Neural Network Emulation for Gravitational Wave Parameter Inference in Dark Matter Environments

Master Thesis

zur Erlangung des akademischen Grades
Master of Science (M.Sc.) in Physik
am Institut für Theoretische Physik (ITP)
der Goethe-Universität Frankfurt am Main

vorgelegt von

Jasmin Hartmann

Erstgutachterin:	Prof. Dr. Laura Sagunski
Zweitgutachter:	Dr. Alessio Spurio Mancini
eingereicht am:	23. Januar 2026

Abstract

Neural Network Emulation for Gravitational Wave Parameter Inference in Dark Matter Environments

In this thesis, we study the use of neural-network-based surrogate models for the efficient generation of gravitational-wave signals from supermassive black hole binary systems evolving in stellar regions of high density. The orbital evolution of such systems is determined not only by gravitational wave emission, but also by environmental effects such as dynamical friction. Accurate modeling of these signals is computationally expensive and poses a major challenge for Bayesian parameter inference.

We develop a deterministic neural emulator trained on reference simulations generated with the SMBHBpy software package. The emulator maps characteristic gravitational wave strain spectra across a multidimensional parameter space that includes initial eccentricity, total mass, and stellar velocity dispersion. The accuracy of the model is verified using appropriate validation metrics and direct waveform comparisons.

We then integrate the emulator into a Markov chain Monte Carlo inference and validate the method using injection recovery tests. The reconstructed parameters agree with the injected values. Emulator-induced deviations are below the statistical uncertainties. At the same time, the computational cost of waveform generation is reduced by orders of magnitude, while the inference time of the trained emulator remains independent of the physical complexity of the underlying reference model. Our results show that neural replacement models are a reliable and efficient tool for the accelerated analysis of gravitational wave data. This enables rapid exploratory testing of extended or newly implemented source models beyond existing waveform templates.

Contents

Abstract **i**

1 Introduction **1**

2 Theoretical Background **5**
2.1 Gravitational Waves and Binary Systems 5
2.2 Supermassive Black Hole Binaries and the SGWB 7
2.3 Dark Matter 8
2.4 Computational modeling and Neural-Network Emulation 10
2.5 Bayesian Inference Framework 12

3 Methodology and Implementation **15**
3.1 Overview of Simulation Processes 16
3.2 SMBHBpy Overview and Capabilities 16
3.3 Parameter Space Design and Sampling 17
3.4 Data Preparation and Signal Preprocessing 19

4 Neural Network Emulator Development **21**
4.1 Model Definitions and Objectives 21
4.2 Network Architecture 21
4.3 Training Methodology 24
4.4 Architecture Comparison and Validation 26

5 Parameter Inference Implementation **30**
5.1 Bayesian Framework for Parameter Estimation 30
5.2 MCMC Sampling Strategy 32
5.3 Convergence Diagnostics 33
5.4 Emulator Accuracy Considerations 34

6 Results and Discussion **36**

6.1 Emulator Training and Validation 36
6.2 MCMC Parameter Recovery 40

7 Conclusions and Future Work **45**
7.1 Conclusions 45
7.2 Future Work 46

Acknowledgements **48**

Selbstständigkeitserklärung **62**

List of Abbreviations

BBH	Binary Black Hole
BNS	Binary Neutron Star
CDM	Cold Dark Matter
CMB	Cosmic Microwave Background
EMRI	Extreme Mass-Ratio Inspiral
GELU	Gaussian Error Linear Unit
GR	General Relativity
GW	Gravitational Wave(s)
IMR	Intermediate Mass-Ratio
LHS	Latin Hypercube Sampling
LIGO	Laser Interferometer Gravitational-Wave Observatory
LISA	Laser Interferometer Space Antenna
ML	Machine Learning
MCMC	Markov Chain Monte Carlo
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
NN	Neural Network
PTA	Pulsar Timing Array
ReLU	Rectified Linear Unit
SGWB	Stochastic Gravitational-Wave Background
SIDM	Self-Interacting Dark Matter
SMBH	Supermassive Black Hole
SMBHB	Supermassive Black Hole Binary
SNR	Signal-to-Noise Ratio

Chapter 1

Introduction

In 2015, Advanced LIGO performed the first direct detection of a gravitational wave (GW) with the observation of the event GW150914 [1]. This signal originated from a binary black hole (BBH) system. Since then, hundreds of GW events have been detected, including the first multi-messenger observation, GW170817, which was produced by a neutron star merger [2, 3].

As GWs couple only weakly to matter, their detection enables studies of regions and epochs of the Universe that are inaccessible to electromagnetic observations [4, 5]. However, as the scope of possible investigations increases, the rapidly growing number of detected GW events introduces substantial computational challenges for data analysis [6].

Current ground-based observatories have established the foundation of GW astronomy but are now approaching the limits of their sensitivity. Although the frequency ranges of Advanced LIGO, Advanced Virgo, and KAGRA allow for the detection of an increasing number of sources, many astrophysical systems remain inaccessible. In particular, supermassive black hole binaries (SMBHBs) and the low-frequency stochastic GW background (SGWB) cannot be observed with current ground-based detectors [7, 8].

Next-generation observatories promise significantly improved sensitivities and access to new frequency bands, thereby opening a new observational window onto previously unobservable GW sources and physical phenomena. Dark matter (DM) is expected to contribute significantly to shaping the environments in which many of these sources evolve, particularly in galactic centers hosting SMBHBs. Such environments can induce subtle modifications to the

orbital dynamics of binaries and, consequently, to the emitted gravitational wave signals.

