Enhanced spectrum of primordial perturbations, galaxy formation, and small-scale structure

Karim H. Seleim * and Amr A. El-Zant †
Centre for Theoretical Physics, The British University in Egypt, Sherouk City 11837, Cairo, Egypt

A. M. Abdel-Moneim
Physics Department, Faculty of Science, Alexandria University, Alexandria 21511, Egypt

(Received 3 March 2020; accepted 12 August 2020; published 8 September 2020)

The standard cosmological structure formation scenario is successful on large scales. Several apparent problems affect it however at galactic scales, such as the small scale problems at low redshift and more recent issues involving early massive galaxy and black hole formation. As these arise where complex baryonic physics becomes important, this is often assumed to be behind the problems. But the same scales are also those where the primordial spectrum is relatively unconstrained, and there are several ways in which it can be modified. We focus on that arising from effects possibly associated with the crossing of high energy cutoff scale by fluctuation modes during inflation. Elementary arguments show that adiabatic evolution cannot modify the near scale invariance, we thus discuss a simple model for the contrary extreme of sudden transition. Numerical calculations and simple arguments suggest that its predictions, for parameters considered here, are more generic than may be expected, with significant modifications requiring a rapid transition. We examine the implications of such a scenario, in this simplest form of sudden jump as well as gradual variants, on the matter power spectrum and halo mass function in light of the limitations imposed by particle production. We show the resulting enhancement and oscillation in the power spectrum on currently nonlinear scales can potentially simultaneously alleviate both the apparent problem of early structure formation and, somewhat counterintuitively, problems at low redshift concerning the abundance of dwarf galaxies, including those too big to fail. We discuss consequences that can observationally constrain the scenario and its parameters, including an inflationary Hubble scale $\lesssim 10^{-8} M_{\rm Pl}$, while touching on the possibility of simultaneous modification of power on the largest scales.

DOI: 10.1103/PhysRevD.102.063505

I. INTRODUCTION

Structure can condense from small density perturbations in a nearly homogeneous universe through gravitational instability. In the context of contemporary cosmology the density perturbations are seeded by quantum fluctuations in a primordial scalar field driving inflation, which later decays as the universe reheats and the standard model particles (and putative dark ones) arise (e.g., [1,2]). The statistical properties of the primordial perturbations thus leave their mark on the cosmic microwave background (CMB) and large scale structure of the galaxy distribution. On these scales, on which they can be inferred with precision, the properties of the perturbations are consistent with a nearly scale invariant primordial spectrum essentially determining their statistics [3].

At a phenomenological level, the simplest models of inflation do predict a near scale invariant primordial power spectrum if the inflaton potential is specified in such a way that the resulting Hubble parameter is nearly constant over a sufficient number of e-folds. Nevertheless, this prediction is not unique [4,5]; indeed, little is known of the microscopic physics of inflation, or the wider particle physics model it may be a part of, and the coupling of the inflaton driving inflation to other fields may lead to changes in the potential that can ruin the predictions of standard slow roll models (e.g., [6–9] and [10–12] for reviews). Stages of singular or rapid evolution of the potential or its derivatives, interrupting slow roll, leave imprints on the primordial spectrum of fluctuations [7,13–22]. Such effects may lead to a variety of "features" and changes that break the scale free spectrum, and may have observable consequences associated with significant enhancement or suppression of power on large scales [23–29], as well as on smaller (galactic and sub-galactic) scales [30–36], including the formation of primordial black holes [37,38]. However the anomalous variation in the inflaton potential needs to be

*karim.selim@bue.edu.eg
Also at Physics Department, Faculty of Science, Alexandria University, Alexandria, Egypt.
†amr.elzant@bue.edu.eg

localized in order for single field slow roll inflation to proceed for a sufficient number of e-folds, and is also limited on large scales by observations suggesting a nearly scale invariant primordial power spectrum [3,39].

In some models, such as DBI inflation, the breaking of scale invariance and associated features in the power spectrum are best represented in terms of sharp temporal variations in the sound speed of perturbations [40–46]. This is still accompanied by corresponding changes in the equation of state of the inflaton and therefore, as in the cases above, by anomalous background evolution. On the other hand, if inflation proceeds modestly longer than the minimal number of e-folds needed to solve problems such as the apparent causal connection in the CMB field ("horizon problem"), primordial perturbations are expected to arise from well within a high energy cutoff scale where new physics may transpire. This gives rise to what is often termed as the trans-Planckian problem. The minimal length scale that it entails can be introduced by modifying the commutation relations or introducing nonlinear dispersion relations for the propagation of fluctuations (early studies include [47–53], while a relatively recent review can be found in [54]).

If the Hubble scale of inflation H is smaller than the scale k_c, where new physics may appear, the background evolution is unmodified by possible new physics; modifications to the scale invariant spectrum may then arise solely from anomalous evolution of the fluctuation modes, rather than of the inflation potential. However, if some form of decoupling is assumed—e.g., motivated by the fact that the power spectrum is evaluated at horizon exit scales $H \ll k_c$—then the corrections can be quite small; of order $(H/k_c)^q$, with $q \gtrsim 1$. In the context of local effective field theory $q = 2$ [55]; if the fluctuation modes are assumed to simply emerge from the quantum foam at conformal time $\eta_c(k)$, depending on the comoving mode wave number k—instead of the usual Bunch-Davies initial conditions taken at $\eta \to -\infty$—then $q = 1$ [56,57]; if the modes emerge from the local adiabatic vacuum then $q = 3$ [58]. A covariant small scale cutoff, introduced by removing field configurations that are off shell by more than a Planck scale leads to $q = 1$ [59].

Much larger corrections are nevertheless possible in principle. This is the case, in particular, if the emergence of the inflating modes from the high energy cutoff scale is assumed to be preceded by nonadiabatic evolution arising from a nonlinear dispersion relation at scales $> k_c$ [47,48,60–62]. Modes then do not emerge from the foam in their lowest energy states. Excited states arise, providing anomalous initial conditions for further evolution and leading to enhancement and oscillations in the power spectrum. The most general parametrization of the effects of a nonadiabatic high energy scale exit would therefore appear to include both phenomena [51,63].

Though physics at these scales is largely unknown, it can in principle be envisioned that the introduction of a cutoff scale in itself can modify the effective dynamics of the fluctuations. Indeed, the introduction of a minimal length scale and a large variety of phenomenological descriptions of "quantum spacetimes," can be characterized by nonlinear dispersion relations [64,65]. At the most intuitive level, a simple hydrodynamic analogy suggests such a modification to the propagation of fluctuations [66–68]. This, much in the same way that an effective macroscopic description of wave propagation through a fluid or lattice may still be employed at wavelengths approaching the interparticle distances, provided this is phenomenologically taken into account through a modified dispersion relation. As, at scales smaller than the interparticle distances waves cannot propagate at all, it is in the transition between such a cutoff scale and the scale on which the standard effective macroscopic description applies that a nonlinear dispersion relation may describe the propagation of fluctuations. At this simplest intuitive level, one may expect the dynamics of a sound wave, initially moving in a medium where interparticle spacing is large and comparable with its wavelength, to keep memory of the anomalous evolution, even after it crosses and propagates into a medium with smaller interparticle spacing, where the effective theory is perfectly valid and "decoupling" is guaranteed. The rough analogy here would be with an inflaton fluctuation mode inflating from wave numbers above the high energy cutoff scale to ones below it, on its way to the horizon. Applications of more sophisticated "analogue" models to the inflationary scenario show that modifications of the dispersion relation can indeed lead to significant changes of power spectrum of field correlations [69].

An important limitation on modifications of the power spectrum through the inclusion of excited states relates to the fact that these are necessarily associated with departures from a vacuum state. And too much excitation can lead to departures significant enough to prohibit inflation from starting and persisting in the first place [70–74]. However, as has been pointed out, the limitations that arise thus may not be very constraining [75–77]. In the present investigation we wish to examine whether, within the limits imposed by particle production, excitations of inflaton modes, stemming from the presence of high energy cutoff scale, can lead to significant and astrophysicaly interesting modifications of the primordial power spectrum.

At present, the nonobservation of significant departures from scale invariance on large (linear) scales, where the primordial spectrum can be rather precisely inferred, seem to embody the main evidence against such modifications. Indeed, observations on scales on which the density perturbation is linear preclude even relatively small modifications [78–82]. Observations are however much less constraining at smaller scales, where they are limited by the Silk damping of the CMB, and by nonlinear structure formation erasing the possibility of directly mapping the observed power spectrum to the primordial one. Precise CMB and large scale structure inference is therefore limited to scales $\gtrsim 10$ Mpc. For a horizon scale of ~ 10 Gpc this spans three orders of magnitudes. More model dependent constraints are available from the Lyman-α forest down to wave numbers roughly corresponding to comoving spatial scales of order of Mpc. Beyond that, the spectrum is currently quite weakly constrained [83–85]. On the other hand, the smallest structures that form in the context of the standard cold dark matter scenario have earth mass and roughly solar system size $\sim 10^{-4}$ pc. From such scales to the smallest scales at which the linear power spectrum can be directly recovered one counts 11 orders of magnitude—nine more than those separating the nonlinear scale to the horizon.

It is not inconceivable that the scale invariance of the primordial power spectrum does not hold in some parts of the aforementioned range. On the contrary, despite the significant successes of the current model of structure formation on large scales [86], through the past couple of decades a variety of problems have arisen on galactic scales. There is a group of quite possibly related longstanding issues connected to the central densities of dark matter halos, and the abundance and dynamical properties of local dwarf galaxies [87,88]; and, in apparent contradiction, more recent issues related to an apparent preponderance of massive old galaxies and supermassive black holes at redshifts $3 \lesssim z \lesssim 9$ that may pose a challenge the current ΛCDM-based structure formation paradigm [89–102].

Such problems appear in the highly nonlinear regime of structure formation; where small density perturbations, born of primordial ones in the presumed inflaton, have sufficiently grown under gravity to form gravitationally bound objects. Since it is also at such scales that complex baryonic physics becomes important, it was natural to suppose that the main determinant lies in complex baryonic physics of galaxy formation and evolution. For example, for the small scale problems at $z = 0$, processes involving energy input to dark matter halos through dynamical friction with baryons [103–111] or through random potential fluctuations driven by starbursts or active galactic nuclei (AGN) [112–118], were invoked. (In addition to suggestions modifying the dark matter particle physics models, as in warm dark matter [119–123], self interacting dark matter [124–129] and fuzzy dark matter [130–134]). Similar attempts are ongoing in the case of the more recent early structure formation issue (some are discussed in Sec. IV C 2).

However, it is also precisely at the nonlinear scales, where baryonic physics becomes important, that the primordial power spectrum is relatively unconstrained. That modifications thereof can be relevant to small scale problems associated with galaxy formation has long been realized [30], but not as extensively investigated as the baryonic solutions discussed above. While feedback from starbursts and AGN is now recognized as a central ingredient of galaxy formation, independently of its possible role in alleviating the aforementioned issues, and whereas massive baryonic clumps, proposed to mediate dynamical friction coupling to dark matter, have since been observed in forming galaxies (e.g., [135,136]), it is also important to further examine mechanisms that address galactic scale problems through modification of the primordial spectrum. Deriving the consequences of such modifications is in itself an interesting tool for understanding the processes from which they may arise in an inflationary era.

In this study we investigate the effect on the power spectrum from field excitations, stemming from nonadiabatic transition through a high energy cutoff regime corresponding to currently nonlinear scales, and within the limits imposed by particle production. We attempt to do this in generic terms, starting from well-defined initial conditions, with linear dispersion relation (but with sound speed different from unity) and examining the effect of the transition. As this solely affects the fluctuation modes, the equation of state of the inflaton, and thus the background evolution, remains unmodified (unlike in cases such as DBI inflation mentioned above), this helps isolate the effect of excitations on the spectrum. We then look for associated effects on the matter power spectrum and dark matter halo mass function.

In the next section, after illustrating in simplest terms how the power spectrum is essentially an adiabatic invariant of the dynamics of inflaton fluctuations, we present and discuss a simple model representing the other extreme of a sudden transition (in an Appendix, we show results that suggest it is generic for a large range of parameters; in a second appendix we discuss the situation when the assumption is relaxed). In Sec. III we discuss what this model entails in more formal terms, evaluating the limits on power spectrum modification in terms of particle production. In Sec. IV, we study, within these limits, the possible modifications on the matter power spectrum and halo mass function. We discuss possible astrophysical consequences and constraints, before presenting our conclusions.

II. ADIABATICITY, SCALE INVARIANCE AND THE SUDDEN EXTREME

A. The evolution of fluctuations

The general quadratic action for inflationary perturbations with sound speed c_s can be expressed in terms of the Mukhanov-Sasaki (MS) variable v as [137–139]:

$$S^{(2)} = \frac{1}{2}\int \mathrm{d}^4x\left(v'^2 - c_s^2(\nabla v)^2 + \frac{z''}{z}v^2\right), \qquad (1)$$

where $z = a\frac{\phi_0'}{\mathcal{H}c_s}$, ϕ_0 is the background inflaton field, $\mathcal{H} = a'/a$, and the primes denote derivative with respect to conformal time η. The evolution of each Fourier mode $v_k(\eta)$ is governed by the MS equation

$$v_k'' + \left(c_s^2 k^2 - \frac{z''}{z}\right) v_k = 0. \quad (2)$$

The MS variable is related to the curvature perturbations by $v = z\mathcal{R}$. This is a quantity of fundamental interest, as it relates primordial quantum fluctuations to the observables, such as CMB anisotropies; the power spectrum of the large scale galaxy distribution; and, ultimately (more indirectly), the formation of smaller scale structures, such as the dark matter halos hosting galaxies. The dimensionless power spectrum of such perturbations is given by

$$\Delta_{\mathcal{R}}^2(k) \equiv \frac{k^3}{2\pi^2} |\mathcal{R}_k|^2, \quad (3)$$

where the right-hand side is evaluated at the horizon ($c_s k = aH$); as in the absence of isocurvature perturbations, the comoving curvature perturbations $\mathcal{R}$ are conserved on superhorizon scales [140]. Scale-invariant perturbations correspond to $\Delta_{\mathcal{R}}^2(k) = \text{const}$. Departures from this can arise if c_s, or the inflationary Hubble scale H, depend on time.

In the standard inflationary scenario, a massless field, and quasi–de Sitter evolution is assumed (and so H is nearly constant throughout the inflationary stage). The associated slow roll parameters, defined as

$$\epsilon \equiv -\frac{\dot{H}}{H^2} \quad \tilde{\eta} \equiv \frac{\dot{\epsilon}}{H\epsilon} \quad \text{and} \quad \kappa \equiv \frac{\dot{c}_s}{Hc_s}, \quad (4)$$

are always much smaller than unity. Any departure from scale invariance is small, and is usually quantified by the spectral tilt parameter n_s:

$$n_s - 1 \equiv \frac{d \ln \Delta_{\mathcal{R}}^2}{d \ln k} = -2\epsilon - \tilde{\eta} - \kappa. \quad (5)$$

The slow roll parameters being small implies that $n_s \approx 1$.

To first order in the slow roll parameters, and assuming canonical kinetic terms, one can also show that

$$\frac{z''}{z} = \frac{1}{\eta^2}\left(2 + 3\epsilon + \frac{3}{2}\tilde{\eta}\right). \quad (6)$$

In this case, to a good approximation, one can rewrite (2) as

$$v_k'' + \left(c_s^2 k^2 - \frac{2}{\eta^2}\right) v_k = 0. \quad (7)$$

If no nonstandard dispersion relation is invoked then $c_s = 1$, and the standard scenario may be fully recovered.

B. Adiabaticity, adiabatic invariants, and primordial power spectrum

1. General context

In the context of Eq. (2), setting $c_s = 1$ can be interpreted as the result of assuming a massless field with linear dispersion relation between the physical frequency ω_{phys} and the physical wave number $k_{\text{phys}} = \frac{k}{a}$, $\omega_{\text{phys}} = k_{\text{phys}}$. However, as discussed in the introduction, this is not a necessity; a modification of the equation of state of the inflaton (e.g., such as in DBI inflation), or modification of the dispersion relation due to modes probing a high energy cutoff scale, beyond which new physics may arise, can change the situation.

In the latter case, beyond a cutoff scale k_c, one can introduce the relevant modification by replacing the square of the comoving wave number k^2 in (2) with

$$k^2 \to k_{\text{eff}}^2(k, \eta) \equiv a^2(\eta)\omega_{\text{phys}}^2\left[\frac{k}{a(\eta)}\right], \quad (8)$$

the main requirement being that the new dispersion relation recovers the linear one for scales $k \ll k_c$ [141]. This dispersion relation is thus necessarily time dependent, as it must transit between two regimes. It can be used to parametrize and reflect the effect of a varying sound speed in Eqs. (2) and (7), the latter applying when the background dynamics is well approximated by standard slow roll. Indeed, in this context, Eq. (2) can be rewritten as

$$v_k'' + \left[k_{\text{eff}}^2(k, \eta) - \frac{z''}{z}\right] v_k = 0. \quad (9)$$

(A more rigorous derivation, based on a variational principle, can be found here [142]).

How does the extra time dependence, that thus arises, affect the power spectrum derived from the above equation? As noticed in several studies, mere time dependence in itself is not sufficient to alter the nearly scale invariant nature of the primordial spectrum of fluctuations. The adiabaticity condition—that is, $|\frac{d\omega}{d\eta}|/\omega^2 < 1$—must be violated. A well known example where this condition is indeed violated invokes the Corley-Jacobson dispersion relation

$$k_{\text{eff}}^2(k, \eta) = k^2 - k^2 |b_m| \left[\frac{k}{k_c a(\eta)}\right]^{2m} \quad (10)$$

(where b_m is a constant and m an integer). The studies [48,141,143], indeed indicated that a modification of the power spectrum, in the form of a change in the spectral index and superimposed oscillations, was possible. However, several criticisms were raised, including the possibility of complex frequencies arising at early times, rendering the quantum field theory ill-defined, and problems related to setting the initial conditions in nonadiabatic regime. To circumvent such issues, a new dispersion relation [142] was proposed, which exhibits linear behavior in the small and large wave numbers, but has intermediate concave region where the adiabaticity is violated locally.

