THE UNIVERSITY *of* EDINBURGH
**School of Physics
and Astronomy**

Senior Honours Project
Exploring New Physics With Neural Networks

Arnav Agarwal
April 2024

Abstract

In this study, I review, refine, and evaluate `BaCoN`, a Bayesian Convolutional Neural Network that classifies input density power spectra into cosmological theories. I streamlined `BaCoN`'s data processing pipeline, reducing training time from days to 1-2 hours – crucial for leveraging the extensive, high-resolution datasets anticipated in Stage-IV studies. Although the revised pipeline reduced accuracy by 6%, this may be rectified through profiling and advanced hyperparameter optimization. I developed a novel method for generating 'random spectra' – designed to detect unseen physics – by applying random filters uniformly across all input spectra, allowing potentially better representation of new physics. Additionally, I integrated the Dark Scattering framework, which initially led to a 15% drop in accuracy when introduced as a separate class due to its overlap with the wCDM theory. Given that this physical degeneracy is pronounced on large scales, and that removing these scales counter-intuitively and drastically reduced accuracy, I suspect flaws in the `BaCoN`'s noise model. Attempts to remove Dark Scattering spectra with lower modification strengths failed to improve performance, prompting me to instead replace the wCDM spectra with Dark Scattering spectra, yielding a modest accuracy of 87.3%. These findings underline the need to refine the power spectrum generation process, consider revising the noise model, and integrate explainable machine learning methods, setting the stage for `BaCoN`'s use in future high-precision cosmology studies.

Declaration

I declare that this project and report is my own work.

Signature: Date: 12/04/2024

Supervisors: Dr. B. Bose and Prof. A. Pourtsidou   10 Weeks

Contents

1 Introduction

A unified theory of the nature of the cosmological forces present in our Universe, including dark matter, dark energy and gravity, remains elusive. Neural networks have shown great promise in tackling such complex scientific challenges by embedding theoretical models directly into their architecture. They thus pave the way for the discovery of underlying physical principles from data, serving as vehicle for the development of theoretical constraints in physics and across science [Beucler *et al.,* Mar. 2021; Cichy & Kaiser, Feb. 2019; Yang *et al.,* Dec. 2020].

The most successful theory of modern cosmology is Lambda – Cold Dark Matter (ΛCDM), which maintains agreement with general relativity through the introduction of non-evolving dark energy and dark matter. However, it is inexhaustive in its explanation of cosmological effects [Cohen-Tannoudji, June 2018]. Attempts have been made in modern cosmology to find alternative models, which include deviations from Einstein's theory of General Relativity.

The cosmological principle, which asserts that the Universe is isotropic and homogeneous over very large distances, is widely accepted. At these largest length scales, we are able to confidently infer the Universe's matter-energy density distribution from various observations. The European Space Agency's recent Planck mission [Planck Collaboration, Sept. 2020] has provided high-precision measurements of the Cosmic Microwave Background (CMB), giving us crucial information on primordial temperature fluctuations which drove the development of structure. On the other hand, the Universe's clustering at smaller scales, often referred to as the *Large Scale Structure* (LSS), remains a contentious area of research.

In cosmology, a distinction is made between the large scale, linear regime, where density perturbations are sufficiently small to be effectively handled with linear perturbation theory, and the non-linear regime, where these perturbations are large enough that non-linear effects must be considered. Upcoming data from Stage-IV cosmological surveys are expected to greatly enhance our precision at these non-linear scales, where in addition to cosmological effects, baryonic processes and the effects of massive neutrinos become significant. Two such surveys are EUCLID satellite and the Large Scale Space Telescope (LSST) [Amendola *et al.,* Sept. 2013; LSST Dark Energy Science Collaboration, Nov. 2012].

[Mancarella *et al.,* Jan. 2022] presented a novel approach for assessing how closely cosmological theories align with data on LSS. They propose the use of their Bayesian Cosmological Network (`BaCoN`) to estimate confidence in cosmological theories. `BaCoN` uses *power spectrum* data, which describes the density distribution of the Universe statistically. The networks classify power spectra according to the theories they are most commensurate with. This allow us to constrain the set of parameters which need be tested against cosmological data. Given their ability to do this effectively without the need to select from a vast set of parameters may make them an effective tool to replace or supplement existing Markov Chain Monte Carlo (MCMC) approaches to analysing cosmological data. This tool is therefore poised to leverage the high quality data to be provided by Stage–IV studies, without the need to over-constrain the theory space.

The authors presented three distinct types of neural networks. Firstly, a five-class model which classifies observed spectra as arising from the ΛCDM, $f(R)$, DGP or wCDM

theories, or as being 'random'. Secondly, a two class model which classifies observations as arising from the standard ΛCDM model or not. Finally, they propose the use of specialised networks which further constrain the theory space to ΛCDM and a promising competitor, or indeed any constrained set of competitors which the five-class models suggest. In this report, I will develop upon on the second type.

I will begin by giving the physical background behind standard cosmology and its competing theories in Section 2, as well as provide context on Stage–IV cosmological studies. Then, in Section 3 I outline of the data generation process for `BaCoN`, give a summary of the networks' Bayesian-Convolutional Network architecture and explain the machine learning principles employed. In Section 4, I will suggest various improvements to and tests on the latest version of `BaCoN` [Thummel *et al.,* Mar. 2024].

The primary improvement I present is the inclusion of the Dark Scattering framework by [Simpson, Oct. 2010] in the network's data. After including this as an independent theory and assessing the network's classification performance, I determine the impact of excluding linear length scales from the data. I do this to determine the physical scales which give the networks the most relevant information for discerning Dark Scattering spectra from the theories previously included in `BaCoN`, which already demonstrates excellent accuracy without Dark Scattering [Thummel *et al.,* Mar. 2024]. I also assess the impact of thresholding the modification strength of the Dark Scattering data from the wCDM theory, in order to determine if this lifts the classification degeneracy between the two theories.

The 'random' spectra included in `BaCoN`, account for new physics beyond theories included in its corpus. To generate these, I propose a slightly modified approach to that presented in the first iteration of `BaCoN` [Mancarella *et al.,* Jan. 2022], which relies on using base spectra from the whole set of tested cosmologies. This is in contrast to the approach in [Thummel *et al.,* Mar. 2024], which only uses ΛCDM base spectra

I conduct these tests with a modification of `BaCoN`'s original data processing and feeding process. These changes may increase the scalability of `BaCoN`, allowing more efficient training on datasets of high resolution and sample size, like those expected from Stage–IV studies. I share the results and discuss the models' performance in Section 5.

This study thereby serves as an extensive review of the key physical and computational concepts behind `BaCoN`, an extension of the theory space it works with, an analysis of its performance in various physical regimes and a set of proposed improvements for future work.

2 Physical Background

2.1 Power Spectra

`BaCoN` was built to study LSS through the *power spectrum*, a metric to capture correlations in the matter-energy density distribution of the Universe. In order to elucidate the physical intuition behind this metric, I will briefly discuss the formalism behind it.

In general, the density ρ of the Universe is dependent on space and time, giving $\rho = \rho(\vec{r}, t)$. If we take a spatial average of the density $\overline{\rho}$ over a sufficiently large volume V, comparable to the biggest structure in the observable Universe, we find

$$\overline{\rho}(t) = \frac{1}{V} \int_V \rho(\vec{r}, t) d^3 r.$$

Taking this as the isotropic, background part of the energy density, we can then define the actual energy density as the sum of this and a perturbation $\delta\rho(\vec{r}, t)$,

$$\rho = \overline{\rho} + \delta\rho = \overline{\rho}(1 + \delta),$$

where

$$\delta(\vec{r}, t) = \frac{\delta\rho(\vec{r}, t)}{\overline{\rho}} = \frac{\rho(\vec{r}, t) - \overline{\rho}}{\overline{\rho}}.$$

Directly measuring $\rho(\vec{r}, t)$, is not possible. Our current understanding posits that the observable density distribution has resulted from the amplification of primordial fluctuations (or *seeds*) by gravitational instability. Since direct observational access to these primordial fluctuations is unfeasible, statistical methods become essential for analyzing cosmic evolution. [Bernardeau *et al.*, Sept. 2002] make this point and present a derivation of the power spectrum from spatial density correlation functions, in the process showing that Fourier transforms can provide the statistical description needed. To this end, we can express our density fluctuation field to be the inverse Fourier transform over its Fourier components $\delta_{\vec{k}}$,

$$\delta(\vec{r}, t) = \frac{V}{(2\pi)^3} \int \delta_k e^{-i\vec{k} \cdot \vec{r}} d^3 k,$$

where individual components are given by

$$\delta_k(t) = \frac{1}{V} \int \delta(\vec{r}, t) e^{i\vec{k} \cdot \vec{r}} d^3 r.$$

This allows us to break down the density fluctuation field to an infinite number of plane waves. As usual with Fourier transforms, larger k modes correspond to smaller physical length scales. Complete knowledge of the Fourier components $\delta_{\vec{k}}(t)$ would allow us to reconstruct the density field $\delta(\vec{r}, t)$. Each Fourier component is a complex number, so we can express it in exponential form,

$$\delta_{\vec{k}}(t) = \left| \delta_{\vec{k}} \right| e^{i\phi_{\vec{k}}}.$$

Assuming that the isotropy of the Universe holds, as predicted by the cosmological principle, we can define a quantity that does not depend on the directions, only the magnitudes, of $\vec{k}$. In order to estimate the variation of the density fluctuations with time, we can take observations of the power spectrum at various redshift bins z, corresponding to the observations at greater distances (as explained by the relativistic doppler effect), and therefore at an earlier time in an expanding Universe. The power spectrum is thus defined as the average over all orientations of the wavenumber $\vec{k}$ of the mean squared amplitude of the Fourier components,

$$P(k, z) = \left\langle \left| \delta_{\vec{k}}(z) \right|^2 \right\rangle.$$

Matter power spectrum observations constructed from observations of visible galaxies, weak lensing and CMB measurements are crucial for evaluating the validity of cosmological theories. The ultimate objective of `BaCoN` is to leverage high-precision power spectrum observables expected from Stage-IV studies (see Section 2.5) to deliver a meaningful confidence in various physical models [Mancarella *et al.,* Jan. 2022; Thummel *et al.,* Mar. 2024].

2.2 Standard Cosmology

The standard model of cosmology, ΛCDM, posits complete agreement with General Relativity through the introduction of dark matter and dark energy - the *dark sector*.

When Einstein was developing his General Theory of Relativity in the early 20th century, he sought to reconcile how the same mass could experience either inertial and gravitational effects depending on the frame of reference, known as Mach's principle. This required postulating a hypothetical matter "of which the total mass is so enormously great, that compared with it all matter known to us is utterly negligible," such that this matter could be taken as the reference matter from which other frames' inertial properties could be inferred [Cohen-Tannoudji, June 2018]. It was later noted by De Sitter that this matter did not display non-gravitational interactions (that were known), leading to it being coined *dark matter*. Furthermore, in order to account for the accelerated expansion of the Universe, a repulsive gravitational force – or 'negative pressure' – was required, which was theorised to arise from the similarly elusive *dark energy* [Kunz & Sapone, Mar. 2007]. This new kind of energy appears as an additional term in Einstein's field equations, linear in the metric tensor $g_{\mu\nu}$ and proportional to the *cosmological constant* Λ, which can be interpreted as the vacuum energy of matter fields. When this formalism is incorporated into the Robertson-Walker metric and the Friedmann-Lemaitre equations, it describes the curvature of the Universe and its rate of expansion [Cohen-Tannoudji, June 2018].

