Primordial gravitational waves from NANOGrav: a broken power-law approach

Micol Benetti,1,2,* Leila L. Graef,3,† and Sunny Vagnozzi4,‡

1Scuola Superiore Meridionale (SSM), Università di Napoli “Federico II”, Largo San Marcello 10, I-80138 Napoli, Italy
2Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Napoli, Via Cinthia 9, I-80126 Napoli, Italy
3Instituto de Física, Universidade Federal Fluminense, Avenida General Milton Tavares de Souza s/n, Gragoatá, 24210-346 Niterói, Rio de Janeiro, Brazil
4Kavli Institute for Cosmology (KICC) and Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom

(Dated: January 10, 2022)

We revisit the possibility that the stochastic common-spectrum process recently detected by the NANOGrav pulsar timing array experiment could be due to primordial gravitational waves (GWs). A naïve extrapolation down to interferometer scales of the blue GW spectrum required to explain NANOGrav consistently with Cosmic Microwave Background (CMB) observations would strongly violate upper limits on the stochastic GW background (SGWB) amplitude from LIGO/Virgo. In combination with the fact that there are over 19 decades in frequency between CMB and interferometer scales, this motivates us to move beyond the commonly adopted approximation of a pure power-law GW spectrum. We consider a broken power-law parametrization for the SGWB spectrum, which turns from blue to red above the break frequency: while phenomenological, this choice maps to various well-motivated early-Universe models, including scenarios featuring non-instantaneous reheating or a non-standard background expansion following reheating. After a detailed discussion of the contribution of the resulting SGWB to the early-Universe radiation energy density, we constrain the broken power-law model against a wide variety of multi-frequency cosmological and GW observations. We find that this phenomenological model is able to explain the NANOGrav signal while remaining in agreement with upper limits on the tensor-to-scalar ratio on CMB scales, Big Bang Nucleosynthesis constraints on the early-Universe radiation energy density, and upper limits on the SGWB amplitude on interferometer scales. We briefly discuss the very bright prospects for testing this model with next-generation probes across the GW frequency landscape, which motivate further exploring connections to specific well-motivated early-Universe models.

I. INTRODUCTION

The first detection of gravitational waves (GWs) from the coalescence of two black holes (BHs) in 2015 [1] has opened an entirely new window onto the Universe, the behavior of its most extreme objects, and the laws of physics at energy scales and regimes which are completely unaccessible on Earth [2–4]. While so far only resolved GW events have been detected, the superposition of numerous incoherent GW sources would instead generate a stochastic GW background (SGWB). The existence of a SGWB is a firm prediction of several well-motivated cosmological and astrophysical scenarios, operating both in the early and late Universe (see e.g. Refs. [5–7] for reviews). Merging supermassive BH (SMBH) binaries are perhaps the best motivated example on the astrophysical side. On the cosmological side, SGWB sources include for instance phase transitions and relics thereof (including topological defects such as cosmic strings). There is no doubt that a direct observation of the cosmological or astrophysical SGWB would be a tremendous achievement, whose implications for astrophysics, cosmology, and high-energy physics would be momentous.

The predicted astrophysical and cosmological SGWB spans an extremely wide frequency range. Significant theoretical and experimental effort has gone into the development of a diverse range of probes (currently running, upcoming, or proposed) which will be able to search for GWs (both resolved GW events or the SGWB) in various frequency bands, from low-frequency \( (f \approx 10^{-20} \text{ Hz}) \) GWs to GWs in the kHz band. From the lowest to the highest frequencies, these probes and/or observables include (but are not limited to) Cosmic Microwave Background (CMB) B-modes [8], spectral distortions [9,10], pulsar timing arrays (PTAs) [11–13], binary resonance [14, 15], and finally direct detection with atomic [16–18], Earth-based [19], and space-based interferometers [20]. Of particular interest to this work are PTAs, which are sensitive to GWs with frequencies in the nHz range, and exploit the fact that millisecond pulsar behave as extremely stable clocks. GWs induce spatially correlated fluctuations in the arrival times of radio pulses from millisecond pulsars [21], which PTAs then search for. The \( \gtrsim \mathcal{O}(\text{nHz}) \) region is of particular interest from the astrophysical and cosmological points of view: the astrophysical SGWB from merging SMBH binaries (SMBHBs) is expected to peak within this band [22], and the same is true for the cosmological SGWB arising within some of the simplest models of cosmic strings [23].

A particularly well-motivated cosmological SGWB source is cosmic inflation [24–30], the leading paradigm for the solution of the flatness, horizon, and monopole

* micol.benetti@unina.it
† leilagraef@id.uff.br
‡ sunny.vagnozzi@ast.cam.ac.uk
problems, as well as the generation of primordial density perturbations. Inflationary vacuum fluctuations become classical on large scales, and induce both scalar and tensor perturbations. Upon horizon re-entry, the latter give rise to the inflationary SGWB, imprinting a distinctive signature in the CMB B-mode polarization, which is therefore among the cleanest probes of the inflationary SGWB [8]. The amplitude and scale dependence of the inflationary SGWB is typically parametrized via the tensor-to-scalar ratio $r$ (characterizing the amplitude of tensor fluctuations relative to scalar ones) and the tensor spectral index $n_T$ respectively. The simplest models of inflation, driven by a single dynamical slow-rolling (scalar) field, predict a spectrum of scalar fluctuations which is nearly scale-invariant (albeit slightly red, with more power on large rather than small scales) and highly Gaussian. These predictions are in excellent agreement with observations, which strongly constrain non-Gaussianity and deviations from scale-invariance, lending very strong support to the inflationary paradigm [31].

Single-field slow-roll models also predict a power spectrum of tensor fluctuations which is slightly red. To leading order in slow-roll parameters, the tensor-to-scalar ratio and tensor spectral index within these models satisfy the so-called inflationary consistency relation [32]:

$$r = -8n_T,$$  \hspace{1cm} (1)

which thus requires $n_T \leq 0$ (hence a red spectrum), since $r \geq 0$. Within these models, and given current constraints on $r$ [33], the amplitude of the inflationary SGWB on PTA and interferometer scales is far too small to be detectable by these probes, which would instead require a strong blue tilt or in any case a strong enhancement on small scales. However, as we will discuss in more detail in Sec. II, several well-motivated inflationary (and non-inflationary) models beyond the simplest ones naturally predict a blue tilt for the tensor power spectrum (see e.g. Refs. [34–71]). Before moving on, let us mention that there is certainly ample theoretical motivation for going beyond the observationally highly successful single-field slow-roll paradigm. For instance, recent work within the “swampland” program has pointed out difficulties in embedding the simplest inflationary models within quantum gravity-consistent UV completions [72–76], whereas such difficulties may be evaded when moving past the single-field slow-roll paradigm [77–92]. Therefore, the possibility that non-minimal inflationary models or alternatives to inflation leading to a blue tensor spectrum may be probed at PTA or interferometer scales is one which is highly worthy of consideration from both the theoretical and observational points of view. This will be the starting point for our work, which, as anticipated earlier, shall be concerned with the SGWB in the $\gtrsim O(\text{nHz})$ frequency range, to which PTAs are sensitive.

The North American Nanohertz Observatory for Gravitational Waves (NANOGrav) is a PTA collaboration which has been collecting pulsar timing data since 2004 [93]. NANOGrav recently released their 12.5-year dataset [94], which contains time-of-arrival measurements for 47 millisecond pulsars observed at the Arecibo Observatory and the Green Bank Telescope between 2004 and 2017. An analysis searching for an isotropic SGWB by the NANOGrav collaboration in their 12.5-year dataset yielded strong evidence for a stochastic common-spectrum process against independent red-noise processes [95], with Bayes factors in found to lie within the range $2.7 \lesssim \log_{10} B \lesssim 4.5$ depending on the solar system ephemeris (SSE) modeling scheme adopted. This result, if genuine and confirmed, raises the tantalizing possibility that NANOGrav may have achieved the first ever SGWB detection. The NANOGrav signal is consistent with a SGWB with characteristic strain amplitude $A_{\text{CP}} \sim 10^{-15}$ \footnote{The subscript “CP” stands for “common(-spectrum) process”.} at a reference frequency $f_{\text{ref}} = 1 \text{yr}^{-1} \approx 3 \times 10^{-8} \text{Hz}$. It is worth pointing out that the NANOGrav detection shows no evidence for quadrupolar spatial correlations [95], as described by the Hellings-Downs (HD) curve [96]. \footnote{The Bayes factor against a spatially uncorrelated common-spectrum process lies within the range $0.37 \lesssim \log_{10} B \lesssim 0.64$, again depending on the SSE model adopted [95].} The lack of evidence for “tell-tale” quadrupolar HD correlations constitutes grounds for caution with regards to claims of the NANOGrav signal constituting a genuine SGWB detection, and the NANOGrav collaboration themselves advocate for a prudent approach on the matter [95]. Intriguingly, the Parkes Pulsar Timing Array (PPTA) collaboration very recently also reported evidence for a common-spectrum process with amplitude and frequency range consistent with the NANOGrav signal in their DR2 dataset [97]. However, this tentative signal once again lacks convincing evidence for quadrupolar HD correlations. Even more recently, similar hints (though once more including the lack of quadrupolar HD correlations) were reported in the European Pulsar Timing Array (EPTA) 24-year dataset [98], strengthening the hints for the possible first detection of a SGWB in the nHz range. \footnote{\label{footnote:a}The possibility that NANOGrav may have observed GW polarization modes other than the standard tensor transverse (TT) ones present in General Relativity (GR) was first discussed in detail in Ref. [99]. The authors claimed strong evidence for scalar transverse (ST) polarization modes, and no evidence for scalar longitudinal (SL) or vector longitudinal (VL) polarization modes. A similar study was performed in Ref. [100] on data from the International Pulsar Timing Array (IPTA), a collaboration comprising the EPTA in addition to NANOGrav and PPTA [101], finding this time weak evidence for ST polarization modes, and placing upper limits on the amplitude of TT polarization modes consistent with the NANOGrav signal. Yet another related study was performed in Ref. [102] on PPTA data, finding no evidence for either of the TT, ST, SL, and VL polarization modes, but only placing upper limits on the amplitude thereof, posing questions as to the origin of the tentative detection of ST polarization modes in earlier works. Finally, the NANOGrav collaboration themselves searched for non-GR polarization modes in Ref. [103], again finding a preference for ST modes over TT modes.}
Keeping in mind the important HD-related caveats discussed above, it is nevertheless worthwhile to contemplate possible scenarios explaining the NANOGrav signal, assuming it constitutes a genuine SGWB detection. As mentioned earlier, the astrophysical SGWB from merging SMBHBs is expected to peak within the range probed by NANOGrav, and remains a possible valid interpretation of the signal [104], although the required local number density of SMBHBs is a factor of five larger than predicted by most astrophysical models [105]. Quite predictably, the NANOGrav detection has stimulated significant activity focused on identifying possible fundamental physics scenarios responsible for the signal, mostly operating in the primordial Universe. Some of the earliest interpretations of the NANOGrav signal involved cosmic strings [106–109] or second-order effects associated to the formation of primordial BHs from the collapse of large curvature perturbations [110, 111], but various other theoretical scenarios have been considered since then (see e.g. Refs. [112–161] for examples, as well as Ref. [162] for a comprehensive characterization of various possible sources for the NANOGrav signal).

