Exploring Generalized Cosmological Expansion Histories with Differentiable Likelihoods

Benicio Gutierrez[1]
Advisor: Francis-Yan Cyr-Racine[1]
[1]*Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87106, USA**

The standard Lambda Cold Dark Matter (ΛCDM) cosmological model successfully describes the observed cosmic microwave background (CMB) spectra, but faces persistent tensions, particularly in the inferred value of the Hubble constant. One proposed resolution introduces a generalized dark fluid (GDF) component in the early universe, characterized by an equation of state $w(a)$ dependent on the scale factor a. In this work, we use a version of the Cosmic Linear Anisotropy Solving System (CLASS) to include such a fluid, allowing for spline-based interpolation of $w(a)$ across the range $-6 \leq \log(a) \leq -2$, and enforce physical constraints to maintain perturbative stability. We train and utilise Cosmopower neural networks to rapidly predict CMB power spectra from high-dimensional parameter inputs. Three Cosmopower emulators were trained: one using six standard ΛCDM parameters, one including eleven additional GDF parameters, and one including four additional GDF parameters. With ΛCDM, the emulator can achieve very accurate results, allowing models to be analyzed for best-fit parameter values. The trained emulators with GDF are not accurate enough for this analysis, highlighting the challenge of incorporating the complexity of a general dark fluid into an emulator. While the trained emulators are not yet accurate enough for precision parameter estimation, they represent an important step toward efficiently incorporating the complexity of a general dark fluid into cosmological analyses, allowing for the search of new physics.

INTRODUCTION

ΛCDM is the current standard model of cosmology, which describes the universe's expansion history along with the formation and evolution of its large-scale structure. ΛCDM describes this as the result of the interplay of ordinary matter, cold dark matter (CDM), dark energy, and radiation.

The Cosmic Microwave Background (CMB) is the result of photons decoupling from matter when the universe was cool enough for electrons and protons to form neutral hydrogen atoms. The CMB presents evidence of very small fluctuations in the early universe, which evolved into the present large-scale structure. Anisotropies, or non-uniformities, in the CMB carry rich information about the early universe. These anisotropies are most commonly characterized through the CMB power spectra, which describe the statistical properties of temperature and polarization fluctuations across different angular scales. Specifically, the power spectra are the Fourier transforms of the two-point correlation functions of these fluctuations, translating spatial correlations into angular frequency space.

The temperature-temperature (TT) power spectrum measures the amplitude of temperature fluctuations, while the polarization-polarization (EE) power spectrum captures the curl-free (E-mode) polarization anisotropies. The temperature-polarization (TE) power spectrum represents the correlation between temperature and E-mode polarization patterns.

By assuming ΛCDM cosmology, Boltzmann codes, such as the Cosmic Linear Anisotropy Solving System (CLASS) /citeclass, can accurately compute cosmological observables such as the CMB, by evolving the perturbation equations for the various components of the universe based on a given set of cosmological parameters, or by solving the Boltzmann equation.

The Boltzmann equation in cosmology describes the evolution of the distribution function of particles in the universe, accounting for both the expansion of the universe and various interactions such as scattering and decay. It plays a crucial role in the study of the Cosmic Microwave Background (CMB), the formation of large-scale structure, and the behavior of different species of particles in the early universe. The equation governs how the number density of particles evolves with time, including the effect of expansion and interactions. In cosmology, the distribution function $f(\mathbf{x}, \mathbf{p}, t)$ typically depends on spatial coordinates $\mathbf{x}$, momentum $\mathbf{p}$, and time t, and is often written in terms of the conformal time τ, which relates to the scale factor of the universe. The collision term in the Boltzmann equation accounts for the scattering processes that change the distribution of particles. The Boltzmann equation is given by [2]:

$$\frac{Df}{d\tau} = \frac{\partial f}{\partial \tau} + \frac{dx^i}{d\tau}\frac{\partial f}{\partial x^i} + \frac{dq}{d\tau}\frac{\partial f}{\partial q} + \frac{dni}{d\tau}\frac{\partial f}{\partial n_i} = \left(\frac{\partial f}{\partial \tau}\right)_C \tag{1}$$

In the cosmological Boltzmann equation, several terms contribute to the evolution of the distribution function. The left-hand side of the equation involves the total derivative of the distribution function f with respect to the proper time τ. The term $\frac{\partial f}{\partial \tau}$ represents the time evolution of the distribution, while the spatial terms like $\frac{dx^i}{d\tau}\frac{\partial f}{\partial x^i}$ represent the effects of particle motion and the expansion of the universe. The term $\frac{dq}{d\tau}\frac{\partial f}{\partial q}$ accounts for the change in momentum due to interactions, and $\frac{dni}{d\tau}\frac{\partial f}{\partial n_i}$ reflects the effect of particle creation or annihilation processes. The right-hand side of the equation contains the collision term C, which models the interactions between particles, such as scattering or decay. These terms together describe how particles in the early universe evolve in terms of both their density and momentum distributions.

Figures 1, 2, and 3 show CLASS's predictions of the TT, EE, and TE spectra of the CMB based on ΛCDM cosmology, in comparison with data from the Planck observational data [4].

Assuming ΛCDM, the Planck satellite [4], which observed anisotropies of the CMB, estimated major cosmological parameters with Bayesian analyses. $\omega_b = \Omega_b h^2$ is the present energy density of baryons and $\omega_{cdm} = \Omega_{cdm} h^2$ is the present energy density of cold dark matter (CDM). Note Ω_b and Ω_{cdm} are expressed as ratios of their energy density to the critical density, $\rho_{crit} = \frac{3H_0^2}{8\pi G}$, the energy density required for spatial curvature to be zero. H_0 is the Hubble constant, the current rate of expansion of the universe. This is usually expressed in units of kilometers per second per megaparsec (km/s/Mpc), meaning that for every megaparsec (about 3.26 million light-years) of distance, the recessional velocity of galaxies increases by H_0 kilometers per second. $h = \frac{H_0}{100}$ is the dimensionless Hubble parameter. $\ln(10^{10} A_\mathrm{s})$ represents the scalar amplitude, where A_s is the amplitude of primordial density fluctuations. n_s, the spectral index of primordial perturbations, quantifies the scale dependence of primordial fluctuations. Finally, τ is optical depth to reionization, which represents how opaque the universe was during reionization, when the universe's hydrogen gas was ionized again by the first stars and galaxies.

However, the ΛCDM model has several issues. For example, the Hubble tension is the discrepancy in the current expansion rate of the universe, which can be acquired from the CMB or from observing the redshift of type 1a supernovae. The value of the Hubble constant is crucial because it plays a role in defining times and distances in the universe.

A method of resolving this is by introducing a general dark fluid in the early universe [5], with additional energy density contribution from $-5 \leq \log(a) \leq -2$. Here, $0 \leq a \leq 1$ is the scale factor, which is the ratio between the universe's size at an earlier time and the current size. $a = 1$ is the present day.

While introducing new components can help alleviate discrepancies, it also substantially increases the dimensionality and complexity of the parameter space, making traditional likelihood analyses for determining parameter values computationally intensive. To address this challenge, emulators based on neural networks have emerged as a powerful tool. Neural networks can efficiently learn the mapping between cosmological parameters and observables such as the CMB power spectra, enabling rapid and accurate predictions without the need to repeatedly solve the full Boltzmann equations. Furthermore, neural network-based emulators can be incorporated into differentiable likelihood frameworks, enabling scalable and flexible inference even in high-dimensional or non-standard cosmological models.