2.2.1 Evidence for the Dark Sector

In order to authoritatively prove the existence of these physical entities, direct physical observation would be required. This remains outside the realm of current research, but experimental evidence exists suggesting the presence of the *dark* components in the Universe. Examples of this lie in galactic rotation curves and CMB measurements.

Galactic rotation curves are often cited as the first robust evidence for the existence of dark matter. In [Persic *et al.,* July 1996], the authors analysed approximately 1100 optical and radio rotation curves. The findings revealed flat or even rising rotation curves are observed at significant distances from galactic centers. Taking Newton's law of gravitation into account, this suggests a invisible form of matter extending far from the galactic core.

CMB measurements are taken as evidence for dark energy. In [Dodelson & Knox, 2000], researchers explored how the anisotropy data from the CMB compellingly constrains the spatial curvature of the Universe to be near zero. A flat universe must satisfy the flatness sum rule, which states that there is a critical density ρ_c which the universe must meet in order to be flat. This is expressed as

$$\rho_R + \rho_M + \rho_\Lambda = \rho_c,$$

where ρ_R is the energy density of radiation and ρ_M is the energy density of non-relativistic matter. These two are not found to sum to ρ_c, a gap which must be filled by the density of an unseen component, ρ_Λ. Whether this component is an actual physical entity, dark energy, or a modification to standard General Relativity remains contentious, as discussed in Section 2.3.

2.2.2   The ΛCDM Model

In ΛCDM, dark energy enters the theoretical framework through the cosmological constant Λ, and dark matter through the postulation of cold dark matter (CDM) - matter at non-relativistic speeds. In ΛCDM, ordinary matter makes up only a small fraction of the universe and cold dark matter dominates. This model has been remarkably successful in explaining a wide array of astronomical observations, including the excellent agreement with primordial CMB measurements from the Planck satellite [Planck Collaboration, Sept. 2020] and the observed acceleration of the expansion of the Universe [Peebles & Ratra, Apr. 2003]. However, several issues with the model remain, particularly at the level of LSS. An example is the cusp-core problem, where ΛCDM predicts that dark matter density should sharply increase towards the centers of galaxies (a "cusp"), but observations suggest a more uniform distribution (a "core") [de Blok, Nov. 2009]. Another challenge is the missing satellites problem, which arises from the prediction of numerous small satellite galaxies around larger ones like the Milky Way, far exceeding the number observed [Klypin *et al.,* Sept. 1999]. These issues, along with other discrepancies and the elusiveness of the dark sector, have given rise to a range of competing cosmological theories.

2.2.3   Parameters

In the computation of power spectra, standard cosmology introduces the parameters $\{\Omega_m, \Omega_b, H_0, n_s, A_s\}$. Ω_m and Ω_b describe the fraction of matter and baryonic matter in the Universe respectively [Thummel *et al.,* Mar. 2024]. H_0 is hubble's constant, which describes how recession velocity in an expanding Universe scales with distance. The spectral index n_s describes the k dependence of fluctuations in the primordial CMB and A_s describes the amplitude of these fluctuations [De Simone, June 2018]. These parameters, and those arising from the competing theories presented in the following section, are required by the power spectrum calculation process described in Section **??**.

2.3   Competing Theories

`BaCoN` compares a set of competing theories against standard ΛCDM cosmology, and each other. These are the Dvali-Gabadadze-Porrati (DGP) model, the $f(R)$ modified gravity model, the wCDM model and the theory of Dark Scattering. Successful theories will demonstrate good agreement Planck CMB data, whilst giving accurate LSS predictions. I will now provide a brief overview of these competing theories and the parameters they introduce.

7

2.3.1 DGP

The Dvali-Gabadadze-Porrati (DGP) theory modifies General Relativity gravity by introducing extra spatial dimensions. This allows for a self-accelerating Universe without the need for dark energy. The model suggests that the leakage of gravity into a higher-dimensional space drives cosmic acceleration [Dvali *et al.*, July 2000]. DGP introduces the parameter Ω_{rc}, which appears in the equation describing the expansion of the Universe, the Friedmann equation. This quantity, known as the crossover density parameter, encapsulates the effect of the extra spatial dimension on the Universe's expansion rate. It is related to the cross-over scale, r_c, which represents the critical length beyond which gravity begins to leak into the higher-dimensional space, marking the transition from 4-dimensional to 5-dimensional gravitational behavior. These regulate the strength and scale at which the modifications to General Relativity become significant, thereby altering cosmic acceleration and the growth of LSS whilst permitting agreement with the primordial CMB at linear scales [Lue & Starkman, Mar. 2002; Luty *et al.*, Oct. 2003]. Despite its promise, DGP fails to explain some empirical observations. One significant issue is its prediction of strong interactions at cosmological scales, leading to conflicts CMB observations [Luty *et al.*, Oct. 2003]. DGP also struggles to simultaneously match LSS observations and detailed data on the Universe's expansion history [Maartens & Majerotto, July 2006].

2.3.2 $f(R)$

$f(R)$ models also modify the General Relativity prescription of gravity, avoiding dark energy. They replace the Ricci scalar, which describes the curvature of spacetime and its gravitational effects, with a more general function $f(R)$. This allows the introduction of additional degrees of freedom and novel effects in the theory of gravity. These aim to recover GR in the high-curvature (linear) regime and deviate from it in the low-curvature (non-linear) regime. The parameter $|f_R|$ quantifies the deviation from R, where $f_R = df(R)/dR$. $f(R)$ functions must be designed so that they lead to the rate of expansion scaling with distance, mimic the cosmological constant at cosmological scales, and remain undetectable in solar system tests [Hu & Sawicki, Sept. 2007]. $f(R)$ theories are prone to predicting LSS formation processes which are not commensurate with observations [Sotiriou & Faraoni, Mar. 2010]. It should also be noted that their effect on the power spectrum can be difficult to distinguish in the linear regime from that of ΛCDM [Hu & Sawicki, Sept. 2007].

2.3.3 wCDM

The wCDM model extends the ΛCDM framework by treating dark energy as a fluid with a dynamical equation of state, characterized by the parameters w and w_a. w denotes the ratio of pressure to density of dark energy $w = P/\rho$, a departure from the cosmological constant scenario where $w = -1$ – a strictly negative pressure [Linder, 2003; Chevalier & Polarski, Apr. 2001]. w_a accounts for the time evolution of w, describing its change with respect to the Universe's scale factor a, expressed as $w(a) = w + w_a(1-a)$ [Linder, 2003]. This allows wCDM to describe a broader spectrum of dark energy behaviors than the constant energy density scenario in ΛCDM [Copeland *et al.*, Nov. 2006]. Constraining the

values of w and w_a from CMB and LSS observational data has proven to be a significant challenge due to the degeneracies among cosmological parameters and the limitations of pre-Stage-IV observational precision [Copeland *et al.*, Nov. 2006].

2.3.4 Dark Scattering

[Simpson, Oct. 2010] proposed an extension to wCDM in the form of dark matter – dark energy scattering. Noting that non-gravitational interactions with all other cosmologically significant particles have been previously explored, he proposes a toy model for the scattering mechanism between dark energy and dark matter. He does so without making any assumptions about the exact particle physics mechanisms occurring, instead focusing on the process of energy transfer between the two entities. By assuming cold, non-relativistic dark matter, and noting the very low theorised density of dark energy, he posits that such scattering would be an elastic process. He introduces the parameter σ_D, the scattering cross section of the process, which quantifies the likelihood of a dark matter – dark energy scattering event. Such a process would, assuming that $w \geq -1$, introduce a drag effect on LSS growth. LSS growth data can thus be used as a constraint on σ_D, which he shows to allow considerably large cross sections compared to other scattering processes. In order to account for the modulation of interaction strength due to particle mass, the Dark Scattering spectra generated for `BaCoN` use the coupling parameter ξ. $\xi = \sigma_D/m_c$ where m_c is the mass of a cold dark matter particle [Carrilho *et al.*, Mar. 2022]. Simpson notes that Dark Scattering would leave the background (linear) cosmology largely unaltered from wCDM, making it difficult to distinguish.

2.4 Non-Cosmological Effects

Forthcoming data from Stage-IV studies will provide measurements at non-linear scales with a unprecedented level of precision. It consequently becomes crucial to include non-cosmological effects, which become significant at these scales, in our models. This is to account for the non-trivial ways that these effects couple with the theories we are studying, particularly those with modifications to gravity [Schneider *et al.*, Mar. 2019]. Namely, these effects are of massive neutrinos and baryonic physics. I will briefly review these effects in this subsection.

Baryonic feedback is a process driven by star formation and Active Galactic Nuclei (AGNs) — the extremely energetic centers of distant galaxies. Star formation and AGN activity expel energy in the form of baryons, heating and displacing interstellar and intergalactic gas. This redistribution alters the gravitational potential and density profiles when compared to dark matter-only simulations. This has a significant effect on *weak lensing* observables, which are used to measure the power spectrum, especially on non-linear scales [Semboloni *et al.*, Sept. 2011; van Daalen *et al.*, Nov. 2019]. [Schneider *et al.*, Mar. 2019] find that neglecting this effect can lead to underestimations of the weak lensing signal by 15-25% on certain scales.

A similar effect on LSS is caused by considering nonzero neutrino mass. Neutrinos suppress structure formation on non-linear scales because they can freely stream out of potential wells, erasing small-scale fluctuations. Cosmological models failing to account for massive neutrinos can misinterpret the growth of structure [Mead, Heymans, *et al.*,

Mar. 2016]. Simulations show that the effect of neutrino mass on the matter power spectrum leads to a measurable suppression of power at small scales [Heitmann *et al.,* Mar. 2016].

`BaCoN-II` is trained on data generated to include these two effects, as described further in Section **??**. The importance of doing so is clear in the context of Stage-IV survey data.

2.5 Stage-IV Surveys

The Dark Energy Task Force has categorised cosmology research into distinct stages of progression. Stage-I represents current knowledge, Stage-II represents the anticipated state of knowledge pending ongoing projects and Stage-III are near-term, medium cost proposed projects. Stage-IV studies represent the latest efforts in this progression, employing large-scale surveys to extract high-precision measurements of the Universe's accelerating expansion. These studies involve collaborations and instruments designed to gather massive datasets over broad sky areas, employing advanced technologies in telescopic and observational capabilities which enable the detection of even the tiniest deviations to the standard model [Albrecht *et al.,* Sept. 2006]. Examples of Stage-IV surveys include DESI [Levi *et al.,* July 2019], the Nancy Grace Roman Space Telescope [Akeson *et al.,* Feb. 2019], the Large Scale Space Telescope at the Vera Rubin Observatory [LSST Dark Energy Science Collaboration, Nov. 2012] and the EUCLID satellite [Amendola *et al.,* Sept. 2013]. I will briefly introduce the latter two.