Three broad classes of GW detectors exist: ground-based interferometers, space-based observatories, and pulsar timing arrays (PTAs). Together, they span an almost continuous GW frequency spectrum. Figure 1.1 illustrates the characteristic frequency ranges and dominant source classes associated with each detector type.

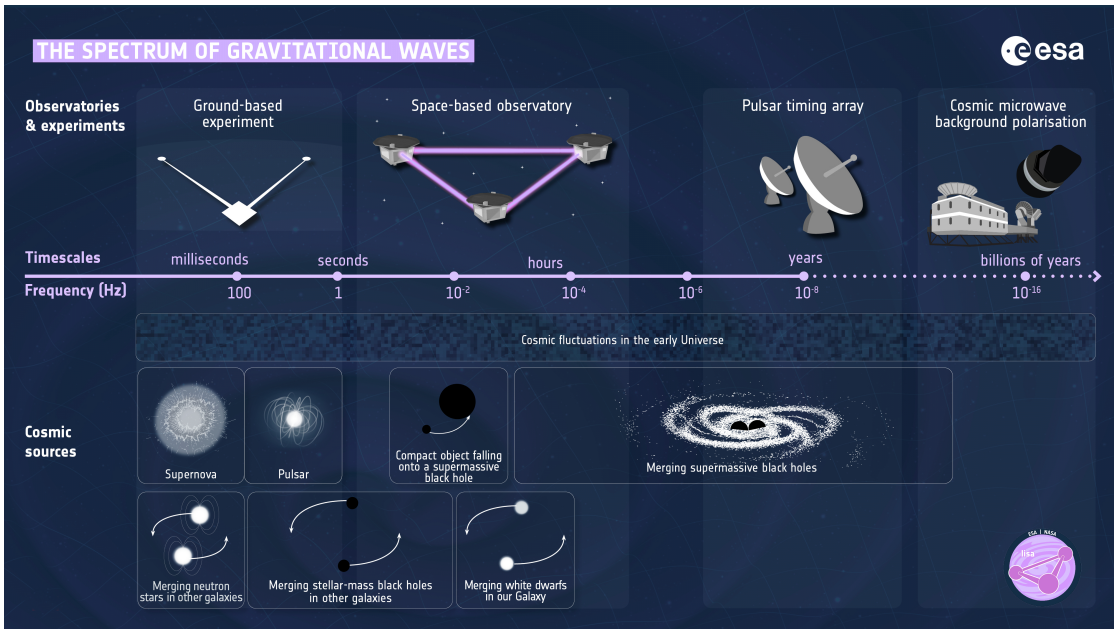

FIGURE 1.1: Spectrum of GWs: ground-based detectors, space-based observatories, and pulsar timing arrays probe complementary frequency bands and astrophysical sources. Taken and adjusted from [9].

The Laser Interferometer Space Antenna (LISA) is a space-based GW observatory operating in the millihertz band (10^{-4}–10^{-2} Hz), with a scheduled launch in the 2030s [10]. LISA is expected to detect hundreds of GW events per year, including signals from SMBHBs, extreme mass-ratio inspirals (EMRIs), and galactic binaries such as double white dwarfs.

Third-generation ground-based detectors include the Einstein Telescope [11] and Cosmic Explorer [12]. Operating at frequencies between ~ 1 and 10^3 Hz [13], these instruments will achieve sensitivity improvements of one to two orders of magnitude compared to current detectors. Their primary targets are

compact binary coalescences involving stellar-mass black holes and neutron stars.

Pulsar timing arrays probe the nanohertz regime (10^{-9}–10^{-7} Hz). Recent results from NANOGrav [14], EPTA [15], and PPTA [16] have reported evidence for a stochastic GW background, which is commonly interpreted as originating from a population of SMBHBs [14].

The physical accuracy of GW source models can be improved by accounting for environmental effects. In dense astrophysical environments, factors such as DM and nearby baryonic structures can influence binary evolution through processes like dynamical friction. DM accounts for approximately 85% of the total matter content of the Universe [17] and is expected to reach particularly high densities in galactic centers [18].

Among environmental effects, DM is of particular interest because different DM scenarios can leave measurable imprints on GW signals. Variations in density profiles and interaction strengths modify the binary dynamics and can affect the emitted gravitational radiation [19]. GWs therefore provide a potentially unique probe of DM properties in galactic centers, where electromagnetic constraints remain challenging. Recent PTA observations have revealed tentative spectral features in the nanohertz SGWB that may be consistent with environmental modifications such as DM, motivating systematic investigations of how DM influences GW signals [14].

Assessing whether extended source models are consistent with GW observations requires confronting their predicted signatures with the data. Bayesian parameter inference provides the standard framework for this task by linking noisy observations to the physical properties of the source and its environment [20]. In practice, this is commonly carried out using Markov Chain Monte Carlo (MCMC) methods [21]. However, MCMC analyses typically require a large number of waveform evaluations, making them computationally expensive for complex source models. While existing waveform approximants reduce this cost, they are largely restricted to well-established vacuum scenarios [22]. Incorporating environmental effects therefore necessitates more flexible and computationally efficient waveform representations [23].

