

Gravitational waves from axion inflation in the gradient expansion formalism. Part II. Fermionic axion inflation

Richard von Eckardstein ^a, Kai Schmitz ^{a,b} and Oleksandr Sobol ^{a,c}

^a*Institute for Theoretical Physics, University of Münster,
Wilhelm-Klemm-Straße 9, 48149 Münster, Germany*

^b*Kavli IPMU (WPI), UTIAS, The University of Tokyo,
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8583, Japan*

^c*Physics Faculty, Taras Shevchenko National University of Kyiv,
64/13, Volodymyrska Street, 01601 Kyiv, Ukraine*

E-mail: richard.voneckardstein@uni-muenster.de,
kai.schmitz@uni-muenster.de, oleksandr.sobol@uni-muenster.de

ABSTRACT: Axion inflation represents an intriguing source of gravitational waves (GWs) from the early Universe. In a companion paper [1], we previously leveraged the gradient expansion formalism (GEF) to investigate pure axion inflation (PAI), i.e., axion inflaton coupled to a pure gauge sector. In this paper, we extend our analysis to fermionic axion inflation (FAI), i.e., we allow for the presence of fermions in the gauge sector. PAI predicts a strongly blue-tilted GW spectrum; in our GEF benchmark study, all parameter regions leading to observable GWs turned out to violate the upper limit on the number of extra relativistic degrees of freedom, ΔN_{eff} . As we demonstrate in this paper, the situation is different for FAI: Schwinger pair creation of the charged fermions results in a damping of gauge-field production, which attenuates the GW signal. As a result, the GW signal from FAI can fall into the sensitivity reach of LISA and ET without violating the upper limit on ΔN_{eff} . This result notably applies to the arguably most realistic variant of Abelian axion inflation, in which the axion couples to the hypercharge sector of the Standard Model. Besides, we discuss GW emission from the fermion gas, which may further enhance the total GW signal but which also requires a more quantitative investigation in future work. Additionally, we identify a new backreaction regime in which fermion production moderates the axion-vector dynamics. In this regime, the axion velocity and all energy-density components exhibit oscillations analogous to the strong backreaction in PAI, but here, the oscillations occur around the slow-roll trajectory and are damped by the presence of charged fermions. These observations define again an interesting GEF benchmark for future lattice studies.

KEYWORDS: Early Universe Particle Physics, Cosmology of Theories BSM, Axions and ALPs

ARXIV EPRINT: [2509.25013](https://arxiv.org/abs/2509.25013)

Contents

1	Introduction	1
2	Fermionic axion inflation	3
3	Tensor power spectrum and gravitational waves	9
4	Model specifications and constraints	11
5	Fermion-tempered backreaction	14
6	Gravitational waves from FAI	14
6.1	Instantaneous reheating	15
6.2	Varying the reheating temperature	19
6.3	Estimating the fermionic contribution to the SGWB	21
7	Conclusion and outlook	23
A	Alternative modeling of fermionic scale dependence	26
A.1	Scale-dependent damping with k_h or k_S	26
A.2	Comparison between k_h - and k_S -damping	27

1 Introduction

The theory of cosmic inflation [2–7] has been extremely successful in explaining the large-scale homogeneity, isotropy, and spatial flatness of our present Universe as well as the origin of the temperature anisotropies in the cosmic microwave background (CMB) [8–12]. However, the simplest models of slow-roll inflation do not address how the inflaton field eventually transfers its energy to a thermal plasma of Standard Model (SM) degrees of freedom. To this end, one needs to introduce a coupling between the inflaton and the SM fields [13–15]. This coupling has to be strong enough in order to allow for an efficient reheating. On the other hand, the radiative corrections to the inflaton potential due to these interactions can easily spoil its flatness and destroy the slow-roll regime of inflation.

A model that proposes to address this issue is axion inflation [16–18]. In this model, the inflaton is an axion-like pseudoscalar coupled to gauge fields via the topological Chern-Simons term $F_{\mu\nu} \tilde{F}^{\mu\nu}$. This construction extends the idea of natural inflation [19] of equipping the inflaton field with an approximate shift symmetry to ensure a naturally flat potential. In the simplest realization of axion inflation, the inflaton field couples to an Abelian vector field, but couplings to non-Abelian vector fields have also been proposed [20].

The Abelian axion inflation model succeeds in sourcing a strong gauge field via a well-known tachyonic instability [16]. If there exist other matter fields that are charged under the gauge group in question, e.g., like in the case of the SM $U(1)_Y$ hypercharge group, the nonperturbative phenomenon known as the Schwinger effect [21–23] implies that the strong

(hyper)electric and (hyper)magnetic fields will create particle-antiparticle pairs of all these charged fields, i.e., every SM fermion field. Even though the Schwinger effect has not been observed experimentally, calculations in de Sitter space [24–35] indicate that this effect should also be relevant during a phase of accelerated expansion like inflation.

A refined version of axion inflation builds upon this notion by including this production mechanism for charged fields [32, 33, 36–46]. Some references instead study the creation of charge carriers for an inflaton field coupled kinetically to the gauge field [47–50]. In both cases, charge carriers are produced in pairs and are thought to induce a conductive medium permeating the inflationary Universe, thereby damping gauge-field production. This effect invariably arises when the gauge bosons generated during inflation also couple to charge carriers. Consequently, any modeling of the production of the hypercharge gauge field during inflation necessarily needs to account for this Schwinger damping.

This article is the second in a two-part series investigating the gauge field generated during Abelian axion inflation as a source for a stochastic gravitational-wave background (SGWB). To achieve this, we employ the gradient expansion formalism (GEF) [36–38, 45, 51–53] to determine the inflationary dynamics. Our first article [1] is dedicated to Abelian axion inflation without additional charge carriers coupled to the gauge field, a model we dub *pure axion inflation* (PAI). In this article, we include the effect of fermions in a model we refer to as *fermionic axion inflation* (FAI). Specifically, we consider a realization of FAI in which the inflaton couples to the SM hypercharge gauge field. Our aim is to determine the prospects for detecting gauge-field-induced gravitational waves (GFIGW) with the Laser Interferometer Space Antenna (LISA) [54, 55], the Einstein Telescope (ET) [56], or within existing pulsar timing array data sets [57–61]. We confront these predictions with cosmological measurements, namely constraints on inflationary models from the PLANCK satellite [62], and limits on the amount of additional dark relativistic degrees of freedom, ΔN_{eff} [63–65]. Additionally, we account for the fact that the LIGO [66, 67] and Virgo [68] detector network did not observe an SGWB in their third observing run [69].

In Part I [1], our analysis demonstrated the difficulties with generating an observable SGWB in the PAI model. We found that sourcing an SGWB with an appreciable amplitude such as to be observable by ET or LISA required a precise compensation between an increase in the axion-vector coupling strength and a corresponding decrease in the inflationary scale. Furthermore, we encountered a tension between an observable SGWB originating from PAI and constraints on ΔN_{eff} . We identified that this conflict arises as a result of the explosive production of gauge bosons backreacting onto the inflaton dynamics, which prolongs the duration of inflation, thereby allowing for more gauge-field generation and an inevitable overproduction of gravitational radiation. As these results are based on the GEF and thus come with conceptual limitations (see below and ref. [1] for an extended discussion), they do not represent a general no-go theorem. Instead, our GEF benchmark study of the PAI model defines an interesting target for future numerical lattice studies.

In the present article, we find that the tension between ΔN_{eff} and an observable GFIGW signal that we had identified in the PAI model is absent for FAI coupled to the hypercharge gauge field. Our results show that the strong (hyper)electromagnetic fields generated due to the axion coupling also produce SM fermions in great abundance, thereby draining energy

from the gauge field. This dampening of the gauge field translates to the SGWB, which is consequently only mildly blue-tilted, but is still loud enough such that it could be detected by both LISA or ET. When these processes happen at a sufficiently high inflation scale, we even find that the signal may explain the evidence for an SGWB that the NANOGrav collaboration finds in their latest 15-year data set [57]. However, this interpretation is in conflict with CMB constraints on the amplitude of scalar and tensor perturbations coming from inflation. As a by-product of our analysis, we also show that the dampened production of gauge bosons due to fermions does not mean that the gauge field cannot backreact onto the inflaton dynamics. When the axion-vector coupling is large enough, backreaction can still occur, however, in a much more temperate manner compared to PAI. We refer to this novel regime of backreaction as *fermion-tempered backreaction*.

Evidently, fermions are crucial in FAI. They effectively suppress gauge-field production, thereby opening up a new region of parameter space which is otherwise in conflict with ΔN_{eff} constraints. However, in this study, we model their dynamics in terms of an effective fermion current expressed in terms of the gauge field, instead of treating them as independent degrees of freedom. Furthermore, we show by means of basic estimates that fermions could contribute a sizeable fraction to the SGWB at high frequencies, which would affect the detection prospects for the individual GW experiments considered in this work. Therefore, our results motivate a more complete treatment of the contribution to the total GW signal from the fermion gas in future work.

The structure of this article is as follows: in section 2, we review the FAI model. Section 3 gives a brief summary of gauge-field induced-gravitational waves. Section 4 discusses model constraints and GW detection characteristics. Then, we explore fermion-tempered backreaction in section 5, before entering an in-depth discussion of the SGWB signal sourced by FAI in section 6, where we assume both instantaneous reheating and the impact of a lowered reheating temperature. In this last section, we also estimate the importance of the fermions as a source for an SGWB contribution. Section 7 finishes with concluding remarks and an outlook. In appendix A, we compare numerical results for the SGWB spectra obtained in two different models of the Schwinger-induced damping.

Notation. Throughout this paper, we assume the background spacetime to be described by the spatially flat Friedmann-Lemaître-Robertson-Walker (FLRW) metric,

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = dt^2 - a^2(t) \delta_{ij} dx^i dx^j = a^2(\eta) \left(d\eta^2 - \delta_{ij} dx^i dx^j \right), \quad (1.1)$$

with scale factor a , physical time t , conformal time η , Greek indices for four-vectors, Latin indices for Euclidean three-vectors, and metric signature $(+, -, -, -)$. Unless stated otherwise, f' denotes the derivative of f with respect to conformal time, while \dot{f} denotes the derivative with respect to physical time. We define the Levi-Civita symbol in four dimensions such that $\varepsilon^{0123} = 1$. All results are given in natural units, $c = \hbar = 1$, and in terms of the reduced Planck mass, $M_{\text{P}} = 1/\sqrt{8\pi G} \simeq 2.435 \times 10^{18} \text{ GeV}$.

2 Fermionic axion inflation

The Lagrangian density for fermionic Abelian axion inflation describes the interactions between a $U(1)$ gauge field, A_μ , coupled via a Chern-Simons term to a pseudoscalar inflaton

field, ϕ , and interacting with fields, χ_i , charged under its U(1) symmetry group,

$$\mathcal{L}_{\text{FAI}} = \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}I(\phi)F_{\mu\nu}\tilde{F}^{\mu\nu} + \mathcal{L}_\chi(\{\chi_i\}, A_\mu). \quad (2.1)$$

In this equation, $V(\phi)$ represents the inflaton potential, $I(\phi)$ is a generic axial coupling function, and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field-strength tensor of the Abelian gauge field. The dual of the field strength tensor is $\tilde{F}^{\mu\nu} = \varepsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}/(2\sqrt{-g})$, with $g = \det g_{\mu\nu}$ being the determinant of the spacetime metric. The last term, \mathcal{L}_χ , is a gauge-invariant Lagrangian density describing the interactions of the gauge field with the charge carriers χ_i .¹ As outlined below, we treat the fermions effectively; thus, we do not further specify \mathcal{L}_χ .

By varying the action $S_{\text{FAI}} = \int d^4x\sqrt{-g}\mathcal{L}_{\text{FAI}}$ with respect to the metric, one derives the energy-momentum tensor for this system, which decomposes into three contributions corresponding to the inflaton field, the gauge field, and the charge carriers:

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi + F_\mu{}^\alpha F_{\alpha\nu} - g_{\mu\nu}\left(\frac{1}{2}\partial_\alpha\phi\partial^\alpha\phi - V(\phi) - \frac{1}{4}F_{\alpha\beta}F^{\alpha\beta}\right) + (T_\chi)_{\mu\nu}, \quad (2.2)$$

where $(T_\chi)_{\mu\nu}$ is the contribution to the energy-momentum tensor from the charge carriers. Assuming the background spacetime to be homogeneous and isotropic, T_ν^μ can be separated into a background contribution, \bar{T}_ν^μ , and a perturbation, δT_ν^μ ,

$$T_\nu^\mu = \bar{T}_\nu^\mu + \delta T_\nu^\mu, \quad \bar{T}_\nu^\mu \equiv \langle T_\nu^\mu \rangle, \quad \delta T_\nu^\mu \equiv T_\nu^\mu - \langle T_\nu^\mu \rangle. \quad (2.3)$$

The average $\langle \cdot \rangle$ depends on the system in question; in our case, it means taking the quantum expectation value with respect to the Bunch-Davies vacuum. The background energy-momentum tensor can be viewed as describing an ideal fluid, whose subcomponents constitute a homogeneous energy density, ρ , and pressure, p ,

$$\rho = \frac{1}{2}\dot{\varphi}^2 + V(\varphi) + \frac{1}{2}\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle + \rho_\chi, \quad p = \frac{1}{2}\dot{\varphi}^2 - V(\varphi) + \frac{1}{6}\langle \mathbf{E}^2 + \mathbf{B}^2 \rangle + \frac{1}{3}\rho_\chi. \quad (2.4)$$

To arrive at these expressions, we assumed that spatial gradients in the inflaton field are negligible such that only the inflaton zero mode contributes to the background expansion, $\varphi(t) = \langle \phi(t, \mathbf{x}) \rangle$. Furthermore, we suppose that all charge carriers are massless, thus $p_\chi = \rho_\chi/3$. For convenience, we also defined electric and magnetic fields, \mathbf{E} and \mathbf{B} , via

$$F_{0i} = aE_i, \quad F_{ij} = -a^2\varepsilon_{ijk}B^k. \quad (2.5)$$

¹These charge carriers can be of arbitrary spin, including scalars, fermions, or vector bosons. However, in the particular case of the SM hypercharge gauge group U(1)_Y, which is of primary interest for us in this paper, fermions are the most relevant charge carriers. More specifically, we assume that the SM Higgs field is stabilized at its origin during axion inflation, e.g., via a nonminimal coupling to the Ricci scalar. In this case, electroweak symmetry is unbroken, all SM fermions are massless, but the SM Higgs field itself is heavy and hence less susceptible to Schwinger pair production. We thus restrict ourselves to fermionic charge carriers in our analysis and refer to our model as *fermionic axion inflation* for precisely this reason.

These background quantities evolve according to the following coupled equations of motion,

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi}(\phi) = \frac{1}{2}I_{,\phi}(\phi) \langle \mathbf{E} \cdot \mathbf{B} + \mathbf{B} \cdot \mathbf{E} \rangle, \quad (2.6a)$$

$$\operatorname{div} \mathbf{E} = 0, \quad \operatorname{div} \mathbf{B} = 0, \quad (2.6b)$$

$$\dot{\mathbf{E}} + 2H\mathbf{E} - \frac{1}{a}\operatorname{rot} \mathbf{B} + I_{,\phi}(\phi)\dot{\phi}\mathbf{B} + \mathbf{J} = 0, \quad (2.6c)$$

$$\dot{\mathbf{B}} + 2H\mathbf{B} + \frac{1}{a}\operatorname{rot} \mathbf{E} = 0, \quad (2.6d)$$

$$H^2 = \frac{\rho}{3M_{\text{P}}^2}, \quad (2.6e)$$

$$\dot{\rho}_\chi + 4H\rho_\chi = \langle \mathbf{J} \cdot \mathbf{E} \rangle. \quad (2.6f)$$

The evolution equation for the energy density of the charge carriers, ρ_χ , follows from covariant energy conservation at the background level. Furthermore, we define the four-current $J^\mu = -\partial\mathcal{L}_\chi/\partial A_\mu = (\rho_c, \mathbf{J}/a)$, with \mathbf{J} the induced current and net charge ρ_c . We set $\rho_c = 0$, since the Schwinger effect produces particles and antiparticles in equal shares.

The effective description in terms of a charged current \mathbf{J} requires further attention. Assuming constant and (anti-)collinear classical electric and magnetic fields in an idealized de Sitter background, the effective current induced via Schwinger pair creation for a single particle species of charge $e|Q|$ and mass m is derived to be [32, 33]

$$|\mathbf{J}'_{\text{SE}}| = \frac{(e|Q|)^3}{6\pi^2 H} |\mathbf{E}'_c| |\mathbf{B}'_c| \psi \left(\frac{\pi |\mathbf{B}'_c|}{|\mathbf{E}'_c|} \right) \exp \left(-\frac{\pi m^2}{e|Q| |\mathbf{E}'_c|} \right). \quad (2.7)$$

By the subscript ‘c’, we indicate that the electric and magnetic fields above are classical and constant and are a priori different from \mathbf{E} and \mathbf{B} appearing in eq. (2.6). Quantities marked with a prime (′) are given in a frame where \mathbf{E}_c and \mathbf{B}_c are (anti-)collinear. The function $\psi(x)$ is given by $\coth x$ or $1/(2 \sinh x)$ for Dirac fermions or complex scalars, respectively. To use this non-trivial result to effectively model \mathbf{J} , we follow considerations outlined in our previous work [45].