2.5.1 EUCLID

The EUCLID Satellite is European Space Agency (ESA) mission – a comprehensive six-year survey at the L2 Sun–Earth Lagrangian point. It plans to scrutinize over 1.5 billion galaxies across a 15,000 square degrees expanse of the sky, leveraging a 1.2m diameter telescope equipped with two sophisticated instruments – the visible imager (VIS) and the near-infrared photometer/spectrograph (NISP). These instruments, operating in visible and near-infrared bands will capture high-resolution images and spectra across spatial frequencies from $k \approx 0.1h\mathrm{MPc}^{-1}$ to $k \approx 10h\mathrm{MPc}^{-1}$ and redshifts $z = 0.9$ to 1.8, facilitating a detailed analysis of the galaxy power spectrum [Laureijs *et al.,* Oct. 2011]. The magnitude of EUCLID's data will be immense, producing "a catalog of up to 100 million galaxy redshifts and an imaging survey... of up to 2 billion galaxy images," at a high spatial resolution [Amendola *et al.,* Sept. 2013]. This underscores the need for computationally efficient and scalable modelling techniques, a key driver for my work on computational optimisations for `BaCoN`, presented in Section 4.1.

2.5.2 LSST

The Vera C. Rubin Observatory's Large Synoptic Survey Telescope (LSST) is set for a 10-year astronomical survey, employing an 8.4-meter primary mirror and a 3.2-gigapixel camera to conduct systematic observations across six optical bands, covering 18,000 square degrees of the sky. This endeavor aims to amass a detailed dataset of over 20 billion galaxies, probing spatial frequencies from $k \approx 0.1h\mathrm{MPc}^{-1}$ to $k \approx 10h\mathrm{MPc}^{-1}$, and spanning redshifts from $z = 0.1$ to $z > 3.0$. The LSST Dark Energy Science Collaboration (DESC) will leverage this vast dataset through several cosmological probes. To navigate

and analyze the anticipated influx of data, which is projected to include up to 100 million galaxy redshifts and 2 billion galaxy images, the DESC emphasises the importance of scalability and efficiency of models working with LSST data [LSST Dark Energy Science Collaboration, Nov. 2012].

3 Computational Background

The raw dataset used in this report is generated by the authors of `BaCoN-II`, the latest iteration of the networks [Thummel *et al.,* Mar. 2024]. The entirety of the dataset is found here. In this section, I will review the method they used for modelling power spectra.

3.0.1 Halo Model

Complete power spectra P_{NL}, incorporating non-linear and linear effects are generated according to the halo model, which assumes that all matter in the Universe is localized in *virialized* structures, clusters at gravitational equilibrium, called *haloes*. The `BaCoN` authors use a halo model reaction function, through which the complete power spectrum can be expressed as

$$P_{\mathrm{NL}} = \mathcal{R}(k,z)P_{\mathrm{pseudo}}(k,z). \tag{1}$$

Here, P_{NL} is the complete non-linear spectrum, $\mathcal{R}$ is the halo model *reaction* and P_{pseudo} is a 'psuedo' power spectrum. P_{pseudo} is constructed from the ΛCDM model evolved with standard gravity up to a final redshift at which the clustering of matter it describes matches that of the target cosmology in P_{NL}, allowing for a smoother transition between linear and non-linear scales [Cataneo *et al.,* July 2019]. $\mathcal{R}$ quantifies corrections to P_{pseudo} arising from non-ΛCDM cosmology, massive neutrinos and non-linear effects. $\mathcal{R}$ is generated using the publicly available code `ReACT` [Bose, Cataneo, *et al.,* Sept. 2020; Bose, Wright, *et al.,* Sept. 2021], which takes the parameter $\sum m_\nu$, the sum of neutrino masses.

A model of P_{pseudo} including the effects of Baryonic feedback is generated through the codes `HMCode2020` [Mead, Heymans, *et al.,* Mar. 2016; Mead, Brieden, *et al.,* Jan. 2021], `MGCAMB` [Zucca *et al.,* May 2019] and `CLASS` [Lesgourgues, May 2011]. `HMCode2020` generates spectra including baryonic feedback, taking the temperature of AGNs (see Section 2.4) T_{AGN} as a parameter to be varied. The program takes an input linear spectrum with ΛCDM or beyond ΛCDM cosmology. These input spectra are generated by `MGCAMB` for ΛCDM, $f(R)$, wCDM and Dark Scattering models, and by `CLASS` for the DGP model [Carrilho *et al.,* Mar. 2022; Thummel *et al.,* Mar. 2024]. However, `HMCode2020`'s accuracy for standard, non-linear spectra is limited. To leverage its modelling of Baryonic feedback and non-ΛCDM physics, its predictions are combined with a highly accurate model of the non-linear ΛCDM spectrum, generated by the power spectrum emulator `EuclidEmulator2` [Euclid Collaboration, May 2021]. This revised approach is expressed as

$$P_{\mathrm{NL}} = B(k,z)P_{\Lambda\mathrm{CDM}}^{\mathrm{EE2}}(k,z), \tag{2}$$

where the superscript "EE2" represents a linear power spectra generated by `EuclidEmulator2` and B is a *boost factor* given by

$$B(k, z) \equiv \frac{\mathcal{R}(k, z) P_{\text{pseudo}}^{\text{HMC}}(k, z)}{P_{\Lambda\text{CDM}}^{\text{HMC}}(k, z)}. \tag{3}$$

The superscript "HMC" indicates power spectra generated by `HMCode2020`. $B/\mathcal{R}$ is then an estimate of the ratio at every (k, z) of the power spectrum – including massive neutrinos and baryonic feedback – to the dark-matter only spectrum [Carrilho *et al.,* Mar. 2022].

3.0.2 Accuracy

In [Bose, Wright, *et al.,* Sept. 2021] it is found that the data generation approach outlined here achieves $\lesssim 5\%$ accuracy compared to hydrodynamical numerical simulations of power spectra for $k \leq 3h\text{MPc}^{-1}$, giving us a strong estimate of their actual accuracy. The authors of `BaCoN` therefore impose a maximum k value of $2.5h\text{MPc}^{-1}$, as do I in this paper [Mancarella *et al.,* Jan. 2022; Thummel *et al.,* Mar. 2024]. The data are generated for a range of redshift bins $z \in 1.5, 0.785, 0.478, 0.1$ and Fourier modes $0.01 \leq k \leq 2.5$ $h\text{MPc}^{-1}$, as expected from a Stage-IV survey (see Section 2.5).

3.0.3 Parameters

Examples are generated for each class by varying the base parameters $\{\Omega_m, \Omega_b, H_0, n_s, A_s\}$, described in Section 2.2; the theory-specific parameters $\{\Omega_{rc}, |f_R|, w, w_a, \xi\}$, described in Section 2.3 and the parameters related to non-cosmological effects $\{T_{AGN}, \sum m_\nu\}$, described in Section 2.4. The base parameters are sampled from a distribution whose mean is taken as the best fit from the Planck data in [Planck Collaboration, Sept. 2020] and whose variance is taken from the uncertainty estimates in EUCLID probes, given in [Blanchard *et al.,* Oct. 2020]. Furthermore, a hard limit is set on $\Omega_b \in [0.04, 0.06]$, as these are the limits imposed by `EuclidEmulator2` [Euclid Collaboration, May 2021]. $\{\Omega_{RC}, |f_R|, \xi\}$ are sampled from distribution with their ΛCDM values as a mean ($\Omega_{RC} = |f_R| = \xi = 0$), with a standard deviation as forecasted by an LSST-like survey [Bose, Cataneo, *et al.,* Sept. 2020; Carrilho *et al.,* Mar. 2022]. However, the actual ΛCDM values are not included. The mean of the distributions for $\{w, w_a\}$ are also taken as their ΛCDM limits $\{-1, 0\}$, with standard deviations taken from EUCLID estimates in [Blanchard *et al.,* Oct. 2020]. The limits $w_0 \in [-1.3, -0.7]$, $w_a \in [-1.5, 0.3]$ are enforced to ensure stability when using `ReACT`. The standard deviation and mean for $\sum m_\nu$ are taken from values given in [Blanchard *et al.,* Oct. 2020]. T_{AGN} is taken to have the mean value from `HMCode2020`, with a standard deviation covering fits to the `BAHAMAS` simulation [Mead, Brieden, *et al.,* Jan. 2021].

3.1 'Random' Spectra

In order to account for new physics which may be present in our observations, but are not accounted for by any of the existing theories in `BaCoN`'s corpus, [Mancarella *et al.,* Jan. 2022] introduced a class of 'random' spectra into the models data. These spectra need to have random features which are not present in the other models, but are still correlated in k and z. In the original version of `BaCoN`, the authors generate these spectra by taking randomly selected examples from the full set of included theories and applying

a set of filters to them. The algorithm to create these filters is outlined in Appendix E of their paper. In `BaCoN-II` the authors revised their approach to instead only use ΛCDM spectra as base spectra [Thummel *et al.,* Mar. 2024]. I suggest reverting to a slightly modified approach from the first `BaCoN` iteration in Section 4.1.

3.2   Noise Model

Noting that BaCoN models are developed with the goal of giving robust classifications when they are tested on Stage–IV observables in the near future, their training data includes modelling of both *aleatoric* and *epistemic* uncertainty [Mancarella *et al.,* Jan. 2022].

Aleatoric uncertainty refers to the stochasticity in the observations arising from our measurement setup. In our case, this refers to the *cosmic variance*, the statistical uncertainty that arises from observing only a finite volume of the Universe to draw conclusions about the Universe as a whole [Wiegand, A. & Schwarz, D. J., Feb. 2012]. Equation 5 of [Thummel *et al.,* Mar. 2024] is used to replicate the cosmic variance from Stage–IV-like measurements.

Epistemic uncertainty refers to the uncertainty arising from the theoretical prescriptions we use in our modelling process. All of the codes and calculations presented earlier have an inherent uncertainty. [Thummel *et al.,* Mar. 2024] find that using `HMCode2020`-only spectra (as in equation 1) as opposed to the combined `EuclidEmulator2` – `HMCode2020` approach (equation 2), introduces a deviation of up to 4%. To prevent training the neural networks to capture these code-specific differences, the `BaCoN` authors created a set of *theoretical error* curves, which attempt to introduce this error as a source of noise into the data. They posit that this error is representative of the error between the generated spectra and those from the real Universe. Their approach for generating these curves is outlined in Section 4.1 of their paper. They argue that their approach yields accurate estimates, as when testing on a network trained with spectra according to equation 2, the accuracy with test spectra generated with either approach converge as the magnitude of the error is increased (see Section 3.3.6 for an explanation of the 'train' vs 'test' split).

