General framework for cosmological dark matter bounds using *N*-body simulations

Keir K. Rogers[1,*] and Hiranya V. Peiris[2,1,†]

[1]*Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, AlbaNova University Center, Stockholm 10691, Sweden*
[2]*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK*

We present a general framework for obtaining robust bounds on the nature of dark matter using cosmological *N*-body simulations and Lyman-alpha forest data. We construct an emulator of hydrodynamical simulations, which is a flexible, accurate and computationally efficient model for predicting the response of the Lyman-alpha forest flux power spectrum to different dark matter models, the state of the intergalactic medium (IGM) and the primordial power spectrum. The emulator combines a flexible parametrization for the small-scale suppression in the matter power spectrum arising in "noncold" dark matter models, with an improved IGM model. We then demonstrate how to optimize the emulator for the case of ultralight axion dark matter, presenting tests of convergence. We also carry out cross-validation tests of the accuracy of flux power spectrum prediction. This framework can be optimized for the analysis of many other dark matter candidates, e.g., warm or interacting dark matter. Our work demonstrates that a combination of an optimized emulator and cosmological "effective theories," where many models are described by a single set of equations, is a powerful approach for robust and computationally efficient inference from the cosmic large-scale structure.

I. INTRODUCTION

Determining the particle physics of the dark matter would not only explain an integral part of the standard cosmological model but can also point to new physics that will deepen and unify our understanding of nature. Following recent null results at high-energy colliders [1] and by direct detection (e.g., [2]), it is timely to consider new theoretically well-motivated avenues that do not rely exclusively on the "WIMP miracle" (weakly interacting massive particles [3]), as well as alternative experimental approaches beyond colliders and direct/indirect detection. Cosmological observations allow constraints on dark matter properties in physical regimes otherwise inaccessible in the laboratory (e.g., [4,5]).

There exist many alternative proposals to the traditional WIMP, each with their own theoretical motivation. What unites a large number of these models are similar phenomenological features in cosmological structure formation (in particular, suppressed small-scale growth) constituting a deviation from the standard paradigm of cold, collisionless dark matter [6–9]. These include, e.g., light axions [10–16]; warm dark matter (particles with an associated free-streaming scale), e.g., sterile neutrinos resonantly produced [17] or as a decay product in the early Universe (e.g., [18]); mixed models, e.g., of cold and warm dark matter [19]; light (sub-GeV) particle candidates with non-negligible scattering with baryons [20], e.g., those with electric and magnetic dipole moments [21] or massive boson exchange [22]; and the wide set of models described by the effective theory of structure formation [23], e.g., with dark radiative components [24] or self-interactions [25].

The Lyman-alpha forest (as a tracer of the small-scale, high-redshift linear matter power spectrum [26]), often in combination with larger-scale probes like the cosmic microwave background (CMB) and baryon acoustic oscillations, provides competitive bounds on these models [19,20,27–35], to which we will refer in general as "noncold" dark matter, following Ref. [36]. The challenge in getting robust bounds using the Lyman-alpha forest and other probes of the cosmic large-scale structure (e.g., galaxies, weak lensing, 21 cm) is the computational cost of modeling the data with sufficient accuracy. This often requires *N*-body simulations [37,38], sometimes with hydrodynamics [39] which adds even further computational cost. This makes "brute-force" sampling of the parameter space [as required by e.g., Markov chain

*keir.rogers@utoronto.ca
†h.peiris@ucl.ac.uk

Monte Carlo methods (MCMC)] computationally infeasible. Further, the many different dark matter and cosmological models that can be tested each require, in principle, a new set of simulations for each analysis, which can make inefficient use of computational resources. Much effort has gone into developing cosmological "effective theories," where many different physical models are described by a single set of equations (e.g., [23,40]). These are designed to address the issue of testing many physical theories in an efficient way.

In this work, we present a framework for cosmological dark matter bounds, specifically using the Lyman-alpha forest, which makes the testing of new models robust, accurate and computationally efficient. This combines a general model for the effect of noncold dark matter (nCDM) on the matter power spectrum presented in Ref. [36] with the Bayesian emulator optimization we presented in Refs. [41,42]. An "emulator" is a flexible model that can characterize the data accurately and is computationally cheap to evaluate [43]. This means it can be easily called within MCMC, making parameter inference feasible. It is optimized using a small set of "training" simulations, the selection of which is critical (see Sec. III B for discussion on Bayesian emulator optimization). Emulators have found wide use in cosmological studies using the large-scale structure [41,42,44–52].

We build an emulator for the Lyman-alpha forest flux power spectrum as a function of the nCDM model and other nuisance parameters. This initial base emulator can then be optimized for the analysis of specific dark matter models. Here, we consider the example of ultralight axion (ULA) dark matter, but the procedure will be equivalent for any of the dark matter candidates which can be characterized by the nCDM model (see above). We optimize the construction of the emulator training set using Bayesian optimization, which iteratively adds training data in order to determine the peak of the posterior distribution accurately while exploring the parameter space fully [53]. This allows tests of the robustness of dark matter bounds with respect to the accuracy of the emulator model. Previous dark matter and cosmological bounds from the Lyman-alpha forest (e.g., [19,20,27–29,31–35,54–58]) have relied on linear interpolation of simulated flux power spectra around a fiducial point. Reference [30] interpolate using "ordinary kriging", which has some similarity to the method we consider here, although they do not model full parameter covariance nor test for convergence using Bayesian optimization (see Sec. III B). The emulator model we use makes fewer assumptions and is more robust in its statistical modeling, which can remove bias in power spectrum estimation and strengthen parameter constraints [42]. Even for cosmological datasets where emulators are already used, the combination of convergence tests using Bayesian optimization and effective theories as described above can provide significant benefits.

In Sec. II, we present the general emulator for nCDM models and then in Sec. III, we present the example of optimizing the emulator for the study of ULA dark matter. We conduct tests of emulator convergence and cross-validation in Sec. IV, discuss in Sec. V and conclude in Sec. VI.

II. NONCOLD DARK MATTER EMULATOR

In this section, we present a general emulator of hydrodynamical simulations for noncold dark matter models using the Lyman-alpha forest flux power spectrum as the constraining data. This initial emulator is not tied to a specific dark matter model, but rather forms the base for Bayesian optimization for a particular model (see Sec. III for the example of ultralight axions). This framework ensures the computational efficiency, theoretical accuracy and maximum precision for dark matter bounds using the cosmic large-scale structure. In Sec. II A, we explain the noncold dark matter model we employ, first introduced by Ref. [36]. We present the details of our cosmological hydrodynamical simulations in Sec. II B and the intergalactic medium and background cosmology model which must be marginalized in Sec. II C. In Sec. II D, we discuss the method of emulation by Gaussian processes and how this connects to Bayesian inference.

A. Noncold dark matter model

In order to build an emulator-inference framework which is as general and powerful as possible, we employ the noncold dark matter (nCDM) model first introduced by Ref. [36]. The model is defined by a "transfer function",

$$T(k) \equiv \sqrt{\frac{P_{\mathrm{nCDM}}(k)}{P_{\mathrm{CDM}}(k)}}$$
$$= [1 + (\alpha k)^{\beta}]^{\gamma}. \qquad (1)$$

$P_{\mathrm{CDM}}(k)$ and $P_{\mathrm{nCDM}}(k)$ are the linear matter power spectra, respectively, for CDM and nCDM models; k is the comoving wave number (length-scale) in $h\,\mathrm{Mpc}^{-1}$; and $[\alpha, \beta, \gamma]$ are the free parameters of the model, where α is in units of h^{-1} Mpc. In our analysis, we define $T(k)$ at the initial conditions of our cosmological simulations (see Sec. II B), i.e., at $z = 99$. It is a generalization of the fitting function for a thermal warm dark matter (WDM) transfer function [59], to which Eq. (1) reduces when $\beta = 2\nu$, $\gamma = -\frac{5}{\nu}$ and $\nu = 1.12$. By having two further degrees of freedom, the general model in Eq. (1) can characterize the matter power spectra for many models beyond the thermal WDM case, e.g., nonthermal WDM, like sterile neutrinos resonantly produced or from particle decays, mixed models of cold and warm dark matter, axionlike particles, light (sub-GeV) dark matter coupling to baryons, dark matter coupling to dark radiation or
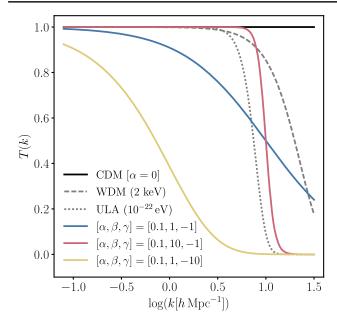
FIG. 1. Transfer functions $T(k)$ [as defined in Eq. (1) at $z = 99$] for various configurations of the noncold dark matter (nCDM) model as a function of comoving distance wave number k. The colored lines show some extremal values of our prior range, while the black lines show nCDM fits for cold dark matter ($\alpha = 0$), 2 keV warm dark matter ($[\alpha, \beta, \gamma] = [0.02, 2.2, -4.5]$) and 10^{-22} eV ultralight axion dark matter $[\alpha, \beta, \gamma] = [0.1, 5.2, -3.4]$.

self-interacting dark matter [17–25]. Many of these models are explored in Ref. [36]; we will consider the example of ultralight axions in Sec. III.

We consider $\alpha \in [0, 0.1]$, $\beta \in [1, 10]$, $\gamma \in [-10, 0]$. Figure 1 shows some example nCDM transfer functions. In particular, we show the cold dark matter limit with $\alpha = 0$; nCDM model fits for the transfer function of a 2 keV WDM particle $[\alpha, \beta, \gamma] = [0.02, 2.2, -4.5]$ and a 10^{-22} eV ULA particle $[\alpha, \beta, \gamma] = [0.1, 5.2, -3.4]$ (more details in Sec. III); and some extremal values of the nCDM parameters within our prior volume. In general, the nCDM model can characterize dark matter candidates which suppress the growth of structure on small scales and hence form a cutoff in the matter power spectrum. By spanning the three free parameters of the model, it is possible to characterize different scales and shapes of this cutoff, which can map to different particle models, as shown by the WDM and ULA examples in Fig. 1. α is the main parameter determining the scale of the suppression and as it increases, the cutoff moves to larger scales (smaller k). β and γ are the main parameters determining the shape of the suppression. As β increases, the cutoff becomes sharper, in particular flatter before $T^2 = 0.5$; as γ increases, the cutoff also becomes sharper, in particular flatter after $T^2 = 0.5$.

B. Cosmological hydrodynamical simulations

Our observable is the Lyman-alpha forest flux power spectrum at $z = [4.2, 4.6, 5.0]$. In order to model this, we run cosmological hydrodynamical simulations of the intergalactic medium (IGM; from which the forest is sourced) using the publicly available [60] code MP-Gadget [39,61,62]. We evolve 512^3 particles each of dark matter and gas in a $(10 \ h^{-1} \ \mathrm{Mpc})^3$ box (see tests of convergence with respect to particle number and box size in the Appendix A) from $z = 99$ to $z = 4.2$, saving snapshots of particle data in particular at $z = [4.2, 4.6, 5.0]$. We include the effect of the nCDM models we study by generating the initial conditions of each simulation using a suppressed matter power spectrum as modified by the transfer function in Eq. (1); we discuss the implications for the nCDM models we can characterize in Sec. V. We focus computational resources on the low-density gas that forms the Lyman-alpha forest by enabling the QUICK-LYA flag, which converts gas particles at overdensities > 1000 and with temperatures $< 10^5$ K into collisionless particles [63]. This yields significant computational efficiency gains with a negligible impact on the accuracy of Lyman-alpha forest statistics.

From each particle snapshot, we generate 32000 mock spectra (with pixel widths $\Delta v = 1 \ \mathrm{km \, s^{-1}}$) with only the Lyman-alpha absorption line and calculate the 1D flux power spectrum using FAKE_SPECTRA [64]. The 1D Lyman-alpha forest flux power spectrum measures correlations in Lyman-alpha absorption lines along the line-of-sight only (i.e., integrated over transverse directions). It is defined as $P_f(k_f) \equiv \int \frac{dk_f^\perp}{(2\pi)^2} P_f^{3D}(\boldsymbol{k_f})$. Here, k_f and $\boldsymbol{k_f^\perp}$ are, respectively, the line-of-sight and transverse components of the full velocity wave vector $\boldsymbol{k_f}$ and have inverse-velocity units (e.g., skm^{-1}); $P_f^{3D}(\boldsymbol{k_f})$ is the 3D flux power spectrum.