Consider that, during axion inflation, quantum gauge-field modes are enhanced, and classicalize once their wavelengths are of the same order as the Hubble horizon. Then, we may suppose that classical electric and magnetic fields with strengths $\langle \mathbf{E}^2 \rangle^{1/2}$ and $\langle \mathbf{B}^2 \rangle^{1/2}$, respectively, permeate a single Hubble patch at every moment in time. Because these classical fields are also the reason for the Schwinger pair creation of charged particles, their presence also implies the existence of an effective conductive medium. Mathematically, the effect of this medium can be quantified in terms of a conductivity σ that relates the induced current either to $\langle \mathbf{E}^2 \rangle^{1/2}$ or to $\langle \mathbf{B}^2 \rangle^{1/2}$ via a generalized form of Ohm’s law, i.e.,

$$\sigma'_E = \frac{|\mathbf{J}'_{\text{SE}}|}{\langle \mathbf{E}'^2 \rangle^{1/2}} \quad \text{or} \quad \sigma'_B = \frac{|\mathbf{J}'_{\text{SE}}|}{\langle \mathbf{B}'^2 \rangle^{1/2}}. \quad (2.8)$$

However, the collinearity of the electric and magnetic fields renders the precise definition of σ a priori ambiguous. In the above equation, \mathbf{J}'_{SE} should be understood as computed

from eq. (2.7) in terms of $\langle \mathbf{E}'^2 \rangle$ and $\langle \mathbf{B}'^2 \rangle$. Since the quantum electric and magnetic fields also feel this conductivity, these considerations motivate a first definition of the quantum induced current \mathbf{J} entering in our EOMs:

$$\mathbf{J} \equiv \sigma_E \mathbf{E} + \sigma_B \mathbf{B}. \quad (2.9)$$

That we consider both magnetic and electric conductivities in this definition is a consequence of the fact that the electric and magnetic fields are not necessarily collinear in the comoving frame in which eq. (2.9) applies. The conductivities $\sigma_{E/B}$ may be derived by boosting to the collinear frame, computing the conductivities from eq. (2.7), and boosting back to the comoving frame. This procedure yields

$$\sigma_E = \sqrt{\frac{\langle \mathbf{E}^2 \rangle - \langle \mathbf{B}^2 \rangle + \Sigma}{\langle \mathbf{E}^2 \rangle + \langle \mathbf{B}^2 \rangle + \Sigma}} \sum_i \mathfrak{F}_i, \quad (2.10)$$

$$\sigma_B = \text{sign}(\langle \mathbf{E} \cdot \mathbf{B} \rangle) \sqrt{\frac{\langle \mathbf{B}^2 \rangle - \langle \mathbf{E}^2 \rangle + \Sigma}{\langle \mathbf{E}^2 \rangle + \langle \mathbf{B}^2 \rangle + \Sigma}} \sum_i \mathfrak{F}_i, \quad (2.11)$$

where $\Sigma = \sqrt{[\langle \mathbf{E}^2 \rangle - \langle \mathbf{B}^2 \rangle]^2 + 4\langle \mathbf{E} \cdot \mathbf{B} \rangle^2}$. The \mathfrak{F}_i encompass the contributions to $\sigma_{E/B}$ of the individual charged fermions and scalars, χ_i , with respective masses m_i and charges Q_i

$$\mathfrak{F}_i = \frac{(e|Q_i|)^3 |\langle \mathbf{E} \cdot \mathbf{B} \rangle|}{6\pi^2 H \sqrt{\Sigma}} \psi_i \left(\pi \sqrt{\frac{\langle \mathbf{B}^2 \rangle - \langle \mathbf{E}^2 \rangle + \Sigma}{\langle \mathbf{E}^2 \rangle - \langle \mathbf{B}^2 \rangle + \Sigma}} \right) \exp \left(-\frac{\pi m_i^2 \sqrt{2}}{e|Q_i|(\langle \mathbf{E}^2 \rangle - \langle \mathbf{B}^2 \rangle + \Sigma)^{1/2}} \right). \quad (2.12)$$

The Ohmic form of the Schwinger-induced current in eq. (2.9) implies that all gauge-field modes are equally affected by the conductive medium. However, as we argued in ref. [45], this description cannot reflect the realistic physical situation. In this simple picture, gauge-field modes with wavelengths much shorter than the typical separation of charge carriers in the conductive medium would feel its effect just as much as modes with far greater wavelengths. To mitigate the severity of this assumption, we account for an additional scale dependence in eq. (2.9) by making the following modification,

$$\mathbf{J}(t, \mathbf{x}) \equiv \frac{1}{(2\pi)^3} \int d^3 \mathbf{y} \left(\sigma_E(t) \mathbf{E}(t, \mathbf{y}) + \sigma_B(t) \mathbf{B}(t, \mathbf{y}) \right) \int d^3 \mathbf{k} \Theta(t, k) e^{i\mathbf{k} \cdot (\mathbf{x} - \mathbf{y})}. \quad (2.13)$$

By convoluting the original current \mathbf{J} with a heuristic function $\int d^3 \mathbf{k} \Theta(t, k) e^{i\mathbf{k} \cdot \mathbf{x}}$, we gain the flexibility to reflect the spectral dependence of the conductive medium by appropriately choosing $\Theta(t, k)$. In ref. [45], we discussed a well-motivated choice for this spectral dependence. This relies on identifying a characteristic momentum scale, $k_S(t)$, which reflects the typical particle-antiparticle separation following their creation. This scale may be estimated from the strength of the electric field at time t , $k_S/a \propto |\mathbf{E}'_c|^{1/2}$. We give more details on this choice for $\Theta(t, k)$ in appendix A.

It is instructive to study the impact of the conductive medium in Fourier space. To this end, consider the gauge field A_μ in radiation gauge ($\text{div} \mathbf{A} = 0$, $A_0 = 0$),

$$\hat{\mathbf{A}}(\eta, \mathbf{x}) = \int \frac{d^3 \mathbf{k}}{(2\pi)^{3/2}} \sum_{\lambda=\pm 1} \left(\epsilon_\lambda(\mathbf{k}) A_\lambda(\eta, k) \hat{a}_\lambda(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{x}} + \epsilon_\lambda^*(\mathbf{k}) A_\lambda^*(\eta, k) \hat{a}_\lambda^\dagger(\mathbf{k}) e^{-i\mathbf{k} \cdot \mathbf{x}} \right), \quad (2.14)$$

in a helicity basis with polarization vectors $\epsilon_\lambda(\mathbf{k})$ obeying

$$\begin{aligned} \mathbf{k} \cdot \epsilon_\lambda(\mathbf{k}) &= 0, & i\mathbf{k} \times \epsilon_\lambda(\mathbf{k}) &= \lambda k \epsilon_\lambda(\mathbf{k}), \\ \epsilon_\lambda(\mathbf{k}) \cdot \epsilon_{\lambda'}^*(\mathbf{k}) &= \delta_{\lambda\lambda'}, & \epsilon_\lambda^*(\mathbf{k}) &= \epsilon_{-\lambda}(\mathbf{k}) = \epsilon_\lambda(-\mathbf{k}), \end{aligned} \quad (2.15)$$

and creation and annihilation operators $\hat{a}_\lambda^\dagger(\mathbf{k})$ and $\hat{a}_\lambda(\mathbf{k})$ satisfying

$$[\hat{a}_\lambda(\mathbf{k}), \hat{a}_{\lambda'}^\dagger(\mathbf{k}')] = \delta_{\lambda\lambda'} \delta^{(3)}(\mathbf{k} - \mathbf{k}'), \quad [\hat{a}_\lambda(\mathbf{k}), \hat{a}_{\lambda'}(\mathbf{k}')] = [\hat{a}_\lambda^\dagger(\mathbf{k}), \hat{a}_{\lambda'}^\dagger(\mathbf{k}')] = 0. \quad (2.16)$$

From Ampère's law in eq. (2.6c) and the induced current \mathbf{J} in eq. (2.13), one may derive an evolution equation for the mode functions $A_\lambda(t, k)$,

$$\ddot{A}_\lambda(t, k) + (H + \sigma_E \Theta(t, k)) \dot{A}_\lambda(t, k) + \left[\left(\frac{k}{a} \right)^2 - \lambda \frac{k}{a} (2\xi H + \sigma_B \Theta(t, k)) \right] A_\lambda(t, k) = 0, \quad (2.17)$$

which reduces to the evolution equation in the PAI model for $\sigma_{E/B} = 0$. Here, as usual, $\xi = I_{,\phi}(\varphi)\dot{\varphi}/(2H)$ denotes the instability parameter or gauge-field production parameter of axion inflation, which, in the absence of fermions, determines the threshold momentum below which gauge-field modes experience the tachyonic enhancement. Qualitatively, the physical interpretation of this equation matches the one of PAI: gauge-field production is a consequence of the non-zero inflaton velocity, $\dot{\varphi}$, which triggers a tachyonic instability in eq. (2.17) for one of the gauge-field helicities, $\lambda = \pm 1$, depending on the sign of $\dot{\varphi}$. However, the Schwinger effect reduces the efficiency of gauge-field production, as σ_B takes an opposite sign to ξ , yielding an impeded effective instability parameter $\xi_{\text{eff}} = \xi + \sigma_B/(2H)$. Meanwhile, the electric conductivity σ_E dampens the gauge-field modes by adding to Hubble friction. The role of the heuristic spectral function $\Theta(t, k)$ is also apparent in eq. (2.17): it determines the range of gauge-field momenta k that are subject to the influence of the purely time-dependent damping terms $\sigma_{E/B}$.

It is important to note the ramifications for the dynamics and phenomenology of axion inflation because of the inhibited gauge-field generation due to the Schwinger effect. The main difference regarding the dynamical evolution is a decrease in $\langle \mathbf{E} \cdot \mathbf{B} \rangle$ such that the backreaction onto the inflaton via the friction term on the right-hand side in eq. (2.6a) is diminished. For a particular benchmark point, ref. [42] demonstrated that a PAI system experiences a phase of strong gauge-field backreaction while a comparable FAI system did not. To understand the implication of this, consider that strong backreaction results in an increased duration of inflation by drawing kinetic energy from the inflaton field and instead allowing for more gauge-field production, which ultimately ends inflation via $V(\varphi) \sim \rho_{EM}$ instead of $V(\varphi) \sim \dot{\varphi}^2$. Therefore, decreasing the impact of backreaction implies that, in FAI, inflation ends on the slow-roll trajectory with a subdominant energy density in ρ_{EM} . This implies that an extended phase of reheating may still take place after the end of axion inflation. A second implication of absent backreaction regards axion inhomogeneities, which we have neglected throughout the derivation presented above. In PAI, these inhomogeneities become large during a phase of strong gauge-field backreaction and can alter the dynamical evolution of the inflaton-gauge-field system [70–73]. At present, this means that studies of PAI that neglect these inhomogeneities, like the GEF, can only qualitatively capture the

phase of strong backreaction [70, 71]. To obtain the full dynamics of PAI, lattice simulations are necessary. However, as we expect an overall reduced dynamical impact of the gauge field during FAI, it is conceivable that also fewer axion inhomogeneities are sourced.

In this regard, we would like to mention the recent work [46], which studies the same physical model by solving the corresponding equations of motion on a lattice using the PENCIL CODE [74]. The impact of the Schwinger effect is incorporated by employing the same Ohmic form of the Schwinger current (2.9) with the conductivities given in (2.11). One important difference between their lattice implementation and the approach used in the present work concerns the treatment of Schwinger damping of the gauge-field modes. Here, we account for the damping only of those modes whose wavelengths exceed the Schwinger pair-production length scale. In contrast, ref. [46] applies the same damping to all gauge-field modes, independently of their momentum, since it is not feasible to perform efficient mode selection in position space on a lattice. With this assumption, the authors of ref. [46] carried out a numerical analysis at three benchmark points in parameter space (specifically, for the axion mass $m = 5.2 \times 10^{-6} M_{\text{P}}$ and axion-vector coupling $\beta = 12, 15, \text{ and } 18$).² They found excellent agreement between the fully inhomogeneous simulations and the homogeneous setup where axion gradients were manually switched off. This outcome confirms our expectations discussed above, at least within the parameter region explored in ref. [46]. Qualitatively, we see no reason for axion gradients to become important at larger coupling values unless the backreaction eventually becomes significant. In section 5, we identify and investigate the properties of this new regime of *fermion-tempered backreaction*, still within the homogeneous approximation. These results provide a GEF benchmark for this region of parameter space, and they should be validated by future lattice simulations.

Generally speaking, lattice simulations have two main advantages compared to semianalytical methods: (i) they can naturally incorporate axion inhomogeneities, and (ii) they allow one to study gauge-field and fermion production after the end of inflation, during the preheating stage. On the other hand, semianalytical approaches such as the GEF offer several important advantages: (a) they are not limited in dynamical range or simulation duration; (b) they allow one to treat different gauge-field modes differently (for example, to include selective damping or to isolate partial contributions to the energy densities); (c) they are significantly faster and enable parameter scans over wide regions of the model space, rather than exploring only isolated benchmark points. Moreover, lattice methods often rely on the output of semianalytical homogeneous approaches for calibration purposes and for obtaining accurate initial conditions for the simulation. Therefore, we believe that combining our GEF results with future lattice simulations will lead to a highly synergistic outcome.

Qualitatively, the obstructed gauge-field production also inhibits the subsequent sourcing of GFIGWs. On the other hand, the energy from the gauge field is instead transferred to the charge carriers, as may be seen from eq. (2.6f). Previous studies found this production of charge carriers to be efficient enough such that ρ_{χ} surpasses ρ_{EM} at the end of inflation [1, 37]. If the fluid associated with these charge carriers is anisotropic, it will add a second induced component to the GW spectrum.

²Note that a direct comparison of our results and the ones of ref. [46] is impossible since we are considering higher values of the coupling constants, $\beta \geq 20$ which are needed to obtain a sufficiently strong GW signal.

3 Tensor power spectrum and gravitational waves

We study the production of tensor perturbations by working with a first-order perturbed FLRW metric, tracking only the transverse and traceless tensor perturbations, h_{ij}^{TT} ,

$$ds^2 = a^2(\eta) \left(d\eta^2 - [\delta_{ij} + h_{ij}^{\text{TT}}] dx^i dx^j \right). \quad (3.1)$$

We describe these perturbations as a quantum field, which we can express in Fourier space using a circular polarisation basis such that

$$\hat{h}_{ij}^{\text{TT}}(\eta, \mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \sum_{\lambda=\pm 1} \left(\epsilon_i^\lambda(\mathbf{k}) \epsilon_j^\lambda(\mathbf{k}) \hat{h}_\lambda(\eta, k) e^{i\mathbf{k}\cdot\mathbf{x}} + \text{h.c.} \right), \quad (3.2)$$

where the polarization vectors $\epsilon^\lambda(\mathbf{k})$ are defined as in eq. (2.15). The Fourier amplitudes, $\hat{h}_\lambda(\eta, k)$, evolve according to

$$\hat{h}_\lambda'' + 2\mathcal{H}\hat{h}_\lambda' + k^2\hat{h}_\lambda = \frac{2}{M_{\text{P}}^2} a^2 \Pi_\lambda^{ij}(\mathbf{k}) \int \frac{d^3\mathbf{x}}{(2\pi)^{3/2}} \sigma_{ij} e^{-i\mathbf{k}\cdot\mathbf{x}}, \quad (3.3)$$

where $\mathcal{H} = a'/a$ is the comoving Hubble rate, $\Pi_\lambda^{ij} = \epsilon_{-\lambda}^i(\mathbf{k}) \epsilon_{-\lambda}^j(\mathbf{k})$ is the projector onto the helicity basis and σ_{ij} is the anisotropic stress of the perturbed energy-momentum tensor δT_ν^μ . In the case of FAI, we derive the anisotropic stress from eq. (2.2), finding

$$\sigma_{ij} = \partial_i \phi \partial_j \phi - (E_i E_j + B_i B_j) + \sigma_{ij}^\chi. \quad (3.4)$$

Assuming axion inhomogeneities to be small, we may neglect the first term. The second term, the anisotropic stress due to the gauge field, is sourced by the tachyonic instability in eq. (2.17), and is analogous to the one obtained in PAI. The final term, σ_{ij}^χ , reflects the anisotropic stress of the charged fields χ_i . As before, we keep this term generic, as we are treating all charged fields effectively without resolving their dynamics.