Before conducting my investigation of `BaCoN`'s performance at various length scales, found in Section 5.5, I conducted some exploratory data analysis to better understand the noise model at various k values.
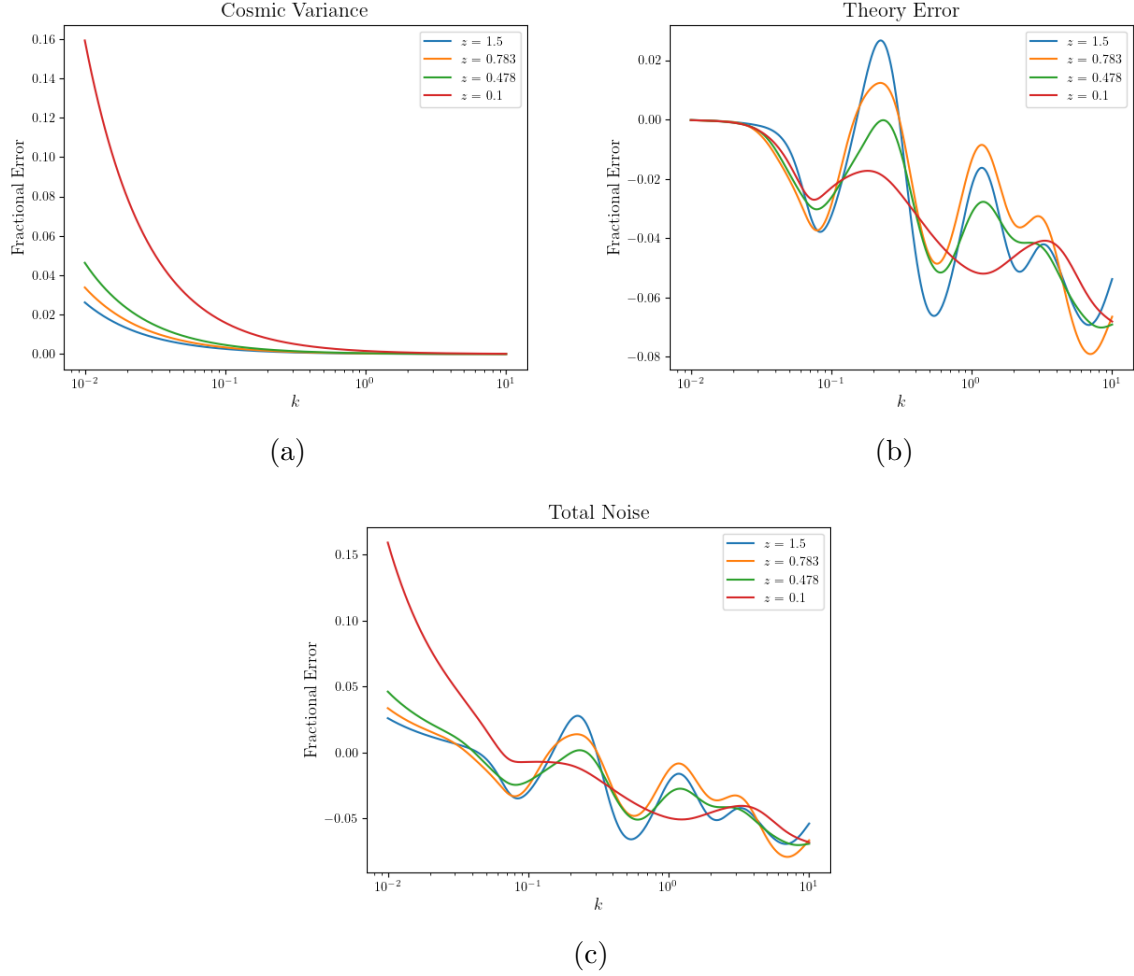
Figure 1: Plots showing a breakdown of noise by source (a, b) as well as total noise (c).

Figure 1 displays plots of both these sources of noise in the top row, and a plot showing the combined total noise in the bottom panel, for all z bins. The plots display noise as fractional error. A single, randomly selected theoretical error curve was used. Figure 1c clearly shows that a transition occurs from $k \approx 0.03$ to $k \approx 0.05$, where the noise goes from being dominated by cosmic variance, to being dominated by theoretical error. This is the case regardless of which theoretical error curves are randomly selected. I will cut this transition region out from the data in Section 5.5, allowing us to determine if BaCoN is more robust to one source of noise than the other.

To train a model with a range of random epistemic uncertainty realisations, 10 copies of each training spectra are created, each with their own randomly selected theoretical error curves. However, as the test spectra are supposed to be representative of a 'real' Stage–IV-like measurement, only one test spectrum is created, accounting only for cosmic variance and not theoretical error [Mancarella *et al.,* Jan. 2022; Thummel *et al.,* Mar. 2024].

3.3 Neural Networks and Machine Learning

The traditional approach to deriving theoretical constraints from cosmological data has been Markov Chain Monte Carlo (MCMC) methods. These sample the posterior probability distribution of cosmological parameters, allowing researchers to infer the parameters that best describe the Universe from observational data [Padilla *et al.*, 2021]. MCMC methods require defining a finite set of parameters to quantify new physics beyond the ΛCDM model. This approach becomes exceedingly complex, especially at non-linear scales, necessitating an immense number of parameters if one remains agnostic to the underlying physical model [Mancarella *et al.*, Jan. 2022; Kennedy *et al.*, Aug. 2019]. Considering statistical model selection principles, MCMC is disadvantaged on the basis of Bayesian evidence; the expansion of parameter space without proportionate support from data ('evidence') reduces the model's likelihood, making it less probable to be an accurate representation of the problem [Trotta, Mar. 2008].

On the other hand, neural networks can adaptively learn from data without needing extensive parameterisation. They present a flexible model to constrain the vast theory space of cosmological models, allowing us to quickly limit the parameter set. This may prove to be a promising alternative method for determining the probability distribution of theoretical parameters, or at least for constraining it sufficiently for a refined MCMC analysis [Mancarella *et al.*, Jan. 2022]. `BaCoN` is a combination of Bayesian and Convolutional neural networks. I will explain the networks' architecture and data handling in this section.

3.3.1 Neural Networks

A neural network is a composite function $f(\vec{x}_0) = f^{(n)}(f^{(n-1)}(\ldots f^{(1)}(\vec{x}_0)))$, where $\vec{x}_0$ is an input vector representing the features of an example in our dataset, which in our case would the $P(k, z)$ values for a single power spectrum. Each $f^{(i)}$ represents the function performed by the ith layer in the network. Each layer comprises of a series of *perceptrons*, which together pass a vector $\vec{x}_i$ to every perceptron in the next layer in response to an input vector $\vec{x}_{i-1}$ from perceptrons in the previous layer. Each layer can have a varying number of perceptrons. The action of the jth perceptron in the ith layer can be modelled as $x_{ij} = \phi(\vec{w}_{ij} \cdot \vec{x}_{i-1} + b)$, where $\vec{w}_i$ is a vector of *weights* - or *parameters* - for each corresponding $\vec{x}_{i-1}$ component. b is a bias term and ϕ is an *activation function*. Note that b is usually included as a component in $\vec{w}_{ij}$, with a matching component with value 1 in $\vec{x}_{i-1}$ The activation function modulates the neuron's response, and is typically a sigmoid-like function imposing an effective threshold above or below which the neuron's output is negligible. Neural networks are able to identify patterns in data which other modelling techniques will miss.

Neural network classifiers, like the ones employed in this report, apply a softmax function $\sigma(\vec{z})_i$ to the final layer of outputs $\vec{z}$,

$$\mu_i \equiv \sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}},$$

where K is the number of classes (the cosmological theories in our case). These values sum to one, so each represents an effective 'confidence' μ_i that the example belongs to

the ith class. An example is classified into the ith class if $\mu_i > \mu_j \ \forall i \neq j$, and μ_i is greater than some threshold confidence P_{th}, which is chosen to be 0.5 [Mancarella *et al.,* Jan. 2022]. If $\mu_i < P_{th}$, the example is considered "Not Classified."

3.3.2 Bayesian Neural Networks

A drawback of standard neural networks is that adding a slight perturbation to the input vector can drastically alter the output classification [Szegedy *et al.,* Feb. 2014]. Bayesian Neural Networks (BNNs) alleviate this issue by allowing for a probabilistic treatment of network weights, modeling them as random variables with probabilities $P(\vec{w}, \mathcal{D})$ given by a posterior distribution $q(\vec{w})$, where $\mathcal{D}$ is the observed data. The final output of the network $\vec{\mu}$ is then also a dependent on the sampled weights for each test. This acts as an effective *regularisation* process, preventing overfitting by encouraging the model to consider a range of weight values rather than fixating on a specific set that fits the training data too closely [Blundell *et al.,* May 2015]. The μ_i values cannot be interpreted directly as probabilities; they change each time the same example is passed through the network. Instead, the `BaCoN` authors propose a novel approach to deriving *more* representative probability scores from the outputs [Mancarella *et al.,* Jan. 2022].

3.3.3 Convolutional Neural Networks

Another drawback of typical neural networks is that they are generally agnostic to spatial hierarchies of features, so they struggle to identify patterns if they occur at different positions (k values) in the data. Convolutional Neural Networks (CNNs) bypass this limitation by applying convolutions on input images through *convolutional layers* and dimensionality reduction through *pooling layers*, before passing the outputs to the conventional neural network perceptron layers described before [Kiranyaz *et al.,* Apr. 2021].

Convolutional layers apply a set of filters, or *kernels*, to the input. Each kernel slides across the input data, performing element-wise multiplication followed by a sum (convolution), generating a feature map. The kernel moves a specified number of k values along in each sliding step, known as the kernel's *stride*. This process captures the local dependencies within the input. Given a one-dimensional input of length L and a kernel of size K, the resulting feature map from each kernel will have a dimensionality of $L - K + 1$, assuming a stride of 1. We can create multiple kernels in each convolutional layer each targeting a specific aspect or pattern within the data.

Pooling layers are applied to reduce the dimensionality of the data, summarizing the features in segments of the feature map. It does this by applying a specified operation, such as averaging or taking the maximum value, on each *pool* in the data. If the input sequence length is L and the pooling operation is applied with a pool size of n and a stride of n, the output sequence length after pooling would be L/n, effectively summarizing the most salient features in each pool. The output of the pooling layer serves as the input to the next convolutional layer or a standard perceptron layer. The final layers of the network flatten the pooled features into a one-dimensional vector to perform classification or regression tasks based on the learned high-level features.

The convolution-pooling process is applied for four *channels* of data, corresponding to the four redshift bins studied by `BaCoN`.

3.3.4 Optimization

The optimization of weights in neural networks is generally achieved through gradient descent. This iterative process adjusts the network's weights to minimize a *loss function* $\mathcal{L}$, which quantifies the difference between the predicted outputs and the actual targets. The loss function used in `BaCoN` is the Evidence Lower Bound (ELBO), which is explained further by [Anand, Dec. 2023]. During each weight update, a step in the direction of $-\vec{\nabla}_w \mathcal{L}$ is taken. Do do this, a *backpropagation* algorithm is employed, which uses the chain rule of calculus to determine $\partial \mathcal{L} / \partial w_{ij}$ for each weight w_{ij}. The weights are then updated according to the equation $\vec{w}_{new} = \vec{w}_{old} - \eta \vec{\nabla}_w \mathcal{L}$ where η is the *learning rate* (note that here $\vec{w}_{old}$, $\vec{w}_{new}$ refer to vectors containing the weights of *all* perceptrons in the network). This process is repeated iteratively over multiple *epochs* (entire passes through the dataset), gradually improving the model's accuracy by refining the weights to minimize the loss function.

Using a learning rate which decays over each epoch helps in avoiding overshooting the minima of the loss function, thereby improving the convergence of the network training. Furthermore, decaying learning rates can help prevent *overfitting*, where the networks learn the specific patterns of the training data to well and then fail to generalise well to unseen data. The networks in this report use exponential learning rate decay.

An extension of standard gradient descent, known as the *Adam optimization algorithm*, is used in this paper. Adam, which stands for Adaptive Moment Estimation, intelligently adjusts the learning rate for each parameter, determining how much the parameters change in response to the estimated error each time the model weights are updated. This prevents converging to local minima in loss and facilitates faster convergence [Kingma & Ba, Jan. 2017].