Another intriguing possibility, which was first explored by one of us in Ref. [163], and then further in Ref. [164], posits that the NANOGrav may be due to an inflationary SGWB. In particular, under the commonly adopted assumption of a pure power-law parametrization for the underlying primordial tensor power spectrum, Ref. [163] finds that explaining the NANOGrav signal requires: a) a very blue ($n_T > 0$) primordial tensor power spectrum, and hence a violation of the consistency relation in order to remain consistent with upper limits on the tensor-to-scalar ratio (and therefore the SGWB amplitude) on CMB scales; b) a very low reheating temperature, in order to not violate Big Bang Nucleosynthesis (BBN) constraints on the radiation energy density in the early Universe. Related findings were also reported in Ref. [164], where late-time entropy production between the end of inflation and the BBN epoch was also taken into consideration. The possibility that the NANOGrav signal may be due to inflationary GWs remains therefore viable, but would require an inflationary model beyond the simplest single-field slow-roll ones, and a non-standard reheating and/or post-reheating scenario.

In this work, we shall keep pursuing the intriguing possibility that the NANOGrav collaboration may indeed have detected a primordial SGWB, remaining agnostic as to whether the latter is related or not to inflation. Our analysis goes beyond these earlier works (and in particular Ref. [163]) in at least three important respects:

- we go beyond widely adopted, but overly simplistic, approximation of a pure power-law for the primordial SGWB spectrum;

- in addition to BBN bounds, we also take into account SGWB constraints at higher frequencies from LIGO/Virgo [165, 166], which the model studied in Ref. [163] would naively violate;

- we perform a fully-fledged likelihood analysis including (besides NANOGrav) other standard cosmological datasets to more properly constrain the primordial SGWB interpretation of the NANOGrav signal, and paying particular attention to the Hubble constant $H_0$, whose inferred value can be particularly sensitive to the existence of a blue GW spectrum, given the contribution of the latter to the effective number of relativistic species $N_{\text{eff}}$, which correlates with $H_0$.

The first of the above three points is particularly important. While the approximation of SGWB power spectrum being described by a pure power-law across various decades in frequency is widespread, it might not be justified when considering probes spanning a wide frequency range and hence providing a large lever arm [167]; for instance, there are 19 decades in frequency between the scales probed by CMB experiments such as Planck and interferometers such as LIGO/Virgo. Here, we shall go beyond the power-law form for the primordial tensor power spectrum, adopting a phenomenological broken power-law ansatz, recently proposed in Ref. [168]. While phenomenological, this parametrization (which we shall refer to as “broken power-law spectrum”) covers a broad range of early-Universe scenarios, ranging from non-instantaneous reheating to a non-standard post-reheating background evolution.

The rest of this paper is then organized as follows. In Sec. II we introduce the standard power-law SGWB spectrum and the associated energy density and number of relativistic species $N_{\text{eff}}$, before presenting the broken power-law SGWB power spectrum and discussing the associated observational quantities such as $N_{\text{eff}}$. In Sec. III we introduce the GW observations against which we constrain the model, placing constraints on the SGWB energy density within various frequency ranges. In Sec. IV we outline the cosmological data and analysis methodology adopted. Our results, and in particular constraints on the broken power-law SGWB spectrum and the possibility of the NANOGrav signal being due to primordial GWs, are presented in Sec. V. Finally, we draw concluding remarks in Sec. VI.

II. STANDARD AND NON-STANDARD PRIMORDIAL GW SPECTRA

A. Standard power-law spectrum

The primordial spectrum of tensor perturbations $P_T(k)$ is a key quantity to connect theoretical early-Universe predictions (including those from inflation) to cosmological observations. A widespread choice in the
literature is that of assuming that the primordial tensor spectrum scales as a pure power-law across various decades in comoving wavenumber $k$:

$$P_T^{\text{prim-pl}}(k) = A_T(k_*) \left( \frac{k}{k_*} \right)^{n_T},$$

(2)

where $A_T$ is the amplitude of the primordial tensor power spectrum (itself related to the tensor-to-scalar ratio $r \equiv A_T/A_s$, where $A_s$ is instead the amplitude of the primordial power spectrum of scalar perturbations) at the pivot scale $k_*$, $n_T$ is the tensor tilt assumed to be a constant throughout the spectrum, and the superscript “pl” stands for “power-law”. It can be convenient to express the above spectrum in frequency space, with frequency $f$ related to $k$ by $k = 2\pi f$.

Within single-field slow-roll inflationary models, the consistency relation implies $r = -8n_T$, and therefore a negative tensor spectral index and correspondingly a red GW spectrum. However, as alluded to earlier, this is not necessarily the case in a number of other (less minimal but not for this reason less motivated) early-Universe scenarios (particularly those involving violations of the null-energy condition), including but not limited to: inflationary models based on modifications to gravity [34–39], non-commutative space-times [40, 41], spatial or temporal diffeomorphism invariance breaking [42–46], and more generally violations of the null-energy condition [47], or models involving couplings to gauge fields and spin-2 fields [48–54], particle production during inflation [55–57], elastic media [58], a non-Bunch-Davies initial state [59], higher order effective gravitational action corrections [60], second-order effects [61] (possibly associated to the formation of primordial BHs), sound speed resonance or a decrease in the GW sound speed during inflation [62, 63], as well as alternatives to inflation such as string gas cosmology [64–68], ekpyrotic scenarios [69, 70], and matter bounces [71].

The tensor power spectrum $P_T(\eta,k)$ at a given conformal time $\eta$ is related to its primordial counterpart $P_T^{\text{prim}}(k)$ as follows:

$$P_T(\eta,k) = T_T^2(\eta,k) P_T^{\text{prim}}(k),$$

(3)

where $T_T(\eta,k)$ is the transfer function, which accounts for the evolution of tensor perturbations across the various epochs of the Universe’s expansion history up to $\eta$. Assuming that reheating at the end of inflation is followed by the standard epochs of radiation domination, matter domination, and dark energy domination, the transfer function admits a simple analytical approximation (see Refs. [172–178]), which we too shall adopt.5

Another key quantity in connecting theory to observations is the GW energy density today $\rho_{GW}$, given by:

$$\rho_{GW} = \int_{k_{IR}}^{k_{UV}} d\ln k \frac{P_T(k)}{32\pi G a^2} [T'(k,\eta_0)]^2,$$

(4)

where $\eta_0$ is the current conformal time, and $'$ denotes a conformal time derivative. The upper and lower integration limits $k_{UV}$ and $k_{IR}$ correspond to physical ultraviolet (UV) and infrared (IR) cutoff scales. For the purpose of comparison to observations, it is also convenient to define the GW dimensionless density parameter $\Omega_{GW}$, given by (see e.g. Refs. [60, 180, 181]):

$$\Omega_{GW} = \frac{1}{\rho_c} \frac{d\rho_{GW}}{d\ln k} = \frac{1}{12} \left( \frac{k}{a_0 H_0} \right)^2 T_T^2(k) P_T^{\text{prim}}(k),$$

(5)

where $H_0$ and $a_0$ are the current values of the Hubble expansion rate and scale factor respectively, and $\rho_c = 3H_0^2/8\pi G$ is the critical density of the Universe today.

In particular, we shall denote by $\Omega_{GW}$ the GW density parameter (as a function of frequency) associated to the pure power-law primordial tensor spectrum $P_T^{\text{prim-pl}}(k)$ given in Eq. (2), i.e. obtained combining Eqs. (2,5).

Mathematically speaking, the IR and UV cutoffs are introduced since the integral in Eq. (4) would diverge in the IR for $n_T \leq -4$ and in the UV for $n_T \geq -2$. Physically speaking, for the IR cutoff, the only modes which contribute to the radiation energy density at any given time are subhorizon modes, as those are the ones which oscillate and propagate as massless modes, hence contributing to the local energy density (although this statement is to some extent ambiguous, as the total energy density is a quantity which can only be measured averaging over several wavelengths). This implies that $k_{IR}$ is a time-dependent quantity, as the horizon itself is time-dependent. In this work we shall be concerned with a blue GW spectrum, for which it is the UV modes which dominate the energy density integral. Therefore, we can safely take the limit $k_{IR} \to 0$ Hz, or more precisely the limit $k_{IR}/k_{UV} \to 0$, with virtually no impact on our results (see Ref. [182] for related discussions).

The choice of UV cutoff is, instead, more arbitrary, although our being interested in a blue GW spectrum implies that this choice will nonetheless have an important impact on our results. If inflation is responsible for the production of primordial tensor modes, one expects an UV cutoff given the size of the horizon at the end of inflation. For GUT-scale inflation and instant reheating, this would correspond approximately to $k_{UV} \sim 10^{23} \text{Mpc}^{-1}$ and hence $f_{UV} \sim 10^8$ Hz [183]. However, a more conservative assumption could be that of not committing to any specific early-Universe model for generating GWs, but simply to require that the power-law spectrum extends over $\sim 60$ e-folds, i.e. the maximum amount of hot Big Bang expansion. For a pivot scale of $k_* = 10^{-2} \text{Mpc}^{-1}$, this implies $k_{UV}/k_* \sim 10^{24}$, leading to $k_{UV} \sim 10^{22} \text{Mpc}^{-1}$ and therefore $f_{UV} \sim 10^8$ Hz, a result which is very close to that obtained from the previous inflation-based argument [182].

---

4 See e.g. Ref. [169] for a comprehensive discussion of inflationary models and alternatives leading to a blue spectrum, motivated by BICEP2’s claimed detection in 2014 [170, 171].

5 See Ref. [179] for a recent re-appraisal of some of these widely adopted approximations.
Less conservative arguments could be used to justify higher values of \( f_{UV} \). For instance, one would in any case expect that the largest possible wavenumber for GWs produced in the early Universe is set by the Planck scale, for which \( k_{UV} \approx 10^{37} \text{Mpc}^{-1} \) and accordingly \( f_{UV} \approx 10^{41} \text{Hz} \). This choice was adopted, for example, in Ref. [67]. A slightly less aggressive choice could be to replace the Planck scale by the GUT scale, from which one gets \( k_{UV} \approx 10^{34} \text{Mpc}^{-1} \) and \( f_{UV} \approx 10^{40} \text{Hz} \). Of course, a sharp cut in \( k \)-space is a simplification and one would in fact generically expect a smooth transition between modes which contribute to the energy density and modes which don’t. However, we do not expect this choice to have a significant impact on our results. We will discuss our specific choice for \( f_{UV} \) later in Sec. IIB, when discussing the choice of broken power-law spectrum we adopt, going beyond the pure power-law in Eq. (2).