This work addresses these challenges by developing a neural-network (NN)–based surrogate model for GW generation. The proposed GW emulator enables fast and accurate waveform evaluation for extended source models that include environmental effects, while remaining compatible with established Bayesian inference techniques. Its reliability is assessed explicitly within an MCMC-based parameter inference context. By significantly reducing waveform evaluation times, this approach facilitates rapid, systematic and comprehensive studies of novel GW source scenarios. Beyond demonstrating computational speedup, this work examines whether NN–based emulators can reliably capture subtle environmental effects, such as those induced by DM, within GW parameter inference.

The resulting framework is implemented as a modular Python package, `gwemu`, designed to integrate seamlessly with standard sampling methods. Its flexible structure allows straightforward adaptation to different waveform generators and supports a wide range of future GW analyses.

Chapter 2

Theoretical Background

The theoretical foundation of this work lies at the intersection of GW physics and machine learning (ML). On the astrophysics side, the modeling of GW signals from SMBHBs is based on general relativity (GR) and incorporates environmental effects such as DM. These components determine the physical structure and evolution of the GW signals considered in this thesis. The methodology is rooted in ML techniques, in particular NNs, which are employed to efficiently approximate complex GW models and enable their use in Bayesian parameter inference.

Given this interdisciplinary scope, the theoretical concepts introduced in this chapter are presented with a focus on the aspects most relevant to GW modeling and inference with NN-based surrogate models. The discussion is intentionally selective and aims to provide a coherent and accessible foundation for the subsequent chapters, allowing readers from different backgrounds to follow both the physical motivation and the methodological assumptions underlying the framework developed in this thesis.

2.1 Gravitational Waves and Binary Systems

GWs are disturbances in spacetime that travel at light speed. Even though their existence was only confirmed through direct observations in 2015 [1], Einstein had already predicted them almost 100 years earlier, in 1916, in his theory of general relativity [24].

GWs are produced from accelerating mass distributions that give rise to a non-axisymmetric time-varying quadrupole moment. The emission of GWs causes

a gradual loss of angular momentum as well as orbital energy [25]. Compact binary systems, such as binaries of black holes or neutron stars, represent the dominant and best-understood sources of gravitational radiation [4, 26].

In the weak-field and slow-motion regime, the quadrupole contribution dominates the GW emission from compact binaries. This regime is relevant for most of the inspiral phase [4, 27]. At leading order, the GW strain amplitude depends on the component masses and their orbital separation. Orbital energy and angular momentum losses follow the same scaling. Here, GW emission is described by the quadrupole formula [28, 29]. It relates the strain amplitude h to the second time derivative of the system's mass quadrupole moment. For a binary system with the component masses m_1 and m_2 on a quasi-circular orbit with angular frequency ω, the leading-order scaling of the GW strain in the post-Newtonian expansion is given by

$$h \sim \frac{G}{c^4} \frac{M_{\text{chirp}}^{5/3} \, \omega^{2/3}}{D}, \tag{2.1}$$

with the chirp mass M_{chirp}

$$M_{\text{chirp}} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}. \tag{2.2}$$

Here, G denotes the gravitational constant, D the luminosity distance to the source, and c is the speed of light. The binary orbit evolves due to the continuous emission of GWs. More specifically, for a quasi-circular orbit the GW energy loss at leading order in the quadrupole approximation is given by

$$\frac{dE}{dt} = -\frac{32}{5} \frac{G^4}{c^5} \frac{m_1^2 m_2^2 M}{a^5}, \tag{2.3}$$

where $M = m_1 + m_2$ is the total mass of the system and a denotes the semi-major axis of the orbit. As a result of this energy loss, the orbital separation decreases according to

$$\frac{da}{dt} = -\frac{64}{5} \frac{G^3}{c^5} \frac{m_1 m_2 M}{a^3}. \tag{2.4}$$

As the orbit shrinks, the orbital angular frequency increases, which is directly reflected in the GW frequency via $f_{\text{GW}} = \omega / \pi$ for the dominant quadrupole

mode. The concurrent growth of frequency and amplitude is denoted as chirp [4].

The binary components subsequently continue to inspiral, accompanied by an increase in their orbital velocity, until the system enters the merger phase. This is followed by the ringdown of the remnant object. Taken together, these stages form the inspiral–merger–ringdown sequence, which can be directly observed in GW signals. The characteristic timescale for orbital decay due to GW emission can be expressed as [29]

$$\tau \propto \frac{a^4}{m_1 m_2 M}, \tag{2.5}$$

where a is the semi-major axis. For eccentric binaries, GW emission also leads to orbital circularisation, with the eccentricity typically decreasing more rapidly than the orbital separation during the inspiral [29, 4].

2.2 Supermassive Black Hole Binaries and the SGWB

SMBHBs are promising sources of low-frequency GW. Recent PTA observations provide evidence for a SGWB that may be caused by the incoherent superposition of GW emitted by a large cosmic population of inspiralling SMBHBs [14, 15, 30].

The operating principle of PTAs is based on the rotational stability of millisecond pulsars to probe nanohertz GWs. More specifically, they monitor correlated variations in pulse arrival times across a galactic-scale pulsar network. The observed correlations follow the quadrupolar angular dependence described by the Hellings–Downs curve [31]. This supports an interpretation in terms of a stochastic background sourced predominantly by SMBHBs [32, 33].