In BNNs, as the weights are drawn from a probability distribution $w_i \sim N(\bar{w}_i, \vec{\sigma})$, $\vec{w}$ cannot be updated directly. Instead a reparameterisation trick is used, where a new random variable ϵ_i is sampled from a standard Gaussian, the parameter $\vec{\theta}$ is also sampled from a distribution and the weights are expressed through a deterministic function $w_i = g(\epsilon_i, \vec{\theta})$, allowing backpropagation [Kingma, Salimans, *et al.,* Dec. 2015]. A drawback of this approach is that the sampled weights are the same for each *batch* (see Section 3.3.6) of data passed through the network, correlating the resulting gradients and slowing the optimisation process [Thummel *et al.,* Mar. 2024]. To mitigate this effect, the `BaCoN` models use *flipout*, a method of generating pseudo-independent weight perturbations for each example in a batch by randomly multiplying perturbations by -1 or 1. This effectively decorrelates the gradients and enhances variance reduction during training [Wen *et al.,* Apr. 2018].

3.3.5 Normalisation

The optimisation of neural networks is far faster when the input data is normalised. Furthermore, normalisation can prevent unwanted effects caused by the differing scales of input features. In order to normalise the spectra, they are divided by a ΛCDM spectrum generated by `EuclidEmulator2` with parameters matching the Planck spectrum in [Planck Collaboration, Sept. 2020], and centred around 1.

3.3.6 Data Segmentation, Batching and Early Stopping

Before the raw data is fed to the `BaCoN` networks, it is segmented and batched. Early stopping is used during the training process.

Data segmentation refers to dividing the dataset into three distinct parts – the training set, used to train the model; the validation set, used to tune model parameters and the test set, used to assess the model's performance after training. After each weight update when training on the training set, the new weights are saved only if the loss on the validation set is less than the best validation loss observed so far. The test set is not used in the training process, and is only used to measure the networks' performance. This ensures that the model not only learns from the training data but also generalizes well to new, unseen data, thereby enhancing its predictive accuracy and robustness. The `BaCoN` data consists of an overall test-validation set of 20,000 examples per class, of which 15% is taken to be the validation set. A test set of 1000 examples per class is generated [Mancarella *et al.,* Jan. 2022].

Batching involves dividing the training data into groups of a specified size. This technique enables the model to update its weights after processing each batch, rather than after each individual example. This introduces a level of randomness in the training process, which can help in avoiding local minima and in speeding up optimisation. Batch sizes that are too low or too high can cause issues.

Early stopping is implemented in `BaCoN`, where if a specified number of epochs (the *patience*) passes without an improvement in the validation loss, training will stop. This is another measure to prevent overfitting.

3.3.7 Hyperparameters

For the specific *hyperparameters* (e.g perceptron number per layer, batch size, validation size, kernel size, stride, pooling size) used by the models in this report, as well as a pictorial depiction of their architecture, see [Mancarella *et al.,* Jan. 2022]. Furthermore, any modifications to these hyperparameters are described in Section 5 and the entire set used for each network is found in the log files associated them. These can be found under `BaCoN-II(i)/models` in the project GitHub repository, found in Section A.1.

4 Methods

4.1 Computational Improvements

Speeding up `BaCoN` models' execution time and reducing their computational cost would allow them to be trained on larger datasets of higher resolution faster. In the context of the vast, high resolution Stage–IV study data described in Section 2.5, this would greatly improve the scalability of the models. The computational improvements which I introduce in this section allow training on very large datasets of the highest resolution quickly.

The resolution read by the models is determined by the *sample pace* hyperparameter. Aside for reading more granular details from the data, a lower sample pace allows the range of k modes to be shortened to a range 4 times shorter without the using *padding*.

The CNN structure described in Section 3.3 requires a minimum number of k modes to prevent dimensionality issues in the convolutional layers. One could prevent this by padding the inputs with zeros, but this could result in the networks storing unwanted positional information of spectral structure [Islam *et al.*, Jan. 2021]. As such, having a higher resolution of k modes prevents such errors from arising when cutting out length scales, as a larger input dimensionality is mantained for a shorter scale range. This is crucial for my investigation of length scales in Section 5.5.

All of the models I present have been trained on the A100 GPU provided on [Google Colab]. The Colab notebook needed to run the training, with the necessary steps explained, can be found in the project GitHub repository under `BaCoN-II(i)/training_colab.ipynb`.

4.1.1 Tensorflow Data Management

The most significant computational improvement I have made is the use of the `tf.data.Dataset` Python class from the Tensorflow library, which the `BaCoN` models are built upon. This optimises data preprocessing, storage and loading [Tensorflow Guide, Sept. 2023a]. Using this class is the recommended approach to feeding data to Tensorflow models.

The class handles data preprocessing through its `.map` method, which allows computations to be done in parallel across Graphics Processing Unit (GPU) cores. Furthermore, `.map` allows preprocessing to be done with Tensorflow's graph execution system, which offers performance improvements by minimising repeated computations [Tensorflow Guide, Jan. 2024]. I apply this method for the normalisation process described in Section 3.3.5, after the input spectra have been loaded and the noise realisations created.

Furthermore, the *prefetching* method offered by the class allows data for each training step to be prepared and loaded to memory before the previous step is complete, preventing data loading to become a bottleneck for execution time. The class allows us to *cache* the data, where the processed data is stored in the device's memory or on its disk. This reduces I/O time - time associated with reading files - and prevents the unnecessary repetition of preprocessing steps.

4.1.2 XLA Compiler

Another optimisation I have implemented is forcing `BaCoN` models to use of the XLA Compiler for Tensorflow. XLA translates high-level Tensorflow operations into code for the LLVM compiler, which works more closely with hardware [OpenXLA Developer Guide, Jan. 2024]. Furthermore, XLA employs *kernel fusion* – merging multiple operations into a single efficient computational *kernel*. This should not be confused with the 'kernel' discussed above in the context of CNNs, instead this refers to a low-level program on a device's operating system responsible for executing commands. XLA allows composite calculations to be completed by a single Kernel, rather than requiring the initialisation, execution and consolidation of results from multiple kernels [Tensorflow Blog, Nov. 2018; Snider & Liang, Jan. 2023]. Kernel fusion significantly reduces computational overhead and enhances the execution speed of TensorFlow models [Snider & Liang, Jan. 2023].

4.2 Updated Random Spectra

In [Mancarella *et al.,* Jan. 2022], the authors generated the class of random spectra by applying a set of filters (see Section 3) to a set of randomly selected base spectra from all of the other theories. I argue that this is a superior approach to that in [Thummel *et al.,* Mar. 2024], which only uses ΛCDM base spectra. This is because the non-ΛCDM spectra serve as potential sources of new physics which may be present in the real cosmology. Including them could allow the random spectra to encode the fingerprints of beyond-General Relativity effects more effectively. If the goal is to identify any unclassifiable spectra which represent physics beyond any of the other theories included in the model, the random class should be created to represent deviations from any one of these theories, not just ΛCDM. As a slight modification to the original approach, instead of randomly selecting spectra from the full set of theories, I take an equal number from each. This may be a better approach as each source of potential physics is then equally represented in the random spectra. Before using this method in my models, I determine if it causes a drop in accuracy in Section 5.2.

4.3 Theoretical Tests

The theoretical tests that I present here are designed to aid our understanding of how varying underlying physical parameters affects `BaCoN`'s performance. Specifically, I consider model performance with various physical length scales and Dark Scattering theory modification strengths.

I began by including data for the Dark Scattering (DS) model into `BaCoN`'s corpus. After assessing the initial impact on accuracy of including DS as a separate class, I split the DS data into four bins of equal size, increasing in χ. At low χ, I hypothesised that the data would be similar to that of wCDM spectra, resulting in higher confusion from the networks between wCDM and DS. The aim here was to find a minimum modification strength at which the model confidently classifies DS data as distinct from wCDM, and therefore the χ value above which this data should be treated as a separate class.

In order to understand the impact of linear scale data on the six-label models' classifying ability, I implemented a k_{min}. Due to the similarity at linear scales between spectra for different theories, specifically $f(R)$ & ΛCDM and DS & wCDM, I hypothesised that cutting out k modes below the transition from cosmic variance dominated noise to theoretical error dominated noise would not reduce (or perhaps improve) the networks' performance. This was based upon the idea that removing the scales with the most overlap would allow the models to focus on the scales where clear distinctions can be made.

5 Results & Discussion

In this section I will compare the training time, accuracy and confusion of the networks after the changes described in the previous section. To facilitate easier identification of the models' files and training data, linked in Section A.1, I number the models with roman numerals from **I** to **XIII**.

In order to visualise confusion, we utilise *confusion matrices*, which display the percentage of 'true' labels classified to belong in each 'predicted' class by the model. In these

figures, "N.C" refers to examples where the output confidence $\mu_i < 0.5 \; \forall i$, so were not classified to any class i. In a perfect model, the confusion matrices will be completely diagonal (excluding N.C rows and columns). Confusion matrices not shown in this section can be found on the project GitHub repository under BaCoN-II(i)/models.

Accuracy here refers to the standard metric, defined as the number of correctly classified examples divided by the total number of examples across all classes. I will also mention the metrics *precision* and *recall*. The precision of a multi-class classifier, for a specific class, refers to the ratio of examples correctly classified as that class to the total number classified to belong to that class. The recall for a class refers to the ratio of examples that were classified as that class to the total number actually belonging to the class in the dataset. The numbers along the diagonals of the confusion matrix (excluding N.C) are recall values.

5.1   New Data Pipeline

Model **I** is a model trained with the exact same hyperparameters as the final model presented in [Thummel *et al.*, Mar. 2024], but with the new data pipeline described in Section 4.1. This resulted in a total test accuracy of 89.1%, compared to the 95.1% in [Thummel *et al.*, Mar. 2024]. The confusion matrix for model **I** is found in Figure 2a. This was a strange result, as the data processing steps are identical in terms of the actual mathematical steps executed, only the computational procedure has been changed. I verified that the spectra being fed to the models is of the correct form by plotting randomly selected processed spectra. One of these is displayed in Section A.2 for all four redshift bins, more can be seen and generated through the files found in the folder Processed Spectra of the GitHub repository. These plots indicate that the spectra being fed to the model are processed in the same way as in the process outlined in Figure 8 of [Thummel *et al.*, Mar. 2024].
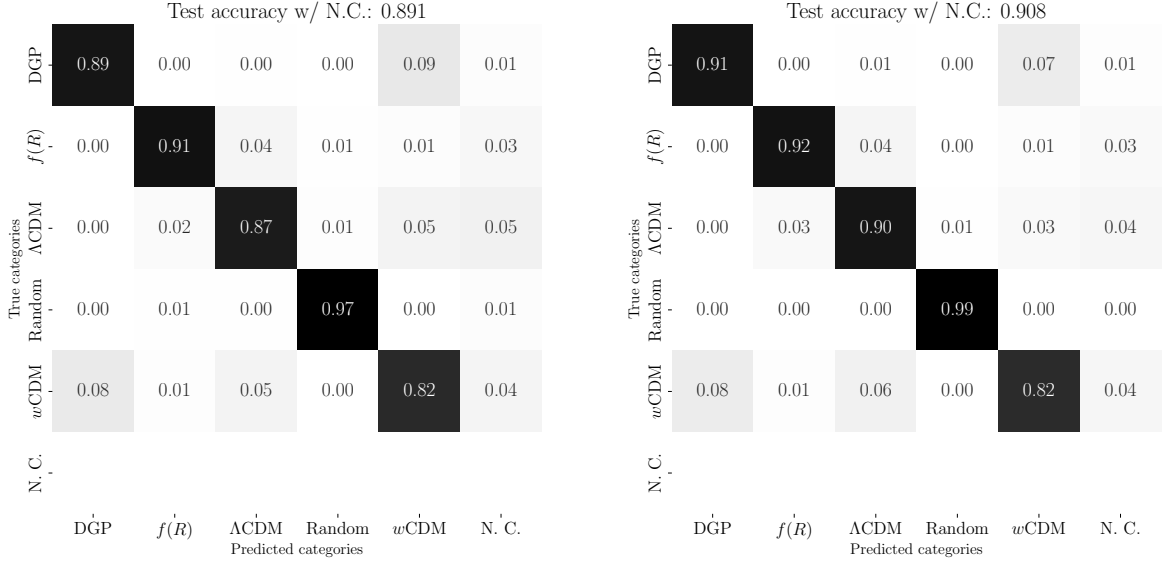
(a) Model **I**: Hyperparameters and classes as in [Thummel *et al.*, Mar. 2024].