From here, we proceed as in our preceding article, ref. [1], writing the solution to eq. (3.3) in terms of a homogeneous (vacuum) and a particular (induced) solution, $\hat{h}_\lambda = \frac{2}{M_{\text{P}}} \left(\hat{u}_\lambda^{\text{vac}} + \hat{u}_\lambda^{\text{ind}} \right)$. In the case of FAI, the induced contribution further separates into two parts, $\hat{u}_\lambda^{\text{ind}} = \hat{u}_\lambda^{\text{GF}} + \hat{u}_\lambda^\chi$, one induced by the gauge field, the other by the charged fields. Defining the tensor power spectrum for the polarization $\lambda = \pm 1$ as

$$\langle \hat{h}_\lambda(\eta, \mathbf{k}) \hat{h}_{\lambda'}(\eta, \mathbf{k}') \rangle = \delta^{(3)}(\mathbf{k} + \mathbf{k}') \delta_{\lambda\lambda'} \frac{\pi^2}{k^3} \mathcal{P}_{T,\lambda}(k), \quad (3.5)$$

the total tensor power spectrum, $\mathcal{P}_T(k) = \sum_{\lambda=\pm 1} \mathcal{P}_{T,\lambda}(k)$, receives three contributions,

$$\mathcal{P}_T(k) = \mathcal{P}_T^{\text{vac}}(k) + \mathcal{P}_T^{\text{GF}}(k) + \mathcal{P}_T^\chi(k), \quad (3.6)$$

assuming that there are no cross-terms between $\hat{u}_\lambda^{\text{GF}}$ and \hat{u}_λ^χ . This separation is expected from Wick's theorem, as both the charged fields and the gauge field are expressed in terms of their own set of annihilation and creation operators. We presented a more detailed derivation of the first two contributions in Part I, ref. [1], the result of which is

$$\mathcal{P}_T^{\text{vac}}(\eta, k) = \frac{4k^3}{\pi^2 M_{\text{P}}^2} |u_\lambda^0(\eta, k)|^2, \quad (3.7a)$$

$$\mathcal{P}_T^{\text{GF}}(\eta, k) = \sum_{\lambda=\pm 1} \frac{k^3}{2\pi^2 M_{\text{P}}^4} \int \frac{d^3 \mathbf{p}}{(2\pi)^3} \sum_{\alpha, \beta=\pm 1} \left(1 + \lambda\alpha \frac{\mathbf{k} \cdot \mathbf{p}}{kp}\right)^2 \left(1 + \lambda\beta \frac{k^2 - \mathbf{k} \cdot \mathbf{p}}{kq}\right)^2 \quad (3.7b)$$

$$\times \left| \int_{-\infty}^0 d\tau \frac{G_k(\eta, \tau)}{a^2(\tau)} \left[A'_\alpha(\tau, p) A'_\beta(\tau, q) + \alpha\beta pq A_\alpha(\tau, p) A_\beta(\tau, q) \right] \right|^2.$$

Here, the gauge-field modes $A_\lambda(t, k)$ are solutions to eq. (2.17), $G_k(\eta, \tau)$ is the retarded Green function to the differential operator $\mathcal{D} = (\partial^2/\partial\eta^2) + 2\mathcal{H}(\partial/\partial\eta) + k^2$, and $q = |\mathbf{k} - \mathbf{p}|$. A detailed computation of the χ -induced contribution, $\mathcal{P}_T^\chi(k)$, is beyond the scope of the current paper. Therefore, we will simply keep this term generic, opting for a simple scaling argument to estimate its importance in the later part of this article.

From the tensor power spectrum, one can directly compute the GW energy density per logarithmic frequency interval in units of the critical energy density,

$$\Omega_{\text{GW}}(f) \equiv \frac{1}{3H_0^2 M_{\text{P}}^2} \frac{d\rho_{\text{GW}}(f)}{d \ln f} = \frac{\pi^2}{3H_0^2} f^2 |\mathcal{T}_{\text{GW}}(f)|^2 \mathcal{P}_T(\eta_{\text{out}}(k_f), k_f), \quad k_f = 2\pi a_0 f, \quad (3.8)$$

where $H_0 = h \times 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is the Hubble constant and the transfer function \mathcal{T}_{GW} describes the evolution of a GW mode with momentum k_f from horizon re-entry when $k_f \simeq \mathcal{H}(\eta_{\text{in}})$ (after horizon exit during inflation when $k_f \simeq \mathcal{H}(\eta_{\text{out}})$) until today [76],

$$|\mathcal{T}_{\text{GW}}(f)|^2 \simeq \frac{H_0^2 \Omega_r}{8\pi^2 f^2} \frac{g_*(T_f)}{g_*(T_0)} \left(\frac{g_{*,S}(T_0)}{g_{*,S}(T_f)} \right)^{4/3} \left[1 + \frac{9}{16} \left(\frac{f_{\text{eq}}}{\sqrt{2}f} \right)^2 \right] |\mathcal{T}_{\text{reh}}(f)|^2. \quad (3.9)$$

The functions $g_*(T)$ and $g_{*,S}(T)$ represent the effective number of relativistic degrees of freedom in the SM thermal plasma at temperature T contributing to the energy density and entropy, respectively. The temperature T_f corresponds to the time when k_f re-entered the horizon and $T_0 \simeq 2.73 \text{ K}$ is the temperature of the CMB photons today. Ω_r is the fractional energy density in radiation today, $h^2 \Omega_r = 4.2 \times 10^{-5}$. The fourth factor describes the transition from radiation to matter domination, with $f_{\text{eq}} = 2.1 \times 10^{-17} \text{ Hz}$ the frequency corresponding to matter-radiation equality. The last factor, $|\mathcal{T}_{\text{reh}}(f)|^2$, affects modes which re-enter the horizon during a period of reheating before the onset of radiation domination. In our previous article, ref. [1], we assumed instantaneous reheating, thus $|\mathcal{T}_{\text{reh}}(f)|^2 = 1$. However, as discussed in the previous section, FAI leaves more room for a period of reheating, which can impact the GW spectrum via this transfer function.

Reheating implies that frequencies experience an additional redshift due to this non-standard phase of cosmological expansion. This is evident when expanding the relationship $k_f = 2\pi a_0 f$, between a frequency f and the corresponding comoving momentum k_f ,³

$$f = \frac{k_f}{2\pi a_0} = \frac{k_f}{2\pi a_{\text{end}}} \frac{a_{\text{end}}}{a_{\text{reh}}} \frac{a_{\text{reh}}}{a_0} = \frac{k_f}{2\pi a_{\text{end}}} e^{-N_{\text{reh}}} \left(\frac{g_{*,S}(T_0)}{g_{*,S}(T_{\text{reh}})} \right)^{1/3} \frac{T_0}{T_{\text{reh}}}, \quad (3.10)$$

³We fix the normalization of the scale factor during the early stages of inflation. In this convention, the present-day value of the scale factor a_0 is not fixed to a certain value (e.g., $a_0 = 1$), but depends on the details of the expansion history. For our purposes, this convention proves advantageous, since it allows us to use the wavenumber k_f as a clock variable during inflation that is independent of the details of reheating.

where a_{end} denotes the end of inflation, which is followed by $N_{\text{reh}} = \ln(a_{\text{reh}}/a_{\text{end}})$ e -folds of reheating until radiation domination sets in at a temperature T_{reh} . By modeling reheating as an expansion phase governed by a non-interacting fluid with equation of state w_{reh} , the duration of reheating, N_{reh} , may be related to the reheating temperature, T_{reh} , via

$$N_{\text{reh}} = \frac{1}{3(1+w_{\text{reh}})} \ln \left(\frac{90M_{\text{P}}H_{\text{end}}^2}{\pi^2 g_*(T_{\text{reh}})T_{\text{reh}}^4} \right). \quad (3.11)$$

Assuming a standard reheating mechanism where the dominant energy density is in inflaton oscillations around its potential minimum, reheating will manifest itself as a period of early matter domination, $w_{\text{reh}} = 0$. For this case, the transfer function $|\mathcal{T}_{\text{reh}}(f)|^2$ has been computed in refs. [77, 78],

$$|\mathcal{T}_{\text{reh}}(f)|^2 = \frac{\theta(f_{\text{end}} - f)}{1 - 0.22 \left(\frac{f}{f_{\text{reh}}}\right)^{1.5} + 0.65 \left(\frac{f}{f_{\text{reh}}}\right)^2}, \quad (3.12)$$

with the frequencies f_{end} and f_{reh} corresponding to the end of inflation and the end of reheating, respectively. As evident from eqs. (3.10) and (3.12), reheating affects the GW spectrum by further redshifting all frequencies with respect to instantaneous reheating, while also suppressing the GW spectrum at high frequencies, $f \gtrsim f_{\text{reh}}$, by a factor f^{-2} .

4 Model specifications and constraints

For the remainder of this article, we focus on the scenario where the gauge field A_μ in eq. (2.1) is the SM hypercharge field. The charge carriers χ_i therefore correspond to the fermions of the SM, which we assume to be massless; see the discussion in Footnote 1. These assumptions are expressed in the individual fermion contributions \mathfrak{F}_i to the conductivities $\sigma_{E/B}$ in eq. (2.11), which sum up to

$$\sum_i \mathfrak{F}_i = C \frac{g'(\mu)^3}{6\pi^2 H} \frac{|\langle \mathbf{E} \cdot \mathbf{B} \rangle|}{\sqrt{\Sigma}} \coth \left(\pi \sqrt{\frac{\langle \mathbf{B}^2 \rangle - \langle \mathbf{E}^2 \rangle + \Sigma}{\langle \mathbf{E}^2 \rangle - \langle \mathbf{B}^2 \rangle + \Sigma}} \right), \quad (4.1)$$

with $C = 41/12$, equal to half the sum of the cubes of hypercharges of all SM fermions. The coupling $g'(\mu)$ is the running hypercharge gauge coupling at one-loop order,

$$g'(\mu) = g'(m_Z) \left(1 + g'(m_Z)^2 \frac{41}{48\pi^2} \ln \frac{m_Z}{\mu} \right)^{-1/2}, \quad (4.2)$$

with the Z -boson mass, $m_Z \simeq 91.2 \text{ GeV}$, $g'(m_Z) \simeq 0.35$ and the characteristic energy scale of pair creation $\mu = \rho_{\text{EM}}^{1/4}$. For brevity, we will refer to this realization of FAI as FAI_{SM}.⁴

⁴Reference [79] argues that gauge-field modes can efficiently interact with each other via fermions and thus thermalize already when the instability parameter satisfies $|\xi| \gtrsim 2.9$ for Standard Model couplings. This argument is based on a computation of the rate for the Breit–Wheeler process $\gamma\gamma \rightarrow e^+e^-$, enhanced by the large occupation numbers of the gauge-field modes. Although we agree that the computation itself may be correct, we believe that its interpretation is misleading. Gauge-field modes with large occupation numbers correspond to a superhorizon classical gauge field rather than on-shell photons, and the computed rate effectively describes the Schwinger pair-production process. Indeed, our numerical solutions confirm that Schwinger pair production becomes efficient when $|\xi_{\text{eff}}|$ is greater than 3. Note also that the inverse process $e^+e^- \rightarrow \gamma\gamma$ is not efficient and does not lead to thermalization, since fermionic states cannot achieve large occupation numbers. Therefore, throughout this work, we assume that the gauge field does not thermalize.

Just as in the analysis of PAI in Part I [1], we work with a linear axion-vector coupling, $I(\phi) = (\beta/M_{\text{P}})\phi$ and assume a simple quadratic scalar potential $V(\phi) = m^2\phi^2/2$. As we argued in Part I, this quadratic potential is representative of a larger class of inflationary models, whose late-time attractor is nearly quadratic below some threshold field value φ_{thr} . Inflation simulated on a quadratic potential yields representative GW spectra for frequencies $f \gtrsim f_{\text{thr}}$, where f_{thr} is the frequency corresponding to modes which exit the horizon when $\varphi = \varphi_{\text{thr}}$. Assuming that the inflaton trajectory for $\varphi < \varphi_{\text{thr}}$ is approximately given by the slow-roll attractor of chaotic inflation, one may deduce the relationship

$$V(\varphi_{\text{thr}}) \simeq m^2 M_{\text{P}}^2 \left[20 - 4.6 \log_{10} \left(\frac{f_{\text{thr}}}{\text{Hz}} \right) + 1.54 \log_{10} \left(\frac{m}{M_{\text{P}}} \frac{T_{\text{reh}}}{\text{GeV}} \right) - 2\Delta N_{\text{BR}} \right], \quad (4.3)$$

which is the generalization of eq. (4.5) in ref. [1] that allows for a period of reheating with $w_{\text{reh}} = 0$. We use this relationship to determine bounds on m by using the measured value of the amplitude of scalar perturbations, $A_S = 2.1 \times 10^{-9}$ [63, 80], bounds on the tensor-to-scalar ratio, $r \lesssim 0.03$, [80, 81], and non-Gaussianity at CMB scales, $|\xi_{\text{CMB}}| \lesssim 2.5$ [82–85],

$$V(\varphi_{\text{thr}}) \lesssim \frac{3}{2} \pi^2 A_S r M_{\text{P}}^4 \lesssim 9.3 \times 10^{-10} M_{\text{P}}^4. \quad (4.4a)$$

$$V(\varphi_{\text{thr}}) \lesssim 48 \pi^2 A_S \frac{|\xi_{\text{CMB}}|}{\beta^2} M_{\text{P}}^4 \lesssim 6.2 \times 10^{-6} M_{\text{P}}^4 / \beta^2. \quad (4.4b)$$

The second bound is only more constraining for $\beta \sim 80 - 100$ (which we never reach in the present study), so we focus entirely on the first bound. Modeling $V(\varphi_{\text{thr}})$ as in eq. (4.3), we can observe that the constraint on m is marginally relaxed when reducing f_{thr} with variations ranging from $m \lesssim (3.1 - 3.4) \times 10^{-6} M_{\text{P}}$ for $f_{\text{thr}} = 10^{-12} - 10^{-9}$ Hz and instantaneous reheating. A lower reheating temperature also only mildly affects the bounds: even for $T_{\text{reh}} \simeq 1$ MeV, we only find $m \lesssim (3.9 - 4.4) \times 10^{-6} M_{\text{P}}$ for $f_{\text{thr}} = 10^{-12} - 10^{-9}$ Hz. The largest effect can be obtained by allowing for an extremely prolonged duration of inflation due to backreaction. For example, for $\Delta N_{\text{BR}} = 20$ and varying $f_{\text{thr}} = 10^{-12} - 10^{-9}$ Hz, one obtains $m \lesssim (4.1 - 4.8) \times 10^{-6} M_{\text{P}}$ assuming instantaneous reheating, and $m \lesssim 6.4 \times 10^{-6} - 10^{-5} M_{\text{P}}$ for $T_{\text{reh}} \simeq 1$ MeV. Evidently, only when allowing for a low reheating temperature and many e -folds of backreaction can the bound on m be substantially relaxed. In practice, we will indicate bounds for $\Delta N_{\text{BR}} = 0$ in our results while still studying inflaton masses as high as $5 \times 10^{-5} M_{\text{P}}$ due to the uncertainties involved in these estimates.

On the side of GWs, we follow the same prescription as in our companion paper. The integrated GW energy density is constrained by Big Bang nucleosynthesis (BBN) and CMB measurements of the effective number of relativistic degrees of freedom beyond those of the SM. We denote this excess above $N_{\text{eff}}^{\text{SM}}$ [86] as

$$\Delta N_{\text{eff}} = N_{\text{eff}} - N_{\text{eff}}^{\text{SM}}, \quad N_{\text{eff}}^{\text{SM}} = 3.0440 \pm 0.0002. \quad (4.5)$$

In principle, we use a conservative limit $\Delta N_{\text{eff}} \lesssim 0.5$ as suggested in ref. [87], accounting for the uncertainty of combined BBN and CMB constraints [62, 64, 65]. In practice, however, we find this limit to be irrelevant in this study, as the corresponding bound on $h^2 \Omega_{\text{GW}}(f)$,

$$\int_{f_{\text{BBN}}}^{f_{\text{end}}} \frac{df}{f} h^2 \Omega_{\text{GW}}(f) \lesssim 5.6 \times 10^{-6} \Delta N_{\text{eff}}, \quad (4.6)$$

is never violated.

We estimate the detection prospects for a GW signal by a given GW observatory using the signal-to-noise ratio (SNR) [88–90],

$$S/N = \left(n_{\text{det}} t_{\text{obs}} \int_{f_{\text{min}}}^{f_{\text{max}}} df \left(\frac{\Omega_{\text{signal}}(f)}{\Omega_{\text{noise}}(f)} \right)^2 \right)^{1/2}. \quad (4.7)$$

The detector sets the frequency band, $[f_{\text{min}}, f_{\text{max}}]$, and if the SGWB search is based on an auto-correlation or a cross-correlation measurement, $n_{\text{det}} = 1$ or 2. For detectors that are not yet operational, we set the observation time to $t_{\text{obs}} = 1$ yr.

In our analysis, we account for three different experiments, each representative of current or next-generation detectors in their respective frequency band. As before, we use ET for the frequency band 1 Hz – 10 kHz, and LISA for the frequency band 10 μ Hz – 1 Hz [54–56]. Different from our previous analysis of the PAI model, we find that GW signals can have a sizable amplitude in the PTA frequency band, 1 nHz – 100 nHz. Correspondingly, we also estimate the potential for a GFIGW observation by the NANOGrav collaboration in their 15-year data set [57], henceforth referred to as NG15. To compute the SNR for this data set, we work with the total observing time span covered by it, $t_{\text{obs}} = 16.03$ yrs. Additionally, we consider any sensitivity by the LIGO-Virgo network as a constraint, given that it did not observe an SGWB in its third observing run [69]. We shall refer to these constraints as HLVO3 corresponding to Hanford (H), Livingston (L), and Virgo (V). These detectors cover a frequency band 10 Hz – 50 kHz [66–68]. For ET and LISA, we use the strain noise spectra Ω_{noise} collected in ref. [91]. For the HLVO3 constraints, we work with the characteristic HLV strain noise spectra available at ref. [92]. For NG15, we take the noise spectrum Ω_{noise} provided by the NANOGrav collaboration [93, 94]. This collection of noise spectra also allows us to compute power-law-integrated sensitivity (PLIS) curves [95] for HLVO3, ET, LISA, and NG15, which we show in figures 3, 5, and 7.