Here we will be considering a simpler scenario, which assumes standard Bunch-Davies type initial conditions, with modified sound speed but still linear dispersion relation. The effective sound speed transits to the standard relation $\omega = k$ as the boundary around k_c is crossed. In this simple controlled context, we wish to estimate the rapidity and steepness of the transition required in order to produce palpable change in the power spectrum. It turns out that such a transition must be quite rapid.

2. The power spectrum as an adiabatic invariant

We now wish to show, in explicit simple terms, that the primordial power spectrum is in fact an adiabatic invariant of the evolution of inflationary perturbation, and thus cannot be significantly modified by any changes in the dispersion relation that keeps the dynamics of the perturbations sufficiently adiabatic.

We will be interested in the case when the dispersion relation is modified due to the fluctuation modes probing scales beyond a high energy cutoff, before they inflate into lower energy scales on their way to horizon exit. Only the effective speed of mode propagation is modified and slow roll inflation of a massless field is assumed to hold in all stages. So, Eq. (7) holds to a good approximation. However, because of the nonstandard dispersion relation assumed, c_s in that equation will not be necessarily unity in all stages. In fact its variation would incorporate changes parametrized by k_{eff} in Eq. (9) above.

In principle c_s in (7) can be either larger or smaller than unity. Perhaps a scenario in which modes do not propagate at all for $k_{\mathrm{phys}} \gg k_c$, and then do so at increasing $c_s \to 1$ as they emerge from the "quantum foam" at $k_{\mathrm{phys}} \sim k_c$, is appealing; it qualitatively connects, for example, to waves propagating in a lattice, which are scattered and dispersed to smaller speeds as one approaches the interparticle spacing, before ceasing to propagate. However analogue models with superluminal speeds $c_s > 1$, beyond the cutoff scale, have also been proposed [68]. As we discuss in Sec. III, for our purposes both situations lead to similar results.

Equation (7) refers to a simple harmonic oscillator with variable frequency. If c_s is constant, the variation solely comes from the second term in the bracket. To separate this effect from that connected to possible variation in c_s at a high energy cutoff transition, we exploit the fact that $k_c \gg H$. This enables one neglect the second term in the brackets of Eq. (7), at scales ($\sim k_c$) around the high energy cutoff transition; as, when modes transit from beyond the cutoff scale k_c to below it, the conformal time $\eta_c = \frac{-1}{a_c H} = \frac{-k_c}{Hk}$. The term in the brackets in the aforementioned equation is then $c_s^2 k^2 (1 - \frac{2}{c_s^2} \frac{H^2}{k_c^2})$. The second term inside this latter bracket is small compared to unity when

$$c_s^2 > 2\left(\frac{H}{k_c}\right)^2. \tag{11}$$

Since we already assume that $k_c \gg H$, this is always the case when $c_s > 1$. We will also assume that this condition is satisfied when considering the case of $c_s < 1$ [144].

This leaves us with an equation of a harmonic oscillator with frequency $\omega = c_s k$. The adiabatic invariant for a standard harmonic oscillator with specific energy E and frequency ω is $J = \frac{E}{\omega}$. Taking the modulus of the amplitude and the velocity $v_k'(\eta) = i\omega_k v_k(\eta)$, the energy of the oscillator is $E = \omega^2 |v_k|^2$. Whatever the evolution at scales above k_c, as long as it is adiabatic J is conserved. Moreover, at scales $< k_c$ one must recover the standard linear dispersion relation, and so $\omega = k$. At such scales, relevant to eventual horizon crossing, one then has

$$J = k|v_k|^2. \tag{12}$$

Comparing this with the standard slow roll inflationary power spectrum

$$\Delta_{\mathcal{R}}^2(k) = \frac{kH^2}{2\pi^2} |v_k|^2, \tag{13}$$

one finds that they are equivalent up to a (nearly) constant factor $\frac{H^2}{2\pi^2}$.

With sufficiently adiabatic evolution through the transition at k_c no significant change to the power spectrum can occur. Any appreciable effect could then result solely from small variations in H, or from second term in bracket of Eq. (7), which also turns out to be quite modest, as may be expected given the quadratic correction at any physical scale $(k_{\mathrm{phys}}/H)^{-2}$. For this implies again that this term is smaller than the first until modes are quite close to existing the horizon. In Appendix A 1, we show numerical calculations that corroborate this contention in the context of the simple model described below.

C. A toy model of the sudden extreme

As we have seen, any adiabatic frequency change, due to nonstandard evolution of modes beyond a high energy cutoff scale, will not alter the nearly scale free form of the resulting power spectrum. We thus consider the opposite extreme; that of a sudden change in the sound speed at k_c, while employing the same approximation of neglecting the second term in the bracket of Eq. (7). The procedure again separates changes in the power spectrum due to variations in c_s at around $k_c \gg H$ from any time dependence connected to the second term in the above equation at much smaller physical wave numbers.

The change in sound speed across the transition is equivalent to a sudden change in frequency. To illustrate such a situation in simplest terms, we consider the effect of such a change on a simple harmonic oscillator, with initial amplitude A, frequency $\omega_{\rm in}$ and phase ϕ. Its evolution is given by

$$X_{\rm in}(t) = A\cos(\omega_{\rm in}t + \phi) \tag{14}$$

$$X'_{\rm in}(t) = -A\,\omega_{\rm in}\sin(\omega_{\rm in}t + \phi), \tag{15}$$

with

$$A = \sqrt{X^2_{\rm in}(t) + \frac{V^2_{\rm in}(t)}{\omega^2_{\rm in}}} \tag{16}$$

$$\phi = \arccos\left(\frac{X_{\rm in}(t)}{A}\right) - \omega_{\rm in}t. \tag{17}$$

Suppose that at some moment $t = t_s$ the spring constant is suddenly altered, and the corresponding frequency of the oscillator changes to $\omega_{\rm out}$. Then, for $t \geq t_s$,

$$X_{\rm out}(t) = B\cos(\omega_{\rm out}(t - t_s)) + C\sin(\omega_{\rm out}(t - t_s)) \tag{18}$$

$$X'_{\rm out}(t) = -B\,\omega_{\rm out}\sin(\omega_{\rm out}(t - t_s)) + C\omega_{\rm out}\cos(\omega_{\rm out}(t - t_s)). \tag{19}$$

Matching the initial and final states at t_s one obtains

$$B = X_{\rm in}(t_s) \tag{20}$$

$$C = \frac{\dot{X}_{\rm in}(t_s)}{\omega_{\rm out}}. \tag{21}$$

So the evolution after the jump can be expressed in terms of the initial state at the jump as

$$X_{\rm out}(t) = X_{\rm in}(t_s)\cos(\omega_{\rm out}(t - t_s)) + \frac{\dot{X}_{\rm in}(t_s)}{\omega_{\rm out}}\sin(\omega_{\rm out}(t - t_s)), \tag{22}$$

with the amplitude and phase changing after the jump.

We now apply this toy model to attempt to mimic the evolution of fluctuations due to a sudden change in frequency (or again, effectively sound speed) of propagation of inflaton fluctuations. In our approximation the evolution is effectively governed by two independent harmonic oscillators, due to the complexity of the mode functions in (7). Thus, for the real part,

$$X_{{\rm in}_r}(\eta) = A_r\cos(\omega_{\rm in}\eta + \phi_r) \tag{23}$$

$$\dot{X}_{{\rm in}_r}(\eta) = -A_r\omega_{\rm in}\sin(\omega_{\rm in}\eta + \phi_r), \tag{24}$$

and for the imaginary part we have

$$X_{{\rm in}_i}(\eta) = A_i\cos(\omega_{\rm in}\eta + \phi_i) \tag{25}$$

$$X'_{{\rm in}_i}(\eta) = -A_i\omega_{\rm in}\sin(\omega_{\rm in}\eta + \phi_i). \tag{26}$$

The sudden step will here correspond to conformal time η_c, when an inflating mode crosses a physical the wave number k_c, where "new physics" may arise. Applying the step condition as previously, for the real part we find

$$X_{{\rm out}_r} = X_{{\rm in}_r}(\eta_c)\cos[\omega_{\rm out}(\eta - \eta_c)] + \frac{X'_{{\rm in}_r}}{\omega_{\rm out}}\sin[\omega_{\rm out}(\eta - \eta_c)]. \tag{27}$$

Similarly, for the imaginary part

$$X_{{\rm out}_i} = X_{{\rm in}_i}(\eta_c)\cos[\omega_{\rm out}(\eta - \eta_c)] + \frac{X'_{{\rm in}_r}}{\omega_{\rm out}}\sin[\omega_{\rm out}(\eta - \eta_c)]. \tag{28}$$

The complete solution then is

$$v_k(\eta) = X_{{\rm out}_r}(\eta) + iX_{{\rm out}_i}(\eta). \tag{29}$$

This can be evaluated for each k, with $\omega_{k_{\rm out}} = k$, given $\omega_{k_{\rm in}} = c_s k$. As mentioned above (and checked in Appendix A 1) usage of Eq. (29), in order to evaluate the effect on the primordial power spectrum of a sudden step in c_s and ω at k_c, returns a good approximation. The results of Appendix A 2 also suggest that the sudden jump scenario itself turns out to be much more generic to any appreciable change in the power spectrum than may seem *a priory*. We now discuss how the power spectrum is evaluated and the modifications to the standard near scale invariant form that arise.

1. The power spectrum of primordial fluctuations

A mode corresponding to comoving wave number k crosses the high energy cutoff scale at $\eta_c = -\frac{k_c}{Hk}$. At $\eta \ll \eta_c$ we assume Bunch-Davies type initial conditions but with $\omega_k = c_s k$, with $c_s \neq 1$. Thus, before the transition in sound speed (and frequency),

$$v_k(\eta) \to \frac{1}{\sqrt{\omega_{k_{\rm in}}}}e^{i\omega_{k_{\rm in}}\eta}. \tag{30}$$

For the initial amplitudes one then has

$$A_r = A_i = \frac{1}{\sqrt{\omega_{k_{\rm in}}}}, \quad \phi_r = 0, \quad \phi_i = \frac{\pi}{2}. \tag{31}$$

Modes with physical wave numbers larger than k_c at the start of inflation undergo a frequency change such that $\omega_{\rm in}/\omega_{\rm out} = c_s$ (where c_s refers to the value, different from unity, before the crossing). Modes with smaller wave numbers do not cross the high energy cutoff scale and their frequency remains unmodified (we discuss how the transition scale is connected to current comoving scales in Sec. IV).

All modes eventually cross the horizon. Using Eqs. (29) and (3), one can evaluate the power spectrum in the context of our simplified model when H is given. This is done at horizon crossing when $\eta = \eta_H$. Alternatively, one can also use Eq. (7) to calculate the power spectrum numerically at superhorizon scales, as done in the Appendix for purpose of comparison and evaluating the relevance of the model.

In de Sitter inflation H is exactly constant, and all modes are assumed to exit the horizon at time $\eta_H = -1/k$. Since again a standard dispersion relation must reign beyond the high energy cutoff scale k_c, one expects $\omega_{\rm out} = k$. All modes then leave the horizon at the same phase and oscillations implied by Eqs. (27) and (28) do not appear in the power spectrum; only enhancement is found at scales undergoing the jumps. Numerically, Eq. (7) can be used to obtain similar results (cf. Appendix A 1).

The Hubble parameter in more realistic models of inflation must vary slowly with time. The variations imply that modes do not leave the horizon at the same phase, and oscillations as well as enhancement appear in the primordial power spectrum. As a simple generic example, we will adopt power-law inflation [145,146] where (in proper time), $a(t) \approx t^p$, with $p > 1$. This corresponds to an inflation potential of the form

$$V(\phi) = V_0 \exp\left[-\frac{1}{M_{\rm Pl}}\sqrt{\frac{2}{p}}(\phi - \phi_i)\right], \tag{32}$$

where $M_{\rm Pl}$ is the reduced Planck mass. with slow-roll parameters given by

$$\epsilon_v = \frac{M_{\rm Pl}^2}{2}\left(\frac{V_\phi}{V}\right)^2 = \frac{1}{p}, \qquad \tilde{\eta}_v = M_{\rm Pl}^2 \frac{V_{\phi\phi}}{V} = \frac{2}{p}, \tag{33}$$

where $V_\phi = \frac{dV}{d\phi}$. The scale factor and the Hubble parameter become

$$a(\eta) = \left(\frac{\eta}{\eta_{\rm in}}\right)^{\frac{p}{1-p}}, \qquad H(\eta) = -\frac{p}{p-1}\left(\frac{\eta}{\eta_{\rm in}}\right)^{\frac{p}{p-1}}\frac{1}{\eta}, \tag{34}$$

with $\eta_{\rm in} = \frac{t_{\rm in}}{1-p}$. With these forms for the evolution of the scale factor and Hubble parameter one can again use Eq. (29) in conjunction with (3) to evaluate the power spectrum, or numerically integrate Eq. (2), which now takes the following form

$$v_k''(\eta) + \left[c_s^2 k^2 - \frac{2p^2 - p}{(1-p)^2}\frac{1}{\eta^2}\right] v_k(\eta) = 0. \tag{35}$$

We will generally use $H = 10^{-4} M_{\rm Pl}$, $p = 55$. and $\eta_i = -10^4 M_{\rm Pl}^{-1}$, in order to get the correct normalization and tilt of the power spectrum on linear scales using power law inflation. The results, when rescaled accordingly, are valid for other values of H, since the relative enhancement of the power spectrum depends only on the in and out frequency ratio of oscillations. The connection between the cutoff scale k_c and the corresponding comoving scale depends on H/k_c rather than the absolute values (Sec. IV A 2). As shown in Appendix A 1, the step enhancement in the power spectrum is accompanied by oscillations in this more generic (as opposed to de Sitter) case.

III. BROKEN INVARIANCE AND PARTICLE PRODUCTION

A. General solution in terms of Bogoliubov coefficients

We now consider what the simplified sudden step model actually implies in terms of quantum fluctuations in an inflaton. For this purpose we translate it to the language Bogoliubov expansion and coefficients. In this context, the high energy cutoff transition will be seen to lead to excitations of the field and particle production. The excitations will take place as a result of transitions between well defined time independent in and out states. They invariably lead to enhancement in the power spectrum, generically accompanied by oscillations.

B. Generic enhancement in power spectrum

Using (7), and again invoking the approximation of neglecting the $\frac{-2}{\eta^2}$ term due to $k_c \gg H$, we get the mode function differential equation of a massless free scalar field in Minkowski spacetime, with $\omega_k = k$. Assuming the field to be initially in the vacuum state $|0\rangle$, the amplitude of the vacuum fluctuations (the square root of the power spectrum) are given in terms of the vacuum mode function corresponding to the Bunch-Davies initial conditions (30). Then, nonadiabatic evolution (whether sudden or not), can transform this initial vacuum state $|0\rangle$ to one with excitations, with respect to the old annihilation operator $\hat{a}_{\rm k}^-$.

To find the effect of such excitations on the power spectrum after the transition is complete, one can proceed as follows. First by writing the mode expansion of the field operator in terms of the annihilation operator $\hat{b}_{\rm k}^-$ of $|0\rangle$ and its complex conjugate

$$\hat{\chi} = \frac{1}{\sqrt{2}}\int (e^{i\mathbf{k}\cdot\mathbf{x}}\mu_k^*\hat{b}_{\mathbf{k}}^- + e^{-i\mathbf{k}\cdot\mathbf{x}}\mu_k\hat{b}_{\mathbf{k}}^+)\frac{d^3\mathbf{k}}{(2\pi)^{3/2}}, \tag{36}$$

and then computing the two point correlation function in the state $|0\rangle$ using this operator. One can then define the amplitude of the quantum fluctuations in terms of the new mode function $\mu_k(\eta)$ as

$$\Delta_\mu(\eta) = \frac{1}{2\pi} k^{3/2} |\mu_k(\eta)|. \tag{37}$$

This new normalized mode function is a linear combination of the old one and its complex conjugate. Using Bogoliubov coefficients, it can be written as

$$\mu_k(\eta) = \alpha_k v_k(\eta) + \beta_k v_k^*(\eta). \tag{38}$$

Thus we have

$$\Delta_\mu(\eta) = \frac{1}{2\pi} \frac{k^{3/2}}{\sqrt{\omega_k}} [|\alpha_k|^2 + |\beta_k|^2 + 2\mathrm{Re}(\alpha_k \beta_k^* e^{2i\omega_k \eta})]^{1/2}. \tag{39}$$

The second coefficient β refers to excitations away from the vacuum state; as we will see below it directly counts particle production. The ratio of the primordial power spectrum after and before the sudden change can be expressed as

$$\frac{\Delta_\mu^2(\eta)}{\Delta_v^2(\eta)} = 1 + 2|\beta_k|^2 + 2\,\mathrm{Re}(\alpha_k \beta_k^* e^{2i\omega_k \eta}). \tag{40}$$

Averaging over a period larger the periodic time of the system (or in generic inflation models, over horizon exists of the different modes with different H), eliminates the oscillating term. The main result is that excitations away from the vacuum state lead to typically larger RMS fluctuations and power spectrum.

An important point to note here is that the generic enhancement in the power spectrum will occur whether the ‘jump’ in frequency is upward—that is whether $\omega_{\mathrm{in}} < \omega_{\mathrm{out}}$—or the downward, with $\omega_{\mathrm{in}} > \omega_{\mathrm{out}}$. Or, assuming a dispersion relation $w = c_s k$ to govern the propagation of fluctuations before and after the jump, the power spectrum will be enhanced whether c_s is larger before the jump, or whether it is larger afterwards. This is seen explicitly below.