For a blue GW spectrum (\( n_T > 0 \)), the integral in Eq. (4) can be solved by adopting the aforementioned analytical approximations for the transfer function \( T(\eta, k) \). Up to corrections of order \( k_{UV}/k_{IR} \), which are extremely tiny, this gives (see e.g. Ref. [182] for the full calculation):

\[
\rho_{GW} = \frac{A_s r}{32\pi G} \left( \frac{k_{UV}}{k_*} \right)^{n_T} \frac{1}{2\pi^2 (a\eta)^2} = \frac{A_s r}{24\pi} \left( \frac{k_{UV}}{k_*} \right)^{n_T} \rho_{\text{tot}},
\]

where the second line is only valid deep during the radiation domination era, and follows from the fact that during this epoch \( 1/(a\eta)^2 = H^2 = 8\pi G \rho_{\text{tot}}/3 \). Here, \( \rho_{\text{tot}} \) is the total energy density of the Universe, which during radiation domination is given by the sum of the photon (\( \gamma \)), neutrino (\( \nu \)), and GW energy densities:

\[
\rho_{\text{tot}} = \rho_\gamma + \rho_\nu + \rho_{GW} \equiv \rho_\gamma \left( 1 + \frac{7}{8} \left( \frac{4}{11} \right)^{4/3} N_{\text{eff}} \right)
\]

\[
= \rho_\gamma \left( 1 + \frac{7}{8} \left( \frac{4}{11} \right)^{4/3} 3.046 \right) + \rho_{GW},
\]

where the first line defines the effective number of relativistic degrees of freedom \( N_{\text{eff}} \) (including contributions from photons, neutrinos, and GWs), and in the second line we are assuming that the three Standard Model neutrino families provide the standard contribution \( N_{\nu_{\text{eff}}} = 3.046 \) [184]. Therefore, the total effective number of relativistic degrees of freedom is given by \( N_{\text{eff}} = N_{\nu_{\text{eff}}} + N_{GW} = 3.046 + N_{GW} \)

which defines the GW contribution to \( N_{\text{eff}} \), given by \( N_{GW} \).

If the GW energy density is a subdominant component of the total radiation energy density, i.e. \( \rho_{GW}/\rho_{\text{tot}} \ll 1 \), we can substitute Eq. (7) into Eq. (6), and solve for \( N_{\text{eff}} \) to obtain an expression which depends only on \( A_s, r, n_T \), and \( k_{UV} \) (see Ref. [182]). Taylor expanding to first order in \( \rho_{GW}/\rho_{\text{tot}} \) leads to the following expression for \( N_{\text{eff}} \):

\[
N_{\text{eff}} \approx 3.046 + \left[ 3.046 + \frac{8}{7} \left( \frac{11}{4} \right)^{4/3} \right] \frac{A_s r}{24\pi} \left( \frac{k_{UV}}{k_*} \right)^{n_T},
\]

an expression which we stress again is valid only if the GW energy density is small compared to the total radiation energy density, which fortunately is the case given observational constraints. From Eq. (8) we can directly read off \( N_{GW} \) given by:

\[
N_{GW} \approx \left[ 3.046 + \frac{8}{7} \left( \frac{11}{4} \right)^{4/3} \right] \frac{A_s r}{24\pi} \left( \frac{k_{UV}}{k_*} \right)^{n_T},
\]

which we recall is only valid for a blue GW spectrum.

In the literature, the assumption of the primordial tensor power spectrum scaling as a pure power-law up to the cutoff frequency \( k_{UV} \) as in Eq. (2) is a widespread one: see e.g. Refs. [163, 182, 183, 195–201], as well as the Planck 2018 constraints on inflation paper [31], for an inevitably incomplete example of recent works adopting this assumption. Other commonly adopted simplifying assumptions include that of a standard thermal history with an instantaneous transition from the early-Universe phase responsible for sourcing the primordial fluctuations (be it inflation or an alternative mechanism, usually taking place at a very high energy scale) to the phase of radiation domination, i.e. an instantaneous reheating phase with no extended matter or stiff matter domination era prior to the usual radiation domination era.

However, all of these are clearly approximations at best, at least for the case of the inflationary SGWB. For instance, by making use of the powerful inflationary flow formalism, it was recently argued in Ref. [167] that in an ensemble of realistic single-field inflationary models, a power-law extrapolation from CMB to interferometer scales can overestimate the amplitude of primordial tensor modes by up to two orders of magnitude. The largest frequencies are sensitive to the non-slow-roll dynamics towards the end of inflation, where the expansion rate and slow roll parameters vary relatively rapidly, typically leading to one or more breaks in the primordial tensor spectrum (see e.g. Refs. [163, 202–204] for related discussions). As a result, the tensor spectral index measured on CMB scales can be completely uncorrelated from the same quantity probed on interferometer scales. We remark, however, that the analysis of Ref. [167] strictly speaking only applies to single-field models which, as discussed in the Introduction, we know cannot accommodate the NANOGrav signal due to the latter requiring a strongly blue GW spectrum.

---

6 This standard value has recently been re-evaluated by several groups to include a more precise treatment of flavour oscillations and finite-temperature effects with the final results converging towards \( N_{\nu_{\text{eff}}} \approx 3.044 \) (see e.g. Refs. [185–191]). While we adopt the older standard value \( N_{\text{eff}} = 3.046 \), we note that adopting the latest value would have virtually no impact on our results, given the precision of current cosmological data, while even the precision of near-future cosmological data will still be more than one order of magnitude worse than \( \Delta N_{\text{eff}} \sim 0.002 \) [192–194].
B. Broken power-law spectrum

To consider a more realistic SGWB spectrum, we go beyond the simple assumption of a pure power-law across various decades in frequency discussed previously. In particular, we consider a phenomenological broken power-law scenario (which however we shall later justify on a theoretical basis) effectively corresponding to the SGWB power-law index changing from \( n_T \) to \( \alpha \) at a characteristic scale \( k_\alpha \) \[168\]. More specifically, the dimensionless GW density parameter \( \Omega_{GW}^{bpl} \) (with “bpl” standing for “broken power-law”) is related to the previously introduced \( \Omega_{GW}^{pl} \) [itself associated to the pure power-law primordial tensor spectrum given in Eq. (3)] as follows:

\[
\Omega_{GW}^{bpl} = \begin{cases} 
\Omega_{GW}^{pl} & (k < k_\alpha) \\
\Omega_{GW}^{pl} \left( \frac{k}{k_\alpha} \right)^\alpha & (k > k_\alpha) 
\end{cases} 
\]  

(10)

where the scale \( k_\alpha \) scale corresponds to a break frequency \( f_\alpha \sim 1.5 \times 10^{-15} (k_\alpha/\text{Mpc}^{-1}) \) Hz. We envisage the break frequency to be \( \mathcal{O}(\text{nHz}) \lesssim f_\alpha \lesssim \mathcal{O}(\text{Hz}) \), so that the NANOGrav signal falls within the first half of the spectrum, and constraints from LIGO/Virgo are relevant to the post-break spectrum. On the other hand, the break can allow us to fit the NANOGrav signal while not running afoul of constraints on the SGWB amplitude on \( \mathcal{O}(10) \) Hz frequencies from LIGO/Virgo, provided \( \alpha < 0 \) (or at the very least \( \alpha < n_T \)), so that the post-break spectrum is less strongly blue, if not red. In our analysis, we shall require \( \alpha < 0 \), so that the high-frequency part of the SGWB spectrum is red, as this is the only way a possible SGWB detection from NANOGrav can be reconciled with LIGO/Virgo’s non-detection.

While the choice of introducing the broken power-law parametrization in Eq. (10) is phenomenological in nature, we shall now argue that it is theoretically justified, as it can actually cover a very broad range of early Universe scenarios (as discussed in more detail in Ref. \[168\]). It is important to note that, in Eq. (10) we have introduced the broken power-law parametrization at the level of dimensionless GW density today. This can be interpreted as arising from at least two fundamentally different classes of scenarios: a first one where it is the transfer function which is modified (for instance due to the propagation of primordial GWs through non-standard epochs of the expansion of the Universe) leading to the broken power-law, and a second one where it is instead the primordial tensor power spectrum which features a broken power-law, where the break could be due to processes inherent to the GW production mechanism. At the level of \( \Omega_{GW}^{bpl} \), these two distinct theoretical origins are completely degenerate. While we shall discuss in detail various theoretical scenarios leading to a break in the GW spectrum, and falling within either of the two categories described above, in our later analysis we shall remain completely agnostic as to the underlying theoretical origin of the break, while simply adopting the phenomenological view that Eq. (10) can cover a broad range of interesting early Universe scenarios.

The broken power-law spectrum is best suited to describe models with a non-standard background evolution, \textit{i.e.} where the Universe does not transition from inflation to radiation domination following an instantaneous reheating process \[164, 168, 205\]. For example, an extended non-instantaneous reheating period, possibly with a low reheating scale, leads to an extended early matter domination before the usual radiation domination: this scenario leads to a break in the GW density which is captured by setting \( \alpha = n_T - 2 \), with \( k_\alpha \) determined by the reheating temperature. \textsuperscript{7}

More generally, if the Universe is dominated by a fluid with effective equation of state \( w_{\text{eff}} \) prior to radiation domination, the effect is also that of a break in the GW power spectrum, captured by setting \( \alpha = n_T + 2(3w_{\text{eff}} - 1)/(1 + 3w_{\text{eff}}) \), with \( f_\alpha \) related to the temperature at which the Universe switches to being radiation dominated. In particular, this expression recovers the case of an extended matter domination era, where \( w_{\text{eff}} = 0 \) and therefore \( \alpha = n_T - 2 \). Another theoretically interesting case is that where the Universe undergoes a kination or stiff matter domination phase, where \( w_{\text{eff}} = 1 \) and therefore \( \alpha = n_T + 1 \) (see e.g. Ref. \[206\]). However, this case will not be of interest to our work, as it makes the GW spectrum even bluer on small scales, thereby worsening the disagreement with LIGO/Virgo. The broken power-law spectrum can also describe the effect of late-time entropy injection (for instance through the decay of an additional scalar field other than the inflaton), although in this case the exact value of \( \alpha \) depends on the specific entropy injection details \[168\].

The scenarios described above effectively lead to a change in the transfer function appearing in Eq. (3), which generally takes a factorizable form with the various factors accounting for different physical processes (e.g. the standard radiation-to-matter transition, or the non-standard scenarios described above). However, as alluded to earlier, the broken power-law spectrum can also effectively describe scenarios where it is the underlying primordial power spectrum which features a break in the spectral index \( n_T \), which is therefore non-uniform across the frequency range. In fact, many early Universe models predicting a blue spectrum on CMB scales typically predict a break in the spectral index on smaller scales. One class of examples are inflationary models where the inflaton is coupled to the field-strength tensor of vector (gauge) fields (see e.g. Refs. \[207–211\]). In these models, production of gauge fields is typically effective over an extended period during inflation, leading to a blue spectrum for modes which exit the horizon during this period (see e.g. \[55–57\]). However, when gauge field production eventually stops being efficient, the associated spectrum returns to being red, decreasing at higher frequencies.