SMBHBs are natural results from galaxy mergers [34]. As their host galaxies coalesce, their central supermassive black holes (SMBHs) lose orbital energy through dynamical friction against the surrounding stellar and DM components and sink toward the center of the merger remnant. At parsec-scale separations, the resulting binary becomes gravitationally bound and continues to

harden through interactions with stars and gas and eventually through GW emission, leading to coalescence [35, 36, 37].

The SGWB spectrum carries information about the SMBHB population it originates from. Both the amplitude and the spectral shape depend on several factors. Among them are the mass and mass-ratio distributions of SMBHBs, the cosmic merger history and the environmental conditions affecting binary evolution [38, 33, 39]. The characteristic strain $h_c(f)$ can be written as the integrated contribution of all inspiralling SMBHBs over cosmic time,

$$h_c^2(f) = \int_0^\infty \int \frac{d^3N}{dz dM dq} \frac{1}{1+z} \frac{dE_{\rm GW}}{d\ln f_r} \frac{dz\, dM\, dq}{4\pi D_L^2(z)}, \tag{2.6}$$

where $d^3N/(dzdMdq)$ denotes the comoving number density of binaries with redshift z, total mass M and mass ratio q. The term $f_r = (1+z)f$ is the rest-frame frequency and $D_L(z)$ is the luminosity distance [40, 36].

2.3 Dark Matter

In the standard cosmological model, DM is the dominant matter component of the Universe. It plays a central role in the formation and evolution of large-scale structure. There are a wide range of astrophysical and cosmological observations that provide evidence for the existence of DM. Phenomena such as flat galaxy rotation curves, gravitational lensing in galaxy clusters, and anisotropies in the cosmic microwave background (CMB) cannot be consistently explained by luminous matter alone [41, 42, 17].

DM is usually modeled as a non-baryonic component that interacts mainly through gravity. On galactic scales, it forms extended halos that surround galaxies and their SMBHs [41]. As a consequence, compact objects in galactic nuclei evolve in DM-rich environments, which motivates treating DM as an environmental effect in GW source modeling.

In its simplest formulation, the cold dark matter (CDM) paradigm assumes that DM particles have negligible self-interactions. This assumption successfully reproduces observations on large cosmological scales [43]. However, alternative scenarios have been proposed to address discrepancies on galactic

scales, such as the core–cusp problem observed in dwarf galaxies and galactic centers [44]. One of these scenarios is self-interacting dark matter (SIDM), in which the DM particles have strong self-interactions. It predicts modified density distributions in the central regions of galaxies, which typically lead to shallower inner density profiles compared to collisionless CDM halos [45, 46]. These differences in DM microphysics can lead to distinct gravitational environments around SMBHs, which in turn result in observable consequences for SMBHB evolution and GW emission.

SMBHBs lose energy and angular momentum through several physical processes as they evolve towards merger. However, this evolution is not necessarily continuous. The so-called *final parsec problem* arises when binaries stall near $\sim$1 pc separations, where stellar and gas interactions become inefficient while GW emission remains too weak to dominate [47]. Within these scales, the binary typically ejects or scatters nearby stars, which depletes the "loss cone" of orbits that could otherwise extract energy from the system. Dynamical friction exerted by DM spikes surrounding SMBHs is a promising proposed mechanism to overcome this longstanding challenge [18, 48, 49].

Due to its slow accretion of mass, a central black hole can adiabatically steepen the surrounding DM density into a spike. The resulting density profile is commonly modeled as a power law, given by

$$\rho_{\rm sp}(r) \propto r^{-\gamma}, \tag{2.7}$$

where $\gamma = 7/3$ corresponds to an ideal adiabatic spike grown from a Navarro–Frenk–White (NFW) halo [18]. It is important to note that effects such as stellar heating can flatten the spike to $\gamma = 3/2$ [50], while SIDM can produce intermediate slopes such as $\gamma = 7/4$ [49, 46].

Dense DM spikes transfer orbital energy and angular momentum from the SMBHB to the surrounding halo and therefore enhance dynamical friction [51, 19]. This additional drag can slow the binary's evolution and modify the predicted GW spectrum, which can potentially produce a low-frequency turnover in the SGWB [52].

For SIDM, thermalisation through self-interactions allows the spike to absorb energy without being destroyed. This allows for an efficient inspiral to the

GW-dominated regime and potentially resolves the final parsec problem [49]. In contrast, CDM spikes tend to be more fragile and less efficient at extracting energy.

DM-induced friction can also influence eccentricity evolution. In particular, steep DM spikes tend to enhance circularisation. Shallower or disturbed halos on the other hand can drive eccentricity growth [51, 19]. These effects can lead to dephasing, frequency shifts, or changes in the inspiral braking index. Similar signatures may arise from stellar scattering or gas accretion, which complicates the attribution of observed SGWB features to DM alone [36, 53].

2.4 Computational modeling and Neural-Network Emulation

The physical modeling of SMBHB evolution involves a range of coupled processes. These include the evolution of orbital eccentricity and cosmological merger rates, as well as several environmental effects such as stellar scattering, gas dynamics, and interactions with DM. Accurately describing these mechanisms requires solving systems of coupled differential equations over Myr–Gyr timescales [32, 37, 38, 33]. High-fidelity simulations are therefore computationally demanding.

The computational cost increases further in population-based studies, for example when modeling stochastic GW signals. Discrete sampling of the cosmic SMBHB population introduces realization variance, which can only be captured through large ensembles of simulations [54, 39]. Performing Bayesian inference under these conditions becomes challenging, as it requires a large number of repeated model evaluations.