C. Relations between the coefficients and the frequencies

The above does not necessarily assume instantaneous transition jumps between the well defined in to out states, just that a time dependent transition occurred. In our simplified model we have two regions connected by a sudden jump, which enables one to calculate the Boguliubov coefficients explicitly in terms of the in and out frequencies.

If one labels the initial vacuum as $|0_{\mathrm{in}}\rangle$, and the final vacuum $|0_{\mathrm{out}}\rangle$. Before the jump, and assumes the scalar field is in the initial vacuum state, the mode function is

$$v_k^{(\mathrm{in})}(\eta) = \frac{1}{\sqrt{\omega_{k_{\mathrm{in}}}}} e^{i\omega_{k_{\mathrm{in}}}\eta}, \tag{41}$$

for $\eta < \eta_c$. Before the jump, the frequency $\omega_{\mathrm{in}} = c_s k$ with $c_s \neq 1$. In order to connect with standard inflationary scenario, the frequency after the jump is $\omega_{k_{\mathrm{out}}} = k$. The final frequency $\omega_{k_{\mathrm{out}}}$ is therefore necessarily different from the initial one $\omega_{k_{\mathrm{in}}}$. This causes excitations in the field, which modify the power spectrum.

After the jump, the mode function $v_k^{(\mathrm{in})}(\eta)$ evolves into the superposition of $v_k^{(\mathrm{out})}(\eta)$ and its complex conjugate:

$$v_k^{(\mathrm{in})}(\eta) = \frac{1}{\sqrt{\omega_{k_{\mathrm{out}}}}} [\alpha_k^* e^{i\omega_{k_{\mathrm{out}}}(\eta-\eta_c)} - \beta_k e^{-i\omega_{k_{\mathrm{out}}}(\eta-\eta_c)}]. \tag{42}$$

The Bogoliubov coefficients α_k, β_k are determined by the requirement that the solution and its first derivative must be continuous at the jump, that is at $\eta = \eta_c$. The result is

$$\alpha_k = \frac{e^{-i\omega_{k_{\mathrm{in}}}\eta_c}}{2} \left(\sqrt{\frac{\omega_{k_{\mathrm{in}}}}{\omega_{k_{\mathrm{out}}}}} + \sqrt{\frac{\omega_{k_{\mathrm{out}}}}{\omega_{k_{\mathrm{in}}}}} \right) \tag{43}$$

$$\beta_k = \frac{e^{i\omega_{k_{\mathrm{in}}}\eta_c}}{2} \left(\sqrt{\frac{\omega_{k_{\mathrm{in}}}}{\omega_{k_{\mathrm{out}}}}} - \sqrt{\frac{\omega_{k_{\mathrm{out}}}}{\omega_{k_{\mathrm{in}}}}} \right) \tag{44}$$

This explicitly shows that the time dependence introduced by assuming a sufficiently rapid transition from in to out states can lead to significant excitation in the inflaton field for large enough frequency ratio. As is clear, the absolute values of β_k and α_k derived above do not depend on whether $\omega_{\mathrm{in}} > \omega_{\mathrm{out}}$ or the reverse. This again shows that generic enhancement in the power spectrum is expected, independent of the direction of the jump.

D. Limits from particle production

As we have seen, excitations of the inflaton generically lead to enhanced power spectrum. In Sec. IV below, we will suggest that these may have important consequences at galactic scales, at both high and low redshifts, pertaining to such apparent problems as the dearth of dwarf galaxies, “too big to fail” and early galaxy formation, while maintaining a standard spectrum at scales where it is highly constrained. But how much excitations of the field can one have without ruining the inflationary scenario itself? Indeed, the exponential expansion during inflation hinges on a dark energy equation of state, too much excitation and particle production can turn it instead into a radiation field, with deceleration replacing the exponential expansion.

The radiation energy density associated with the relativistic particles, which can be assumed to be produced through excitations of the field, may be expressed as [76]

$$\langle\rho\rangle = \int_{k_{\rm phys}=H}^{k_{\rm phys}=k_c} d^3\mathbf{k}_{\rm phys}\omega_{\rm phys}(k_{\rm phys})n_{k_{\rm phys}}, \tag{45}$$

where $k_{\rm phys}$ and $\omega_{\rm phys}$ are the physical wave numbers and frequencies. The occupation number of excited states can be expressed in terms of the second Bogoliubov coefficient as $n(k) = |\beta_k|^2$. In the relevant integration range the relation between the wave numbers and frequencies is linear, and the integral is dominated by larger values of $k_{\rm phys}$. In this case, $\langle\rho\rangle \approx \beta^2 k_c^4$, where β corresponds to β_k at larger values of k dominating the integral.

In the context of the sudden step scenario β_k is some nonzero constant for modes affected by the jump (and zero otherwise), and the above estimate is rigorously justified. In order for inflation to start and proceed then, $\beta^2 k_c^4$ must be smaller than the energy density scale of inflation $H^2 M_{\rm Pl}^2$. This leads to the condition

$$|\beta| < \frac{M_{\rm Pl}H}{k_c^2}. \tag{46}$$

If $k_c = M_{\rm Pl}$ this is small for $H \ll k_c$. However much smaller cutoff scales may in principle be allowed (and claimed all the way down to the TeV scale e.g., [147–149]; also [150] for review). For the largest field inflation allowed by recent data, with $H \lesssim 3\times 10^{-5} M_{\rm Pl}$ and relatively conventional high scale $k_c \gtrsim 10^{-3} M_{\rm Pl}$, one finds $|\beta| \lesssim 30$ as an upper limit. In general, one only needs $k_c/M_{\rm Pl} \approx H/k_c$ to get a Bugoliubov coefficient of order one. The backreaction condition above may thus in principle allow for large modifications that could be tested and constrained observationally, even in the nonlinear regime of structure formation.

As long as (46) is satisfied inflation can start and proceed, but in order to obtain a near invariant spectrum on large scales, the time derivative of the backreaction energy must also be small as the large scale modes exit the horizon. The limits of integration in Eq. (45) is an upper limit on backreaction energy, which assumes that the whole interval between k_c and H is filled with excited states corresponding to modes that have already crossed k_c. In this case the time derivative $\frac{d\langle\rho\rangle}{dt} \sim \beta^2 H^3 \dot{H}$ is much smaller in absolute value than the change in energy density of the inflaton $\sim M_{\rm Pl}^2 H\dot{H}$ for values of β^2 of interest. However, at earlier times, as modes are crossing k_c and filling up the interval down to H, the integration interval is variable, the time derivative of the backreaction energy $d\frac{\langle\rho\rangle}{dt} \sim \beta^2 k_{\rm phys}^3(k_c)\dot{k}_{\rm phys}(k_c) \approx \beta^2 H k_{\rm phys}^4(k_c)$ can be much larger (here $k_{\rm phys}(k_c)$ refers to the physical wave number of the first scale that crosses k_c; it decreases as the mode inflates toward the horizon, when $k_{\rm phys}(k_c) = H$). This leads to the constraint $\beta^2 \lesssim \epsilon \frac{H^2 M_{\rm Pl}^2}{k_c^4}(\frac{k_c}{k_{\rm phys}})^4$, where $\epsilon = -\dot{H}/H^2$. A more detailed treatment gives a similar constraint, $\beta^2 \lesssim 2(6\pi)^2 \epsilon \frac{H^2 M_{\rm Pl}^2}{k_c^4}(\frac{k_c}{k_{\rm phys}})^4$ [151].

The effect of the changing energy density as the excited states are filling up the interval between k_c and H can be quite complicated, as it would require evaluation of the modified evolution, taking into account the rescaling of the energy density (which itself can act as vacuum energy [76]). Here we just point out that, simply assuming the usual relation $\epsilon = \frac{H^2}{8\pi^2 P_0}$ to hold when the effect is small enough, leads to the condition

$$|\beta| \lesssim \left(\frac{2\times 10^{-9}}{P_0}\right)^{1/2} \times 6.7\times 10^4 \left(\frac{H}{k_{\rm phys}(k_c)}\right)^2, \tag{47}$$

with $P_0 \approx 2\times 10^{-9}$ the standard characteristic value of the standard primordial power spectrum of scalar fluctuations, This rough estimate suggests that $|\beta|$ can be of order one, without affecting the power spectrum on larger scales exiting the horizon, if these scales exit when the spatial physical scale that first crosses the high energy threshold has inflated enough to be about 0.004 times the size of the horizon. We further discuss the possible interpretation of this constraint in Sec. IVA 2.

As the ratio of the power spectrum modified by excitations to the vacuum power spectrum scales as $1 + 2\beta^2$, considerable modifications may be allowed in principle, if $|\beta|$ is of order one or larger. In the following we consider possible consequences of, and constraints on, such modification on currently nonlinear scales, where existing observational constraints are relatively weak and apparent problems with galaxy formation at low and high redshift arise.

IV. MATTER POWER SPECTRUM AND HALO MASS FUNCTION

In this section we examine some possible astrophysical implication of the sudden change of frequency at a high energy cutoff. For this purpose we compute the linear matter power spectrum and the dark matter halo mass function. The modified halo mass function will be of interest, particularly in terms of its possible observable consequences on the galaxy stellar mass function. For the actual calculations we assume a λCDM universe with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $h = 0.7$, and RMS dispersion in the density field at 8 h^{-1} Mpc at $z = 0$, $\sigma_8 = 0.8$.

A. The matter power spectrum

1. Evaluation procedure

The power spectrum of perturbation in the CDM is evaluated from

$$P(k,a) = \frac{4}{9}\frac{k^4 P_i(k)}{\Omega_m^2 H_0^4} T^2(k) D^2(a), \tag{48}$$

where $P_i(k) \equiv \Delta^2_{\mathcal{R}}(k)$ is the primordial power spectrum, $D(a)$ the linear growth factor, and H_0 is the present value of the Hubble parameter.

As we will be primarily interested in generic consequences, rather than detailed comparison with data, for this purpose we generally use the BBKS fitting form [152]

$$T\left(x \equiv \frac{k}{k_{eq}}\right) = \frac{\ln(1+0.171x)}{0.171x}[F(x)]^{-1/4}, \tag{49}$$

with $k_{\rm eq} = 0.073\ \Omega_m h^2\ {\rm Mpc}^{-1}$ and

$$F(x) = 1 + 0.284x + (1.18x)^2 + (0.399x)^3 + (0.490x)^4. \tag{50}$$

For the growth factor, we use [153]

$$D(z) = \frac{D^+(z)}{D^+(z=0)}, \tag{51}$$

where

$$D^+(z) = \frac{5\Omega_m}{2}\frac{H(z)}{H_0}\int_z^\infty \frac{(1+z')dz'}{[H(z')/H_0]^3}, \tag{52}$$

with

$$H(z) = H_0\sqrt{\Omega_m(1+z)^3 + (1-\Omega_m)}. \tag{53}$$

We have verified our results against full solution of the perturbation equations using the public code CLASS (class-code.net), and show results using this code in examining the predictions of the enhanced primordial spectrum scenario in the more general case when the assumption of sudden jump is relaxed (Appendix B and Fig. 8).

2. Choice of jump scale

CMB and large scale structure observations place quite tight constraints on the amplitude of the primordial power spectrum on scales ∼10 Mpc or larger, We now show how smaller scales can be affected by a modified power spectrum, while larger scales remain unaffected, if inflation proceeds for approximately the number of e-folds needed to solve the horizon problem.

Observable inflation takes place after the comoving spatial scale $k^{-1}(H_0) \sim H_0^{-1}$ exits the horizon; it is characterized by the minimum number of e-folds needed to solve the horizon problem: $N[k(H_0)] = \ln[a_{\rm end}/a_{k(H_0)}]$, where the subscripts denote the end of inflation and the epoch of horizon exit of the scale $k^{-1}(H_0)$. The condition can be written as

$$\frac{a_{k(H_0)}}{a_0} = \frac{H_0}{H}, \tag{54}$$

where, a_0 is the current scale factor. H is the Hubble parameter at the horizon exit of the scale $k(H_0)^{-1}$ during inflation.

Given a comoving scale k, one may ask when it was equal to a given physical scale k_c during inflation. This gives the following condition

$$\frac{k}{a_0 H_0} = \frac{a_c k_c}{a_{k(H_0)} H}. \tag{55}$$

As an example, we set $k(k_c) = 1\ {\rm Mpc}^{-1}$, $H = 10^{-4} M_{\rm Pl}$, and $k_c = M_{\rm Pl}$. We then find that $a_c \sim a_{k(H_0)}$; that is, at the time the current horizon scale H_0^{-1} exits the horizon during inflation, the comoving spatial scale ∼1 Mpc is of the order of the Planck length.

This general picture is reproduced even if the cutoff scale k_c is not the Planck scale. All one needs is $H/k_c \approx 10^{-4}$. If inflation proceeds for a number of e-folds larger than the number $N_{\rm min} = N[k(H_0)]$ required to solve the horizon problem, then the "jump scale" can still correspond to $k(k_c)) = 1\ {\rm Mpc}^{-1}$ if $H/k_c < 10^{-4}$. In general, the number of e-folds allowed, with $k(k_c)$ corresponding to the smallest comoving spatial scale affected by the high energy cutoff transition, is

$$N = N_{\rm min} + \ln\left[\left(\frac{k_c}{H}\right)\left(\frac{k(H_0)}{k(k_c)}\right)\right]. \tag{56}$$

The scale $k(H_0) \approx 10^{-4}\ {\rm Mpc}^{-1}$ is fixed by the present size of the horizon, while $k(k_c) = 1\ {\rm Mpc}^{-1}$ happens to roughly correspond to the largest scale on which significant modification of the power spectrum would not affect its inference from galaxy cluster counts and lensing surveys (but, depending on the exact scale, not necessarily Lyman-α bounds, as discussed in Sec. IV C 3). Larger values of $k(k_c)$ are in principle possible, and in this case the power spectrum can be modified on smaller scales, affecting smaller nonlinear structures. However, if one takes into account our crude estimate of the time variation of the backreaction, this may be constrained. For, as mentioned in relation to Eq. (47), to maintain $|\beta|$ of order 1, one may need $H/k_{\rm phys}(k_c) \gtrsim 0.004$. If $k(k_c) \approx 1\ {\rm Mpc}^{-1}$, the comoving scale exiting the horizon when this is satisfied is $\approx 0.004\ {\rm Mpc}^{-1}$. Larger scales, with smaller wave numbers, can be affected if one insists on $|\beta| \gtrsim 1$. In the context of the simplest scenario with constant β_k beyond the cutoff regime, the power spectrum may be modified on such scales. This may be allowed on comoving scales $k < 0.004\ {\rm Mpc}^{-1}$, and may even be relevant to supposed anomalies of the CMB on large scales, but not on smaller spatial scales, where modifications are tightly constrained. That changes in the power spectrum on the largest scales

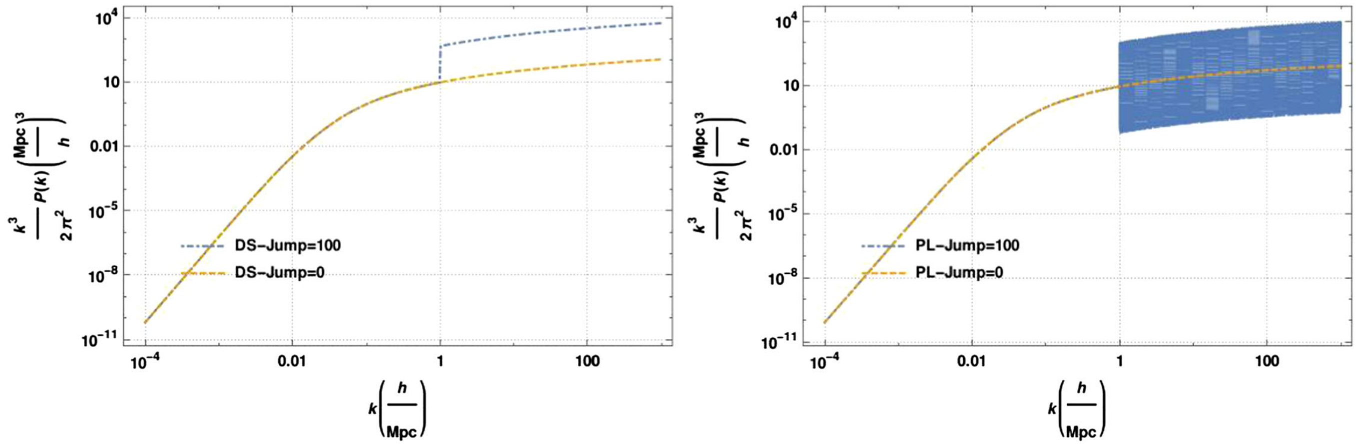

FIG. 1. Dimensionless matter power spectra at $z = 0$. The perturbed cases correspond to sudden jumps of a factor of 100 in mode frequency (and therefore sound speed) due to shift in dispersion relation at the high energy cutoff scale (chosen to correspond to a comoving wave number $k = 1h/\mathrm{Mpc}$ as discussed in the text). Spectra are shown for a de Sitter background (left) and corresponding power law inflation model as a generic example (cf. Sec. II C 1). The oscillations in the latter case are absent in the de Sitter one due to all modes leaving the horizon at the same phase. The frequency ratio corresponds to a Boguliubov coefficient $|\beta| = 4.95$ [Eq. (44)]. Note that, for power law inflation, there is net enhancement despite the strong oscillations, which appear symmetric around the unperturbed spectrum on the logscale. This will result in similar mass dispersions in the de Sitter and power law models, where the smoothing also leads to gradual enhancement despite the sudden jump at the cutoff scale (Fig. 2).

may be connected with backreaction associated with initial evolution has already been noted (e.g., [154]), and may be of interest in the present context, but its proper examination is beyond our present scope.