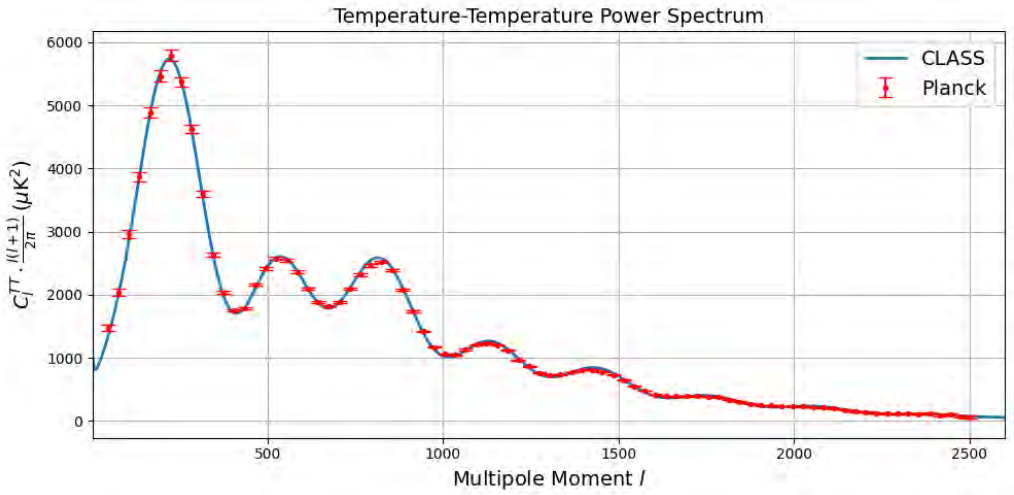


FIG. 1: (a) Temperature (TT) spectrum

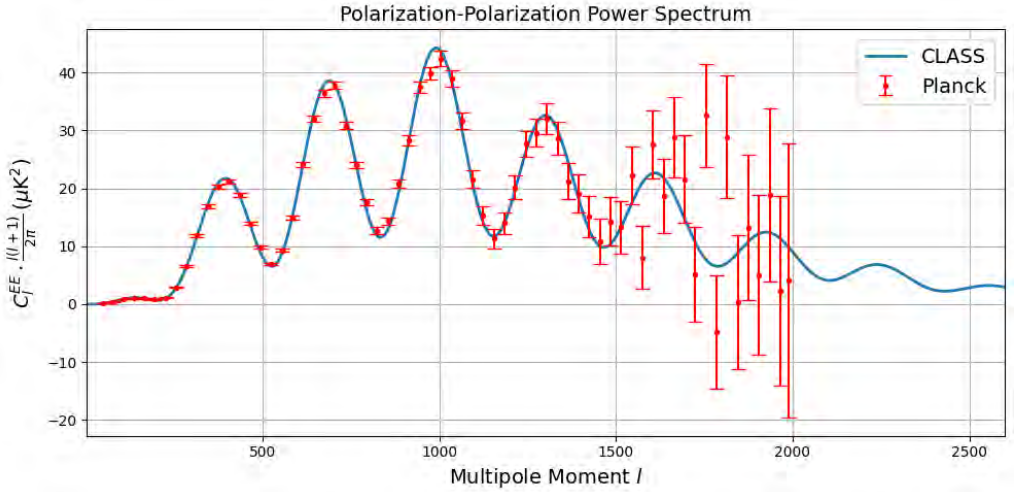


FIG. 2: (b) E-mode polarization (EE) spectrum

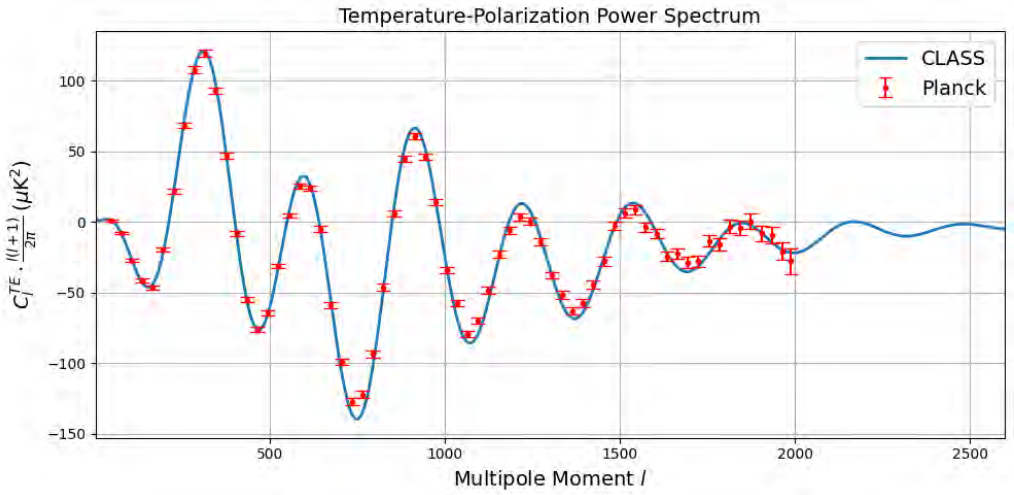


FIG. 3: (c) Temperature-polarization correlation (TE) spectrum

FIG. 4: CMB power spectra, ordered as follows: (a) temperature (TT) spectrum, (b) E-mode polarization (EE) spectrum, and (c) temperature-polarization correlation (TE) spectrum. The blue curve represents theoretical predictions with ΛCDM according to CLASS. The red points with error bars are binned Planck 2018 data [4].

GENERALIZED DARK FLUID (GDF)

The different components of the Universe (baryons, photons, dark matter, etc.) each evolve with the expansion of the Universe according to their own equation of state w, which describes the ratio of their pressure to their energy density. Table I shows the equation of state for the different energy components making up the Universe today according to ΛCDM. Ω is the energy density ratio between a particular component and the critical density ρ_{crit}.

TABLE I: Equation of state for different energy components in the ΛCDM model.

Energy	w	Ω
Baryonic Matter	0	$\Omega_b \approx 0.05$
Dark Matter	0	$\Omega_{cdm} \approx 0.27$
Radiation	$\frac{1}{3}$	$\Omega_g + \Omega_{ur} \approx 10^{-5}$
Dark Energy	-1	$\Omega_\Lambda \approx 0.69$

The evolution of these energy components is a consequence of energy conservation, which is described by the continuity equation:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0 \tag{2}$$

Substituting $P = w\rho$, the equation can be solved to observe the energy density of various parameters over time.

$$\rho(a) = \rho_0 \left(\frac{a_0}{a}\right)^3 \exp\left(-3\ln(10)\int_{a0}^{a} w(\log(a'))d(\log(a'))\right). \tag{3}$$

Figure 5a shows the energy densities as a function of the scale factor. Note that there are three separate time regimes where different components dominate. From $-14 \leq \log(a) < -4$ (corresponding to), radiation is dominating. From $-4 < \log(a) < -0.14$, matter dominates. Finally, from $-0.14 < \log(a) \leq 0$ (present day), dark energy dominates.

A potential modification to ΛCDM is the introduction of a new form of energy, which we will refer to as a general dark fluid (GDF). The major properties of GDF are its equation of state w and its present energy density Ω_{GDF}; to generalize this fluid, w is allowed to change over time. This can be used to represent major cosmological events, such as phase transitions, that this fluid may have experienced in the past.

GDF is designed to be highly flexible, capable of interpolating between behaviors of various known cosmic components such as cold dark matter, radiation, and dark energy. In doing this, GDF can represent a wide range of. Its treatment in this work is strictly linear and perturbative. The model includes free parameters for the equation of state $w(a)$, which influences the fluid's behavior at the level of cosmological perturbations.

It is important to note that standard fluid dynamics concepts such as the Reynolds number and turbulence are not meaningful in the context of the GDF as modeled here. The Reynolds number quantifies the relative importance of inertial to viscous forces in a fluid and is used to predict the onset of turbulence. However, in cosmology, and especially in linear perturbation theory, the relevant scales are so large and the amplitudes of perturbations so small that the dynamics remain linear and smooth. The GDF is evolved within this linear regime, where perturbations do not grow large enough to become nonlinear or turbulent. Thus, one cannot define a Reynolds number for the GDF in any conventional way, nor is turbulence expected or permitted within the scope of the model. These simplifications are justified both by the nature of the cosmological observables under consideration and by the goal of keeping the parameter space tractable in high-dimensional exploration.

A modified version of CLASS includes this GDF, which can produce a cubic spline for $w(a)$ with several controlled points [3].

In cosmology, small deviations from a perfectly uniform universe are studied using perturbation theory. One of the most fundamental quantities in this framework is the *density contrast*, denoted by δ, which quantifies fractional deviations in energy density relative to the homogeneous background. It is defined as

$$\delta = \frac{\delta\rho}{\bar{\rho}}, \tag{4}$$

where $\delta\rho$ is the perturbation in the energy density, and $\bar{\rho}$ is the background energy density. A positive value of δ indicates an overdensity (i.e., a region with more matter than average), while a negative value represents an

underdensity. This variable is central to understanding how structures such as galaxies and galaxy clusters form through gravitational instability.

Alongside density perturbations, the motion of matter in response to gravitational forces is described by the divergence of the velocity field, θ. This quantity captures how the velocity of the fluid is converging or diverging at a given point in space. In Fourier space, it is related to the perturbed energy-momentum tensor as

$$\bar{\rho}(1+w)\theta = ik^i \delta T^0_i, \tag{5}$$

where $w = \bar{p}/\bar{\rho}$ is the background equation-of-state parameter, and δT^0_i represents the perturbation to the momentum density. The quantity k^i is the Fourier wavenumber vector. The divergence θ thus acts as a source for density perturbations and plays a crucial role in determining how matter flows under the influence of gravity in the early universe.

In addition to isotropic pressure, more general fluids like GDF can support anisotropic stress, denoted by σ. This term accounts for directional differences in internal stresses, similar to shear stress in classical fluid dynamics. The scalar anisotropic stress in Fourier space is defined as

$$\sigma = -\frac{1}{\bar{\rho}(1+w)} \left(\hat{k}_i \hat{k}_j - \frac{1}{3}\delta_{ij} \right) \Sigma_{ij}, \tag{6}$$

where Σ_{ij} is the traceless spatial part of the stress-energy tensor perturbation, and $\hat{k}_i$ are the normalized components of the wavevector. The tensor $\hat{k}_i \hat{k}_j - \frac{1}{3}\delta_{ij}$ acts as a projector that isolates the shear-like (anisotropic) part of the perturbation. This quantity becomes especially relevant in models where dark matter or dark energy behaves more like a relativistic fluid or a free-streaming particle species.

Since the conservation equations of general relativity do not naturally specify how anisotropic stress evolves, a *closure relation* is required. One such relation, inspired by models of neutrino free-streaming, describes the time evolution of σ as

$$\dot{\sigma}_g = -3H\sigma + \frac{8}{3(1+w)} k^2 c_v^2 \Theta_g, \tag{7}$$

where $\dot{\sigma}_g$ is the time derivative of the gauge-invariant anisotropic stress, H is the Hubble expansion rate, and c_v^2 is a model parameter that acts similarly to a shear viscosity coefficient. The term Θ_g is a gauge-invariant combination of velocity perturbations and metric fluctuations that acts as a source for anisotropic stress. The first term on the right-hand side represents the decay of anisotropic stress due to cosmic expansion, while the second term provides a source due to the bulk motion of the fluid and gravitational perturbations.

The source term Θ_g in the evolution equation above is defined by the expression

$$\Theta = \frac{\hat{\theta}}{k^2} - \zeta - \frac{1}{2}\nu, \tag{8}$$

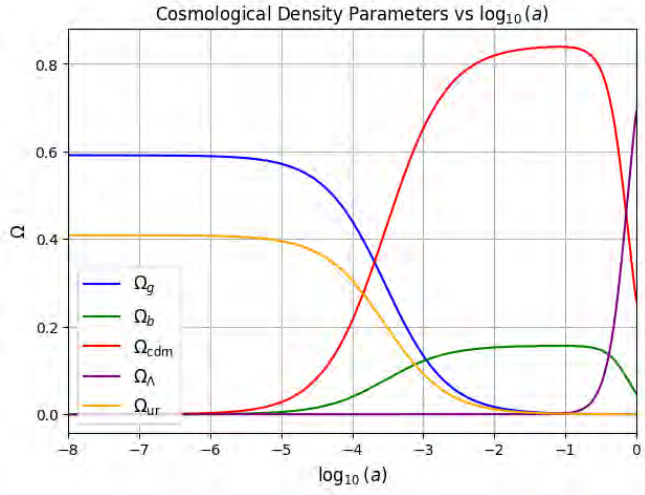
where $\hat{\theta}$ is the gauge-invariant version of the velocity divergence, and ζ and ν are scalar perturbations to the metric. Specifically, ζ arises from the time-space component of the perturbed metric, while ν comes from the traceless part of the spatial metric perturbations. In certain commonly used gauges, such as the conformal Newtonian gauge, both ζ and ν vanish, and Θ reduces to $\hat{\theta}/k^2$. This simplification allows for a direct interpretation of Θ as a rescaled velocity divergence.