More generally, inference problems in GW astrophysics are often limited by the cost of waveform generation and likelihood evaluation. Each evaluation can involve complex differential equations. Population synthesis from first principles may also be required. Consequently, this restricts the dimensionality of the parameter space that can be explored in practice [55, 37]. This is the motivation for surrogate models that approximate the relation between physical parameters and GW signals at much lower computational cost.

NNs provide a flexible class of surrogate models inspired by biological neural systems. Their conceptual foundation is based on the perceptron model introduced by Rosenblatt [56]. The perceptron is a simplified mathematical abstraction of a biological neuron. It computes a weighted sum of input features and applies a nonlinear activation function,

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right). \tag{2.8}$$

Here, x_i are the input features, w_i are the corresponding weights, and b denotes a bias term. The activation function f introduces nonlinearity into the model. Common choices include the sigmoid function,

$$\sigma(x) = \frac{1}{1+e^{-x}}, \tag{2.9}$$

the hyperbolic tangent,

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}, \tag{2.10}$$

and the rectified linear unit (ReLU),

$$\mathrm{ReLU}(x) = \max(0, x). \tag{2.11}$$

Among these, ReLU is widely used due to its computational simplicity and robust performance in many practical applications [57, 58].

Multi-layer perceptrons (MLPs) consist of stacked perceptrons, where each layer applies a linear transformation followed by a nonlinear activation,

$$\mathbf{h} = f(\mathbf{W}\mathbf{x} + \mathbf{b}), \tag{2.12}$$

where $\mathbf{W}$ is the weight matrix and $\mathbf{b}$ the bias vector.

The expressive power of such architectures is formalised by the universal approximation theorem. It states that feedforward NNs with at least one hidden layer are capable of approximating any continuous function on a compact domain [59]. It is important to emphasize that this does not provide guarantees on practical training efficiency or generalisation.

Training a NN involves tuning its parameters to match a target output. This is achieved by minimising a loss function. A loss function measures the difference between the NN predictions and the training data [58]. Backpropagation efficiently computes the gradients of the loss function with respect to all network parameters by recursively applying the chain rule from the output layer to the input layer. These gradients are then used by gradient-based optimization methods to update the network parameters accordingly [57]. The mean-squared error is a common choice for regression tasks, given by

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2. \tag{2.13}$$

Several strategies are commonly used to improve generalisation and reduce overfitting in NNs. These include regularisation techniques such as weight decay, which penalizes large weights, batch normalization, which stabilises and accelerates training by normalising intermediate activations, and dropout, which randomly disables neurons during training to reduce co-adaptation [57]. In addition, adaptive optimization algorithms such as Adam automatically adjust learning rates based on first- and second-order gradient moments and can improve convergence behavior [60].

NNs have become an established tool in GW astrophysics, with applications ranging from waveform surrogate modeling and noise characterisation to accelerated parameter estimation [61, 62, 63, 64, 65]. The replacement of expensive numerical simulations with faster approximations enables orders-of-magnitude speedups in Bayesian inference [61, 66]. In our work, NN emulation is introduced as a tool to address computational bottlenecks in GW parameter inference, with a particular focus on SMBHBs and environmental effects.

2.5 Bayesian Inference Framework

Bayesian inference forms the basis of parameter estimation in GW astrophysics. In particular, it aims to determine the posterior distribution $P(\theta|d)$ from observed data and prior information [67, 68, 69].

Despite its conceptual clarity, Bayesian parameter inference poses several practical challenges in GW astrophysics [20, 21]. Repeated waveform evaluations render Markov Chain Monte Carlo (MCMC) methods computationally expensive for complex source models, while prior selection and sampler convergence require careful consideration. These challenges motivate the use of fast surrogate waveform models that remain compatible with established inference pipelines.

Bayes' theorem relates the posterior distribution to the likelihood and the prior,

$$P(\theta|d) = \frac{P(d|\theta)P(\theta)}{P(d)}, \tag{2.14}$$

where $P(\theta)$ is the prior distribution and $P(d)$ is the Bayesian evidence. Even though the evidence plays a central role in Bayesian model comparison, it acts only as a normalisation constant for parameter estimation and can therefore be neglected [70].

The likelihood function $P(d|\theta)$ essentially evaluates how well a given set of model parameters explains the observed data. The GW detector noise is usually assumed to follow a Gaussian distribution, with a corresponding likelihood of

$$\log L(\theta) = -\frac{1}{2}\sum_i \frac{(d_i - m_i(\theta))^2}{\sigma_i^2} - \frac{1}{2}\sum_i \log(2\pi\sigma_i^2). \tag{2.15}$$

Here, d_i are the measured strain values, $m_i(\theta)$ are the corresponding model predictions and σ_i are the noise variances [71, 72]. This expression reduces to a chi-squared likelihood if the detector noise is assumed to be independent and Gaussian.

Because posterior distributions tend to be high-dimensional in practical applications, an analytical solution is often not possible. Instead, MCMC methods provide a numerical way to sample the posterior and quantify uncertainties [73, 74, 75]. The expectations of functions $f(\theta)$ are here approximated as

$$\int f(\theta)P(\theta|d)\mathrm{d}\theta \approx \frac{1}{N}\sum_{i=1}^{N} f(\theta_i). \tag{2.16}$$

Here, the samples θ_i are drawn from the posterior distribution.