We estimate $P_f(k_f)$ first in each mock spectrum using a fast Fourier transform (FFT): $P_f^i(k_f) = |\hat{\delta}_f(k_f)|^2$, where $\hat{\delta}_f(k_f)$ is the 1D FFT of $\delta_f(v) = \frac{\mathcal{F}}{\langle \mathcal{F} \rangle} - 1$. $\mathcal{F} = e^{-\tau}$ is the fraction of emitted quasar flux transmitted, $\langle \mathcal{F} \rangle$ is the mean flux across all spectra in a given box, τ is the optical depth and the effective optical depth $\tau_{\mathrm{eff}} = -\ln\langle \mathcal{F} \rangle$; v is the line-of-sight velocity. In Ref. [65], the flux is normalized to a "rolling mean flux" (averaged at each pixel using a finite window). This attempts to correct for redshift evolution within individual spectra and mismodeling of the quasar emission continuum; we test the impact on our results in Appendix C. Finally, we then average the flux power spectra over all mock spectra and correct for the pixel window function [66]: $P_f(k_f) = \langle \frac{P_f^i(k_f)}{W^2(k_f, \Delta v)} \rangle_i$. The pixel window function is $W(k_f, \Delta v) = \frac{\sin(k_f \Delta v / 2)}{k_f \Delta v / 2}$; note that we do not simulate the finite resolution of the spectrograph as this does not commute with the rescaling of mean fluxes (see Sec. II C 4) and we compare only to data where the spectrographic resolution has been corrected (we test the impact of imperfect resolution modeling in Appendix B).

We calculate (and emulate; see Sec. II D) the flux power spectrum at the (real-valued) discrete Fourier transform sample frequencies of the simulation box with wave numbers in inverse-distance units ($h\,\mathrm{Mpc}^{-1}$). This is because the conversion to velocity units depends on redshift and the energy density by the Hubble law. Since the matter energy density will be varied in our base emulator (see Sec. II C 5), this would mean that the velocity wave numbers of each simulation would vary across the emulator and the modes would not be consistently distributed. In the likelihood function (see Sec. II D), we linearly interpolate to velocity wave numbers.

C. Intergalactic medium and cosmological model

In order to study nCDM models using the flux power spectrum, we must marginalize over uncertainty in the thermal and ionization history of the IGM (which partly derives from the uncertain nature of cosmic reionization), as well as relevant aspects of the cosmological model.

1. Pressure smoothing

The energy deposited in the IGM by the background of ionizing ultraviolet (UV) photons and the finite temperature of the IGM ($\sim 10^4$ K at $z \sim 5$) leads to a few different observable effects in the flux power spectrum. The heat in the IGM smoothes the spatial distribution of the gas below a characteristic scale, where the gas pressure balances gravitational clustering. A simple, classical interpretation defines a Jeans pressure smoothing scale, which is a function of the instantaneous gas temperature [e.g., [67]]. However, in an expanding universe with an evolving thermal state, the relevant pressure smoothing scale depends on the integrated history of the gas heating as spatial fluctuations from earlier times expand or fail to collapse depending on the temperature at that epoch. This "filtering length" in general deviates from the Jeans length, being typically greater before reionization and less afterwards (~ 100 comoving kpc at $z \sim 5$).

In order to track the heat deposited in our simulations and the resulting filtering scale, we use the integrated energy deposited into the gas per unit mass at the mean density,

$$u_0(z) = \int_{z_{\mathrm{end}}}^{z_{\mathrm{ini}}} \frac{\mathrm{d}z}{H(z)(1+z)} \frac{\sum_i n_i \epsilon_i}{\bar{\rho}}, \qquad (2)$$

first introduced by Ref. [68]. Here, n_i are the number densities for each species $i \in$ [HI, HeI, HeII], $\bar{\rho}$ is the mean mass density, $H(z)$ is the Hubble parameter and $[z_{\mathrm{end}}, z_{\mathrm{ini}}]$ sets the redshift range for integration. Reference [65] determined redshift ranges for integration at $z = [4.2, 4.6, 5.0]$, such that u_0 most strongly correlates with the flux power spectrum. We use those redshift ranges: at $z = 4.2$,

$[z_{\mathrm{end}}, z_{\mathrm{ini}}] = [4.2, 12]$; at $z = 4.6$, $[z_{\mathrm{end}}, z_{\mathrm{ini}}] = [4.6, 13]$; at $z = 5.0$, $[z_{\mathrm{end}}, z_{\mathrm{ini}}] = [6, 13]$. In general, a delay between heat injection and a change in the gas distribution is expected, but this delay can be reduced in the flux power spectrum owing to nonlinear redshift-space distortions arising from gas peculiar velocities; Ref. [65] posits this as an explanation for the different z_{end}. Otherwise, the values of z_{ini} are consistent with the gas gradually losing sensitivity to heating at earlier times.

2. Temperature-density relation

The small-scale flux power spectrum is sensitive to the instantaneous thermal state of the IGM in particular owing to the thermal broadening of absorption lines. After the gas is heated by the (re)ionization front, it cools (for $\Delta z \sim 1$–2) towards a thermal asymptote set largely by the balance between UV background (UVB) photoheating and adiabatic cooling owing to cosmological expansion (e.g., [69]). Further, the overwhelming majority of the low-density gas ($\Delta \lesssim 10$) to which the forest is sensitive follows a tight power-law temperature-density relation [70], where gas temperature

$$T(z) = T_0(z)\Delta^{\tilde{\gamma}(z)-1}. \qquad (3)$$

This relation has two free parameters: the temperature at mean density $T_0(z)$ and a slope $\tilde{\gamma}(z)$. The latter is generally expected to be ~ 1 (in particular immediately after an ionization event) and to asymptote towards 1.6; however, $\tilde{\gamma}(z)$ is generally poorly constrained by data with sometimes inconsistent measurements [e.g., [71]]. We discuss in Sec. V where this model breaks down, e.g., a double power law during reionization or temperature fluctuations arising from an inhomogeneous reionization [72,73], and the resulting implications for dark matter bounds.

3. Ionization state

The ionization state of each species is determined by the balance between the UVB ionization rates and recombination effects. Uncertainty in the ionization rates is degenerate in the flux power spectrum with the mean amount of absorption in the transmitted flux or the effective optical depth τ_{eff}. We therefore marginalize over the uncertainty in the ionization state by varying τ_{eff}; we parametrize τ_{eff} at each redshift by a multiplicative correction $\tau_0(z)$ to a fiducial model [65]:

$$\tau_{\mathrm{eff}}(z) = \tau_0(z) \times 1.4 \times 10^{-3} \times (1+z)^4. \qquad (4)$$

4. Simulation input and output

The thermal and ionization history in our hydrodynamical simulations (see Sec. II B) is determined by a set of

input spatially uniform UV ionization and heating rates for each primordial species [HI, HeI, HeII] as a function of redshift; our default set of rates comes from Ref. [74].[1] In order to generate a wide range of IGM histories, we vary the default photoheating rates $\epsilon_{0,i}$ by an amplitude H_A and an overdensity Δ–dependent rescaling H_S such that heating rates $\epsilon_i = H_A \epsilon_{0,i} \Delta^{H_S}$ are used. This uniformly scales the heating rates with redshift. In order to vary the redshift dependence, we employ the reionization model of Ref. [76]. We vary two free parameters of the model: the redshift of hydrogen reionization z_{rei}, defined as the redshift at which the volume-averaged ionization fraction in HII = 0.5; and, for the first time in studying nCDM models using the flux power spectrum, the total heat injection during hydrogen reionization T_{rei}. The reionization model of Ref. [76] self-consistently generates heating and ionization rates for chosen reionization parameters, only transitioning to the default rates once ionization is complete. The model, in particular, removes spurious heating at high redshift before and during reionization, which has been present in previous work (e.g., [74,77–79]).

We vary $H_A \in [0.05, 3.5]$, $H_S \in [-1.3, 0.7]$, $z_{rei} \in [6, 15]$, $T_{rei} \in [1.5 \times 10^4, 4 \times 10^4]$ K. In order to be further agnostic about the IGM states over which we marginalize and to have as robust nCDM bounds as possible, in our emulator and likelihood function (see Sec. II D), we allow IGM parameters to vary independently with redshift. To achieve this, we label each simulation at each redshift with the parameter set $[\tau_0, T_0, \tilde{\gamma}, u_0]$. This no longer restricts the redshift evolution of our IGM model to that determined by the input model alone, which may be incomplete (see, e.g., other estimations of UVB heating and ionization rates [80]). In particular, this allows us to account for uncertainty in HeII reionization, which may start heating the IGM as early as $z \sim 4.5$, without explicitly varying helium parameters in the input model. This would otherwise further extend the dimensionality of the model, which has important implications regarding the construction of the emulator (see Sec. II D).

At each redshift we restrict our sample values of $[T_0, \tilde{\gamma}, u_0]$ to those generated in our simulations; i.e., we do not employ the practice of rescaling the temperature-density relation after a simulation snapshot has been saved (e.g., [65,81]). This practice may be inaccurate when using the smallest scales in the flux power spectrum, e.g., as rescaling the temperatures of simulation particles ignores their peculiar velocities. The effect of this on the flux power spectrum may not be fully captured by other IGM parameters (i.e., u_0). Future work will assess the impact of this. However, we do employ the common practice of rescaling

mock spectrum optical depths in order to increase our sampling of τ_{eff}; Ref. [82] tested the accuracy of this procedure. This process is relatively computationally cheap and means that, for each simulation snapshot, we can have an arbitrary number of effective optical depth values; we discuss this further in Sec. II D.

5. Cosmology

The flux power spectrum at these redshifts ($z \sim 5$) and scales ($k_f \sim 10^{-2} - 10^{-1}$ s km^{-1}) is sensitive to aspects of the cosmological model, which must be marginalized for robust nCDM bounds. In particular, to search for a cutoff in the power spectrum arising from an nCDM model, we marginalize over the amplitude $A_s \in [1.2 \times 10^{-9}, 2.5 \times 10^{-9}]$ and slope $n_s \in [0.9, 0.995]$ of the primordial power spectrum. We note that these differ from the standard cosmological parameters [83] as we define them at a pivot scale $k_p = 2$ Mpc^{-1}, which lies in the range of scales we are considering. This requires translating CMB priors from the larger pivot scale used by the Planck Collaboration. In the data analysis we present in Ref. [84], we use *Planck*-derived priors on the primordial power spectrum. It follows that during the acquisition of training points during Bayesian optimization (see Sec. III B), these are largely drawn from that prior distribution (through the exploitation term) and so are clustered around a *Planck* fiducial cosmology (see Figs. 4 and 6). We also vary in the initial set of simulations, the fractional matter energy density $\Omega_m \in [0.26, 0.33]$, although in our initial optimization analysis presented below, we fix it to a fiducial value from *Planck* [83,85], $\Omega_m = 0.3209$. All other cosmological parameters are fixed to fiducial values from the *Planck* 2018 results [85], in particular, $\Omega_b h^2 = 0.022126$ and $\Omega_c h^2 = 0.12068$.

D. Gaussian process emulation and likelihood function

In order to infer bounds on nCDM parameters and to marginalize over the nuisance parameters identified in Sec. II C, we want to be able to model data as a function of theory parameters. A challenge arises because each evaluation of our theory vector naively requires the calculation of a hydrodynamical simulation (see Sec. II B) and numerical methods of inference [e.g., Markov-chain Monte Carlo sampling (MCMC)] may require millions of these evaluations. This would be computationally infeasible given the numerical requirements for sufficiently converged simulations (see Appendix A). We follow the solution we presented in Refs. [41,42], which is to construct a "surrogate model" or "emulator" for the simulations using a Gaussian process (e.g., [43]) and to ensure convergence and sufficient accuracy in its construction by the method of Bayesian emulator optimization. A Gaussian process emulator is sufficiently computationally cheap to evaluate that standard inference methods like MCMC can be used.

---

[1]This assumes the IGM to be optically thin; we discuss in Sec. V where this model breaks down, e.g., the impact of radiative transfer (e.g., [75]), as well as the implications for dark matter bounds.

We emulate separately at each redshift $z = [4.2, 4.6, 5.0]$, the flux power spectrum as a function of the nCDM parameters (Sec. II A) and the nuisance IGM and cosmological parameters (Sec. II C): $\boldsymbol{\theta} = [\alpha, \beta, \gamma, \tau_0(z = z_i), T_0(z = z_i), \tilde{\gamma}(z = z_i), u_0(z = z_i), n_s, A_s, \Omega_m]$. As discussed in Sec. II B, we emulate the flux power spectrum at 45 sample wave numbers in inverse-distance units ($h\,\mathrm{Mpc}^{-1}$). This accounts for the fact that the velocity wave numbers of a given simulation depend on Ω_m (by the Hubble law) and so emulating in velocity space would be complicated by having simulation modes binned differently at each training point. In the likelihood function, we then linearly interpolate the flux power spectrum to velocity wave number bins.

