields ϕ_{λ_0} and ϕ_{λ_1}, and map out how these trajectories depend on the different defining parameters of our model. However, an exhaustive study of this toy model is not our purpose in this paper. Rather, as stated in the Introduction, in this paper our interest in this toy model stems from the fact that—despite its simplicity—it gives rise to certain features which have the potential to transcend our typical expectations when interpreted in a cosmological setting. It is therefore to these new features that we now turn.

In keeping with the above observations, in the rest of this paper we shall mostly concentrate on those regions of parameter space in which our fields are already underdamped—or are in the process of *becoming* underdamped—during the mass-generating phase transition. This is important, since the time-dependent effects of our mass-generating phase transition tend to be washed out if our fields remain overdamped while it occurs. Thus, in this way, we shall be focusing on precisely those parameter-space regions of interest: those which are likely

to exhibit a nontrivial interplay between the width of the mass-generating phase transition, the transition between overdamped and underdamped regimes, and the mixing between all of the fields experiencing these effects.

III. TOTAL LATE-TIME ENERGY DENSITY

The total energy density ρ is the quantity of central interest in this paper. The role it plays as a cosmological observable gives it direct importance in any analysis of our toy model—particularly at late times, after our phase transition has been completed. In general, we are interested not only in the total late-time energy density $\bar{\rho}$, but also in its distribution between the two individual components $\bar{\rho}_\lambda$. Moreover, in each case, we are particularly interested in knowing the extent to which our mass-generating phase transition might leave imprints on these late-time energy densities. These are the issues that we shall study in this section.

Calculating the energy density of the system proceeds directly from the equations of motion for our two fields and their derivatives. In the original $\{\phi_0, \phi_1\}$ basis, the total energy density is given by

$$\rho = \frac{1}{2}\left(\sum_i \dot{\phi}_i^2 + \sum_{ij} \phi_i \mathcal{M}_{ij}^2 \phi_j\right). \tag{3.1}$$

By contrast, in the mass-eigenstate basis $\{\phi_{\lambda_0}, \phi_{\lambda_1}\}$ introduced in Eq. (2.23), the total energy density is given by

$$\rho = \frac{1}{2}\sum_\lambda [\dot{\phi}_\lambda^2 + (\lambda^2 + \dot{\theta}^2)\phi_\lambda^2] + \dot{\theta}\sum_{\lambda\lambda'} \dot{\phi}_\lambda \epsilon_{\lambda\lambda'} \phi_{\lambda'}, \tag{3.2}$$

where $\epsilon_{\lambda\lambda'}$ is the Levi-Civita symbol with $\epsilon_{\lambda_0\lambda_1} \equiv +1$.

The first thing we observe is that in neither case can we express our total energy density as a sum of individual contributions. Indeed, we cannot write ρ in the form $\rho = \sum_i \rho_i$ or $\rho = \sum_\lambda \rho_\lambda$. The reason for this, as most evident from Eq. (3.2), is ultimately that our fields experience a *time-dependent* mixing, with $\dot{\theta} \neq 0$. Without mixing—or even with only a constant mixing—such a decomposition could have been done and individual contributions identified. *Thus, in this paper we shall never refer to the individual contributions to the total energy density except at late times when the mixing has essentially stabilized and* $\dot{\theta} = 0$. Indeed, at late times, we can then identify

$$\bar{\rho}_\lambda = \frac{1}{2}(\dot{\phi}_\lambda^2 + \bar{\lambda}^2 \phi_\lambda^2), \tag{3.3}$$

where the overbar indicates late-time values in accordance with Eq. (2.27).

As already discussed at the end of Sec. II, the effects of the mass-generating phase transition are washed out if our fields are still overdamped when it occurs. Likewise, given the formulation of our toy model, there is no meaningful way for the mass-generating phase transition to occur very much later than the critical overdamped/underdamped transition for either of our two fields: the lighter field is massless prior to the mass-generating phase transition and as such remains overdamped during this entire period, while the heavier field, either formally overdamped or underdamped, has no amplitude of its own until the mass-generating phase transition. For this reason, in this section we shall focus on those regions of parameter space in which our fields become underdamped near the time at which the phase transition begins.

The most complete way of surveying physics within this regime is to vary both τ_G and Δ_G, with τ_G limited to the range $\tau_G \gtrsim \tau_\zeta^{(i)}$, where each $\tau_\zeta^{(i)}$ is implicitly defined (as in Fig. 2) as the time at which $3H = 2\lambda_i$. This method, which by construction surveys all possibilities, thus involves variations in two independent parameters. However, another way of surveying many aspects of the physics within this regime is to fix τ_G, and instead to vary the width of the phase transition Δ_G. By considering fiducial values of τ_G which are carefully chosen with respect to the $\bar{m}_{ij}^2$ values, we can reach situations in which $\tau_\zeta^{(i)}$ are near τ_G or just prior to it. These values of $\tau_\zeta^{(0)}$ are illustrated in Fig. 9, and we see that this method also allows us to reach the desired values of $\tau_\zeta^{(0)}$. It is therefore this latter approach that we shall follow in the next two sections.

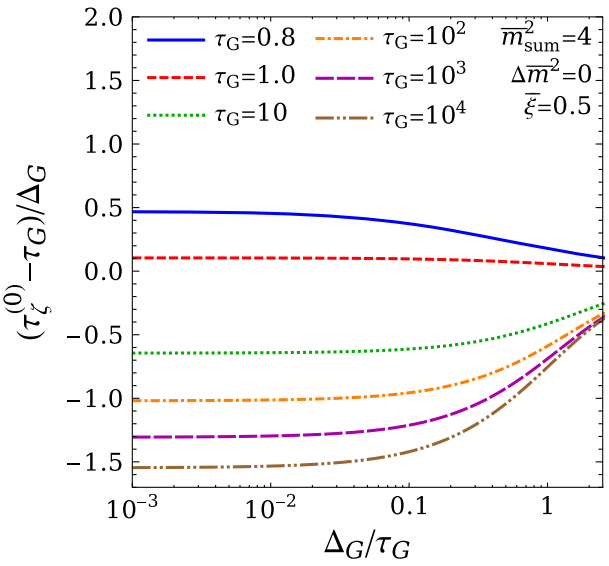

FIG. 9. The critical damping time $\tau_\zeta^{(0)}$ for the lighter field ϕ_{λ_0} in a matter-dominated universe, expressed as a number of widths Δ_G prior to the central phase-transition time τ_G and plotted as a function of Δ_G/τ_G for different values of τ_G. For $\tau_G \sim 1$, the critical damping transition essentially occurs during the central portion of the mass-generating phase transition for all Δ_G; for large τ_G, by contrast, the critical damping transition occurs just prior to the mass-generating phase transition. These two cases thus survey our main regions of interest in this paper.

Given our choices for τ_G and Δ_G, we can then calculate the late-time total energy density $\bar{\rho}$ for different values of the $\bar{m}_{ij}^2$ parameters—i.e., for different values of $\bar{m}_{\rm sum}^2$, $\Delta\bar{m}^2$, and $(\bar{\xi}, \bar{\theta}_{\rm max})$. Our goal is to understand the effects on the total late-time energy densities $\bar{\rho}$ that emerge when a nonzero mixing, parametrized by $\bar{\xi}$, and a nonzero width Δ_G for our mass-generating phase transition are present simultaneously.

Our results are shown in Figs. 10–12: In Figs. 10 and 11 we assume $\tau_G = 1$ (for radiation- and matter-dominated universes respectively), while in Fig. 12 we assume $\tau_G \gg 1$ (for which matter- and radiation-dominated universes yield the same results). In each of these figures we survey values of the late-time mixing saturation $\bar{\xi}$ within the range $0 \leq \bar{\xi} < 1$; note that we refrain from considering the actual limiting case with $\bar{\xi} = 1$ when discussing late-time quantities such as the late-time energy density $\bar{\rho}$ because the lighter field ϕ_{λ_0} remains massless at late times when $\bar{\xi} = 1$ (as is evident from Fig. 6) and therefore never technically enters the asymptotic underdamped region which characterizes our definition of "late" times. We also note that although the left, center, and right panels within each figure correspond to different values of the maximum allowed mixing angle $\bar{\theta}_{\rm max}$, we see that it is only in terms of the mixing *saturation* $\bar{\xi}$ that we can make sensible comparisons across the different panels. Indeed, although a fixed change in $\bar{\xi}$ within a panel with large $\bar{\theta}_{\rm max}$ corresponds to a much larger change in the absolute mixing angle $\bar{\theta}$ than it does within a panel with small $\bar{\theta}_{\rm max}$, we see that it is only the changes in $\bar{\xi}$ rather than $\bar{\theta}$ which can be compared meaningfully across panels. For this reason, even though the value of $\theta_{\rm max}$ is important in order to express our results in terms of absolute mixing angles, in the following we shall simply describe our mixings as small or large depending on the corresponding values of $\bar{\xi}$.

These figures illustrate the effects of turning on a finite width Δ_G for the mass-generating phase transition in conjunction with nonzero mixing between our two fields. For the small-τ_G regime plotted in Figs. 10 and 11, we observe that *small* mixing actually *enhances* the late-time energy density $\bar{\rho}$. By contrast, we see that *large* mixing actually *suppresses* the late-time energy density $\bar{\rho}$. Indeed, we see from Figs. 10 and 11 that these effects (in both directions) are more pronounced for matter-dominated universes than radiation-dominated universes and for situations in which $\Delta\bar{m}^2$ is positive rather than zero or negative.

For the large-τ_G regime plotted in Fig. 12, by contrast, we see that matter- and radiation-dominated universes give rise to identical results. All late-time energy densities $\bar{\rho}$, regardless of the mixing saturation $\bar{\xi}$, share a common value when $\Delta_G = 0$. However, unlike the small-τ_G case, the late-time energy densities experience no enhancement at all, even in the absence of mixing ($\bar{\xi} = 0$). Indeed, $\bar{\rho}$ experiences only a suppression for nonzero Δ_G—a suppression which, given the logarithmic $\bar{\rho}$ axes within Fig. 12, grows much more severe than it was for small τ_G. Moreover, we see that this suppression of the late-time energy density $\bar{\rho}$ is stronger for *negative* $\Delta\bar{m}^2$ (i.e., for negative $\bar{\eta}$) than for zero or positive—a feature which is completely reversed relative to the small-τ_G case! Finally, we observe the emergence of a nonmonotonic "oscillatory" behavior for $\bar{\rho}$ as a function of Δ_G when τ_G is large and when the mixing saturation $\bar{\xi}$ grows close to 1. This feature is most pronounced when $\Delta\bar{m}^2$ is positive, but exists for all $\Delta\bar{m}^2$. These oscillations will be discussed in Sec. V, and indicate a strong sensitivity of the suppression factor relative to even small variations in the phase-transition width Δ_G.

All of these results illustrate the dramatic consequences that ensue when we consider the timescales associated with both our mass-generating phase transition and the mixing it generates. For example, we see from the $\Delta\bar{m}^2 > 0$ plot in Fig. 11 that mixing has no effect on the total late-time energy density $\bar{\rho}$ when the mass-generating phase transition is rapid (i.e., when $\Delta_G = 0$). *Thus, it is only the existence of a nonzero phase-transition timescale which allows the mixing to leave a nontrivial imprint at late times!* Moreover, we see from these figures that the enhancements and suppressions experienced by $\bar{\rho}$ are typically quite large, stretching from 20% or 30% in the case of enhancements all the way to many orders of magnitude in the case of suppressions! These effects can thus have dramatic implications for the relative size of the corresponding slice of the "cosmic pie"—i.e., for the overall composition of the total late-time energy budget of the universe.

It is also instructive to understand those limits of the above results in which one or the other of our two variables $\bar{\xi}$ and Δ_G is taken to zero. We begin by focusing on the effects that emerge solely due to the presence of a nonzero width Δ_G for the mass-generating phase transition (i.e., the effects that occur when $\bar{\xi} = 0$). Results are shown in Fig. 13, where we plot contours of the total normalized late-time energy density $\bar{\rho}(\sigma)/\bar{\rho}(0)$ within the (τ_G, σ) plane, where $\sigma \equiv \Delta_G/(\sqrt{\pi}\tau_G)$ as in Eq. (2.13). Generally, we see from Fig. 13 both suppression and enhancements in the late-time energy density are possible, depending on whether $\tau_G \gtrsim 5$ or $\tau_G \lesssim 1$ respectively. These results therefore explain the behaviors of the $\bar{\xi} = 0$ curves in Figs. 10, 11, and 12. Furthermore, we see from Fig. 13 that $\bar{\rho}(\Delta_G)$ is largely insensitive to σ, and instead depends almost exclusively on the value of τ_G. Indeed, for $\tau_G \gtrsim 1$, we see from Fig. 13 that $\bar{\rho}(\Delta_G)$ scales approximately as τ_G^{-1}. This in turn implies that $\bar{\rho}(\Delta_G) \sim \Delta_G^{-1}$ in this region. This observation will be discussed further in Sec. IV.

Conversely, we can also study the effects that arise due to mixing alone (i.e., with $\Delta_G = 0$). It turns out that when $\Delta_G = 0$, the late-time energy density $\bar{\rho}$ can actually be calculated *analytically* for arbitrary mixing; the general result is given in Eq. (A8). In Fig. 14, we choose values $\Delta\bar{m}^2 = 3$ and $\bar{m}_{\rm sum}^2 = 4$, and plot the corresponding late-time energy density $\bar{\rho}(\bar{\xi})$ as a function of the late-time mixing saturation $\bar{\xi}$, normalized to its unmixed value $\bar{\rho}(\bar{\xi} = 0)$, for a variety of different τ_G. For $\bar{\xi} = 0$, this

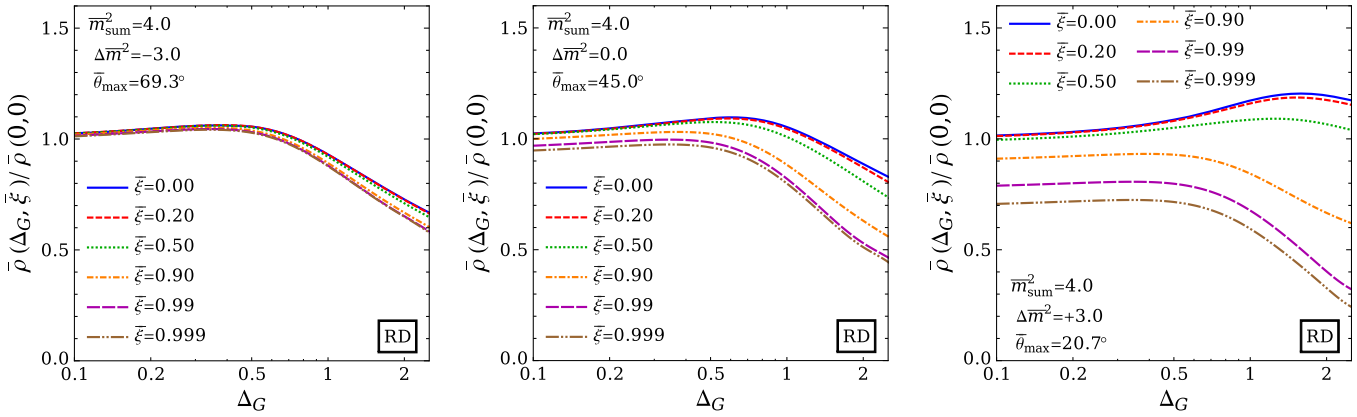

FIG. 10. Behavior of the total late-time energy density $\bar{\rho}(\Delta_G,\bar{\xi})$ when mixing effects, parametrized by $\bar{\xi}$, are combined with a nonzero width Δ_G for the mass-generating phase transition. For these plots we adopt a relatively *small* fiducial time $\tau_G = 1.0$ and assume a radiation-dominated universe; we also hold $\bar{m}^2_{\rm sum} = 4$ and take $\Delta\bar{m}^2 = -3$, 0, $+3$ for the left, center, and right panels, respectively. In each case we plot the late-time energy density $\bar{\rho}$ as a function of the phase-transition width Δ_G for different values of the mixing saturation $\bar{\xi}$, where in each panel $\bar{\rho}$ is normalized to its value for $\Delta_G = 0$ and $\bar{\xi} = 0$.

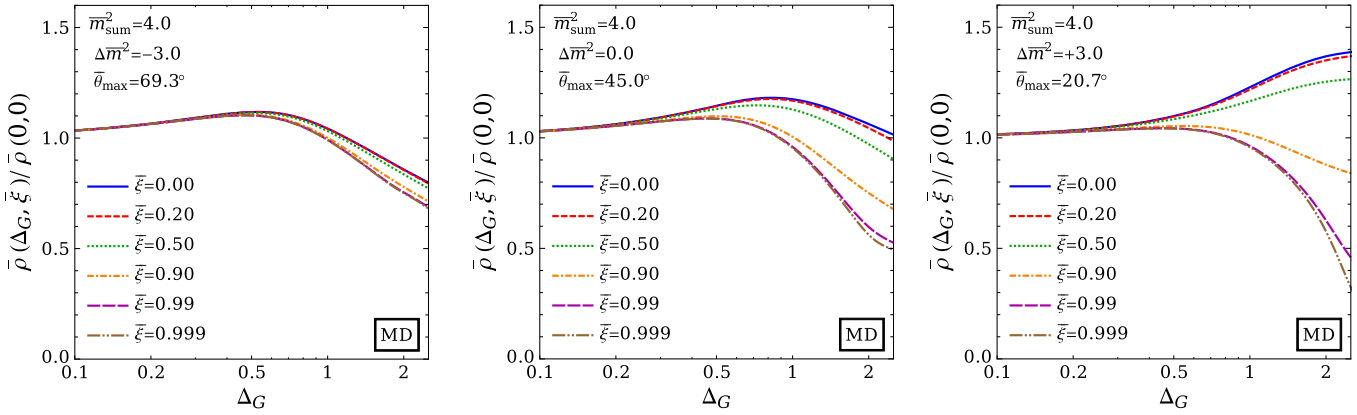

FIG. 11. Same as Fig. 10, but for a matter-dominated universe.