\textsuperscript{7} In this case the exact relation between the break wavenumber and the reheating temperature is given by Eq. (14) in Ref. \[168\].
III. CONSTRAINTS ON GRAVITATIONAL WAVES ACROSS THE FREQUENCY SPECTRUM

The landscape of current and future probes of GWs is extremely vast, diverse, and complementary, allowing to probe the signatures of GWs across a wide range of times and frequencies. Some of these probes (e.g. interferometers) are sensitive to the SGWB in a certain relatively narrow frequency band, whereas other probes carry an integral sensitivity to the SGWB energy density in a wide frequency range (e.g. BBN). Below we briefly present the GW probes we will make use of in this work.

A. Interferometers

Large laser interferometers can be used for GW direct detection, through the effect of a passing GW shortening or lengthening the two arms of the interferometers, in turn affecting the resulting interferometric patterns. Two state-of-the-art current GW interferometers are the Laser Interferometer Gravitational-Wave Observatory (LIGO) and the Virgo interferometer [222, 223], both of which are sensitive to GWs in the $10^{-4}$ Hz frequency range. To date, the LIGO and Virgo collaborations have detected $\sim 90$ resolved GW events resulting from the mergers of stellar objects (black holes and neutron stars) [224], and in doing so opened a remarkable window onto the Universe and revolutionized our understanding of gravity, in part thanks to coincidental multi-messenger observations across the electromagnetic spectrum [225–239].

Besides resolved events, interferometers can be used to search for a possible astrophysical or cosmological SGWB, such as the one we are considering in this work. LIGO and Virgo have placed an upper limit on the amplitude of the SGWB in the frequency range $20 \lesssim f / \text{Hz} \lesssim 86$. Following other works, we shall take the following 95% confidence level (C.L.) upper limit [165, 166]:

$$\Omega_{GW}(k_{LV}) \lesssim 1.7 \times 10^{-7} ,$$

where $k_{LV} \sim 2.3 \times 10^{16} \text{Mpc}^{-1}$ is the comoving wavenumber corresponding to a frequency $f_{LV} \sim 35 \text{Hz}$ which we take as representative for LIGO/Virgo’s limit (following earlier work), and “LV” stands for “LIGO/Virgo”. More recent constraints place an even tighter limit, $\Omega_{GW}(k_{LV}) \lesssim 6.6 \times 10^{-9}$, at a frequency of $f \sim 25 \text{Hz}$ [240].

We note that the limit in Eq. (11) is sensitive to the SGWB energy density at a specific wavenumber. In particular, the frequency $f_{LV} \sim 35 \text{Hz}$ falls within the post-break part of the broken power-law GW spectrum of

$$\frac{\Omega_{GW}(k_{LV})}{\Omega_{GW}(k_{LV})} = f_{LV}^{-\alpha} \ ,$$

where $f_{LV} \sim 2.3 \times 10^{16} \text{Mpc}^{-1}$ is the comoving wavenumber corresponding to a frequency $f_{LV} \sim 35 \text{Hz}$ which we take as representative for LIGO/Virgo’s limit (following earlier work), and “LV” stands for “LIGO/Virgo”. More recent constraints place an even tighter limit, $\Omega_{GW}(k_{LV}) \lesssim 6.6 \times 10^{-9}$, at a frequency of $f \sim 25 \text{Hz}$ [240].

We note that the limit in Eq. (11) is sensitive to the SGWB energy density at a specific wavenumber. In particular, the frequency $f_{LV} \sim 35 \text{Hz}$ falls within the post-break part of the broken power-law GW spectrum of

8 Strictly speaking, this limit assumes a fiducial value for $H_0$ (which should scale in such a way that $\Omega_{GW} \propto H_0^{-2}$ when scanning over $H_0$ in our subsequent analysis). However, we do not expect this point to have a significant impact on our results.

9 The data products of LIGO/Virgo O3 are available at: https://dcc.ligo.org/LIGO-G2001287/public
Eq. (10). Therefore only the part of the spectrum for \( k > k_\alpha \), or equivalently \( f > f_\alpha \), needs to be considered in order for theoretical predictions to be compared to LIGO/Virgo’s upper limit in Eq. (11). Hence, we expect LIGO/Virgo constraints on the SGWB amplitude to mostly constrain the break frequency \( f_\alpha \) and post-break spectral index \( \alpha \). For purely illustrative purposes, in Fig. 1 we show a benchmark example of the broken power-law SGWB spectrum, indicating the NANOGrav signal and LIGO/Virgo’s upper limit. As can be seen, a break in the GW spectrum is clearly required to reconcile a pre-break blue spectrum explaining NANOGrav with LIGO/Virgo’s upper limit.

Looking to the near future, prospects for direct detection of resolved GW events and the SGWB with interferometers and other types of surveys are very bright. The frequency window between \( \mathcal{O}(10^{-7}) \) and \( \mathcal{O}(10^{3}) \) Hz will be covered by a diverse range of experiments, including but not limited to space-based laser interferometers, next-generation ground-based detectors, binary resonance probes, and space-based atom interferometry. In Sec. VI we will discuss in more detail future probes of GWs (including the broken power-law SGWB spectrum we are considering) across the GW frequency landscape.

### B. Pulsar Timing Arrays

Pulsar timing arrays (PTAs) aim for a SGWB detection by exploiting the fact that millisecond pulsars behave as extremely stable clocks. A SGWB travelling between an ensemble of pulsars and us would leave its imprint through fluctuations in the arrival times of radio pulses, which would be spatially correlated \([11–13]\). By searching for these correlations, PTAs can search for a SGWB in the \( 10^{-9} - 10^{-7} \) Hz frequency range. Current and planned PTA surveys include NANOGrav \([95]\), PPTA \([241]\), and EPTA \([242]\), collectively constituting the IPTA \([243]\). Moreover, the Square Kilometer Array is expected to potentially be able to detect thousands of millisecond pulsars, and thus might play an important role in the landscape of future PTA surveys \([244]\).

It is customary to report the results of PTA searches in terms of the GW strain power spectrum as a function of frequency, \( h_c(f) \), which is related to \( \Omega_{GW}(f) \) by:

\[
\Omega_{GW}(f) = \frac{2\pi^2}{3H_0^2} f^2 h_c^2(f). \tag{12}
\]

The GW strain power spectrum is typically approximated as a power-law at a reference frequency \( f_{\text{yr}} = 1 \text{ yr}^{-1} \) (a numerically convenient value given the frequencies to which PTAs are most sensitive), with amplitude and spectral index given by \( A_{CP} \) and \( \alpha_{CP} \) respectively:

\[
h_c(f) = A_{CP} \left( \frac{f}{f_{\text{yr}}} \right)^{\alpha_{CP}} \equiv A_{CP} \left( \frac{f}{f_{\text{yr}}} \right)^{3 - \gamma_{CP}}, \tag{13}
\]

where the spectral index \( \alpha_{CP} \) is often exchanged for the related quantity \( \gamma_{CP} = 3 - 2\alpha_{CP} \). The SGWB resulting from merging SMBHBs is expected to be described by \( \gamma_{CP} = 13/3 \) or, equivalently, \( \alpha_{CP} = -2/3 \) \([245–247]\).

In this work we shall consider the NANOGrav signal \([95]\), which with all the caveats discussed earlier...
we shall interpret as a genuine SGWB detection. The NANOGrav collaboration fitted the power-law approximation of the strain power spectrum in Eq. (13) to their 5 lowest frequency bins with highest signal-to-noise, in the frequency range $2.5 \times 10^{-9} \lesssim f / \text{Hz} \lesssim 9.0 \times 10^{-8}$, and obtained joint constraint on $\log_{10} A_{\text{CP}}$ and $\gamma_{\text{CP}}$. We note that, for a SGWB spectrum behaving as a pure power-law at least up to the NANOGrav frequency range, such as the one we are considering since the break occurs for $f_\alpha \gtrsim O(\text{nHz})$, the tensor spectral index $n_T$ is related to $\gamma_{\text{CP}}$ and/or $\alpha_{\text{CP}}$ as $n_T = 5 - \gamma_{\text{CP}} = 2 + 2\alpha_{\text{CP}}$, whereas $A_{\text{CP}}$ scales as $A_{\text{CP}} \propto \sqrt{T}$ [163, 178].

**C. Big Bang Nucleosynthesis**

Big Bang Nucleosynthesis (BBN), taking place in the very earliest stages of our universe, is the process responsible for the production of light nuclei other than those of $^1\text{H}$ (see e.g. Ref. [248] for a review). BBN is primarily responsible for the production of $^4\text{He}$, alongside smaller amounts of $^3\text{He}$, deuterium ($^2\text{H}$) and tritium ($^3\text{H}$), $^7\text{Li}$, and $^7\text{Be}$, with $^3\text{H}$ and $^7\text{Be}$ later decaying to $^4\text{He}$ and $^7\text{Li}$ respectively. The final yield of light elements, tightly constrained observationally, is highly sensitive to the expansion rate of the Universe, usually assumed to proceed as in a radiation-dominated Universe following reheat- ing in the standard scenario. Therefore, BBN will be highly sensitive to a non-standard SGWB spectrum such as the one considered, given that blue primordial GWs contribute to the energy density of the Universe as an extra radiation component.

Unlike the previously discussed constraints from interferometers (LIGO/Virgo) and PTAs (NANOGrav), which are mostly sensitive to the SGWB spectral density at a certain frequency ($f_{\text{UV}}$ and $f_{\text{IR}}$ respectively) or within a narrow frequency band, BBN carries integrated sensitivity to a wide range of frequencies from $f_{\text{IR,BBN}}$ to $f_{\text{UV}}$, as captured by an effective number of relativistic species $N_{\text{eff,BBN}}$ defined analogously to Eq. (7) starting from Eq. (4). In this case, the relevant IR cutoff is given by $f_{\text{IR,BBN}} \sim 10^{-10} \text{Hz}$, approximately corresponding to the comoving horizon at the time of BBN, when the temperature of the Universe was $T \sim O(\text{MeV})$. For the broken power-law model, it is relatively straightforward to show that $N_{\text{eff,BBN}}$ is approximately given by:

$$N_{\text{eff,BBN}} \sim 0.3 \frac{rA_8}{n_T} \left[ \left( \frac{f_{\text{IR,BBN}}}{f_\alpha} \right)^{n_T} f_\alpha \right] + 0.3 \frac{rA_8}{\alpha} \left[ \left( \frac{f_{\text{UV}}}{f_\alpha} \right)^{\alpha} f_\alpha \right],$$

with pivot scale given by $k_\ast = 0.01 \text{ Mpc}^{-1}$ (or equivalently $f_\alpha = 1.5 \times 10^{-17} \text{Hz}$), and as stated earlier we fix $f_{\text{IR,BBN}} = 10^{-10} \text{Hz}$ and $f_{\text{UV}} = 10^8 \text{Hz}$ [182, 183]. Inserting numbers, the above expression becomes:

$$N_{\text{eff,BBN}} \sim 0.3 \frac{rA_8}{n_T} \left[ \left( \frac{2}{3} 10^{17} \frac{f_\alpha}{\text{Hz}} \right)^{n_T} - \left( \frac{2}{3} 10^{7} \right)^{n_T} \right] + 0.3 \frac{rA_8}{\alpha} \left[ \left( 10^{-8} \frac{f_\alpha}{\text{Hz}} \right)^{\alpha} - 1 \right],$$

where $|\alpha| = -\alpha$, since we require $\alpha < 0$ in order for the high-frequency part of the SGWB spectrum to be red and satisfy LIGO/Virgo limits, while simultaneously explaining the NANOGrav signal with the blue low-frequency part of the SGWB spectrum.