In high-dimensional parameter spaces, sampling the posterior distribution efficiently poses a significant challenge. MCMC methods address this by constructing a Markov chain whose stationary distribution corresponds to the target posterior, which enables efficient exploration of parameter space. A common implementation is the `emcee` package [76], which employs an affine-invariant ensemble sampler that evolves multiple walkers simultaneously.

To ensure that the sampled distributions faithfully represent the target posterior, it is necessary to verify the convergence of the sampler before interpreting the results. Ways to do this are to monitor the autocorrelation time of the chains, compare results across independent runs and visually inspect the trace plots [77, 76].

Chapter 3

Methodology and Implementation

This chapter introduces the methodological framework employed in this thesis to model GW signals from SMBHBs and to perform efficient Bayesian parameter inference. Building on the theoretical foundations outlined in Chapter 2, the focus now shifts to the practical implementation of the simulation pipeline, the NN-based surrogate model, and the inference strategy used to analyse the resulting GW signals.

The methodology combines physically motivated GW modeling with ML techniques in order to address the computational challenges inherent to traditional simulation-based approaches. In particular, NN emulation is employed to significantly reduce waveform evaluation times, thereby enabling MCMC-based parameter inference within a feasible computational budget. Each component of the pipeline is described in detail to clarify its role within the overall framework and to ensure reproducibility.

NN emulation refers to the use of trained NNs as surrogate models that approximate the input–output behavior of computationally expensive physical simulations. In contrast to traditional modeling approaches based on explicit numerical integration or analytical approximations, NN emulators learn this mapping directly from data and enable fast waveform predictions once trained. The neural network models are implemented using the PyTorch framework, which provides efficient automatic differentiation and flexible model construction for training and evaluation [78].

3.1 Overview of Simulation Processes

The aim of our work is to develop a flexible computational framework to train NN-based GW emulators in order to speed up waveform generation and subsequent parameter inference. The resulting emulator framework is designed to be robust and capable of capturing physics beyond the standard model.

The simulation process consists of five main steps. In the first step, the `SMBHBpy` framework is chosen as a concrete application case to simulate SMBHB evolution under environmental effects, including stellar dynamical friction and DM halo models [79, 80]. Subsequently, systematic parameter space sampling is applied to efficiently cover the full parameter space. Next, the simulated signals are preprocessed and normalized via logarithmic transformation and z-score standardization to prepare the data for stable and effective NN training [81]. In the fourth step, the NN-based waveform emulator is trained, learning the mapping between physical source parameters and GW signals. Finally, a subsequent MCMC analysis [76] is used to validate that the trained emulator can be employed within established GW inference pipelines.

The NN-based emulation framework can significantly reduce the computational cost of waveform evaluation, even for scenarios involving new physics models. The developed framework is modular and packaged as a Python library, referred to as `gwemu`. Its structure is designed such that alternative GW source models can be integrated with minimal code changes.

3.2 SMBHBpy Overview and Capabilities

The Python-based `SMBHBpy` package was developed by Daniel [79]. It is used to model the orbital evolution of SMBHBs and compute the associated GW emission. The framework extends the earlier `IMRIpy` code base for intermediate mass-ratio (IMR) inspirals developed by Becker [80], adapting it to the SMBHB regime and to the low-frequency GW signals typically associated with binaries embedded in dense stellar and DM environments.

The framework is implemented in geometrized units ($G = c = 1$) with internal unit conversions to parsecs. Its architecture enables the simultaneous numerical solution of coupled differential equations governing the temporal evolution

of the semi-major axis $a(t)$ and eccentricity $e(t)$ under the influence of multiple energy dissipation mechanisms. The orbital evolution is determined via orbit-averaged energy and angular momentum balance equations,

$$\frac{dE_{\rm orb}}{dt} = \left\langle \frac{dE_{\rm GW}}{dt} \right\rangle + \left\langle \frac{dE_{\rm DF}}{dt} \right\rangle, \tag{3.1}$$

$$\frac{dL_{\rm orb}}{dt} = \left\langle \frac{dL_{\rm GW}}{dt} \right\rangle + \left\langle \frac{dL_{\rm DF}}{dt} \right\rangle, \tag{3.2}$$

from which numerical integration yields the corresponding $a(t)$ and $e(t)$ trajectories.

The GW contribution follows the Peters–Mathews formalism [28, 29]. Environmental effects are incorporated through dynamical friction terms. In stellar environments, these depend on the velocity dispersion σ, which defines the radius of influence,

$$r_h = \frac{GM}{\sigma^2}, \tag{3.3}$$

where $M = m_1 + m_2$ denotes the total binary mass. DM environments introduce additional dynamical friction contributions that depend on the assumed density profile and interaction model. `SMBHBpy` supports vacuum configurations as well as CDM and SIDM scenarios [18, 49].

3.3 Parameter Space Design and Sampling

The parameter space considered in this work consists of three continuous input variables that directly influence the resulting GW signal. The logarithmic total mass range is chosen to cover physically sensible SMBHB systems in the PTA regime. The initial eccentricity is restricted to moderate values consistent with galaxy merger simulations while maintaining numerical solver stability. The velocity dispersion range corresponds to typical values observed in galactic nuclei, spanning lower-mass systems to massive ellipticals.

An overview of the parameter ranges is given in Table 3.1 .