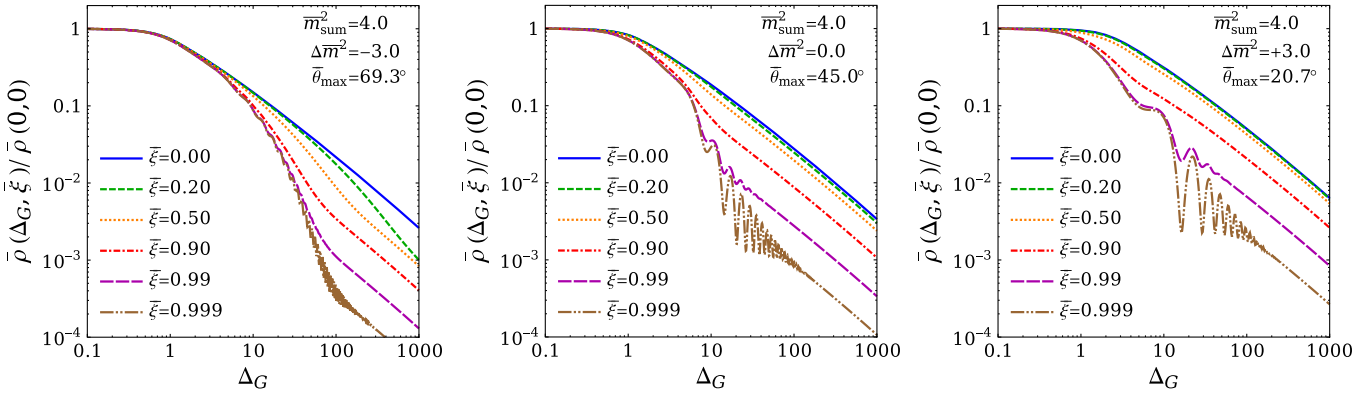

FIG. 12. Same as Figs. 10 and 11, except that we now adopt a relatively *large* fiducial time $\tau_G = 10^4$. For such large values of τ_G, the results are nearly the same for both radiation- and matter-dominated universes. Note that unlike Figs. 10 and 11, the vertical axes here are *logarithmic*.

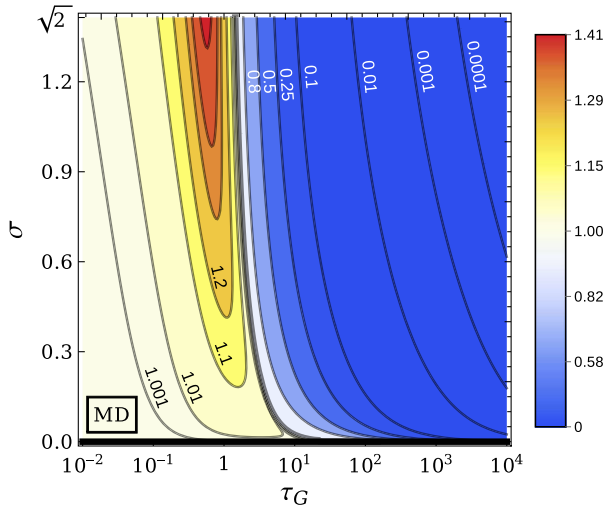

FIG. 13. Contours of the total late-time energy-density enhancement/suppression factor $\bar{\rho}(\Delta_G)/\bar{\rho}(0)$, plotted in the (τ_G, σ) plane, assuming a matter-dominated universe without mixing ($\bar{\xi} = 0$) and $\bar{m}_{00}^2 = 1$. The thick black line running vertically and along the $\sigma = 0$ axis is the contour for which we have neither enhancement nor suppression. The corresponding contours for a radiation-dominated universe are similar. We observe that in the absence of mixing, a finite width for the mass-generating phase transition tends to modify the late-time energy density $\bar{\rho}(\sigma)$ compared to what it would have been for an instantaneous phase transition; this suppresses $\bar{\rho}(\sigma)$ rather significantly for $\tau_G \gtrsim 5$, but enhances $\bar{\rho}(\sigma)$ for $\tau_G \lesssim 1$. Remarkably, these results are generally insensitive to σ as long as $\sigma \gtrsim 0.1$, and follow an approximate scaling behavior $\bar{\rho} \sim \Delta_G^{-1}$ in this region.

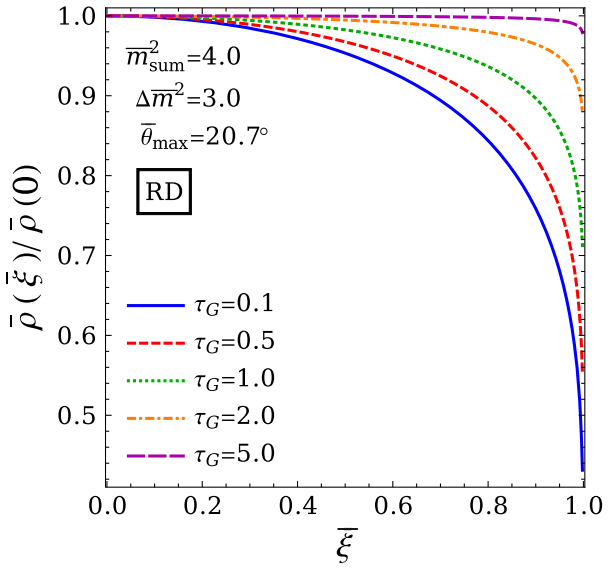

FIG. 14. Suppression of the late-time energy density $\bar{\rho}$ due to mixing effects alone, with $\Delta_G = 0$, in a radiation-dominated universe. We assume an instantaneous phase transition at $\tau = \tau_G$, and plot the corresponding late-time energy density $\bar{\rho}(\bar{\xi})$ as a function of the late-time mixing saturation $\bar{\xi}$, normalized to its value at $\bar{\xi} = 0$, for different choices of τ_G. We see that in all cases the late-time energy density $\bar{\rho}(\bar{\xi})$ experiences a suppression which grows increasingly severe as the mixing saturation is increased. As discussed in the text, the magnitude of this effect increases as $\bar{\eta} \equiv \Delta\bar{m}^2/\bar{m}^2_{\rm sum} \to 1$ (i.e., as $\bar{m}^2_{00} \to 0$), and is entirely absent if the phase transition instead occurs during a matter-dominated epoch or more generally if $\tau_G \gg 1$.

corresponds to a situation in which our two original unmixed fields $\phi_{0,1}$ remain uncoupled, with all of the energy density remaining in ϕ_0; by contrast, as $\bar{\xi}$ increases, increasing amounts of energy density are shared between our two fields during the mass-generating phase transition. Surprisingly, we see from Fig. 14 that increasing $\bar{\xi}$ results in an *enhanced dissipation* of the total energy density, so that the total late-time energy density $\bar{\rho}(\bar{\xi})$ is suppressed relative to what it would have been in the absence of mixing. Indeed, we see from Fig. 14 that this suppression is strongest for relatively small τ_G and relatively large mixing saturations $\bar{\xi}$. Moreover, as evident from the calculation in Appendix A, this effect exists in all cases except for a matter-dominated epoch. This effect was first observed in Ref. [35] within the context of an infinite tower of Kaluza-Klein axion modes, where it was exploited in order to permit a higher-dimensional loosening of the usual four-dimensional overclosure bounds on the Peccei-Quinn scale $f_{\rm PQ}$. However, we now see that this effect is completely general, and persists even when only two modes are involved.

We thus conclude that for phase transitions occurring during a radiation-dominated epoch, mixing alone induces a significant suppression in the late-time energy density. Indeed, this is true even if the phase transition is treated as instantaneous. This effect grows in significance as $\Delta\bar{m}^2 \to \bar{m}^2_{\rm sum}$ *and* $\bar{\xi} \to 1$. *This effect is also largest when* $\tau_G \sim \tau_\zeta^{(i)}$, *and ultimately vanishes for* $\tau_G \gg \tau_\zeta^{(i)}$.

Note that all of these observations are consistent with the plots shown in Figs. 10, 11, and 12. For $\Delta_G = 0$, we see that all of the plotted curves begin at $\bar{\rho}(\Delta_G, \bar{\xi})/\bar{\rho}(0,0) = 1$ *except* for those in Fig. 10, where the mere fact of having $\bar{\xi} \neq 0$ induces a suppression of the late-time energy density, even for $\Delta_G = 0$.

Finally, along the same lines, it is interesting to contemplate what happens for more general universes beyond those that are radiation-dominated. The corresponding results for universes with general values of κ are shown in Fig. 15. We see that mixing alone indeed produces a suppression of the late-time energy density for universes with $\kappa < 2$, but this suppression actually becomes an *enhancement* for universes with $\kappa > 2$! Moreover, as expected, both effects become stronger as the mixing saturation $\bar{\xi}$ increases—strong enough to change the late-time energy density by factors of two or three or even more. Indeed, it is only for a matter-dominated universe (corresponding to $\kappa = 2$) that this effect disappears.

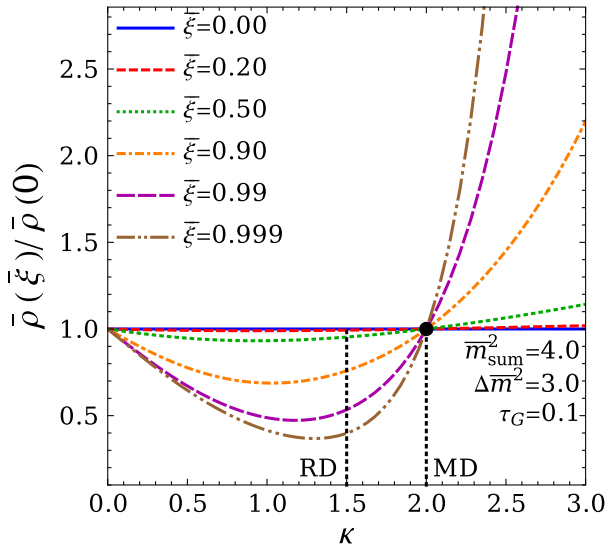

FIG. 15. Suppression of the late-time energy density $\bar{\rho}$ due to mixing effects alone, with $\Delta_G = 0$, for arbitrary universes parametrized by κ. As in Fig. 14, we assume an instantaneous phase transition at $\tau = \tau_G$, but we now plot the corresponding late-time energy density $\bar{\rho}(\bar{\xi})$ as a function of κ, normalized to its value at $\bar{\xi} = 0$, for different choices of $\bar{\xi}$. We see that mixing causes a *suppression* of the late-time energy density within $\kappa < 2$ universes but an *enhancement* within $\kappa > 2$ universes—indeed, these effects disappear only in the special case of a matter-dominated universe. These effects can therefore produce significant modifications of the final late-time energy density, even for instantaneous mass-generating phase transitions.

IV. INDIVIDUAL LATE-TIME ENERGY DENSITIES

In the previous section, we focused on only one quantity: the total late-time energy density $\bar{\rho}$. However, another important feature is the actual *distribution* of this total energy density $\bar{\rho}$ amongst the two fields of our system. We shall now proceed to study this issue.

In Fig. 16, we show the behavior of the individual late-time energy densities associated with ϕ_{λ_0} (the lighter field) and ϕ_{λ_1} (the heavier field) for a mass-generating phase transition occurring at $\tau_G = 10^4$. It is therefore the sum of these two energy densities which produces the results in Fig. 12; recall that these results are the same for matter- and radiation-dominated universes. It turns out that there are many features illustrated within Fig. 16 which will be important for our future results. We shall therefore step through these features, one by one.

We begin by concentrating on the special case without mixing (i.e., with $\bar{\xi} = 0$). This will allow us to study the effects of the nonzero phase-transition width Δ_G for a single component alone without the complications due to mixing. For $\Delta\bar{m}^2 \geq 0$, all of the energy density arising due to the mass-generating phase transition accrues to the lighter field ϕ_{λ_0}. For $\Delta\bar{m}^2 < 0$, by contrast, all of the late-time energy density is associated with the heavier field ϕ_{λ_1}, as discussed at the end of Sec. II.

In either case, we observe from Fig. 16 that the corresponding late-time energy density $\bar{\rho}_\lambda$ remains essentially constant (i.e., independent of Δ_G) for $\Delta_G \lesssim 2\pi/\lambda$. This makes sense, as the effects of introducing a nonzero width for our phase transition will be essentially invisible if the mass generation occurs more rapidly than the natural timescale of oscillations of our (underdamped) field ϕ_λ. In other words, for $\Delta_G \lesssim 2\pi/\lambda$, the process of pumping energy density into our system occurs with what may be considered to be maximum efficiency, since the phase transition appears to be effectively instantaneous with respect to the natural oscillation timescale. However, for $\Delta_G \gtrsim 2\pi/\lambda$, the field oscillations tend to compete against the process of mass generation. As a result, the mass-generating phase transition is less efficient in pumping energy density into our system, thereby inducing a suppression in the late-time energy $\bar{\rho}_\lambda$ which ultimately scales as an inverse power of Δ_G:

$$\bar{\rho}_\lambda(\Delta_G, 0) \sim 1/\Delta_G \quad \text{for } \Delta_G \gg 2\pi/\lambda. \tag{4.1}$$

Thus, once Δ_G exceeds the natural oscillation period of our mass eigenstate, increasing the width of the phase transition has the effect of introducing a power-law suppression of the corresponding late-time energy density.

These results hold for the unmixed scenarios with $\bar{\xi} = 0$. However, for the lighter fields ϕ_{λ_0} (corresponding to the top row of Fig. 16), we see that the above asymptotic behavior continues to hold *regardless of the value of the mixing*:

$$\bar{\rho}_{\lambda_0}(\Delta_G, \bar{\xi}) \sim 1/\Delta_G \quad \text{for } \Delta_G \gg 2\pi/\lambda_0. \tag{4.2}$$

Thus, we see that increasing the width of the phase transition continues to be associated with a power-law suppression of the late-time energy density of the lighter field—even in the presence of nonzero mixing.

Turning on a mixing between our two fields also has a number of other important effects on their individual late-time energy densities. For the lighter fields ϕ_{λ_0} (as considered along the top row of Fig. 16), the most prominent effect is of course the set of very strong oscillations which are induced for very large mixing saturations $\bar{\xi} \lesssim 1$ and "intermediate" widths $1 \lesssim \Delta_G \lesssim 100$. We shall defer our discussion of these oscillations until the next section, but we see from Fig. 16 that these oscillations are relatively large, occasionally enhancing the corresponding late-time energy densities above what they would have been in the absence of mixing *by several orders of magnitude*! Indeed, this enhancement of the corresponding late-time energy densities above their $\bar{\xi} = 0$ values persists even for values of Δ_G which lie beyond the actual oscillations themselves.

As a result, we see that in the presence of a nonzero width for the phase transition, mixing has a general tendency to *enhance* the late-time energy density of the

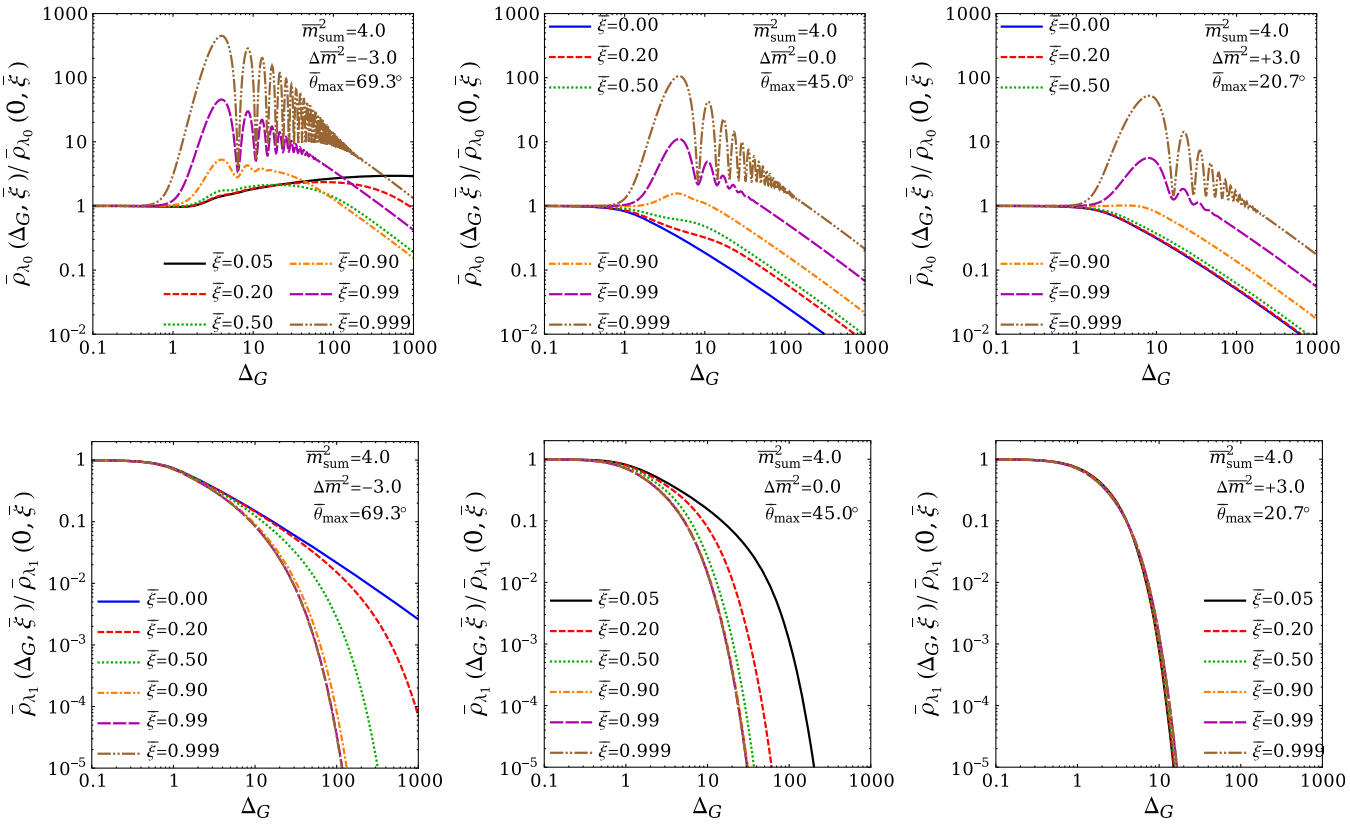

FIG. 16. Behavior of the *individual* late-time energy densities $\bar{\rho}_{\lambda_{0,1}}(\Delta_G,\bar{\xi})$ when mixing effects, parametrized by $\bar{\xi}$, are combined with a nonzero width Δ_G for the mass-generating phase transition. For these plots we adopt the large fiducial time $\tau_G = 10^4$ utilized for Fig. 12, which leads to identical results for radiation- and matter-dominated universes. The top row shows results for $\bar{\rho}_{\lambda_0}$ (associated with the lighter state in all cases), while the bottom row shows results for $\bar{\rho}_{\lambda_1}$ (associated with the heavier state); we have also held $\bar{m}^2_{\rm sum} = 4$ and take $\Delta\bar{m}^2 = -3$, 0, $+3$ for the left, center, and right columns, respectively. In each case we plot the late-time energy density $\bar{\rho}_\lambda$ as a function of the phase-transition width Δ_G for different values of the mixing saturation $\bar{\xi}$, normalized to its value for $\Delta_G = 0$. Note that we refrain from plotting curves for $\bar{\xi} = 0$ (and instead plot curves for $\bar{\xi} = 0.05$) in cases where the corresponding energy densities vanish.

lighter field—all without disturbing the power-law suppression discussed above. Indeed, for the lighter fields ϕ_{λ_0}, the effects of the nonzero transition width Δ_G and the nonzero mixing saturation $\bar{\xi}$ tend to pull in opposite directions: the width tends to suppress the corresponding late-time energy density while the mixing tends to enhance it. As evident from the plots along the top row of Fig. 16, the nature of the net result (either an overall enhancement or overall suppression) therefore depends nontrivially on the precise values of Δ_G and $\bar{\xi}$ involved.