Here we will be mainly interested in the enhancement of the power spectrum on large nonlinear scales, corresponding to $k \approx \mathrm{Mpc}^{-1}$, because of the particularly interesting consequences for galaxy formation we discuss in Section IV C. Fig. 1 shows the resulting dimensionless matter power spectrum for a jump corresponding to ratio of sound speeds (or in and out frequencies) of 100 on such scales. The ratio is associated with a Boguliubov coefficient $|\beta|$ of about 5. The large value is chosen as to clearly delineate phenomena associated with significant excitation on nonlinear structure formation. This fixes our basic fiducial model. We will examine, in addition, the effect of smaller enhancements and comoving spatial cutoff scales (Sec. IV C 3), as well as the effect of relaxing the sudden jump assumption (Appendix B; Sec. IV C 3).

B. Halo mass function

1. Evaluation procedure

On nonlinear scales, modifications of the primordial power spectrum are primarily encoded in the mass function of self gravitating dark matter objects, the halos hosting galaxies. We evaluate this function using the Press-Schecter formalism, which estimates the number of dark matter halos per unit mass and comoving volume, given the linear matter power spectrum *via* a spherical collapse model [155,156]. This is given by

$$\frac{dn}{dm} = \frac{\rho_0}{M^2} f(\sigma) \left| \frac{d \ln \sigma}{d \ln M} \right| \tag{57}$$

where ρ_0 is the mean matter density at $z = 0$ and $f(\sigma)$ is given by

$$f(\sigma) = \sqrt{\frac{2}{\pi}} \nu \exp\left(\frac{-\nu^2}{2}\right), \tag{58}$$

where $\nu = \delta_c/\sigma$, with $\delta_c = 1.686$ the critical overdensity for spherical collapse and σ the RMS variance of mass fluctuations within a sphere of radius R and containing mass $M = \vartheta_f \rho_0 R^3$, where ϑ_f a constant that depends on the filter function W. For Gaussian filter it is $\vartheta_f = (2\pi)^{3/2}$. The filter function is characterized by its size R or mass M. In the case of Gaussian filter we use here, the relation between them is

$$M = 4.37 \times 10^{12}\, \Omega_m\, h^{-1} \left(\frac{R}{h^{-1}\,\mathrm{Mpc}}\right)^3 M_\odot. \tag{59}$$

As our primary aim is to illustrate generic consequences of enhanced small scale power spectrum, we generally kept to the aforementioned simplest form of the Press-Schecter formalism. However we have also verified the insensitivity of our results to that choice by comparing with an ellipsoidal collapse fitting function, which provides better fits to mass functions of halos identified in cosmological simulations [157–160],

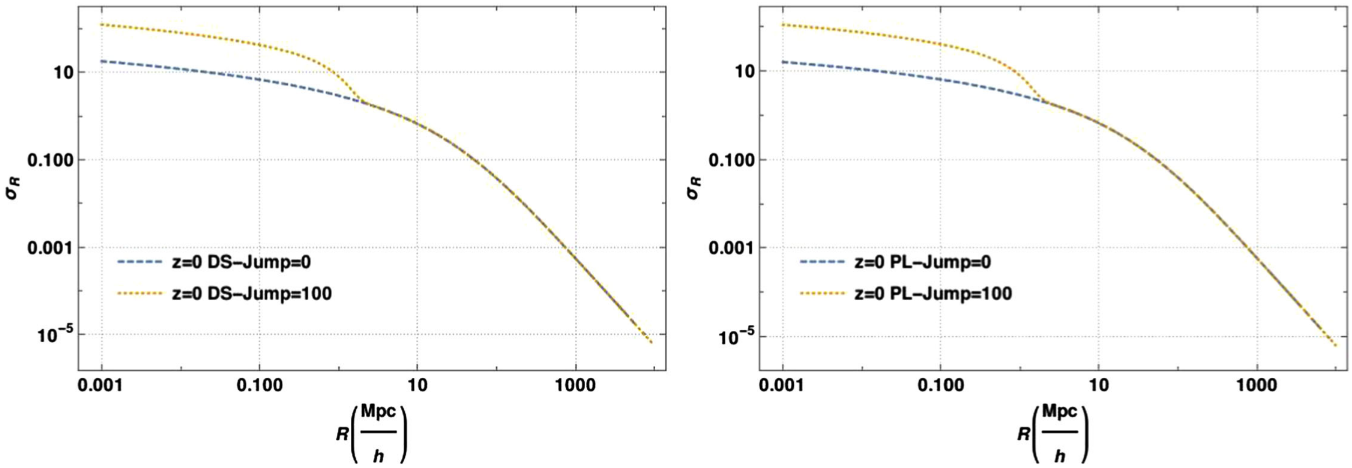

FIG. 2. RMS mass fluctuations corresponding to power spectra shown in Fig. 1. Note that the strong oscillations in the power spectrum, in the case of power law inflation, have little effect here, as they are smoothed over and integrated out as the dispersions are extracted from the power spectra. Despite the sharp jump in linear power spectra, the change in the RMS mass fluctuations beyond the cutoff scale is also gradual.

$$f(\sigma) = A\sqrt{\frac{2a_s}{\pi}}\left[1+\left(\frac{\nu^2}{a_s}\right)^{p_s}\right]\nu \exp\left(-\frac{a_s\nu^2}{2}\right), \quad (60)$$

where we set $p_s = 0.3$, $A = 0.3222$ and $a_s = 0.707$. Results using that form are shown in Appendix B, where we examine implication of an enhanced small scale spectrum when the assumption of sudden jump is relaxed, and also in Fig. 8.

The mass variance is calculated through the integral,

$$\sigma^2(R) = \frac{1}{2\pi^2}\int_0^\infty k^2 P(k) W^2(kR)dk, \quad (61)$$

where $P(k)$ is the linear power spectrum and $W(kR)$ is the Fourier transform of the Gaussian filter function

$$W(kR) = \exp\left(-\frac{(kR)^2}{2}\right). \quad (62)$$

Figure 2 shows the thus calculated dispersion for de Sitter and power low inflation models. As can be seen the strong oscillations in the power spectrum of the latter case are smoothed and integrated over, and the results are quite similar in the two cases. Also, despite the sharp jump in the corresponding power spectra, the change in the RMS mass fluctuations in the nonlinear regime beyond the cutoff scale is gradual.

2. *Mass function at redshift zero*

Figure 3 shows the resulting Press-Schecter halo multiplicity function, which estimates the fraction of mass in halos of mass M, corresponding to the unperturbed and perturbed (with jump) matter power spectra shown in Fig. 1. As may be expected given the mass dispersions shown in Fig. 2, the results are virtually identical in case of de Sitter and power law inflation, despite the strong oscillations in the spectrum in the latter case. Perhaps also expected is the enhancement at larger masses embodied in the bump encompassing a scale around a few $10^{12}\,M_{\odot}$, when the power spectrum is boosted. More counterintuitively, there is a dearth of small halos when the power spectrum is perturbed. This is due to those smaller halos more rapidly merging into larger ones, as we discuss further below (next subsection).

The bump at higher masses is not in itself directly observable; as it can be accounted for by changing the galaxy-halo occupation numbers. Indeed it roughly corresponds to the highest mass to light ratio inferred when

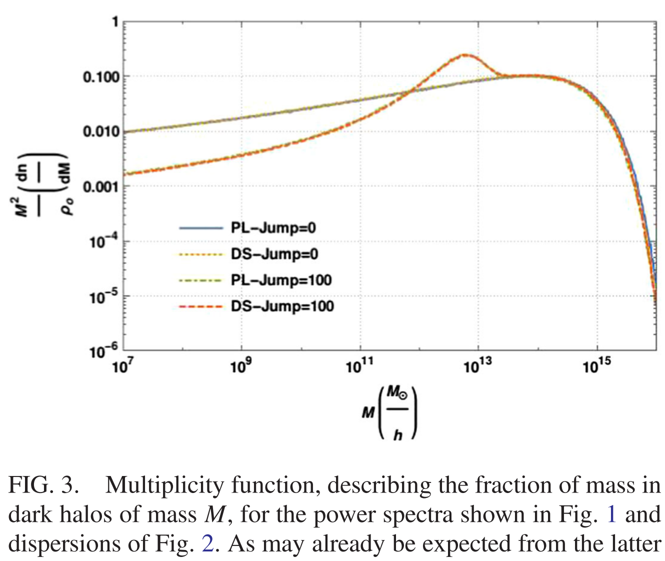

FIG. 3. Multiplicity function, describing the fraction of mass in dark halos of mass M, for the power spectra shown in Fig. 1 and dispersions of Fig. 2. As may already be expected from the latter figure, the results are similar in de Sitter and power law inflation (labeled DS and PL, respectively), due to the smoothing and integration over the oscillations as the mass function is derived.

fitting galaxies to halos in context of the standard model. Nevertheless, the compatibility of such enhancement with data can be tested through a combination of abundance matching and dynamical modeling. In fact, for galaxies with stellar mass above $5 \times 10^{10}\,M_\odot$, abundance matching with standard power spectrum seems to overpredict the observed stellar masses for a given dynamical mass [161]. As the galaxy number density (determined by the stellar mass function), is a decreasing function of mass, the discrepancy may in principle be accommodated in our current context as follow: by increasing the abundance of halos with larger dynamical masses (with the bump at higher masses), the galaxy population associated with such halos will then be one with correspondingly larger number density, and hence smaller masses. This is essentially the same effect that may help alleviate the apparent problems with early massive galaxy formation as we discuss in Sec. IV C 2. Thus the solution to such potential problems does not only appear consistent with the distribution of dynamical masses at low redshift but may even resolve certain problems there; at both high and low mass scales.

3. Enhancement at high z and large M, and suppression at the opposite ends

For the same parameters used above, Fig. 4 shows the multiplicity function at selected redshifts. Four trends are clear. First, the enhancement in the primordial power spectrum leads to enhancement in the number of halos at intermediate masses, culminating approximately at the mass scale corresponding to the length scale to where the jump in the power spectrum is placed. The second is that the effect is larger, and is apparent for a larger range in masses, at higher redshifts. Indeed, Fig. 5 shows a maximum enhancement of almost four orders of magnitude at $z = 8$, compared to only a factor of a few at $z = 0$. This is because high mass halos are rarer at higher redshifts and thus the relative increase due to the enhancement resulting from the discontinuity in the power spectrum is larger. It also results in the rate of change in the mass function with redshift, at fixed mass, being smaller in the case of enhanced spectrum than in the unperturbed case.

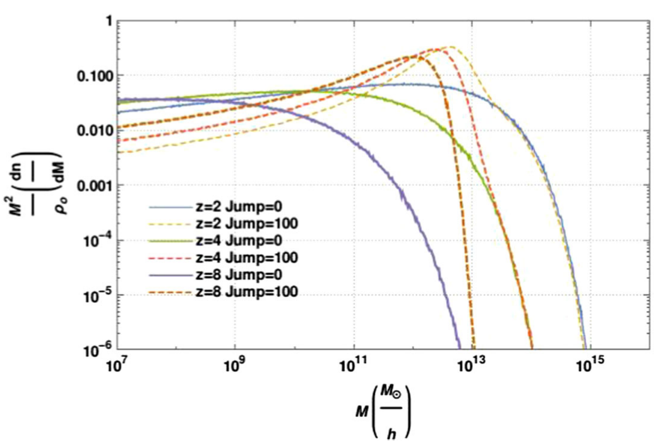

FIG. 4. Same as in Fig. 3 but at the indicated redshifts (using power law inflation model). Solid lines show the results with standard (nearly) scale invariant primordial spectrum, which are compared to those obtained when the power spectrum is boosted at smaller scales, as a result of an imposed step (corresponding to a ratio of hundred fold) in mode frequency and propagation speed (cf. Fig. 1).

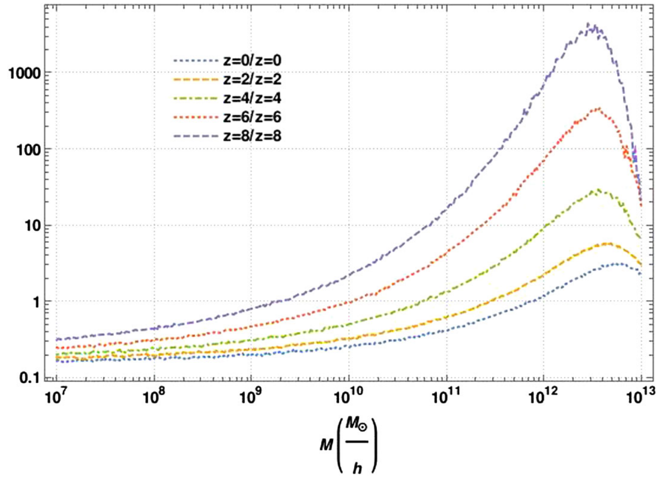

FIG. 5. Ratios of mass functions with modified power spectrum to those resulting from unmodified power spectrum, at the same redshift. The results correspond to the ratios of dashed to solid lines in Fig. 4 and Fig. 3.

Finally, there is the somewhat counterintuitive effect, mentioned at the conclusion of the previous subsection, of significant *suppression* of the multiplicity function contribution of halos at smaller masses, below a few $10^{11}\,M_\odot$. This interesting result may be understood by recalling that enhancement in the power spectrum at comoving scales associated with masses of $\gtrsim 10^{12}\,M_\odot$, implies that all smaller scales are also enhanced. And enhancement at smaller (length and mass) scales in turn implies that smaller halos form at higher redshift and that by the redshifts considered here they have already been typically subsumed in larger ones; that is, the typical mass scale, for a given fluctuation level at a given redshift, shifts up. This leads to a relative decrease in the number of halos with masses $\lesssim 10^9\,M_\odot$. As opposed to the case of the enhancement of the multiplicity of relatively high mass halos, the deenhancement is here relatively larger at smaller z, as the lower mass halos are now those that are rarer at such redshifts.

C. Interpretation and possible consequences

As, in the current model of structure formation, galaxies form in seeds provided by the potential wells of dark matter halos, the significant modifications to the halo mass function are expected to leave imprints on the associated galaxy stellar mass function. Pertinent questions here thus include whether those modifications have consequences for problems arising at small scales within the current standard scenario of structure formation, as outlined in the introduction; or, in contrast, whether such modifications can constrained also on nonlinear scales.

1. Small scale problems at low redshift and the dearth of dwarf galaxies

One straightforward consequence of the suppression of halo multiplicity at small scales pertains to the longstanding issue of the dearth of dwarf galaxies in the standard scenario: a galaxy like the Milky Way is expected, in the context of the ΛCDM with a standard primordial power spectrum, to have hundreds of satellites that are not observed, and some of the predicted hosting halos are too "large" to have "failed" to form galaxies. These are aspects of the so-called small scale problems of the standard scenario has given rise to various explanations, e.g., in terms of baryonic physics, warm dark matter, fuzzy dark matter, as well as direct suppression of the small scale power spectrum.

Our somewhat counterintuitive result, on the other hand, is that an enhancement of power on small scales can also lead to a suppression in the number of small halos (as these 'over-merge' into larger entities). This suppression at $z = 0$ at the scales where issues such as the dearth of small galaxies ($M_h \lesssim 10^9\,M_\odot$) and too big to fail ($10^9 \lesssim M_h \lesssim 10^{11}\,M_\odot$) problems appear, can therefore be of relevance to apparent small scale crises arising in the context the standard model of structure formation. The order of magnitude suppression at smaller masses is directly relevant to resolving the apparent discrepancy between the number of observed small satellite galaxies and large number of small halos found in cold dark matter simulations. The suppression on the larger mass scales, on the other hand, may help alleviate the too big to fail issue; when this is posed as an abundance matching problem, whereby the abundance of simulated halos is too large at the masses inferred from the dynamics of observed galaxies in the range $10^9 \lesssim M_h/M \odot \lesssim 10^{11}$ (e.g., ref. [162]).

2. The excess of early massive galaxies

The enhancement at higher mass scales may, on the other hand, have consequences for the more recently raised issues associated with early galaxy formation. These are extensions of longstanding phenomena related to what is referred to as 'downsizing' (e.g., [163]), required to account for preponderance of early massive galaxies; a phenomenon that does not appear entirely natural in a hierarchical structure formation scenario, where the smaller halos embodying the potential wells hosting the galaxies form first.

The problem of early galaxy formation has been termed 'impossibly early' in the context of the standard ΛCDM scenario of structure formation [93]. In that work, the authors attempt to infer the halo mass function at high-z, primarily from stellar mass functions derived using photometric spectral energy distribution templates and ultraviolet luminosity functions. The halo mass is then inferred by assuming a stellar to halo mass of $M_*/M_h = 1/70$. If this local value of M_*/M_h is used, then Fig. 1 of the aforementioned work suggests that the number density of massive galaxies can greatly exceed that of the halos they should inhabit for $z \gtrsim 4$ in the standard ΛCDM structure formation scenario. The discrepancy becomes more severe as one moves up in redshift and mass, reaching four orders of magnitude or more.

The above would seem to rule out the standard scenario of structure formation in the context of ΛCDM cosmology. However, a couple of caveats have been pointed out. First, regarding the assumption that M_*/M_h does not vary with redshift. For, as can also be seen from Fig. 1 of [93], instead of moving the points inferred from the observed stellar number densities down orders of magnitude to fit the corresponding halo number densities, one can move the points horizontally to the left by an order of magnitude. This fitting procedure in effect invokes a z (and M_*) dependent M_*/M_h, to replace the fiducial local value of $M_*/M_h = 1/70$ assumed by the authors. The procedure, requiring $M_*/M_h \sim 1/7$, is still in principle consistent with a universal baryon fraction of $1/6.3$, associated with the standard cosmological scenario, but only just [98,164].