This analysis focuses on the difference between vacuum and CDM environments. The DM type is treated as a discrete choice during data generation, and separate datasets with 1,000 parameter combinations are generated for each

Parameter	**Symbol**	**Range**
Initial eccentricity	e_0	$[0.0, 0.3]$
Total mass	$\log_{10}(M_{\text{tot}}/M_\odot)$	$[8.0, 10.0]$
Velocity dispersion	σ	$[100, 400]$ km/s

TABLE 3.1: Parameter ranges for the three-dimensional continuous parameter space. e_0 controls the initial orbital shape, M_{tot} is the combined black hole mass, and σ characterizes the stellar environment.

scenario. While a unified emulator including the DM type as an additional input parameter would be possible in principle, separating the environments allows for a more controlled analysis of environmental effects at fixed computational cost. Although the framework supports additional interaction models such as SIDM, this study strictly focuses on vacuum and CDM configurations.

Efficient coverage of the parameter space is essential, as naive random sampling can lead to clustering and gaps. Latin Hypercube Sampling (LHS) [82], as implemented in `scipy.stats.qmc`, ensures approximately uniform marginal distributions for each parameter while avoiding the computational redundancy of grid-based sampling. Each parameter range is divided into N equally sized intervals, with one sample drawn from each interval in every dimension. A common construction is

$$x_{ij} = \frac{\pi_j(i) - U_{ij}}{N}, \quad i = 1, \ldots, N, \quad j = 1, \ldots, d, \tag{3.4}$$

where π_j denotes a random permutation of $\{1, \ldots, N\}$ and $U_{ij} \sim \mathcal{U}(0, 1)$ are independent uniform random variables. The resulting samples are subsequently scaled to the physical parameter bounds.

FIGURE 3.1: Coverage of the three-dimensional parameter space achieved by Latin Hypercube Sampling. The scatter plot illustrates the $(e_0, \log_{10} M_{\text{tot}})$ plane, with color indicating the velocity dispersion σ. Uniform coverage without clustering artifacts is observed.

3.4 Data Preparation and Signal Preprocessing

The GW signals generated with `SMBHBpy` are represented in terms of the characteristic strain $h_c(f)$, which provides a compact frequency-domain description of the GW amplitude and is commonly used in studies of low-frequency sources such as SMBHBs. In this work, $h_c(f)$ constitutes the primary signal representation used for dataset construction, NN training, and Bayesian parameter inference.

Before NN training, the generated waveform signals undergo several preprocessing steps. The characteristic strain amplitudes typically span multiple orders of magnitude, ranging from 10^{-18} to 10^{-14}. To compress this dynamic range and prevent large-amplitude regions from dominating the loss function, a logarithmic transformation is applied,

$$\tilde{h}_c(f_i) = \log_{10}\left(h_c(f_i)\right). \tag{3.5}$$

This transformation emphasizes relative differences between spectral features across frequency scales.

As `SMBHBpy` can produce signals of variable length depending on the orbital evolution timescale, all signals are resampled to a common frequency grid consisting of 128 points between 10^{-10} and 10^{-6} Hz. For signals longer than 128 points, uniform downsampling is performed,

$$f_{\text{resampled}}[k] = f_{\text{original}}\left[\left\lfloor k \cdot \frac{N_{\text{orig}} - 1}{127} \right\rfloor\right], \quad k = 0, \ldots, 127. \tag{3.6}$$

For shorter signals, linear interpolation is used to upsample the data,

$$h_{\text{resampled}}(f_k) = \text{interp1d}\left(f_{\text{original}}, h_{\text{original}}\right)(f_k). \tag{3.7}$$

The dataset is subsequently split into training and validation sets. No separate test set is used, as the model performance is assessed through validation metrics and downstream inference results. Z-score normalization is applied to both the input parameters and the GW signals to center the data around zero

with unit variance,

$$\hat{x} = \frac{x - \mu_x}{\sigma_x}, \quad \hat{y} = \frac{y - \mu_y}{\sigma_y}, \tag{3.8}$$

where the mean and standard deviation are computed exclusively on the training set [81]. The normalization parameters are stored as part of the trained model to ensure consistent transformations during inference.

Failed simulations are discarded prior to training. Numerical instabilities at extreme parameter combinations can lead to non-physical solutions, and removing such cases ensures that all training data represent physically valid orbital evolutions. Additional quality control checks verify finite signal values, appropriate frequency coverage, and physically reasonable amplitude scaling with mass. After completion of all preprocessing steps, NN training is performed on the resulting dataset.

Chapter 4

Neural Network Emulator Development

This chapter introduces the NN emulator, specifically the model definition and objectives, details of the network architecture as well as the training methodology.

4.1 Model Definitions and Objectives

The emulator approximates a smooth but highly non-linear mapping $\mathcal{F} : \mathbb{R}^3 \to \mathbb{R}^{128}$ from three physical parameters to the 128-dimensional log-strain spectrum:

$$\hat{\mathbf{y}} = \mathcal{F}_\theta \left(e_0, \log_{10} M_{\text{tot}}, \sigma \right). \tag{4.1}$$

Here, $\hat{\mathbf{y}} = (\hat{y}_1, \ldots, \hat{y}_{128})$ represents the predicted normalized log-strain at each of the frequency points. The network learns the parameters θ during its training.

The emulator must interpolate smoothly across the parameter space while significantly accelerating GW signal generation.

4.2 Network Architecture

The NN architecture adopts a fully connected feedforward design. While the output dimension (128) greatly exceeds the input dimension (3), the physical

dependence of the GW spectrum on the source parameters is smooth and low-dimensional. As a result, intermediate representations of moderate complexity are sufficient to capture the relevant structure of the mapping.