Let us examine more closely the two contributions to $N_{\text{eff,BBN}}$ in Eq. (15), with the first line corresponding to the low-frequency part of the spectrum ($f < f_\alpha$), and the second line corresponding to the high-frequency one ($f > f_\alpha$). Unsurprisingly, we see that the contribution to $N_{\text{eff,BBN}}$ from the low-frequency part of the broken power-law spectrum is dominated by the highest available frequencies (i.e. those around the break $f_\alpha$), and similarly the contribution to $N_{\text{eff,BBN}}$ from the high-frequency part is dominated by the lowest available frequencies (i.e. again those around the break $f_\alpha$). As anticipated earlier in Sec. II B, this occurs as we are within the regime where $n_T > 0$ and $\alpha < 0$, i.e. where the low-/high-frequency part of the SGWB spectrum is respectively blue/red. Therefore, the blue (low-frequency) part of the spectrum will be most sensitive to the highest available frequencies, and conversely for the red (high-frequency) part, implying that $N_{\text{eff,BBN}}$ ends up being dominated by modes around $f_\alpha$.

Neglecting sub-dominant terms, we find that Eq. (15) can be well approximated by:

$$N_{\text{eff,BBN}} \approx \frac{0.3 rA_8}{n_T} \left[ \left( \frac{2}{3} 10^{17} \frac{f_\alpha}{\text{Hz}} \right)^{n_T} - \frac{n_T}{\alpha} \right].$$

Observational determinations of the abundances of light elements severely restrict $N_{\text{eff,BBN}}$. Different probes return different limits, but we can safely take $N_{\text{eff,BBN}} \lesssim 0.4$ as an indicative upper limit [249–253]. Moreover, we find that for $f_\alpha$ within the range $10^{-9} \text{Hz} \lesssim f_\alpha \lesssim 35 \text{Hz}$, i.e. within the NANOGrav and LIGO/Virgo frequencies, the $-n_T/\alpha$ term is always several orders of magnitude smaller than the $f_\alpha$-dependent term, and hence can be safely neglected. In any case we stress that, within our

10 To see this, one can estimate the last term on the right-hand side of Eq. (16) with the aid of Eq. (10). A lower limit to $|\alpha|$ can be obtained by considering the lowest possible value of $f_\alpha$, corresponding to the NANOGrav frequency $f \sim 3 \times 10^{-8} \text{Hz}$, and saturating the inequality set by the LIGO/Virgo limit. Setting the left-hand side of Eq. (10) to the LIGO/Virgo upper limit ($\Omega_{\text{GW}} \sim 1.7 \times 10^{-7}$), relating this to the SGWB energy density $\rho_{\text{GW}}$ indicated by NANOGrav ($\Omega_{\text{GW}} \sim 5.6 \times 10^{-9}$), and using the fact that $k/k_\alpha = f_{\text{IR}}/f_\alpha = 35/(3 \times 10^{-8})$, we find the rough lower limit $|\alpha| \gtrsim 0.15$. It is trivial to show that for any value
FIG. 2. Evolution of \( N_{\text{eff,BBN}}^{\text{GW}} \) as a function of \( n_T \), \( r \), and \( f_\alpha \), as captured by Eq. (16), neglecting the \(-1/\alpha\) term which is subdominant as argued in the main text. Left panel: curves of constant \( N_{\text{eff,BBN}}^{\text{GW}} \) in the \( \log_{10} r-n_T \) plane, saturating the \( N_{\text{eff,BBN}}^{\text{GW}} \lesssim 0.4 \) limit, indicatively corresponding to the upper limit on the amount of extra radiation set by BBN, for different values of the break frequency \( f_\alpha \) (as per the color coding). Right panel: evolution of \( N_{\text{eff,BBN}}^{\text{GW}} \) as a function of \( n_T \) for different values of the tensor-to-scalar ratio \( r \) and break frequency \( f_\alpha \) (as per the color coding).

of the tensor spectral index \( 0.7 \lesssim n_T \lesssim 1.3 \) (required to explain NANOGrav, as shown in Ref. [163]), the last term on the right-hand side of Eq. (16) is always at least 17 orders of magnitude smaller than the first term.

D. Cosmic Microwave Background

The Cosmic Microwave Background (CMB), the relic radiation from the epoch of recombination, is an extremely important probe of early-Universe physics. For what concerns our study, we will be interested in the CMB at two different levels. Firstly, the CMB probes GWs at extremely low frequencies (\( f \lesssim 10^{-16} \) Hz), through their imprint on the B-mode polarization pattern: therefore, measurements of the B-mode power spectrum, and in particular limits on the tensor-to-scalar ratio \( r \) (e.g. Ref. [251]), can directly constrain the low-frequency part of our SGWB spectrum.

Secondly, the CMB is highly sensitive to the energy density of any extra radiation component, particularly through the effect of the latter on the damping of tail of the small-scale (large multipole \( \ell \) ) power spectra (Silk damping) [254] and on the early integrated Sachs-Wolfe (eISW) effect, both of which are tightly constrained by Planck CMB data [255–258]. As with BBN, the CMB is sensitive to a large range of modes, from \( f_{\text{IR,CMB}} \) to \( f_{\text{UV}} \). Therefore, the expression for \( N_{\text{eff,CMB}}^{\text{GW}} \) ends up being analogous to the BBN one, given in Eq. (14), except for the lower integration limit in the first line (low-frequency contribution) being given by \( f_{\text{IR,CMB}} \) instead of \( f_{\text{IR,BBN}} \).

In principle not only is \( f_{\text{IR,CMB}} \) much lower than \( f_{\text{IR,BBN}} \sim 10^{-10} \) Hz, but it is a time-dependent quantity, as at any given time only sub-horizon modes oscillate and propagate as extra radiation components. However, to simplify the discussion, we can again resort to the fact that the low-frequency part of the SGWB spectrum is blue. This leads to the expectation that the most important low-frequency contributions to \( N_{\text{eff,CMB}}^{\text{GW}} \) come from modes around the break frequency \( f_\alpha \), and...
similarly for the high-frequency contributions, given that the high-frequency part of the SGWB is red. In practice, this implies that $N_{\text{eff, CMB}}^{\text{GW}}$ is mostly insensitive to $f_{\text{IR, CMB}}$ and $f_{\nu, UV}$, given that $f_{\text{IR, CMB}} \ll 10^{-10}\,\text{Hz} \ll f_\alpha$ and $f_{\nu, UV} \sim 10^{8}\,\text{Hz} \gg f_\alpha$. Formally, this means that we can safely set $f_{\text{IR, CMB}} \sim 0\,\text{Hz}$, and that the expression for $N_{\text{eff, CMB}}^{\text{GW}}$ (relevant for the computation of the CMB temperature and E-mode polarization anisotropy power spectra) is identical to that for $N_{\text{eff, BBN}}^{\text{GW}}$ in Eq. (14).

IV. DATASETS AND METHODOLOGY

We now place observational constraints on the broken power-law SGWB spectrum we consider, with energy density spectrum given by Eq. (10), to examine whether it is possible to explain the NANOGrav signal while remaining consistent with a wide range of observational datasets, including those discussed in Sec. III. In our baseline analysis, we only consider CMB measurements alongside the constraints on the SGWB spectrum mentioned in Sec. III and the NANOGrav signal, which we treat as a genuine SGWB detection (with all the caveats discussed in Sec. I). In a later extended analysis, we further consider late-time cosmological measurements, to test whether they significantly improve cosmological parameter constraints, finding the answer to be negative.

In our most general analysis, we consider a ten-parameter model extending the standard six-parameter $\Lambda$CDM model by allowing four additional parameters to vary: the tensor-to-scalar ratio $r$, the pre-break tensor spectral index $n_T$ (not satisfying the inflationary consistency relation $r = -8n_T$, and hence independent of $r$), the break frequency $f_\alpha$, and the pre-break tensor spectral index $\alpha$. Note that we consider purely adiabatic initial conditions, while restricting our analysis to a spatially flat Universe, and fixing the sum of the neutrino masses to 0.06 eV, the minimum value allowed within the normal ordering, as the neutrino mass is known to correlate very weakly with inflationary parameters [259–261]. Finally, we set the scalar and tensor pivot scales to 0.05 Mpc$^{-1}$ and 0.01 Mpc$^{-1}$ respectively.

Theoretical predictions for cosmological observables in the presence of the broken power-law SGWB spectrum are obtained through a modified version of the Boltzmann solver CAMB [262]. We add a module to CAMB which calculates the SGWB energy density as a function of wavenumber/frequency for a given choice of cosmological parameters, and the quantities $N_{\text{eff, BBN}}^{\text{GW}}$ and $N_{\text{eff, CMB}}^{\text{GW}}$ (which as argued earlier are approximately identical). Moreover, this module computes the amplitude of the SGWB strain power spectrum $A_{\nu}$ at the NANOGrav reference frequency $f_{\nu}$, as parametrized by Eqs. (12,13), so that we can later treat it as a derived parameter.

We perform a Bayesian statistical analysis to constrain the model in question. We set flat priors on all cosmological parameters, except for $r$ and $f_\alpha$, with the prior ranges chosen to be wide enough as to not cut the associated marginalized posterior distributions where these are appreciably non-zero. For $r$, we instead set a prior flat in $\log_{10} r$. This choice makes the exploration of the parameter space much more efficient, particularly since explaining the NANOGrav signal while complying with upper limits from Planck and BICEP2/Keck Array requires a small but non-zero $r$, as shown by one of us in Ref. [163]. These considerations suggest that a prior flat in $\log_{10} r$ may more correctly encapsulate our prior knowledge than a prior flat in $r$ does. Similarly, as $f_\alpha$ can span a wide range encompassing several orders of magnitude, we set a prior flat in $\log_{10} f_\alpha$, which better captures our prior information on the break frequency. \footnote{More precisely, we actually set a prior flat in $\log_{10} k_\alpha$, with $k_\alpha$ the break wavenumber. However, given the proportionality relation between $k_\alpha$ and $f_\alpha$, a prior flat in $\log_{10} k_\alpha$ translates to a prior flat in $\log_{10} f_\alpha$.}

The posterior distributions for the cosmological parameters are sampled by means of Markov Chain Monte Carlo (MCMC) methods. We generate MCMC chains by using the cosmological MCMC sampler CosmoMC [263], modified to interface itself with the modified version of CAMB discussed previously, and to include the additional observations (including NANOGrav) we further discuss below. The convergence of the generated MCMC chains is monitored via the Gelman-Rubin parameter $R - 1$ [264], and we set the requirement $R - 1 < 0.01$ in order for the chains to be considered converged.