(b) Model **II**: Same classes; tweaked batch size and initial learning rate.

Figure 2: Comparison of models **I** and **II**.

This strange result may indicate a change in the training dynamics due to the data feeding process. Noting the graphs in [Tensorflow Guide, Sept. 2023a] pertaining to how the data is fed to the models, this result may be due to how Tensorflow may be handling the order and stage at which the Mapping, Batching, Caching and Prefetching operations are applied. If this is the case, tweaking training hyperparameters to accommodate the new setup would likely increase accuracy. To this end, I reduced the batch size from 12500 to 2000 in model **II**, to allow more granular weight updates. I also increased the starting learning rate from 0.01 to 0.05 and increased the number of epochs from 50 to 100, allowing the model to make bigger weight changes at later epochs and have more epochs to train on. As shown in Figure 2b, the test accuracy of model **II** was indeed slightly higher than that of model **I**, up to 90.8%. The responsiveness of accuracy to changes in hyperparameters leads me to conclude that the performance decline could be resolved through a proper hyperparameter optimisation procedure. However, closer inspection of the new data pipeline may be worthwhile. I discuss this further in Section 6.

The training of models **I** and **II** was significantly faster than previous versions of `BaCoN`. The final model in [Thummel *et al.*, Mar. 2024] took nearly three days to train, whereas model **II** took under an hour, whilst still using all 20,000 available spectra per class and the same resolution. These times include data loading, data preprocessing, model initialisation and training. It is unclear what computational equipment the former was trained using, making this direct comparison somewhat unfair. Regardless, on most computers, the `BaCoN` models previously took days to train, as mentioned on the GitHub repository for [Thummel *et al.*, Mar. 2024], whereas this implementation takes a matter of hours. The bulk of the execution time is during data loading, data processing and model initialisation phases, with the training stage only accounting for 3 out of 55 minutes. I

suggest a potential method of further reducing execution time in Section 6.

Noting this dramatic training speed improvement, that the drop in accuracy is likely a hyperparameter issue and that the accuracy of the models using the new data pipeline is still $\sim 90\%$ for the original classes, I continued using the new pipeline in the remaining models I investigated. I also continued to use the increased initial learning rate in models **III** and **IV**.

5.2 New Random Class

In model **III**, I replaced the random class with spectra generated with the process described in Section 4.1. Although this model presented a slight drop in accuracy of 1% and a 3% drop in recall for the random class, the model still performed remarkably well. This demonstrates that the method I suggest does not significantly affect the network's ability to identify random spectra. Given that these spectra may embed information on unseen new physics more effectively, and do not cause a significant drop in accuracy, I continued to use this random class prescription in the rest of the models I trained.

Figure 3: Model **III**: Modified random class prescription.

5.3 Dark Scattering as a Sixth Class

For model **IV**, I added the data for the Dark Scattering theory as an additional class to the networks, transforming them from five-label to six-label classifiers. I increased the batch size to 3000 for this model, to accommodate the additional class. The confusion matrix for this model is shown in Figure 4.

Figure 4: Model **IV**: Initial six-label model including Dark Scattering class.

A significant drop in overall accuracy is noted, down to 74.2%, a drop of 15.6% from model **III**. The main cause of this seems to be the high classification degeneracy between DS and wCDM. Due to examples from both classes being incorrectly classified as the other, the precision and recall for both has fallen. I mentioned [Simpson, Oct. 2010]'s prediction that these two classes' spectra will be very similar in Section 2.3.4. Indeed, the Dark Scattering model is simply an extension of the dynamical dark energy framework of wCDM. As [Thummel *et al.*, Mar. 2024] mention, such physical degeneracy can result in a high clasification degeneracy from the networks.

Furthermore, a significant rise in examples from every class that were unclassifiable (looking at the final column of Figure 4) is noted. It seems that the inclusion of the DS data has reduced the ability of the network to distinguish between classes in general.

5.4 Increased Resolution

Given the greatly improved training speeds with the new data pipeline, it was practical to train models with a higher resolution – a sample pace of 1 rather than 4. Continuing on from model **III**, I trained a model with six classes and a sample pace of 1. Initially, the training and validation accuracy was non-convergent and sporadic over epochs after this change. Suspecting that this was a hyperparameter issue, I increased the batch size fourfold to 12000 in order to account for the quadrupling of the data size, and reverted to the initial learning rate in model **I** of 0.01. The reduction of the learning rate was to adjust for the noisiness in accuracy initially observed over later epochs. Furthermore, I increased the number of epochs from 50 to 100 to allow a longer time for convergence, whilst retaining a patience of 20. This adjusted model, model **V**, had an accuracy only $\sim 0.3\%$ less than model **IV**. Given the fine-grained data expected from Stage–IV studies, as discussed in Section 2.5, and the lack of a significant accuracy drop from using the higher resolution, I trained the remaining models with this lower sample pace.

5.5   Accuracy over Length Scales

Noting [Simpson, Oct. 2010]'s prediction of background, linear length scales being those at which the biggest degeneracy between DS and wCDM would exist, I trained models **VI**, **VII** and **VIII** with increasingly higher k_{min} modes – $0.15, 0.4$ and $1.0h\text{MPc}^{-1}$ respectively. Figure 5 shows the evolution of testing accuracy over each $k_{\min}$ value, including model **V** for $k = 0h\text{MPc}^{-1}$. The figure includes a quadratic fit for visualisation of the trend only.
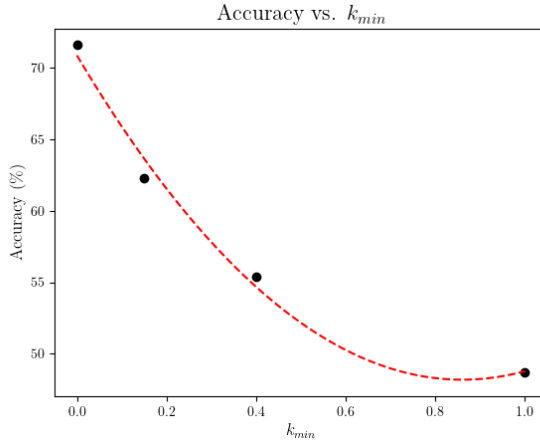
Figure 5: Models **V-VIII**: Accuracy of models with increasing $k_{\min}$. Circles are actual values, red line is quadratic fit for visualisation.

The figure shows that accuracy decreases as k is increased, and most dramatically so as the most linear scales are cut out. It is logical that as k values are cut out, the accuracy will decrease, as the model has less information available to make a classification. However, given the prediction by [Simpson, Oct. 2010], it is surprising that the steepest decline occurs as the longest length scales are cut. This does not seem to arise from an increase in misclassifications between theories, but rather an increase in unclassified spectra. Figure 6 shows the confusion matrix for model **VIII**, displaying this effect. The column for N.C has far higher scores for each true class.

Figure 6: Model **VIII**: Highest $k_{\min}$ tested, $1.0h\mathrm{MPc}^{-1}$.

The model appears to make distinctions between theories on linear scales and struggles to do so on non-linear scales. To explain this effect, I refer back to Section 3.2, specifically the observed transition from length scales with noise dominated by cosmic variance, to those dominated by theoretical error.

Although the theoretical error model may effectively represent the error between the generation processes, it may be that it results in the loss of fine power spectrum structure on non-linear and intermediate scales. This would explain why accuracy is lower when theories with strong degeneracies on linear scales are included, such as wCDM and DS. I discuss this further in Section 6.

Given that cutting out k values did not result in an improvement in accuracy, I continued with no $k_{\min}$ in the remaining models I investigated.

5.6 Binning of Dark Scattering Modification Strength

As an alternative approach to lifting the degeneracy between DS and wCDM, I trained models **IX**-**XII** with only a subset of the dark scattering data. In order to ensure that an equal number of examples were present for each class (which can cause issues in machine learning models when not handled correctly [Google for Developers, June 2023]), this required reducing the number of examples in all classes from 20000 to 5000, and the batch size back down to 3000. The DS data was split into bins A-D, with $\chi \in [7.85 \times 10^{-4}, 5.16]; [5.16, 10.50]; [10.50, 17.47]$ and $[17.47, 30.00]$ bGeV^{-1} respectively. The training accuracy increased from bin A (66.4%) to bin D (71.5%), however the overall confusion did not improve dramatically, as shown by model **XII**'s confusion matrix in Figure 7. Furthermore, the increase in accuracy was not monotonic across the bins; there was a slight drop in accuracy from bin B (69.7%) to bin C (69.0%). This indicates overall that the networks were not responsive to the modification strengths.

Figure 7: Model **XII**: Using only bin D of Dark Scattering data ($17.47 \leq \xi \leq 30.00$ bGeV^{-1}).

5.7 New Five-Label Model

Noting that there was no significant reduction in confusion from binning the Dark Scattering data, I trained a final five-label model (model **XIII**) where I replaced the wCDM spectra with DS spectra. Again, this was motivated by [Simpson, Oct. 2010]'s finding that these theories would be highly physically degenerate. The confusion matrix is shown in Figure 8. For this model, I used the full set of 20,000 spectra per class and a batch size of 10000 to support it.

Figure 8: Model **XIII**: Five-label model with wCDM spectra replaced by Dark Scattering spectra.

Here, confusion is drastically reduced compared to six-label models including both DS and wCDM. Accuracy rose to 87.3%, 13.1% more than model **IV**. The recall for DS specifically rose from 41% to a far more reasonable 75%. The DS class has a similar precision and recall ($\sim 5\%$ lower) to the wCDM class in model **III**, as shown in Figure 3. The DS data in this new five-label model demonstrates the same degeneracies as the old five-label model, with DS showing degeneracies with DGP and ΛCDM, just like wCDM did. This serves as further evidence of the substitutability of the theories.

This is a promising result, highlighting that the inclusion of Dark Scattering into the BaCoN corpus yields better results when it replaces wCDM spectra. However, the overall accuracy and the recall of the DS class is still lower than the original five-model **II**. Potential solutions are discussed further in Section 6.

6 Future Work and Conclusions

In this work I have reviewed the state of current research in cosmology (Section 2), reviewed the computational background behind `BaCoN` and its potential solution to the cosmological model selection problem (Section 3), presented a refactoring of `BaCoN`'s data generation process (Sections 4.1 and 5.1), proposed a new random class generation process (Sections 4.2 and 5.2) and assessed the model's performance with the inclusion of the Dark Scattering theory, accounting for underlying physical parameters (Sections 4.3, 5.5, 5.6, 5.7).