By contrast, for the *heavier* fields ϕ_{λ_1} (for which the corresponding energy densities are plotted along the bottom row of Fig. 16), the effects of nonzero mixing are quite the opposite: the enhancement discussed above is gone, and instead there is now an *additional suppression* which helps to drive the corresponding energy densities to even smaller values, as functions of Δ_G, than we had for the lighter fields!

It is important to understand the origins of this additional suppression. Unlike the lighter field ϕ_{λ_0}, which obtains its energy density directly from the phase transition through the generation of a nonzero mass, the heavier field ϕ_{λ_1} actually gains its energy density when it gains an overall field amplitude through mixing with the lighter field. The magnitude of the resulting energy density is then governed by two factors: the magnitude of the field amplitude generated, and the masses that are also generated by the phase transition.

Thus, while the natural timescale governing the rate at which the energy density is originally pumped into the lighter field is given directly by Δ_G—and thus by the function $h(\tau)$ plotted in Fig. 3—the natural timescale governing the rate at which the energy density is pumped into the heavier field is governed not only by Δ_G but also by the rate of change of the *mixing angle* θ plotted in Fig. 8. In analogy with Eq. (2.12), we may even define a corresponding width Δ_θ for the mixing angle θ via the slope of $\theta(\tau)$ at its midpoint $\bar{\theta}/2$:

$$\Delta_\theta \equiv (\bar{\theta}/\dot{\theta})|_{\tau_\theta}; \tag{4.3}$$

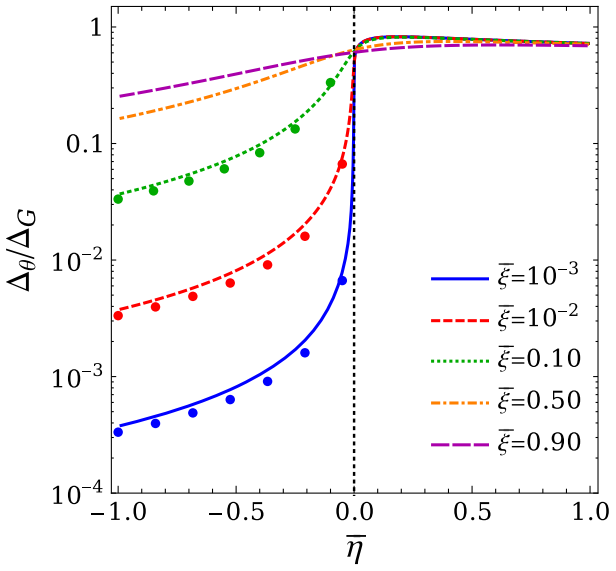

FIG. 17. The ratio of the two timescales Δ_G and Δ_θ, plotted as a function of $\bar{\eta}$ for different mixing saturations $\bar{\xi}$. In all cases, we see that $\Delta_\theta < \Delta_G$, with $\Delta_\theta \ll \Delta_G$ for $\bar{\eta} < 0$ and $\bar{\xi} \ll 1$. The dots represent the approximate analytical result in Eq. (4.4). The distinction between Δ_θ and Δ_G has important implications for the energy density associated with the heavier field ϕ_{λ_1}, since Δ_G governs the rate at which the *mass* of this field is modified during the phase transition while Δ_θ governs the rate at which the *amplitude* of this field is generated as a result of its mixing with the lighter field.

here τ_θ, as defined in Eq. (2.33), is defined analogously to τ_G as the time at which $\theta = \bar{\theta}/2$. A rough analytical approximation for Δ_θ when $\bar{\eta} < 0$ is given by

$$\Delta_\theta \approx \frac{\bar{\xi}}{3|\bar{\eta}|}\Delta_G \quad \text{for } \bar{\eta} < 0. \tag{4.4}$$

The magnitudes of Δ_G and Δ_θ are compared in Fig. 17, from which we see that Δ_θ never exceeds Δ_G. Indeed, we see from Fig. 17 that Δ_θ can be much smaller than Δ_G for $\bar{\eta} < 0$ and $\bar{\xi} \ll 1$. We thus see that during our mass-generating phase transition, the mixing angle θ always changes at least as rapidly as do the eigenvalues λ_i, and can in fact under certain circumstances change even more rapidly, the latter despite the fact that both sets of changes arise due to the same phase transition.

This observation is directly relevant for the manner in which the heavier field accrues its energy density from the mass-generating phase transition. As such, this feature is thereby directly relevant for the resulting behavior of the late-time energy density $\bar{\rho}_{\lambda_1}$. There are three distinct cases to consider.

(i) $\Delta_G \ll 2\pi/\lambda_1$: In this case, the oscillations of the heavier field are so slow during the phase transition that all aspects of the phase transition appear effectively instantaneous. This is true for the generation of the *amplitude* of the heavier field as well as the change in its *mass*. Energy density is thereby delivered to the heavier field with maximum efficiency, in a manner which is completely insensitive to the nonzero timescales Δ_θ and Δ_G associated with the phase transition and which is thereby protected against all of their associated dissipating effects. This behavior is evident for sufficiently small Δ_G within the plots along the lower row of Fig. 16.

(ii) $\Delta_\theta \lesssim 2\pi/\lambda_1 \lesssim \Delta_G$: In this case, the oscillations of the heavier field are sufficiently slow that the generation of its field amplitude appears to be effectively instantaneous; these oscillations are nevertheless sufficiently rapid that full oscillations can occur during the change in field mass. The net result is that the heavy-field energy density $\bar{\rho}_{\lambda_1}$ experiences the same suppression as does the light-field energy density $\bar{\rho}_{\lambda_0}$, with both depending nontrivially on Δ_G and ultimately scaling inversely with Δ_G. Note that this case exists only for situations in which Δ_θ is significantly smaller than Δ_G, which according to Fig. 17 tend to emerge only for $\Delta\bar{m}^2 < 0$ and relatively small mixing.

(iii) $2\pi/\lambda_1 \lesssim \Delta_\theta$: In this case, the field oscillations of the heavier field are sufficiently rapid that full oscillations are occurring not only during the change in its mass *but also during the generation of its amplitude*. This latter feature leads to an additional source of suppression for the late-time energy density beyond what the lighter field experiences, and causes $\bar{\rho}_{\lambda_1}$ to exhibit an even more dramatic suppression as a function of Δ_G than that exhibited by $\bar{\rho}_{\lambda_0}$. This behavior is clearly evident for sufficiently large Δ_G in the plots along the lower row of Fig. 16.

The above discussion describes the behaviors of the individual components $\bar{\rho}_\lambda$ as functions of Δ_G and $\bar{\xi}$. However, our final task is to determine how much each of these individual components contributes to the total energy density $\bar{\rho}$. Of course, to do so requires that we understand not just the intrinsic behaviors of these individual components as functions of Δ_G and $\bar{\xi}$, but also the relative sizes of these individual components. In other words, we need to understand the *relative normalization* of the curves for $\bar{\rho}_{\lambda_0}$ in the top row of Fig. 16 relative to those for $\bar{\rho}_{\lambda_1}$ in the bottom row. However, this relative normalization may easily be determined in the $\Delta_G = 0$ limit, for which analytical results are given in Appendix A. Indeed, use of the virial approximation $\lambda_i\tau_G \gg 1$, which is valid for the plots in Fig. 16, yields the ratio

$$\frac{\bar{\rho}_{\lambda_0}(0,\bar{\xi})}{\bar{\rho}(0,\bar{\xi})} = \frac{\bar{\lambda}_0^2\cos^2\bar{\theta}}{\bar{\lambda}_0^2\cos^2\bar{\theta} + \bar{\lambda}_1^2\sin^2\bar{\theta}} = \cos^2\bar{\theta}\left(\frac{1-\bar{\eta}\sec 2\bar{\theta}}{1-\bar{\eta}}\right) \tag{4.5}$$

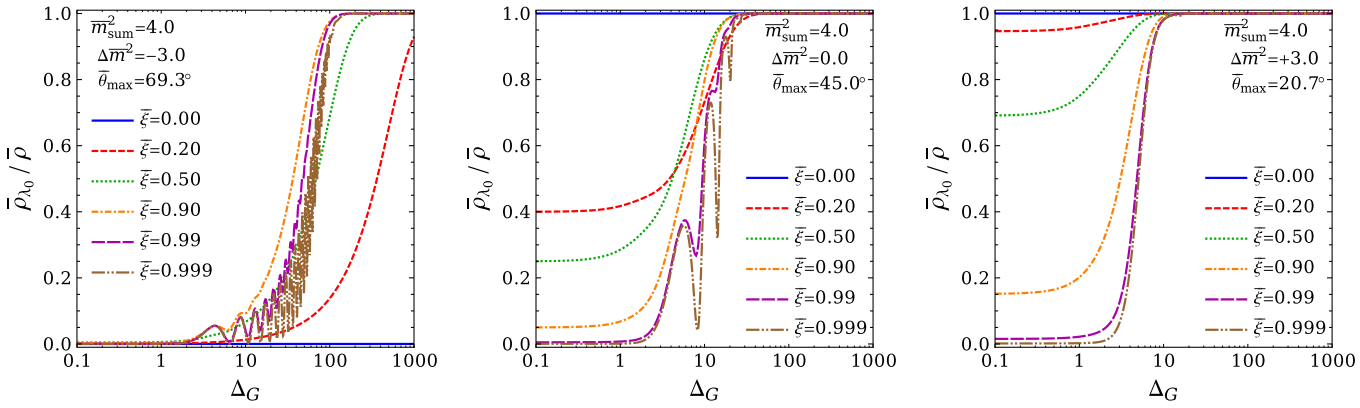

FIG. 18. The fraction of the total late-time energy density $\bar{\rho}(\Delta_G, \xi)$ which is associated with the lighter field ϕ_{λ_0}, plotted as functions of Δ_G for different values of ξ, with $\tau_G = 10^4$. Regardless of the small-Δ_G values of this fraction [as given analytically in Eq. (4.5)], we see that increasing the width Δ_G of the phase transition generally has the net effect of transferring more and more of the total energy density into the lighter field. Thus, simply by adjusting the width of the phase transition, we see that we can often entirely reverse the distribution of the late-time energy density.

as the fraction of the total late-time energy density remaining in the lighter field. As expected, this quantity evolves monotonically from unity at $\theta = 0$ to zero at $\theta = \theta_{\rm max}$ [where $\theta_{\rm max}$ is defined in Eq. (2.26)]. It may thus be greater or less than 50%, depending on the eigenvalues and mixing angles in question. Note that in the limit of small mixing (which corresponds to $\theta \approx 0$ or $\theta \approx \pi/2$), the result in Eq. (4.5) implies that whether the bulk of the energy density winds up at late times associated with the lighter or heavier field depends simply on the sign of η rather than its magnitude. (Recall, in this connection, that $\eta \to -\eta$ implicitly changes $\theta \to \pi/2 - \theta$.)

The result in Eq. (4.5) holds only for $\Delta_G = 0$. However, as Δ_G grows larger, this ratio will change. For the total energy densities plotted in Fig. 12, the corresponding fractional results are shown in Fig. 18 as a function of Δ_G. It is easy to understand the behavior shown in Fig. 18. Regardless of the fraction of the total late-time energy density associated with the lighter field ϕ_{λ_0} when $\Delta_G = 0$, we have already seen that the energy densities associated with the heavier fields are typically more suppressed as a function of Δ_G than are the energy densities associated with the lighter fields. Thus, *regardless of the late-time energy-density configuration when the phase transition is instantaneous, we see that increasing the width of the phase transition has the effect of throwing an increasingly large share of the total late-time energy density into the lighter field*. Indeed, in some cases, *we see that we can entirely reverse the distribution of the total energy density from the heavier field to the lighter field, simply by adjusting the timescale over which the phase transition occurs*!

V. PARAMETRIC RESONANCE

We now turn to what is perhaps the most prominent feature within the plots shown in Figs. 12, 16, and 18: the appearance of nonmonotonicities in the late-time energy densities ρ_λ associated with the lighter fields ϕ_{λ_0} as functions of the phase-transition width Δ_G. As indicated in Sec. III, these oscillations grow particularly large when τ_G is large and when the mixing saturation $\bar{\xi}$ grows close to 1. These nonmonotonicities indicate that within this region of parameter space there exists an extremely strong sensitivity of the late-time energy density $\bar{\rho}_{\lambda_0}$ to even small variations in the phase-transition width Δ_G. Moreover, as evident from Fig. 16, these nonmonotonicities can often enhance $\bar{\rho}_{\lambda_0}$ by several orders of magnitude!

These nonmonotonicities are yet another consequence of the interplay between the width of our phase transition and the mixing it induces. It is not difficult to understand the origin of these oscillations: ultimately they are *parametric resonances* triggered by our mass-generating phase transition. Recall that a parametric resonance generically occurs when the mass of an oscillator itself exhibits an oscillatory behavior whose frequency is approximately twice the natural frequency of the oscillator. (An example of this is a child "pumping" a swing by alternatively standing and squatting on the swing seat at the correct moments during the period of the swing.) In such systems, the amplitude of oscillation grows exponentially and without bound unless the oscillator also has a frictional damping term. A similar behavior can also emerge if the friction term (rather than the mass term) experiences the oscillatory behavior.

At first glance, it may seem that our system of coupled scalar fields has no means of experiencing a parametric resonance. The equations of motion of our model are given in Eq. (2.7), and although both the Hubble damping term $H(\tau)$ and the mass matrix $\mathcal{M}^2_{ij}$ in our model are time-dependent, the Hubble term is monotonically falling. Likewise, the $h(\tau)$-function—which governs the time

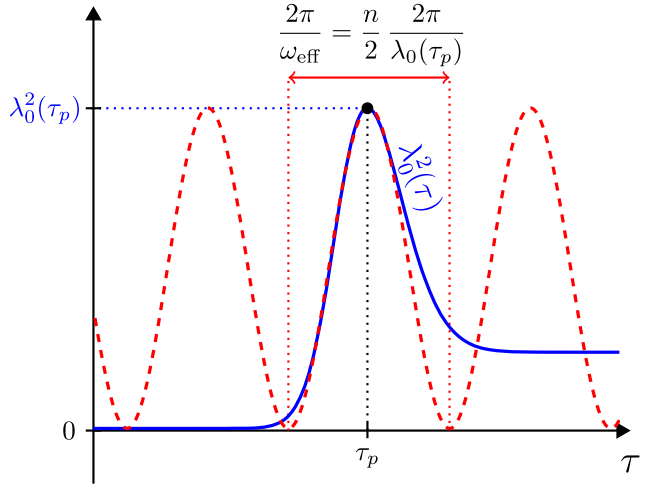

FIG. 19. A schematic illustration of the "pulse" (solid blue curve) experienced by the smaller eigenvalue $\lambda_0(\tau)$ near $\tau \approx \tau_p$ and a full sinusoidal function (dashed red curve) to which it may be approximated in the vicinity of the pulse. As discussed in the text, this approximation allows us to extract the effective frequency $\omega_{\rm eff}$ associated with the pulse and thereby determine the mathematical condition under which an nth-order parametric resonance occurs.

development of our mass matrix, as in Eq. (2.9)—is also monotonic.

However, this is where the mixing plays a critical role. Because of the nonzero mixing inherent in this system, the aspects of the mass matrix which are important are not its individual components m_{ij}^2 but rather its *eigenvalues* λ_i. Indeed, even though the mass-matrix components $m_{ij}^2(\tau)$ are all monotonic as functions of time τ, the eigenvalues λ_i need not be. These eigenvalues were plotted as functions of time in Fig. 7, and we see that for the smaller eigenvalue λ_0 the mixing induces a level-repulsion that often results in a nonmonotonic "pulse." This pulse is more prominent for $\Delta\bar{m}^2 < 0$ than for $\Delta m^2 > 0$, but it is ultimately a generic feature of the behavior of the lower eigenvalue for sufficiently saturated mixing. Indeed, this pulse grows increasingly dramatic as $\bar{\xi} \to 1$, and has a width governed by Δ_G. In general, it is straightforward to show that a pulse will always appear for the lightest eigenvalue if the extra contribution to the squared-mass matrix that comes from the mass-generating phase transition has negative determinant—i.e., $\det \bar{m}_{ij}^2 < 0$.

Strictly speaking, a single pulse does not oscillatory behavior make. However, during the relevant time interval near τ_G, this pulse may be regarded as one oscillation within a full sinusoidal pattern. This situation is illustrated schematically in Fig. 19. Of course, variations in Δ_G directly affect the width of this pulse, and thereby change the effective pulse frequency. As a result, there will exist a single value of Δ_G for which this frequency is exactly twice the natural frequency of our oscillator, and for which our primary parametric resonance emerges. Even greater values of Δ_G then correspond to the higher harmonics of this resonance.

This, then, is the origin of the parametric oscillations apparent in the plots shown in Figs. 12, 16, and 18. As such, this entire phenomenon is a prime example of the interplay between the mixing $\bar{\xi}$ and the width of the mass-generating phase transition: the nonzero mixing produces the level repulsion that leads to the pulse, while the nonzero width of the phase transition endows this pulse with the specific width/frequency needed for it to potentially trigger an actual resonance. Moreover, with this explanation, we now understand why these resonances occur only for certain values of Δ_G, and only for the lighter field. This also explains why these resonances grow stronger as $\bar{\xi} \to 1$. Furthermore, although this pulse exists for both $\Delta\bar{m}^2 < 0$ and $\Delta\bar{m}^2 > 0$ (as evident in Fig. 7), we have already learned in the previous section that for the intermediate values of Δ_G relevant for the parametric resonance, the energy density of the lighter field contributes a greater fraction of the total energy density for $\Delta\bar{m}^2 > 0$ than it does for $\Delta\bar{m}^2 < 0$. This behavior is illustrated, for example, in Fig. 18. It is for this reason that the *total* energy density $\bar{\rho}$ exhibits a stronger parametric resonance for $\Delta\bar{m}^2 > 0$ than for $\Delta\bar{m}^2 < 0$ (as evident in Fig. 12), even though the *individual* component $\bar{\rho}_{\lambda_0}$ exhibits a stronger parametric resonance for $\Delta\bar{m}^2 < 0$ than for $\Delta\bar{m}^2 > 0$ (as is evident in Fig. 16).