The GDF models used in this study have parameters representing the current-day energy density of the GDF Ω_{GDF}, along with nodes controlling $w(a)$. A cubic spline is constructed to produce a function of $w(log(a))$ based on these controlled points, along with constraints. The constructed splines for w have several requirements. Firstly, $w = 1/3$ from $-14 \leq \log(a) \leq -8$, so that the GDF behaves like relativistic particles during radiation domination. To prevent splines from entering the regime $|w| > 1$, the maximum value of w at a particular node is set to 0.8. Examples of splines produced are shown in figure 5b.

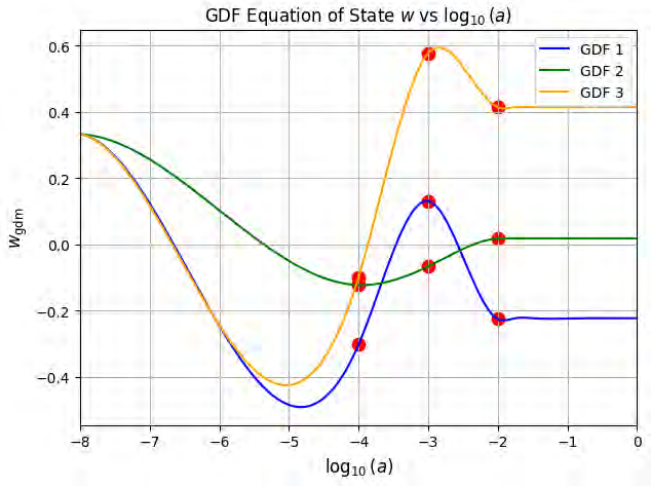
Perturbations in the GDF propagate at the adiabatic speed of sound [3]:

$$c_a^2 = w - \frac{a\dot{w}}{3\dot{a}(1+w)} \tag{9}$$

$\dot{a}$ and $\dot{w}$ are derivatives with relation to conformal time. A major concern is that for negative w values or if $\dot{w}$ is too large, the adiabatic speed of sound becomes negative, causing perturbations to grow exponentially, which is unphysical. To prevent this, the minimum bound for nodes of w is $w \geq -0.4$. Additionally, if $c_a^2 < 0$, c_a^2 is set to zero.



(a) Plot of the energy densities of various components versus the log scale factor a.



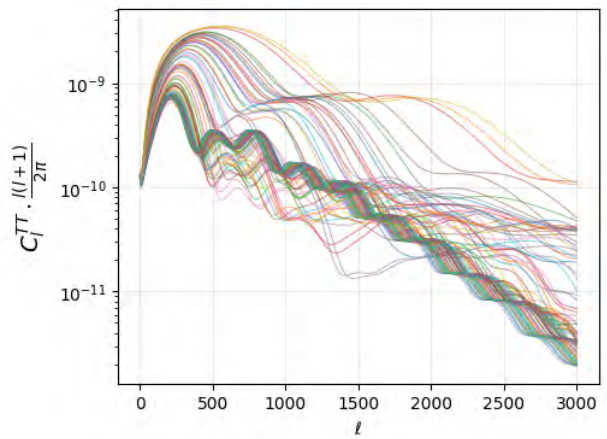
(b) Examples of cubic splines for $w(a)$ for the GDF, constructed from given points (red).

FIG. 5

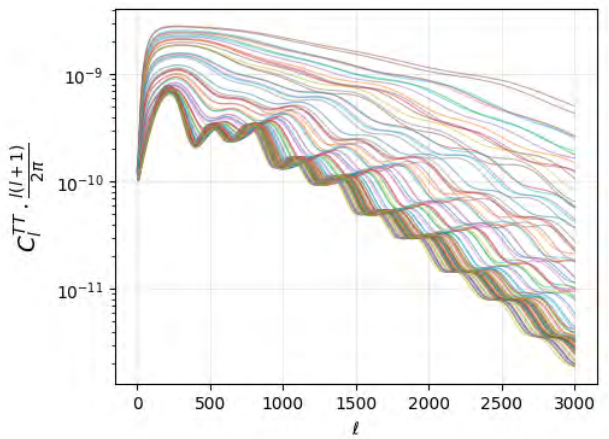
In this implementation of GDF, nonadiabatic pressure is ignored, meaning the adiabatic speed of sound c_a^2 is equal to the speed of sound in the rest frame c_s^2. This is done to ignore the speed of sound in the rest frame as an additional parameter. Additionally, the viscosity parameter is set to zero, to ignore this as a parameter, which makes the fluid ideal.

A GDF model is controlled with 4 parameters: the current-day energy density of the GDF Ω_{GDF}, and $w(a)$ nodes at $\log(a) = -4, -3, -2$.

The node placements were determined by producing 100 TT spectra with ΛCDM parameters set to the best fit averages according to the Planck survey [4], with Latin hypercube sampling of Ω_{GDF} from the range $(0, 0.1)$ and a single node for $w(a)$, selected at $\log(a) = -6, -5, -4, -3,$ or -2 ranging from $(-0.4, 0.8)$. The constant nodes are set to 0, $w = 1/3$ in the region of $-14 \leq \log(a) \leq -8$, and $w = 0$ in the region of $-1.75 \leq \log(a) \leq 0$. Figures **??** to 6b show these 100 generated spectra from the varied nodes at $\log(a) = -3$ and -2 as examples for how the GDF affects the TT spectrum. The plots with the node placements for all five points are provided in the appendix. The three node placements corresponding to the greatest variation in the spectra are selected, which are $\log(a) = -4, -3, -2$.



(a) 100 TT spectra with varying node at $\log(a) = -3$



(b) 100 TT spectra with varying node at $\log(a) = -2$

Table III contains the GDF parameters that are introduced, along with their ranges.

COSMOPOWER

The addition of GDF introduces several parameters to the model, and hence much more complexity. To perform accurate and efficient analyses, a neural network or emulator (from Cosmopower) is trained based on data generated from the new model. Then, differentiable likelihood analyses can be performed on the neural network with a tool such as candl [7]. For more details on Cosmopower, see the original paper [6].

Cosmopower helps to create neural networks that map cosmological parameters to the CMB spectra. A neural network consists of an input layer of nodes, several hidden layers, and an output layer. Each node receives input from the previous layer, applies a weighted sum to those inputs (where each input is multiplied by a weight and added to a bias), and passes the result through a nonlinear activation function. Figure 7 presents the structure of a neural network, from [6]. This nonlinearity allows the network to approximate highly complex functions that linear models cannot. The weights and biases are adjusted during training to minimize the difference between the network's predictions and the true data, by minimizing the loss (root mean squared error) function:

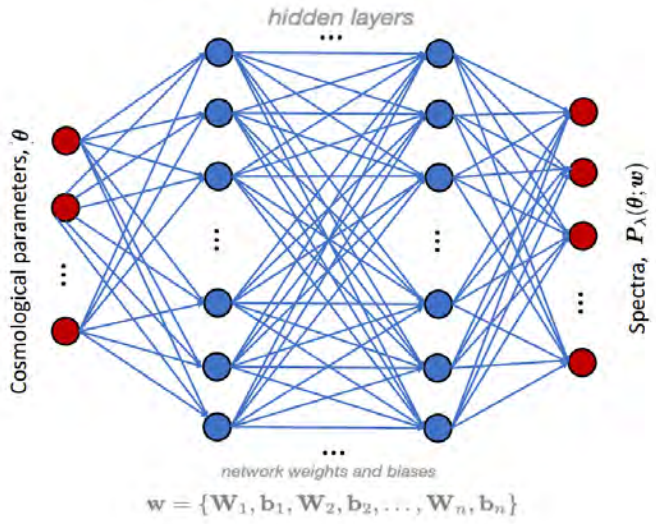


FIG. 7: Representation of a Cosmopower neural network mapping the cosmological parameters directly to spectra. On the left is the input layer, for the cosmological parameters. On the right are the output nodes for the cosmological spectra. In between the input and output layers are hidden layers, where each of the nodes has different weights and biases. Figure from [6].

$$\mathcal{L} = \sqrt{\frac{1}{\mathrm{N_{points}}} \sum_{i=1}^{\mathrm{N_{points}}} (\hat{y}_i - y_i)^2} \tag{10}$$

N is the number of points of the spectra, $\hat{y}_i$ is the predicted value of the ith point of the spectrum, and y_i is the actual value of the spectrum for that point.

Cosmopower has two methods for creating neural networks. The first is a direct mapping from cosmological parameters to CMB spectra. The second, principal component analysis (PCA) compresses the power spectra into a smaller number of components before training. Instead of having the neural network learn the entire spectrum directly, it learns to predict the PCA coefficients, which describe how much of each principal component is needed to reconstruct the spectrum.

The neural networks trained in this report consist of four hidden layers with 512 nodes. The TT and EE spectra neural networks use the direct mapping from cosmological parameters to log-spectra to reduce the emulators' output ranges. The TE spectra cannot be expressed with log-spectra, because the power spectra include negative values; as an alternative, to capture the structure of the TE spectra, the emulators utilize PCA compression.

The ranges for the ΛCDM parameters generated are in table II. Table III contains the constraints on the GDF, with the bold lines representing controllable parameters for the neural network.

TABLE II: Ranges for ΛCDM Parameters

Parameter	Avg.	σ	Range
ω_b	0.02236	0.00015	$(-25\sigma, +25\sigma)$, (0.01861, 0.02611)
ω_{cdm}	0.1202	0.0014	$(-25\sigma, +25\sigma)$, (0.0852, 0.1552)
h	0.6727	0.0060	$(-25\sigma, +25\sigma)$, (0.5227, 0.8227)
n_s	0.9649	0.0044	$(-25\sigma, +25\sigma)$, (0.8549, 1.0749)
$\ln(10^{10}A_s)$	3.045	0.016	$(-25\sigma, +25\sigma)$, (2.645, 3.445)
τ	0.0544	(+0.0070, -0.0081)	$(-6\sigma, +10\sigma)$, (0.0058, 0.1244)

TABLE III: GDF constraints used in the 10-parameter model, with 4 GDF parameters. The bold lines indicate parameters for the emulators.