We solve the dynamics of inflation using the gradient expansion formalism (GEF). For more details on the GEF, see appendix A of Part I [1] and our previous works on FAI [42, 45]. We initialize the system on the slow-roll attractor, $\varphi(0) = 15.55 M_{\text{P}}$ and $\dot{\varphi}(0) = -\sqrt{2/3} m M_{\text{P}}$ assuming zero gauge-field and fermion energy densities, $\rho_{\text{EM}}(0) = \rho_{\chi}(0) = 0$. The SGWB spectra are computed based on the background evolution determined by the GEF. First, we determine the gauge-field spectra by solving eq. (2.17). This computation is performed using our newly available Gradient Expansion Formalism Factory (GEFF) [75], available at <https://github.com/richard-von-eckardstein/GEFF>. Then, we use these mode functions to compute the induced tensor power spectrum, eq. (3.7). More details on this computation are also given in Part I, appendix B.

To completely determine the modeling of FAI in our effective scheme, we need to define the heuristic scale dependence $\Theta(t, k)$ of the induced current \mathbf{J} in eq. (2.13). In ref. [45], we presented two particular choices, one which is physically well motivated, and an approximation of the same that allows for easier numerical implementation. In section 6, we explore the SGWB production in the parameter space spanned by our model parameters β and m . To make such a scan feasible, we opt for the simpler, approximate implementation of $\Theta(t, k)$. We discuss both the modeling of $\Theta(t, k)$, as well as the impact of this approximate model in appendix A. We find that the approximate modeling applied in the main body of the text does not affect our conclusions.

5 Fermion-tempered backreaction

Before we delve into our main results, i.e., estimating the detectability of GFIGWs from FAI_{SM} , we want to comment on a novel dynamical regime of FAI, which we dub *fermion-tempered backreaction*. As the name suggests, the mechanism behind fermion-tempered backreaction is linked to the damping of the gauge field by the conductive medium.

To understand the new backreaction regime, we first need to consider the usual backreaction during PAI. It typically proceeds via the gauge-friction term $I_{,\phi}\langle\mathbf{E}\cdot\mathbf{B}\rangle$ becoming large compared to Hubble friction, as quantified by the backreaction parameter

$$\delta_{\text{KG}} \equiv \frac{|I_{,\phi}(\varphi)\langle\mathbf{E}\cdot\mathbf{B}\rangle|}{|3H\dot{\varphi}|}. \quad (5.1)$$

As this friction reduces the inflaton velocity, which itself is responsible for gauge-field production, the entire system enters an oscillatory stage of deceleration and acceleration of the axion field with a retarded response by the gauge field to the change in velocity [52, 96]. During this process, the kinetic energy in the inflaton is reduced, and the energy density stored in the gauge field ends inflation by becoming dominant enough to stop the accelerated expansion.

Now consider the added impact of fermions. Once the gauge-friction term becomes relevant, the system will attempt to enter strong backreaction. However, upon completing the first cycle of oscillations in $\dot{\varphi}$, gauge-field production is enhanced, but so are the conductivity terms $\sigma_{E/B}$. Their effect is to dampen the gauge field, thus limiting the amount of friction generated for the axion. This stabilizes the oscillations of the system such that the kinetic energy is not completely drained from the axion. Instead, it performs an oscillatory motion around the original slow-roll trajectory, which increases the duration of inflation due to this stuttering motion, but without an explosive production of gauge bosons.

An example of these dynamics is given in figure 1. There, we show the evolution of the energy densities ρ_i for the individual fluids in FAI_{SM} for $\beta = 59$ and $m = 2.5 \times 10^{-5} M_{\text{P}}$. Initially, the inflaton velocity starts feeling $\langle\mathbf{E}\cdot\mathbf{B}\rangle$, inducing small oscillations due to the retarded response between the two fluids. However, every local maximum in the gauge-field abundance is mimicked by a rise in fermion production, as seen by comparing ρ_χ to ρ_E and ρ_B . The damping of ρ_E and ρ_B due to these fermions adds to the already reduced production of gauge quanta following the oscillating inflaton velocity. This results in a loss of friction for the inflaton, allowing it to accelerate and repeat this cycle. The lower panel in figure 1 supports this interpretation. There, we show the evolution of the backreaction parameter, δ_{KG} . Clearly, whenever δ_{KG} becomes close to unity, the friction term is immediately depleted, leading to temperate oscillations in all fluids.

6 Gravitational waves from FAI

To examine the detection prospects of GFIGWs from FAI_{SM} , we perform a linear parameter scan in β and $\log_{10} \frac{m}{M_{\text{P}}}$. Unlike the analysis of PAI in Part I, we find no need to perform this parameter scan in rotated coordinates. The reason for rotating the parameter space in the first analysis was that, in the PAI model, the strength of gauge-field backreaction scales as

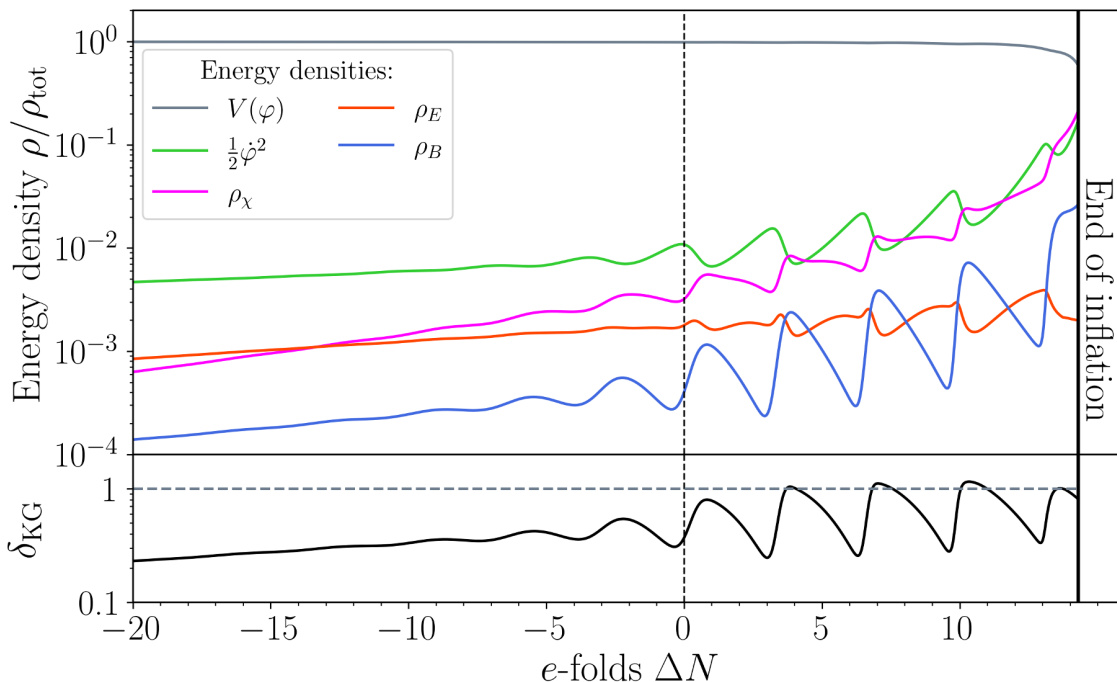


Figure 1. The evolution of FAI_{SM} for $\beta = 59$ and $m = 2.5 \times 10^{-5} M_{\text{P}}$. The upper panel shows the evolution of the energy densities $V(\varphi)$ (gray), $\frac{1}{2}\dot{\varphi}^2$ (green), ρ_E (red), ρ_B (blue), and ρ_χ (magenta) normalized to the total energy density $\rho_{\text{tot}} = 3H^2 M_{\text{P}}^2$. The lower panel shows the corresponding evolution of the backreaction parameter δ_{KG} defined in eq. (5.1). The horizontal axis indicates e -folds with respect to the expected end of slow-roll inflation at $\Delta N = 0$. The evolution of this system is representative for fermion-tempered backreaction.

$\sim \frac{m}{M_{\text{P}}} 10^{\kappa\beta}$ with some coefficient $\kappa > 1$. Thus, an increase in β needs to be compensated by a decrease in $\log_{10} \frac{m}{M_{\text{P}}}$ according to the slope κ , which we determined heuristically in Part I. However, these considerations no longer hold for FAI since the overall gauge-field amplitude is suppressed as a consequence of fermion production. Hence, also backreaction is inhibited, altering the direct dependence between β and m . In fact, the well-known analytical formula for the induced tensor power spectrum of PAI, $\mathcal{P}_{T,\lambda}^{\text{ind}} \propto H^4 \exp(4\pi\xi)$, no longer applies to FAI due to this dampened production. The generic scaling of $\mathcal{P}_{T,\lambda}^{\text{ind}}$ with $H^4 \sim m^4$ still holds, however, as it is a consequence of eq. (3.7), and is independent of the precise mode functions $A_\lambda(t, k)$. Conversely, the instability parameter ξ is inhibited by gauge-field production (see section 2). Consequently, we expect that increasing the inflaton mass enhances $\mathcal{P}_{T,\lambda}^{\text{ind}}$, while an increase of β has a reduced effect compared to the PAI case.

Following these considerations, the results in this section are based on a parameter scan over inflaton masses ranging between $10^{-7.5} - 10^{-5.5} M_{\text{P}}$ and β between 20 – 60.

6.1 Instantaneous reheating

Let us first focus on the observational prospects for detecting GFIGWs from FAI_{SM} assuming instantaneous reheating. For a discussion of a lowered reheating temperature, T_{reh} , see section 6.2. Our findings are summarized in figure 2, where we indicate regions with

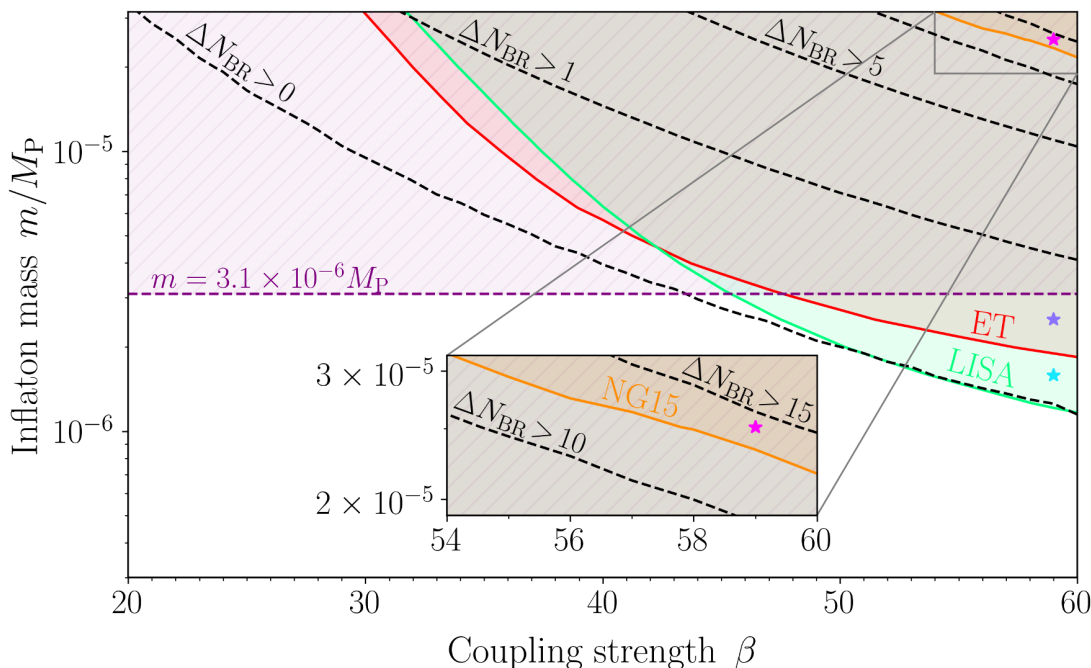


Figure 2. Regions with $S/N > 1$ for ET (red), LISA (green), and NG15 (orange) and instantaneous reheating. PLANCK constraints on the inflaton mass are shown as a purple shaded region. In dashed black, we show contour lines for ΔN_{BR} , the prolonged duration of inflation due to gauge-field backreaction. The benchmark points marked in cyan, violet, and magenta correspond to the GW spectra shown in figure 3. For $\beta \gtrsim 45$, we find a viable region for a detection by ET or LISA, while an FAI interpretation of the NG15 signal appears to be excluded by PLANCK.

$S/N > 1$ for ET, LISA, and NG15 in red, green, and orange, respectively, together with $m \lesssim 3.1 \times 10^{-6} M_P$, the CMB bound, as a purple shaded region. Evidently, for $\beta \gtrsim 30$, FAI_{SM} supports a sufficiently strong gauge field such that the resulting GWs would be detectable by both LISA and ET. In fact, we find a large overlap between regions of $S/N > 1$ for ET and LISA, indicating that the GW amplitude spans a large frequency range. For $\beta \gtrsim 54$ and $m \gtrsim 2 \times 10^{-5} M_P$, the signal can even extend down to the nanohertz regime, such that it is potentially observed by NG15. However, to support large enough amplitudes in Ω_{GW} , the inflaton mass needs to be sufficiently high. Hence, a large part of the LISA and ET region of interest is in tension with the CMB bound on m , but a small viable region still remains for $\beta \gtrsim 45$. The region of interest for NG15 is clearly excluded by this bound. In contrast to the CMB bounds, we find that neither ΔN_{eff} nor HLVO3 impose any constraint in the region of parameter space that we analyze.

Alongside the SNR regions in figure 2, we also indicate the relevance of gauge-field backreaction by showing contour lines for the extended duration of inflation due to backreaction, ΔN_{BR} . As expected, larger masses and couplings increase the relevance of backreaction. However, in stark contrast to our findings for PAI, strong backreaction is not required to reach $S/N > 1$. This is clearly a result of efficient fermion damping prohibiting backreaction even when increasing the inflaton mass. However, even when backreaction effects do arise, their impact during FAI is manifestly different. For PAI, we found that strong backreaction

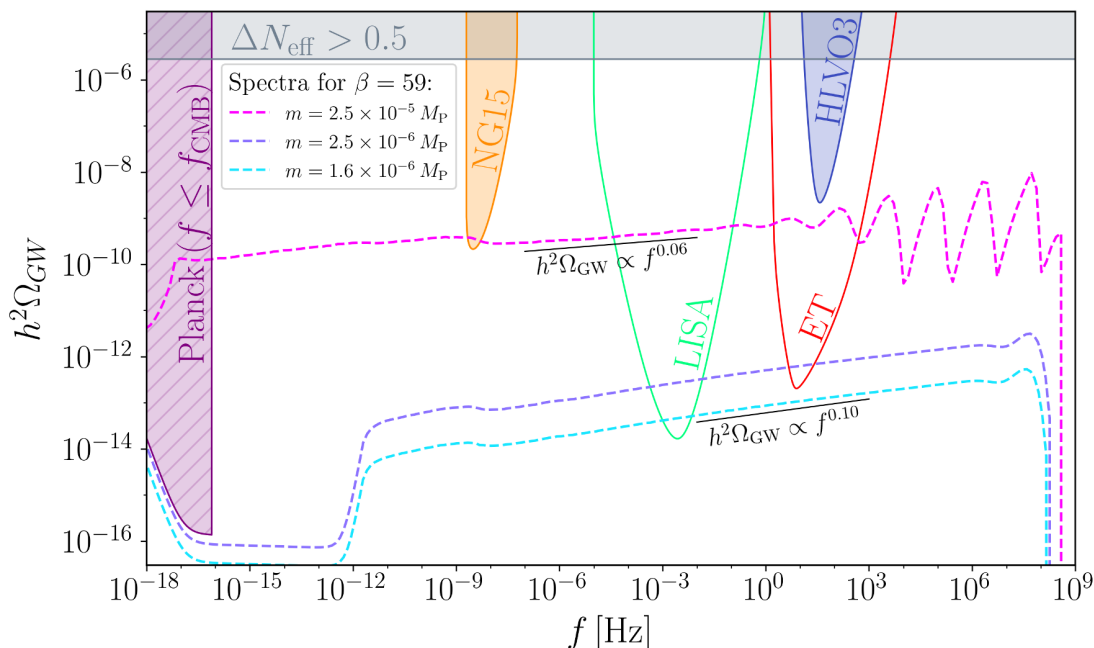


Figure 3. GW spectra for the benchmark points $\beta = 59$ with $m = 1.5 \times 10^{-6} M_{\text{P}}$ (cyan), $m = 2.5 \times 10^{-6} M_{\text{P}}$ (violet), and $m = 2.5 \times 10^{-5} M_{\text{P}}$ (magenta). PLIS curves are shown for ET, LISA, and NG15 with the same colors as in figure 2. In addition, we indicate constraints from HLVO3 in blue and from PLANCK (see text for more details) in purple. Clearly, all spectra avoid ΔN_{eff} constraints. However, the magenta spectrum is in clear conflict with the PLANCK observations.

inevitably leads to an overproduction of GWs in tension with $\Delta N_{\text{eff}} \lesssim 0.5$. However, these constraints are entirely irrelevant for FAI_{SM} , even for an extremely increased duration of inflation, $\Delta N_{\text{BR}} \gtrsim 15$. This is a consequence of the fermion-tempered backreaction that we discussed in section 5. It allows for a delayed end of inflation without being accompanied by a drastic production of gauge bosons.