1. Gaussian process

Our emulator model is a Gaussian process; this is a stochastic process such that a finite number of simulation outputs (i.e., the flux power spectrum) forms a multivariate Gaussian distribution: $P_f(\boldsymbol{\theta}) \sim \mathcal{N}(0, K(\boldsymbol{\theta}, \boldsymbol{\theta}'; \boldsymbol{\psi}))$. Here, for clarity, we have dropped in the notation the dependence of P_f on k_f and z. The free hyperparameters $\boldsymbol{\psi}$ of the emulator model for $P_f(\boldsymbol{\theta})$ will be optimized (or "trained") using a set of "training simulations" $P_f(\boldsymbol{\theta}')$ evaluated at a set of points $\boldsymbol{\theta}'$. The zero mean condition is approximated by normalizing the flux power spectra by the median value in the training set. Having chosen the training set of simulations (see Sec. II D 2), the final part of the model to determine is the covariance kernel $K(\boldsymbol{\theta}, \boldsymbol{\theta}'; \boldsymbol{\psi})$. We use a linear combination of a squared exponential kernel [or radial basis function (RBF); $\sigma_{\mathrm{RBF}}^2 \exp\left(-\frac{(\boldsymbol{\theta}-\boldsymbol{\theta}')^2}{2l^2}\right)$], a linear kernel [$\sigma_{\mathrm{linear}}^2 \boldsymbol{\theta}.\boldsymbol{\theta}'$] and a constant noise kernel [$\sigma_{\mathrm{noise}}^2$]; i.e., $\boldsymbol{\psi} = [\sigma_{\mathrm{RBF}}^2, \sigma_{\mathrm{linear}}^2, \sigma_{\mathrm{noise}}^2, l]$. The Gaussian process model constitutes a wide prior in function space and so allows emulation of the flux power spectrum without any strong prior knowledge of its parameter dependence. Moreover, the model is probabilistic and so any prediction of the flux power spectrum away from the training points will have an associated uncertainty which can be propagated to our statistical model. The choice of covariance kernel is motivated by wanting a flexible model (hence the squared exponential), recognizing that linear interpolation has been a reasonable approach in previous efforts (e.g., [27]) (hence the linear kernel) and having a term which can account for noise in the training data (although we find that this is always negligible).

2. Training simulations

In the construction of the training data, we adapt the methods we presented in Refs. [41,42] to this more general parametrization. We run an initial set of fifty simulations spanning the wide prior range on the nCDM, IGM and cosmological parameters set out above (Secs. II A and II C). To achieve this, we choose the parameter values

of these simulations from a Latin hypercube sampling scheme [86]. This sampling scheme is chosen because it has good space-filling properties with a low number of sampling points. It distributes N training points such that along each parameter axis, the full space is sampled in N equally divided subspaces. We generate many different hypercubes and use the one that maximizes the minimum Euclidean distance between samples. This Latin hypercube is constructed for our input simulation parameters, i.e., $[\alpha, \beta, \gamma, H_A, H_S, z_{\mathrm{rei}}, T_{\mathrm{rei}}, n_s, A_s, \Omega_m]$, since we do not know *a priori* the mapping from input to output IGM parameters. This means that in our initial set of training simulations, the IGM parameters $[T_0(z = z_i), \tilde{\gamma}(z = z_i), u_0(z = z_i)]$ are not sampled by a Latin hypercube. We find, however, that these parameters are still evenly sampled (see Fig. 4 for the distribution of training points) and the subsequent application of Bayesian emulator optimization will correct for any loss of precision. A partial degeneracy appears between $T_0(z = z_i)$ and $u_0(z = z_i)$ such that snapshots with high T_0 and low u_0 (and vice versa) are not produced. This constitutes a physical prior on these parameters as very hot IGMs with little previous heating (and vice versa) would be contrived. We therefore set a prior probability distribution that excludes parameter space outside the convex hull of the training points in this plane (at each redshift). The full details of the prior distribution specific to the data analysis we present in Ref. [84] can be found in that reference.

We also note that $\tau_0(z = z_i)$ does not form part of the Latin hypercube since (as discussed in Sec. II C), we can have an arbitrary number of τ_0 samples by rescaling mock spectrum optical depths. For each training simulation, we have ten values of $\tau_0(z = z_i) \in [0.75, 1.25]$ (i.e., allowing for 25% corrections to the fiducial model). These are evenly distributed such that across all the training points, we have $10N = 500$ different values of τ_0 for maximal support in this dimension.

The framework we propose is as follows: the initial base nCDM emulator can be used for parameter inference on the many particular nCDM models it can characterize (see Sec. II A). This means that the process of Bayesian emulator optimization (the iterative addition of training simulations until parameter inference has converged) should be applied for each particular set of dark matter model parameters (and nuisance IGM/cosmological parameters). In Sec. III, we demonstrate the example of optimizing the emulator for inferring a bound on the mass of the ultralight axion. Other dark matter model parameters can be inferred by equivalent analyses. We therefore defer a discussion of the specifics of the Bayesian emulator optimization to Sec. III B.

3. Emulator training

We now discuss how a given emulator (whether optimized or not) propagates to the final parameter inference.

With a given training set, we optimize (or "train") the hyperparameters $\boldsymbol{\psi}$ of the Gaussian process by maximizing the marginal likelihood of the training data. We train each redshift separately but use the same set of optimized hyperparameters for each wave number within each redshift bin. We do not model correlations between different wave numbers; Ref. [81] found this to be a sufficient approximation. We can then make predictions for the flux power spectrum at arbitrary values of $\boldsymbol{\theta}^*$ by evaluating the posterior predictive distribution (PPD) for the flux power spectrum conditional on the training data. This is a Gaussian distribution with mean $\boldsymbol{\mu} = K_* K^{-1} P_f(\boldsymbol{\theta}')$ and variance $\boldsymbol{\sigma}^2 = K_{**} - K_* K^{-1} K_*^{\mathrm{T}}$, where $K_* = K(\boldsymbol{\theta}^*, \boldsymbol{\theta}'; \boldsymbol{\psi})$ and $K_{**} = K(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*; \boldsymbol{\psi})$. The generic behavior of the Gaussian process PPD is that its uncertainty increases as a function of distance to the nearest training points. It follows that more precise theoretical estimates of the flux power spectrum will in general be made in regions of parameter space which are more densely sampled by training data; the Bayesian emulator optimization exploits this feature of Gaussian processes (see Sec. III B).

4. Likelihood function

In this analysis, we use a Gaussian likelihood function, although this initial emulator could be used with any likelihood function that can robustly propagate the theoretical uncertainty arising from the Gaussian process emulation. The theory vector is taken to be the PPD mean $\boldsymbol{\mu}(\boldsymbol{\theta})$ and the uncertainty in the flux power spectrum emulation is propagated by adding in quadrature the diagonal PPD covariance $\boldsymbol{\sigma}^2(\boldsymbol{\theta})$ to the data covariance.

III. ULTRALIGHT AXION EMULATOR

In this section, we present the construction of the Bayesian-optimized emulator for inferring a bound on the mass of ultralight axions as the dark matter using the Lyman-alpha forest flux power spectrum. This emulator is built using the base nCDM emulator presented in Sec. II. It constitutes an example of how to construct an optimized emulator for a particular dark matter model characterized by the nCDM parameters. Other dark matter candidates could be studied in an equivalent way. In Sec. III A, we explain the ultralight axion model we employ; and in Sec. III B, we discuss the method of Bayesian emulator optimization and how this connects to estimation of the posterior probability distribution.

A. Ultralight axion model

Figures 2 and 3 demonstrate the construction of our ultralight axion dark matter model. We fit (using the Levenberg-Marquardt algorithm [87,88]) the nCDM parameters $[\alpha, \beta, \gamma]$ to the ULA transfer functions (as calculated by the Boltzmann code axionCAMB [89]) as a function of axion mass m_a. We use polynomial functions
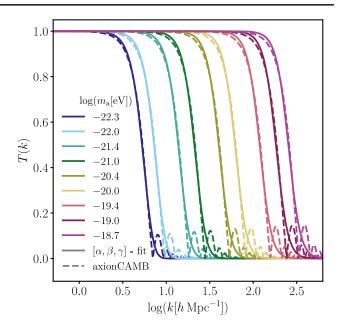

FIG. 2. The transfer function $T(k)$ [as defined in Eq. (1) at $z = 99$] for the ultralight axion dark matter model as a function of comoving distance wave number k and axion mass m_a. The dashed lines show $T(k)$ as calculated by the Boltzmann code axionCAMB [89]; the solid lines show our nCDM model fit using the $[\alpha, \beta, \gamma]$ parameters. The nCDM model cannot characterize the small-scale oscillations in the axionCAMB transfer functions but in the power spectrum, these are strongly suppressed relative to the initial cutoff scale.

to model the dependences of $[\alpha, \beta, \gamma]$ on m_a.[2] We fit the α dependence at the fiducial value of $h = 0.6686$ [85], to which h is fixed in this part of the analysis (see Sec. III B). No particular physical meaning should be ascribed to the polynomial model we have employed; it serves as a simple mapping from axion mass to our nCDM parameters. The polynomial model strictly only applies within our prior bounds on axion mass, in which our model parameters are fit, for $\log(m_a[\mathrm{eV}]) \in [-22, -19]$.

Our ground truth for the ULA transfer functions is given by axionCAMB, [90] a modified version of the Einstein-Boltzmann solver CAMB [91]. It solves the coupled Friedmann/Klein-Gordon system of equations for a homogeneous ULA field in an expanding universe. At early times, when $m_a \ll a\mathcal{H}$ (a is the cosmological scale factor; $\mathcal{H}$ is the conformal Hubble parameter), the evolving axion equation of state $w(a)$ and adiabatic sound speed are used to evolve axion energy and pressure perturbations along with the other cosmological energy components. At late times, when $m_a \gg a\mathcal{H}$, the equations become stiff and the axion field is well-approximated as a pressureless fluid ($w = 0$) and the perturbations are evolved in the

---

[2]$\log(\alpha[h^{-1}\,\mathrm{Mpc}]) = 5.5 \times 10^{-3} \mathcal{M}^3 + 0.33 \mathcal{M}^2 + 6.2 \mathcal{M} + 33.1$; $\beta = -0.026 \mathcal{M}^2 - 0.82 \mathcal{M} - 0.45$; $\log(-\gamma) = -0.013 \mathcal{M}^3 - 0.75 \mathcal{M}^2 - 14.7 \mathcal{M} - 9.6$; where $\mathcal{M} = \log(m_a[\mathrm{eV}])$.
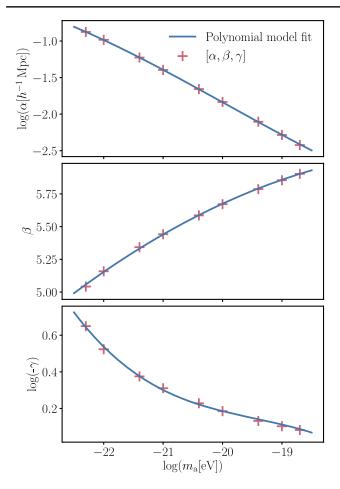
FIG. 3. The nCDM model fit for ultralight axion dark matter transfer functions as a function of axion mass m_a. *From top to bottom*, the three nCDM parameters $[\alpha, \beta, \gamma]$, showing a polynomial model fit (see text for more details).

Wentzel-Kramers-Brillouin approximation for ULAs with scale-dependent sound speed (e.g., [92]). This approximation captures the suppression of small-scale axion perturbations, observed as the sharp cutoff in the transfer functions in Fig. 2, without requiring the resolution of the short ULA oscillation timescale ($\sim m_a^{-1}$). We find that the nCDM model [Eq. (1)] can well characterize the power spectrum cutoff produced by axionCAMB (Fig. 2). It is not able to characterize the suppressed small-scale oscillations in the transfer function. However, even in the initial conditions of our simulations (Sec. II B) in the matter power spectrum, these oscillations will be further suppressed since $P(k) \propto T^2(k)$. They will be erased further again in propagation to our observable, the flux power spectrum at $z \sim 5$ by nonlinearities in the density field and the projection from a 3D to a 1D line-of-sight power spectrum. Considering that current data measure the flux power spectrum at the precision of $\sim 10\%$–25% (e.g., [65]), we conclude that our bounds on axion mass will be insensitive to this approximation and driven by the

nondetection of the initial sharp cutoff in the power spectrum. We will discuss this further and compare to the approximations in previous efforts in Sec. V. We consider only the case where ULAs form the entirety of the dark matter; future analyses can extend our parameter space to consider subdominant contributions to the dark matter energy density (see Sec. V).