Parameter	Range
$\boldsymbol{\Omega_{GDF}}$	$\mathbf{(0.0, 0.10)}$, lin. spaced
$w(-14.0 \leq \log(a) \leq -8.0)$	0.3333
$\boldsymbol{w(\log(a) = -4.0)}$	$\mathbf{(-0.4, 0.8)}$
$\boldsymbol{w(\log(a) = -3.0)}$	$\mathbf{(-0.4, 0.8)}$
$\boldsymbol{w(\log(a) = -2.0)}$	$\mathbf{(-0.4, 0.8)}$
$w(-2.0 \leq \log(a) \leq 0)$	$w(\log(a) = -2.0)$

The training process of the neural network begins with the model passing through a series of epochs, where each epoch represents a full pass through the entire training dataset. During each epoch, the data is divided into smaller batches, with each batch containing 1024 spectra. These batches are sequentially processed by the model, which performs a forward pass to make predictions, computes the loss (root mean squared error), and updates the weights by calculating the gradients. The batch size of 1024 ensures that the model can update its weights in manageable steps, preventing it from being overwhelmed by the entire dataset. The rate that the nodes' weights are updated depends on the learning rate hyperparameter. Initially, a relatively higher learning rate of 10^{-2} is used to allow the model to make significant progress in adjusting its weights. As training progresses, the learning rate gradually decreases in successive stages, with rates of 10^{-3}, 10^{-4}, 10^{-5}, and finally 10^{-6}. This progressive reduction, known as cooling, ensures that the model makes rapid adjustments in the early stages and then fine-tunes the weights more carefully as it approaches an optimal solution. After each epoch, the model's performance is evaluated on the validation set, which comprises 10% of the data. If the validation loss does not improve for 100 consecutive epochs (known as the patience value), the model will stop training early. This strategy prevents overfitting by halting training once the model's performance stagnates, even if it hasn't reached the maximum number of epochs. Through this combination of epochs, batches, learning rate adjustments, and early stopping, the model gradually improves its performance while avoiding overfitting and ensuring efficient training.

The training parameters were generated using Latin Hypercube Sampling, which divides each input parameter's range into equal intervals and randomly selecting one value from each interval without repeating, ensuring that every part of the range is covered. Then, it combines these values across all parameters in a way that each combination is unique and well-distributed across the full parameter space.

I generated the training data and trained the neural networks on the University of New Mexico's Center for Advanced Research Computing (CARC) Wheeler and Xena clusters. Generating 100,000 spectra requires ~ 40 hours. The training data of 100,000 spectra is about 2 Gb. Because of RAM limitations, this meant limitations of using $\sim 300,000$ spectra per training session. Training an emulator with the suggested hyperparameters typically requires less than one hour with GPUs.

DIFFERENTIABLE LIKELIHOOD ANALYSIS: CANDL

Differentiable likelihood analysis, as implemented in the candl [7] framework, provides substantial benefits over traditional methods in CMB data analysis. The primary advantage lies in the ability to automatically and robustly compute derivatives of the likelihood function with respect to both cosmological and nuisance parameters. This is made possible through JAX, a high-performance Python library that supports automatic differentiation. Unlike finite-difference methods, which can be numerically unstable and require careful tuning of step sizes, JAX allows for

accurate and efficient derivative evaluation. This makes calculations like Fisher matrices, sensitivity forecasts, and error propagation much faster and more reliable. In practice, this means one can easily explore how changes in data processing (like bin width choices) affect cosmological parameter uncertainties or understand correlations between different data subsets, leading to better-informed survey design and analysis strategies.

A specific and powerful use of differentiable likelihoods is in gradient-based optimization and sampling. One such optimization approach is the Newton-Raphson method, a classic technique for finding the roots (or in this case, maxima) of functions. In the context of likelihood analysis, it is used to locate the best-fit point in parameter space. Newton-Raphson requires both the gradient and the Hessian (the matrix of second derivatives) of the likelihood function. With candl, these quantities are readily available via automatic differentiation, allowing the Newton-Raphson algorithm to efficiently and directly converge to the likelihood maximum. Moreover, the computed Hessian can be repurposed—for example, as a proposal distribution in Markov chain Monte Carlo (MCMC) sampling or as an approximation to the posterior covariance, offering further computational advantages.

With the efficiency and accuracy of emulation, it is possible to perform differentiable likelihood analysis using the emulators. Likelihood analysis is often used to compare theoretical models to observational data, involving computing the likelihood of a set of parameters given the observed data. To compute the likelihood, one must calculate the Cosmic Microwave Background (CMB), along with the derivatives of the likelihood with respect to the cosmological parameters. This derivative is essential for tasks like Markov Chain Monte Carlo (MCMC) sampling or gradient-based optimization methods, which require the gradients of the likelihood function with respect to the model parameters. Computing these derivatives directly can be computationally expensive, as it would involve recalculating the power spectra for small changes in the cosmological parameters and performing a finite difference calculation. However, by training a neural network to predict the power spectra given a set of cosmological parameters, the emulator can efficiently generate not only the power spectra themselves but also their derivatives with respect to the cosmological parameters. Neural networks have the advantage of being differentiable, meaning that the gradients (or derivatives) of the model's output with respect to the input parameters can be computed automatically and efficiently using backpropagation. This makes the emulator a powerful tool for performing differentiable likelihood analysis, as it significantly speeds up the process of calculating the required derivatives and enables fast, accurate parameter estimation.

ΛCDM EMULATION AND DIFFERENTIABLE LIKELIHOOD ANALYSIS

To test the overall pipeline and probe the simplest case, we started with training TT, EE, and TE emulators with $\sim$ 40,000 spectra with 6 ΛCDM parameters: ω_b, ω_{cdm}, h, $\ln(10^{10}A_\mathrm{s})$, n_s, and τ. We selected a multipole moment $2 \leq \ell \leq 3000$. These parameters were selected using Latin hypercube sampling. The selected parameter ranges are in table IV.

TABLE IV: Ranges for ΛCDM Training Parameters, for ΛCDM Emulator

Parameter	Avg.	σ	Range
ω_b	0.02236	0.00015	(0.01865, 0.02625)
ω_{cdm}	0.1202	0.0014	(0.05, 0.255)
h	0.6727	0.0060	(0.64, 0.82)
n_s	0.9649	0.0044	(0.84, 1.1)
$\ln(10^{10}A_s)$	3.045	0.016	(1.61, 3.91)
τ	0.0544	(+0.0070, -0.0081)	(0.01, 0.8)

We tested the emulators' accuracy with $\sim$ 8,000 samples that were generated similarly to the training data. For TT, EE, and TE spectra, figures 8, 9, and 10 show the emulators' error over the multipole moment. For TT and EE spectra, the relative errors are calculated. Because TE spectra have regions where the spectra is 0, the absolute error is calculated instead. Specifically, the figures show the 68th, 95th, and 99th percentiles for the test's relative error. For the TT spectra, with the exception of the ℓ near 0 and 2400, the majority of tests have a relative error less that 0.1%. For the EE spectra, with the exception of the $\ell < 300$, the majority of tests have a relative error less that 0.2%. Lastly, for the TE spectra, the majority of tests resulted in an absolute error less than 10^{-14}. Considering that the original TE spectra is of the order of magnitude of 10^{-11}, the absolute error is $\sim$ 1000 times smaller than the predicted spectra.

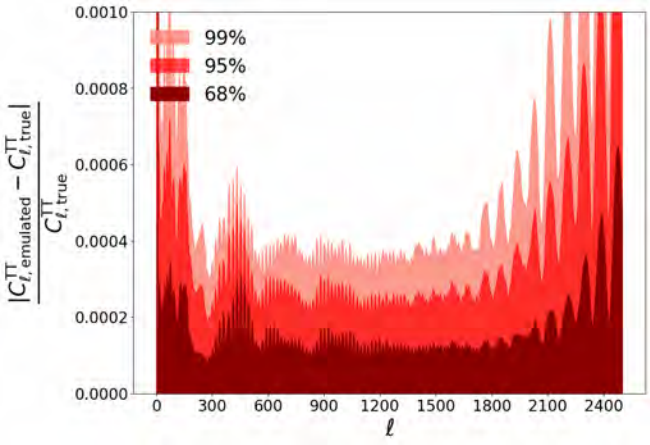


FIG. 8: ΛCDM TT emulator relative error based on ∼8,000 tests.

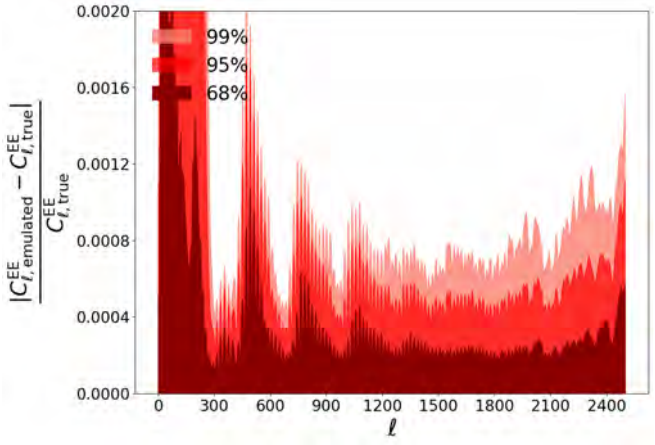


FIG. 9: ΛCDM EE emulator relative error based on ∼8,000 tests.

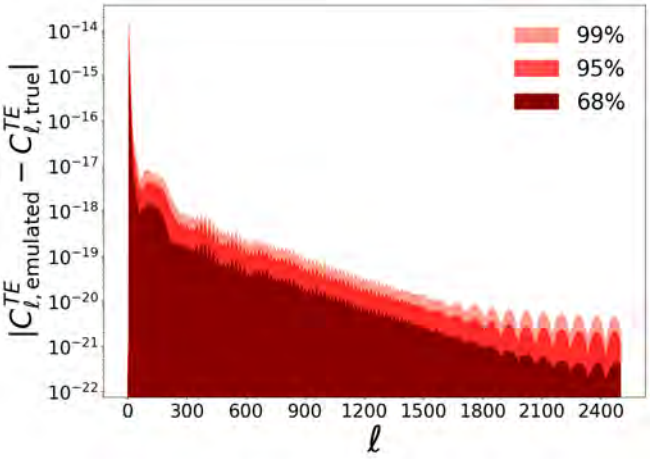


FIG. 10: ΛCDM TE emulator absolute error based on ∼8,000 tests.