The region of interest for NG15 notably coincides with $\Delta N_{\text{BR}} \gtrsim 10$. Given that the CMB bound, eq. (4.4), depends on ΔN_{BR} , one may wonder if this backreaction sufficiently weakens the constraints on m , such that the NG15 region of interest would become viable. However, even for $\Delta N_{\text{BR}} = 20$, the bound is still far too constraining; $m \lesssim 4 \times 10^{-6} M_{\text{P}}$.

In figure 2, we highlight three benchmark points for $\beta = 59$ with $m = 1.5 \times 10^{-6} M_{\text{P}}$, $2.5 \times 10^{-6} M_{\text{P}}$ and $2.5 \times 10^{-5} M_{\text{P}}$ in cyan, violet, and magenta, respectively. The full GW spectra for these points are presented in figure 3 alongside the PLIS curves for ET, LISA, and NG15. In addition, we show constraints from HLVO3 in terms of its PLIS curve. Constraints from PLANCK on the tensor-to-scalar ratio are also indicated. We do so in a particular way that requires further explanation: as discussed in section 4, we only demand that the inflaton potential be of the form $V(\phi) = m^2 \phi^2 / 2$ for field values below some threshold value, $\phi \leq \phi_{\text{thr}}$, while at field values above this threshold, the inflaton potential is free to receive model-dependent corrections. The upper limits on $V(\phi_{\text{thr}})$ in eq. (4.4) notably apply to this more flexible model, in which $V(\phi) = m^2 \phi^2 / 2$ for $\phi \leq \phi_{\text{thr}}$ and $V(\phi) = V_{\text{CMB}}(\phi)$ for $\phi > \phi_{\text{thr}}$, where $V_{\text{CMB}}(\phi)$ represents a monotonically increasing continuation of the quadratic

potential at small values, but which is otherwise arbitrary. In figure 3, on the other hand, we need to make a choice for $V_{\text{CMB}}(\phi)$. For definiteness, we simply assume that, for the purposes of drawing GW spectra at all frequencies in figure 3, the potential is of the form $V(\phi) = m^2\phi^2/2$ at all field values, including field values above φ_{thr} (frequencies below f_{thr}). This particular model may be referred to as the $m^2\phi^2$ -equivalent of the actual, more flexible model that we assume in our discussion of parameter space in figure 2. The CMB constraints in figure 3 therefore should be interpreted as follows: if an $m^2\phi^2$ -equivalent GW spectrum does or does not violate the CMB bound shown in figure 3, the corresponding more flexible model does or does not violate the observational constraints on A_s , r , and $|\xi_{\text{CMB}}|$, respectively. This construction has several consequences. First of all, the CMB limit in figure 3 is not a model-independent sensitivity curve on the same footing as the PLIS curves in the plot. By construction, it represents a specific constraint that only applies to the GW spectra in the $m^2\phi^2$ -equivalent model. Second, these $m^2\phi^2$ -equivalent GW spectra do not need to satisfy the observational limit on the amplitude of the tensor power spectrum, $rA_S \lesssim 0.03 \times 2.1 \times 10^{-9}$; this bound applies to the GW spectra in our actual, more flexible model, in which $V_{\text{CMB}}(\phi)$ is left unspecified. To determine the upper (“would-be”) limit on rA_S in the $m^2\phi^2$ -equivalent model, we work with eq. (4.4) together with eq. (4.3) evaluated for instantaneous reheating and $f_{\text{thr}} = 10^{-12}$ Hz. The resulting excluded region is then constructed by assuming a flat tensor power spectrum up to $f_{\text{CMB}} = 7.7 \times 10^{-17}$ Hz, which we plug into eq. (3.8).

Comparing figure 3 to its counterpart In Part I [1], figure 3, confirms our previous description of the differences between PAI and FAI. For PAI, we found peaked spectra with a strongly blue tilted GW spectrum, surpassing the vacuum contribution by many orders of magnitude. For FAI_{SM} on the other hand, GFIGWs are abundant enough to be observable by ET, LISA, or even NG15, but never in threat of violating bounds on ΔN_{eff} . Instead of the drastic amplification of GWs due to strong backreaction, a temperate amplitude in GFIGWs is sustained over many orders of magnitude in frequency. A first initial burst in gauge-field production lifts the spectrum above the vacuum contribution, where it then remains with only a mild blue tilt. This is clearly a consequence of the dampened efficiency of gauge-field production due to Schwinger pair creation.

The two lower-mass spectra, $m = 1.5 \times 10^{-6} M_{\text{P}}$ and $2.5 \times 10^{-6} M_{\text{P}}$, are nearly identical, except for the increase in amplitude for the higher-mass spectrum. For $f \gtrsim 10^{-11}$ Hz, the spectrum rises with a mild blue tilt, $h^2\Omega_{\text{GW}} \propto f^{0.1}$ (except for a small decrease due to the change in $g_{*,S}$ around the quantum chromodynamics (QCD) crossover at $f \sim 10^{-9}$ Hz). The spectra not only reach the sensitivity curves for ET and LISA, but are also safe regarding CMB constraints at $f \sim f_{\text{CMB}}$. The GFIGW contribution only becomes important at $f \sim 10^{-12}$ Hz and the vacuum contribution is not in tension with the upper limit on the tensor-to-scalar ratio, $r \lesssim 0.03$, as indicated by the purple shaded region in figure 2.

The spectrum at high mass, $m = 2.5 \times 10^{-5} M_{\text{P}}$, corresponds to the evolution of the energy density shown in figure 1. It evidently is the product of a background evolution with significant gauge-field backreaction. This observation allows us to explain all the apparent differences between this spectrum and the ones for $m = 1.5 \times 10^{-6} M_{\text{P}}$ and $2.5 \times 10^{-6} M_{\text{P}}$. Firstly, note the drastic increase to the duration of inflation, $\Delta N_{\text{BR}} \simeq 14$, as inferred from figures 1 and 2. It implies that the GW frequency is redshifted by a factor of 10^5 , explaining

why the rise in this GW spectrum happens at much lower frequencies than for the other two spectra. Secondly, we can see novel oscillations in $h^2\Omega_{\text{GW}}$ at high frequencies mimicking those of the gauge-field energy densities in figure 1.

The rise in the GFIGW amplitude for $m = 2.5 \times 10^{-5}M_{\text{P}}$ around f_{CMB} clearly is in conflict with PLANCK constraints. However, we want to stress that this parameter point was simulated assuming a perfectly quadratic inflaton potential at all field values. If we were to flatten the inflaton potential at early times, one would expect gauge-field production to set in later due to the lower inflaton velocity on the flatter sections. Consequently, the production of GFIGWs would be observed at higher frequencies. Of course, this would not save the GW spectrum presented here from being ruled out, as even the vacuum contribution for this spectrum is too high, which can be both read off from figure 2 and figure 3, remembering our special construction of the CMB exclusion contour in figure 3. Still, this model dependence is worth remembering when viewing these GW spectra.

However, even when disregarding PLANCK constraints, the spectrum for $m = 2.5 \times 10^{-5}M_{\text{P}}$ would still not be suitable for explaining the NG15 data, although it reaches a sufficient amplitude. The NG15 data favors a steeper spectral index, $h^2\Omega_{\text{GW}} \propto f^{5-\gamma}$, with $\gamma = 3.2 \pm 0.6$ (median and 90% credible interval) [57]. Clearly, the spectrum we show falls short of this spectral slope with $h^2\Omega_{\text{GW}} \propto f^{0.06}$ around the relevant frequencies. In fact, the background slope of our spectrum is so small that the changes in the effective number of relativistic degrees of freedom around the QCD phase transition even lead to momentary drops of the spectral index below zero (i.e., $\gamma > 0$), which is disfavored by the NG15 data.

In summary, we find that, while PAI is primarily constrained by ΔN_{eff} due to the sudden and drastic transition into the strong-backreaction regime, this is no longer true for FAI. The additional friction from Schwinger pair production tempers these violent processes, allowing for observable signals by both ET and LISA. However, these same damping effects also imply that a larger Hubble rate at production is required in order to supply the necessary amplitude in GWs. Therefore, we find that the most relevant constraints on the SGWB from FAI come from the PLANCK measurements of A_S and the limit on r , in contrast to PAI, where the most stringent constraints came from ΔN_{eff} .

6.2 Varying the reheating temperature

To elaborate on our findings in the previous section, we include a simple modeling of reheating effects, assuming a reheating phase of early matter domination, i.e., $w_{\text{reh}} = 0$. We treat the reheating temperature, T_{reh} , as a free parameter, which we can lower down to $T_{\text{reh}} \sim T_{\text{BBN}} \sim 1 \text{ MeV}$. Of course, this is a generous assumption and the precise initial conditions present at the end of inflation are relevant here. For certain parameter regions of FAI_{SM}, especially for high couplings β and masses m , we find that a large fraction of energy is already stored in SM fermions and $U(1)_Y$ gauge bosons. Therefore, one would assume that the duration of reheating may not be an arbitrary parameter. However, for illustrating the effect of reheating, it is easiest to treat all parameter points on equal footing, instead of setting a lower threshold on T_{reh} on a point-by-point basis.

In figure 4, we show the change in the $S/N > 1$ regions for LISA, ET, and NG15 when lowering the reheating temperature to $T_{\text{reh}} = 10^9 \text{ GeV}$, 10^6 GeV , 10^3 GeV , and 1 GeV .

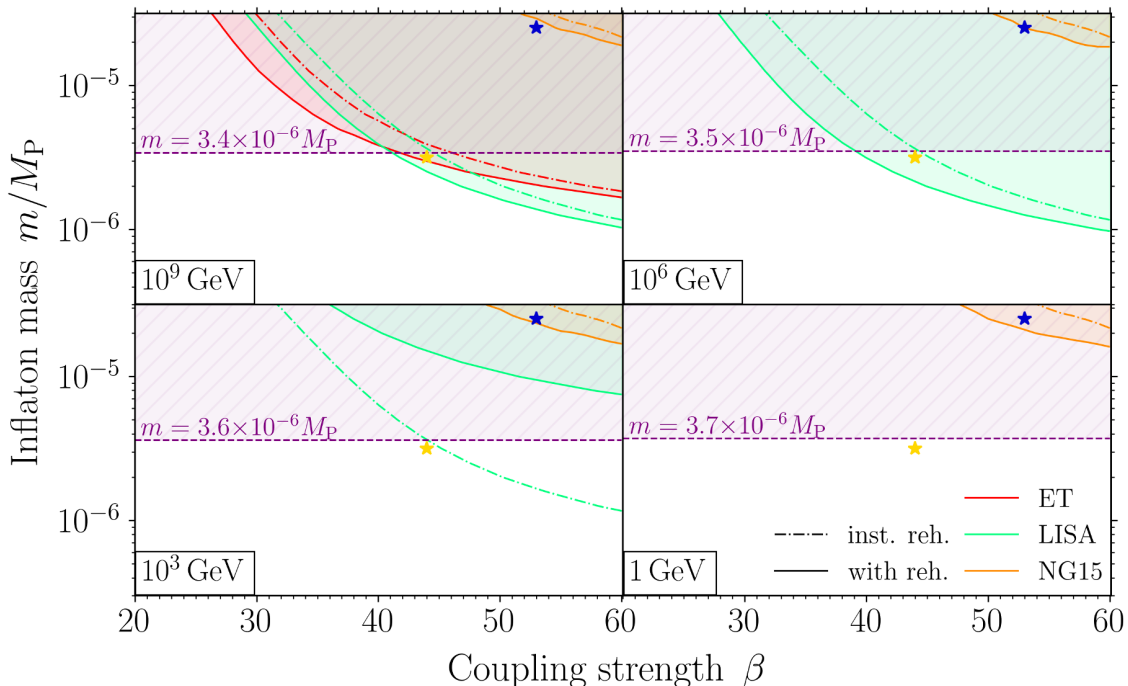


Figure 4. Regions with $S/N > 1$ for ET (red), LISA (green), and NG15 (orange) including variations in the reheating temperature, T_{reh} . From left to right and top to bottom, T_{reh} is 10^9 GeV, 10^6 GeV, 10^3 GeV, and 1 GeV as indicated in the lower left corner of each panel. For reference, the contours of $S/N > 1$ assuming instantaneous reheating (i.e., the ones in figure 2) are shown as dashed-dotted lines. A lower T_{reh} first increases the sensitivity of a detector, until it falls below the temperature corresponding to the frequency band of the detector. For points marked in yellow, $\beta = 44$, $m = 3.2 \times 10^{-6} M_{\text{P}}$, and blue, $\beta = 53$, $m = 2.5 \times 10^{-5} M_{\text{P}}$, we show $h^2 \Omega_{\text{GW}}(f)$ in figure 5.

For comparison, we also show the results of figure 2 in dashed-dotted lines. Evidently, lowering the reheating temperature enlarges the SNR regions before they vanish entirely. This is easily understood: lowering T_{reh} corresponds to redshifting the blue-tilted spectrum. Correspondingly, a higher amplitude in $h^2 \Omega_{\text{GW}}$ can be achieved for lower frequencies. However, a lowered reheating temperature also implies a drop-off at the frequency corresponding to T_{reh} . Therefore, once the reheating temperature drops below temperatures corresponding to the frequency band of a detector, the detector loses all sensitivity. This happens first for ET, with the lower frequency roughly corresponding to $T \sim 4 \times 10^7$ GeV. Then, LISA loses sensitivity around $T \sim 400$ GeV. For NG15, we do not observe this in figure 4 as its critical temperature is $T \sim 100$ MeV, i.e., below $T_{\text{reh}} = 1$ GeV, the lowest temperature we consider in figure 4. We also indicate the mild drift in the PLANCK limits on the inflaton mass, eq. (4.4), for $\Delta N_{\text{BR}} = 0$ as indicated by the respective label.

To elaborate on these results, we again indicate two qualitatively different benchmark points in figure 4, $\beta = 44$, $m = 3.2 \times 10^{-6} M_{\text{P}}$, and $\beta = 53$, $m = 2.5 \times 10^{-5} M_{\text{P}}$, for which we show the effect of reheating on $h^2 \Omega_{\text{GW}}(f)$ in figure 5. For each parameter point, the spectrum for instantaneous reheating is depicted in solid black, while the colored dashed curves show the effect of lowering T_{reh} . PLIS curves are given exactly as in figure 3. We again

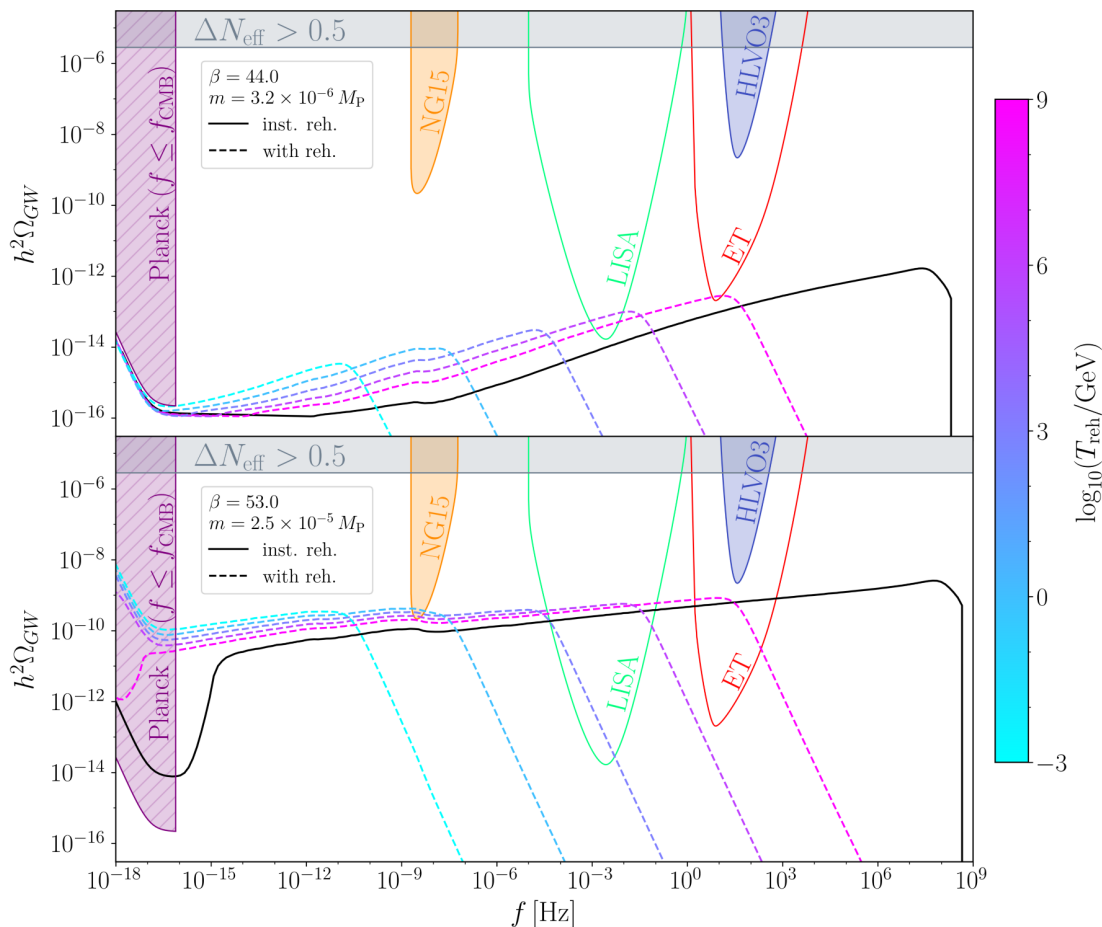


Figure 5. GW spectra for $\beta = 44$, $m = 3.2 \times 10^{-6} M_{\text{P}}$ (upper panel) and $\beta = 53$, $m = 2.5 \times 10^{-5} M_{\text{P}}$ (lower panel) including their variations when lowering the reheating temperature. The black solid spectra correspond to instantaneous reheating, while the color of the dashed spectra correspond to $T_{\text{reh}} = 10^9 \text{ GeV}$, 10^6 GeV , 10^3 GeV , 1 GeV , and 1 MeV , respectively; see the color bar on the right. PLIS curves and PLANCK constraints are shown with the same color code as in figure 3.

compute the PLANCK bound as in figure 3, but assuming the lowest reheating temperature for which we show spectra, $T_{\text{reh}} = 1 \text{ MeV}$. The spectra match our expectations: a lower reheating temperature redshifts the spectrum, first increasing the sensitivity of a detector, before the spectral peak is redshifted out of its frequency band for a low enough T_{reh} .