The new data generation process has demonstrated an improvement in training time from a matter of days to $1 - 2$ hours. This has made leveraging datasets with a large number of examples and a high resolution more feasible, such as those expected from Stage–IV studies. I was able to quickly train models with all 20,000 available examples

per class, sampling all k modes. However, when training models with the exact same data and hyperparameters as the previous process, this resulted in an accuracy drop of 6%. Since the data output from the data generation process is the same in the old methods and new methods, as evidenced by the processed spectra available in Section A.2 and on the project GitHub, this is an unusual result. Despite consistent hyperparameters, the new pipeline appears to have altered the training dynamics of the models. It would therefore be worthwhile to investigate the cause of this by using tool that visualises the data feeding process, such as the Tensorflow Profiler [Tensorflow Guide, Nov. 2022]. It is likely that the drop in accuracy can be resolved by employing a hyperparameter optimisation process such as K-fold Cross Validation, or even using advanced tools for this, such as Weights & Biases [Weights & Biases Docs, 2024] or Tensorboard [Tensorboard Guide, Jan. 2022].

The training time can be further reduced by implementing Tensorflow-friendly file formats [Tensorflow Guide, Sept. 2023b] for the raw input data, given that my results show that the majority of the execution time is involved in initial data reading. Furthermore, future `BaCoN` iterations can use Tensor Processing Units, specialised machine learning processors [Google Cloud Documentation, Sept. 2023], whose support I have largely implemented already.

I found that using my proposed method for generating random spectra resulted in no significant change in model accuracy or random class recall, whilst potentially being more physically representative. I will note that, given that this class is designed to represent unseen physical models, a 'correct' prescription for its generation may remain conjecture in the near future.

Upon adding the Dark Scattering theory as an additional class to the networks, I noted a significant drop in accuracy (of $\sim 15\%$), as well as a drop in precision and recall for the Dark Scattering and wCDM classes. This was not an altogether unexpected result, noting that Dark Scattering is an extension of wCDM, and that these theories were expected to be physically degenerate [Simpson, Oct. 2010]. Given that this prediction of degeneracy was primarily on the linear scale, I investigated the effect of cutting out lower k modes from the data. Surprisingly, the removal of the lowest k values resulted in the steepest drop in accuracy, with increasingly more examples becoming unclassifiable. I hypothesise that this is due to the current data generation process used for `BaCoN`. My results suggest that the models extract more relevant information from linear scales, where cosmic variance dominates the noise mdoel, than non-linear scales, where theoretical error dominates. I suggest that although the theoretical model used by [Thummel *et al.*, Mar. 2024] allows for the reconciliation of results when testing with spectra generated by different prescriptions, the theoretical error model is resulting in a loss of fine structure on non-linear and intermediate scales. As [Thummel *et al.*, Mar. 2024] state themselves, there are more improvements yet to be made to the data generation process. Revisiting the theoretical error models and studying again the impact of length scales is a worthwhile area of future research.

Furthermore, it was found that binning the dark scattering data by modification strength, quantified by the mass-reduced scattering cross section ξ, did not result in a significant change in classification performance. The models trained on the bin with the highest modification strengths was only 5.1% more accurate than those trained on the bin with the lowest; confusion was not significantly better either. For this reason, I decided to treat the DS class as a replacement for the original wCDM class in a five-label

model. The performance of this model was remarkably better, but was still worse than the original five-label model. This result may improve as the data generation process is refined. Alternatively, noting that the DS data generated here are sampled from a distribution where the mean is the ΛCDM limit, but the actual limit is not included [Thummel *et al.*, Mar. 2024], it may be worthwhile to create a combined class with both wCDM spectra and DS spectra, and then constrain this further if real data suggests that this new combined class is representative of observations. This could be done with a refined MCMC analysis, or through the use of specialised networks, as proposed in [Mancarella *et al.*, Jan. 2022].

Considering these results as a whole, the importance of implementing explainable machine learning methods, as proposed in [Thummel *et al.*, Mar. 2024], is underscored. These would allow us to determine which class-specific features the networks are identifying, and enable us to more precisely determine the impact of underlying physical assumptions and parameters.

This investigation has served as a source of possible improvements for `BaCoN` aimed at increasing its scalability and physical representativeness, as well as a probe into the impact of the underlying physics on the networks. This is a further step in refining this powerful new approach to problem facing modern cosmology. `BaCoN` models, combined with upcoming Stage–IV survey data, may prove to be a key component in developing our understanding of the structure of the Universe.

7 References

References

1. Akeson, R. *et al.* The Wide Field Infrared Survey Telescope: 100 Hubbles for the 2020s. *ArXiv.* 2019. doi:10.48550/arXiv.1902.05569.

2. Albrecht, A. *et al.* Report of the Dark Energy Task Force. *ArXiv.* 2006. doi:10.48550/arXiv.astro-ph/0609591.

3. Amendola, L. *et al.* Cosmology and Fundamental Physics with the Euclid Satellite. *Living Reviews in Relativity.* **16.** 6. 2013. doi:10.12942/lrr-2013-6.

4. Anand, V. J. Evidence Lower Bound. *Medium.* 2023. https://medium.com/@vjanand/evidence-lower-bound-3f7382e90c81.

5. Bernardeau, F., Colombi, S., Gaztañaga, E. & Scoccimarro, R. Large-scale structure of the Universe and cosmological perturbation theory. *Physics Reports.* **367.** 1–3. 1–248. 2002. doi:10.1016/s0370-1573(02)00135-7.

6. Beucler, T. *et al.* Enforcing Analytic Constraints in Neural Networks Emulating Physical Systems. *Physical Review Letters.* **126.** 9. 2021. doi:10.1103/PhysRevLett.126.098302.

7. Blanchard, A. *et al.* Euclid preparation: VII. Forecast validation for Euclid cosmological probes. *Astronomy & Astrophysics.* **642.** 66. 2020. doi:10.1051/0004-6361/202038071.

8. Blundell, C., Cornebise, J., Kavukcuoglu, K. & Wierstra, D. Weight Uncertainty in Neural Networks. *ArXiv.* 2015. doi:`10.48550/arXiv.1505.05424`.

9. Bose, B., Cataneo, M., *et al.* On the road to percent accuracy IV: ReACT – computing the non-linear power spectrum beyond ΛCDM. *Monthly Notices of the Royal Astronomical Society.* **498.** 4. 4650–4662. 2020. doi:`10.1093/mnras/staa2696`.

10. Bose, B., Wright, B. S., *et al.* On the road to per cent accuracy – V. The non-linear power spectrum beyond ΛCDM with massive neutrinos and baryonic feedback. *Monthly Notices of the Royal Astronomical Society.* **508.** 2. 2479–2491. 2021. doi:`10.1093/mnras/stab2731`.

11. Carrilho, P. *et al.* On the road to percent accuracy VI: the non-linear power spectrum for interacting dark energy with baryonic feedback and massive neutrinos. *Monthly Notices of the Royal Astronomical Society.* **512.** 3. 3691–3702. 2022. doi:`10.1093/mnras/stac641`.

12. Cataneo, M. *et al.* On the road to percent accuracy: non-linear reaction of the matter power spectrum to dark energy and modified gravity. *Monthly Notices of the Royal Astronomical Society.* **488.** 2. 2121–2142. 2019. doi:`10.1093/mnras/stz1836`.

13. Chevalier, M. & Polarski, D. Accelerating Universes with Scaling Dark Matter. *International Journal of Modern Physics D.* **10.** 2. 213–224. 2001. doi:`10.1142/S0218271801000822`.

14. Cichy, R. M. & Kaiser, D. Deep Neural Networks as Scientific Models. *Trends in Cognitive Sciences.* **23.** 4. 305–317. 2019. doi:`10.1016/j.tics.2019.01.009`.

15. Cohen-Tannoudji, G. Lambda, the fifth foundational constant considered by Einstein. *Metrologia.* **55.** 4. 486–498. 2018. doi:`10.1088/1681-7575/aac10b`.

16. Copeland, E. J., Sami, M. & Tsujikawa, S. Dynamics of dark energy. *International Journal of Modern Physics D.* **15.** 11. 1753–1935. 2006. doi:`10.1142/S021827180600942X`.

17. De Simone, A. Introduction to cosmology and dark matter. *Proceedings of the European School for High Energy Physics.* **6.** 2019. 2018. doi:`10.23730/CYRSP-2019-006.145`.

18. De Blok, W. J. G. The Core-Cusp Problem. *Advances in Astronomy.* **2010.** 1–14. 2009. doi:`10.1155/2010/789293`.

19. Dodelson, S. & Knox, L. Dark Energy and the Cosmic Microwave Background Radiation. *Physical review letters.* **84.** 16. 3523–3526. 2000. doi:`10.1103/PhysRevLett.84.3523`.

20. Dvali, G., Gabadadze, G. & Porrati, M. 4D gravity on a brane in 5D Minkowski space. *Physics Letters B.* **485.** 1–3. 208–214. 2000. doi:`10.1016/s0370-2693(00)00669-9`.

21. Euclid Collaboration. Euclid preparation: IX. EuclidEmulator2 – power spectrum emulation with massive neutrinos and self-consistent dark energy perturbations. *Monthly Notices of the Royal Astronomical Society.* **505.** 2. 2840–2869. 2021. doi:`10.1093/mnras/stab1366`.

22. Google Cloud Documentation. Introduction to Cloud TPU. 2023. `https://cloud.google.com/tpu/docs/intro-to-tpu`.

23. Google Colab. https://colab.research.google.com/.

24. Google for Developers. Imbalanced Data. *Machine Learning.* 2023. https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data.

25. Heitmann, K. *et al.* The Mira-Titan Universe: Precision Predictions for Dark Energy Surveys. *The Astrophysical Journal.* **820.** 2. 108. 2016. doi:10.3847/0004-637X/820/2/108.

26. Hu, W. & Sawicki, I. Models of $f(R)$ cosmic acceleration that evade solar system tests. *Phys. Rev. D.* **76.** 6. 064004. 2007. doi:10.1103/PhysRevD.76.064004.

27. Islam, M. A., Kowal, M., Jia, S., Derpanis, K. & Bruce, N. D. B. Position, Padding and Predictions: A Deeper Look at Position Information in CNNs. *ArXiv.* 2021. doi:10.48550/arXiv.2101.12322.

28. Kennedy, J., Lombriser, L. & Taylor, A. Screening and degenerate kinetic self-acceleration from the nonlinear freedom of reconstructed Horndeski theories. *Phys. Rev. D.* **100.** 4. 2019. doi:10.1103/PhysRevD.100.044034.

29. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv.* 2017. doi:10.48550/arXiv.1412.6980.

30. Kingma, D. P., Salimans, T. & Welling, M. Variational Dropout and the Local Reparameterization Trick. *ArXiv.* 2015. doi:10.48550/arXiv.1506.02557.

31. Kiranyaz, S. *et al.* 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing.* **151.** 2021. doi:https://doi.org/10.1016/j.ymssp.2020.107398.

32. Klypin, A., Kravtsov, A. V., Valenzuela, O. & Prada, F. Where Are the Missing Galactic Satellites? *The Astrophysical Journal.* **522.** 1. 82–92. 1999. doi:10.1086/307643.