Another caveat that has been pointed out concerns the extraction of M_* and associated number densities from the ultraviolet luminosity function at high z, which some of the data points of [93] relies on [168]. However, a multiwavelength analysis of a sample of massive galaxies at $z > 3$ also leads to a cumulative mass function that can be consistent (within estimated errors) with M_*/M_h approaching the universal baryon fraction at $z \sim 5.5$ and $M_* > 10^{11}\,M_\odot$ [169]. That work also shows (Fig. 14) that the number densities of massive galaxies are very difficult to reproduce in hydrodynamic numerical simulations—with significant underestimate for $z > 3$—which may be expected, as their reproduction would seem to require that all available baryons reside inside galaxies, and their near total conversion to stars over a short time (~Gyr). This would have as consequence the presence of a significantly "quenched," quiescent population of massive galaxies already at high redshift. The presence of such a population, which is indeed observed, poses significant challenges. Synthesizing the stellar populations of one such object, observed at $z = 3.717$, for example, seems to again require prior evolution involving a M_*/M_h reaching the universal baryon fraction [170] (see also [171]). There now appears to be a substantial population of such galaxies, observed at increasing redshift [172–178], and not easily reproduced by either hydrodynamic simulations [175,176] or semianalytical models [173,177].

Although questions as to the ultimate severity of these problems will only be settled with the next generation surveys (e.g., with the James Webb Space Telescope

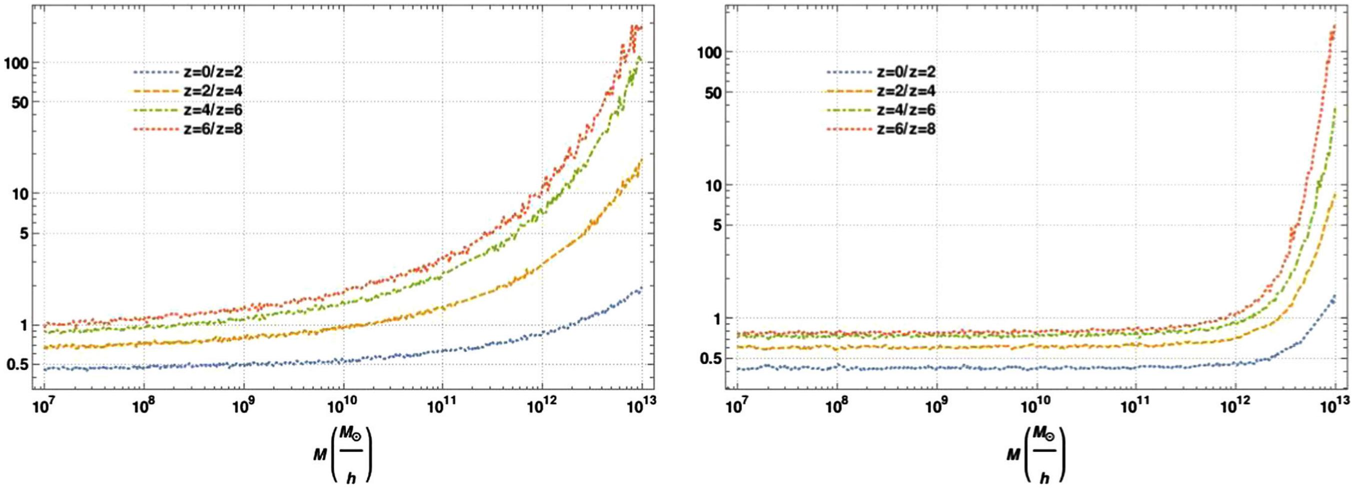

FIG. 6. Ratios of the mass functions at different redshifts, for the standard case (left) and that with modified power spectrum (right). Note the slower evolution (reflected in the smaller ratios) at higher z for most of the mass range in the modified case.

(JWST)), the situation warrants pointing out that they can in principle be alleviated by invoking small scale enhancement of the primordial power spectrum examined here.

Figure 5 shows that significant enhancements can be achieved at mass scales $10^{12}\,M_\odot \lesssim M \lesssim 10^{13}\,M_\odot$, with a peak at a scale corresponding to highest dark matter to stellar mass ratio in standard modeling, at which the enhancements reach even the "impossibly" large levels claimed in [93]. Perhaps no less important is the slower evolution of the mass function for $z \gtrsim 4$, observed in Fig. 6, which is more consistent with the redshift evolution of the inferred stellar mass densities in [93] than the much faster evolution in the standard case (the slow evolution of the stellar mass function for $4 \lesssim z \lesssim 7$ was also observed for example by Song *et al.* [179]). This would seem to waive the apparent requirement of a M_*/M_h that is high dependent on redshift in order to fit the data.

With better statistics, and firmer grip on observational systematics, it should be possible to distinguish between scenarios involving enhancement in the primordial power spectrum, such as the one presented here and reconciliation with data through improvement of the baryonic model; by invoking further "downsizing" physics input, in terms of feedback, quenching and other "subgrid" physics (assuming the data remain consistent with the strict upper bounds placed in the context of ΛCDM [98]). As the baryonic models become better constrained, there may be particular consequences that could also constrain (or confirm) the sort of scenario discussed here. We now discuss some of these.

3. Other observables, constraints and variation on basic model

In the context of the enhanced spectrum scenario presented here, the clustering of halos, on mass scales and redshifts where numbers are predicted to be significantly enhanced, may be measurably different from the standard case. This is because the biasing with respect to the matter distribution would be expected to be different (since they would correspondent to less rare density peaks). Combined clustering and abundance matching analysis in the context of a 'halo model' (e.g., [180]), particularly at higher redshifts [181], could thus in principle test, and place constraints on, scenarios involving primordial power spectrum enhancements. The galaxy-matter correlation function, entering into calculations of galaxy-galaxy weak lensing signals, should also be different in the present scenario from the standard case. The difference should again be especially significant at higher redshifts, where the abundance of high mass halos is strongly increased, making for a relatively clumpy matter distribution. Tests are also possible at low redshift. particularly as regards to the peak in abundance of Milky Way sized halos, which seems consistent with observations (Sec. IV B 2). Some observations on the other hand suggest that the standard model itself may overpredict the halo mass function at scales $10^{13}\,M_\odot \lesssim M \lesssim 10^{14}\,M_\odot$ [182]; further enhancements of the mass function at such scales may be thus constrained.

Another observable that can potentially place immediate constraints on the scenario discussed here is the Lyman-α forest. Here, detailed comparison with data involves complex simulations that depend on assumptions regarding the state of the intergalactic medium, which become less robust at nonlinear scales [134,183]. In the nonlinear regime the modifications in the power spectrum are primarily imprinted in the RMS dispersion and the halo mass function, where the complex pattern of enhancement and suppression at different scales and redshifts would contribute to the mass fluctuations probed by one dimensional Lyman-α spectra. The large enhancement at higher masses and redshifts may also affect the thermal history of the intergalactic medium. It may therefore be worth investigating if and how such changes affect the standard

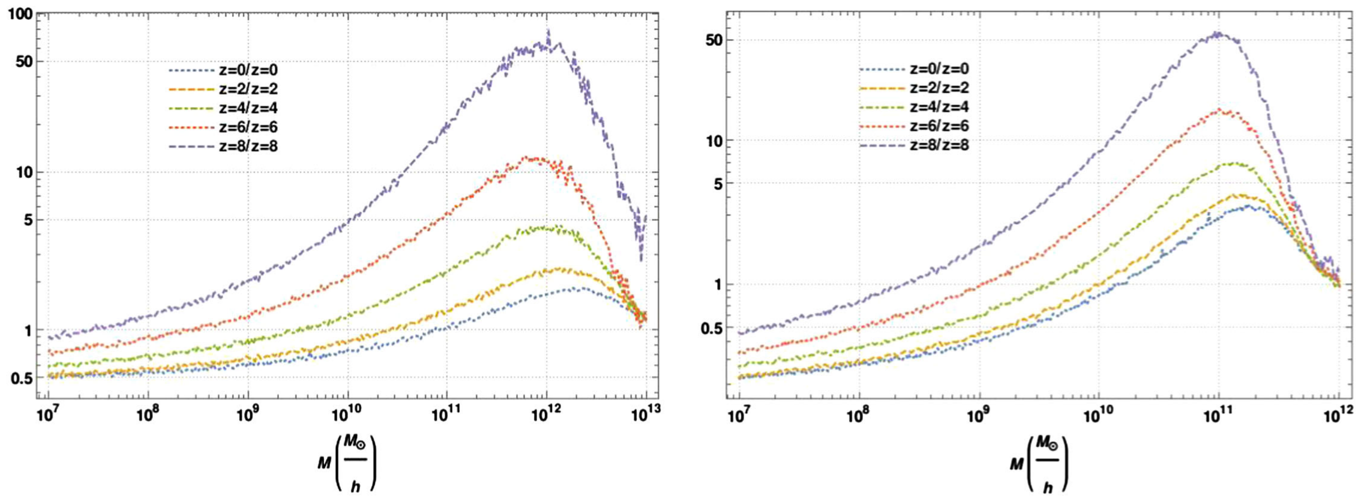

FIG. 7. Same as in Fig. 5 but with step frequency ratios of 10 instead of 100, leading about to about an order of magnitude less modification in the power spectrum, with Bogulibov coefficient $|\beta| = 1.42$ instead of 4.95 (left); and for power spectrum comoving modification scale $k = 3\ h/\mathrm{Mpc}$, instead of $k = 1\ h/\mathrm{Mpc}$ (right).

constraints regarding the power spectrum. Pending such investigation, as the modifications to the linear power spectrum and mass dispersion considered above are large and fall within the region relevant to Lyman-α observations, it apt to probe what is to be expected if more modest modifications are made.

In Fig. 7 (left panel) we show the relative change in the mass function for (about an order of magnitude) smaller perturbation in the power spectrum, as well as on scales deeper in the nonlinear regime. As can be seen, in the former case, significant enhancement in the mass function can still occur at the right scale at higher redshift (where halos are exceedingly rare in the standard scenario), so as to alleviate the apparent early galaxy formation problem. The reduction in number of small halos, relevant to the dearth of small galaxies and too big to fail problems at low z, is smaller however.

When the modification in the power spectrum is placed on a smaller spatial scales, deeper in the nonlinear regime (Fig. 7, right panel), the decrease in number of small halos at $z = 0$ is again significant and relevant to the dearth of dwarf galaxies, but does not cover all the mass range relevant to the too big to fail problem. As may be expected, the enhancement at high redshifts happens at a smaller mass scale (they are also smaller because halos in the standard scenario are already more abundant at such scales). Enhancement at such scales is not directly applicable to the problem of high M_*/M_h at $M_* \gtrsim 10^{10.5}\ M_\odot$ and $z \gtrsim 4$, as discussed in a previous subsection. It may nevertheless be relevant at higher redshift, as the progenitors of massive quiescent galaxies were assembled (especially if the star formation rate density does not steeply decline beyond $z = 8$, as suggested by some authors; c.f. Ref. [171], particularly the discussion in Sec. 7.3 and references therein). As one further increases the comoving wave number associated with the high energy cutoff k_c, this general trend persists. The suppression on small scales at $z = 0$ are found to correspond to masses of small halos that are overabundant in CDM up to comoving cutoff $k \sim 9\ \mathrm{Mpc}^{-1}$, which essentially avoids Lyman-α bounds. At larger modification wave numbers, however, one finds enhancement rather than suppression at mass scales $\lesssim 10^9\ M_\odot$, relevant to the dearth of dwarf galaxies issue. Although enhancing the power spectrum at such smaller spatial scales would not appear to alleviate any of the issues regarding galaxy formation discussed here, it could still have consequences for early black hole formation and the epoch of ionization (see also [92]).

On the other hand, relaxing the assumption of a sudden jump transition leads to suppression and enhancement of halo abundances on larger mass scales. This allows for retaining the advantages of a sudden cutoff at comoving wave number 1 h/Mpc, while keeping the mass function at $10^{13}\ M_\odot$ at $z = 0$ unchanged, and modifying the power spectrum much more modestly at $1h/\mathrm{Mpc}$ (cf. Appendix B, Fig. 12). Figure 8 shows the relative change in the mass function for such a gradual transition in the power spectrum around a characteristic wave number of 3 h/Mpc comoving. As can be seen, significant suppression at $z = 0$ is again recovered at mass scales of order $10^{11}\ M_\odot$, relevant to both the dearth of dwarf galaxies and the too big to fail problems. Enhancement at higher z also occurs at scales significantly larger than the corresponding sudden jump case (shown on right panel of Fig. 7), which renders the enhancement more directly relevant to early massive galaxy formation issues. The mass function is unmodified at scales $10^{13}\ M_\odot$ at $z = 0$.

Thus, potential resolution of all or some of the galactic and subgalactic scale issues through modification of the power spectrum, rather than (or in addition to) baryonic

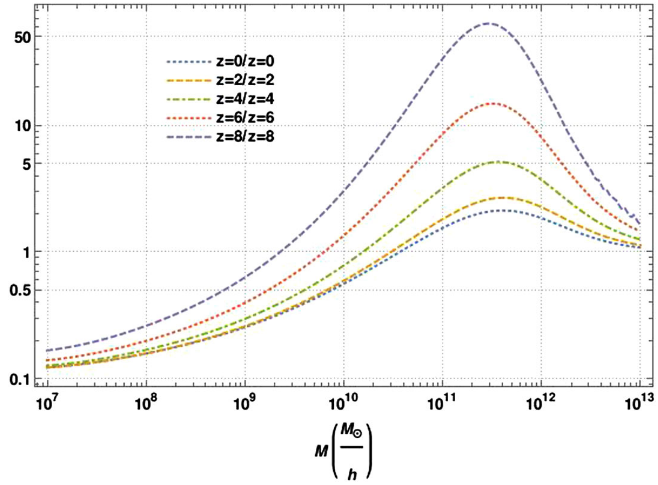

FIG. 8. Same as in Figs. 5 and 7 but with gradually modified power spectrum, rather than sudden jump. The spectrum is modified using Eqs. (B1) and (B2) with $S = 200$, $b = 2$ and $k_c = 3\ h\,\mathrm{Mpc}^{-1}$ comoving. The results correspond to ratios of dashed and solid lines in Fig. 13 (taking the dashed line at $b = 2$ for the left hand panel). The associated modifications to the power spectra are those shown in Fig. 12.

physics input, may in principle be tested and constrained through distinctive predictions. This is true in general and is not confined to our particular simple model of a sudden jump; such tests will thus become more relevant if the small scale issues connected to the standard structure formation scenario are confirmed to persist with incoming observations. In the context of the present scenario, such observations can potentially probe imprints (or lack thereof) of high energy cutoff physics on the relevant astrophysical scales, and place constraints on the duration of inflation, as the ratio of the Hubble scale of inflation to the high energy cutoff scale and the number of inflationary e-folds fix the scale at which the matter power spectrum and halo mass function is modified (cf. Sec. IV A 2). For the minimal number of e-folds required to solve the horizon problem for example, $H/k_c \approx 10^{-4}$ is required to address the galactic scale issues discussed. Significant power spectrum modification also require $k_c \lesssim H/k_c$ (Sec. III D). These tests can be stringent; as, given this scale, and the level of excitation determined by the Boguliubov β_k, the predictions of the simplest scenario of sudden jump through a high energy transition scale, are unique in terms of the expected effect on the power spectrum.

V. CONCLUSION

Slow roll inflation predicts a nearly scale invariant spectrum of primordial fluctuations, which is borne out by precise observations of the cosmic microwave background and large scale structure in the universe. Nevertheless, that prediction is not unique, a variety of effects invoking discontinuous or phased evolution between slow rolls, for example, can lead to anomalous "features" in the spectrum. Excited states arising from modes crossing a high energy cutoff scale can also lead to significant modifications to the scale free spectrum. Although these are essentially ruled out at the scales where the aforementioned observations are effective, the primordial spectrum is relatively unconstrained on smaller, currently nonlinear scales, where the matter distribution has collapsed into bound self-gravitating objects, washing out the primordial signature by largely encoding it in the halo mass function.

On the contrary, at such scales—which span many more octaves of observable structure than the three that are probed in the linear regime—a variety of issues arise in the context of the standard model of structure formation; such as the "small scale problems" at low redshift and the apparent problems involving early galaxy and supermassive black hole formation at higher z, which can be seen as extension of longstanding phenomena requiring "downsizing" in galaxy formation. As these issues arise precisely at the scales where complex baryonic physics comes to play a central role in the standard scenario of structure formation, it was natural that extensive investigation of solutions in these terms have been pursued. However, as these are also the scales where the primordial spectrum of fluctuations is relatively weakly constrained, this aspect, with its effects and consequences, may also warrant further investigation.

Here we considered the effects of excited states arising from the transiting of fluctuation modes through a high energy cutoff scale. As the power spectrum of primordial fluctuations is effectively an adiabatic invariant of inflaton dynamics (Sec. II B 2), adiabatic evolution necessarily leaves the nearly scale free spectrum intact. We next considered a simple model of the opposite extreme; of a sudden jump across the transition. The initial conditions for the fluctuations before the jump are well defined, taking the Bunch-Davies form, but with propagation speed $c_s \neq 1$. An intuitive, simple analogue model approximated by such a transition corresponds to the case of a gas or lattice where sound waves do not propagate at all below the interparticle distance, then propagate at an anomalous speed in an effective macroscopic approximation, before finally propagating with the standard sound speed and dispersion relation as the wavelength become progressively larger than the interparticle distance.

In this context, the primordial spectrum is invariably enhanced rather than suppressed (whether the initial $c_s > 1$ or is < 1), for all scales undergoing the transition through the high energy cutoff (Sec. III). This is accompanied by strong, tightly spaced oscillations in the power spectrum of generic (as opposed to pure de Sitter) models of inflation, where modes exit the horizon at different phases. Numerical calculations suggest that sufficiently nonadiabatic evolution, leading to significant modification of the power spectrum implies an effectively sudden transition for all $c_s > 1$ and for $0.01 \lesssim c_s < 1$ (Appendix A 2). The simple model of sudden jump, and its predictions, are in this range thus generic. We also considered the possibility of a more gradual transition when the aforementioned conditions are not satisfied (Appendix B).