In general, it is easy to determine the values of Δ_G for which such parametric resonances occur. As indicated above, we imagine that $\lambda_0^2(\tau)$ experiences a "pulse" centered at $\tau = \tau_p$, as sketched in Fig. 19. Within the region of the pulse, we then approximate the behavior of the pulse as part of a sinusoidal function with an effective frequency $\omega_{\rm eff}$:

$$\lambda_0^2(\tau) \approx \frac{1}{2}\lambda_0^2(\tau_p)\{1 + \cos[\omega_{\rm eff}(\tau - \tau_p)]\} \tag{5.1}$$

for $\tau \approx \tau_p$. From Eq. (5.1) we then find that

$$\omega_{\rm eff}^2 = -4\left(\frac{\ddot{\lambda}_0}{\lambda_0}\right)\Bigg|_{\tau=\tau_p}. \tag{5.2}$$

A pulse with this frequency will lead to an nth-order parametric resonance only if $\omega_{\rm eff}$ is $(2/n)$ times the natural oscillation frequency of our system. The case with $n = 1$ produces the primary parametric resonance, while the cases with $n \in \mathbb{Z} > 1$ produce its higher-order harmonics. However, near $\tau \approx \tau_p$, the natural oscillation frequency associated with the lightest field is nothing but $\lambda_0(\tau_p)$, since it is the mass of the field which drives its oscillations. We thus must demand

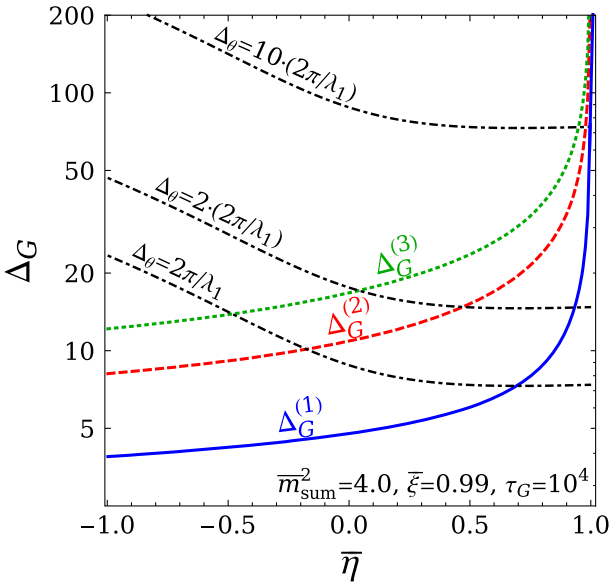

FIG. 20. The phase-transition widths $\Delta_G^{(n)}$ $(n=1,2,3)$ at which the first three parametric resonances occur, plotted as functions of $\bar{\eta}$. Contours showing $2\pi/\lambda_1$ in relation to Δ_θ are also superimposed. The region with large $\bar{\eta}$ is thus one in which our parametric resonances occur while $\Delta_\theta \gg 2\pi/\lambda_1$—i.e., a region in which the total late-time energy density $\bar{\rho}$ is dominated by the contribution from the lighter field experiencing the parametric resonance.

$$\omega_{\rm eff} = \frac{2}{n}\cdot\lambda_0(\tau_p). \tag{5.3}$$

In other words, a parametric resonance will occur only when two timescales are balanced: the timescale $1/\omega_{\rm eff}$ of the pulse, and the timescale $1/\lambda_0(\tau_p)$ of the field oscillations in the vicinity of the pulse. Combining this with Eq. (5.2), we then obtain a condition for an nth-order parametric resonance:

$$\left(\frac{\ddot{\lambda}_0}{\lambda_0^3}\right)\Bigg|_{\tau=\tau_p} = -\frac{1}{n^2}, \quad n\in\mathbb{Z}^+. \tag{5.4}$$

Variations in Δ_G will modify the value of the left side of this equation. Thus, there exist a discrete set of values $\Delta_G^{(n)}$ for which this equation can be satisfied, one for each value of n. These are then the phase-transition widths for which parametric resonances exist.

The solutions for $\Delta_G^{(n)}$ can be obtained numerically, and our results are shown in Fig. 20. We see that the values of $\Delta_G^{(n)}$ generally increase with n, as expected. Likewise, these critical widths also increase as functions of $\bar{\eta}$ and ultimately diverge as $\bar{\eta}\to 1$. Moreover, superimposed on this plot are contours showing the values of $2\pi/\lambda_1$ in relation to Δ_θ. As noted in Sec. IV, the smaller $2\pi/\lambda_1$ is in relation to Δ_θ, the more suppressed is the energy-density contribution of the heavier field relative to that of the lighter field and hence the more the lighter field (which feels the parametric resonance) dominates the total energy density. This then confirms our previous expectation that the parametric resonance, through stronger within $\bar{\rho}_{\lambda_0}$ when $\Delta\bar{m}^2<0$, is nevertheless more pronounced within $\bar{\rho}$ when $\Delta\bar{m}^2>0$.

Thus, we conclude that parametric resonances not only occur, but can also dominate the total energy density of our system. Such parametric resonances can distort the resulting late-time energy densities by several orders of magnitude, and thus may play an important role in early-universe cosmology.

VI. REOVERDAMPING

We now turn to another novel feature which results from the confluence between mixing and a finite phase-transition width: a phenomenon which we refer to as "reoverdamping."

As discussed at the beginning of Sec. II, any scalar field ϕ with mass $\lambda(t)$ will experience overdamped behavior (and thus function as vacuum energy) at times t for which $3H(t)/2 \gtrsim \lambda(t)$. By contrast, such a field will experience underdamped behavior (and thus function as matter) when $3H(t)/2 \lesssim \lambda(t)$. Of course, $H(t)\sim 1/t$ in any postinflationary epoch. Thus, in situations for which $\lambda(t)$ is either constant or monotonically increasing, as illustrated in

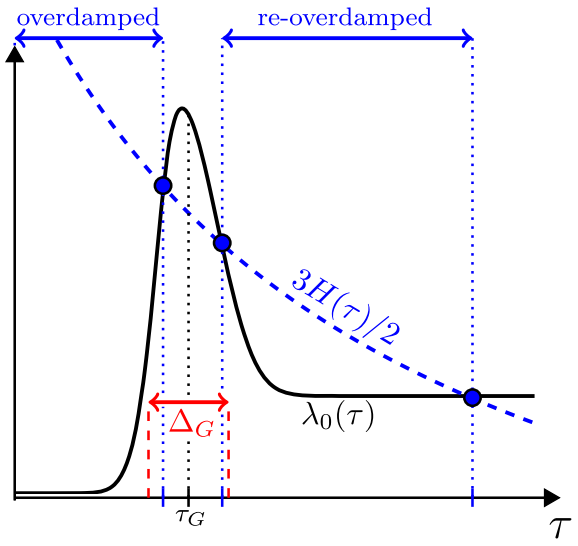

FIG. 21. A schematic illustrating the behavior of the lower-mass eigenvalue $\lambda_0(t)$ in the presence of a "pulse," plotted as a function of time with the Hubble parameter $3H(t)/2$ superimposed. The corresponding field ϕ_{λ_0} therefore experiences not just the usual transition from overdamped to underdamped behavior, as in Figs. 1 and 2, but also a novel *reoverdamping* transition. As shown, this reoverdamped phase can persist for considerably longer than the width Δ_G of the pulse that produced it, and thus can have a significant impact on the resulting cosmological history.

Fig. 2, there exists a single, relatively short era of times t for which $3H(t)/2 \approx \lambda(t)$. Indeed, this is the period, as discussed in Sec. II, during which our scalar field ϕ is transitioning from overdamped to underdamped behavior. After this transition, our field then remains underdamped for all future times.

This is the standard situation, and assumes that $\lambda(t)$ is either constant or monotonically increasing. However, as we discussed in Sec. V, this need not be true when multiple fields mix in the presence of a mass-generating phase transition of nonzero width. Indeed, as we have seen in Sec. V for our two-component toy model, the mass of our lighter field can experience a "pulse," as illustrated in Fig. 19; this is the phenomenon underlying the parametric resonances discussed in Sec. V. In the presence of a pulse, however, the $3H(t)/2$ curve may cross the $\lambda(t)$ curve not just once but *three* times, resulting in a field which begins overdamped, then becomes underdamped, but then experiences a *reoverdamped* phase before eventually settling back into a final underdamped state. This behavior is illustrated in Fig. 21. Moreover, as sketched in Fig. 21, the period during which the field exists in a reoverdamped phase can be of extremely long duration—even relative to the width of the pulse that produced it. Indeed, as we shall discuss below, this duration can become arbitrarily long, depending on the values of the underlying input parameters of our model. As a result, the reoverdamped phase can persist throughout a sizable portion of cosmological history.

It is important to stress that unlike the original overdamped phase, this *reoverdamped* phase of the theory should not be associated with vacuum energy. Indeed, at the moment of transition into the reoverdamped phase, oscillatory behavior ceases but our field value reenters the overdamped phase carrying a nonzero velocity $\dot{\phi}$. Thus, the reoverdamped phase is one in which our field value has a nonzero but steadily decreasing first derivative. In this

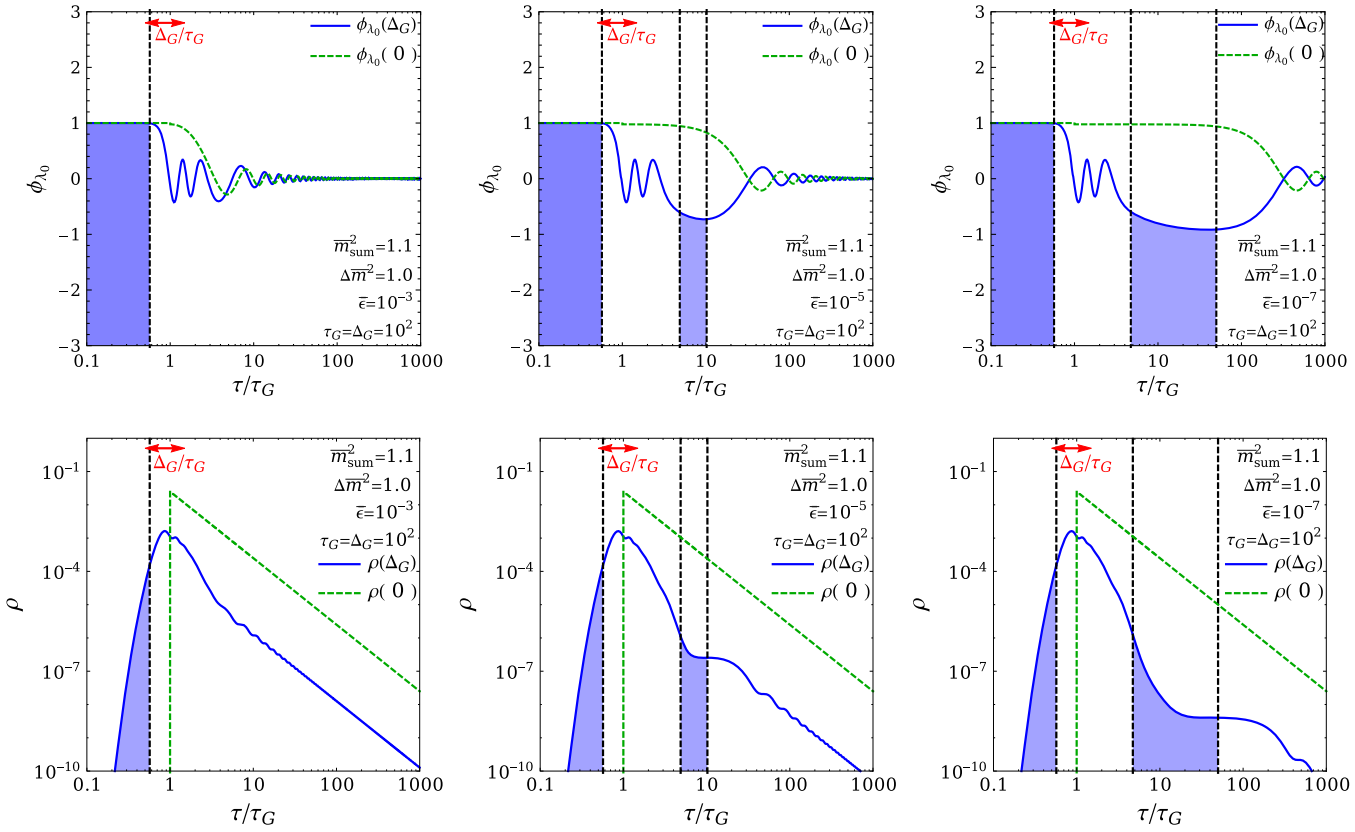

FIG. 22. The lighter field ϕ_{λ_0} (top row) and corresponding total energy density ρ (bottom row), plotted as functions of time for situations (left to right) in which $\xi \to 1$ (or $\epsilon \equiv 1-\xi \to 0$). Note that the fields and corresponding energy densities are normalized so that $\phi_{\lambda_0} = 1$ for $\tau \ll \tau_G$. Time intervals during which the system is overdamped are shaded (blue), while the width Δ_G of the original mass-generating phase transition is also indicated (red). As the mixing saturation increases towards its maximum value (a point of enhanced symmetry for the system), we see that a reoverdamped phase emerges with increasing duration. For comparison purposes, superimposed on each plot are the field values and/or total energy densities that our system would have exhibited in the $\Delta_G \to 0$ limit (green); recall that in this limit it is the energy density of the heavier field which dominates the total energy density. We thus see that the reoverdamped phase can easily persist throughout a large period of cosmological history, with a phenomenology that is quite different from those of the traditional overdamped or underdamped phases.

sense, then, the reoverdamped phase represents a truly different behavior for our field values—one which is neither vacuum energy (constant field values) nor matter (oscillatory field values), but something completely different which transcends our traditional expectations for scalar fields in an expanding universe.

The reoverdamped phase tends to emerge in systems in which the mixing is close to its maximum allowed value, with $\xi \to 1$ (or $\epsilon \equiv 1-\xi \to 0$). The resulting behavior for field values and energy densities in this limit are illustrated in Fig. 22. As expected, we see that the reoverdamped phase has a phenomenology which is quite distinct from those of the traditional overdamped or underdamped phases.

On the one hand, this reoverdamped phase can be considered to be a mere "transient." After all, it cannot last forever: as long as the field ϕ has nonzero mass at late times, this phase must always give way to a final underdamped phase. In other words, the reoverdamped phase can never represent the eventual end-state for our system as $t \to \infty$. Moreover, even within this phase, the nonzero initial field velocity $\dot{\phi}$ which characterizes its unique properties always eventually dissipates to $\dot{\phi} \approx 0$ as long as this phase has sufficient duration. In such cases, this phase eventually comes to resemble vacuum energy. On the other hand, however, this attitude can be viewed as unnecessarily stringent. Even though the reoverdamped phase must always eventually give way to one which is underdamped, we have seen that it can, depending on the specific model under study, be made to persist for arbitrarily long duration. Likewise, this phase can begin with arbitrarily large initial velocity $\dot{\phi}$. Thus, in purely practical terms, this phase can be considered alongside "vacuum energy" and "matter" as an equally valid new behavioral phase that a scalar field can experience during a significant portion of cosmological evolution.

As a final comment, we remark that this reoverdamping phenomenon also leads to a resonancelike behavior wherein quantities such as the late-time energy density experience a nonmonotonic sensitivity to quantities such as Δ_G. This is because the behavior of our field ϕ during the reoverdamped epoch depends very sensitively on the particular field velocity $\dot{\phi}$ that happened to exist at the precise moment the reoverdamping transition occurs. This gives rise to a periodic sensitivity of the final energy density to any quantity which affects the positioning of the reoverdamping transition moment. However, it turns out that the mathematical conditions underlying this periodicity are essentially the same as those underlying the parametric resonance discussed in Sec. V; this makes sense as both concern the fitting of (half-)integral numbers of field oscillations within the same eigenvalue pulse. Moreover, it turns out that the parametric resonance necessarily exists for all situations in which reoverdamping exists (even though the converse is not true). As a result, in such situations both features are inherent in the solutions to our equations of motion and the effects of this reoverdamping-induced oscillation are difficult to disentangle from those of the parametric resonance.

VII. PHENOMENOLOGICAL EXAMPLE: A SECOND AXION

As a phenomenological example of some of the main results of this paper, let us consider the effects that might come from adding a second axion to the standard QCD axion theory. As we shall see, the introduction of a second possible axion can severely distort the physics normally associated with the ordinary QCD axion, even if the second axion is itself associated with a relatively heavy Peccei-Quinn scale or is only mildly coupled to the first.

Our theoretical set up is as follows. We assume that ϕ_0 and ϕ_1 are both axionlike particles—i.e., the pseudo-Nambu-Goldstone bosons associated with the breaking of global symmetries $U(1)_0$ and $U(1)_1$ at energy scales f_0 and f_1, respectively. We assume that both fields are coupled to the QCD gauge fields via the chiral anomaly, with $\mathcal{O}(1)$ model-dependent coefficients ξ_0 and ξ_1. Moreover, in keeping with the assumptions of our toy model, we assume that at early times ϕ_0 is massless while ϕ_1 has a nonzero mass M which possibly arises due to some earlier dynamics. Eventually, nonperturbative QCD instanton effects give rise to a temperature dependent effective potential $V(\phi_1,\phi_2)$ which takes the form

$$V(T) = \mu_\Lambda^4 h_{\rm QCD}^2(T)\left[1-\cos\left(\frac{\xi_0\phi_0}{f_0}+\frac{\xi_1\phi_1}{f_1}+\bar{\Theta}\right)\right] \quad (7.1)$$

where $\mu_\Lambda^4 \equiv g_s^2\Lambda_{\rm QCD}^4/(32\pi^2)$, where $h_{\rm QCD}(T)$ describes the temperature dependence of this effective potential as it turns on (as illustrated in Fig. 4), and where $\bar{\Theta}$ is the QCD theta-angle. Thus we see that the instanton-induced QCD confining phase transition plays the role of our mass-generating phase transition: it occurs at time $t_G = t_{\rm QCD}$ with width $\Delta_G = \Delta_{\rm QCD} = \sqrt{\pi}\sigma t_{\rm QCD}$, and gives rise to an additional contribution to the squared-mass matrix of our axionlike fields which at late times takes the form

$$\bar{m}_{ij}^2 \equiv \left.\frac{\partial^2 V}{\partial\phi_i\partial\phi_j}\right|_{\langle\phi\rangle} = \frac{\mu_\Lambda^4}{f_i f_j} \quad (7.2)$$

where we have set $\xi_0 = \xi_1 = 1$ for simplicity.