6.3 Estimating the fermionic contribution to the SGWB

Up until now we have exclusively considered GWs sourced by gauge bosons. However, figure 1 indicates that more energy density may be stored in fermions than the gauge field at the end of inflation. If this fermion gas is sufficiently anisotropic, one would expect that it contributes correspondingly to the total amplitude of the GW spectrum.

A computation of this additional contribution, on equal footing to the one for the gauge field, is beyond the scope of this article. However, we can make an educated estimate of

its relevance,

$$\Omega_{\text{GW}}^{\chi} \sim \beta_* \left(\frac{\rho_{\chi}}{\rho_{\text{EM}}} \right)^2 \Omega_{\text{GW}}^{\text{GF}}. \quad (6.1)$$

Here, $\Omega_{\text{GW}}^{\chi}$ is the contribution from fermions to the GW spectrum, while $\Omega_{\text{GW}}^{\text{GF}}$ is that of the gauge field, which we compute from eq. (3.7). We assume that the overall spectral dependence for the fermion contribution matches that of the gauge field, simply relating the two contributions via their respective energy densities. This reflects the fact that we model the conducting fermionic medium as direct functions of the gauge-field expectation values, $\langle \mathbf{E}^2 \rangle$, $\langle \mathbf{B}^2 \rangle$, and $\langle \mathbf{E} \cdot \mathbf{B} \rangle$. The coefficient β_* describes how inherently anisotropic (quadrupolar and beyond) the fermion energy density is, i.e., its ability to source GWs.

Next, we try to estimate β_* from the microphysics hidden behind ρ_{χ} . In ref. [32], it was computed that the electric and magnetic field source two distinct fermion populations. Firstly, chiral fermions are directly produced from the electric field via the chiral anomaly. These fermions quickly thermalize due to internal scattering, which we take to imply that they quickly become isotropic and are thus irrelevant for GWs. However, a second population of left-right symmetric fermions occupies higher Landau levels due to the presence of the magnetic field. For these fermions, scattering is negligible, and they do not thermalize. Reference [32] estimated that this second population is dominant in ρ_{χ} , finding that the chiral fermions only make up a small fraction of around $10^{-2} - 10^{-3}$. We take this to imply that, in principle, a large fraction of ρ_{χ} may be anisotropic. We will thus assume an optimistic value, $\beta_* = 1$ for our estimates.

In figure 6, we show the relevance of the fermionic contribution by showing how the SNR-regions for LISA, ET, and NG15 are affected when adding the estimate for $h^2\Omega_{\text{GW}}^{\chi}$ from eq. (6.1) to $h^2\Omega_{\text{GW}}^{\text{GF}}$ from eq. (3.7). All results are computed assuming instantaneous reheating. The $S/N > 1$ regions enveloped by dotted lines indicate the result with $h^2\Omega_{\text{GW}}^{\chi}$ included, while the dashed-dotted lines are computed from $h^2\Omega_{\text{GW}}^{\text{GF}}$ alone. Besides LISA, ET, and NG15, also HLVO3 now shows sensitivity to $h^2\Omega_{\text{GW}}$.

LISA and ET become considerably more sensitive to the GW signal from FAI, especially for large coupling, β . The explanation is obvious: stronger gauge-field production yields stronger fermion production. Interestingly, we find that the sensitivity region for NG15 has not been significantly increased. We attribute this to the observation that fermion production occurs towards the end of inflation, as suggested by figure 1. Therefore, at low frequencies, $h^2\Omega_{\text{GW}}^{\chi}$ does not dominate over $h^2\Omega_{\text{GW}}^{\text{GF}}$, and the fermions do not add to the signal in the nanohertz band. The fact that the fermions contribute primarily at high frequencies also explains the appearance of an $S/N > 1$ region for HLVO3: because $h^2\Omega_{\text{GW}}^{\chi}$ is more blue-tilted than $h^2\Omega_{\text{GW}}^{\text{GF}}$, $h^2\Omega_{\text{GW}}$ gains in amplitude at higher frequencies. Based on our simple estimates alone, including $h^2\Omega_{\text{GW}}^{\chi}$ thus further weakens the FAI interpretation of the NG15 signal, as the HLVO3 region largely overlaps with the one for NG15.

These simple estimates clearly highlight the importance of a more refined computation of the fermionic GW contribution, $h^2\Omega_{\text{GW}}^{\chi}$, seeing as it may be dominant over $\Omega_{\text{GW}}^{\text{GF}}$ at high frequencies. A precise computation including the spectral information on the fermions is, however, beyond the scope of this analysis.

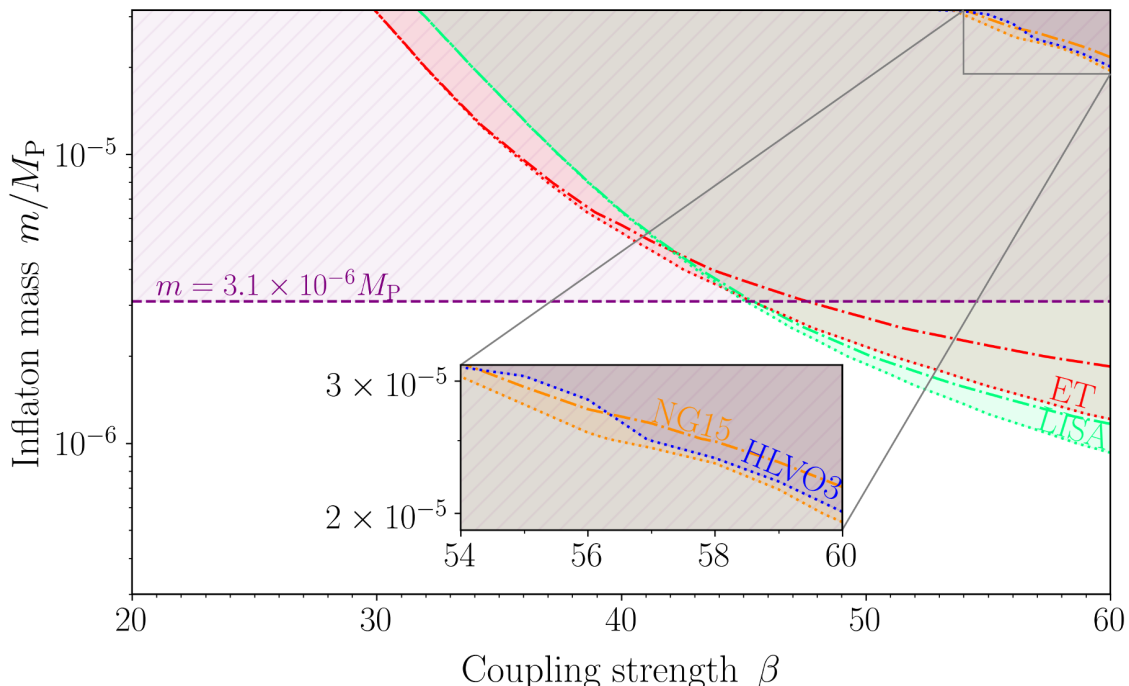


Figure 6. Sensitivity regions when estimating the additional contribution to $h^2\Omega_{\text{GW}}$ due to the fermion gas. The regions with $S/N > 1$ are shown for ET (red), LISA (green), NG15 (orange), and HLVO3 (blue). These results are computed assuming instantaneous reheating. The fermion contribution may boost the detectability region for ET and LISA, thus dominating over the GFIGW contribution at high frequencies. At low frequencies, the extra contribution matters little, as evident from the NG15 region being nearly insensitive to the additional source. More so, the enhanced signal at high frequencies implies that HLVO3 is in conflict with a FAI interpretation of the NG15 signal. These results are a simple estimate assuming an optimistically large anisotropic fermion fraction.

7 Conclusion and outlook

Axion inflation is a promising model of inflation that attempts to reconcile the flatness of the inflaton potential with non-trivial couplings of the inflaton to gauge fields. In the simplest models, an Abelian gauge field couples to the inflaton via a shift-symmetric interaction, $\phi F\tilde{F}$. This results in abundant gauge-field production that may in turn source a stochastic gravitational-wave background (SGWB). Models of fermionic axion inflation (FAI) extend the field content of axion inflation by charged matter fields. These charged particles are generated via the Schwinger effect as a consequence of the strong electric and magnetic fields sourced by the axion. The resulting conductive medium of charge carriers then dampens the production of the gauge field.

In this article, we investigated the imprint of this non-trivial interplay between the inflaton, the gauge field, and fermions onto the SGWB sourced by the gauge field during axion inflation. To this end, we considered a particularly well-motivated realization of the FAI model in which the axion couples to the Abelian $U(1)_Y$ hypercharge field. Therefore, the fermions produced by the Schwinger effect are those of the Standard Model (SM). This article

is the second part following a preceding analysis of gravitational-wave (GW) production during pure Abelian axion inflation (PAI) [1].

To estimate the phenomenological viability of the resulting SGWB signal, we considered its detectability by current and next-generation GW observatories. This includes experiments spanning a large frequency range, the Laser Interferometer Space Antenna (LISA), the Einstein Telescope (ET) as well as the latest data set of the NANOGrav collaboration (NG15), representative of the sensitivity of pulsar timing array observations. The constraints on the signal come from the non-observation of an SGWB in the third observing run of the LIGO-Virgo network (HLVO3), PLANCK measurements, and bounds on the effective number of dark relativistic degrees of freedom, ΔN_{eff} .

We studied the production of GWs from FAI assuming a linear coupling between the inflaton and the gauge field, $(\beta/M_{\text{P}})\phi F\tilde{F}$, and assuming a simple quadratic potential for the inflaton. In the two-dimensional parameter space spanned by the inflaton-gauge-field coupling strength, β , and the inflaton mass, m , we find large regions where an appreciable amount of gauge-field-induced gravitational waves (GFIGWs) may be produced during FAI coupled to the SM. Especially for couplings $\beta \gtrsim 45$ and inflaton masses below the PLANCK bound, $m \lesssim 3.1 \times 10^{-6} M_{\text{P}}$, an SGWB with sufficient amplitude can be generated such that it would be observable by both LISA and ET, thus marking a particular region of interest. To reach the sensitivity range of NG15, larger couplings, $\beta \gtrsim 54$, and larger masses, $m \gtrsim 2 \times 10^{-5} M_{\text{P}}$, are required, in tension with existing PLANCK data. However, even when disregarding the bound on the inflaton mass, the spectral slope of spectra reaching NG15 sensitivity are too shallow to be consistent with NG15; $h^2 \Omega_{\text{GW}} \propto f^{0.06}$.

We also analyzed the effect of a lowered reheating temperature, T_{reh} , finding qualitative agreement with the results for instantaneous reheating. Unsurprisingly, when the reheating temperature is lowered below the corresponding frequency range of a detector, the detector loses sensitivity. This occurs for ET at $T_{\text{reh}} \sim 4 \times 10^7$ GeV, then for LISA at $T_{\text{reh}} \sim 400$ GeV and again for NG15 at $T_{\text{reh}} \sim 100$ MeV.

Finally, we estimated the additional contribution to the SGWB that is sourced by the SM fermions produced during FAI. We find that this novel component to the GW spectrum could enhance the signal at high frequencies, as the fermion energy density surpasses that of the gauge field towards the end of inflation. This implies an increased sensitivity of LISA and ET to GWs from FAI. At the same time, this additional component of the SGWB puts further stress on the interpretation of the NG15 signal as coming from FAI. We find that, as the SGWB signal rises more steeply at high frequencies due to the fermionic contribution, most signals that would explain the NG15 data should also have been detected in HLVO3. Therefore, the region of parameter space with an appreciable signal-to-noise ratio for NG15 is in tension with the non-detection of an SGWB by HLVO3. These results rely on a simple estimate (see eq. (3.4) of the SGWB signal with an optimistic assumption on the anisotropy of the fermionic gas. The relevance of the fermionic contribution based on these simple estimates clearly highlights the importance of a proper computation of this novel SGWB component. We leave this for future work.

The results for FAI are in stark contrast to the results for PAI in our companion paper [1]. In our GEF benchmark study, we found that an observable signal for PAI was in tension

with bounds on ΔN_{eff} , while a signal from FAI is primarily constrained by the maximal inflation scale inferred from the PLANCK data. The reason for this difference is in the uninhibited production of gauge bosons during PAI, which implies that strong backreaction effects can trigger a dynamical instability in the dynamics of the inflaton, leading to an explosive production of gauge bosons and GWs. The damping effect of the fermionic medium in FAI tempers these effects, allowing for the generation of a moderately blue-tilted GW spectrum. We find that, in the most interesting region of parameter space, the SGWB amplitude grows as $\propto f^{0.1}$ around the LISA and ET frequency bands. While this damping does not imply the absence of backreaction, it does moderate it, resulting in a qualitatively new backreaction regime that we call fermion-tempered backreaction.

Both of our analyses were performed using the gradient expansion formalism (GEF). In our previous paper, we discussed at length the limitations of this method when studying the dynamical evolution of PAI. These limitations concern the generation of axion gradients sourced by the gauge field. The effect of this additional dynamical degree of freedom has so far only been successfully captured on the lattice. However, the dampened production of gauge bosons during FAI implies that this effect could be less significant, as recently indicated by lattice simulations [46]. Therefore, the GEF appears to be ideally suited for studying the dynamics of FAI.

In summary, in this article, we studied the prospects of detecting GWs originating from the production of SM particles during axion inflation. We find that a GW signal sourced by $U(1)_Y$ hypercharge bosons generated via their coupling to the axion-inflaton field would be detectable by both LISA and ET. We also estimated, for the first time, the additional contribution to the GW spectrum sourced by the fermions produced during FAI, finding it to make a relevant contribution at high frequencies.

We hope that our results are going to encourage other researchers to further pursue the study of fermion generation in the early Universe, in particular during axion inflation. Our results highlight the importance of accounting for the non-trivial interplay between the fermions and the gauge field in this model to make accurate phenomenological predictions. The fact that an SGWB sourced by SM fields during inflation may be detected by next-generation GW observatories underpins the relevance of improving various aspects of the computation that we have performed here. First, one should strive towards computing fermion backreaction onto the gauge field from first principles in place of the effective modeling that we applied in this article. Second, the importance of axion inhomogeneities should be reassessed for the model of FAI to ensure that their impact can indeed be neglected. In particular, we hope our results will spark the interest of researchers to validate this assumption using lattice techniques. Third, in this article, our focus has only been on the gauge field as a source for an SGWB. We already pointed out that the fermions themselves could give a relevant contribution to the SGWB. However, these same fields can also source scalar metric perturbations, which in turn can induce an additional SGWB component at second order in perturbation theory. This third major source could then further affect the spectra that we have shown here. At the present stage, we are not aware of any computations of the scalar power spectrum for FAI, in particular not one that also accounts for fermions as a source of density fluctuations. Overall, FAI is a rich research field with the potential of many important improvements to be made in future work.

Acknowledgments

K. S. is an affiliate member of the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU) at the University of Tokyo and as such supported by the World Premier International Research Center Initiative (WPI), MEXT, Japan (Kavli IPMU). The work of O.S. has received funding through the SAFE — Supporting At-Risk Researchers with Fellowships in Europe project, which is funded by the European Union. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union, the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them. Part of this work was conducted using the High Performance Computing Cluster PALMA II at the University of Münster (<https://www.uni-muenster.de/IT/HPC>).