Given the excitation level induced in the inflaton field, and the current comoving scale corresponding to the jump across the high energy cutoff scale during inflation, the predictions of the simple sudden jump models are essentially unique (in terms of its effect on the matter power spectrum, mass variance and the dark matter halo mass function). The level of excitation can be quantified through a Bogoliubov coefficient $\beta_k \neq 0$ for scales that undergo the jump, and is easily evaluated in terms of the in and out frequency ratio (or equivalently c_s ratio; Sec. III C). If assumed to be within a few orders of magnitude of the Planck scale, the jump scale corresponds to currently nonlinear scales if inflation proceeds for approximately the number of e-folds necessary to solve the horizon problem. In general, the comoving jump scale corresponds to currently nonlinear scales for minimal inflation if $H/k_c \sim 10^{-4}$, with smaller ratios allowing for larger e-folds (Sec. IV A 2). In this context, the nonlinear scales can be modified, while leaving the standard spectrum intact on linear ones.

Backreaction bounds on $|\beta|$ must be imposed, as "over-excitation" of the inflaton would result in radiation domination rather than inflation; these may however still allow for major enhancements of the power spectrum $\sim 1 + 2\beta^2$ (and oscillations in the generic inflation case). As we discuss in Sec. III D, this would be generally the case if $k_c \lesssim (H/k_c) M_{\rm Pl}$. Such enhancements can have observable consequences, confirming or constraining the effect of excitations on structure formation on nonlinear scales. In order to impose modifications on such scales in particular, and still keep the excitations from overwhelming the inflaton vacuum state, one thus requires $H/k_c \lesssim 10^{-4}$ and $k_c \lesssim (H/k_c) M_{\rm Pl}$. This implies $k_c \lesssim 10^{-4} M_{\rm Pl}$ and $H \lesssim 10^{-8} M_{\rm Pl}$. A rough estimate of the derivative of the backreaction suggests possible modification of the power spectrum on the largest scales, and may place tight constraints on the comoving scale at which enhancement of the small scale power spectrum can occur (to about a comoving Mpc; Sec. III D). That modification on the largest scales can accompany the changes on small, nonlinear ones, is an interesting possibility that may be worth studying in detail.

To probe for possible characteristic signatures of the modifications on nonlinear scales, we evaluate (in Sec. IV) the dark halo multiplicity function, quantifying the fraction of mass in halos of mass M. In our fiducial example, the peak, resulting from power spectrum enhancement, is chosen to correspond to a few times $10^{12} M_\odot$. This is the mass scale where the highest mass to light ratio is inferred when associating galaxies with halos in the context of halo models derived within the standard scenario. It is also the scale where issues related to the apparent preponderance of early massive galaxies, particularly quiescent ones, appear (Sec. IV C 2). For relatively small enhancements at small redshifts z, the enhancement at larger $z \sim 8$ is dramatic, as such massive halos are very rare at these redshifts in the standard scenario. The change in the number densities of massive galaxy-hosting halos with redshift is also much smaller than in the standard case. Combined, these effects may alleviate the apparent impossibly early galaxy formation problem, even in the most extreme form claimed.

Perhaps more surprisingly, an enhancement of the spectrum at these intermediate nonlinear scales leads to *suppression* of small halos at low z, thus potentially alleviating longstanding issues related to the dearth of small galaxies, including those too big to fail, in the standard structure formation scenario. This is due to the enhanced spectrum leading to overmerging of small mass objects at high z, so as to lead to a suppression of such objects at low z.

The halo mass function, in itself, cannot place strong constraints on enhancements of the primordial power spectrum on currently nonlinear scales, as one can vary the galaxy halo occupation number to match the data (in the standard scenario, the early galaxy formation issues at high M and z arises because this seems to sometimes require very large stellar mass fraction, which the enhanced halo mass function here may resolve; Sec. IV C 2). However, combined abundance matching and dynamical analysis at low-z can. Halo abundance enhancement at the scales considered here appears consistent with such analyses; it may in fact alleviate the apparent overprediction of stellar masses for given dynamical mass for galaxies with stellar mass $\gtrsim 5 \times 10^{10} M_\odot$ ([161]; Sec. IV B 2).

Major modifications in the spectrum of primordial fluctuation are eventually encoded in more minor modifications to the nonlinear matter power spectrum, as these enter primarily through the modified halo mass function rather than the statistics of the spatial distribution. Nevertheless, the scenario of an enhanced primordial power spectrum at scales corresponding to currently nonlinear ones, can also be tested through its signature on halo biasing. The fact that more massive halos would be less rare may be expected to particularly impact such observables as galaxy-mass correlations and leasing signals (especially at higher redshift where the effects of enhancement at higher mass scales are more prominent). To address both aforementioned issues—of massive high-z massive galaxies and small local ones—simultaneously in the most severe forms claimed, through sudden transitions, also entails significant modifications at scales probed by Lyman-α observations (the required modifications are more modest, or at scales that may be less constrained, if only partial resolution of both issues is sought or if the assumption of sudden transition is relaxed; Sec. IV C 3).

Thus, observations, coupled with modeling and simulations with modified spectrum, may place constraints on scenarios invoking enhanced power on currently nonlinear scales, distinguishing them from baryonic solutions to the same problems. In the context of the analytical sudden step model primarily considered here, this includes constraints on H/k_c, k_c, $|\beta|$, and the number of inflationary e-folds, as discussed above. Given the field excitation levels (i.e.,$|\beta|$) and the comoving scale of the high energy transition, the consequences for the matter spectrum and halo mass function are essentially unique. Variants that could also be tested include those involving phased or discontinuous stages of inflation with relatively localized peaks in the primordial spectrum. This will become perhaps more pressing if next generation surveys (e.g., employing the JWST) confirm problems related to early galaxy formation. On smaller (subgalactic) scales still, primordial power spectrum enhancement may be relevant to early supermassive black hole formation, and the formation of the first dark matter objects, and may be tested through such effects as CMB spectral distortions.

ACKNOWLEDGMENTS

We wish to thank Adel Awad, Alexey Golovnev, Sergio E. Joras, Giovanni Marozzi, Jerome Martin and Joe Silk for useful discussions, and the referee for constructive comments and suggestions. This project was supported financially by the Science and Technology Development Fund (STDF), Egypt. Grant No. 25859.

APPENDIX A: COMPARISON OF SIMPLIFIED MODEL WITH NUMERICAL SOLUTION, AND THE EFFECT OF RELAXING THE SUDDEN JUMP CONDITION

In this Appendix we test the approximation of the simplified model of Sec. II C, introduced to evaluate the effect of nonadiabatic transition at a high energy cutoff scale k_c on the primordial power spectrum. There are two approximations that were invoked; the sudden step and the neglect the term proportional to $-\frac{1}{\eta^2}$ in the Mukhanov-Sasaki equations (2) and (7). We start by examining the latter, then we discuss the former.

1. Model versus numerical solution of Mukhanov-Sasaki equation with step

We evolve the dynamics of fluctuation modes numerically, using the MS equation (2) for de Sitter and power law inflationary backgrounds, while replacing the term k^2 with

$$k^2 \to k_{\rm eff}^2(k,\eta) \equiv a^2(\eta)\omega_{\rm phys}^2\left[\frac{k}{a(\eta)}\right]. \tag{A1}$$

where

$$\omega_{\rm phys}^2\left[\frac{k}{a(\eta)}\right] = \left(\frac{k}{a(\eta)} + \frac{\delta k}{a(\eta)}\mathcal{H}\left[\frac{k}{k_c a(\eta)} - 1\right]\right)^2, \tag{A2}$$

and $\mathcal{H}$ is the Heaviside step function. The parameter δ quantifies the size of the step, such that the sound speed past the step given by $c_s = 1 + \delta$; it can be positive or negative, corresponding to an upward and downward jump in sound speed respectively. As discussed in Sec. III, in the context of the sudden step scenario they are equivalent in terms of the effect on the power spectrum.

The primordial spectra are evaluated, for both the analytical model and numerical calculations, as described in Sec. II C 1. The results are shown and compared in Figs. 9 and 10, for the case of de Sitter and power law inflation respectively. As noted in Sec. II C 1, in the de Sitter case all modes exit the horizon at exactly the same phase. And any initial shift in phase, due to change in effective frequency related to the second term in bracket of the MS equation, leads to corresponding constant difference in the final power spectrum. This leads to a difference between the numerical and analytical results, where the aforementioned term is neglected. Nevertheless, the relative error in the ratio of the perturbed to unperturbed power spectrum is still of order 25% when the step frequency ratio is 10. It is an order of magnitude lower still when the change of frequency is hundred folds.

In the case of power law inflation the Hubble scale H is not exactly constant. The modes exit the horizon at different phases, and this leads to the oscillations, which accompany the enhancement in Fig. 10. The corresponding error is then primarily in phases, with the maxima and minima of the oscillations practically equal in the simplified model and the numerical calculations. The change in phase is generally unimportant for calculating quantities with observable consequences; such as the mass fluctuations at a given spatial or mass scale, and halo mass multiplicity function. For these depend on integrals of the matter power spectrum (as discussed in Sec. IV). The simple analytical model—with its simple interpretation in terms of well defined in and out states; Sec. III—thus turns out to be a good approximation.

2. Nonadiabaticity versus sudden jump condition

Oscillations can, in general, be considered adiabatic if $\omega(\eta)$ changes only slightly over a characteristic time $\Delta\eta$ of order of one oscillation period. If the frequency ω changes from a value ω_1 to another value ω_2, on a characteristic timescale $\Delta\eta$, the change may thus be adiabatic if

$$|\omega_1 - \omega_2| < \omega^2\Delta\eta. \tag{A3}$$

In the context of our sudden step model, ω_1 and ω_2 will correspond to $\omega_{\rm in}$ and $\omega_{\rm out}$, respectively. We take the typical

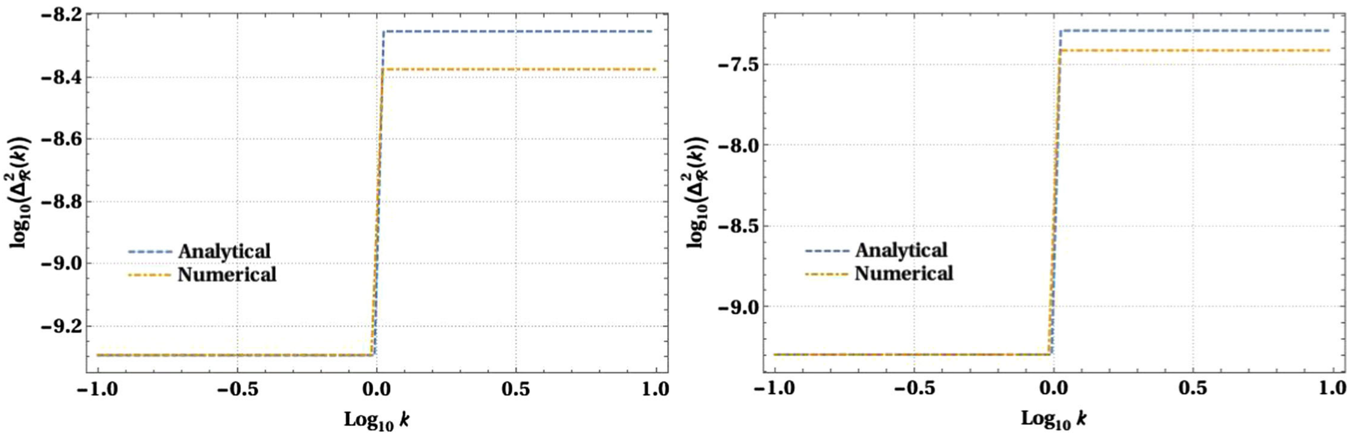

FIG. 9. Comparison of primordial power spectrum obtained from simplified analytical model with full numerical calculation, for sudden jumps corresponding to ratios of the in and out sound speed, or equivalently frequencies, of 10 (left) and 100 (right). This is done for a de Sitter background with $H/k_c = 10^{-4}$ and k is shown in units of k_c.

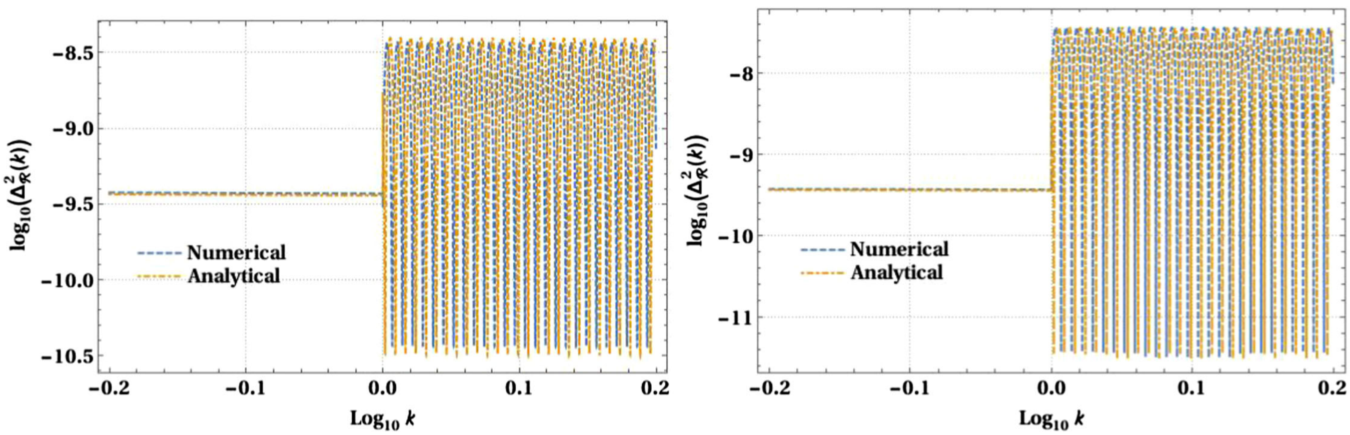

FIG. 10. Same as in Fig. 9, but for power law inflation model (discussed in Sec. II C 1).

ω on the right hand side to correspond to the minimal frequency; supposing that the dynamics may be affected nonadiabatically if the change in frequency is larger than this. To examine to what extent that model may describe a more general situation, where change may be more gradual, we use the logistic function to parametrize the transition:

$$k_{\mathrm{eff}} \equiv \omega(\eta) = k + \frac{\delta k}{1 + \exp\left[-\gamma\left(\frac{\eta}{\eta_c} - 1\right)\right]}. \tag{A4}$$

Here the parameter γ describes the stiffness of the transition, this being steep and steplike for $\gamma \gg 1$, and δ (which may be positive or negative) the scale of the step in the transition. Thus, in the high energy regime limit, the sound speed $c_s = 1 + \delta$, while $c_s = 1$ when the transition to the standard low-energy physics regime is complete. In these terms, the characteristic time over which ω changes between its initial and final value is $\frac{\eta_c}{\gamma}$. The adiabaticity violation condition can then be written as

$$\gamma > \frac{\mathrm{Min}(c_s^2)}{|c_s - 1|}\left(\frac{k_c}{H}\right), \tag{A5}$$

where we have used $\eta_c = \frac{-k_c}{Hk}$, and $c_s \neq 1$ corresponds to the high energy limit sound speed. Since, as we have seen in Sec. II B 2, the power spectrum is essentially an adiabatic invariant of the dynamics of inflationary perturbations, it is necessary to satisfy this condition in order to modify the standard power spectrum.

Two cases are of particular interest in the context of the present study: $c_s \gg 1$, so that $\mathrm{Min}(c_s) = 1$ and $c_c \ll 1$, when $\mathrm{Min}(c_s) = c_s$. For sound speeds considered here, the adiaticity condition (A5) is violated at smallest possible when the sound speed is minimal, that is $c_s = 0.01$. Still, even in this case, γ is of order one or larger if $k_c/H \geq 10^4$, as required to keep the significant changes in power spectrum in the nonlinear regime of structure formation (Sec. IV A 2). For our parameters, the transition is thus necessarily stiff.

We now show that the transition is stiff even form much smaller $k_c/H = 300$, which is a minimal value, in the sense that with smaller values the effect of the second term in brackets of Eq. (7) becomes important at $\eta \sim \eta_c$ for $c_s = 0.01$ (see discussion relating to inequality (11). For it turns out that the adiabaticity condition needs to be quite strongly violated for sufficient change in the dynamics significantly affects the power spectrum (that significant changes occurs well beyond the adiabaticity breaking condition is common in dynamical systems [184]). This can be seen from Fig. 11, where we show that large changes only occur when γ is orders of magnitudes above the value estimated from (A5). This is the case for both the $c_s = 100$ and $c_s = 0.01$, with the former being stiffer still as expected from (A5) [185]. The transition is stiffer still for smaller values of $c_s > 1$ and larger $c_s < 1$. Thus, for sound speed ratios considered here, our simple model of a sudden jump, and accompanying signature of a sudden break in the power spectrum, appears much more generic than may seem *a priori*.

APPENDIX B: MODIFIED MASS FUNCTION FROM NONSUDDEN SPECTRUM ENHANCEMENTS

As we have seen in the previous Appendix, the sudden jump transition in the power spectrum at smaller scales is a good approximation for the parameters primarily considered in this paper, namely for initial $c_s \geq 0.01$; for such values, the sufficient violation of the adiabaticity condition, required for significant modification of the power spectrum, practically implies a sudden transition. Nevertheless, as is already apparent in Fig. 11, for $c_s = 0, 01$ (right hand panel), the sudden jump approximation becomes less accurate as a predictor of significant change at smaller c_s. If one envisages a transition starting at significantly smaller sound speed still, it may then take place more gradually while still imparting a palpable effect on the power spectrum.