Given this setup, we thus have two mass eigenstates, ϕ_{λ_0} and ϕ_{λ_1}, with late-time masses $\bar{\lambda}_0$ and $\bar{\lambda}_1$ respectively. As we survey the range from $M=0$ to $M=\infty$, we find that these late-time masses increase monotonically within the ranges

$$0 \le \bar{\lambda}_0 \le \mu_\Lambda^2/f_0 \quad (7.3)$$

and

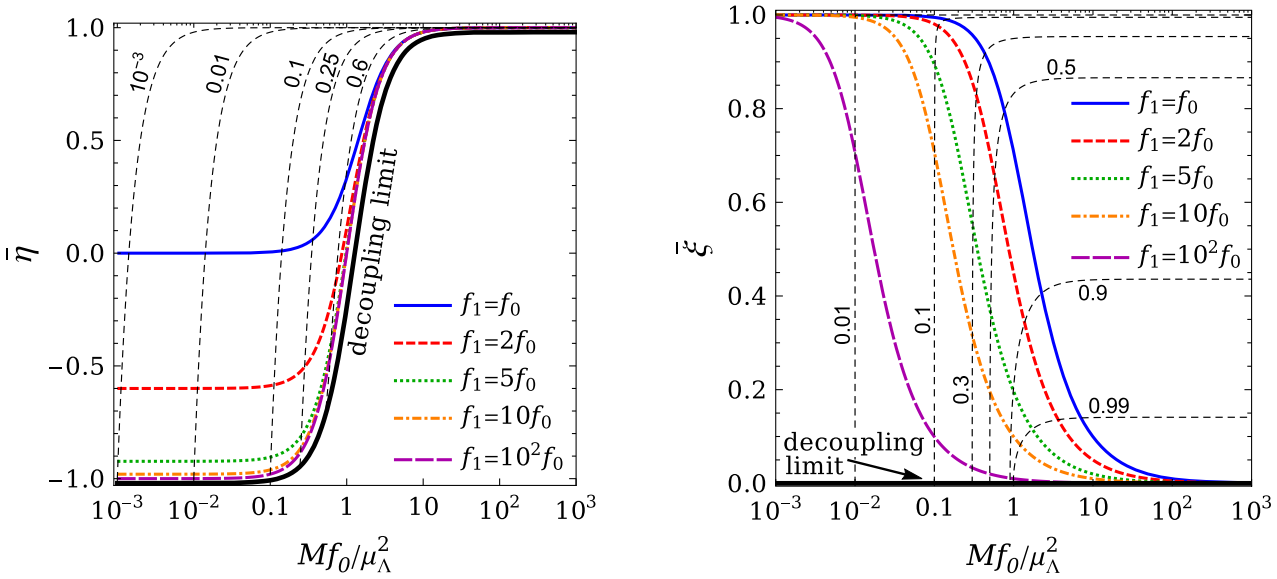

FIG. 23. The quantities $\bar{\eta}$ (left panel) and $\bar{\xi}$ (right panel), plotted as functions of M for different values of f_1/f_0. Superimposed on each plot are contours of $\bar{x}$ (dashed black lines). The decoupling regions in each case correspond to taking $f_1/f_0 \to \infty$ for any M as well as taking $M \to \infty$ for any f_1/f_0; these smoothly connect to form the curves labeled "decoupling limit" on each plot. Note that we have $\bar{\xi} = 0$ along the entire decoupling-limit line, as expected since the decoupling limit is one in which there is no mixing between the two axions. By contrast, the decoupling-limit line surveys all possibilities of $\bar{\eta}$. We also observe from either panel that $\bar{x} = 1$ along the decoupling line for all $M \geq \mu_\Lambda^2/f_0$, while $\bar{x} < 1$ along the decoupling line for all $M < \mu_\Lambda^2/f_0$; this is easier to see in the right panel but is true as a general statement about the relationship between M and $\bar{x}$ in the decoupling limit. Thus, an infinite M implies that $\bar{x} = 1$ but $\bar{x} = 1$ implies only that $M \geq \mu_\Lambda^2/f_0$.

$$\mu_\Lambda^2 \sqrt{\frac{1}{f_0^2} + \frac{1}{f_1^2}} \leq \bar{\lambda}_1 \leq \infty. \tag{7.4}$$

For reasons to be discussed shortly, we shall identify the lighter mass eigenstate as our QCD axion, which means that we shall identify $\bar{\lambda}_0$ as the axion mass m_a. We shall also define $x \equiv \lambda_0 f_0/\mu_\Lambda^2$; thus $0 \leq x \leq 1$. Note that in this section we are keeping M explicit in our expressions; thus, all quantities such as $\bar{\lambda}_i$ have their true physical mass dimensions. Our goal is then to understand how the late-time energy density associated with this field is influenced by the presence of the additional axion.

Note that there are two physically distinct ways in which to realize a limit in which the second axion decouples from the first: either we can take $M \to \infty$ and leave f_1/f_0 arbitrary, or we can take $f_1/f_0 \to \infty$ and leave M arbitrary. In the first case, we find that $\bar{x} = 1$ regardless of the value of f_1. In the second case, by contrast, we find that $\bar{x} = 1$ for all $M \geq \mu_\Lambda^2/f_0$, but $\bar{x} = Mf_0/\mu_\Lambda^2 < 1$ for all $M < \mu_\Lambda^2/f_0$. Of course both decoupling limits smoothly merge together in the region for which both M and f_1/f_0 are taken to infinity.

In order to quickly survey the expected phenomenologies of this two-axion system as functions of $\{M, f_0, f_1\}$—and also to understand the properties of these different decoupling limits—let us calculate the corresponding values of $\bar{\eta}$ and $\bar{\xi}$. In terms of $\{M, f_0, f_1\}$, these are given by

$$\bar{\eta} = \frac{1 + (f_1/f_0)^2[2(Mf_0/\mu_\Lambda^2)^2 - 1]}{1 + (f_1/f_0)^2[2(Mf_0/\mu_\Lambda^2)^2 + 1]} \tag{7.5}$$

and

$$\bar{\xi} = \frac{1}{\sqrt{1 + 2(f_1/f_0)^2(Mf_0/\mu_\Lambda^2)^2}}. \tag{7.6}$$

These quantities are shown in Fig. 23. Although both decoupling limits result in $\bar{\xi} \to 0$ (the absence of mixing), as expected, we immediately see that the decoupling limit with $f_1 \gg f_0$ and arbitrary M generally gives rise to all possible values of $\bar{\eta}$ in the range $-1 \leq \bar{\eta} \leq 1$. By contrast, the decoupling limit with $M \to \infty$ gives rise to $\bar{\eta} = 1$ regardless of the value of f_1/f_0.

This is an important distinction because we have already seen in Sec. IV that $\bar{\eta} > 0$ is the regime in which the majority of the energy density is found in the lighter field at late times, whereas $\bar{\eta} < 0$ is the regime in which the majority of the energy density is transferred to the heavier field at late times. This behavior was in fact discussed explicitly below Eq. (4.5) in the limit of an instantaneous phase transition. Thus, it is only the decoupling limit with

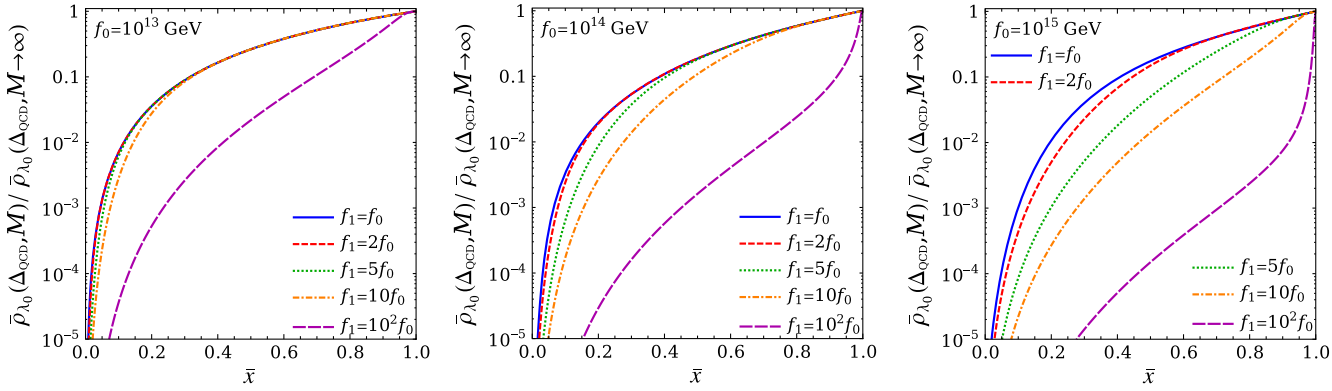

FIG. 24. The late-time energy density $\bar{\rho}_{\lambda_0}(\Delta_{\rm QCD}, M)$ associated with the QCD axion field ϕ_{λ_0}, plotted as a function of the late-time axion mass $\bar{x} \equiv \bar{\lambda}_0 f_0/\mu_\Lambda^2$ and normalized to what it would have been in the decoupling limit in which the second axion plays no role. These plots thus illustrate the effects of the nonzero mixing between our two axion fields in the presence of a nonzero QCD phase transition width $\Delta_{\rm QCD}$. Note that although each normalization factor $\bar{\rho}_{\lambda_0}(\Delta_{\rm QCD}, M \to \infty)$ is in principle different for each different f_1/f_0 curve, physics in the decoupling limit $M \to \infty$ is insensitive to the value of f_1/f_0. Thus all curves in each panel share the same normalization factor.

$M \to \infty$ and arbitrary f_1/f_0 for which the energy density actually remains in the lighter field where it started, as we would expect for a proper decoupling limit.

We shall therefore adopt the limit with $M \to \infty$ and arbitrary f_1/f_0 as our benchmark against which to measure the effects that come from the presence of the second axion. Note that in the $M \to \infty$ limit, the physics of our system actually becomes *independent* of f_1/f_0; we have already seen this behavior for the specific quantities $\bar{\eta}$ and $\bar{\xi}$ in Fig. 23, but this feature indeed holds more generally for all properties of the system.

Although we have been considering the lighter mass eigenstate ϕ_{λ_0} as our QCD axion, it is important to realize that the entire notion of a "QCD axion" no longer strictly applies. For example, there is no mass eigenstate in our model whose mass is inversely proportional to f_0 or f_1, as might be taken to characterize a QCD axion. Likewise, there is no mass eigenstate in this model which solves the strong *CP* problem; in particular, we see from Eq. (7.1) that it is only the linear combination $\phi_0/f_0 + \phi_1/f_1$ which dynamically relaxes against the strong *CP* angle $\bar{\Theta}$. However, in the $M \to \infty$ decoupling limit, our model reduces to the standard single-axion model: the lighter mass eigenstate remains massless prior to the instanton-induced phase transition, gathers a mass μ_Λ^2/f_0 after this transition, and solves the strong *CP* problem. For this reason we shall continue to refer to the lighter mass eigenstate as our "QCD axion," even in the presence of mixing.

Given this, we now calculate the late-time energy density associated with the lighter (QCD) axion as a function of the lighter (QCD) axion mass $\bar{x} \equiv \bar{\lambda}_0 f_0/\mu_\Lambda^2$, normalized to our decoupled benchmark limit with $M = \infty$ (for which $\bar{x} = 1$). (As a technical point, note that we describe this benchmark through the condition $M = \infty$ rather than the condition $\bar{x} = 1$, since the former implies the latter regardless of f_1/f_0 whereas the latter fails to imply the former when $f_1/f_0 \to \infty$.) In all cases we take $t_G = t_{\rm QCD}$ and $\Delta_G = \Delta_{\rm QCD}$ to have the specific fixed values that correspond to the instanton-induced QCD phase transition, as plotted in Fig. 4.

Our results are shown in Fig. 24. We see from Fig. 24 that the late-time energy density associated with the QCD axion is suppressed relative to what would have been expected in the $M \to \infty$ decoupling limit. Indeed, this suppression is significant, occasionally stretching over many orders of magnitude, and arises regardless of the mass $\bar{x}$ of the QCD axion, its corresponding Peccei-Quinn scale f_0, or the Peccei-Quinn scale f_1 associated with the second axion. Indeed, none of this would have occurred were it not for the nonzero width of the instanton-induced QCD phase transition. This then demonstrates the phenomenological relevance of this nonzero width, and the need to incorporate these kinds of effects in studies of multiaxion theories.

Note that this suppression of the late-time energy density due to the presence of the second axion field persists *even in the limit for which* $f_1 \gg f_0$. However, this somewhat counter-intuitive result should not come as a surprise. As we know, the limit $f_1/f_0 \to \infty$ corresponds to one in which $\bar{\xi} \to 0$—i.e., one in which the late-time mixing vanishes. However, this does not imply that the mixing is small throughout the time evolution—in fact, both θ and its time derivative can be quite sizable *during* the mass-generating phase transition. Indeed, this is particularly true for the relatively small values of $\bar{x}$ which correspond to $\bar{\eta} < 0$. Thus, it is *during* the mass-generating phase transition that the bulk of the energy density can be transferred to the second axion—even when $f_1 \gg f_0$! This is therefore an effect that arises as a direct consequence of the nonzero width of the instanton-induced QCD phase transition and its associated timescale.

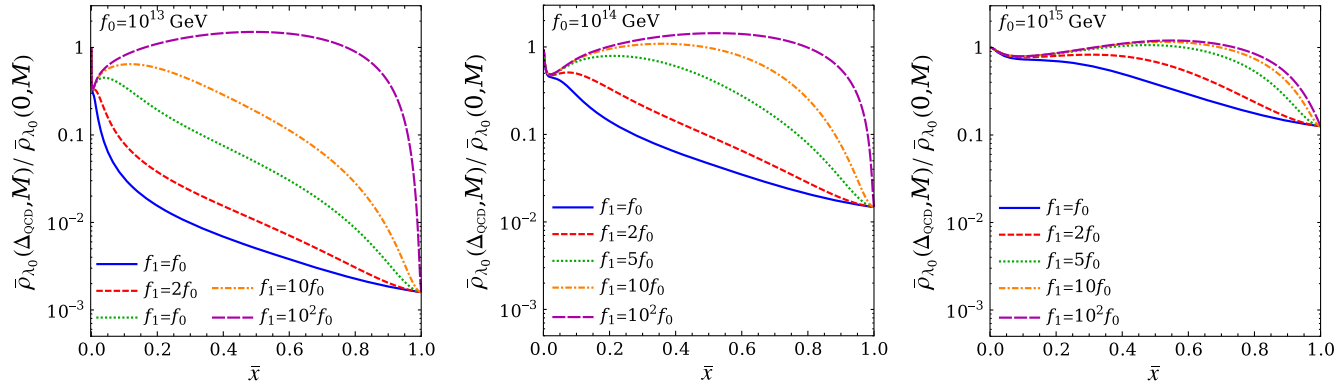

FIG. 25. Same as Fig. 24 except that the late-time energy densities $\bar{\rho}_{\lambda_0}(\Delta_{\rm QCD}, M)$ are now normalized to the values that they would have had if the QCD phase transition had been treated as instantaneous. These plots thus illustrate the effects that emerge due to the actual nonzero width of the QCD phase transition—effects which survive even in the $M \to \infty$ decoupling limit (for which $\bar{x} \to 1$). Note that unlike the plots in Fig. 24, here the normalization factors $\bar{\rho}_{\lambda_0}(0, M)$ are different for each f_1/f_0 curve and in fact even vary with $\bar{x}$ *along* each curve.

Interesting as these results are, they only address the issue of how our late-time energy density compares with what would arise in the decoupled case—i.e., the case with no mixing ($\bar{\xi} = 0$). In this paper, however, we have also repeatedly studied another comparison, namely to the case of an *instantaneous* phase transition—i.e., the case in which we imagine our phase transition to have $\Delta_{\rm QCD} = 0$. The corresponding results are shown in Fig. 25.

Once again, it is not difficult to understand the features in these plots. First, let us discuss the endpoints at $\bar{x} = 1$. Of course, these endpoints correspond to the decoupling limit (because $\bar{x} = 1$ implies $\bar{\xi} = 0$, as we have seen in Fig. 23); thus the physics at these endpoints includes the effects from only the lighter (QCD) axion itself. As we see from Fig. 25, even in this limit the resulting late-time energy density is significantly suppressed by factors ranging from 5 to 500 as f_0 ranges from 10^{15} to 10^{13} GeV. Thus the effects of properly accounting for the nonzero width $\Delta_{\rm QCD}$ of our instanton-induced phase transition are extremely important, even for a single QCD axion!

Moving away from the decoupling limit towards smaller values of $\bar{x}$, we now begin to incorporate the additional effects of the mixing with the second axion field. As we see from Fig. 25, these effects *enhance* the resulting late-time energy density above what it would have been in the decoupling limit. In so doing, they can even possibly overcome the above suppression completely—a prospect that depends on the value of f_1/f_0—and thereby potentially *increase* the late-time energy density above what it would have been for an instantaneous phase transition! This enhancement ultimately reflects the same physics that we have already seen in Figs. 10 and 11 for small mixing saturations in the $t_G \approx t_\zeta^{(0)}$ region. Indeed, we see from Fig. 23 that larger values of f_1/f_0 correspond to smaller mixing saturations $\bar{\xi}$.

Finally, we can also understand why these curves all begin at 1 for $\bar{x} = 0$. At this endpoint of these plots, our axion fields remain overdamped during the mass-generating phase transition. They thus remain insensitive to the transition width $\Delta_{\rm QCD}$.

We can also consider the fraction of the *total* energy density of our two-axion system which is associated with the lighter axion field. These results are shown in Fig. 26. In general, we see that as f_1/f_0 increases, an increasing fraction of the total energy density is associated with the heavier axion field. However, as $\bar{x} \to 1$, we see that virtually all of the total energy density can be associated with the lighter axion field.