FIG. 11: Accuracy of ΛCDM emulators based on ∼8,000 tests.

Since all three emulators display low error, with a differentiable likelihood tool, such as candl [7], we efficiently determined parameter values that best fit CMB data. With the candl built-in Newton-Raphson method, the likelihood corresponding to the CMB South Pole Telescope SPT-3G 2018 TT/TE/EE dataset [9] is maximized. In this case, both the likelihood corresponding to the multi-frequency data and the lite [8] data are maximized. A multi-frequency likelihood is the full, uncompressed likelihood that includes the raw cross-spectra between these frequency channels (e.g., 90×90 GHz, 90×150 GHz, 150×220 GHz, etc.), and it includes both the CMB signal and all the nuisance effects from foregrounds, beams, calibration, and noise. The lite-likelihood is a compressed version of the full multi-frequency CMB likelihood, where the complex, multi-channel data (which includes various foreground contaminations and frequency-specific effects) is reduced to a simpler, foreground-marginalized data vector. By compressing the complex, multi-frequency CMB data into foreground-marginalized CMB-only band powers, the lite likelihood significantly reduces computational costs while retaining the essential cosmological information. This makes it much faster to evaluate and easier to interpret, which is especially useful for exploratory analyses, parameter estimation, or when computational resources are limited. Despite this simplification, when constructed carefully—such as with automatic differentiation methods—the lite likelihood maintains a high level of accuracy, closely matching the results of the full multi-frequency analysis.

Table V shows the best-fit values that maximize both the multi-frequency likelihood and the lite-likelihood. The parameter estimations with both datasets are similar, except for $\ln(10^{10}A_s)$, indicated by small values for Δ/σ, the

normalized shift in the best-fit values. Additionally, the ratio between both standard deviations $\sigma_{\rm lite}/\sigma_{\rm mf} \approx 1$ for all parameters, indicating a similarity in the standard deviation for both datasets. The multi-frequency and lite likelihood results show good agreement with the published SPT-3G constraints. The H_0, ω_b, $\omega_{\rm cdm}$, and $n_{\rm s}$ values are consistent within uncertainties, with shifts well below 1σ. However, the estimation for $\ln(10^{10}A_s)$ with both methods seems to be off from the SPT-3G [8, 9] estimation by $\sim 6\sigma$.

Figure 12 shows the triangle plot of the posterior distributions for the cosmological parameters obtained using the multi-frequency and lite-likelihood from candl. All of the contours between the multi-frequency and lite analyses overlap very well.

Figure 13 compares the SPT-3G lite data and the curve fit with the best-fit parameters.

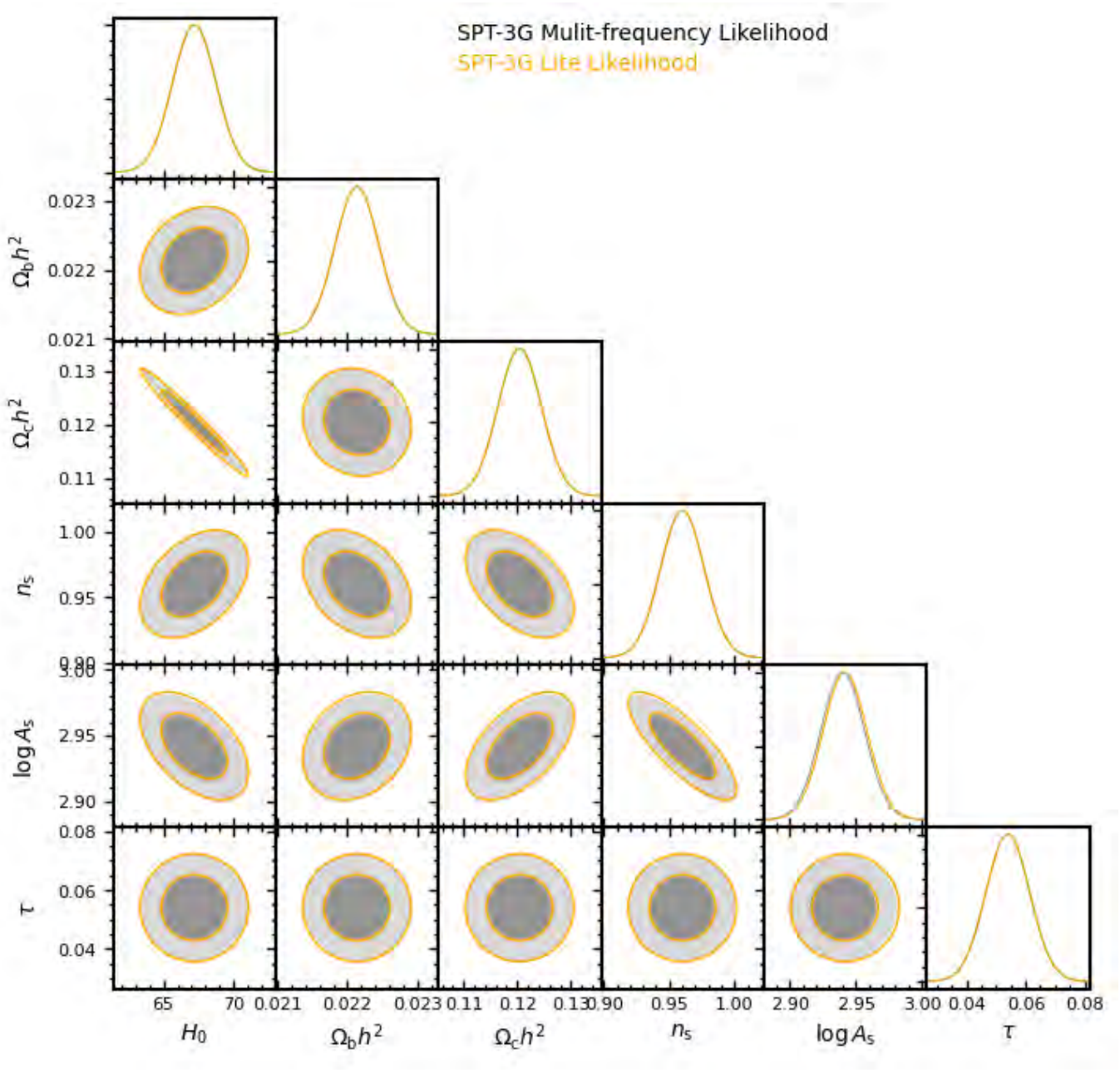


FIG. 12: Plot of posterior distributions of ΛCDM parameters based on candl SPT-3G data [9]. The gray contours are from the multifrequency ($\nu\mu$) likelihood. The yellow contours are from the lite-likelihood. For each set of contours, the inner contour corresponds to the 68% confidence region, and the outer contour to the 95% confidence region.

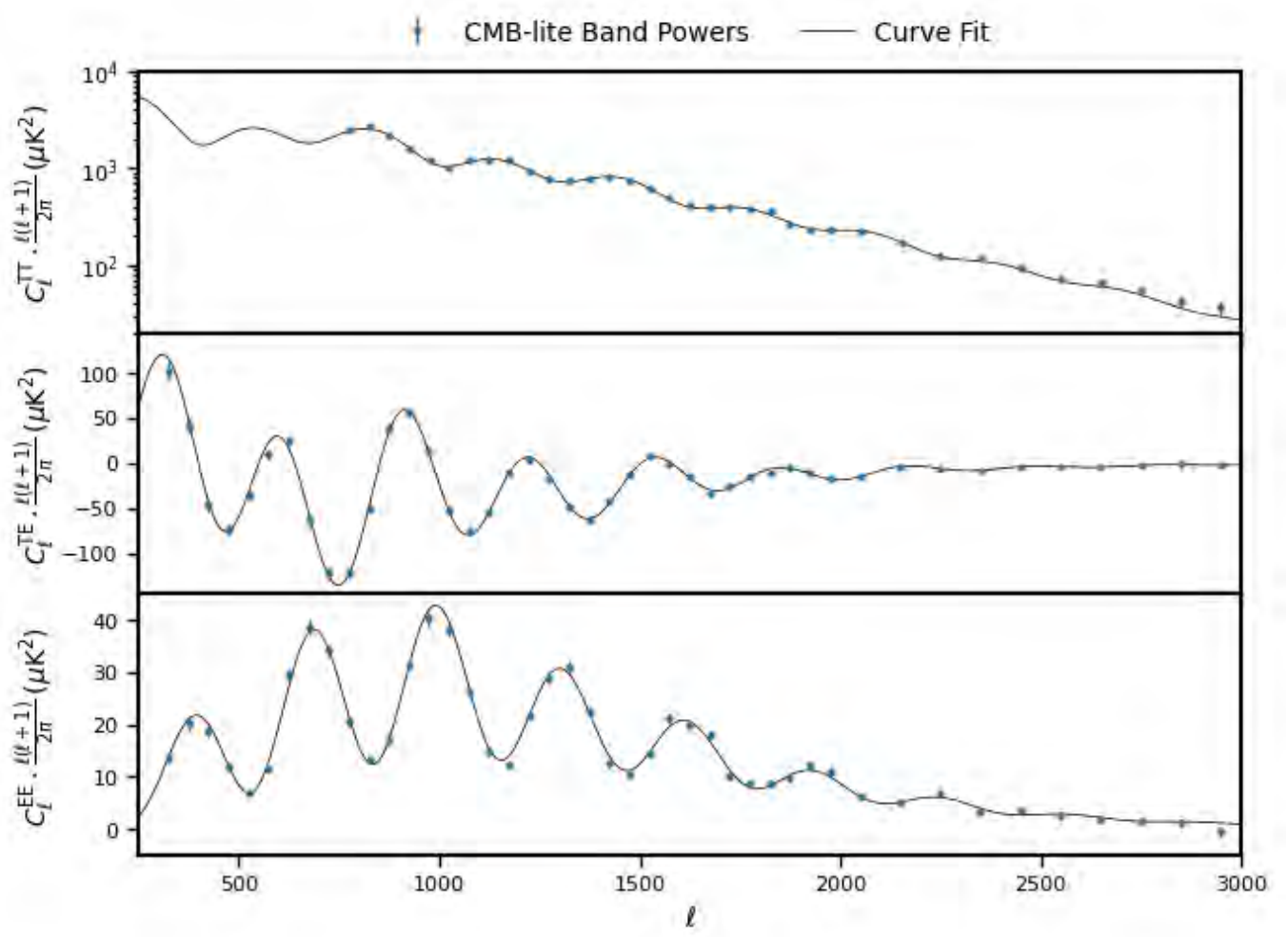


FIG. 13: Plot of best fit curve for binned SPT-3G Lite data.