B. Bayesian-optimized emulator and posterior distribution

In order to infer a bound on the mass of ultralight axions as the entirety of the dark matter (and marginalize over nuisance IGM/cosmological parameters; see Sec. II C), we must numerically sample the posterior probability distribution for the full set of inference parameters $\boldsymbol{\phi} = [\log(m_a[\text{eV}]), \tau_0(z=z_i), T_0(z=z_i), \tilde{\gamma}(z=z_i), u_0(z=z_i), n_s, A_s]$, for $z_i = [4.2, 4.6, 5.0]$. We use the Gaussian likelihood function presented in Sec. II D.

Although Ω_m is varied in our emulator (Sec. II C), since relevant data do not constrain this parameter, here we fix $\Omega_m = 0.3209$, which is equivalent to $h = 0.6686$ [85]. We estimate the posterior probability distribution (with a given prior distribution) using the affine-invariant MCMC ensemble sampler EMCEE [93].

In order to evaluate the theory flux power spectrum, we use the nCDM emulator described in Sec. II. We map from $\log(m_a[\text{eV}])$ to $[\alpha, \beta, \gamma]$ using the model constructed in Sec. III A. However, as mentioned in Sec. II D, the initial base emulator with fifty training simulations is unoptimized and not guaranteed to give converged estimates of the posterior distribution. This can arise when the uncertainty in emulator prediction (the variance in its posterior predictive distribution) dominates over the data error (covariance) at the peak of the true posterior distribution (which is not *a priori* known). This can lead to weakened parameter constraints (as the emulator error broadens the peak of the likelihood) or even bias determination of the maximum posterior point due to the nonstationarity of the emulator variance. We account for this problem by optimizing the emulator training set (iteratively adding to it) such that the true posterior peak is determined; this is the Bayesian emulator optimization we introduced in Refs. [41,42].

At each iteration of the Bayesian emulator optimization, after training the emulator on the existing training set and evaluating the posterior distribution by MCMC (see above), an acquisition function is maximized in order to determine the position in parameter space for the next training simulation. Following Ref. [41], we use the modified GP-UCB (Gaussian process upper confidence bound) acquisition function [94–97] $\mathcal{A}(\tilde{\boldsymbol{\phi}}) = \mathcal{P}(\tilde{\boldsymbol{\phi}}|\boldsymbol{d}) + \tilde{\alpha}\boldsymbol{\sigma}^{\mathrm{T}}(\tilde{\boldsymbol{\phi}})\Sigma^{-1}\boldsymbol{\sigma}(\tilde{\boldsymbol{\phi}})$. $\mathcal{P}(\tilde{\boldsymbol{\phi}}|\boldsymbol{d})$ is the natural logarithm of the posterior probability given data $\boldsymbol{d}$; Σ is the data covariance matrix; and $\boldsymbol{\sigma}(\tilde{\boldsymbol{\phi}})$ is the emulator error vector (or standard deviation of the Gaussian process PPD; see Sec. II D). Following Refs. [98,99], we set the acquisition hyperparameter $\tilde{\alpha} = 0.85$; this is roughly

equivalent to having $\sim 1\sigma$ confidence in estimation of the posterior distribution during optimization. Since, as discussed in Sec. II, we can have an arbitrary number of training points in the $\tau_0(z = z_i)$ dimensions due to a computationally cheap simulation postprocessing, we do not optimize the training set in these dimensions. In the acquisition function, $\tilde{\phi}$ is the truncated set of inference parameters without $\tau_0(z = z_i)$. It can be seen that the acquisition function has two terms which balance (in the first term) *exploitation* of approximate knowledge of our objective function (the posterior distribution) and (in the second term) *exploration* of the full prior volume. The exploitation term prefers adding training simulations at the peak of the posterior (which is the most important region to characterize accurately). The exploration term prefers adding simulations where uncertainty in the estimation of the true (zero emulator error) posterior is large, which is equivalent to regions where the ratio of emulator to data covariance is large. The linear combination balances these two components. In order to prevent the exploitation term getting stuck in local or spurious maxima of the true posterior, the exploration term and the stochastic element in the acquisition (see below) are vital. It is possible to vary the balance between exploitation and exploration using the acquisition hyperparameter.

We find the maximum of the acquisition function using a truncated Newton algorithm [100]. In order to ensure we explore the full peak of the posterior distribution (i.e., the 95% to 99% credible region), we then choose the position of the next training simulation by adding a random displacement (drawn from a multivariate Gaussian distribution) to the maximum acquisition point [101]. The size of this displacement is tuned to the size of the credible region which we want to characterize accurately. A further subtlety arises since the acquisition function is defined on the space of the (truncated) inference parameters, which include the output IGM parameters $[T_0, \tilde{\gamma}, u_0]$. In order to determine which input IGM parameters to use for the next training simulation, we map back to $[H_A, H_S, z_{rei}, T_{rei}]$ using radial basis function interpolation between the existing training set [102]. This method is related to, but less general than, the Gaussian process emulation and serves as a simple yet robust method for this mapping. Once the next training simulation is run, the $\tau_0(z = z_i)$ samples are redistributed such that the $10N$ (N is the total number of training simulations including the optimization set) points evenly sample the full prior range. The emulator is then retrained (the emulator hyperparameters ψ are reoptimized) and the posterior distribution is reestimated by MCMC using the new iteration of the emulator.

The process of Bayesian emulator optimization continues and training simulations are added to the emulator until convergence is observed. We use improved "double-lock" convergence criteria. As in Ref. [41], one criterion is that successive estimations of the posterior distribution (e.g., as summarized by marginalized statistics like the mean and 1σ

and 2σ constraints) do not change within a given tolerance; we show the results of this convergence test in Sec. IV and Fig. 7. The second criterion is that the exploration term of the acquisition function evaluated at the parameter positions of optimization simulations converges towards zero. This ensures that the acquisition of new training data is then dominated by exploitation, i.e., that training points are being added at the peak of the posterior and that the ratio of emulator to data covariance is tending towards zero in this region. The results of this convergence test are shown in Sec. IV and Fig. 5.

In this analysis, we use the batch version of Bayesian emulator optimization [41]. This involves, at each iteration, proposing a batch of optimization simulations at once, running these in parallel and then retraining the emulator only after the entire batch has been added to the training set. The first simulation in a batch of a given size is chosen as detailed above. Subsequent members of a batch are chosen in a similar way, except that the output (flux power spectra) from previous members are not yet known. However, the positions in parameter space for these members are known and this information can at least update the emulator error distribution, since the PPD variance depends only on the positions of training data and not the data themselves (see Sec. II D). This partial information updates the exploration term of the acquisition function for iterations within a batch. Batch optimization can be preferred to balance efficient parallel use of computational resources with the loss of optimality in the acquisition of new training data. In this study, we start with three batches of five and then switch to batches of two, although we note that the third and sixth batches are short of one simulation each, owing to hardware issues that led to an incomplete calculation.

IV. TESTS

In this section, we conduct tests of convergence (Sec. IV A) and cross-validation (Sec. IV B) on the Bayesian-optimized ULA emulator presented in Sec. III.

A. Tests of convergence

Figure 4 indicates the distribution of emulator training points, projected onto each parameter plane. It can be seen that the initial Latin hypercube set of fifty (see Sec. II; most lightly shaded in Fig. 4) fully spans the parameter space. The optimization simulations (43 in total; see Sec. III; more darkly shaded in Fig. 4) concentrate in a smaller subspace, which corresponds to the peak of the posterior distribution (see Fig. 6 and below). There is however a degeneracy between $T_0(z = z_i)$ and $u_0(z = z_i)$ since it is unphysical to generate simulations with high T_0 and low u_0 and vice versa. We reiterate that these are not samples from the posterior distribution and that the posterior is always estimated with fully converged MCMC chains.

Figure 5 shows the first of the two convergence tests we use. It indicates how the exploration term of the Bayesian
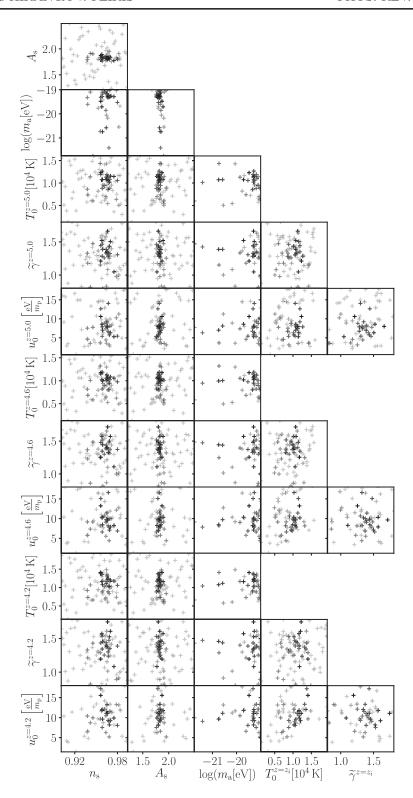
FIG. 4. The positions of the Bayesian-optimized ULA emulator training simulations projected onto each parameter plane. The shade of the crosses indicates the iteration at which the training point is added, such that the initial set is shaded lightest and the final optimization point is shaded darkest. We show only the planes where parameters are emulated together (i.e., not the planes with parameters from different redshift bins; or the $\tau_0(z = z_i)$ planes which are densely sampled by simulation postprocessing, see Sec. II B). It follows that on the x axis, z_i corresponds to the redshift indicated on the y axis. We note that for $\log(m_a[\mathrm{eV}])$, the initial set fully spans the $[\alpha, \beta, \gamma]$ volume and does not project onto that axis; however, the initial set contributes to the final emulator model as we always emulate in $[\alpha, \beta, \gamma]$. It can be seen how the darker-shaded optimization points concentrate in a smaller subvolume, which corresponds to the peak of the posterior distribution (see Fig. 6).
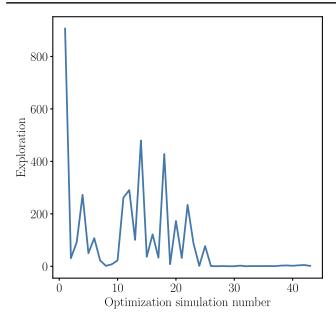
FIG. 5. The convergence of the exploration term of the acquisition function (see Sec. III B) as a function of optimization simulation number for the Bayesian-optimized ULA emulator.

optimization acquisition function (see Sec. III B) tends towards zero as more training simulations are added to the emulator. The exploration term is effectively the ratio of emulator to data covariance. Its convergence to zero reflects how the optimization becomes more dominated by exploitation, i.e., the proposal of training simulations is at the peak of the posterior and the ratio of emulator to data covariance is negligible there. This means that estimation of the posterior peak (by MCMC) tends towards the true posterior, with negligible propagated emulator error.

Figure 6 indicates the convergence of the posterior distribution (see Sec. III B) as summarized by the set of marginalized 1D and 2D distributions. For clarity, we show only the projections in the IGM parameters at $z = 4.2$; we observe similar convergence at the other redshifts that we consider. It can be seen in each projection that the posterior estimation converges as more optimization simulations are added to the emulator training set. In most projections, there is a significant shift from using the initial emulator (Sec. II) to having added 19 optimization points. In particular, the axion mass m_a goes from being (spuriously) detected to setting a lower bound. However, after this point, the shifts in the posterior are much less significant; and following the iteration after which 35 optimization points have been added, the shifts in the posterior are statistically negligible. We track this more quantitatively using summaries of the marginalized posterior (marginalized mean and 1σ and 2σ constraints).

Figure 7 shows for each model parameter, the number of sigma shift (defined by the marginalized posteriors at a given optimization epoch) in these quantities between iterations as a function of optimization simulation number.

It can be seen that by the end of the optimization, the shifts are small, indicating the desired robustness in the estimation of the posterior distribution. In Fig. 6, in the planes of parameters that are emulated together (remembering that different redshifts are emulated separately), we indicate by colored crosses the projected positions of optimization simulations. It can be seen that by Bayesian optimization, the optimization simulations are more concentrated in the posterior peak than the initial set which is evenly distributed in the full prior volume. This reduces the emulator error within the posterior peak, leading to accurate determination of the flux power spectrum (see Sec. IV B) and hence the likelihood function in this region.

B. Tests of cross-validation

We carry out leave-one-out cross-validation of the Bayesian-optimized emulator. This involves, in turn, removing one of the training simulations, retraining the reduced emulator and then comparing its prediction to the truth of the validation simulation that was removed.[3] We remove simultaneously the training simulation from each redshift bin since these are emulated separately. Figure 8 shows the distribution of the ratio of empirical-to-predicted emulator error in the flux power spectrum at the validation simulations, as a function of wave number k and redshift z. The empirical error is the difference between the emulator PPD mean (see Sec. II D) and the true flux power spectrum as measured from the validation simulation. The predicted emulator error is the standard deviation of the emulator PPD. If the emulator model is fitting well, the distribution of the error ratio should be a unit Gaussian. When the distribution is fatter than this, this indicates underfitting, i.e., that the emulator error is being underestimated and that the Gaussian process model does not have sufficient flexibility to describe accurately the training or the validation data. When the distribution is narrower than a unit Gaussian, this indicates overfitting, i.e., that the emulator error is being overestimated and that the Gaussian process has too much flexibility and is fit too closely to the training data. This has negative implications for the generalizability of the emulator model for unseen (e.g., validation) data.