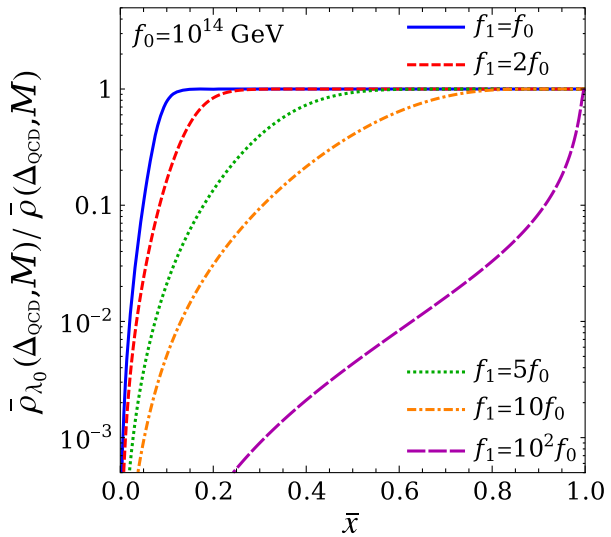

FIG. 26. The late-time fraction $\bar{\rho}_{\lambda_0}/\rho$ of the total late-time energy density which is associated with lighter axion field, plotted as a function of $\bar{x}$. We see that this fraction generally decreases as f_1/f_0 increases, and also increases as a function of $\bar{x}$.

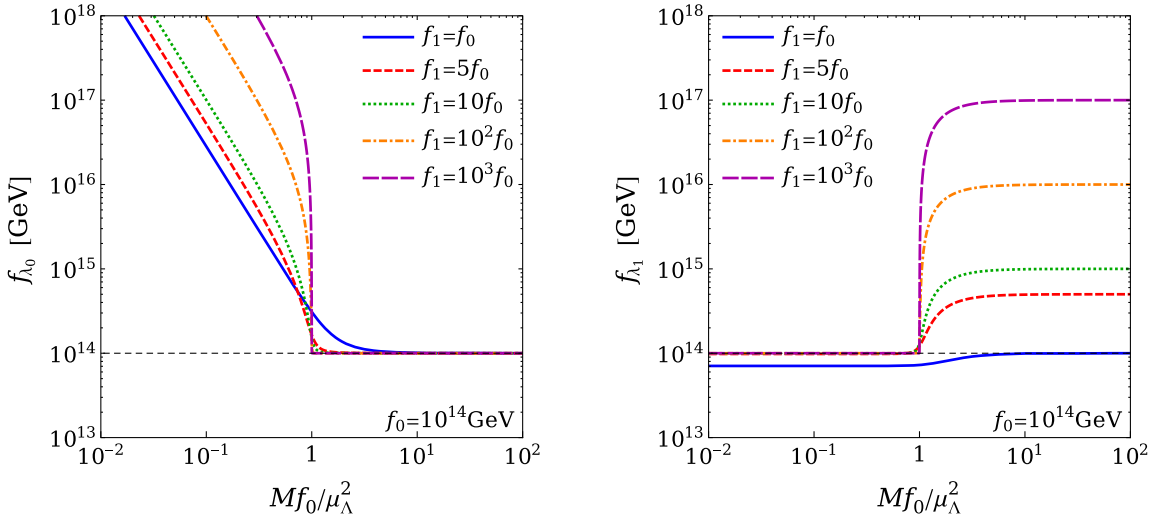

FIG. 27. The effective coupling scales f_{λ_0} and f_{λ_1} associated with our axion mass eigenstates ϕ_{λ_0} and ϕ_{λ_1}, plotted as functions of Mf_0/μ_Λ^2. We see that these scales are never significantly smaller than the f_i coupling scales associated with our original unmixed states, and that either one or the other typically takes a value which significantly exceeds this when $f_1/f_0 \gg 1$.

As a final comment, we remark that it is the combination $\phi_0/f_0 + \phi_1/f_1$ which not only solves the strong-CP problem but which also couples to gluons. Depending on the particular model under study, this combination may also couple to photons and other Standard-Model (SM) particles. We can therefore decompose this combination into our late-time mass eigenstates in order to determine the effective mass scales f_{λ_0} and f_{λ_1} that govern the relevant SM couplings of our mass eigenstates at any moment in time:

$$\frac{\phi_0}{f_0} + \frac{\phi_1}{f_1} \equiv \frac{\phi_{\lambda_0}}{f_{\lambda_0}} + \frac{\phi_{\lambda_1}}{f_{\lambda_1}}. \tag{7.7}$$

The late-time values of these effective coupling scales are plotted in Fig. 27.

Several features of these plots are worthy of note. For example, we observe that in all cases the $\bar{f}_{\lambda_i}$ coupling scales associated with our mass eigenstates are never significantly smaller than the f_i coupling scales associated with our original unmixed states. Thus neither of our mass eigenstates will decay too rapidly to SM states, a situation which might potentially have resulted in phenomenological difficulty. Likewise, combining the results in Fig. 27 with those of Figs. 23 and 26, we see that there are a variety of different phenomenological situations which can emerge, depending on the underlying parameters in our model.

For example, let us consider the case with $Mf_0/\mu_\Lambda^2 \approx 0.4$ and $f_1 \approx 5f_0$. We immediately learn from Fig. 27 that our QCD axion coupling scale is $f_{\lambda_0} \approx 10^{15}$ GeV, which is a full order of magnitude higher than the value that it would have had in the decoupling limit. This implies that the QCD axion has an enhanced invisibility as assessed through its interactions with SM states. However, we see from Fig. 23 that these parameter combinations correspond to $\bar{x} \approx 0.4$, whereupon we see from Fig. 26 that this in turn corresponds to an energy-density fraction $\bar{\rho}_{\lambda_0}/\bar{\rho} \approx 80\%$. This is thus a situation in which our QCD axion is invisible but nevertheless carries the bulk of the total late-time energy density.

A somewhat different situation emerges if we keep $f_1 \approx 5f_0$ but now take $Mf_0/\mu_\Lambda^2 \approx 0.1$. In this case, we learn from Fig. 27 that our QCD axion coupling scale is even higher —$f_{\lambda_0} \approx 10^{17}$ GeV—which suppresses its couplings to SM states by three orders of magnitude compared with the decoupling limit. However, we see from Fig. 23 that these parameter combinations now correspond to $\bar{x} \approx 0.1$, whereupon we see from Fig. 26 that this in turn corresponds to an energy-density fraction $\bar{\rho}_{\lambda_0}/\bar{\rho} \approx 3\%$. This is thus a situation in which our QCD axion is invisible and carries almost *none* of the total energy density—all because of its mixing with the second axion in the presence of a nonzero phase transition width!

The opposite situation, of course, emerges for large M. For example, with $Mf_0/\mu_\Lambda^2 \approx 10$ we find that the physics of this model is roughly insensitive to the value of f_1/f_0; we find $\bar{x} \approx 1$ and $\bar{\rho}_{\lambda_0}/\bar{\rho} \approx 100\%$. This, of course, signifies nothing but the approach to our decoupling limit. However, as long as we are not precisely *at* the decoupling limit, we still find from Fig. 24 that the energy density associated with this lighter axion field remains considerably suppressed compared with what we would have found at the full decoupling limit. Thus in this sense our QCD axion has regular couplings to SM states but nevertheless carries relatively little total energy density.

Needless to say, our goal in this section has not been to propose a complete phenomenological model of axion physics. Rather, it has merely been to illustrate the rich implications that can emerge for a QCD axion in the presence of mixing with a second axion when the nonzero width of the QCD phase transition is properly taken into account. As we have seen, the total energy density of such a two-axion system can be significantly suppressed relative to what might have been expected in the decoupling limit, and the individual energy densities associated with each axion can be severely distorted. Indeed, this lesson has been one of the primary themes of this paper.

VIII. BEYOND TWO FIELDS

In Sec. II, a toy model was constructed in order to study the effects that arise in multiscalar models due to the simultaneous presence of two key ingredients: mass-generating phase transitions with finite widths, and nonzero mixing between the different components. For the sake of simplicity, our toy model consisted of only two scalar fields, as this is the minimum number that could be chosen for our purposes. However there is no reason that we are limited to two fields, and indeed the number of fields may be extended arbitrarily. Therefore, in order to gain a quick sense of what possibilities might emerge with an increased number of fields, we shall now briefly consider several aspects of the case with three fields. A more general study of the N-field case will be presented in Ref. [38].

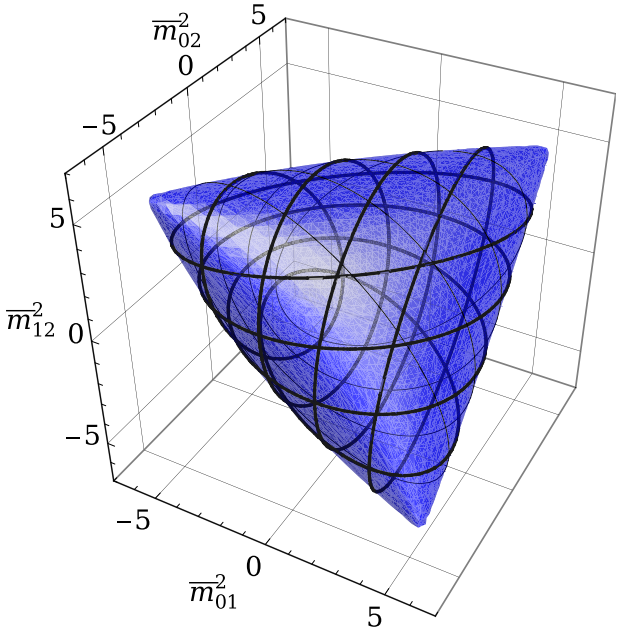

FIG. 28. The region within the three-dimensional mixing space parametrized by $\{\bar{m}^2_{01}, \bar{m}^2_{02}, \bar{m}^2_{12}\}$ for which the mass matrix $\mathcal{M}^2$ in Eq. (8.1) is positive-semidefinite. For this plot we have taken $\bar{m}^2_{00} = 5$, $\bar{m}^2_{11} = 5$, and $\bar{m}^2_{22} = 1$.

In analogy with Eq. (2.8), we shall begin by assuming a rescaled (dimensionless) mass matrix $\mathcal{M}^2(t)$ which takes the form

$$\mathcal{M}^2(\tau) = \begin{bmatrix} m^2_{00}(\tau) & m^2_{01}(\tau) & m^2_{02}(\tau) \\ m^2_{01}(\tau) & 1 + m^2_{11}(\tau) & m^2_{12}(\tau) \\ m^2_{02}(\tau) & m^2_{12}(\tau) & 4 + m^2_{22}(\tau) \end{bmatrix}. \tag{8.1}$$

Here $\tau \equiv Mt$ (where M is the overall mass parameter which has been scaled out, as in Sec. II), and we assume that all components have a common time dependence associated with a mass-generating phase transition whose properties are exactly those of our two-component model. Note that the diagonal components of this matrix imply that our three fields have masses 0, M, and $2M$ at early times prior to the phase transition. These values are chosen for simplicity, and are also motivated by the Kaluza-Klein masses that might result in a theory with a flat extra dimension (in which case we would identify M as the inverse length of this dimension).

Inherent in this mass matrix are now *three* independent mixing angles which correspond to m^2_{01}, m^2_{02}, and m^2_{12}. However, just as in the two-field case in Sec. II, there are ultimately constraints on these mixing parameters which ensure that our mass matrix $\mathcal{M}^2$ remains positive-semidefinite. Such constraints are illustrated in Fig. 28 for the case with $\bar{m}^2_{00} = 5$, $\bar{m}^2_{11} = 5$, and $\bar{m}^2_{22} = 1$—a fixed

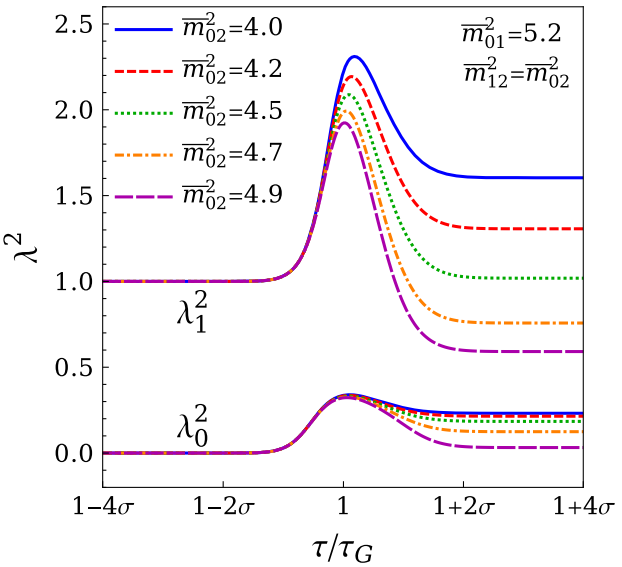

FIG. 29. The two smaller mass eigenvalues λ^2_0 and λ^2_1 in our three-field model, plotted as a function of time for different values of $\bar{m}^2_{02} = \bar{m}^2_{12}$. As in Fig. 28, we have taken $\bar{m}^2_{00} = 5$, $\bar{m}^2_{11} = 5$, and $\bar{m}^2_{22} = 1$. We see that each field can now separately have its own pulse with its own effective frequency, resulting in the possibility of *two* parametric resonances which can either reinforce or interfere with each other.

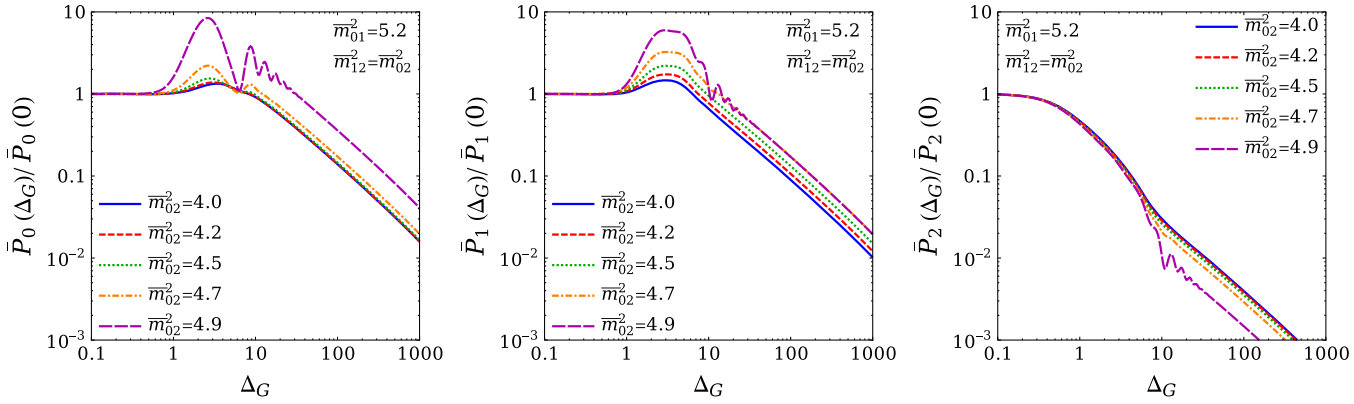

FIG. 30. The partial-sum energy densities $\bar{P}_\ell(\Delta_G)$, $\ell = 0, 1, 2$, plotted as functions of the phase-transition width Δ_G and normalized to the values $\bar{P}_\ell(0)$ that they would have had for an instantaneous phase transition. As in Figs. 28 and 29, we have again taken $\bar{m}_{00}^2 = 5$, $\bar{m}_{11}^2 = 5$, and $\bar{m}_{22}^2 = 1$. We observe many features familiar from our two-component toy model, including parametric resonances which enhance the corresponding energy densities as well as suppressions by many orders of magnitude that develop for sufficiently large Δ_G. However, there are also new features which arise due to the fact that we now have more than two fields in our system. The center panel, in particular, illustrates the manner in which the separate resonances from the two lighter fields reinforce each other and produce a *plateau* along which the partial-sum energy density $\bar{P}_1$ is enhanced by nearly an order of magnitude over a significant Δ_G-interval.

parameter choice we shall adopt for all of the plots in this section.

In our two-component toy model, we have seen that the lighter field can occasionally develop a "pulse" which is ultimately responsible for the parametric-resonance and reoverdamping phenomena in Secs. V and VI respectively. However, in a *three*-component model, it turns out that the *two* lightest fields can each develop a pulse. This is illustrated in Fig. 29. Indeed these two pulses need not have the same effective frequencies or correspond to the same values of Δ_G. As we shall see, this can then give rise to multiple resonances which can reinforce or interfere with each other and thereby produce new effects which transcend those realizable with only a single pulse. We also note that the pulse of our second mass eigenvalue λ_1^2 can leave this eigenvalue with a smaller magnitude at late times than at early times—a phenomenon which was not possible for λ_0^2 in the two-field case. This has the effect of allowing a further elongation of the time interval during which this field can remain in a reoverdamped or parametrically resonant state.

To illustrate the interplay between the different resonances, we can consider the late-time energy densities of this system. Since we now have three individual contributions to the total energy density, we shall highlight their separate effects by considering the partial sums

$$P_\ell \equiv \sum_{i=0}^{\ell} \rho_{\lambda_i}. \tag{8.2}$$

In Fig. 30 we have plotted these partial sums, with each normalized to the value that it would have had if the mass-generating phase transition had been instantaneous. This demonstrates the combined effects of the mixing and the finite phase-transition width. The left panel of Fig. 30 illustrates the behavior of $\bar{P}_0 \equiv \bar{\rho}_{\lambda_0}$ alone: this behavior is similar to what we have already seen in our two-field toy model, for which the pulse in the evolution of λ_0^2 produces resonant peaks in the corresponding energy-density component at late times. These peaks grow stronger for larger $\bar{m}_{02}^2$, with the energy density in this field dissipating inversely with Δ_G once $\Delta_G \gtrsim 2\pi/\lambda_0$. However, the center panel shows the combined contribution to the total energy density from the two lighter fields and thus exhibits something new: the contributions of the two separate parametric resonances combine to reinforce each other and even produce an effective enhancement *plateau* which stretches over an extended interval in Δ_G. Indeed, all along this plateau, the energy densities of our system can be enhanced by as much as nearly an order of magnitude! Finally, the right panel of Fig. 30 illustrates the behavior of the total energy density $\bar{\rho}$. Note that in this case, the heaviest field effectively dominates the energy density for $\Delta_G \lesssim 5$, thereby washing out those parametric resonances that appear in that range, while leaving those that appear for $\Delta_G \gtrsim 5$.

The purpose of this section has merely been to provide several short examples of new phenomena that can arise when more than two scalar fields are involved. We shall defer a more detailed study of these general cases to Ref. [38].

IX. CONCLUSIONS

In this paper we have examined some of the novel effects that emerge when multiple scalars mix with each other and experience a time dependent mass-generating phase transition within the context of a time-dependent

background cosmology. As we have explicitly demonstrated within a relatively simple two-field toy model, these effects include both enhancements and suppressions of the total late-time energy density of the system as well as the late-time energy densities of the individual scalar components. These enhancements and suppressions need not be mere $\mathcal{O}(1)$ effects; indeed, we have seen that the resulting late-time energy densities can often be altered by many orders of magnitude. These alterations are even sufficient to produce wholesale shifts in the identity of the field carrying the bulk of the late-time cosmological abundance.