As far as CMB data is concerned, we consider measurements of CMB temperature anisotropy and polarization power spectra, and their cross-spectra, from the Planck 2018 legacy data release [251]. We combine the high-$\ell$ $P_{lTT}$ likelihood for TT within the multipole range $30 \leq \ell \leq 2500$ with the TE and EE and low-$\ell$ TT-only likelihoods based on the Commander component-separation algorithm in pixel space within the multipole ranges $30 \leq \ell \leq 2000$ and $2 \leq \ell < 29$ respectively, and the low-$\ell$ EE-only SimAll likelihood in the multipole range $2 \leq \ell < 29$ [265]. We also make use of the likelihood for the Planck CMB lensing power spectrum reconstructed from the temperature 4-point function [266]. Finally, we include the BICEP2/Keck Array B-mode polarization likelihood, which strongly constrains the amplitude of the tensor-to-scalar ratio at CMB scales [33, 267, 268] ($r_{0.05} < 0.06$ at 95% C.L. when combined with Planck). Note that the high-$\ell$ part of the CMB TT and EE likelihoods indirectly constrains $N_{\text{eff, CMB}}^{\text{GW}}$ through the effect of extra radiation on the damping tail and the eISW effect, as discussed in Sec. III.D.

Following earlier work [60], we treat the LIGO/Virgo upper limit on $\Omega_{\text{GW}}(f_{1})$ reported in Eq. (11) as a half-Gaussian prior on the amplitude of the SGWB energy density at the frequency in question, which is thus treated as a derived parameter. We choose to use this constraint and not the most recent ones [240] to obtain the most conservative results possible.
We instead treat the BBN limit as a hard upper limit on $N^{GW}_{eff, BBN} < 0.4$. However, we stress that this information is essentially redundant, given that within the assumed SGWB spectrum $N^{GW}_{eff, CMB} \approx N^{GW}_{eff, BBN}$, and the adopted CMB likelihoods already set a tighter limit on $N^{GW}_{eff, CMB}$, through the effect of the latter on the damping tail and the eISW effect.

For what concerns the NANOGrav signal, we approximate this as a multivariate Gaussian on the two derived parameters $\log_{10} A_{CP}$ and $\gamma_{CP}$, defined in Eqs. (12,13), and evaluated at the reference frequency $f_{r}$ relevant for NANOGrav. Specifically, if we denote by $x(\theta)$ the vector of derived parameters $x = \{\log_{10} A_{CP}(\theta), \gamma_{CP}(\theta)\}$, where $\theta$ is the 10-dimensional vector of cosmological parameters, we approximate the NANOGrav log-likelihood as being given by:

$$\ln L_{\text{NANOGrav}}(\theta) = -\frac{\chi^{2}_{\text{NANOGrav}}(\theta)}{2},$$

where $\chi^{2}_{\text{NANOGrav}}$ is given by:

$$\chi^{2}_{\text{NANOGrav}} = (x(\theta) - \mu)^{T} \Sigma^{-1} (x(\theta) - \mu).$$

In Eq. (18), $\mu \approx (-15.2, 5.3)$ is the vector of mean values for the derived parameters $\log_{10} A_{CP}$ and $\gamma_{CP}$, and $\Sigma$ is the covariance matrix (with $T$ denoting the transpose operation), which we numerically determine to be $\Sigma \approx \left(\begin{array}{cc} 0.1 & -0.182 \\ -0.182 & 0.4 \end{array}\right)$. To test the goodness of the multivariate Gaussian likelihood approximation in Eqs. (17,18), with the approximate values of $\mu$ and $\Sigma$ we determined, we draw 10000 random samples from the probability distribution function defined by $L(x)$, and verify that the associated 68% and 95% C.L. constraints on $\log_{10} A_{CP}$ and $\gamma_{CP}$ fall to very good approximation within the corresponding 1$\sigma$ and 2$\sigma$ contours given in the right panel of Fig. 1 of Ref. [95]. While this is of course an approximation, for the purposes of our analysis it provides a sufficiently precise compression of the information content of the tentative NANOGrav detection. Finally, while in Eqs. (17,18) we have retained the dependence on the full set of cosmological parameters $\theta$, we stress that to very good approximation the NANOGrav likelihood mostly depends on the tensor-to-scalar ratio $r$ (with a minor dependence on $A_{s}$ as well as on $\Omega_{m}$ and $H_{0}$, themselves depending on the fundamental parameters $\omega_{b}$, $\omega_{c}$, and $\delta_{s}$, all of which are very well constrained by the CMB), and the tensor spectral index $n_{T}$, with $\gamma_{CP}$ depending only on $n_{T}$ since $\gamma_{CP} = 5 - n_{T}$ as discussed earlier. In any case, we numerically determine $\log_{10} A_{CP}$ and $\gamma_{CP}$ at each step of our MCMC analysis.

Our baseline analysis combines CMB data from Planck and BICEP2/Keck Array, constraints on the SGWB energy density from LIGO/Virgo and BBN, and the tentative NANOGrav detection. We refer to this combination of likelihoods as “base”.

At a later stage, we combine the base dataset with additional late-time distance and expansion rate measurements. In particular, we consider Baryon Acoustic Oscillation (BAO) measurements from the 6dFGS [269], SDSS DR7 MGS [270], and BOSS DR12 samples [271], distance moduli measurements from the Pantheon Type Ia Supernovae (SNeIa) compilation [273], and measurements of the expansion rate $H(z)$ from cosmic chronometers (CC), using the relative ages of massive, early-time, passively evolving galaxies as first proposed in Ref. [274], and reported in Refs. [275–282] (see Tab. 1 in Ref. [283]). We refer to the combination of BAO, SNeIa, and CC measurements as “late”.

Our goal in combining the base and late dataset combinations is to check whether the latter can improve the determination of the SGWB parameters, in particular $\alpha$ and $f_{a}$. Since BAO, SNeIa, and CC measurements do not directly constrain the SGWB spectrum, one might legitimately wonder why to expect an improvement in first place. The reason is that these measurements help improving constraints on parameters which may be degenerate with the SGWB ones, and in particular the tensor-to-scalar ratio $r$. Moreover, as we stated earlier, $\log_{10} A_{CP}$ carries some amount of dependence (albeit very weak) on $\Omega_{m}$ and $H_{0}$. In any case, we find a posteriori that adding the late dataset combination to the base one does not improve the constraints on the SGWB parameters, implying that the aforementioned degeneracies play a very marginal role in our analysis.

| $f_{a}$, r | $f_{a}$ [Hz] | $r$ | $n_{T}$ |
|------------|-------------|-----|---------|
| Fixed      | $10^{-6}$/$10^{-3}$ | $10^{-6}$ | $1.31 \pm 0.03$ |
| Fixed      | $10^{-6}$/$10^{-3}$ | $10^{-3}$ | $0.98 \pm 0.03$ |
| Fixed      | $10^{-6}$/$10^{-3}$ | $0.06$ | $0.79 \pm 0.03$ |
| Free       | $< 10^{-0.7}$ | $> 10^{-0.6}$ | $0.97 \pm 0.19$ |

TABLE I. Constraints on the SGWB parameters using the base dataset combination. In the first three analyses $f_{a}$ and $r$ are fixed, while in the fourth analysis these two parameters are varied (while fixing $\alpha = -3$). The constraints obtained within the latter analysis are very stable against the further introduction of $\alpha$ as a free parameter. Quoted intervals correspond to 68% C.L. intervals, whereas quoted upper/lower limits correspond to 95% C.L. upper/lower limits.

---

12 Alternatively, we note that the $\log_{10} A_{CP}$ and $\gamma_{CP}$ posterior samples for the NANOGrav signal are available at: https://github.com/nanograv/12p5yr_stochastic_analysis

13 When this project was initiated, the likelihood for the completed lineage of SDSS experiments, including in particular the legacy eBOSS measurements [272], was not yet publicly available. In any case, we expect that the inclusion of the latter measurements would not quantitatively alter our results, given our a posteriori finding that the inclusion of late-time measurements returns cosmological constraints which are virtually identical to those obtained from the base dataset combination alone.

14 We have conservatively chosen to omit the measurements reported in Ref. [284], given the concerns on these measurements recently raised in Ref. [285].
In our analysis, we choose to proceed in two distinct steps, to better understand to what extent the data is actually sensitive to the beyond-ΛCDM parameters (in particular $f_\alpha$, $r$, and $\alpha$), and to aid the convergence of our MCMC chains. In a first instance, we fix $f_\alpha$ and $r$ to certain sets of values, and examine the resulting constraints on $n_T$ and $\alpha$ (as well as the other cosmological parameters). We find that the data is not strongly sensitive to $\alpha$, so in the second instance we fix this parameter, and vary all the other 9 cosmological parameters. However, we perform a final sanity check to ensure that varying $\alpha$ does not significantly affect our results, finding this to indeed be the case. We would also like to stress that, as we are taking into account rather negative values of $\alpha$, considering the more stringent constraints on the SGWB amplitude given by the more recent LIGO/VIRGO results [240] would not qualitatively affect our results.

V. RESULTS AND DISCUSSION

As stated above, we start by understanding to what extent the data is sensitive to the break frequency $f_\alpha$ and the tensor-to-scalar ratio $r$. We do so by fixing these parameters, and considering the base dataset combination. Specifically, we fix $f_\alpha$ to the values $10^{-3}$, $10^{-6}$ Hz, and $r$ to the values 0.06 (saturating constraints from Planck and BICEP2/Keck Array within a ΛCDM+$r$ model), $10^{-3}$, and $10^{-6}$. In particular, the break frequency is chosen to lie between the NANOGrav and LIGO/Virgo frequencies. The results of these tests are reported in Tab. I.

We find, as expected, that the smaller the (fixed) value of $r$, the larger the inferred value of $n_T$. This is not unexpected, and in line with the findings in Ref. [163]: if the SGWB amplitude on CMB frequencies is smaller, the tensor tilt will necessarily have to be larger in order for the SGWB amplitude to grow more quickly with increasing frequency and match the amplitude of the NANOGrav signal at the respective frequency. This behavior is clearly shown in Tab. I, where the first three rows report the inferred value of $n_T$ as a function of the (fixed) value of $r$, and we can see that the former decreases (increases) while the latter increases (decreases).

For what concerns the break frequency $f_\alpha$, we find that, as long as this parameter lies below a certain value, it has little effect on our results: in other words, within this range, the likelihood is roughly flat along the $f_\alpha$ direction. Larger frequencies, however, are unable to match the amplitude of the NANOGrav signal while respecting $N_{\text{eff}}$ constraints (again in line with what was found earlier in Ref. [163], focusing on the reheating temperature $T_{\text{rh}}$ which can be related to $f_\alpha$). For instance we find that $f_\alpha = 10 \text{Hz}$ requires a very low value of $A_s$ and is thus unable to explain the NANOGrav signal, therefore being disfavored by our analysis.