In Fig. 8, the distributions are split into two sets: the left-handed halves of each violin consider the validation test only for the simulations added by Bayesian optimization, while the right-handed halves consider all the simulations including the initial Latin hypercube. In each case, we consider only the final, converged emulator and in cross-validation, optimize the emulator on all other 92 training

---

[3]This is actually leave-*ten*-out cross-validation since we remove in turn all ten effective optical depth samples (at each redshift) associated with a given training simulation. We do not remove optical depth samples one-by-one since this would be a trivial test considering that these dimensions can be densely sampled by a simulation postprocessing step. The number of optical depth samples was tested in Ref. [42].
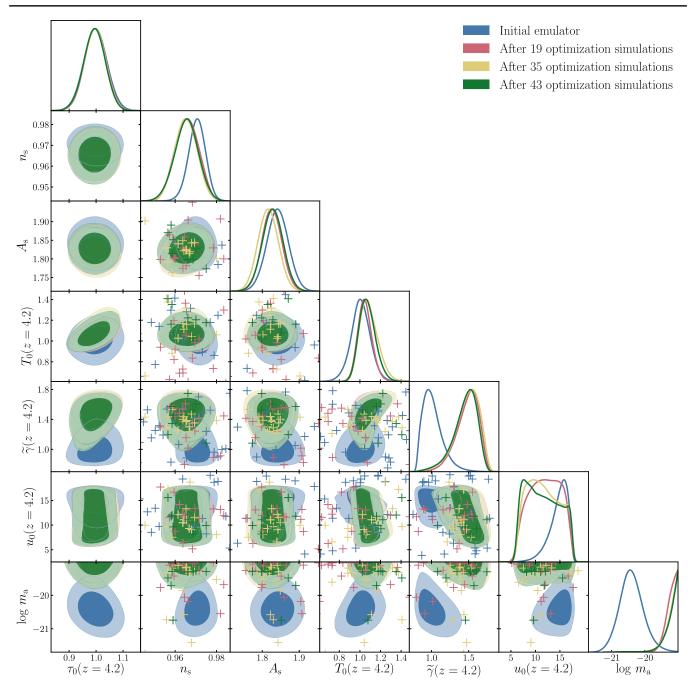
FIG. 6. The convergence of the posterior distribution (set of marginalized 1D and 2D distributions) using the Bayesian-optimized ULA emulator. Each set of colored contours shows the estimate of the posterior at various iterations of the emulator. The darker and lighter shaded regions, respectively, indicate the 68% and 95% credible regions. In each plane [except the $\tau_0(z = 4.2)$ planes which are densely sampled by simulation postprocessing, see Sec. II B], crosses indicate the projected positions of emulator training simulations. The crosses are color-coded by the iteration of the emulator by which the training points are added. We note that for $\log(m_a[\mathrm{eV}])$, the initial set fully spans the $[\alpha, \beta, \gamma]$ volume and does not project onto that axis; however, the initial set contributes to the final emulator model as we always emulate in $[\alpha, \beta, \gamma]$. $T_0(z = 4.2)$ is in units of 10^4 K; $u_0(z = 4.2)$ is in units of $\mathrm{eV}m_p^{-1}$, where m_p is the proton mass; and m_a is in units of eV. For clarity, we show only the projections onto the IGM parameters at $z = 4.2$; we observe similar convergence at the other redshifts we consider.

simulations. In splitting the distributions that are shown, we indicate the two populations that are formed. When the initial set of simulations is included, the overall distribution tends towards overestimating the emulator error, in particular on the largest scales (smallest k). However, when considering only the optimization simulations (and hence
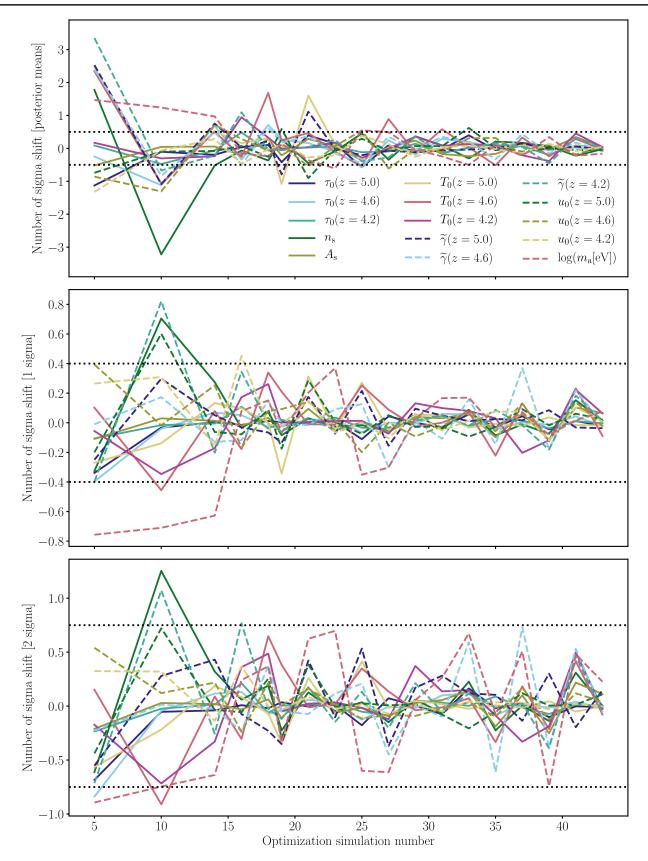
FIG. 7. The convergence of summary statistics of the posterior distribution using the Bayesian-optimized ULA emulator. *From top to bottom*, the number of sigma shift (defined by the marginalized posteriors at a given optimization epoch) between emulator iterations for the 1D marginalized posterior means, 1σ and 2σ constraints. Each colored line shows the convergence for each model parameter.
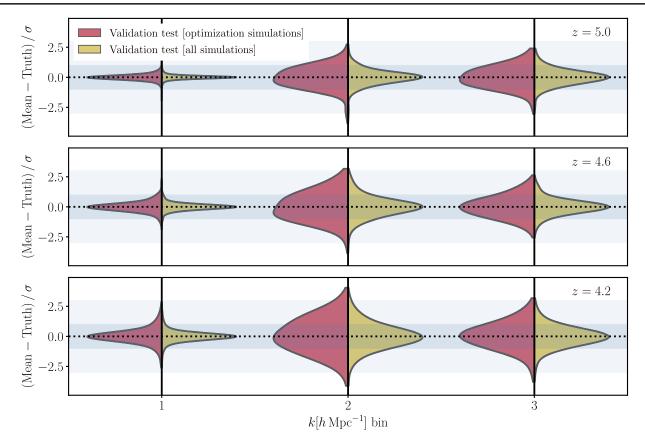
FIG. 8. A violin plot (set of kernel density estimation histograms) showing the ratio of empirical (emulator mean—truth) to predicted emulator error for leave-one-out cross-validation simulations, (*from left to right*) as a function of binned wave number k and (*from top to bottom*) redshift z. The left set of PDFs considers only the simulations added by Bayesian optimization; the right set considers all simulations including the initial set. For comparison, we indicate 1σ and 3σ limits, respectively, by the darker and lighter shaded bands. If the emulator model was fitting perfectly, each distribution would be a unit Gaussian; the distribution being fatter indicates underfitting (underestimating emulator error) and the distribution being narrower indicates overfitting (overestimating emulator error). The wave number bins span the following ranges (where k is in $h\,\mathrm{Mpc}^{-1}$): 1 $[0.6 < k < 10]$; 2 $[10 < k < 19]$; 3 $[19 < k < 28.3]$.

the training points at the peak of the posterior distribution), the Gaussian process fit is better, although the error is still overestimated on the largest scales; there is some mild underestimation on intermediate scales. Overestimating the emulator error will tend to weaken parameter constraints as large errors broaden the peak of the posterior.

Figure 9 compares the empirical and predicted emulator error (i.e., respectively, the numerator and denominator in the ratio illustrated in Fig. 8) to the data error (this is the square root of the diagonal elements of the data covariance used in our data analysis presented in Ref. [84]). The distributions of (the logarithm of) the ratio of emulator-to-data error are in general bi-modal; the lower-valued modes come from the optimization simulations, while the other modes come from the initial Latin hypercube. This reflects how the ratio of emulator-to-data error is significantly lower for the optimization simulations (and therefore at the peak of the posterior distribution) than for the initial Latin hypercube [in fact, for clarity, we only show the lower-valued optimization set mode for the empirical ratios in the right-hand probability density function (PDFs)]. This is a direct consequence (and

desired outcome) of the Bayesian optimization, which builds up training information at the posterior peak and locally reduces emulator uncertainty in the flux power spectrum, which leads to more accurate determination of the peak of the posterior distribution. The left-handed and right-handed distributions in each violin show the ratio of emulator-to-data error, respectively, for the predicted emulator error (PPD standard deviation) and the empirical emulator error (difference between PPD mean and truth). As is consistent with the test in Fig. 8, the empirical ratio is consistently shifted to lower values than the predicted ratio, reflecting how the emulator tends towards overfitting the training data and overestimating the emulator uncertainty.

Nonetheless, for the optimization simulations, the mode average of the predicted emulator-to-data error ratio is always less than one. This means that in the posterior peak ($\sim 95\%$ credible region), uncertainty in the flux power spectrum (i.e., the diagonal of the covariance matrix in the likelihood function; see Sec. II D) is dominated by uncertainty in the data. As discussed in Ref. [41], this is a key requirement for converged estimation of the posterior
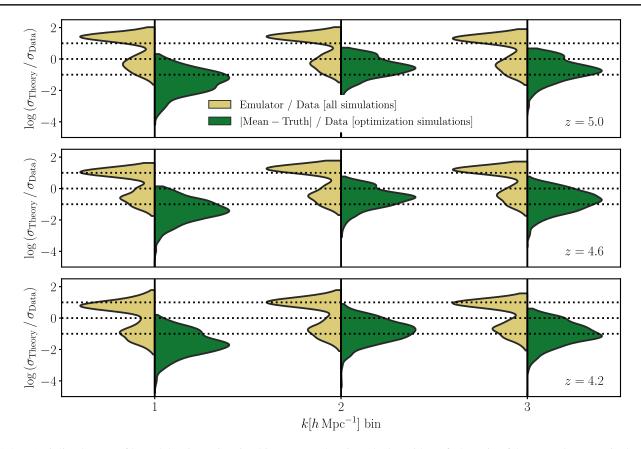
FIG. 9. A violin plot (set of kernel density estimation histograms) showing (the logarithm of) the ratio of theory to data error in the flux power spectrum, (*from left to right*) as a function of binned wave number k and (*from top to bottom*) redshift z. *For each element*, the left PDF shows the ratio of emulator error to data error; for the majority of training data in the optimization set (which constitutes the lower-valued mode in each case), this is < 1. Since Fig. 8 shows through cross-validation a tendency to overestimate the prediction uncertainty, the right PDFs show the ratio of the empirical error (for validation simulations only in the optimization set for clarity) to data error. This ratio is often $\ll 1$ (and less than the theoretical emulator-to-data error ratio, see also Fig. 8), indicating no statistically significant bias in emulation of the flux power spectrum. See more details in the text. The wave number bins span the following ranges (where k is in $h\,\mathrm{Mpc}^{-1}$): 1 [$0.6 < k < 10$]; 2 [$10 < k < 19$]; 3 [$19 < k < 28.3$].

distribution as it means that any residual uncertainty in the emulator model is not statistically significant. Further, a key motivation for the Bayesian emulator optimization that we use is to move beyond relying on heuristic assumptions on the accuracy requirements of an emulator. The tests of convergence that we present in Sec. IV A provide a stringent test that the addition of more training points (and so further reducing the emulator error) makes no statistically significant change to the inferred parameter constraints. Considering the tendency to overestimate the emulator error discussed above (in particular on the largest scales), the ratio of empirical emulator-to-data error is often < 0.1. This shows that there is no statistically significant bias (relative to the data uncertainty and accounting for the theoretical emulator error included in the likelihood function) in the emulation of the flux power spectrum.

Figure 10 shows a test of cross-validation on the inferred posterior distribution using the Bayesian-optimized ULA emulator. We show the (marginalized) posterior distribution on the ULA model and nuisance IGM and cosmological parameters given mock simulated data, which have been removed from the emulator training set and the emulator retrained on the reduced set. This is a test as to whether parameter values can be inferred without bias using an emulator for the flux power spectrum that has not seen the mock data. We see that the truth is almost perfectly recovered at the maximum (marginalized) posterior values (we expect no scatter from the mock data as we do add any noise realization) and always within a fraction of a sigma. The same is observed for the nuisance IGM parameters at other redshifts, which we do not show for clarity in the figure. We do not reoptimize (by Bayesian optimization) the emulator for the mock data, and the mock data are taken from the optimization set of the training simulations (those added by Bayesian optimization as shown in Sec. IV A).