We also found that our late-time energy densities can exhibit large parametric resonances which make them extremely sensitive to relatively small changes in the degree of mixing or in the width of the phase transition. Thus, any effects which might change—even slightly—the rate at which the phase transition unfolds can have a significant impact on the late-time abundances of our fields. We also demonstrated that our system can enter a so-called "reoverdamped" phase in which the field values and energy densities behave in a manner quite different from what would normally be associated with pure dark matter or vacuum energy. Indeed, we saw that the system can remain in this reoverdamped phase for a considerable length of time, suggesting that this phase be considered alongside "vacuum energy" and "matter" as an equally valid new behavioral phase that a scalar field can experience during a significant portion of cosmological evolution. Finally, we illustrated the importance of some of these effects within the context of a simple model in which a second axion is incorporated into the standard QCD axion theory.

As indicated in the Introduction, our results can be taken as an explicit demonstration of how critical it can be to properly incorporate the nonzero but finite widths of mass-generating phase transitions when discussing late-time energy densities in cosmological settings involving multiple mixed scalars. Indeed, treating these widths as being either zero or infinite (the latter corresponding to a so-called "adiabatic" approximation in the axion case) can lead to results for late-time quantities such as cosmological abundances which are significantly distorted from their true values, often by many orders of magnitude! Given the sensitivity with which the different slices of the so-called "cosmic pie" have already been measured, such effects will inevitably play critical roles in determining whether a particular cosmological scenario is viable or ruled out.

Needless to say, the features we have discussed in this paper give rise to many new possibilities for the phenomenology of the early universe and cosmological evolution. As such, new approaches to model-building can readily be imagined. For example, many extensions to the Standard Model involve scalar fields which are massless (or effectively massless) at high scales, but for which masses are generated dynamically at lower scales. Examples include string moduli, string axions, and other axionlike particles. However, these particles can be cosmologically problematic if these dynamically-generated masses are sufficiently large because they can overclose the universe once they begin oscillating and act as massive matter. Indeed, this is the crux of the cosmological moduli problem [41–43]. Moreover, particles of this sort are often produced through mechanisms such as vacuum misalignment which are completely decoupled from the dynamics of inflation. In such cases, spatial fluctuations in the energy density of these particles therefore give rise to isocurvature perturbations, the observational limits on which are quite severe [44].

Fortunately, mechanisms such as those we have discussed in this paper can suppress the resulting late-time energy densities by many orders of magnitude. This enhanced energy dissipation can therefore provide a novel way of alleviating such constraints on models of new physics, and may represent an alternative solution to the moduli and isocurvature problems. They may also allow a general weakening of the cosmological bounds on the energy scales associated with the new physics (such as the Peccei-Quinn scale associated with axions).

As we have discussed, the features explored in this paper can lead not only to a reduction in the total late-time energy density of our system, but also to changes in the late-time *partitioning* of that total energy density amongst the fields in our theory. In new physics scenarios which involve multiple scalar fields, this too can have observable phenomenological consequences. For example, certain classes of string models are generically expected to give rise to large numbers of light axionlike particles [30–33], the masses of which must be generated dynamically via nonperturbative effects. In situations in which the same dynamics contributes to the generation of mass terms for multiple such particles, the relative cosmological abundances of those particles can be significantly altered by the effects we have studied here.

There are many ways in which the analysis in this paper might be extended and generalized. For example, in this paper we have focused primarily on scalar fields which have a negligible back-reaction on the background cosmology, at least during the epoch of dynamical mass generation. In other words, we have assumed that our scalars do not contribute significantly to the total energy density of the universe until well after their masses have settled into their late-time values. However, in principle, we could relax this assumption and broaden this study to include scalars which *do* have a significant impact on the background cosmology. This would then establish a nonlinear back-reaction which could possibly enrich the dynamics of this system even further, with the evolution of the Hubble parameter $H(t)$ itself depending on the behaviors of the scalar energy densities as functions of time.

This possible extension of our work could be particularly important if the scalars involved are those which are directly responsible for triggering periods of rapid cosmological expansion, including cosmic inflation. Indeed, a wide variety of inflationary models involving multiple scalar fields exist in the literature; these include hybrid inflation [45,46], assisted inflation [47–50], N-flation [51], and multifield stochastic inflation [52]. Likewise, several constructions have been proposed through which a potential for the fields responsible for cosmic inflation is generated dynamically [53–56]. Such constructions are typically motivated by the possibility of realizing a viable inflationary potential without the introduction of extremely small couplings or arbitrary mass scales. It would therefore be extremely interesting to explore the cosmological consequences of a time dependent, dynamically-generated inflaton potential in the context of multifield inflation.

In this connection, it is perhaps worth emphasizing that the parametric resonance we have discussed in Sec. V is entirely unrelated to the parametric resonance which gives rise to an epoch of explosive production of light scalars in inflationary models with a preheating phase [57–61]. Indeed, the effect we have discussed in this paper involves a resonance between the oscillation frequency of the light scalar, as given by its mass λ_0, and the effective modulation frequency $\omega_{\rm eff}$ of that mass. As such, this parametric resonance is ultimately driven by the dynamics of the mass-generation mechanism and leads to an enhancement in the amplitude for coherent oscillations of the zero-momentum mode of ϕ_{λ_0}. By contrast, the effect discussed in preheating scenarios involves a resonance between the frequency of the coherently-oscillating inflaton field and the frequency associated with a given momentum mode of the lighter scalar field into which it decays. Such a resonance is ultimately driven by inflaton dynamics and results in particlelike excitations of this lighter scalar.

Other extensions and generalizations of our work are also possible. For example, our original dynamical equations in Eq. (2.1) assume that there are no significant spatial variations in our field values, but in general it is possible to allow our field values to vary not only in time but also in space. This could then give rise to distinctly different physics in different spatial regions, with domain walls potentially separating different regions experiencing different phases and different energy densities. Such spatial variations and domain walls would inevitably introduce additional length scales into the problem, and thereby likely give rise to new complex dynamical behaviors for our system as a whole.

In a similar vein, we could likewise imagine that our fundamental dynamical equations in Eq. (2.1) are modified by the inclusion of *source terms* reflecting possible interactions with more complex cosmological environments (e.g. with other fields). Such source terms could then give rise to additional features which effectively modify our mass eigenvalues and eigenstates, trigger enhancements or suppressions of late-time energy densities, and produce additional kinds of resonances. Indeed, such effects would be analogous to the well-known "matter effects" (i.e., MSW effects) that arise when neutrinos propagate through matter rather than through a pure vacuum.

Another possible generalization of our toy model is to consider more than two scalar fields. Indeed, as briefly discussed in Sec. VIII, there is no limit to the number of scalar fields which may be considered. These models can therefore grow in both complexity and sophistication compared with the two-component model we have studied here.

One natural possibility along these lines is to consider the case in which the different fields ϕ_i are nothing but the Kaluza-Klein (KK) modes of a single higher-dimensional field Φ. This is indeed a well-motivated possibility, as string theory naturally gives rise to many scalars (axions, geometric/gravitational moduli, gauge-neutral singlets, and so forth) that populate the "bulk" volume transverse to the brane on which gauge interactions reside. As such, the different KK modes of such bulk fields are natural dark-matter candidates, even though they might be unstable—an observation which has motivated the Dynamical Dark Matter framework of Refs. [28,29,36]. In this case, the masses that we have assumed to exist prior to the phase transition would be nothing but the corresponding KK masses, while the extra time-dependent contributions $\bar{m}_{ij}h(\tau)$ are presumably generated by a phase transition which breaks the higher-dimensional symmetries that would otherwise have aligned our mass eigenstates with KK eigenstates. The resulting mixing structure is therefore determined according to the intrinsically higher-dimensional geometric symmetries of the theory and the manner in which these symmetries might be broken by the phase transition. (For example, in the five-dimensional axion scenario of Refs. [29,35,36] the instanton-induced phase transition takes place purely on the brane and thereby breaks a translation symmetry that otherwise existed along the extra dimension.) In general, the cosmological properties of such KK systems can be extremely rich, and will be discussed in more detail in Ref. [38].

Given these observations, it might at first glance seem that the two-component toy model we have studied in this paper—while illustrative for pedagogical purposes—might not have been worthy of the detailed attention we have afforded it. However, this model can always be viewed as the low-energy limit of a more complete model (such as the model of Kaluza-Klein cosmology discussed above) in which only the contributions from the lightest two modes are considered. Indeed, in many cosmological settings, the contributions of the lightest fields are likely to dominate the resulting phenomenology and/or carry the largest abundances. Thus, in this sense, our two-component toy model

can be viewed as the common root underlying a variety of much more general models—a "kernel" which gives rise to rather universal features and behaviors that will continue to appear (perhaps with further embellishments) in more complete or realistic settings. As such, then, this toy model can be viewed as representing a universal low-energy limit of a wide class of models of this type whose leading-order phenomenologies will be exactly those we have investigated in this paper.

ACKNOWLEDGMENTS

We would like to thank D. Chung, S. Su, and S. Watson for useful discussions. The research activities of K. R. D and J. K. were supported in part by the Department of Energy under Grant No. DE-FG02-13ER-41976, while the research activities of B. T. were supported in part by internal funding from Reed College. The research activities of K. R. D. were also supported in part by the National Science Foundation through its employee IR/D program. The opinions and conclusions expressed herein are those of the authors, and do not represent any funding agencies.

APPENDIX A: NARROW-WIDTH BEHAVIOR OF THE LATE-TIME ENERGY DENSITY: ANALYTICAL RESULTS

As we have seen, the presence of a nonzero width Δ_G for the mass-generating phase transition prevents us from obtaining analytical results for most of our quantities of interest. However, in the case with $\Delta_G = 0$ (corresponding to an instantaneous phase transition), the resulting dynamics can be treated analytically. In this appendix we provide analytical results for the late-time energy density $\bar{\rho}$.

For $\Delta_G = 0$, the behavior of each of our two fields is relatively simple. Prior to the phase transition at τ_G, only the massless field has a nonzero value A_0, as in Eq. (2.16); this field is necessarily overdamped, however, and thus remains effectively constant until the phase transition at τ_G. At that point, the mass matrix receives the extra contribution which mixes the two fields with an angle θ; the initial massless field value is thus redistributed at $\tau = \tau_G$ such that

$$\phi_{\lambda_0}(\tau_G) = A_0 \cos\theta, \qquad \phi_{\lambda_1}(\tau_G) = A_0 \sin\theta. \tag{A1}$$

Each field then evolves forward in time independently.

At first glance, motivated by the behavior illustrated in Fig. 1, we might try to estimate the total late-time energy density $\bar{\rho}$ using a virial approximation in which we assume that the energy density associated with each field ϕ_{λ_i} remains constant until $\tau_\zeta^{(i)} \equiv \kappa/2\lambda_i$ [where $\kappa \equiv 2$ (or $3/2$) for a matter- (radiation-)dominated universe], after which it scales as $\sim\tau^{-\kappa}$. Defining $\tau_\lambda \equiv \max\{\tau_G, \kappa/2\lambda\}$, we would then write

$$\bar{\rho} \approx \sum_\lambda \rho_\lambda(\tau_\lambda)\left(\frac{\tau}{\tau_\lambda}\right)^{-\kappa} = \frac{1}{2}\sum_\lambda \lambda^2[\phi_\lambda(\tau_G)]^2\left(\frac{\tau}{\tau_\lambda}\right)^{-\kappa}. \tag{A2}$$

Unfortunately, this approximation breaks down for any fields ϕ_i for which the threshold of critical damping $\tau_\zeta^{(i)} = \kappa/2\lambda_i$ is close to the phase transition time τ_G.

For this reason, we shall instead evaluate our field values and corresponding energy densities exactly. In the limit of an instantaneous phase transition occurring at τ_G, the exact solution for each field is given by

$$\phi_\lambda(\tau) = \frac{\pi}{2}[\phi_\lambda(\tau_G)]\lambda\tau_G\left(\frac{\tau}{\tau_G}\right)^{-\kappa_-} \times [J_{\kappa_+}(\lambda\tau_G)Y_{\kappa_-}(\lambda\tau) - Y_{\kappa_+}(\lambda\tau_G)J_{\kappa_-}(\lambda\tau)] \tag{A3}$$

where $\kappa_\pm \equiv (\kappa \pm 1)/2$, where λ refer to the mass eigenvalues for $\tau \geq \tau_G$, and where $J_\nu(z)$ and $Y_\nu(z)$ are Bessel functions of the first and second kind, respectively. From this result we may directly calculate the corresponding energy densities for all $\tau \geq \tau_G$, obtaining

$$\rho_\lambda(\tau) = \frac{\pi^2}{8}[\phi_\lambda(\tau_G)]^2\lambda^4(\tau_G\tau)\left(\frac{\tau}{\tau_G}\right)^{-\kappa}\mathcal{J}_\lambda(\tau) \tag{A4}$$

where

$$\mathcal{J}_\lambda = [J_{\kappa_+}(\lambda\tau_G)Y_{\kappa_-}(\lambda\tau) - Y_{\kappa_+}(\lambda\tau_G)J_{\kappa_-}(\lambda\tau)]^2 + [J_{\kappa_+}(\lambda\tau_G)Y_{\kappa_+}(\lambda\tau) - Y_{\kappa_+}(\lambda\tau_G)J_{\kappa_+}(\lambda\tau)]^2. \tag{A5}$$

Given these exact results, we can now extract the late-time behavior of ρ_λ through the asymptotic expansions

$$J_\alpha(z) \sim \sqrt{\frac{2}{\pi z}}\cos\left(z - \frac{\alpha\pi}{2} - \frac{\pi}{4}\right) + \ldots$$
$$Y_\alpha(z) \sim \sqrt{\frac{2}{\pi z}}\sin\left(z - \frac{\alpha\pi}{2} - \frac{\pi}{4}\right) + \ldots \tag{A6}$$

which hold for $z \gg 1$. With this approximation we find that

$$\bar{\mathcal{J}}_\lambda \approx \frac{2}{\pi\lambda\tau}[J^2_{\kappa_+}(\lambda\tau_G) + Y^2_{\kappa_+}(\lambda\tau_G)] + \ldots, \tag{A7}$$

whereupon use of Eq. (A1) yields the total late-time energy density

$$\bar{\rho} = \frac{\pi^2}{8}A_0^2(\tau_G\tau)\left(\frac{\tau}{\tau_G}\right)^{-\kappa} \times (\lambda_0^4\bar{\mathcal{J}}_{\lambda_0}\cos^2\theta + \lambda_1^4\bar{\mathcal{J}}_{\lambda_1}\sin^2\theta). \tag{A8}$$

The result in Eq. (A8) is an exact expression for the late-time energy density $\bar{\rho}$ in the $\Delta_G \to 0$ limit. Given this result, several things are immediately clear. First, we

observe that in the case of matter-dominated universe (corresponding to $\kappa = 2$), the total late-time energy density simplifies to

$$\bar{\rho} = \frac{A_0^2}{4\tau^2}[2 + (\lambda_0^2 + \lambda_1^2)\tau_G^2 + (\lambda_0^2 - \lambda_1^2)\tau_G^2\cos(2\theta)] = \frac{A_0^2}{2\tau^2}(1 + m_{00}^2\tau_G^2), \tag{A9}$$

where in passing to the second line we have used the results in Eq. (2.32). As a result, all θ-dependence drops out of the final result for the total late-time energy density. In other words, *for a matter-dominated universe, the corresponding late-time energy density is independent of mixing*!

Second, we see that the virial approximation in Eq. (A2) corresponds to the limit in which $\tau_G\lambda_i \gg 1$ for both fields. Indeed, in this limit we see that $\tau_\lambda = \tau_G$ for each field (since both fields are already significantly underdamped when the phase transition occurs), and we can likewise apply the asymptotic expansions in Eq. (A6) to the remaining Bessel functions in Eq. (A7), thereby reproducing the result in Eq. (A2). Further use of the results in Eq. (2.32) then gives the total late-time energy density

$$\bar{\rho} \approx \frac{A_0^2}{2} m_{00}^2 \left(\frac{\tau}{\tau_G}\right)^{-\kappa}, \tag{A10}$$

which is again independent of the mixing—now regardless of the value of κ! However, comparing this result with the exact result in Eq. (A9) for the case of a matter-dominated universe illustrates the imperfect nature of the virial approximation.

APPENDIX B: LATE-TIME ENERGY DENSITIES: AN ALTERNATIVE APPROACH

In the main body of this paper, we have consistently viewed the matrix elements m_{ij}^2 as our independent variables. These have been reparametrized in other forms, giving rise to variables such as $m_{\rm sum}^2$, Δm^2, η, θ, and ξ, but in all cases these independent variables have served as inputs. We then calculated the values of dependent quantities such as the eigenvalues λ_i, energy densities ρ, and so forth. Thus, when we have studied how a certain dependent variable such as a late-time energy density $\bar{\rho}$ depends on the mixing, we have implicitly held the diagonal matrix elements $\bar{m}_{00}^2$ and $\bar{m}_{11}^2$ (or equivalently $\bar{m}_{\rm sum}^2$ and $\bar{\eta}$) fixed, and varied the off-diagonal matrix element $\bar{m}_{01}^2$ (or equivalently the mixing angle $\bar{\theta}$ or mixing saturation $\bar{\xi}$).

However, another possibility is to study the mixing-dependence of quantities such as $\bar{\rho}$ when we hold the late-time *eigenvalues* $\bar{\lambda}_i$ fixed instead. This approach can be useful in situations where the masses of our states are known and it is only the mixing between such states that we wish to vary.

In order to implement this approach, we must first invert our usual algebraic relations and express $\{m_{00}^2, m_{11}^2\}$ in terms of $\{\lambda_0^2, \lambda_1^2\}$. However, we must also recognize that our previous assumption of holding $\{m_{00}^2, m_{11}^2\}$ fixed is also implicitly buried within the definition of another variable: the mixing saturation ξ. Although m_{01}^2 or θ parametrize the mixing in absolute terms and are thus independent of such assumptions, the mixing *saturation* ξ defined in Eq. (2.28) describes the mixing as a fraction of the *maximum allowed* mixing, and this maximum allowed mixing has always been determined relative to a fixed $\{m_{00}^2, m_{11}^2\}$. What we now need, by contrast, is a new variable ξ_λ which describes the mixing as a fraction of the maximum mixing allowed for a given fixed $\{\lambda_0^2, \lambda_1^2\}$. Indeed, this is an entirely different quantity.