Here, we have allowed $\alpha$ to vary, requiring this parameter to be negative so that the high-frequency part of the broken SGWB spectrum is red. We find, as expected, that we can only set an upper limit on this parameter. This is not surprising since, leaving aside $N_{\text{GW}}^\text{eff}$ constraints for the moment, an arbitrarily large negative $\alpha$ can be invoked in order for the high-frequency part of the SGWB spectrum to be consistent with LIGO/Virgo’s upper limit. For example, for $r = 10^{-5}$ we find $\alpha < -2.8$ at 95% C.L., whereas this limit changes to $\alpha < -0.9$ and $\alpha < -0.7$ for $r = 10^{-3}$ and $r = 0.06$ respectively. It is worth mentioning that these results, alongside those reported in Tab. I, can easily be seen to satisfy the relation $\alpha = -2 + n_T$, expected from models with non-instantaneous reheating: this confirms that in light of our observational constraints, the broken power-law SGWB spectrum adopted can be interpreted in the context of these models [164, 168, 205].

In the second part of our analysis, we vary all cosmological parameters (including $f_\alpha$ and $r$), except for $\alpha$. The reason is that we expect the likelihood to be roughly flat along the $\alpha$ direction: in fact, any sufficiently negative $\alpha$ will fit the data, and more precisely the upper limit set by LIGO/Virgo, equally well. If we do not commit to any specific theory, there is in principle no well-motivated theoretical lower limit to the value $\alpha$ can take within our phenomenological parametrization of Eq. (10). Since we are mostly interested in constraints on the pre-break part of the SGWB spectrum, i.e. the one responsible for explaining the NANOGrav signal, we decide to fix $\alpha = -3$. This arbitrary value is sufficiently negative as to not impact our results: in other words, as per the results of the first part of our analysis, this choice guarantees that the broken power-law SGWB spectrum will be consistent with the LIGO/Virgo upper limit for any choice of parameters explaining NANOGrav.

We present a triangular plot of the constraints on the main parameters of interest in Fig. 3, including the derived parameter $N_{\text{GW}}^\text{eff}$, which quantifies the contribution of the SGWB spectrum to the radiation energy density. We find that this parameter is very strongly constrained to $N_{\text{GW}}^\text{eff} \lesssim 0.11 [0.28]$ at 68% [95%] C.L., given the impact of extra radiation on the damping tail and eISW effect, both tightly constrained by CMB data. Let us stress that this result strictly depends on the assumptions made, including the standard $\eta$-dependence in the transfer function. This implies that our results may not in general be applicable to models which, while being compatible with the broken power-law SGWB spectrum, have a different $\eta$-dependence in the transfer function (as these may provide different $N_{\text{GW}}^\text{eff}$ constraints). However, to understand whether this is the case requires committing to a specific non-standard model of inflation, whereas here we have taken a model-agnostic phenomenological stand.

Moreover, we obtain a detection of non-zero $r$, with the lower limit $r > 2.5 \times 10^{-7}$ at 95% C.L., consistent with earlier findings in Ref. [163], as smaller values of $\alpha$ would require too large a tilt, disfavored by the NANOGrav signal (and in particular the constraints on $\gamma_{\text{CP}}$). Our
FIG. 3. Triangular plot showing 2D joint and 1D marginalized posterior probability distributions for $\log_{10} r$, $n_T$, $\log_{10}(f_\alpha/\text{Hz})$, and $N_{\text{eff}}^{\text{GW}}$ from the base (red contours) and base+late (green contours) dataset combinations, with the former only including CMB and GW observations (including the tentative NANOGrav signal), and the latter also including late-time distance and expansion rate BAO, SNeIa, and CC measurements. The corresponding constraints have been obtained by varying all the cosmological parameters except for the post-break frequency $\alpha$, fixed to $\alpha = -3$. It is clear that the inclusion of late-time measurements has a negligible effect on the resulting constraints.

The results discussed so far were obtained using the base dataset combination. We include the late dataset combination to see whether this improves constraints on any of the cosmological parameters. We find that this mostly leads to improvements in the inferred values of non-SGWB derived parameters, such as $\Omega_m$ and $H_0$. For instance, the inferred 68% C.L. constraints on $H_0$ shift from $H_0 = (67.38 \pm 0.53) \text{ km s}^{-1} \text{ Mpc}^{-1}$ to $H_0 = (67.78 \pm 0.40) \text{ km s}^{-1} \text{ Mpc}^{-1}$. Therefore, unlike what was found by some of us using earlier data in Ref. [200], a blue GW spectrum is unable to solve or even alleviate the Hubble tension [46]. It is known that a larger value of $N_{\text{eff}}$ is in principle a viable way of alleviating the $H_0$ tension, as a larger radiation component reduces the sound horizon, allowing for a larger value of $H_0$ (see e.g. Refs. [286–300]). However, the updated datasets

analysis also confirms that we require an extremely blue spectrum for frequencies below the break, as we infer $n_T = 0.97 \pm 0.19$. In addition, we find an upper limit on the break frequency, with $f_\alpha < 0.2 \text{ Hz}$ at 95% C.L.: this is again as expected given the results of our previous simplified analysis with $f_\alpha$ and $r$ fixed, as larger values of $f_\alpha$ would run afoul of constraints on $N_{\text{eff}}^{\text{GW}}$.

The results discussed so far were obtained using the base dataset combination. We include the late dataset combination to see whether this improves constraints on any of the cosmological parameters. We find that this mostly leads to improvements in the inferred values of non-SGWB derived parameters, such as $\Omega_m$ and $H_0$. For instance, the inferred 68% C.L. constraints on $H_0$ shift from $H_0 = (67.38 \pm 0.53) \text{ km s}^{-1} \text{ Mpc}^{-1}$ to $H_0 = (67.78 \pm 0.40) \text{ km s}^{-1} \text{ Mpc}^{-1}$. Therefore, unlike what was found by some of us using earlier data in Ref. [200], a blue GW spectrum is unable to solve or even alleviate the Hubble tension [46]. It is known that a larger value of $N_{\text{eff}}$ is in principle a viable way of alleviating the $H_0$ tension, as a larger radiation component reduces the sound horizon, allowing for a larger value of $H_0$ (see e.g. Refs. [286–300]). However, the updated datasets

15 This is broadly consistent with the earlier findings of Ref. [163], which however did not perform a full Bayesian analysis but just a parameter scan.
we have used here severely constrain $N_{\text{eq}}^{\text{GW}}$ (primarily through its impact on Silk damping [254] and the eISW effect [255–258]), therefore strongly limiting the possibility of appreciably raising $H_0$ within this model. We also find that constraints on the SGWB parameters are virtually unaffected, only improving very slightly (and to an extent which is in principle compatible with a statistical fluctuation given the convergence level of the MCMC chains), implying that degeneracies with non-SGWB parameters whose determination is improved by including late-time datasets plays no major role in our analysis.

Finally, as a sanity check, we perform a run where we vary all 10 cosmological parameters including $\alpha$, for which we set a flat prior with a purely indicative lower prior limit of $-10$. We find that the constraints on the other 9 cosmological parameters are very stable against the introduction of $\alpha$ as a free parameters. Moreover, we infer a rather loose upper limit on $\alpha < -0.6$ at 95\% C.L., limit which was well satisfied in our previous analysis where we chose to fix $\alpha = -3$. This final analysis confirms that the likelihood is roughly flat in the $\alpha$ direction provided $\alpha$ is sufficiently negative, and that fixing this parameter to a sufficiently negative value does not have an appreciable effect on our results, which are mainly constraining the parameters governing the pre-break part of the broken power-law SGWB spectrum (parameters which we are also most interested in).

Overall, our results show that the broken power-law SGWB spectrum can be brought in agreement with a wide multi-frequency range of observations and constraints, ranging from CMB and BBN to interferometers, while at the same time potentially explaining the tentative NANOGrav detection. In particular, the break in the power spectrum, which can be justified by several well-motivated fundamental physics scenarios as discussed in Sec. II B, can significantly reduce the SGWB contribution to the radiation energy density in the early Universe as captured by $N_{\text{eq}}^{\text{GW}}$, which would otherwise be very large within the simple power-law model [163, 200], while allowing the SGWB spectrum to remain in agreement with constraints from LIGO/Virgo. All of this is possible for a wide range of break frequencies $f_a$ and post-break tilt $\alpha$, as long as the latter is negative so that the post-break SGWB spectrum is red, while being blue for lower frequencies (to explain the NANOGrav signal while remaining in agreement with upper limits on the tensor-to-scalar ratio on CMB scales).

VI. CONCLUSIONS AND PROSPECTS

NANOGrav’s possible SGWB detection [95] could constitute a significant milestone towards our understanding of the physics of the very early Universe. While the most natural candidate for a signal in the nHz range is an astrophysical SGWB arising from the merger of SMBHs, it is worth investigating whether NANOGrav’s signal could instead be cosmological in nature. One of the best motivated candidates in this sense is a SGWB generated during an early period of inflation. In this paper, we have re-examined the possibility that NANOGrav may have detected primordial GWs, arising from inflation or early-Universe alternatives to inflation. In particular, we have extended earlier works by going beyond the overly simplistic assumption of a pure power-law SGWB spectrum, considering constraints on the SGWB amplitude from a variety of sources, and performing a more complete analysis adopting various precision cosmological datasets.

We have considered a broken power-law SGWB spectrum [Eq. (10)], with the spectral index changing from $n_T$ to $\alpha$ above a characteristic break frequency $f_a$. While we have taken a phenomenological stance, without committing to any specific early-Universe model, we have argued (see Sec. II B) that such a phenomenological parametrization is actually flexible enough to cover a wide range of interesting early Universe models. These include scenarios altering the SGWB transfer function such as inflationary models with low-scale reheating and/or a non-standard background evolution following reheating, and late-time entropy injection, as well as scenarios altering the primordial SGWB power spectrum, such as models involving particle production, or alternatives to inflation.

This particular choice of SGWB spectrum allows us to explain the tentative NANOGrav detection in the $O(\text{nHz})$ range while complying with upper limits on the tensor-to-scalar ratio on CMB scales [$f \ll O(10^{-15}) \text{Hz}$], and on the SGWB amplitude on interferometer scales [$f \sim O(10) \text{Hz}$]. In particular, we require a spectrum which is blue for frequencies below the break, and red above, with the break frequency lying between the PTA and interferometer ranges. The break helps a) ensuring consistency with LIGO/Virgo’s upper limits (a requirement which had been missed earlier in Ref. [163]), and b) suppressing the SGWB contribution to the radiation energy density in the early Universe (as captured by the parameter $N_{\text{eq}}^{\text{GW}}$), which in turn is key in order not to run afoul of CMB and BBN constraints.