As a test of the sensitivity of the parameter inference to the prior distribution, we vary the prior on $T_0(z = z_i)$. The blue set of contours (obscured by the red set) uses a
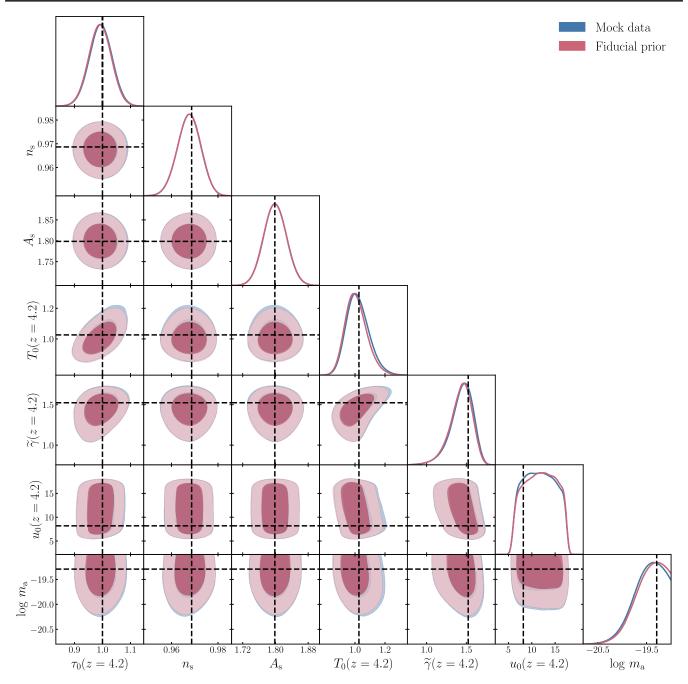
FIG. 10. Cross-validation test on the inferred posterior distribution using the Bayesian-optimized ULA emulator: set of marginalized 1D and 2D distributions given mock simulated data removed from the training set. The darker and lighter shaded regions, respectively, indicate the 68% and 95% credible regions. The blue set of contours (obscured by the red set) uses a Gaussian prior on $T_0(z = z_i)$ with means matched to the true values of the mock data (and standard deviations of 3000 K): respectively at $z = [5.0, 4.6, 4.2]$, $[10208, 9471, 10262]$ K. As a test of sensitivity to the prior distribution, the red set of contours instead uses $T_0(z = z_i)$ prior means set to the fiducial simulation values from our data analysis in Ref. [84]: respectively, at $z = [5.0, 4.6, 4.2]$, $[8022, 7651, 8673]$ K. Dashed lines indicate the true parameter values of the mock data. $T_0(z = 4.2)$ is in units of 10^4 K; $u_0(z = 4.2)$ is in units of $eV m_p^{-1}$, where m_p is the proton mass; and m_a is in units of eV. For clarity, we show only the projections onto the IGM parameters at $z = 4.2$; we observe similarly unbiased inference at the other redshifts we consider.

Gaussian prior on $T_0(z = z_i)$ with means matched to the true values of the mock data (and standard deviations of 3000 K; see caption). The red set of contours instead uses

$T_0(z = z_i)$ prior means set to the fiducial simulation values from Ref. [84] (also see caption). We see almost no resulting shift in the posterior.

V. DISCUSSION

A. Dark matter model

The model we employ to describe the effect of noncold dark matter on the matter power spectrum (first introduced in [36]) can characterize many different dark matter candidates with different physical mechanisms suppressing the small-scale power spectrum. These include, e.g., warm dark matter with a free streaming scale (e.g., [103]) or dark matter interacting with itself or baryons (e.g., [20,25]); this is discussed further in Ref. [36]. For the nCDM model to apply, the detectable effect on the power spectrum relative to the cold dark matter limit must be fully captured by modified initial conditions at $z \sim 100$. We consider here and in an accompanying data analysis [84], the example of ultralight axion dark matter (Sec. III). We fit the nCDM model to ULA matter power spectra calculated by the modified Boltzmann code axionCAMB ([89]; see Sec. III A). We find the fit to be good aside from suppressed small-scale power spectrum oscillations which cannot be captured by the nCDM model. We argue, however, that given the precision of flux power spectrum measurements on the relevant scales ($\sim 10\%$ to 25%; [65]), these suppressed oscillations are currently undetectable and so the nCDM model captures the relevant phenomenology. This compares to the approximate ULA power spectrum transfer function employed in previous Lyman-alpha forest studies [33], which is a semianalytical approximation presented in Ref. [16]. Although this model (imperfectly) captures small-scale oscillations (which do not persist to the flux power spectrum at $z \sim 5$), the initial cutoff in the power spectrum is sharper than for the full calculation made by axionCAMB. We find that the nCDM model better captures the relevant cutoff in the power spectrum (see Fig. 2).

B. Intergalactic medium model

Robust dark matter bounds using the Lyman-alpha forest must marginalize over uncertainty in the thermal and ionization state of the intergalactic medium, which leads to a similar, but distinguishable, suppression in the small-scale flux power spectrum (see Sec. II C). We model the IGM as obeying a single power-law temperature-density relation [TDR; Eq. (3)]. We also argue that pressure smoothing can be tracked by the cumulative energy deposited per unit mass at the mean density $u_0(z)$; and exploit the fact that the strength of the photoionizing UV background is degenerate in the flux power spectrum with the effective optical depth of Lyman-alpha absorption. This compares to previous Lyman-alpha forest studies of dark matter (e.g., [27,33]) which have usually tracked the pressure smoothing by varying the redshift of hydrogen reionization. When varying this parameter simultaneously with the TDR at each redshift, this can lead to marginalizing over unphysical IGMs where the energy deposited by (re)ionization is inconsistent with the instantaneous gas

temperature. This may spuriously weaken dark matter bounds. In order to avoid this, we model the IGM using only "output" parameters that describe the state of the IGM at a fixed time-slice. In the data analysis presented in Ref. [84], we prevent unphysical redshift evolution of the IGM by preventing changes in T_0 and u_0 between adjacent redshift bins greater than a conservative threshold value. We also improve the input for our simulations by employing the reionization model of Ref. [76]. Compared to previous prescriptions (e.g., [74,77–79]), it removes spurious high-redshift heating and provides additional physical flexibility by varying both the timing of and total heat injection during reionization.

However, an underlying assumption in our (and previous) IGM models is that reionization is a spatially homogeneous process. This is an approximation, as ionization and heating are expected to expand in overlapping bubbles around the (uncertain) sources of ionizing radiation (e.g., [104]). The resultant temperature and ionizing background spatial fluctuations that can persist to the IGM at $z \sim 5$ (e.g., [72,105]) have been suggested as a possible source of systematic error in Lyman-alpha forest dark matter bounds (e.g., [106]). Reference [73] found, using simulations which couple gas hydrodynamics with radiative transfer [107], that the relevant effect on the flux power spectrum is below the sensitivity of current data (e.g., [65]). Nonetheless, as modeling and data improve, future versions of the emulator can benefit from explicitly marginalizing over the effect of inhomogeneity in both hydrogen and helium reionizations (e.g., [72,105,108]).

VI. CONCLUSIONS

We have presented a general framework for testing the nature of dark matter using cosmological data and N-body simulations, specifically implemented for searching for deviation from cold dark matter using the Lyman-alpha forest. It consists of a Gaussian process emulator, trained using hydrodynamical simulations to predict how the Lyman-alpha forest flux power spectrum responds to different dark matter models, the state of the intergalactic medium and the spectrum of primordial fluctuations. This is necessary since MCMC methods for inference would otherwise be impractical owing to the computational cost of N-body simulations. The emulator exploits a flexible parametrization of the effect of noncold dark matter on the matter power spectrum (introduced by [36]). This can accurately capture the suppression in the power spectrum arising from many different dark matter models (relative to the cold dark matter limit), e.g., warm (e.g., [103]) or interacting (e.g., [20,25]) dark matter or ultralight axion dark matter [16], among others (see [36]). The emulator can be optimized for the analysis of particular dark matter models by the Bayesian emulator optimization we presented in Ref. [41]. This tests for convergence in the fidelity

of the emulator model by adaptively building up the emulator's training information.

We demonstrate the example of constructing a Bayesian-optimized emulator for the mass of ultralight axion dark matter. We carry out tests of convergence with respect to the number of hydrodynamical simulations input to the emulator; and cross-validation for the accuracy of flux power spectrum prediction. These tests confirm the robustness of the data analysis we carry out in Ref. [84]. The emulator framework we present here will facilitate the future analysis of other dark matter candidates by an equivalent procedure. We argue that the application of Bayesian optimization to emulators of cosmological "effective theories," where many different models are described by a single set of equations, can be a powerful approach for robust and computationally efficient parameter inference from the cosmic large-scale structure.

ACKNOWLEDGMENTS

APPENDIX A: NUMERICAL CONVERGENCE

In this analysis, we use a simulation mass resolution and box volume that at least matches (and often improves upon) the numerics of previous Lyman-alpha forest studies of the
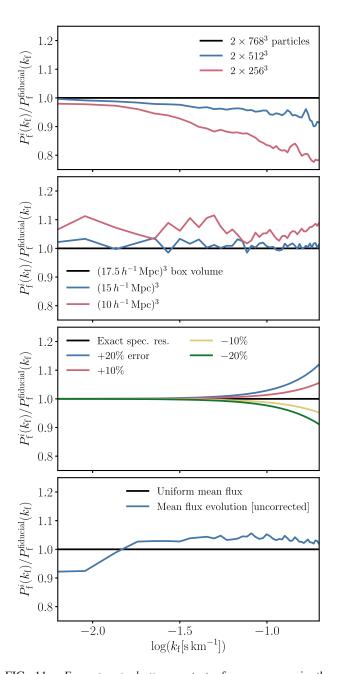

FIG. 11. *From top to bottom*: a test of convergence in the simulated flux power spectrum at $z = 5$ with respect to the number of simulation particles (for fixed comoving box volume); a test of convergence with respect to the comoving volume of the simulation box [for fixed particle mass resolution; the emulator is built with simulations with a comoving volume of $(10 \ h^{-1} \ \mathrm{Mpc})^3$ and 2×512^3 particles]; a test of the systematic error arising from misestimating the spectral resolution of the data (up to $\pm 20\%$; the uncertainty on the spectral resolution is generally estimated to be $\sim 10\%$, e.g., [65]); and a test of the systematic error arising from ignoring the redshift evolution of the mean flux in sections of Lyman-alpha forest (the residual error in our analysis will be much less since data are corrected for this effect by the "rolling mean" estimator; e.g., [65]). For comparison, uncertainty in flux power spectrum data are $\sim 10\%$ to 25% depending on wave number (e.g., [65]).

small-scale matter power spectrum ($k_f \sim 10^{-1}$ s km^{-1}; e.g., [65]); here, we test convergence in these properties. The top panel of Fig. 11 shows a test of convergence in the simulated flux power spectrum with respect to the number of simulation particles, while fixing the comoving volume of the simulation box to $(10 \ h^{-1} \, \mathrm{Mpc})^3$. This is equivalently a test of convergence with respect to the mass resolution of the simulation. In each case, we rescale mock spectrum optical depths such that each simulation has the same mean transmitted flux in order to remove this parameter dependence. We observe convergence as the number of particles is increased (and hence the mass resolution is improved). We find the correction from having more particles than our nominal value (2×512^3) never to be greater than 10% and always less than the typical uncertainty in flux power spectrum data (e.g., [65]). We carry out this test for fiducial model parameters corresponding to a 2 keV warm dark matter particle, since it is known that convergence criteria are more stringent in the presence of suppressed initial conditions (e.g., considering the effect of numerical fragmentation; [109]. We therefore consider this a robust test of numerical convergence.

The second panel from the top of Fig. 11 shows a test of convergence with respect to the comoving volume of the simulation box, while fixing the mass resolution of the simulation. We observe convergence as the box volume is increased. We find the correction from having a larger box than our nominal value $(10 \ h^{-1} \, \mathrm{Mpc})^3$ to be $\sim 10\%$ or less, which is less than the typical data uncertainty. We find no statistically significant change in the convergence in the observationally relevant redshift window $z \sim 4 - 5$.