A Alternative modeling of fermionic scale dependence

In this appendix, we discuss the modeling of the function $\Theta(t, k)$ in eq. (2.13) by means of which we account for the scale dependence of Schwinger damping in the mode equations for the gauge field. We compare two models suggested in ref. [45], which we call k_h -damping and k_S -damping, respectively. In the main body of the text, we relied on k_h -damping, even though this model is an approximation of the more well-motivated k_S -damping model. In the following, we will first introduce both models and then discuss how our results in the main text change when replacing k_h -damping by k_S -damping on the basis of the benchmark points shown in figure 3, i.e., $\beta = 59$ and $m = 1.5 \times 10^{-6} M_P$, $2.5 \times 10^{-6} M_P$, and $2.5 \times 10^{-5} M_P$.

A.1 Scale-dependent damping with k_h or k_S

In the k_S -damping model, the function $\Theta(t, k)$ is given by

$$\Theta(t, k) = \theta(k_S(t) - k)\theta(k_S(t) - a(t)H(t)). \quad (\text{A.1})$$

Here, the wavenumber $k_S(t)$ characterizes the typical particle-antiparticle separation following their creation. Hence, the first factor in eq. (A.1) ensures that gauge-field modes with $k > k_S$ are not affected by the conductive medium, while the second factor prohibits super-Hubble particle creation.

The scale k_S can be determined from a kinematic analysis. For FAI_{SM} , one has [45]

$$k_S = a \frac{C^{1/3} g'(\mu)}{2^{1/4}} \left[\langle \mathbf{E}^2 \rangle - \langle \mathbf{B}^2 \rangle + \sqrt{\langle \mathbf{E}^2 - \mathbf{B}^2 \rangle^2 + 4 \langle \mathbf{E} \cdot \mathbf{B} \rangle^2} \right]^{1/4}. \quad (\text{A.2})$$

However, as explained in ref. [45], the k_S -damping model is more challenging to implement in the GEF. In particular, in the present analysis, we found it to be unsuitable for a parameter scan like the one in section 6. The challenge arises as there is a second important scale affecting gauge-field evolution besides the damping scale k_S : the instability scale k_h , which determines if a gauge-field mode $A_\lambda(t, k)$ is tachyonically amplified. It is given by [45]

$$k_h(t) = \max_{t' < t} \left\{ a(t') H(t') \left(|\xi_{\text{eff}}| + \sqrt{\xi_{\text{eff}}^2 + s_E^2 + s_E} \right) \right\}, \quad s_E = \frac{\sigma_E \Theta}{2H}. \quad (\text{A.3})$$

Since k_h and k_S are evidently different, when modeling $\Theta(t, k)$ as in eq. (A.1), one consistently needs to distinguish between gauge-field modes that have already experienced the tachyonic instability before feeling the presence of the damping conductive medium, and those modes that are damped before being tachyonically amplified. Therefore, the GEF modeling becomes significantly more involved, as we have explained in detail in ref. [45].

To facilitate the numerical implementation of $\Theta(t, k)$, one can make the approximate identification $k_h \simeq k_S$, such that⁵

$$\Theta(t, k) = \theta(k_h(t) - k)\theta(k_S(t) - a(t)H(t)). \tag{A.4}$$

This second choice is what we refer to as the k_h -damping model.

Just comparing eqs. (A.2) and (A.3), it may not be evident why this identification is justified. One approach is to verify the assumption *a posteriori*. However, given that both quantities scale with aH , it is not surprising that one will find them to differ only by an $\mathcal{O}(1)$ factor, which is insignificant when considering the exponential growth of the scale factor during inflation.

A.2 Comparison between k_h - and k_S -damping

In order to justify that we rely on the simpler k_h -damping model in the main body of the text, we now study how our main results are affected by this choice.

First, we re-analyze the onset of the regime of fermion-tempered backreaction by comparing the results for the benchmark point $\beta = 59$ and $m = 2.5 \times 10^{-5} M_P$ as given in figure 1. The updated version of this plot can be seen in figure 7, where we overlay the previous results for k_h -damping (solid lines) with those for k_S -damping (dotted lines). Notably, the backreaction qualitatively remains the same: an oscillation of the inflaton’s kinetic energy around the slow-roll attractor accompanied by oscillations in ρ_χ and ρ_{EM} . However, we do see that backreaction sets in marginally earlier in the k_S -damping model. We can understand this by considering that $k_S < k_h$ at early times, and, hence, some gauge modes $A_\lambda(t, k)$ will be tachyonically amplified without immediately feeling the damping effect due to the fermion fluid. This implies that the integrated gauge field can start to affect the inflaton dynamics earlier. At the same time, the end of inflation occurs marginally earlier in the k_S -damping model (the vertical dotted line at $\Delta N \simeq 13$). We may interpret this by considering that, once $k_S > k_h$, some gauge modes will be damped even before feeling the enhancing effect of the rolling inflaton field. Overall, we are reassured in the fact that fermion-tempered backreaction is not qualitatively affected by our modeling of $\Theta(t, k)$.

Next, we show how the SGWB spectra in figure 3 are affected by our modeling of $\Theta(t, k)$. We show the updated version of this plot in figure 8. Again, we overlay the original result for k_h -damping (dashed lines) with those for k_S -damping (dotted lines). Again, we find good qualitative agreement for the GW spectra corresponding to the benchmark points $\beta = 59$ and $m = 1.5 \times 10^{-6} M_P$ (cyan), $2.5 \times 10^{-6} M_P$ (violet). In both models, they reach roughly the same amplitude, cover the same frequency range, are detectable by the same observatories,

⁵Note that only in the first Heaviside function k_h is identified with k_S . The second, k -independent Heaviside function affects all gauge modes $A_\lambda(t, k)$ equally. Thus, we do not gain anything in terms of the numerical implementation by approximating $k_S \simeq k_h$ for this second term.

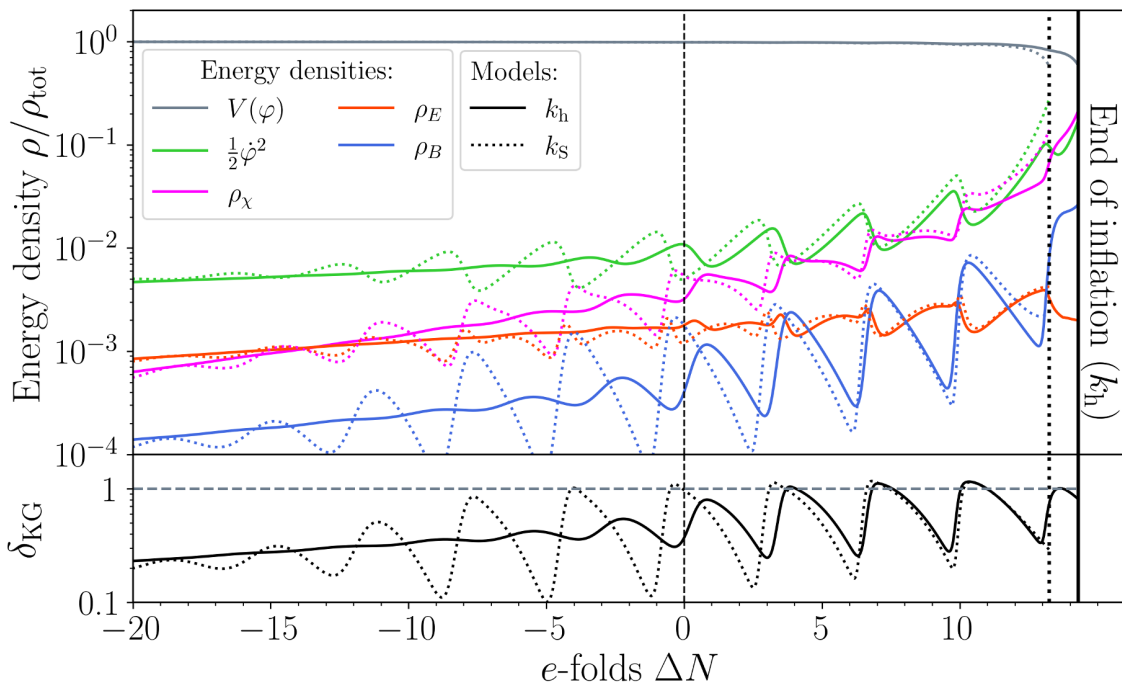


Figure 7. Same as in figure 1, but we overlay the previous results for k_h -damping (solid lines) with those for k_S -damping (dotted lines).

and feature the same spectral slope. For the k_S -damping model, the amplitudes of the spectra are marginally decreased, and the growth is marginally less stable, including some small oscillations. For $\beta = 59$ and $m = 2.5 \times 10^{-5} M_P$ we also find qualitative agreement between the results for k_h -damping and k_S -damping. Both spectra reach NG15 sensitivity, but are ruled out by PLANCK. However, as one can anticipate based on our results in figure 7, the spectra for k_S -damping feature strong oscillations already at lower frequencies. Also, one can see the slight reduction of redshift due to ΔN_{BR} from the later rise in the GFIGW component at low frequencies for the k_S -damping model. Again, the results are in overall good qualitative agreement, and we find that the results are consistent with the interpretation given in the main body of the text.

The fact that the two models for $\Theta(t, k)$ give such similar results can also be understood quantitatively. We find that, once $k_S > k_h$, they only ever differ at most by a factor of two, while both scales vary exponentially with time. Therefore, it does not matter much if one distinguishes between k_S or k_h when modeling the scale dependence of the Schwinger induced current in eq. (2.13). The agreement between the two models is especially satisfactory given that we treat the fermions effectively instead of relying on a first-principles computation. These encouraging observations notwithstanding, we also argue that it would be interesting to investigate the new regime of fermion-tempered backreaction on the lattice in order to resolve the remaining conceptual uncertainties inherent to our approach.

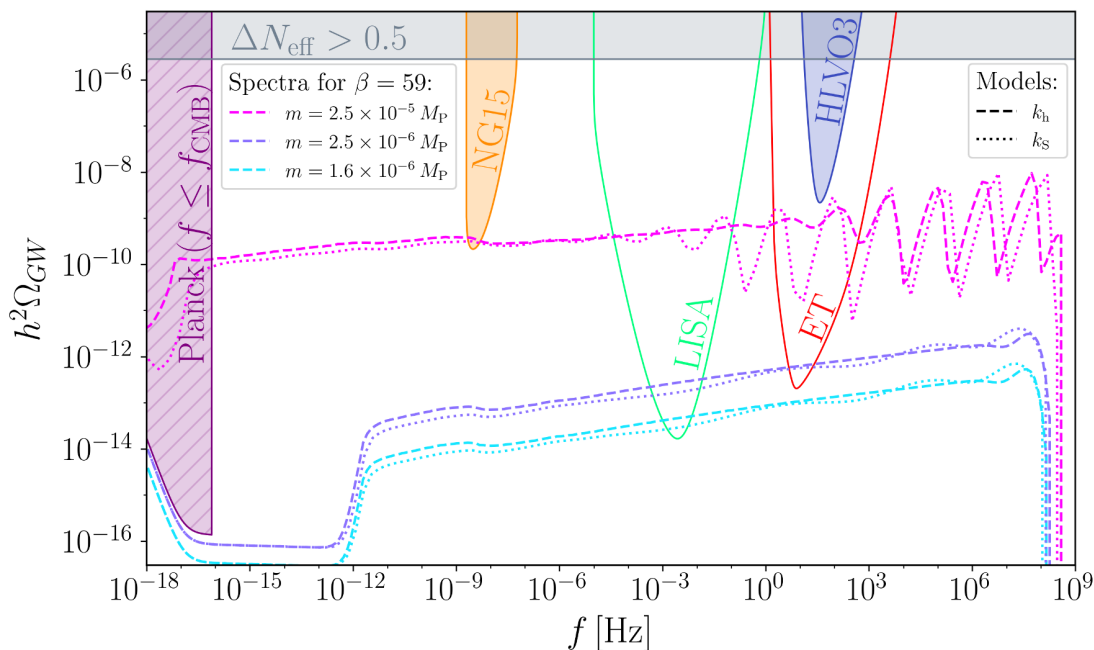


Figure 8. Same as in figure 3, but we overlay the previous results for k_h -damping (dashed lines) with those for k_S -damping (dotted lines).

Data Availability Statement. This article has no associated data or the data will not be deposited.

Code Availability Statement. The source code used for the numerical solutions of this study, the Gradient Expansion Formalism Factory (GEFF), is freely available at <https://github.com/richard-von-eckardstein/GEFF>.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] R. von Eckardstein, K. Schmitz and O. Sobol, *Gravitational waves from axion inflation in the gradient expansion formalism. Part I. Pure axion inflation*, *JHEP* **01** (2026) 018 [[arXiv:2508.00798](https://arxiv.org/abs/2508.00798)] [[INSPIRE](#)].
- [2] A.A. Starobinsky, *A New Type of Isotropic Cosmological Models Without Singularity*, *Phys. Lett. B* **91** (1980) 99 [[INSPIRE](#)].
- [3] A.H. Guth, *The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems*, *Phys. Rev. D* **23** (1981) 347 [[INSPIRE](#)].
- [4] A.D. Linde, *A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems*, *Phys. Lett. B* **108** (1982) 389 [[INSPIRE](#)].

- [5] A. Albrecht and P.J. Steinhardt, *Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking*, *Phys. Rev. Lett.* **48** (1982) 1220 [INSPIRE].
- [6] A.A. Starobinsky, *Dynamics of Phase Transition in the New Inflationary Universe Scenario and Generation of Perturbations*, *Phys. Lett. B* **117** (1982) 175 [INSPIRE].
- [7] A.D. Linde, *Chaotic Inflation*, *Phys. Lett. B* **129** (1983) 177 [INSPIRE].
- [8] V.F. Mukhanov and G.V. Chibisov, *Quantum Fluctuations and a Nonsingular Universe*, *JETP Lett.* **33** (1981) 532 [INSPIRE].
- [9] V.F. Mukhanov and G.V. Chibisov, *The vacuum energy and large scale structure of the universe*, *Sov. Phys. JETP* **56** (1982) 258 [INSPIRE].
- [10] S.W. Hawking, *The Development of Irregularities in a Single Bubble Inflationary Universe*, *Phys. Lett. B* **115** (1982) 295 [INSPIRE].
- [11] A.H. Guth and S.Y. Pi, *Fluctuations in the New Inflationary Universe*, *Phys. Rev. Lett.* **49** (1982) 1110 [INSPIRE].
- [12] J.M. Bardeen, P.J. Steinhardt and M.S. Turner, *Spontaneous Creation of Almost Scale — Free Density Perturbations in an Inflationary Universe*, *Phys. Rev. D* **28** (1983) 679 [INSPIRE].
- [13] L. Kofman, A.D. Linde and A.A. Starobinsky, *Towards the theory of reheating after inflation*, *Phys. Rev. D* **56** (1997) 3258 [hep-ph/9704452] [INSPIRE].
- [14] R. Allahverdi, R. Brandenberger, F.-Y. Cyr-Racine and A. Mazumdar, *Reheating in Inflationary Cosmology: Theory and Applications*, *Ann. Rev. Nucl. Part. Sci.* **60** (2010) 27 [arXiv:1001.2600] [INSPIRE].
- [15] M.A. Amin, M.P. Hertzberg, D.I. Kaiser and J. Karouby, *Nonperturbative Dynamics Of Reheating After Inflation: A Review*, *Int. J. Mod. Phys. D* **24** (2014) 1530003 [arXiv:1410.3808] [INSPIRE].
- [16] W.D. Garretson, G.B. Field and S.M. Carroll, *Primordial magnetic fields from pseudoGoldstone bosons*, *Phys. Rev. D* **46** (1992) 5346 [hep-ph/9209238] [INSPIRE].
- [17] M.M. Anber and L. Sorbo, *N-flationary magnetic fields*, *JCAP* **10** (2006) 018 [astro-ph/0606534] [INSPIRE].
- [18] M.M. Anber and L. Sorbo, *Naturally inflating on steep potentials through electromagnetic dissipation*, *Phys. Rev. D* **81** (2010) 043534 [arXiv:0908.4089] [INSPIRE].
- [19] K. Freese, J.A. Frieman and A.V. Olinto, *Natural Inflation with Pseudo-Nambu-Goldstone Bosons*, *Phys. Rev. Lett.* **65** (1990) 3233 [INSPIRE].
- [20] P. Adshead and M. Wyman, *Chromo-Natural Inflation: Natural inflation on a steep potential with classical non-Abelian gauge fields*, *Phys. Rev. Lett.* **108** (2012) 261302 [arXiv:1202.2366] [INSPIRE].
- [21] F. Sauter, *Über das Verhalten eines Elektrons im homogenen elektrischen Feld nach der relativistischen Theorie Diracs*, *Z. Phys.* **69** (1931) 742 [INSPIRE].
- [22] W. Heisenberg and H. Euler, *Consequences of Dirac's theory of positrons*, *Z. Phys.* **98** (1936) 714 [physics/0605038] [INSPIRE].
- [23] J.S. Schwinger, *On gauge invariance and vacuum polarization*, *Phys. Rev.* **82** (1951) 664 [INSPIRE].
- [24] M.B. Fröb et al., *Schwinger effect in de Sitter space*, *JCAP* **04** (2014) 009 [arXiv:1401.4137] [INSPIRE].