We constrain the broken power-law SGWB spectrum using a wide range of cosmological and astrophysical datasets, including CMB (Planck and BICEP2/Keck Array) and BBN, limits from LIGO/Virgo, the tentative NANOGrav signal which we interpret as a genuine detection, and late-time measurements of the expansion history. This dataset combination strongly constrains the GW contribution to the radiation energy density, with the 95\% C.L. upper limit $N_{\text{eq}}^{\text{GW}} \lesssim 0.11$. At the same time, we find that explaining the NANOGrav detection requires a very blue spectrum below the break, with $n_T = 0.97 \pm 0.19$, and of course a non-zero value for the tensor-to-scalar ratio, with the 95\% C.L. lower limit $r > 2.5 \times 10^{-7}$, perfectly in agreement with upper limits from the non-observation of a primordial B-mode polarization signal. We find that the model is able to satisfy BBN and LIGO/Virgo limits for a wide range of values for the post-break tilt $\alpha < 0$, while we find the 95\% C.L. upper limit on the break frequency $f_a < 0.2 \text{Hz}$. 
The constraints we obtained on the SGWB spectrum parameters are still rather loose, primarily due to the sparsity and still limited precision of the available datasets constraining the SGWB across 20 decades in frequency (although we recall that CMB and BBN carry integrated sensitivity to the SGWB spectrum throughout a wide frequency range). Several next-generation experiments will fill the gap between PTA and interferometer frequencies [301] as shown in Fig. 4. Here we plot three benchmark example SGWB spectra such as the one we considered (with three values of $n_T$ respectively consistent, marginally consistent, and inconsistent with the NANOGrav signal), alongside the expected noise curves of some of these upcoming experiments (with and without the contribution of astrophysical foregrounds, whose inclusion lowers the expected sensitivity level of these experiments), as obtained in the very detailed analysis of Ref. [301]. It is clear that the sensitivities of these experiments (featuring improvements of up to 2-3 orders of magnitude compared to existing ones) and their expanded band width will significantly aid our ability to constrain the SGWB spectrum considered, particularly for what concerns the break frequency and tensor tilt below and above the break.

Examples of upcoming surveys which will be extremely helpful towards further testing the phenomenological model we considered include but are not limited to the following (in brackets the frequency range these will probe): the SKA, which is expected to detect thousands of millisecond pulsars ($10^{-9}$-10$^{-7}$ Hz) [244], $\mu$Ares ($10^{-6}$-10$^{-2}$Hz) [302], LISA ($10^{-4}$-10$^{-1}$ Hz) [303], BBO [304] and DECIGO ($10^{-4}$-10Hz) [305], DO ($10^{-5}$-10 Hz) [306], AEDGE ($10^{-2}$-10$^3$ Hz) [17], as well as the Einstein Telescope ($10^{9}$-10$^{3}$ Hz) [307]. Binary resonance, for instance through laser ranging of the Moon, artificial satellites around the Earth, and timing of binary pulsars, will be able to probe a wide frequency range encompassing the $O(\mu Hz)$ gap between the most sensitive bands of PTA and interferometers [14, 15]. Future CMB missions such as the space-based LiteBIRD [308] and the ground-based CMB-S4 [192] and Simons Observatory [193, 194] will be able to probe values of the tensor-to-scalar ratio as low as $O(10^{-3})$ through the associated primordial B-mode polarization signal. The combination of all these probes will enable a precise characterization of the SGWB spectrum.
including BBN information will potentially extend the range to 21 decades in frequency,\footnote{including BBN information will potentially extend the range to 29 decades in frequency,\footnote{including BBN information will potentially extend the range to 29 decades in frequency,\footnote{including BBN information will potentially extend the range to 29 decades in frequency,\footnote{potentially allowing to distinguish between different theoretical origins for the phenomenological spectrum we have assumed, and in particular whether the latter is sourced by quantum vacuum fluctuations in the metric tensor, sources such as production of gauge fields coupled to the inflaton, or primordial scenarios alternative to inflation.}

While a detection of spatial quadrupolar correlations is required for the NANOGrav signal to be confirmed as a genuine SGWB detection, we believe there is nonetheless reason to be cautiously optimistic, particularly given PPTA and EPTA’s tentative detections in a similar frequency range [97, 98]. With these caveats in mind, our work reinforces the possibility (also discussed in previous works [156, 163, 164]) that the NANOGrav signal may be primordial in origin, associated to early-Universe scenarios such as inflation or alternatives thereto. Fitting a wide range of precision cosmological and astrophysical datasets within a phenomenological (but representative of several well-motivated theoretical scenarios) broken power-law SGWB spectrum [168], we have confirmed that this interpretation requires several deviations from the standard picture (e.g. instantaneous reheating and standard post-reheating background evolution), and reinforced the importance of multi-frequency GW observations towards better understanding the dynamics of the primordial Universe. While we await a definitive confirmation of NANOGrav’s signal, there are several interesting avenues for follow-up work, particularly related to connections to more specific primordial scenarios (inflationary or not) and observational implications thereof, such as primordial BHs, observable non-Gaussianity, or connections to the possible non-tensorial nature of the NANOGrav signal [99, 100, 102, 103]. We defer the investigation of these and related issues to future work.

ACKNOWLEDGMENTS

MB is supported by the Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Napoli, iniziativa specifica QGSKY. L.L.G is supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Grant No. 307052/2019-2, and by the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro–FAPERJ, Grant No. E-26/201.297/2021. S.V. is supported by the Isaac Newton Trust and the Kavli Foundation through a Newton-Kavli Fellowship, and by a grant from the Foundation Blancheflor Boncompagni Ludovisi, née Bililt. S.V. acknowledges a College Research Associateship at Homerton College, University of Cambridge. We acknowledge the use of the CAMB [262] and CosmoMC [263] codes. We acknowledge the use of computational facilities provided by the High Performance Data Center (DCON) of the Observatório Nacional.

\begin{thebibliography}{100}

\bibitem{Abbott2016} B. P. Abbott \textit{et al.} (LIGO Scientific, Virgo), \textit{Phys. Rev. Lett.} \textbf{116}, 061102 (2016), arXiv:1602.03837 [gr-qc].
\bibitem{Sathyaprakash2009} B. S. Sathyaprakash and B. F. Schutz, \textit{Living Rev. Rel.} \textbf{12}, 2 (2009), arXiv:0903.0338 [gr-qc].
\bibitem{Cai2016} R.-G. Cai, Z. Cao, Z.-K. Guo, S.-J. Wang, and T. Yang, \textit{Natl. Sci. Rev.} \textbf{4}, 687 (2017), arXiv:1703.00187 [gr-qc].
\bibitem{Bian2017} L. Bian \textit{et al.}, (2021), arXiv:2106.10235 [gr-qc].
\bibitem{Maggiore2000} M. Maggiore, \textit{Phys. Rept.} \textbf{331}, 283 (2000), arXiv:gr-qc/9909001.
\bibitem{Caprini2018} C. Caprini and D. G. Figueroa, \textit{Class. Quant. Grav.} \textbf{35}, 163001 (2018), arXiv:1801.04268 [astro-ph.CO].
\bibitem{Giovannini2020} M. Giovannini, \textit{Prog. Part. Nucl. Phys.} \textbf{112}, 103774 (2020), arXiv:1912.07065 [astro-ph.CO].
\bibitem{Kamionkowski2016} M. Kamionkowski and E. D. Kovetz, \textit{Ann. Rev. Astron. Astrophys.} \textbf{54}, 227 (2016), arXiv:1510.06042 [astro-ph.CO].
\bibitem{Ota2014} A. Ota, T. Takahashi, H. Tashiro, and M. Yamaguchi, \textit{JCAP} \textbf{10}, 029 (2014), arXiv:1406.0451 [astro-ph.CO].
\bibitem{Kite2021} T. Kite, A. Ravenni, S. P. Patil, and J. Chluba, \textit{Mon. Not. Roy. Astron. Soc.} \textbf{505}, 4396 (2021), arXiv:2010.00040 [astro-ph.CO].
\bibitem{Sazhin1978} M. V. Sazhin, \textit{Soviet Astr.} \textbf{22}, 36 (1978).
\bibitem{Detweiler1979} S. L. Detweiler, \textit{Astrophys. J.} \textbf{234}, 1100 (1979).
\bibitem{Foster1990} R. S. Foster and D. C. Backer, \textit{Astrophys. J.} \textbf{361}, 300 (1990).
\bibitem{Blas2021} D. Blas and A. C. Jenkins, (2021), arXiv:2107.04063 [gr-qc].
\bibitem{Bian2021} D. Blas and A. C. Jenkins, (2021), arXiv:2107.04601 [astro-ph.CO].
\bibitem{Graham2017} P. W. Graham, J. M. Hogan, M. A. Kasevich, S. Rajendran, and R. W. Romani (MAGIS), (2017), arXiv:1711.02225 [astro-ph.IM].
\bibitem{El-Neaj2020} Y. A. El-Neaj \textit{et al.} (AEDGE), \textit{EPJ Quant. Technol.} \textbf{7}, 6 (2020), arXiv:1908.00802 [gr-qc].
\bibitem{Ellis2020} J. Ellis and V. Vaskonen, \textit{Phys. Rev. D} \textbf{101}, 124013 (2020), arXiv:2003.13480 [gr-qc].
\bibitem{Abramovici1992} A. Abramovici \textit{et al.}, \textit{Science} \textbf{256}, 325 (1992).
\bibitem{Barausse2020} E. Barausse \textit{et al.}, \textit{Gen. Rel. Grav.} \textbf{52}, 81 (2020), arXiv:2001.09793 [gr-qc].
\bibitem{Yokoyama2021} J. Yokoyama, \textit{AAPPS Bull.} \textbf{31}, 17 (2021), arXiv:2105.07629 [gr-qc].
\bibitem{Sesana2004} A. Sesana, F. Haardt, P. Madau, and M. Volonteri, \textit{Astrophys. J.} \textbf{611}, 623 (2004), arXiv:astro-ph/0401543.
\bibitem{Blanco-Pillado2019} J. J. Blanco-Pillado, K. D. Olum, and X. Siemens, \textit{Phys. Lett. B} \textbf{778}, 392 (2019), arXiv:1709.02494 [astro-ph.CO].
\bibitem{Kazanas1980} D. Kazanas, \textit{Astrophys. J. Lett.} \textbf{241}, L59 (1980).
\bibitem{Starobinsky1980} A. A. Starobinsky, \textit{Phys. Lett. B} \textbf{91}, 99 (1980).
\bibitem{Sato1981} K. Sato, \textit{Phys. Lett. B} \textbf{99}, 66 (1981).
\bibitem{Guth1981} A. H. Guth, \textit{Phys. Rev. D} \textbf{23}, 347 (1981).
\bibitem{Mukhanov1981} V. F. Mukhanov and G. V. Chibisov, \textit{JETP Lett.} \textbf{33}, 532 (1981).
\end{thebibliography}