We consider potentially observable consequences of this effect by examining a series of progressively steeper transitions. We do this by modifying the primordial power spectrum in a parametric manner, such that

$$\Delta^2_{\mathcal{R}}(k) = \Delta^2_{\mathcal{R}\mathrm{St}}(k)\left[(S-1)G\left(\frac{k}{k_c}\right)+1\right], \tag{B1}$$

where $\Delta^2_{\mathcal{R}\mathrm{St}}(k)$ is the standard, nearly scale invariant, power spectrum of scalar perturbations. The transition function G tends to unity as $k \gg k_c$ and vanishes as $k/k_c \to 0$, and $S \geq 1$ is the enhancement factor (it determines the ratio of the asymptotic values of the power spectrum at small and large scales). We have tried several forms for G, and the resulting trends were verified to be generic. Here we show results for the following form (also used in [186] for the purpose of suppression of the spectrum rather than enhancement):

$$G(x) = \frac{1}{2}[\tanh(b\log x)+1] = \frac{1}{2}\left[\frac{x^{2b}-1}{x^{2b}+1}+1\right], \tag{B2}$$

where $b > 0$ determines the steepness of the transition around $x = k/k_c$. In the following we will take k_c to correspond to a comoving scale of $3h\,\mathrm{Mpc}^{-1}$. As Eqs. (39)) and (44) show S to be about 50 for a sudden step scenario with initial $c_s = 0.01$, the requirement that the initial $c_s < 0.01$ implies $S > 50$ (otherwise, in line with the aforementioned considerations, a gradual transition may not have a significant effect on the power spectrum). In what follows we take $S = 200$.

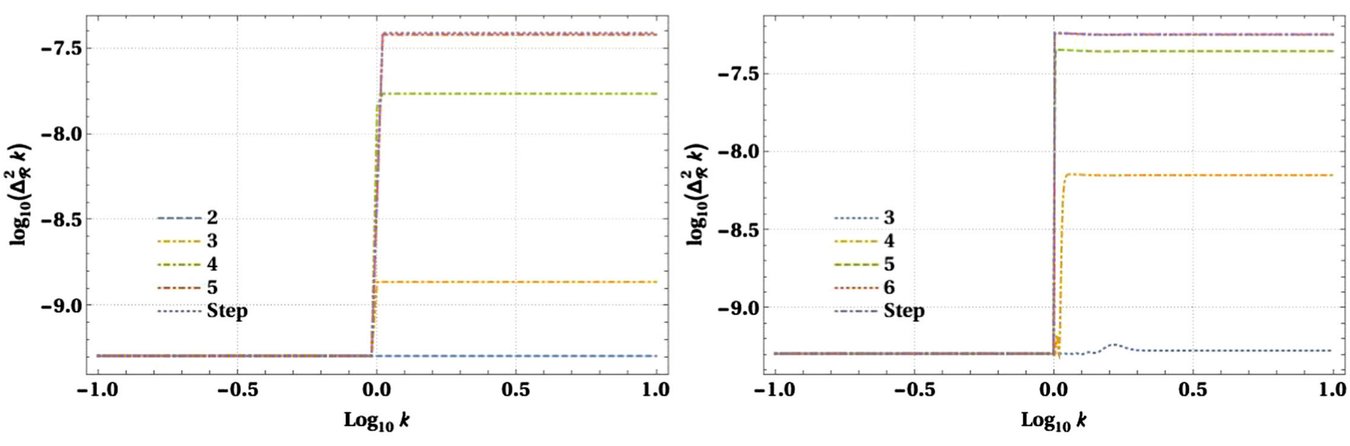

FIG. 11. The primordial power spectrum evaluated at different levels of violations of the adiabaticity condition (A3), when transition across the high energy cutoff scale is interpolated using a logistic function [Eq. (A4)], and numerically integrated. The numbers in the legend keys refer to the order of magnitude above the critical value of γ required to violate the adiabaticity condition [Eq. (A5)]. Left panel: interpolation between high energy sound speed $c_s = 100$ and standard regime ($c_s = 1$). Right panel: interpolation between high energy sound speed $c_s = 0.01$ and standard regime. The results are shown for de Sitter inflation, and wave numbers on the horizontal axis are expressed in terms of the high energy cutoff scale k_c, with $k_c/H = 300$.

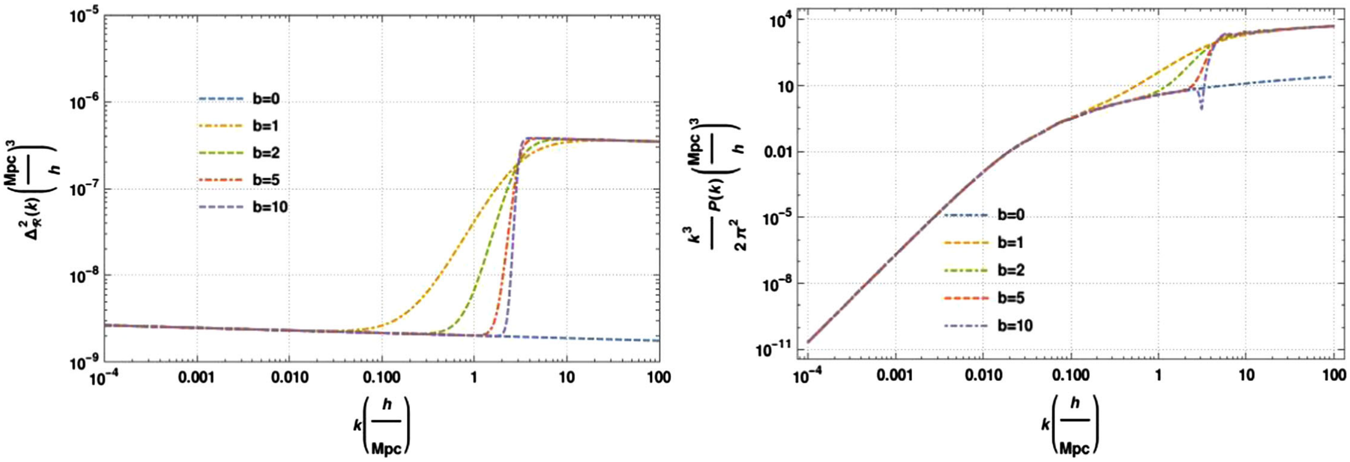

FIG. 12. Primordial (left) and matter (right) power spectra, modified at small scales using Eqs. (B1) and (B2), showing progressively steeper transitions with increasing b. The characteristic high energy transition scale is taken as $k_c = 3h\,\text{Mpc}^{-1}$ comoving, and $S = 200$.

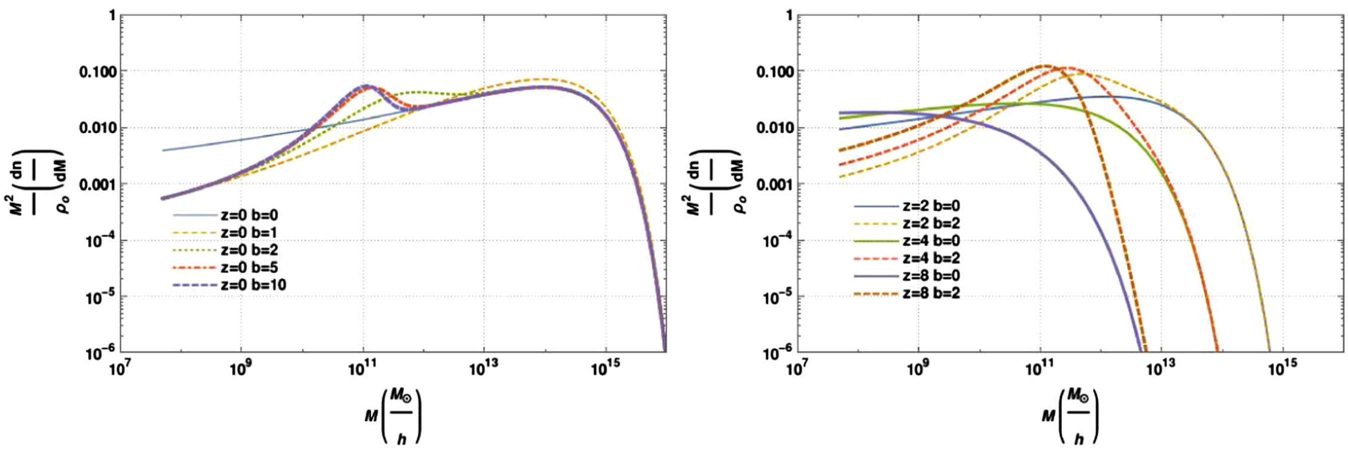

FIG. 13. Left panel: the halo multiplicity functions at $z = 0$ for the spectra shown in Fig. 12. Right panel: the multiplicity functions at shown redshifts for the case $b = 2$.

Figure 12 shows the primordial power spectrum, as well as the dimensionless matter power spectrum calculated using publicly available CLASS code (class-code.net), for several values of b. The corresponding multiplicity functions, calculated using (57) and (60), are shown in Fig. 13, where the left hand panel displays results at $z = 0$. For large b, those results are similar to the sudden jump case. For $b = 1$ the effect is smeared out, with increase at halo masses $\gtrsim 10^{14}\,M_\odot$, which would increase tension with cluster counts, which is already present in thee standard model. The suppression at small mass scales is enhanced by the gradual transition.

Of particular interest is the intermediate, $b = 2$, case. The enhancement takes place at larger masses than the corresponding case with sudden jump (at $3h\,\text{Mpc}^{-1}$). The suppression at $z = 0$ takes place at larger masses as well. This allows for largely retaining the advantages of the sharp cutoff at $1h\,\text{Mpc}^{-1}$—in terms of simultaneously alleviating both the dearth of dearth of dwarf galaxy and too big to fail problems at $z = 0$, as well as accounting for early galaxy formation at higher redshifts—while avoiding any enhancement at scales of order $\gtrsim 10^{13}\,M_\odot$ at $z = 0$, and relatively mildly perturbing the matter power spectrum at $1h\,\text{Mpc}^{-1}$ (Fig. 12; see also Sec. IV C 3). The enhancement of the mass function at $z = 0$ may also be relevant for explaining the overestimation of abundance matching within the standard model of stellar masses of massive galaxies ([161]; Sec. IV B 2).

[1] D. H. Lyth and A. R. Liddle, *The Primordial Density Perturbation: Cosmology, Inflation and the Origin of Structure*, Revised Version (Cambridge University Press, Cambridge, England, 2009).
[2] P. Peter and J.-P. Uzan, *Primordial Cosmology*, Oxford Graduate Texts (Oxford University Press, Oxford, 2009).
[3] Y. Akrami *et al.* (Planck Collaboration), arXiv:1807.06211.
[4] D. S. Salopek, J. R. Bond, and J. M. Bardeen, Phys. Rev. D **40**, 1753 (1989).
[5] S. Dodelson and E. Stewart, Phys. Rev. D **65**, 101301 (2002).
[6] J. A. Adams, G. G. Ross, and S. Sarkar, Phys. Lett. B **391**, 271 (1997).
[7] D. J. H. Chung, E. W. Kolb, A. Riotto, and I. I. Tkachev, Phys. Rev. D **62**, 043508 (2000).
[8] C. P. Burgess, J. M. Cline, F. Lemieux, and R. Holman, J. High Energy Phys. 02 (2003) 048.
[9] A. E. Romano and M. Sasaki, Phys. Rev. D **78**, 103522 (2008).
[10] D. H. Lyth and A. A. Riotto, Phys. Rep. **314**, 1 (1999).
[11] M. Yamaguchi, Classical Quantum Gravity **28**, 103001 (2011).
[12] D. Baumann and L. McAllister, *Inflation and String Theory* (Cambridge University Press, Cambridge, UK, 2015).
[13] A. A. Starobinskij, Sov. J. Exp. Theor. Phys. Lett. **55**, 489 (1992), https://ui.adsabs.harvard.edu/abs/1992JETPL..55..489S/abstract.
[14] D. Polarski and A. A. Starobinsky, Nucl. Phys. **B385**, 623 (1992).
[15] J. Adams, B. Cresswell, and R. Easther, Phys. Rev. D **64**, 123514 (2001).
[16] J. Lesgourgues, Nucl. Phys. **B582**, 593 (2000).
[17] M. Yamaguchi, Phys. Rev. D **64**, 063502 (2001).
[18] B. Feng, M.-z. Li, R.-J. Zhang, and X.-m. Zhang, Phys. Rev. D **68**, 103511 (2003).
[19] O. Elgaroy, S. Hannestad, and T. Haugboelle, J. Cosmol. Astropart. Phys. 09 (2003) 008.
[20] J.-O. Gong, J. Cosmol. Astropart. Phys. 07 (2005) 015.
[21] C. Dvorkin and W. Hu, Phys. Rev. D **81**, 023518 (2010).
[22] M. H. Namjoo, H. Firouzjahi, and M. Sasaki, J. Cosmol. Astropart. Phys. 12 (2012) 018.
[23] J. Lesgourgues, D. Polarski, and A. A. Starobinsky, Mon. Not. R. Astron. Soc. **297**, 769 (1998).
[24] M. Gramann and G. Hütsi, Mon. Not. R. Astron. Soc. **316**, 631 (2000).
[25] M. Gramann and G. Hütsi, Mon. Not. R. Astron. Soc. **327**, 538 (2001).
[26] C. R. Contaldi, M. Peloso, L. Kofman, and A. Linde, J. Cosmol. Astropart. Phys. 07 (2003) 002.
[27] J.-O. Gong, J. Cosmol. Astropart. Phys. 07 (2005) 015.
[28] R. Sinha and T. Souradeep, Phys. Rev. D **74**, 043518 (2006).
[29] M. Joy, A. Shafieloo, V. Sahni, and A. A. Starobinsky, J. Cosmol. Astropart. Phys. 06 (2009) 028.
[30] M. Kamionkowski and A. R. Liddle, Phys. Rev. Lett. **84**, 4525 (2000).
[31] J. Yokoyama, Phys. Rev. D **62**, 123509 (2000).
[32] A. R. Zentner and J. S. Bullock, Phys. Rev. D **66**, 043003 (2002).
[33] B. Little, A. Knebe, and R. R. Islam, Mon. Not. R. Astron. Soc. **341**, 617 (2003).
[34] K. Enqvist, T. Sawala, and T. Takahashi, arXiv:1905.13580.
[35] V. Berezinsky, V. Dokuchaev, Y. Eroshenko, M. Kachelrieß, and M. A. Solberg, Phys. Rev. D **81**, 103529 (2010).
[36] M. Leo, C. M. Baugh, B. Li, and S. Pascoli, J. Cosmol. Astropart. Phys. 12 (2018) 010.
[37] M. Yamaguchi, Phys. Rev. D **64**, 063503 (2001).
[38] T. Bringmann, C. Kiefer, and D. Polarski, Phys. Rev. D **65**, 024008 (2002).
[39] R. Easther and H. Peiris, J. Cosmol. Astropart. Phys. 09 (2006) 010.
[40] R. Bean, X. Chen, G. Hailu, S. H. H. Tye, and J. Xu, J. Cosmol. Astropart. Phys. 03 (2008) 026.
[41] M. Nakashima, R. Saito, Y. Takamizu, and J. Yokoyama, Prog. Theor. Phys. **125**, 1035 (2011).
[42] V. Miranda, W. Hu, and P. Adshead, Phys. Rev. D **86**, 063529 (2012).
[43] V. Miranda and W. Hu, Phys. Rev. D **89**, 083529 (2014).
[44] M. Park and L. Sorbo, Phys. Rev. D **85**, 083520 (2012).
[45] S. Gariazzo, O. Mena, H. Ramírez, and L. Boubekeur, Phys. Dark Universe **17**, 38 (2017).
[46] G. Ballesteros, J. Beltran Jimenez, and M. Pieroni, J. Cosmol. Astropart. Phys. 06 (2019) 016.
[47] J. Martin and R. H. Brandenberger, Phys. Rev. D **63**, 123501 (2001).
[48] R. H. Brandenberger and J. Martin, Mod. Phys. Lett. A **16**, 999 (2001).
[49] R. Easther, B. R. Greene, W. H. Kinney, and G. Shiu, Phys. Rev. D **64**, 103502 (2001).
[50] J. C. Niemeyer, Phys. Rev. D **63**, 123502 (2001).
[51] J. Martin and R. Brandenberger, Phys. Rev. D **68**, 063513 (2003).
[52] R. Easther, W. H. Kinney, and H. Peiris, J. Cosmol. Astropart. Phys. 08 (2005) 001.
[53] A. Kempf and L. Lorenz, Phys. Rev. D **74**, 103517 (2006).
[54] R. H. Brandenberger and J. Martin, Classical Quantum Gravity **30**, 113001 (2013).
[55] N. Kaloper, M. Kleban, A. E. Lawrence, and S. Shenker, Phys. Rev. D **66**, 123510 (2002).
[56] U. H. Danielsson, Phys. Rev. D **66**, 023511 (2002).
[57] R. Easther, B. R. Greene, W. H. Kinney, and G. Shiu, Phys. Rev. D **66**, 023518 (2002).
[58] J. C. Niemeyer, R. Parentani, and D. Campo, Phys. Rev. D **66**, 083510 (2002).
[59] A. Chatwin-Davies, A. Kempf, and R. T. W. Martin, Phys. Rev. Lett. **119**, 031301 (2017).
[60] R. H. Brandenberger and J. Martin, Int. J. Mod. Phys. A **17**, 3663 (2002).
[61] T. Zhu, A. Wang, K. Kirsten, G. Cleaver, and Q. Sheng, Phys. Rev. D **93**, 123525 (2016).
[62] A. Ashoorioon, R. Casadio, G. Geshnizjani, and H. J. Kim, J. Cosmol. Astropart. Phys. 09 (2017) 008.
[63] C. Armendariz-Picon and E. A. Lim, J. Cosmol. Astropart. Phys. 12 (2003) 006.
[64] G. Amelino-Camelia, Living Rev. Relativity **16**, 5 (2013).