- [25] T. Kobayashi and N. Afshordi, *Schwinger Effect in 4D de Sitter Space and Constraints on Magnetogenesis in the Early Universe*, *JHEP* **10** (2014) 166 [[arXiv:1408.4141](#)] [[INSPIRE](#)].
- [26] C. Stahl, E. Strobel and S.-S. Xue, *Fermionic current and Schwinger effect in de Sitter spacetime*, *Phys. Rev. D* **93** (2016) 025004 [[arXiv:1507.01686](#)] [[INSPIRE](#)].
- [27] T. Hayashinaka and J. Yokoyama, *Point splitting renormalization of Schwinger induced current in de Sitter spacetime*, *JCAP* **07** (2016) 012 [[arXiv:1603.06172](#)] [[INSPIRE](#)].
- [28] T. Hayashinaka, T. Fujita and J. Yokoyama, *Fermionic Schwinger effect and induced current in de Sitter space*, *JCAP* **07** (2016) 010 [[arXiv:1603.04165](#)] [[INSPIRE](#)].
- [29] E. Bavarsad, C. Stahl and S.-S. Xue, *Scalar current of created pairs by Schwinger mechanism in de Sitter spacetime*, *Phys. Rev. D* **94** (2016) 104011 [[arXiv:1602.06556](#)] [[INSPIRE](#)].
- [30] T. Hayashinaka and S.-S. Xue, *Physical renormalization condition for de Sitter QED*, *Phys. Rev. D* **97** (2018) 105010 [[arXiv:1802.03686](#)] [[INSPIRE](#)].
- [31] M. Banyeres, G. Domènech and J. Garriga, *Vacuum birefringence and the Schwinger effect in (3+1) de Sitter*, *JCAP* **10** (2018) 023 [[arXiv:1809.08977](#)] [[INSPIRE](#)].
- [32] V. Domcke and K. Mukaida, *Gauge Field and Fermion Production during Axion Inflation*, *JCAP* **11** (2018) 020 [[arXiv:1806.08769](#)] [[INSPIRE](#)].
- [33] V. Domcke, Y. Ema and K. Mukaida, *Chiral Anomaly, Schwinger Effect, Euler-Heisenberg Lagrangian, and application to axion inflation*, *JHEP* **02** (2020) 055 [[arXiv:1910.01205](#)] [[INSPIRE](#)].
- [34] M. Bastero-Gil et al., *Schwinger Current in de Sitter Space*, [arXiv:2503.01981](#) [[INSPIRE](#)].
- [35] M. Bastero-Gil et al., *Classical constant electric fields and the Schwinger effect in de Sitter*, [arXiv:2508.14973](#) [[INSPIRE](#)].
- [36] O.O. Sobol, E.V. Gorbar and S.I. Vilchinskii, *Backreaction of electromagnetic fields and the Schwinger effect in pseudoscalar inflation magnetogenesis*, *Phys. Rev. D* **100** (2019) 063523 [[arXiv:1907.10443](#)] [[INSPIRE](#)].
- [37] E.V. Gorbar, K. Schmitz, O.O. Sobol and S.I. Vilchinskii, *Gauge-field production during axion inflation in the gradient expansion formalism*, *Phys. Rev. D* **104** (2021) 123504 [[arXiv:2109.01651](#)] [[INSPIRE](#)].
- [38] E.V. Gorbar, K. Schmitz, O.O. Sobol and S.I. Vilchinskii, *Hypermagnetogenesis from axion inflation: Model-independent estimates*, *Phys. Rev. D* **105** (2022) 043530 [[arXiv:2111.04712](#)] [[INSPIRE](#)].
- [39] T. Fujita, J. Kume, K. Mukaida and Y. Tada, *Effective treatment of U(1) gauge field and charged particles in axion inflation*, *JCAP* **09** (2022) 023 [[arXiv:2204.01180](#)] [[INSPIRE](#)].
- [40] Y. Cado and M. Quirós, *Numerical study of the Schwinger effect in axion inflation*, *Phys. Rev. D* **106** (2022) 123527 [[arXiv:2208.10977](#)] [[INSPIRE](#)].
- [41] V. Domcke et al., *Wash-in leptogenesis after axion inflation*, *JHEP* **01** (2023) 053 [[arXiv:2210.06412](#)] [[INSPIRE](#)].
- [42] E.V. Gorbar et al., *Chirality Production during Axion Inflation*, *Ukr. J. Phys.* **68** (2023) 717 [[arXiv:2111.05848](#)] [[INSPIRE](#)].
- [43] M. Bastero-Gil, P.B. Ferraz, L. Ubaldi and R. Vega-Morales, *Schwinger dark matter production*, *JCAP* **10** (2024) 078 [[arXiv:2312.15137](#)] [[INSPIRE](#)].

- [44] M. Bastero-Gil, P.B. Ferraz, L. Ubaldi and R. Vega-Morales, *Super heavy dark matter from inflationary Schwinger production*, *Phys. Rev. D* **110** (2024) 095019 [[arXiv:2311.09475](#)] [[INSPIRE](#)].
- [45] R. von Eckardstein, K. Schmitz and O. Sobol, *On the Schwinger effect during axion inflation*, *JHEP* **02** (2025) 096 [[arXiv:2408.16538](#)] [[INSPIRE](#)].
- [46] O. Iarygina, E.I. Sfakianakis and A. Brandenburg, *Schwinger effect in axion inflation on a lattice*, [arXiv:2506.20538](#) [[INSPIRE](#)].
- [47] O.O. Sobol, E.V. Gorbar, M. Kamarpour and S.I. Vilchinskii, *Influence of backreaction of electric fields and Schwinger effect on inflationary magnetogenesis*, *Phys. Rev. D* **98** (2018) 063534 [[arXiv:1807.09851](#)] [[INSPIRE](#)].
- [48] H. Kitamoto, *Schwinger Effect in Inflaton-Driven Electric Field*, *Phys. Rev. D* **98** (2018) 103512 [[arXiv:1807.03753](#)] [[INSPIRE](#)].
- [49] E.V. Gorbar, A.I. Momot, O.O. Sobol and S.I. Vilchinskii, *Kinetic approach to the Schwinger effect during inflation*, *Phys. Rev. D* **100** (2019) 123502 [[arXiv:1909.10332](#)] [[INSPIRE](#)].
- [50] O.O. Sobol, E.V. Gorbar, A.I. Momot and S.I. Vilchinskii, *Schwinger production of scalar particles during and after inflation from the first principles*, *Phys. Rev. D* **102** (2020) 023506 [[arXiv:2004.12664](#)] [[INSPIRE](#)].
- [51] R. Durrer, O. Sobol and S. Vilchinskii, *Backreaction from gauge fields produced during inflation*, *Phys. Rev. D* **108** (2023) 043540 [[arXiv:2303.04583](#)] [[INSPIRE](#)].
- [52] R. von Eckardstein et al., *Axion inflation in the strong-backreaction regime: decay of the Anber-Sorbo solution*, *JHEP* **11** (2023) 183 [[arXiv:2309.04254](#)] [[INSPIRE](#)].
- [53] V. Domcke, Y. Ema and S. Sandner, *Perturbatively including inhomogeneities in axion inflation*, *JCAP* **03** (2024) 019 [[arXiv:2310.09186](#)] [[INSPIRE](#)].
- [54] LISA collaboration, *Laser Interferometer Space Antenna*, [arXiv:1702.00786](#).
- [55] J. Baker et al., *The Laser Interferometer Space Antenna: Unveiling the Millihertz Gravitational Wave Sky*, [arXiv:1907.06482](#) [[INSPIRE](#)].
- [56] M. Punturo et al., *The Einstein Telescope: A third-generation gravitational wave observatory*, *Class. Quant. Grav.* **27** (2010) 194002 [[INSPIRE](#)].
- [57] NANOGrav collaboration, *The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background*, *Astrophys. J. Lett.* **951** (2023) L8 [[arXiv:2306.16213](#)] [[INSPIRE](#)].
- [58] EPTA and INPTA: collaborations, *The second data release from the European Pulsar Timing Array — III. Search for gravitational wave signals*, *Astron. Astrophys.* **678** (2023) A50 [[arXiv:2306.16214](#)] [[INSPIRE](#)].
- [59] D.J. Reardon et al., *Search for an Isotropic Gravitational-wave Background with the Parkes Pulsar Timing Array*, *Astrophys. J. Lett.* **951** (2023) L6 [[arXiv:2306.16215](#)] [[INSPIRE](#)].
- [60] H. Xu et al., *Searching for the Nano-Hertz Stochastic Gravitational Wave Background with the Chinese Pulsar Timing Array Data Release I*, *Res. Astron. Astrophys.* **23** (2023) 075024 [[arXiv:2306.16216](#)] [[INSPIRE](#)].
- [61] M.T. Miles et al., *The MeerKAT Pulsar Timing Array: the first search for gravitational waves with the MeerKAT radio telescope*, *Mon. Not. Roy. Astron. Soc.* **536** (2024) 1489 [[arXiv:2412.01153](#)] [[INSPIRE](#)].
- [62] PLANCK collaboration, *Planck 2018 results. X. Constraints on inflation*, *Astron. Astrophys.* **641** (2020) A10 [[arXiv:1807.06211](#)] [[INSPIRE](#)].

- [63] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [*Erratum ibid.* **652** (2021) C4] [[arXiv:1807.06209](#)] [[INSPIRE](#)].
- [64] T.-H. Yeh, K.A. Olive and B.D. Fields, *The impact of new $d(p, \gamma)3$ rates on Big Bang Nucleosynthesis*, *JCAP* **03** (2021) 046 [[arXiv:2011.13874](#)] [[INSPIRE](#)].
- [65] O. Pisanti, G. Mangano, G. Miele and P. Mazzella, *Primordial Deuterium after LUNA: concordances and error budget*, *JCAP* **04** (2021) 020 [[arXiv:2011.11537](#)] [[INSPIRE](#)].
- [66] LIGO SCIENTIFIC collaboration, *Advanced LIGO: The next generation of gravitational wave detectors*, *Class. Quant. Grav.* **27** (2010) 084006 [[INSPIRE](#)].
- [67] LIGO SCIENTIFIC collaboration, *Advanced LIGO*, *Class. Quant. Grav.* **32** (2015) 074001 [[arXiv:1411.4547](#)] [[INSPIRE](#)].
- [68] VIRGO collaboration, *Advanced Virgo: a second-generation interferometric gravitational wave detector*, *Class. Quant. Grav.* **32** (2015) 024001 [[arXiv:1408.3978](#)] [[INSPIRE](#)].
- [69] KAGRA et al. collaborations, *Upper limits on the isotropic gravitational-wave background from Advanced LIGO and Advanced Virgo's third observing run*, *Phys. Rev. D* **104** (2021) 022004 [[arXiv:2101.12130](#)] [[INSPIRE](#)].
- [70] D.G. Figueroa, J. Lizarraga, A. Urrio and J. Urrestilla, *Strong Backreaction Regime in Axion Inflation*, *Phys. Rev. Lett.* **131** (2023) 151003 [[arXiv:2303.17436](#)] [[INSPIRE](#)].
- [71] D.G. Figueroa et al., *Nonlinear dynamics of axion inflation: A detailed lattice study*, *Phys. Rev. D* **111** (2025) 063545 [[arXiv:2411.16368](#)] [[INSPIRE](#)].
- [72] A. Caravano, E. Komatsu, K.D. Lozanov and J. Weller, *Lattice simulations of axion- $U(1)$ inflation*, *Phys. Rev. D* **108** (2023) 043504 [[arXiv:2204.12874](#)] [[INSPIRE](#)].
- [73] R. Sharma, A. Brandenburg, K. Subramanian and A. Vikman, *Lattice simulations of axion- $U(1)$ inflation: gravitational waves, magnetic fields, and scalar statistics*, *JCAP* **05** (2025) 079 [[arXiv:2411.04854](#)] [[INSPIRE](#)].
- [74] PENCIL CODE collaboration, *The Pencil Code, a modular MPI code for partial differential equations and particles: multipurpose and multiuser-maintained*, *J. Open Source Softw.* **6** (2021) 2807 [[arXiv:2009.08231](#)] [[INSPIRE](#)].
- [75] R. von Eckardstein, *GEFF: The Gradient Expansion Formalism Factory — A tool for inflationary gauge-field production*, [arXiv:2510.12644](#) [[INSPIRE](#)].
- [76] C. Caprini and D.G. Figueroa, *Cosmological Backgrounds of Gravitational Waves*, *Class. Quant. Grav.* **35** (2018) 163001 [[arXiv:1801.04268](#)] [[INSPIRE](#)].
- [77] S. Kuroyanagi, T. Takahashi and S. Yokoyama, *Blue-tilted Tensor Spectrum and Thermal History of the Universe*, *JCAP* **02** (2015) 003 [[arXiv:1407.4785](#)] [[INSPIRE](#)].
- [78] S. Kuroyanagi, T. Takahashi and S. Yokoyama, *Blue-tilted inflationary tensor spectrum and reheating in the light of NANOGrav results*, *JCAP* **01** (2021) 071 [[arXiv:2011.03323](#)] [[INSPIRE](#)].
- [79] R.Z. Ferreira and A. Notari, *Thermalized Axion Inflation*, *JCAP* **09** (2017) 007 [[arXiv:1706.00373](#)] [[INSPIRE](#)].
- [80] G. Galloni, S. Henrot-Versillé and M. Tristram, *Robust constraints on tensor perturbations from cosmological data: A comparative analysis from Bayesian and frequentist perspectives*, *Phys. Rev. D* **110** (2024) 063511 [[arXiv:2405.04455](#)] [[INSPIRE](#)].
- [81] M. Tristram et al., *Improved limits on the tensor-to-scalar ratio using BICEP and Planck data*, *Phys. Rev. D* **105** (2022) 083524 [[arXiv:2112.07961](#)] [[INSPIRE](#)].

- [82] N. Barnaby and M. Peloso, *Large Nongaussianity in Axion Inflation*, *Phys. Rev. Lett.* **106** (2011) 181301 [[arXiv:1011.1500](#)] [[INSPIRE](#)].
- [83] N. Barnaby, R. Namba and M. Peloso, *Phenomenology of a Pseudo-Scalar Inflaton: Naturally Large Nongaussianity*, *JCAP* **04** (2011) 009 [[arXiv:1102.4333](#)] [[INSPIRE](#)].
- [84] PLANCK collaboration, *Planck 2015 results. XVII. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **594** (2016) A17 [[arXiv:1502.01592](#)] [[INSPIRE](#)].
- [85] PLANCK collaboration, *Planck 2018 results. IX. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **641** (2020) A9 [[arXiv:1905.05697](#)] [[INSPIRE](#)].
- [86] M. Drewes et al., *Towards a precision calculation of N_{eff} in the Standard Model. Part III. Improved estimate of NLO contributions to the collision integral*, *JCAP* **06** (2024) 032 [[arXiv:2402.18481](#)] [[INSPIRE](#)].
- [87] G. Agazie et al., *The NANOGrav 15 yr Data Set: Running of the Spectral Index*, *Astrophys. J. Lett.* **978** (2025) L29 [[arXiv:2408.10166](#)] [[INSPIRE](#)].
- [88] B. Allen, *The stochastic gravity wave background: Sources and detection*, in the proceedings of the *Les Houches School of Physics: Astrophysical Sources of Gravitational Radiation*, Les Houches, France, September 26 – October 06 (1995) [[gr-qc/9604033](#)] [[INSPIRE](#)].
- [89] B. Allen and J.D. Romano, *Detecting a stochastic background of gravitational radiation: Signal processing strategies and sensitivities*, *Phys. Rev. D* **59** (1999) 102001 [[gr-qc/9710117](#)] [[INSPIRE](#)].
- [90] M. Maggiore, *Gravitational wave experiments and early universe cosmology*, *Phys. Rept.* **331** (2000) 283 [[gr-qc/9909001](#)] [[INSPIRE](#)].
- [91] K. Schmitz, *New Sensitivity Curves for Gravitational-Wave Signals from Cosmological Phase Transitions*, *JHEP* **01** (2021) 097 [[arXiv:2002.04615](#)] [[INSPIRE](#)].
- [92] LIGO SCIENTIFIC et al. collaborations, *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Data behind the figures*, (2021) [[DOI:10.5281/zenodo.5571767](#)].
- [93] NANOGrAV collaboration, *The NANOGrav 15 yr Data Set: Detector Characterization and Noise Budget*, *Astrophys. J. Lett.* **951** (2023) L10 [[arXiv:2306.16218](#)] [[INSPIRE](#)].
- [94] NANOGrAV collaboration, *Why Gravitational Wave Science Needs Pulsar Timing Arrays And Why Pulsar Timing Arrays Need Both Arecibo and the GBT: A response to the NSF-AST Portfolio Review from the NANOGrav Collaboration*, [arXiv:1210.5998](#) [[INSPIRE](#)].
- [95] E. Thrane and J.D. Romano, *Sensitivity curves for searches for gravitational-wave backgrounds*, *Phys. Rev. D* **88** (2013) 124032 [[arXiv:1310.5300](#)] [[INSPIRE](#)].
- [96] V. Domcke, V. Guidetti, Y. Welling and A. Westphal, *Resonant backreaction in axion inflation*, *JCAP* **09** (2020) 009 [[arXiv:2002.02952](#)] [[INSPIRE](#)].