Parameter	**Multi-frequency**	**Lite**	Δ/σ	$\sigma_{\rm lite}/\sigma_{\rm mf}$
H_0	67.2 ± 1.6	67.2 ± 1.6	0.0089	0.998
ω_b	0.02214 ± 0.00031	0.02214 ± 0.00031	0.0023	1.000
$\omega_{\rm cdm}$	0.1205 ± 0.0041	0.1205 ± 0.0041	0.0004	1.001
n_s	0.960 ± 0.017	0.960 ± 0.017	-0.0012	0.996
$\ln(10^{10} A_s)$	2.940 ± 0.016	2.942 ± 0.016	-0.112	0.999
τ	0.0540 ± 0.0074	0.0540 ± 0.0074	1.636×10^{-6}	1.000

TABLE V: Comparison of cosmological parameter estimates from multi-frequency and lite likelihoods. Δ/σ indicates the normalized shift in best-fit values; $\sigma_{\rm lite}/\sigma_{\rm mf}$ compares the the standard deviations of the two results.

EMULATION WITH GDF

Initially, 17-parameter emulators are trained with 6 ΛCDM parameters and 11 GDF parameters: the current-day energy density of the GDF Ω_{GDF}, and $w(a)$ at 10 evenly spaced nodes from $-6 \leq \log(a) \leq -2$. 14-parameter TT, EE, and TE emulators are trained with 303,000 spectra. The accuracy is determined with calculating the error (relative for TT, EE, and absolute for TE) with 20,000 testing spectra. The resulting emulators are too inaccurate to analyze. 17 parameters wis possibly too complex for the neural network to accurately learn given the limited training data.

Because of the complexity of the 17-parameter model, the problem is simplified to 10 parameters. The 10-parameter emulators are trained with 6 ΛCDM parameters and 4 GDF parameters: the current-day energy density of the GDF Ω_{GDF}, and $w(a)$ at 3 nodes at $\log(a) = -4, -3, -2$.

TT, EE, and TE emulators were trained on $\sim 100,000$ spectra. With 10,000 tests, the relative errors were calculated and displayed in figures 14, 15, and 15. The relative errors are of a similar order of magnitude as the 17-parameter model. The neural networks' relative errors are too large to analyze.

A major potential issue with this model is a high degree of complexity, and the GDF is contributing significant changes to the spectra, which are not properly accounted for in training due to the limited training data. The amount of training data used is limited due to the large size of the dataset and the constraints of available RAM and storage.

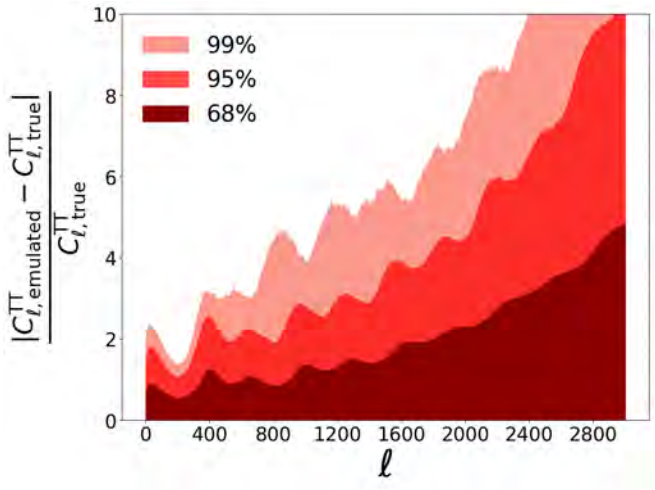


FIG. 14: 10-parameter GDF TT emulator relative error based on ∼10,000 tests.

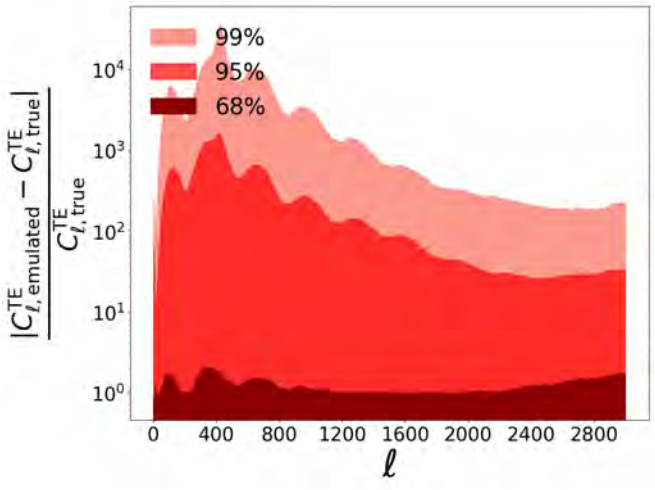


FIG. 15: 10-parameter GDF EE emulator relative error based on ∼10,000 tests.

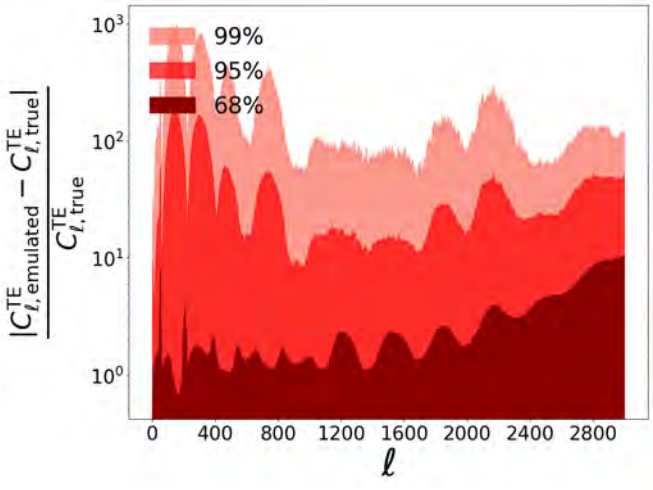


FIG. 16: 10-parameter GDF TE emulator absolute error based on ∼10,000 tests.

DISCUSSION AND CONCLUSION

This work highlights that, despite efforts to leverage cosmological emulation using neural networks—specifically through the Cosmopower framework—the challenge of efficiently estimating parameters in complex extensions of the ΛCDM model persists. By incorporating a Generalized Dark Fluid (GDF) with a time-dependent equation of state, we attempted to explore a rich class of early-universe expansion histories that have the potential to address persistent cosmological problems, such as the Hubble tension.

The ΛCDM emulator results confirm the effectiveness of emulation and differentiable likelihood techniques. Using ∼40,000 spectra for training and 8,000 for testing, the ΛCDM emulators for TT, EE, and TE spectra achieved sub-percent accuracy in nearly all multipole ranges. When used with the candl likelihood framework, these emulators produced cosmological parameter estimates that fit to SPT-3G data and are in agreement with the estimates from SPT-3G [9], validating both the reliability and computational advantage of the multi-frequency and lite approaches. These results demonstrate the practical utility of automatic differentiation in modern cosmological inference.

In contrast, the 14-parameter GDF emulators were not viable for analysis. Despite training on over 300,000 spectra, the high dimensionality introduced severe degeneracies, yielding relative errors in TT and EE spectra ranging from 40% to 80% even in the best percentile ranges. This suggests that the expressiveness of the model, although theoretically rich, introduces parameter correlations that are difficult for the emulator to resolve effectively. This is likely due to the limitations of the size of training data, the training schedule hyperparameters, and the neural network architecture. Additionally, the complexity and degeneracy of many parameter combinations possibly make it difficult for the neural network's loss to converge and achieve accuracy.

To reduce the complexity of the GDF model, the number of parameters was reduced to a 10-parameter GDF model, which limited the spline nodes for w(a) to three strategically chosen points. However, the accuracy of the 10-parameter GDF model still did not meet the required standards for differentiable analysis. Despite optimizing the training procedure, including adjusting learning schedules and refining the neural network architecture, the model continued to show high relative errors in the parameter estimates.

The high relative errors suggest that the neural networks' architectures may not be sufficiently tailored to handle the intricacies of a cosmological model with such a large parameter space. This raises questions about whether further simplifications, such as reducing the number or range of parameters, are necessary to achieve more reliable results. Alternatively, improvements in the learning schedules—such as adjusting learning rates, incorporating adaptive optimizers, or utilizing techniques like gradient accumulation—could potentially enhance the training stability. Moreover, a deeper or more specialized neural network architecture may be needed to handle the complex correlations and degeneracies inherent in the high-dimensional cosmological data.

The failure of the 14 and 10-parameter models underscores the inherent challenges when working with extensions to the ΛCDM model, particularly in the presence of Generalized Dark Fluid. While the increased parameter space offers richer possibilities for modeling the early universe, it also makes the task of efficiently and accurately emulating these models far more challenging. Thus, it is clear that future work must either involve simplifying the problem further or refining the model's architecture to better capture the dynamics of cosmological parameters, ensuring more effective differentiation and analysis in cosmological inference tasks.

ACKNOWLEDGEMENTS

I would like to thank my research advisor, Professor Francis-Yan Cyr-Racine, who provided constant guidance and support throughout the project. Additionally, I would like to thank the Rayburn Reaching Up Fund for funding undergraduate research such as this.

I would also like to thank the UNM Center for Advanced Research Computing, supported in part by the National Science Foundation, for providing the high-performance computing, large-scale storage, and computational resources used in this work.

REFERENCES

* Electronic address: bgutierrez1@unm.edu

[1] Julien Lesgourgues. (2011). The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.1104.2932

[2] Ma, C.-P., Bertschinger, E. (1995). Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges. 455, 7–7. https://doi.org/10.1086/176550

[3] Meiers, M., Knox, L., Schöneberg, N. (2023). Exploration of the Pre-recombination Universe with a High-Dimensional Model of an Additional Dark Fluid. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.2307.09522.