[65] S. Hossenfelder, Living Rev. Relativity **16**, 2 (2013).
[66] W. G. Unruh, Phys. Rev. Lett. **46**, 1351 (1981).
[67] W. G. Unruh, Phys. Rev. D **51**, 2827 (1995).
[68] C. Barcelo, S. Liberati, and M. Visser, Living Rev. Relativity **8**, 12 (2005); **14**, 3 (2011).
[69] S.-Y. Ch and U. R. Fischer, Phys. Rev. Lett. **118**, 130404 (2017); **118**, 179901(A) (2017).
[70] T. Tanaka, arXiv:astro-ph/0012431.
[71] J. C. Niemeyer and R. Parentani, Phys. Rev. D **64**, 101301 (2001).
[72] A. A. Starobinsky, Pis'ma Zh. Eksp. Teor. Fiz. **73**, 415 (2001) [JETP Lett. **73**, 371 (2001)].
[73] M. Giovannini, Classical Quantum Gravity **20**, 5455 (2003).
[74] M. Porrati, Phys. Lett. B **596**, 306 (2004).
[75] B. R. Greene, K. Schalm, G. Shiu, and J. P. van der Schaar, J. Cosmol. Astropart. Phys. 02 (2005) 001.
[76] R. H. Brandenberger and J. Martin, Phys. Rev. D **71**, 023504 (2005).
[77] U. H. Danielsson, Phys. Rev. D **71**, 023516 (2005).
[78] K. Goldstein and D. A. Lowe, Phys. Rev. D **67**, 063502 (2003).
[79] R. Easther, W. H. Kinney, and H. Peiris, J. Cosmol. Astropart. Phys. 05 (2005) 009.
[80] T. Okamoto and E. A. Lim, Phys. Rev. D **69**, 083519 (2004).
[81] J. Martin and C. Ringeval, J. Cosmol. Astropart. Phys. 01 (2005) 007.
[82] J. Martin and C. Ringeval, Phys. Rev. D **69**, 127303 (2004).
[83] J. Chluba, A. L. Erickcek, and I. Ben-Dayan,Astrophys. J. **758**, 76 (2012).
[84] J. Chluba, R. Khatri, and R. A. Sunyaev, Mon. Not. R. Astron. Soc. **425**, 1129 (2012).
[85] R. Emami and G. Smoot, J. Cosmol. Astropart. Phys. 01 (2018) 007.
[86] C. S. Frenk and S. D. M. White, Ann. Phys. (Amsterdam) **524**, 507 (2012).
[87] A. Del Popolo and M. Le Delliou, Galaxies **5**, 17 (2017).
[88] J. S. Bullock and M. Boylan-Kolchin, Annu. Rev. Astron. Astrophys. **55**, 343 (2017).
[89] G. De Rosa *et al.*, Astrophys. J. **790**, 145 (2014).
[90] X.-B. Wu, F. Wang, X. Fan, W. Yi, W. Zuo, F. Bian, L. Jiang, I. D. McGreer, R. Wang, J. Yang, Q. Yang, D. Thompson, and Y. Beletsky, Nature (London) **518**, 512 (2015).
[91] D. Watson, L. Christensen, K. K. Knudsen, J. Richard, A. Gallazzi, and M. J. Michałowski, Nature (London) **519**, 327 (2015).
[92] S. Hirano, N. Zhu, N. Yoshida, D. Spergel, and H. W. Yorke, Astrophys. J. **814**, 18 (2015).
[93] C. L. Steinhardt, P. Capak, D. Masters, and J. S. Speagle, Astrophys. J. **824**, 21 (2016).
[94] R. Valiante, R. Schneider, M. Volonteri, and K. Omukai, Mon. Not. R. Astron. Soc. **457**, 3356 (2016).
[95] M. Stefanon, R. J. Bouwens, I. Labbé, A. Muzzin, D. Marchesini, P. Oesch, and V. Gonzalez, Astrophys. J. **843**, 36 (2017).
[96] K. Glazebrook, C. Schreiber, I. Labbé, T. Nanayakkara, G. G. Kacprzak, P. A. Oesch, C. Papovich, L. R. Spitler, C. M. S. Straatman, K.-V. H. Tran, and T. Yuan, Nature (London) **544**, 71 (2017).
[97] B. P. Venemans, F. Walter, R. Decarli, E. Bañados, C. Carilli, J. M. Winters, K. Schuster, E. da Cunha, X. Fan, E. P. Farina, C. Mazzucchelli, H.-W. Rix, and A. Weiss, Astrophys. J. Lett. **851**, L8 (2017).
[98] P. Behroozi and J. Silk, Mon. Not. R. Astron. Soc. **477**, 5382 (2018).
[99] J.-J. Tang, T. Goto, Y. Ohyama, C. Jin, C. Done, T.-Y. Lu, T. Hashimoto, E. Kilerci Eser, C.-Y. Chiang, and S. J. Kim, Mon. Not. R. Astron. Soc. **484**, 2575 (2019).
[100] R. Cecchi, M. Bolzonella, A. Cimatti, and G. Girelli, Astrophys. J. **880**, L14 (2019).
[101] A. Smith and V. Bromm, Contemp. Phys. **60**, 111 (2019).
[102] T. Wang *et al.*, Nature (London) **572**, 211 (2019).
[103] A. El-Zant, I. Shlosman, and Y. Hoffman, Astrophys. J. **560**, 636 (2001).
[104] A. A. El-Zant, Y. Hoffman, J. Primack, F. Combes, and I. Shlosman, Astrophys. J. **607**, L75 (2004).
[105] E. Romano-Díaz, I. Shlosman, Y. Hoffman, and C. Heller, Astrophys. J. **685**, L105 (2008).
[106] T. Goerdt, B. Moore, J. I. Read, and J. Stadel, Astrophys. J. **725**, 1707 (2010).
[107] D. R. Cole, W. Dehnen, and M. I. Wilkinson, Mon. Not. R. Astron. Soc. **416**, 1118 (2011).
[108] S. Inoue and T. R. Saitoh, Mon. Not. R. Astron. Soc. **418**, 2527 (2011).
[109] C. F. P. Laporte and S. D. M. White, Mon. Not. R. Astron. Soc. **451**, 1177 (2015).
[110] C. Nipoti and J. Binney, Mon. Not. R. Astron. Soc. **446**, 1820 (2015).
[111] A. Del Popolo, F. Pace, M. Le Delliou, and X. Lee, Phys. Rev. D **98**, 063517 (2018).
[112] J. I. Read and G. Gilmore, Mon. Not. R. Astron. Soc. **356**, 107 (2005).
[113] S. Mashchenko, J. Wadsley, and H. M. P. Couchman, Science **319**, 174 (2008).
[114] S. Peirani, S. Kay, and J. Silk, Astron. Astrophys. **479**, 123 (2008).
[115] A. Pontzen and F. Governato, Mon. Not. R. Astron. Soc. **421**, 3464 (2012).
[116] R. Teyssier, A. Pontzen, Y. Dubois, and J. I. Read, Mon. Not. R. Astron. Soc. **429**, 3068 (2013).
[117] A. Pontzen and F. Governato, Nature (London) **506**, 171 (2014).
[118] A. A. El-Zant, J. Freundlich, and F. Combes, Mon. Not. R. Astron. Soc. **461**, 1745 (2016).
[119] P. Colín, V. Avila-Reese, and O. Valenzuela, Astrophys. J. **542**, 622 (2000).
[120] P. Bode, J. P. Ostriker, and N. Turok, Astrophys. J. **556**, 93 (2001).
[121] A. V. Macciò, S. Paduroiu, D. Anderhalden, A. Schneider, and B. Moore, Mon. Not. R. Astron. Soc. **424**, 1105 (2012).
[122] M. R. Lovell, C. S. Frenk, V. R. Eke, A. Jenkins, L. Gao, and T. Theuns, Mon. Not. R. Astron. Soc. **439**, 300 (2014).
[123] A. El-Zant, S. Khalil, and A. Sil, Phys. Rev. D **91**, 035030 (2015).

[124] D. N. Spergel and P. J. Steinhardt, Phys. Rev. Lett. **84**, 3760 (2000).
[125] A. Burkert, Astrophys. J. **534**, L143 (2000).
[126] C. S. Kochanek and M. White, Astrophys. J. **543**, 514 (2000).
[127] J. Miralda-Escudé, Astrophys. J. **564**, 60 (2002).
[128] J. Zavala, M. Vogelsberger, and M. G. Walker, Mon. Not. R. Astron. Soc. **431**, L20 (2013).
[129] O. D. Elbert, J. S. Bullock, S. Garrison-Kimmel, M. Rocha, J. Oñorbe, and A. H. G. Peter, Mon. Not. R. Astron. Soc. **453**, 29 (2015).
[130] J. Goodman, New Astron. **5**, 103 (2000).
[131] W. Hu, R. Barkana, and A. Gruzinov, Phys. Rev. Lett. **85**, 1158 (2000).
[132] H.-Y. Schive, M.-H. Liao, T.-P. Woo, S.-K. Wong, T. Chiueh, T. Broadhurst, and W.-Y. P. Hwang, Phys. Rev. Lett. **113**, 261302 (2014).
[133] D. J. E. Marsh and J. Silk, Mon. Not. R. Astron. Soc. **437**, 2652 (2014).
[134] L. Hui, J. P. Ostriker, S. Tremaine, and E. Witten, Phys. Rev. D **95**, 043541 (2017).
[135] R. Genzel *et al.*, Astrophys. J. **733**, 101 (2011).
[136] A. Cava, D. Schaerer, J. Richard, P. G. Pérez-González, M. Dessauges-Zavadsky, L. Mayer, and V. Tamburello, Nat. Astron. **2**, 76 (2018).
[137] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, Phys. Rep. **215**, 203 (1992).
[138] C. Armendariz-Picon, T. Damour, and V.-i. Mukhanov, Phys. Lett. B **458**, 209 (1999).
[139] J. Garriga and V. F. Mukhanov, Phys. Lett. B **458**, 219 (1999).
[140] D. Salopek and J. Bond, Phys. Rev. D **42**, 3936 (1990).
[141] J. Martin and R. H. Brandenberger, Phys. Rev. D **63**, 123501 (2001).
[142] M. Lemoine, M. Lubo, J. Martin, and J.-P. Uzan, Phys. Rev. D **65**, 023510 (2001).
[143] J. Martin and R. H. Brandenberger, Phys. Rev. D **65**, 103514 (2002).
[144] Indeed, as discussed in Sec. IV A 2, in order to get modification of the matter power spectrum and halo mass function at scales of the order of a comoving Mpc, which are of particular interest, one needs $H/k_c \lesssim 10^{-4}$, much smaller than $c_s = 0.01$, which is the smallest we consider.
[145] F. Lucchin and S. Matarrese, Phys. Rev. D **32**, 1316 (1985).
[146] D. H. Lyth and E. D. Stewart, Phys. Lett. B **274**, 168 (1992).
[147] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, Phys. Lett. B **429**, 263 (1998).
[148] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, Phys. Lett. B **436**, 257 (1998).
[149] L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999).
[150] S. Abel and J. Santiago, J. Phys. G **30**, R83 (2004).
[151] Cf. Ref. [75]. Their more rigorous formulation using effective field theory invokes an "earliest time," which is defined as the time the smallest CMB scale leaves the cutoff scale. The earliest time here would correspond to that when the first scales cross the high energy cutoff. In both situations the origin of the time variation of the backreaction lies in the same change in number of excited states in the interval between k_c and H.
[152] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).
[153] Z. Lukić, K. Heitmann, S. Habib, S. Bashinsky, and P. M. Ricker, Astrophys. J. **671**, 1160 (2007).
[154] D. Boyanovsky, H. J. de Vega, and N. G. Sanchez, Phys. Rev. D **74**, 123006 (2006).
[155] W. H. Press and P. Schechter, Astrophys. J. **187**, 425 (1974).
[156] J. Bond, S. Cole, G. Efstathiou, and N. Kaiser, Astrophys. J. **379**, 440 (1991).
[157] R. K. Sheth, H. J. Mo, and G. Tormen, Mon. Not. R. Astron. Soc. **323**, 1 (2001).
[158] R. K. Sheth and G. Tormen, Mon. Not. R. Astron. Soc. **329**, 61 (2002).
[159] G. Despali, C. Giocoli, R. E. Angulo, G. Tormen, R. K. Sheth, G. Baso, and L. Moscardini, Mon. Not. R. Astron. Soc. **456**, 2486 (2016).
[160] J. Comparat, F. Prada, G. Yepes, and A. Klypin, Mon. Not. R. Astron. Soc. **469**, 4157 (2017).
[161] A. V. Macciò, S. Courteau, N. N. Q. Ouellette, and A. A. Dutton, Mon. Not. R. Astron. Soc. **496**, L101 (2020).
[162] S. H. Miller, R. S. Ellis, A. B. Newman, and A. Benson, Astrophys. J. **782**, 115 (2014).
[163] R. S. Somerville and R. Davé, Annu. Rev. Astron. Astrophys. **53**, 51 (2015).
[164] Note that in [98] the more conservative cumulative stellar mass function, which averages over the increasing $\frac{M_*}{M_h}(M_*)$, is employed. The data for $M_* = 10^8\,M_\odot$ to $10^{10}\,M_\odot$ are consistent with ΛCDM. Tension still arises for the data point shown for $M_* \geq 10^{11.7}\,M_\odot$ at $z \approx 5.5$ (their Fig. 2 top-left). The results in Fig. 18 of [165] show data consistent with $M_*/M_h = 1/6.3$, or even larger, already at $z \sim 5$, for $M_* \gtrsim 10^{11}\,M_\odot$. In general, it seems that galaxies can be accommodated into halos with $M_*/M_h \ll 1/6.3$ for $z \gtrsim 5$ if one keeps to $M_* \lesssim 10^{10.5}\,M_\odot$ (e.g., [166], and Fig. 9 of [167]). On the other hand, the most extreme cases in the data presented by Steinhardt *et al.* [93] (where $M_*/M_h \approx 1/7$ seems required), are just barely consistent with standard ΛCDM.
[165] I. Davidzon *et al.*, Astron. Astrophys. **605**, A70 (2017).
[166] P. S. Behroozi, R. H. Wechsler, and C. Conroy, Astrophys. J. **770**, 57 (2013).
[167] P. Behroozi, R. H. Wechsler, A. P. Hearin, and C. Conroy, Mon. Not. R. Astron. Soc. **488**, 3143 (2019).
[168] C. Mancuso, A. Lapi, J. Shi, J. Gonzalez-Nuevo, R. Aversa, and L. Danese, Astrophys. J. **823**, 128 (2016).
[169] B. Alcalde Pampliega, P. G. Pérez-González, G. Barro, H. Domínguez Sánchez, M. C. Eliche-Moral, N. Cardiel, A. Hernán-Caballero, L. Rodriguez-Muñoz, P. Sánchez Blázquez, and P. Esquej, Astrophys. J. **876**, 135 (2019).
[170] K. Glazebrook, C. Schreiber, I. Labbé, T. Nanayakkara, G. G. Kacprzak, P. A. Oesch, C. Papovich, L. R. Spitler, C. M. S. Straatman, K.-V. H. Tran, and T. Yuan, Nature (London) **544**, 71 (2017).
[171] K. Mawatari, A. K. Inoue, T. Hashimoto, J. Silverman, M. Kajisawa, S. Yamanaka, T. Yamada, I. Davidzon, P. Capak,

L. Lin, B.-C. Hsieh, Y. Taniguchi, M. Tanaka, Y. Ono, Y. Harikane, Y. Sugahara, S. Fujimoto, and T. Nagao, Astrophys. J. **889**, 137 (2020).

[172] C. M. S. Straatman *et al.*, Astrophys. J. **783**, L14 (2014).

[173] C. Schreiber, K. Glazebrook, T. Nanayakkara, G. G. Kacprzak, I. Labbé, P. Oesch, T. Yuan, K. V. Tran, C. Papovich, L. Spitler, and C. Straatman, Astron. Astrophys. **618**, A85 (2018).

[174] P. Santini, E. Merlin, A. Fontana, B. Magnelli, D. Paris, M. Castellano, A. Grazian, L. Pentericci, S. Pilo, and M. Torelli, Mon. Not. R. Astron. Soc. **486**, 560 (2019).

[175] E. Merlin, F. Fortuni, M. Torelli, P. Santini, M. Castellano, A. Fontana, A. Grazian, L. Pentericci, S. Pilo, and K. B. Schmidt, Mon. Not. R. Astron. Soc. **490**, 3309 (2019).

[176] F. Valentino *et al.*, Astrophys. J. **889**, 93 (2020).

[177] G. Girelli, M. Bolzonella, and A. Cimatti, Astron. Astrophys. **632**, A80 (2019).

[178] B. Forrest *et al.* Astrophys. J. Lett. **890**, L1 (2020).

[179] M. Song *et al.*, Astrophys. J. **825**, 5 (2016).

[180] J. A. Peacock, *Cosmological Physics*, edited by J. A. Peacock (Cambridge University Press, Cambridge, England, 1999), pp. 704.

[181] Y. Harikane *et al.*, Astrophys. J. **821**, 123 (2016).

[182] A. Leauthaud *et al.*, Mon. Not. R. Astron. Soc. **467**, 3024 (2017).

[183] A. Arinyo-i Prats, J. Miralda-Escud, M. Viel, and R. Cen, J. Cosmol. Astropart. Phys. 12 (2015) 017.

[184] J. Binney and S. Tremaine, *Galactic Dynamics: Second Edition*, edited by J. Binney and S. Tremaine (Princeton University Press, Princeton, NJ, 2008).

[185] Note also, in the latter case of $c_s = 0.01$, the small bump on the right, visible in the low γ spectra beyond the cutoff scale. These are due to nonadiabaticity at the crossing of the high energy transition at η_c. They are much less prominent for smaller $k_c/H \geq 10^4$ values, used in the rest of this work. They nevertheless represent another potentially interesting effect associated with nonadiabatic evolution.

[186] T. Nakama, J. Chluba, and M. Kamionkowski, Phys. Rev. D **95**, 121302 (2017).