Thus, we seek to express $\{m_{00}^2, m_{11}^2, \xi\}$ in terms of $\{\lambda_0^2, \lambda_1^2, \xi_\lambda\}$. Equivalently, defining

$$\lambda_{\rm sum}^2 \equiv \lambda_0^2 + \lambda_1^2 \tag{B1}$$

and

$$\eta_\lambda \equiv \frac{\Delta\lambda^2}{\lambda_{\rm sum}^2} \equiv \frac{\lambda_1^2 - \lambda_0^2}{\lambda_0^2 + \lambda_1^2} \tag{B2}$$

in analogy with $m_{\rm sum}^2$ and η, we would like to express $\{m_{\rm sum}^2, \eta, \xi\}$ in terms of $\{\lambda_{\rm sum}^2, \eta_\lambda, \xi_\lambda\}$. Surprisingly, these relations ultimately take a relatively simple form:

$$\begin{aligned} m_{\rm sum}^2 &= \lambda_{\rm sum}^2 \\ \eta^2 &= \eta_\lambda^2\left(\frac{1-\xi_\lambda^2}{1-\eta_\lambda^2\xi_\lambda^2}\right) \\ \xi &= \eta_\lambda\xi_\lambda. \end{aligned} \tag{B3}$$

These relations in turn imply that

$$m_{01}^2 = \frac{1}{2}(\Delta\lambda^2)\xi_\lambda\left(\frac{1-\eta_\lambda^2}{1-\eta_\lambda^2\xi_\lambda^2}\right). \tag{B4}$$

Likewise, following Eq. (2.30), we can also determine the absolute mixing angle θ via

$$\tan^2(2\theta) = \frac{\xi^2}{\eta^2}(1-\eta^2) = \xi_\lambda^2\left(\frac{1-\eta_\lambda^2}{1-\xi_\lambda^2}\right). \tag{B5}$$

Note that $0 \leq \eta_\lambda \leq 1$, whereas $-1 \leq \eta \leq 1$. Also note that the eigenvalues λ_i^2 are independent of the sign of η (as evident from Fig. 6). It is for this reason that our inverse relations in Eq. (B3) determine η [and likewise $\tan(2\theta)$] only up to an overall sign. Thus, strictly speaking, our variable map really trades $\{\eta, \xi\}$ for $\{\eta_\lambda, \xi_\lambda, \epsilon\}$ where $\epsilon \equiv \text{sign}(\eta) = \pm 1$ is a discrete $\mathbb{Z}_2$ phase. Having $\epsilon > 0$ corresponds to $\eta > 0$ and $\tan(2\theta) > 0$. However, we shall avoid

writing ϵ in what follows, understanding the proper phase choice to be implicit.

In principle, given this variable map, we could now work our way through this entire paper again using the variables $\{\lambda^2_{\rm sum}, \eta_\lambda, \xi_\lambda\}$ rather than the variables $\{m^2_{\rm sum}, \eta, \xi\}$. Indeed, every curve in every plot which has been labeled in terms of certain numerical values for the first set of parameters could now equivalently be relabeled in terms of different numerical values for the second set of parameters. In cases where these curves describe the behavior of a certain quantity such as an energy density as a function of time, or the behavior of a quantity such as a late-time energy density as a function of a phase-transition width Δ_G, these curves would themselves remain intact and continue to take the same shapes as before because our time and Δ_G parameters are independent of our variable change. All that would change are the labels associated with each such curve. Indeed, every feature that appears in our plots for certain fixed values of $\{\bar{m}^2_{\rm sum}, \bar{\eta}\}$ (such as the oscillations due to parametric resonances or the reoverdamping of certain fields) would continue to exist when $\{\bar{\lambda}^2_{\rm sum}, \bar{\eta}_\lambda\}$ are held fixed instead. As a result, nothing we have said in this paper concerning the parametric-resonance or the reoverdamping phenomena would be affected by this change in variables.

The only changes that could potentially occur due to this variable exchange concern the effects of mixing as discussed in Secs. III and IV. Of course, each individual curve within the plots in Figs. 10, 11, 12, 16, and 18 would remain exactly as before and would merely be relabeled using $\{\lambda^2_{\rm sum}, \eta_\lambda, \xi_\lambda\}$ rather than $\{m^2_{\rm sum}, \eta, \xi\}$. However, what *would* change is the resulting *grouping* of these curves into distinct panels in which all curves share common values of, say, η_λ rather than η. Indeed, although the space of all possible curves is not changed, these curves would experience a reshuffling into differently-grouped subsets. As a result, the only physics questions whose answers might change in this process are those which rely on comparisons between the different curves *within* a given grouping. Clearly, questions pertaining to the effects of mixing (i.e., the effects of changing the values of ξ or ξ_λ while holding η or η_λ fixed) are in this category.

Rather than provide an exhaustive study of such regroupings and their outcomes, we shall here content ourselves with providing a concrete example of the effects that can emerge by focusing on a single pair of curves which we shall call A and B. These curves can be taken to show, for example, a late-time energy density $\bar{\rho}$ as a function of Δ_G, but they will differ in their absolute mixing by an angle $\Delta\bar{\theta}$ when $\{\bar{m}^2_{\rm sum}, \bar{\eta}\}$ are held fixed. We will then calculate a third curve B' which also differs from the curve A by the same absolute angle $\Delta\bar{\theta}$ when $\{\bar{\lambda}^2_{\rm sum}, \bar{\eta}_\lambda\}$ are held fixed instead. In this way, comparing curves B and B' will then give us an idea of the difference that can come from the choice of which variables to hold fixed.

TABLE I. Parameters defining the curves A, B, and B' shown in Fig. 31. All curves correspond to $\bar{m}^2_{\rm sum} = \bar{\lambda}^2_{\rm sum} = 4$ and $\epsilon = +$.

	$\bar{\eta}$	$\bar{\xi}$	$\bar{\eta}_\lambda$	$\bar{\xi}_\lambda$	$\bar{\theta}$
A	0.75	0.90	0.96	0.94	19.22°
B	0.75	0.50	0.82	0.61	11.90°
B′	0.88	0.80	0.96	0.84	11.90°

For concreteness, we select curves A and B to be the $\bar{\xi} = 0.9$ and $\bar{\xi} = 0.5$ curves within the right panel of Fig. 11. Within this panel, $\Delta\bar{m}^2 = 4$ and $\bar{\eta} = 0.75$ for both curves. Use of Eq. (B3) then provides us with the corresponding $\{\bar{\lambda}^2_{\rm sum}, \bar{\eta}_\lambda, \bar{\xi}_\lambda\}$ values for each of these curves: we find $\{4.0, 0.96, 0.94\}$ for curve A and $\{4.0, 0.82, 0.61\}$ for curve B. Likewise, use of Eq. (B5) with either the $\{\bar{\eta}, \bar{\xi}\}$ or $\{\bar{\eta}_\lambda, \bar{\xi}_\lambda\}$ variables then tells us that $\bar{\theta}_A \approx 19.22°$ and $\bar{\theta}_B \approx 11.90°$. [The fact that these values emerge in each case using either set of variables is a useful cross-check that the absolute mixing angle $\bar{\theta}$ is indeed invariant across the mapping in Eq. (B3).] We now wish to define a third curve B' such that $\bar{\eta}_\lambda$ is held fixed relative to curve A (i.e., $\bar{\eta}^{(B')}_\lambda = \bar{\eta}^{(A)}_\lambda \approx 0.96$) while the mixing-angle shift $\Delta\bar{\theta}_{AB'} \equiv \bar{\theta}_{B'} - \bar{\theta}_A$ matches $\Delta\bar{\theta}_{AB} \equiv \bar{\theta}_B - \bar{\theta}_A$. This latter condition implies that $\bar{\theta}_{B'} = \bar{\theta}_B \approx 11.90°$, whereupon use of Eq. (B5) tells us that $\bar{\xi}^{(B')}_\lambda \approx 0.84$. Use of Eq. (B3) then tells us that $\{\bar{m}^2_{\rm sum}, \bar{\eta}, \bar{\xi}\} \approx \{4, 0.88, 0.80\}$ for curve B'. These parameter values for curves A, B, and B' are summarized in Table I.

In Fig. 31 we plot curves A, B, and B'. The left panel of Fig. 31 shows these curves for a matter-dominated universe, while the right panel shows these same curves for a radiation-dominated universe. Thus, while curves A and B in the left panel of Fig. 31 respectively correspond to the $\bar{\xi} = 0.9$ and $\bar{\xi} = 0.5$ curves in the right panel of Fig. 11, curves A and B in the right panel of Fig. 31 respectively correspond to the $\bar{\xi} = 0.9$ and $\bar{\xi} = 0.5$ curves in the right panel of Fig. 10.

As evident from Fig. 31, there are significant differences between curves B and B'. Indeed, these curves are sufficiently distinct in the radiation-dominated case (right panel) to turn what would have been an overall *enhancement* of the normalized late-time energy density (curve B) into an overall *suppression* (curve B')! We thus conclude that the question of which variables to hold fixed can indeed be a critical one for certain types of physics questions within certain regions of parameter space. This also highlights the need to be extremely careful about specifying which parameters are held fixed when making statements concerning the effects of varying some parameters relative to others. In this paper, in

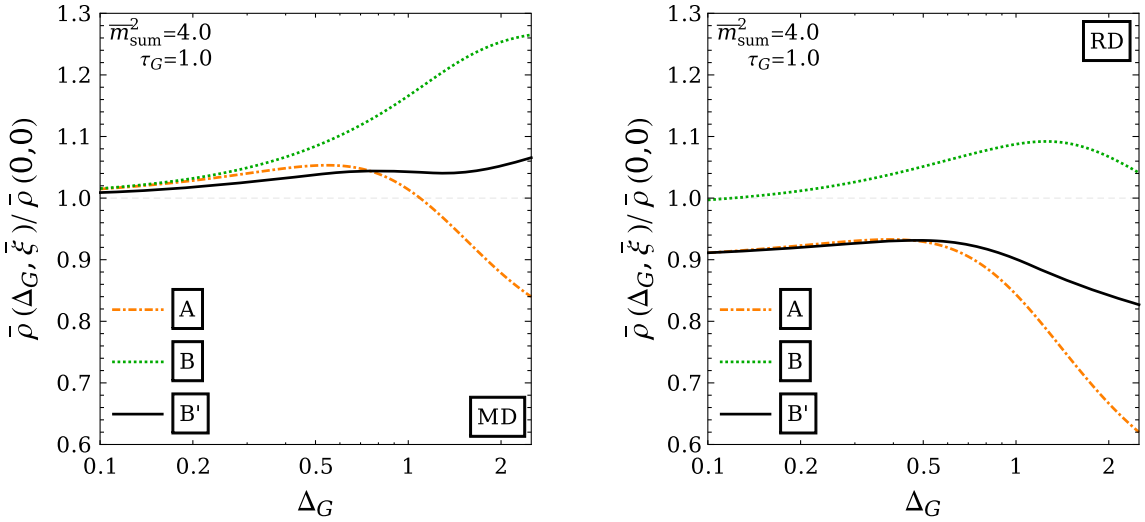

FIG. 31. The late-time energy density $\bar{\rho}$, plotted as a function of Δ_G in a matter-dominated universe (left panel) or radiation-dominated universe (right panel) and normalized to its value for an instantaneous phase transition with zero mixing. In each case, the curves A, B, and B' correspond to the parameter choices in Table I: curves A and B differ by a fixed $\Delta\bar{\theta} = \bar{\theta}_B - \bar{\theta}_A$ while holding $\bar{\eta}$ fixed, while curves A and B' differ by the same $\Delta\bar{\theta}$ but instead hold $\bar{\eta}_\lambda$ fixed. Thus, the difference between curves B and B' in each case illustrates the impact of holding $\bar{\eta}$ fixed versus holding $\bar{\eta}_\lambda$ fixed when performing a shift of the mixing angle $\Delta\theta$ relative to the common curve A.

order to study the effects of mixing, we have consistently held our diagonal mass-matrix elements m^2_{00} and m^2_{11} fixed while varying m^2_{01} (or equivalently varying θ or ξ). This ultimately represents a choice which stems from our desire to perturb the potential generated by our underlying mass-generating phase transition as minimally as possible. However, this issue will be explored further in Ref. [38].

[1] R. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).
[2] R. Peccei and H. R. Quinn, Phys. Rev. D **16**, 1791 (1977).
[3] S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978).
[4] F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978).
[5] See, e.g., J. Jaeckel and A. Ringwald, Annu. Rev. Nucl. Part. Sci. **60**, 405 (2010), 1002.0329 and references therein.
[6] P. Binetruy, G. Girardi, and P. Salati, Nucl. Phys. **B237**, 285 (1984).
[7] K. Griest and D. Seckel, Phys. Rev. D **43**, 3191 (1991).
[8] G. Jungman, M. Kamionkowski, and K. Griest, Phys. Rep. **267**, 195 (1996).
[9] J. L. Feng, eConf **C0307282**, L11 (2003) [arXiv:hep-ph/0405215].
[10] For an introduction, see M. Green, J. Schwarz, and E. Witten, *Superstring Theory, Vols. 1 and 2*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, 1988).
[11] J. Polchinski, *String Theory, Vols. 1 and 2* (Cambridge University Press, Cambridge, 2001).
[12] K. Becker, M. Becker, and J. Schwarz, *String Theory and M-Theory: A Modern Introduction* (Cambridge University Press, Cambridge, 2006).
[13] S. R. Coleman, Nucl. Phys. **B262**, 263 (1985); **B269**, 744 (E) (1986).
[14] A. Kusenko, Phys. Lett. B **405**, 108 (1997).
[15] B. Ratra and P. J. E. Peebles, Phys. Rev. D **37**, 3406 (1988).
[16] J. Khoury and A. Weltman, Phys. Rev. Lett. **93**, 171104 (2004).
[17] J. Khoury and A. Weltman, Phys. Rev. D **69**, 044026 (2004).
[18] P. Brax, C. van de Bruck, A.-C. Davis, J. Khoury, and A. Weltman, Phys. Rev. D **70**, 123518 (2004).
[19] F. Wilczek, Phys. Rev. Lett. **49**, 1549 (1982).
[20] Y. Chikashige, R. N. Mohapatra, and R. D. Peccei, Phys. Lett. **98B**, 265 (1981).
[21] G. B. Gelmini and M. Roncadelli, Phys. Lett. **99B**, 411 (1981).
[22] N. G. Deshpande and E. Ma, Phys. Rev. D **18**, 2574 (1978).
[23] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, Phys. Lett. B **513**, 232 (2001).
[24] R. R. Caldwell, R. Dave, and P. J. Steinhardt, Phys. Rev. Lett. **80**, 1582 (1998).
[25] C. Boehm, T. Ensslin, and J. Silk, J. Phys. G **30**, 279 (2004).
[26] C. Boehm and P. Fayet, Nucl. Phys. **B683**, 219 (2004).
[27] A. Kusenko, M. Loewenstein, and T. T. Yanagida, Phys. Rev. D **87**, 043508 (2013).
[28] K. R. Dienes and B. Thomas, Phys. Rev. D **85**, 083523 (2012).

[29] K. R. Dienes and B. Thomas, Phys. Rev. D **85**, 083524 (2012).
[30] E. Witten, Phys. Lett. **149B**, 351 (1984).
[31] J. P. Conlon, J. High Energy Phys. 05 (2006) 0778.
[32] P. Svrcek and E. Witten, J. High Energy Phys. 06 (2006) 051.
[33] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, and J. March-Russell, Phys. Rev. D **81**, 123530 (2010).
[34] M. S. Turner, Phys. Rev. D **33**, 889 (1986).
[35] K. R. Dienes, E. Dudas, and T. Gherghetta, Phys. Rev. D **62**, 105023 (2000).
[36] K. R. Dienes and B. Thomas, Phys. Rev. D **86**, 055013 (2012).
[37] K. Babu, S. M. Barr, and D. Seckel, Phys. Lett. B **336**, 213 (1994).
[38] K. R. Dienes, J. Kost, and B. Thomas (in preparation).
[39] O. Wantz and E. Shellard, Nucl. Phys. **B829**, 110 (2010).
[40] O. Wantz and E. Shellard, Phys. Rev. D **82**, 123508 (2010).
[41] G. Coughlan, W. Fischler, E. W. Kolb, S. Raby, and G. G. Ross, Phys. Lett. **131B**, 59 (1983).
[42] T. Banks, D. B. Kaplan, and A. E. Nelson, Phys. Rev. D **49**, 779 (1994).
[43] B. de Carlos, J. Casas, F. Quevedo, and E. Roulet, Phys. Lett. B **318**, 447 (1993).
[44] P. Ade *et al.* (Planck), arXiv:1502.01589.
[45] A. D. Linde, Phys. Rev. D **49**, 748 (1994).
[46] E. J. Copeland, A. R. Liddle, D. H. Lyth, E. D. Stewart, and D. Wands, Phys. Rev. D **49**, 6410 (1994).
[47] A. R. Liddle, A. Mazumdar, and F. E. Schunck, Phys. Rev. D **58**, 061301 (1998).
[48] P. Kanti and K. A. Olive, Phys. Rev. D **60**, 043502 (1999).
[49] P. Kanti and K. A. Olive, Phys. Lett. B **464**, 192 (1999).
[50] N. Kaloper and A. R. Liddle, Phys. Rev. D **61**, 123513 (2000).
[51] S. Dimopoulos, S. Kachru, J. McGreevy, and J. G. Wacker, J. Cosmol. Astropart. Phys. 08 (2008) 003.
[52] P. Adshead, R. Easther, and E. A. Lim, Phys. Rev. D **79**, 063504 (2009).
[53] S. Dimopoulos, G. Dvali, and R. Rattazzi, Phys. Lett. B **410**, 119 (1997).
[54] K. Izawa, M. Kawasaki, and T. Yanagida, Phys. Lett. B **411**, 249 (1997).
[55] K. Izawa, Prog. Theor. Phys. **99**, 157 (1998).
[56] K. Harigaya, M. Ibe, K. Schmitz, and T. T. Yanagida, Phys. Lett. B **720**, 125 (2013).
[57] J. H. Traschen and R. H. Brandenberger, Phys. Rev. D **42**, 2491 (1990).
[58] A. D. Dolgov and D. P. Kirilova, Sov. J. Nucl. Phys. **51**, 172 (1990) [Yad. Fiz. **51**, 273 (1990)].
[59] Y. Shtanov, J. H. Traschen and R. H. Brandenberger, Phys. Rev. D **51**, 5438 (1995).
[60] L. Kofman, A. D. Linde, and A. A. Starobinsky, Phys. Rev. Lett. **73**, 3195 (1994).
[61] L. Kofman, A. D. Linde, and A. A. Starobinsky, Phys. Rev. D **56**, 3258 (1997).