33. Kunz, M. & Sapone, D. Dark energy versus modified gravity. *Physical Review Letters.* **98.** 12. 2007. doi:10.1103/PhysRevLett.98.121301.

34. Laureijs, R. *et al.* Euclid Definition Study Report. *ArXiv.* 2011. doi:10.48550/arXiv.1110.3193.

35. Lesgourgues, J. The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. *ArXiv.* 2011. doi:10.48550/arXiv.1104.2932.

36. Levi, M. E. *et al.* The Dark Energy Spectroscopic Instrument (DESI). *ArXiv.* 2019. doi:https://doi.org/10.48550/arXiv.1907.10688.

37. Linder, E. V. Exploring the expansion history of the universe. *Physical Review Letters.* **90.** 9. 2003. doi:10.1103/PhysRevLett.90.091301.

38. LSST Dark Energy Science Collaboration. Large Synoptic Survey Telescope: Dark Energy Science n. *ArXiv.* 2012. doi:10.48550/arXiv.1211.0310.

39. Lue, A. & Starkman, G. Gravitational leakage into extra dimensions: Probing dark energy using local gravity. *Physical Review D.* **67.** 6. 2002. doi:10.1103/PhysRevD.67.064002.

40. Luty, M., Porrati, M. & Rattazzi, R. Strong Interactions and Stability in the DGP Model. *Journal of High Energy Physics.* **2003.** 09. 029. 2003. doi:10.1088/1126-6708/2003/09/029.

41. Maartens, R. & Majerotto, E. Observational constraints on self-accelerating cosmology. *Physical Review D.* **74.** 2. 2006. doi:10.1103/PhysRevD.74.023004.

42. Mancarella, M., Kennedy, J., Bose, B. & Lombriser, L. Seeking new physics in cosmology with Bayesian neural networks: Dark energy and modified gravity. *Physical Review D.* **105.** 2. 2022. doi:10.1103/PhysRevD.105.023531.

43. Mead, A. J., Brieden, S., Tröster, T. & Heymans, C. HMCode2020: improved modelling of non-linear cosmological power spectra with baryonic feedback. *Monthly Notices of the Royal Astronomical Society.* **502.** 1. 1401–1422. 2021. doi:10.1093/mnras/stab082.

44. Mead, A. J., Heymans, C., *et al.* Accurate Halo-model Matter Power Spectra with Dark Energy, Massive Neutrinos and Modified Gravitational Forces. *Monthly Notices of the Royal Astronomical Society.* **459.** 2. 1468–1488. 2016. doi:10.1093/mnras/stw681.

45. OpenXLA Developer Guide. XLA Architecture. 2024. https://openxla.org/xla/architecture.

46. Padilla, L. E., Tellez, L. O., Escamilla, L. A. & Vazquez, J. A. Cosmological Parameter Inference with Bayesian Statistics. *Universe.* **7.** 7. 213. 2021. doi:10.3390/universe7070213.

47. Peebles, P. & Ratra, B. The Cosmological Constant and Dark Energy. *Reviews of Modern Physics.* **75.** 2. 559–606. 2003. doi:10.1103/RevModPhys.75.559.

48. Persic, M., Salucci, P. & Stel, F. The Universal Rotation Curve of Spiral Galaxies: I. the Dark Matter Connection. *Monthly Notices of the Royal Astronomical Society.* **281.** 1. 27–47. 1996. doi:10.1093/mnras/278.1.27.

49. Planck Collaboration. Planck 2018 results - VI. Cosmological parameters. *Astronomy & Astrophysics.* **641.** A6. 2020. doi:10.1051/0004-6361/201833910.

50. Schneider, A. *et al.* Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation. *Journal of Cosmology and Astroparticle Physics.* **2019.** 03. 020–020. 2019. doi:10.1088/1475-7516/2019/03/020.

51. Semboloni, E., Hoekstra, H., Schaye, J., van Daalen, M. P. & McCarthy, I. G. Quantifying the effect of baryon physics on weak lensing tomography: Baryon physics and weak lensing tomography. *Monthly Notices of the Royal Astronomical Society.* **417.** 3. 2020–2035. 2011. doi:10.1111/j.1365-2966.2011.19385.x.

52. Simpson, F. Scattering of dark matter and dark energy. *Physical Review D.* **82.** 8. 2010. doi:10.1103/physrevd.82.083505.

53. Snider, D. & Liang, R. Operator Fusion in XLA: Analysis and Evaluation. *ArXiv.* 2023. doi:10.48550/arXiv.2301.13062.

54. Sotiriou, T. P. & Faraoni, V. f(R) theories of gravity. *Reviews of Modern Physics.* **82.** 1. 451–497. 2010. doi:10.1103/RevModPhys.82.451.

55. Szegedy, C. *et al.* Intriguing properties of neural networks. *ArXiv.* 2014. doi:10.48550/arXiv.1312.6199.

56. Tensorboard Guide. Hyperparameter Tuning with the HParams Dashboard. 2022. https://www.tensorflow.org/tensorboard/hyperparameter_tuning_with_hparams.

57. Tensorflow Blog. Pushing the Limits of GPU Performance With XLA. 2018. https://blog.tensorflow.org/2018/11/pushing-limits-of-gpu-performance-with-xla.html.

58. Tensorflow Guide. Analyze tf.data performance with the TF Profiler. 2022. https://www.tensorflow.org/guide/data_performance_analysis.

59. Tensorflow Guide. Better performance with the tf.data API. 2023. https://www.tensorflow.org/guide/data_performance.

60. Tensorflow Guide. Intro to Graphs. 2024. https://www.tensorflow.org/guide/intro_to_graphs.

61. Tensorflow Guide. TFRecord and tf.train.Example. 2023. https://www.tensorflow.org/tutorials/load_data/tfrecord.

62. Thummel, L., Bose, B., Pourtsidou, A. & Lombriser, L. Beyond-ΛCDM Cosmologies with Bayesian Neural Networks II: Massive Neutrinos, Baryonic Feedback, and the Theoretical Error. *ArXiv.* 2024. doi:10.48550/arXiv.2403.16949.

63. Trotta, R. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics.* **49.** 2. 104–71. 2008. doi:10.1080/00107510802066753.

64. Van Daalen, M. P., McCarthy, I. G. & Schaye, J. Exploring the effects of galaxy formation on matter clustering through a library of simulation power spectra. *Monthly Notices of the Royal Astronomical Society.* **491.** 2. 2424–2446. 2019. doi:10.1093/mnras/stz3199.

65. Weights & Biases Docs. Tune Hyperparameters. 2024. https://docs.wandb.ai/guides/sweeps.

66. Wen, Y., Vicol, P., Ba, J., Tran, D. & Grosse, R. Flipout: Efficient Pseudo-Independent Weight Perturbations on Mini-Batches. *ArXiv.* 2018. doi:10.48550/arXiv.1803.04386.

67. Wiegand, A. & Schwarz, D. J. Inhomogeneity-induced variance of cosmological parameters. *Astronomy & Astrophysics.* **538.** A147. 2012. doi:10.1051/0004-6361/201118137.

68. Yang, S., He, X. & Zhu, B. Learning Physical Constraints with Neural Projections. *ArXiv.* 2020. doi:doi.org/10.48550/arXiv.2006.12745.

69. Zucca, A., Pogosian, L., Silvestri, A. & Zhao, G. MGCAMB with massive neutrinos and dynamical dark energy. *Journal of Cosmology and Astroparticle Physics.* **2019.** 05. 1. 2019. doi:10.1088/1475-7516/2019/05/001.

A Appendix

A.1 Code and Data Sources

The code for this project can be found on my GitHub repository, at
https://github.com/arnava13/SeniorHonoursProject/. The updated version of BaCoN is
found in the folder `BaCoN-II(i)`, with the training scripts, testing scripts, Google Colab
notebook and numbered model files used in this project. The code used to plot the noise
model is found in the folder `Noise Plots`, the scripts used to bin the Dark Scattering
data in `Parameter Binning`, and the processed spectra from the new data generation
pipeline in `processed spectra`.

The entire dataset with 20,000 examples per class, the normalisation spectrum, the
random class filters and the theoretical error curves, generated by [Thummel *et al.*, Mar.
2024], can be found at https://doi.org/10.5281/zenodo.10628849. The spectra used are
found under `data/B2/EE2`.

The random spectra used in this report, as well as the binned Dark Scattering data
can be found at https://doi.org/10.5281/zenodo.10963220

A.2 Processed Spectra

Figure 9 displays the processed spectra after the new data generation process for each
redshift bin of a randomly selected input example. These appear to be of the same form
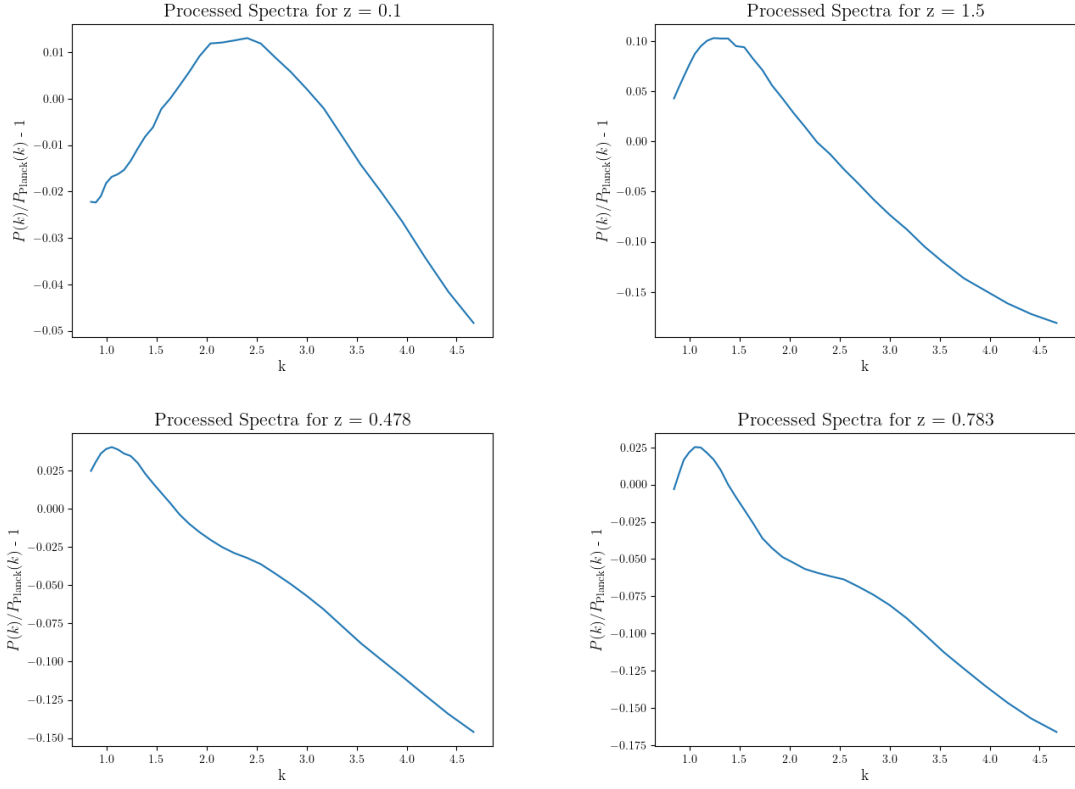found in Figure 1 of [Thummel *et al.*, Mar. 2024] – properly normalised and noisy.

Figure 9: Processed spectra output from the new data generation process.