[4] Planck Collaboration, P.A.R. Ade et al., "Planck 2018 results. I. Overview and the cosmological legacy of Planck," *Astronomy Astrophysics* 641, A1 (2020). https://doi.org/10.1051/0004-6361/201833880.

[5] Kou, R., Lewis, A. (2025). A flexible parameterization to test early physics solutions to the Hubble tension with future CMB data. Journal of Cosmology and Astroparticle Physics, 2025(01), 033. https://doi.org/10.1088/1475-7516/2025/01/033

[6] A. Spurio Mancini, Piras, D., Alsing, J., Joachimi, B., Hobson, M. (2022). CosmoPower: emulating cosmological power spectra for accelerated Bayesian inference from next-generation surveys. Monthly Notices of the Royal Astronomical Society, 511(2), 1771–1788. https://doi.org/10.1093/mnras/stac064

[7] L. Balkenhol, C. Trendafilova, K. Benabed, Galli, S. (2024). candl: cosmic microwave background analysis with a differentiable likelihood. Astronomy and Astrophysics, 686, A10–A10. https://doi.org/10.1051/0004-6361/202449432.

[8] Balkenhol, L. (2025). Compressed "CMB-lite" Likelihoods Using Automatic Differentiation. The Open Journal of Astrophysics, 8. https://doi.org/10.33232/001c.129886

[9] L. Balkenhol, Dutcher, D., Alessio Spurio Mancini, Aristide Doussot, K. Benabed, Galli, S., Ade, R., Anderson, A. J., B. Ansarinejad, Archipley, M., Bender, A. N., Benson, B. A., Bianchini, F., Bleem, L. E., Bouchet, F. R., Bryant, L., Étienne Camphuis, Carlstrom, J. E., Cecil, T. W., Chang, C. L. (2023). Measurement of the CMB temperature power spectrum and constraints on cosmology from the SPT-3G 2018 TT , TE , and . Physical Review, 108(2). https://doi.org/10.1103/physrevd.108.023510

Appendix: Spectra with Varying Nodes

The plots present the effects of individually varying w at each of the five selected values of $\log(a) = -6, -5, -4, -3, -2$, along with the associated splines for $w(a)$. The constant nodes are set to 0, $w = 1/3$ in the region of $-14 \leq \log(a) \leq -8$, and $w = 0$ in the region of $-1.75 \leq \log(a) \leq 0$. These plots illustrate the 100 spectra generated using the Planck [4] best fit ΛCDM parameters, with variations in Ω_{GDF} sampled through Latin hypercube sampling. Specifically, Ω_{GDF} was sampled from the range $(0, 0.1)$, and the value of $w(a)$ at the node was varied over the interval $(-0.4, 0.8)$. Each plot corresponds to a different node placement and shows the resulting spectra with varying values of $w(a)$. The selected nodes at $\log(a) = -4, -3, -2$ correspond to the greatest variation observed in the spectra and are highlighted in the plots.

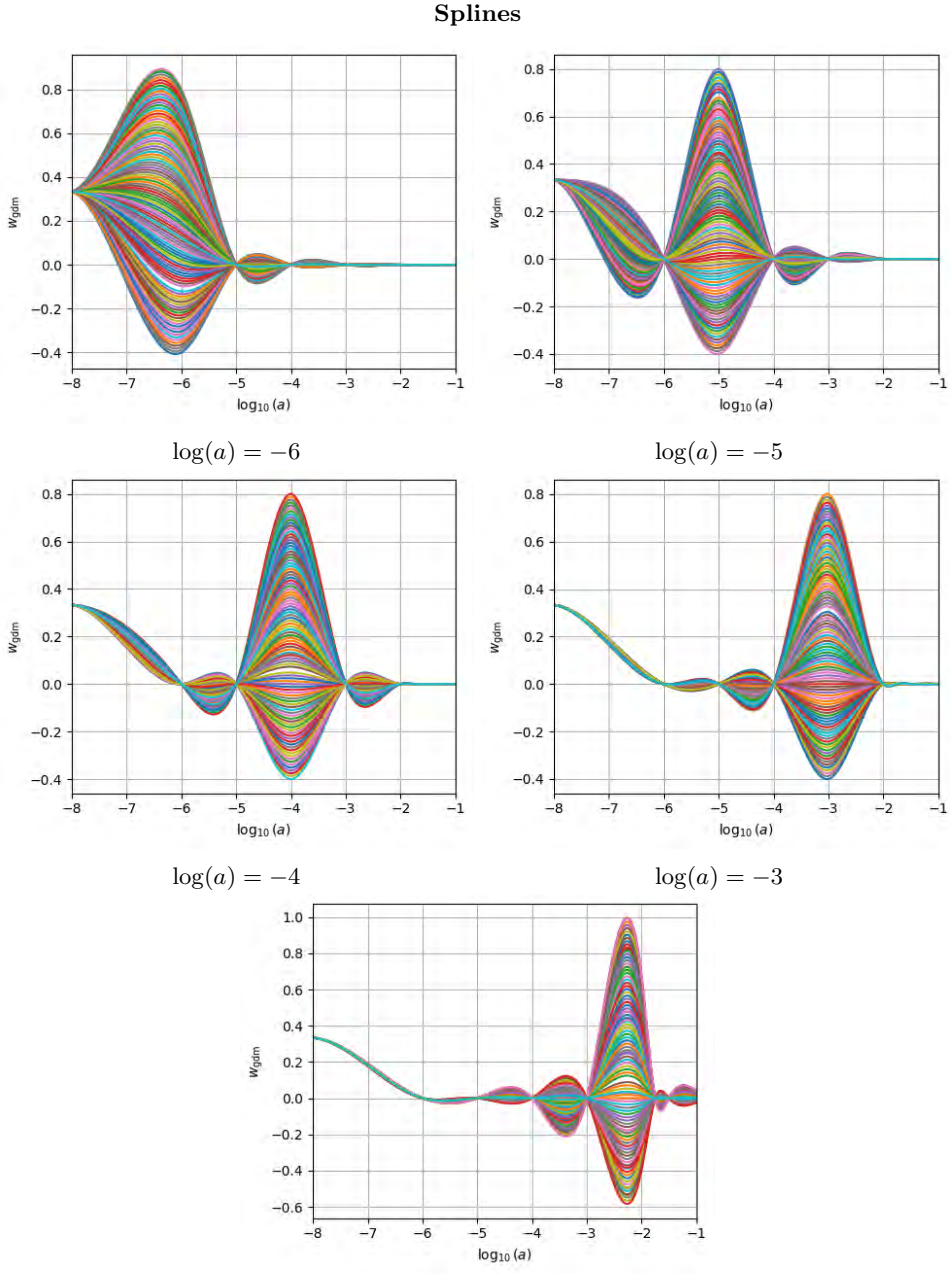
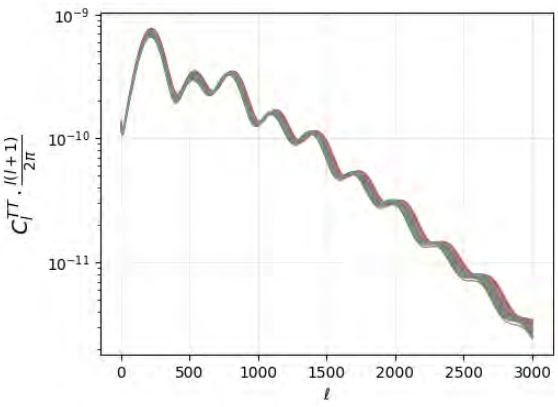
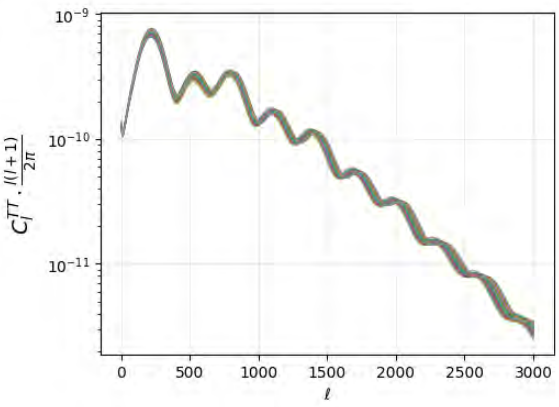


FIG. 17: 100 generated $w(a)$ splines with a varying node at the respective $\log(a)$.