APPENDIX B: SPECTRAL RESOLUTION

The second panel from the bottom of Fig. 11 shows a test of the systematic error in the flux power spectrum arising from a misestimation of the spectral resolution. We find that overestimating the resolution by 20% (we take the fiducial full-width-at-half-maximum to be 6 km s^{-1}, e.g., [65]) increases by $\sim 10\%$ the smallest-scale power spectrum mode that we consider. This arises since the smoothing in the transmitted flux that comes from a finite spectral resolution is overcorrected in the flux power spectrum (see Ref. [66] for the full model) and too much power is added on small scales. We find the opposite effect when underestimating the spectral resolution. The typical estimated uncertainty on the spectral resolution is $\sim 10\%$, which we find would have no statistically significant effect on the flux power spectrum ($< 5\%$) compared to typical data uncertainty ($\sim 10\%$ to 25%; e.g., [65]). This would remain true even if the resolution uncertainty was underestimated by a factor of two.

APPENDIX C: MEAN FLUX EVOLUTION

The bottom panel of Fig. 11 shows a test of the systematic error in the flux power spectrum at $z = 5$ arising from ignoring the redshift evolution of the mean transmitted flux in individual sections of Lyman-alpha forest. We model this (without resorting to radiative transfer) by injecting a redshift dependence in the effective optical depth to mock spectra generated from a simulation snapshot at $z = 5$, using the fiducial model presented in Ref. [65]. We then calculate the flux power spectrum from this snapshot ignoring the redshift dependence, i.e., following the default procedure of normalizing the flux by the mean in the whole box. We compare this to the flux power spectrum measured from the same snapshot without the injected redshift dependence, but with the optical depths uniformly rescaled to match the global mean flux in the two settings. We find an effect on all scales since the 1D flux power spectrum is a nonlinear mapping of the optical depths that we modify. However, as expected, the largest effect is on large scales since we are effectively modifying a long wavelength mode in the optical depth. Nonetheless, the effect on the power spectrum is always $< 10\%$ and less than the typical data error (e.g., [65]). Further, we expect the residual error in our analysis to be much less than this since the effect is corrected in the data by using a "rolling mean" estimator [65]. Rather than normalizing the transmitted flux by a globally estimated mean, the mean flux is estimated locally at each pixel using a finite window centered on that point. This corrects for the redshift evolution in the mean flux within sections of Lyman-alpha forest. We therefore expect only a residual error in comparing to simulations at a fixed time-slice. We expect the effect to weaken at lower redshifts. This is because the effective optical depth evolves more slowly at later times and the redshift path length for a given comoving interval also decreases with redshift.

[1] A. Boveia and C. Doglioni, Dark matter searches at colliders, Annu. Rev. Nucl. Part. Sci. **68**, 429 (2018).

[2] P. J. Fox, G. Jung, P. Sorensen, and N. Weiner, Dark matter in light of LUX, Phys. Rev. D **89**, 103526 (2014).

[3] G. Jungman, M. Kamionkowski, and K. Griest, Supersymmetric dark matter, Phys. Rep. **267**, 195 (1996).

[4] V. Gluscevic, Y. Ali-Haimoud, K. Bechtol, K. K. Boddy, C. Boehm, J. Chluba, F.-Y. Cyr-Racine, C. Dvorkin, D. Grin,

J. Lesgourgues *et al.*, Cosmological probes of dark matter interactions: The next decade, arXiv:1903.05140.

[5] A. Drlica-Wagner, Y.-Y. Mao, S. Adhikari, R. Armstrong, A. Banerjee, N. Banik, K. Bechtol, S. Bird, K. K. Boddy, A. Bonaca *et al.*, Probing the fundamental nature of dark matter with the large synoptic survey telescope, arXiv:1902.01055.

[6] P. J. E. Peebles, Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations, Astrophys. J. Lett. **263,** L1 (1982).

[7] G. R. Blumenthal, S. Faber, J. R. Primack, and M. J. Rees, Formation of galaxies and large-scale structure with cold dark matter, Nature (London) **311,** 517 (1984).

[8] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. White, The evolution of large-scale structure in a universe dominated by cold dark matter, Astrophys. J. **292,** 371 (1985).

[9] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, The statistics of peaks of Gaussian random fields, Astrophys. J. **304,** 15 (1986).

[10] R. D. Peccei and H. R. Quinn, CP Conservation in the Presence of Pseudoparticles, Phys. Rev. Lett. **38,** 1440 (1977).

[11] S. Weinberg, A New Light Boson?, Phys. Rev. Lett. **40,** 223 (1978).

[12] F. Wilczek, Problem of Strong p and t Invariance in the Presence of Instantons, Phys. Rev. Lett. **40,** 279 (1978).

[13] J. Preskill, M. B. Wise, and F. Wilczek, Cosmology of the invisible axion, Phys. Lett. **120B,** 127 (1983).

[14] L. Abbott and P. Sikivie, A cosmological bound on the invisible axion, Phys. Lett. **120B,** 133 (1983).

[15] M. Dine and W. Fischler, The not-so-harmless axion, Phys. Lett. **120B,** 137 (1983).

[16] W. Hu, R. Barkana, and A. Gruzinov, Fuzzy Cold Dark Matter: The Wave Properties of Ultralight Particles, Phys. Rev. Lett. **85,** 1158 (2000).

[17] K. Enqvist, K. Kainulainen, and J. Maalampi, Resonant neutrino transitions and nucleosynthesis, Phys. Lett. B **249,** 531 (1990).

[18] J. König, A. Merle, and M. Totzauer, keV sterile neutrino dark matter from singlet scalar decays: The most general case, J. Cosmol. Astropart. Phys. 11 (2016) 038.

[19] A. Boyarsky, J. Lesgourgues, O. Ruchayskiy, and M. Viel, Lyman-α constraints on warm and on warm-plus-cold dark matter models, J. Cosmol. Astropart. Phys. 05 (2009) 012.

[20] C. Dvorkin, K. Blum, and M. Kamionkowski, Constraining dark matter-baryon scattering with linear cosmology, Phys. Rev. D **89,** 023519 (2014).

[21] K. Sigurdson, M. Doran, A. Kurylov, R. R. Caldwell, and M. Kamionkowski, Dark-matter electric and magnetic dipole moments, Phys. Rev. D **70,** 083501 (2004).

[22] A. Loeb and N. Weiner, Cores in Dwarf Galaxies from Dark Matter with a Yukawa Potential, Phys. Rev. Lett. **106,** 171302 (2011).

[23] F.-Y. Cyr-Racine, K. Sigurdson, J. Zavala, T. Bringmann, M. Vogelsberger, and C. Pfrommer, ETHOS—An effective theory of structure formation: From dark particle physics to the matter distribution of the Universe, Phys. Rev. D **93,** 123527 (2016).

[24] L. Ackerman, M. R. Buckley, S. M. Carroll, and M. Kamionkowski, Dark matter and dark radiation, Phys. Rev. D **79,** 023519 (2009).

[25] D. N. Spergel and P. J. Steinhardt, Observational Evidence for Self-Interacting Cold Dark Matter, Phys. Rev. Lett. **84,** 3760 (2000).

[26] S. Chabanier, M. Millea, and N. Palanque-Delabrouille, Matter power spectrum: from Ly α forest to CMB scales, Mon. Not. R. Astron. Soc. **489,** 2247 (2019).

[27] V. Iršič, M. Viel, M. G. Haehnelt, J. S. Bolton, S. Cristiani, G. D. Becker, V. D'Odorico, G. Cupani, T.-S. Kim, T. A. M. Berg *et al.*, New constraints on the free-streaming of warm dark matter from intermediate and small scale Lyman-α forest data, Phys. Rev. D **96,** 023522 (2017).

[28] A. Garzilli, O. Ruchayskiy, A. Magalich, and A. Boyarsky, How warm is too warm? Towards robust Lyman-α forest bounds on warm dark matter, arXiv:1912.09397.

[29] N. Palanque-Delabrouille, C. Yèche, N. Schöneberg, J. Lesgourgues, M. Walther, S. Chabanier, and E. Armengaud, Hints, neutrino bounds, and WDM constraints from SDSS DR14 Lyman-α and Planck full-survey data, J. Cosmol. Astropart. Phys. 04 (2020) 038.

[30] R. Murgia, V. Iršič, and M. Viel, Novel constraints on noncold, nonthermal dark matter from Lyman-α forest data, Phys. Rev. D **98,** 083540 (2018).

[31] W. L. Xu, C. Dvorkin, and A. Chael, Probing sub-GeV dark matter-baryon scattering with cosmological observables, Phys. Rev. D **97,** 103530 (2018).

[32] M. Archidiacono, D. C. Hooper, R. Murgia, S. Bohr, J. Lesgourgues, and M. Viel, Constraining dark matter-dark radiation interactions with CMB, BAO, and Lyman-α, J. Cosmol. Astropart. Phys. 10 (2019) 055.

[33] V. Iršič, M. Viel, M. G. Haehnelt, J. S. Bolton, and G. D. Becker, First Constraints on Fuzzy Dark Matter from Lyman-α Forest Data and Hydrodynamical Simulations, Phys. Rev. Lett. **119,** 031302 (2017).

[34] E. Armengaud, N. Palanque-Delabrouille, C. Yèche, D. J. E. Marsh, and J. Baur, Constraining the mass of light bosonic dark matter using SDSS Lyman-α forest, Mon. Not. R. Astron. Soc. **471,** 4606 (2017).

[35] T. Kobayashi, R. Murgia, A. De Simone, V. Iršič, and M. Viel, Lyman-α constraints on ultralight scalar dark matter: Implications for the early and late universe, Phys. Rev. D **96,** 123514 (2017).

[36] R. Murgia, A. Merle, M. Viel, M. Totzauer, and A. Schneider, "Non-cold" dark matter at small scales: A general approach, J. Cosmol. Astropart. Phys. 11 (2017) 046.

[37] A. G. Doroshkevich, E. V. Kotok, A. N. Poliudov, S. F. Shand arin, I. S. Sigov, and I. D. Novikov, Two-dimensional simulation of the gravitational system dynamics and formation of the large-scale structure of the universe, Mon. Not. R. Astron. Soc. **192,** 321 (1980).

[38] G. Efstathiou and J. W. Eastwood, On the clustering of particles in an expanding universe, Mon. Not. R. Astron. Soc. **194,** 503 (1981).

[39] V. Springel, N. Yoshida, and S. D. M. White, GADGET: A code for collisionless and gasdynamical cosmological simulations, New Astron. **6,** 79 (2001).

[40] M. Cataneo, L. Lombriser, C. Heymans, A. J. Mead, A. Barreira, S. Bose, and B. Li, On the road to percent accuracy: Non-linear reaction of the matter power spectrum to dark energy and modified gravity, Mon. Not. R. Astron. Soc. **488,** 2121 (2019).

[41] K. K. Rogers, H. V. Peiris, A. Pontzen, S. Bird, L. Verde, and A. Font-Ribera, Bayesian emulator optimisation for cosmology: application to the Lyman-alpha forest, J. Cosmol. Astropart. Phys. 02 (2019) 031.

[42] S. Bird, K. K. Rogers, H. V. Peiris, L. Verde, A. Font-Ribera, and A. Pontzen, An emulator for the Lyman-α forest, J. Cosmol. Astropart. Phys. 02 (2019) 050.

[43] C. E. Rasmussen, Gaussian processes in machine learning, in *Summer School on Machine Learning* (Springer, New York, 2003), pp. 63–71.

[44] K. Heitmann, D. Higdon, M. White, S. Habib, B. J. Williams, E. Lawrence, and C. Wagner, The Coyote Universe. II. Cosmological models and precision emulation of the nonlinear matter power spectrum, Astrophys. J. **705,** 156 (2009).

[45] J. Kwan, K. Heitmann, S. Habib, N. Padmanabhan, H. Finkel, E. Lawrence, N. Frontiere, and A. Pope, Cosmic emulation: Fast predictions for the Galaxy power spectrum, Astrophys. J. **810,** 35 (2015).

[46] Z. Zhai, J. L. Tinker, M. R. Becker, J. DeRose, Y.-Y. Mao, T. McClintock, S. McLaughlin, E. Rozo, and R. H. Wechsler, The Aemulus Project III: Emulation of the Galaxy correlation function, Astrophys. J. **874,** 95 (2019).

[47] J. Liu, A. Petri, Z. Haiman, L. Hui, J. M. Kratochvil, and M. May, Cosmology constraints from the weak lensing peak counts and the power spectrum in CFHTLenS data, Phys. Rev. D **91,** 063507 (2015).

[48] A. Petri, J. Liu, Z. Haiman, M. May, L. Hui, and J. M. Kratochvil, Emulating the CFHTLenS weak lensing data: Cosmological constraints from moments and Minkowski functionals, Phys. Rev. D **91,** 103511 (2015).

[49] W. D. Jennings, C. A. Watkinson, F. B. Abdalla, and J. D. McEwen, Evaluating machine learning techniques for predicting power spectra from reionization simulations, Mon. Not. R. Astron. Soc. **483,** 2907 (2019).