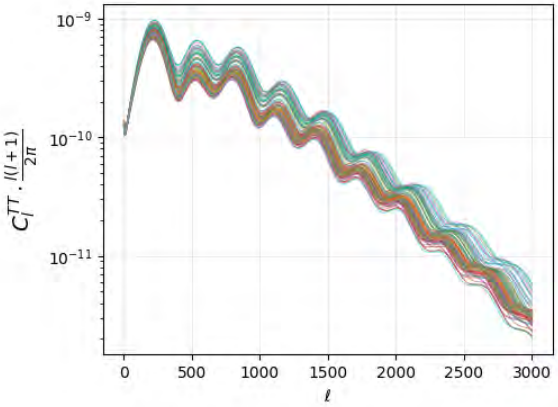
TT Spectra



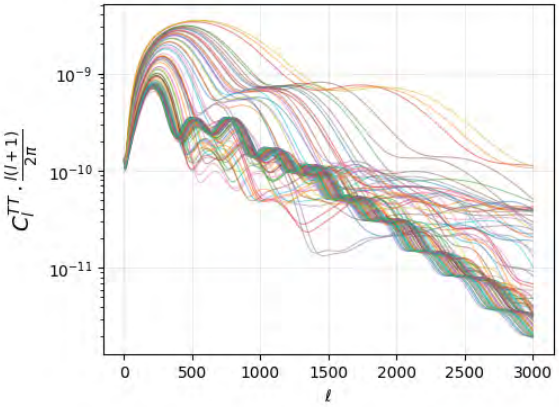
$\log(a) = -6$



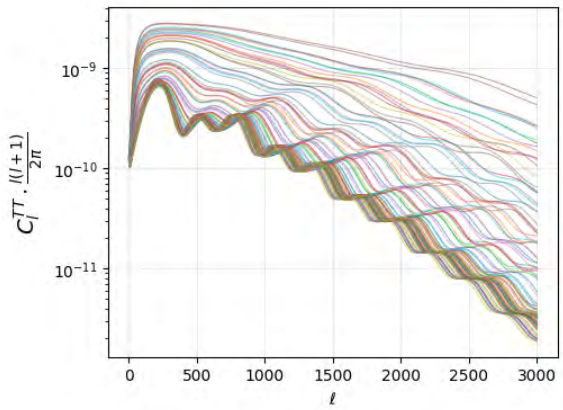
$\log(a) = -5$



$\log(a) = -4$



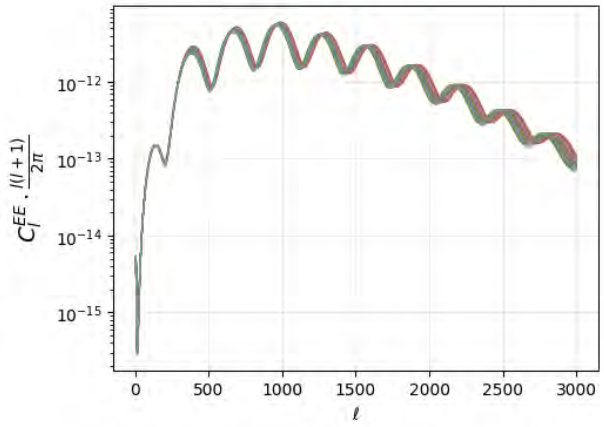
$\log(a) = -3$



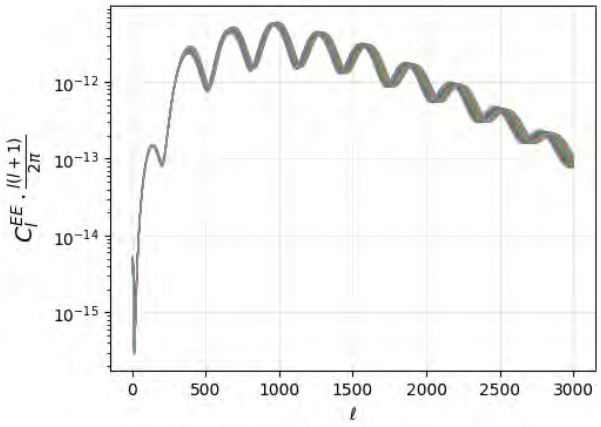
$\log(a) = -2$

100 TT spectra with varying node at the respective $\log(a)$.

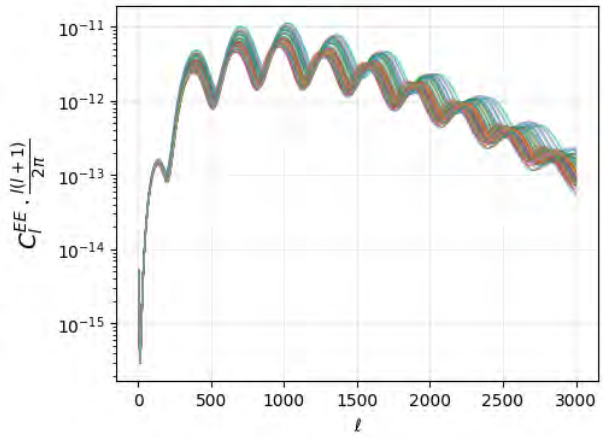
EE Spectra



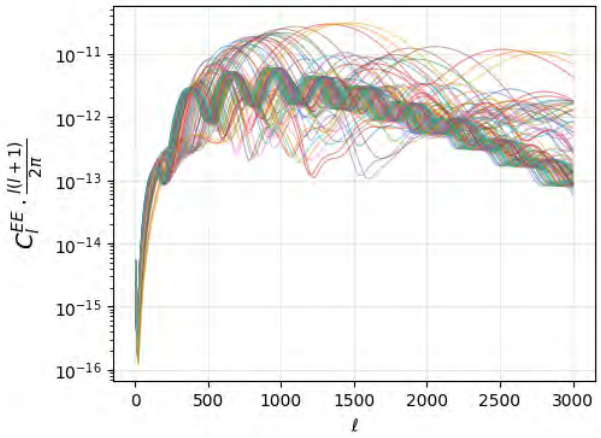
$\log(a) = -6$



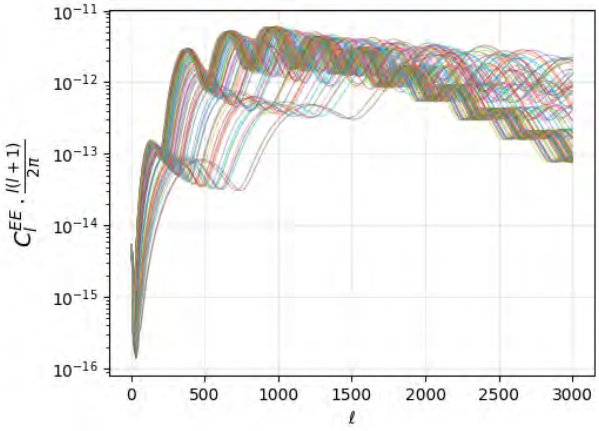
$\log(a) = -5$



$\log(a) = -4$



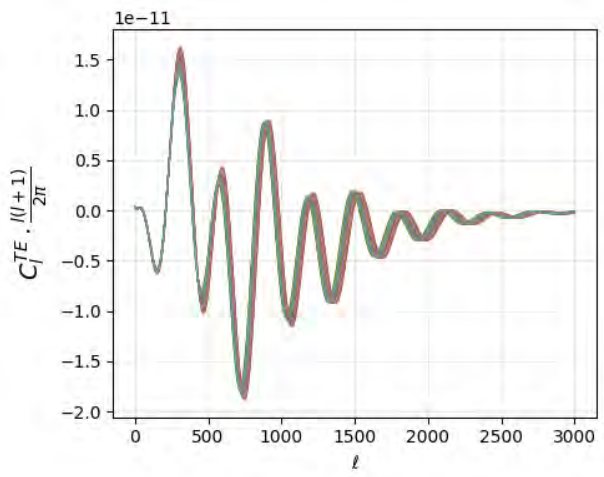
$\log(a) = -3$



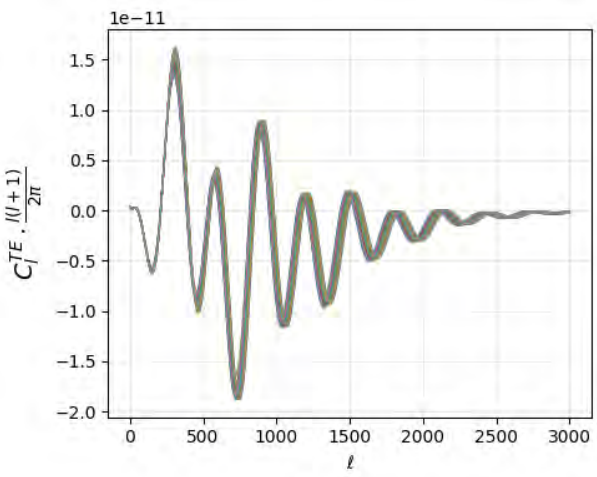
$\log(a) = -2$

FIG. 19: 100 EE spectra with varying node at the respective $\log(a)$.

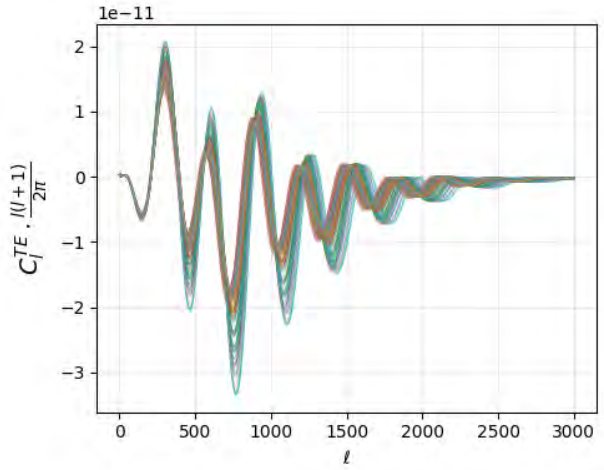
TE Spectra



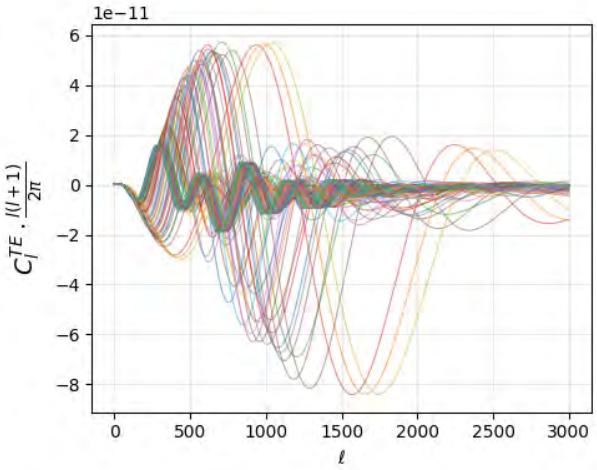
$\log(a) = -6$



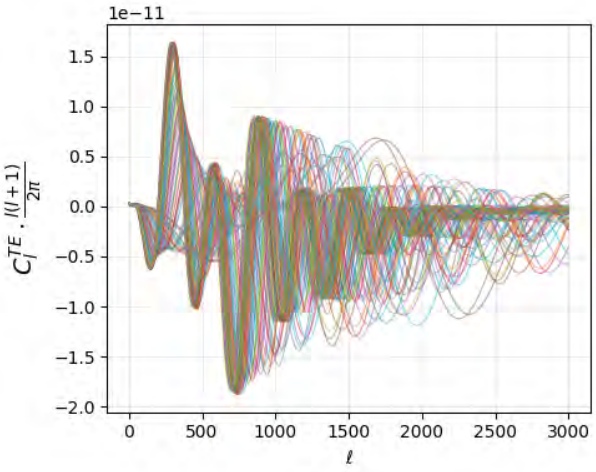
$\log(a) = -5$



$\log(a) = -4$



$\log(a) = -3$



$\log(a) = -2$

FIG. 20: 100 TE spectra with varying node at the respective $\log(a)$.