[50] T. McClintock, E. Rozo, M. R. Becker, J. DeRose, Y.-Y. Mao, S. McLaughlin, J. L. Tinker, R. H. Wechsler, and Z. Zhai, The Aemulus Project II: Emulating the halo mass function, Astrophys. J. **872,** 53 (2019).

[51] M. Walther, J. F. Hennawi, H. Hiss, J. Oñorbe, K.-G. Lee, A. Rorai, and J. O'Meara, A new precision measurement of the small-scale line-of-sight power spectrum of the Lyα forest, Astrophys. J. **852,** 22 (2018).

[52] B. Giblin, M. Cataneo, B. Moews, and C. Heymans, On the road to per cent accuracy—II. Calibration of the non-linear matter power spectrum for arbitrary cosmologies, Mon. Not. R. Astron. Soc. **490,** 4826 (2019).

[53] H. J. Kushner, A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise, J. Basic Eng. **86,** 97 (1964).

[54] U. Seljak, A. Makarov, P. McDonald, S. F. Anderson, N. A. Bahcall, J. Brinkmann, S. Burles, R. Cen, M. Doi, J. E. Gunn et al., Cosmological parameter analysis including SDSS Lyα forest and galaxy bias: Constraints on the primordial spectrum of fluctuations, neutrino mass, and dark energy, Phys. Rev. D **71,** 103515 (2005).

[55] N. Palanque-Delabrouille, C. Yèche, J. Lesgourgues, G. Rossi, A. Borde, M. Viel, E. Aubourg, D. Kirkby, J.-M. LeGoff, J. Rich et al., Constraint on neutrino masses from SDSS-III/BOSS Lyα forest and other cosmological probes, J. Cosmol. Astropart. Phys. 02 (2015) 045.

[56] N. Palanque-Delabrouille, C. Yèche, J. Baur, C. Magneville, G. Rossi, J. Lesgourgues, A. Borde, E. Burtin, J.-M. LeGoff, J. Rich et al., Neutrino masses and cosmology with Lyman-alpha forest power spectrum, J. Cosmol. Astropart. Phys. 11 (2015) 011.

[57] C. Yèche, N. Palanque-Delabrouille, J. Baur, and H. du Mas des Bourboux, Constraints on neutrino masses from Lyman-alpha forest power spectrum with BOSS and XQ-100, J. Cosmol. Astropart. Phys. 06 (2017) 047.

[58] S. Chabanier, N. Palanque-Delabrouille, C. Yèche, J.-M. Le Goff, E. Armengaud, J. Bautista, M. Blomqvist, N. Busca, K. Dawson, T. Etourneau et al., The one-dimensional power spectrum from the SDSS DR14 Lyα forests, J. Cosmol. Astropart. Phys. 07 (2019) 017.

[59] P. Bode, J. P. Ostriker, and N. Turok, Halo formation in warm dark matter models, Astrophys. J. **556,** 93 (2001).

[60] https://github.com/MP-Gadget/MP-Gadget

[61] Y. Feng, S. Bird, L. Anderson, A. Font-Ribera, and C. Pedersen, Mp-gadget/mp-gadget: A tag for getting a doi (2018).

[62] V. Springel, The cosmological simulation code GADGET-2, Mon. Not. R. Astron. Soc. **364,** 1105 (2005).

[63] M. Viel, M. G. Haehnelt, and V. Springel, Inferring the dark matter power spectrum from the Lyman-alpha forest in high-resolution QSO absorption spectra, Mon. Not. R. Astron. Soc. **354,** 684 (2004).

[64] S. Bird, FSFE: Fake spectra flux extractor (2017).

[65] E. Boera, G. D. Becker, J. S. Bolton, and F. Nasir, Revealing reionization with the thermal history of the intergalactic medium: new constraints from the Lyα flux power spectrum, Astrophys. J. **872,** 101 (2019).

[66] N. Palanque-Delabrouille, C. Yèche, A. Borde, J.-M. Le Goff, G. Rossi, M. Viel, É. Aubourg, S. Bailey, J. Bautista, M. Blomqvist et al., The one-dimensional Lyα forest power spectrum from BOSS, Astron. Astrophys. **559,** A85 (2013).

[67] J. Binney and S. Tremaine, *Galactic dynamics* (Princeton University Press, Princeton, NJ, 2011).

[68] F. Nasir, J. S. Bolton, and G. D. Becker, Inferring the IGM thermal history during reionization with the Lyman α forest power spectrum at redshift $z \simeq 5$, Mon. Not. R. Astron. Soc. **463,** 2335 (2016).

[69] M. McQuinn, A. Lidz, M. Zaldarriaga, L. Hernquist, P. F. Hopkins, S. Dutta, and C.-A. Faucher-Giguère, He II reionization and its effect on the intergalactic medium, Astrophys. J. **694,** 842 (2009).

[70] L. Hui and N. Y. Gnedin, Equation of state of the photoionized intergalactic medium, Mon. Not. R. Astron. Soc. **292,** 27 (1997).

[71] J. S. Bolton, M. Viel, T.-S. Kim, M. G. Haehnelt, and R. F. Carswell, Possible evidence for an inverted temperature-density relation in the intergalactic medium from the flux

distribution of the Ly-alpha forest, Mon. Not. R. Astron. Soc. **386,** 1131 (2008).

[72] J. Oñorbe, F. B. Davies, Z. Lukić, J. F. Hennawi, and D. Sorini, Inhomogeneous reionization models in cosmological hydrodynamical simulations, Mon. Not. R. Astron. Soc. **486,** 4075 (2019).

[73] X. Wu, M. McQuinn, R. Kannan, A. D'Aloisio, S. Bird, F. Marinacci, R. Davé, and L. Hernquist, Imprints of temperature fluctuations on the $z \sim 5$ Lyman-α forest: A view from radiation-hydrodynamic simulations of reionization, Mon. Not. R. Astron. Soc. **490,** 3177 (2019).

[74] F. Haardt and P. Madau, Radiative transfer in a clumpy Universe. IV. New synthesis models of the cosmic UV/X-ray background, Astrophys. J. **746,** 125 (2012).

[75] J. Chardin, M. G. Haehnelt, D. Aubert, and E. Puchwein, Calibrating cosmological radiative transfer simulations with Ly α forest data: Evidence for large spatial UV background fluctuations at $z \sim 5.6 - 5.8$ due to rare bright sources, Mon. Not. R. Astron. Soc. **453,** 2943 (2015).

[76] J. Oñorbe, J. F. Hennawi, and Z. Lukić, Self-consistent Modeling of reionization in cosmological hydrodynamical simulations, Astrophys. J. **837,** 106 (2017).

[77] F. Haardt and P. Madau, Radiative transfer in a clumpy Universe. II. The ultraviolet extragalactic background, Astrophys. J. **461,** 20 (1996).

[78] F. Haardt and P. Madau, Modelling the UV/X-ray cosmic background with CUBA, in *Clusters of Galaxies and the High Redshift Universe Observed in X-Rays*, edited by D. M. Neumann and J. T. V. Tran (e-published as XXXVI Rencotres de Moriond, Savoie, 2001), p. 64, https://ui .adsabs.harvard.edu/abs/2001cghr.confE..64H/abstract.

[79] C.-A. Faucher-Giguère, A. Lidz, M. Zaldarriaga, and L. Hernquist, A new calculation of the ionizing background spectrum and the effects of He II reionization, Astrophys. J. **703,** 1416 (2009).

[80] E. Puchwein, F. Haardt, M. G. Haehnelt, and P. Madau, Consistent modelling of the meta-galactic UV background and the thermal/ionization history of the intergalactic medium, Mon. Not. R. Astron. Soc. **485,** 47 (2019).

[81] T. Takhtaganov, Z. Lukic, J. Mueller, and D. Morozov, Cosmic inference: Constraining parameters with observations and highly limited number of simulations, Astrophys. J. **906,** 74 (2021).

[82] Z. Lukić, C. W. Stark, P. Nugent, M. White, A. A. Meiksin, and A. Almgren, The Lyman α forest in optically thin hydrodynamical simulations, Mon. Not. R. Astron. Soc. **446,** 3697 (2015).

[83] Planck Collaboration *et al.*, Planck 2018 results. I. Overview and the cosmological legacy of Planck, Astron. Astrophys. **641,** A1 (2020).

[84] K. K. Rogers and H. V. Peiris, companion Letter, Strong Bound on Canonical Ultralight Axion Dark Matter From the Lyman-Alpha Forest, Phys. Rev. Lett. **126,** 071302 (2021).

[85] Planck Collaboration, Planck 2018 results. VI. Cosmological parameters, Astron. Astrophys. **641,** A67 (2020).

[86] M. D. McKay, R. J. Beckman, and W. J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics **21,** 239 (1979).

[87] K. Levenberg, A method for the solution of certain nonlinear problems in least squares, Q. Appl. Math. **2,** 164 (1944).

[88] D. W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, J. Soc. Ind. Appl. Math. **11,** 431 (1963).

[89] R. Hložek, D. J. E. Marsh, D. Grin, R. Allison, J. Dunkley, and E. Calabrese, Future CMB tests of dark matter: Ultralight axions and massive neutrinos, Phys. Rev. D **95,** 123511 (2017).

[90] https://github.com/dgrin1/axionCAMB

[91] A. Lewis and S. Bridle, Cosmological parameters from CMB and other data: A Monte Carlo approach, Phys. Rev. D **66,** 103511 (2002).

[92] D. J. E. Marsh and P. G. Ferreira, Ultralight scalar fields and the growth of structure in the Universe, Phys. Rev. D **82,** 103528 (2010).

[93] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, emcee: The MCMC Hammer, Publ. Astron. Soc. Pac. **125,** 306 (2013).

[94] D. D. Cox and S. John, Sdo: A statistical method for global optimization, in *Multidisciplinary Design Optimization: State-of-the-Art* (Society for Industrial and Applied Mathematics, Philadelphia, 1997), pp. 315–329, https://www .worldcat.org/title/multidisciplinary-design-optimization-state-of-the-art/oclc/36673710.

[95] P. Auer, Using confidence bounds for exploitation-exploration trade-offs, J. Mach. Learn. Res. **3,** 397 (2002), https://www.jmlr.org/papers/volume3/auer02a/ auer02a.pdf.

[96] P. Auer, N. Cesa-Bianchi, and P. Fischer, Finite-time analysis of the multiarmed bandit problem, Mach. Learn. **47,** 235 (2002).

[97] V. Dani, T. P. Hayes, and S. M. Kakade, Stochastic linear optimization under bandit feedback, *21st Annual Conference on Learning Theory, Helsinki, Finland* (2008), https://www.k4all.org/event/21st-annual-conference-on-learning-theory-conference/.

[98] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, Gaussian process optimization in the bandit setting: No regret and experimental design, IEEE Trans. Inf. Theory **58,** 3250 (2012).

[99] E. Brochu, V. M. Cora, and N. de Freitas, A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning, arXiv:1012.2599.

[100] S. G. Nash, Newton-type minimization via the lanczos method, SIAM J. Numer. Anal. **21,** 770 (1984).

[101] M. Järvenpää, M. U. Gutmann, A. Pleska, A. Vehtari, and P. Marttinen, Efficient acquisition rules for model-based approximate Bayesian computation, Bayesian Anal. **14,** 595 (2019).

[102] M. Buhmann and N. Dyn, Spectral convergence of multi-quadric interpolation, Proc. Edinb. Math. Soc. **36,** 319 (1993).

[103] M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese, and A. Riotto, Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with

WMAP and the Lyman-α forest, Phys. Rev. D **71**, 063534 (2005).

[104] P. Madau, F. Haardt, and M. J. Rees, Radiative transfer in a Clumpy Universe. III. The nature of cosmological ionizing sources, Astrophys. J. **514**, 648 (1999).

[105] F. B. Davies and S. R. Furlanetto, Large fluctuations in the hydrogen-ionizing background and mean free path following the epoch of reionization, Mon. Not. R. Astron. Soc. **460**, 1328 (2016).

[106] L. Hui, J. P. Ostriker, S. Tremaine, and E. Witten, Ultralight scalars as cosmological dark matter, Phys. Rev. D **95**, 043541 (2017).

[107] R. Kannan, M. Vogelsberger, F. Marinacci, R. McKinnon, R. Pakmor, and V. Springel, AREPO-RT: Radiation hydrodynamics on a moving mesh, Mon. Not. R. Astron. Soc. **485**, 117 (2019).

[108] T. Suarez and A. Pontzen, Large-scale fluctuations in the cosmic ionizing background: The impact of beamed source emission, Mon. Not. Roy. Astron. Soc. **472**, 2643 (2017).

[109] M. R. Lovell, C. S. Frenk, V. R. Eke, A. Jenkins, L. Gao, and T. Theuns, The properties of warm dark matter haloes, Mon. Not. Roy. Astron. Soc. **439**, 300 (2014).