The NN consists of an input layer, followed by three hidden layers, and an output layer:

$$\mathbf{z}^{(1)} = \mathbf{W}^{(1)}\hat{\mathbf{x}} + \mathbf{b}^{(1)}, \qquad \mathbf{h}^{(1)} = \mathrm{GELU}\left(\mathrm{LayerNorm}\left(\mathbf{z}^{(1)}\right)\right), \tag{4.2}$$

$$\mathbf{z}^{(2)} = \mathbf{W}^{(2)}\mathrm{Dropout}\left(\mathbf{h}^{(1)}\right) + \mathbf{b}^{(2)}, \quad \mathbf{h}^{(2)} = \mathrm{GELU}\left(\mathrm{LayerNorm}\left(\mathbf{z}^{(2)}\right)\right), \tag{4.3}$$

$$\mathbf{z}^{(3)} = \mathbf{W}^{(3)}\mathrm{Dropout}\left(\mathbf{h}^{(2)}\right) + \mathbf{b}^{(3)}, \quad \mathbf{h}^{(3)} = \mathrm{GELU}\left(\mathrm{LayerNorm}\left(\mathbf{z}^{(3)}\right)\right), \tag{4.4}$$

$$\hat{\mathbf{y}} = \mathbf{W}^{(4)}\mathbf{h}^{(3)} + \mathbf{b}^{(4)}. \tag{4.5}$$

Here, $\hat{\mathbf{x}} \in \mathbb{R}^3$ denotes the normalized input parameter vector, consisting of the initial eccentricity e_0, the logarithmic total mass $\log_{10} M_{\mathrm{tot}}$, and the velocity dispersion σ. The vectors $\mathbf{z}^{(l)}$ represent the affine transformations at layer l, while $\mathbf{h}^{(l)}$ denote the corresponding hidden layer activations after normalization and non-linear transformation. The network output $\hat{\mathbf{y}} \in \mathbb{R}^{128}$ corresponds to the predicted, normalized log-strain evaluated on the fixed frequency grid introduced in Chapter 3.

When consecutive layers have matching dimensionality, skip connections are introduced. Here, $\mathrm{GELU}(\cdot)$ denotes the Gaussian Error Linear Unit activation function, which is defined explicitly in Equation 4.7:

$$\mathbf{h}^{(i+1)} = \mathrm{GELU}\left(\mathrm{LayerNorm}\left(\mathbf{z}^{(i+1)}\right)\right) + \mathbf{h}^{(i)}. \tag{4.6}$$

These residual connections serve two purposes: preserving gradient flow during backpropagation, and allowing low-level parameter combinations to directly influence the output. Both properties prove beneficial for learning smooth physical mappings.

Listing 4.1 shows the core implementation of the forward pass. The full training and inference scripts follow standard PyTorch and `emcee` usage patterns and are summarized in the methodology and inference chapters.

```python
def forward(self, x: torch.Tensor) -> torch.Tensor:
    h1 = self.activation(self.ln1(self.fc1(x)))

    h2 = self.fc2(self.dropout(h1))
    h2 = self.activation(self.ln2(h2)) + h1   # Residual

    h3 = self.fc3(self.dropout(h2))
    h3 = self.activation(self.ln3(h3)) + h2   # Residual

    return self.fc4(h3)
```

LISTING 4.1: Core NN forward pass with residual connections

All hidden layers use the Gaussian Error Linear Unit (GELU) activation function [83]:

$$\mathrm{GELU}(x) = x \cdot \Phi(x) = x \cdot \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right], \tag{4.7}$$

where $\Phi(x)$ is the standard normal cumulative distribution function.

Unlike ReLU (Eq. 2.11), GELU provides a smooth, continuously differentiable activation, which is well suited for regression on smooth GW spectra and leads to more stable training.

Additionally, each hidden layer includes Layer Normalization [84]:

$$\mathrm{LayerNorm}(\mathbf{h}) = \gamma \odot \frac{\mathbf{h} - \mu_{\mathbf{h}}}{\sqrt{\sigma^2_{\mathbf{h}} + \epsilon}} + \beta, \tag{4.8}$$

where $\mu_{\mathbf{h}}$ and $\sigma^2_{\mathbf{h}}$ are the mean and variance over the feature dimension, γ and β are learnable scale and shift parameters, and $\epsilon = 10^{-5}$ ensures numerical stability.

Layer Normalization is preferred over Batch Normalization for two reasons that are particularly relevant in a Bayesian inference setting. First, it behaves the same during training and inference, which is important when evaluating individual samples in MCMC. Second, it does not rely on running estimates of activation statistics, which simplifies the model storage and reuse [84, 85].

Other techniques like dropout regularization [86] were explored during initial experiments, but did not lead to the best-performing configuration. Stochastic

regularization was not required, due to the smoothness of the target function and the use of strong weight decay.

In addition, L2 weight regularization (weight decay) with coefficient $\lambda = 10^{-4}$ is applied during optimization,

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}} + \lambda \sum_{l} \left\| \mathbf{W}^{(l)} \right\|_F^2, \tag{4.9}$$

where $\| \cdot \|_F$ denotes the Frobenius norm. This regularization term penalizes large weight magnitudes and thereby reduces the risk of overfitting to noise in the training data.

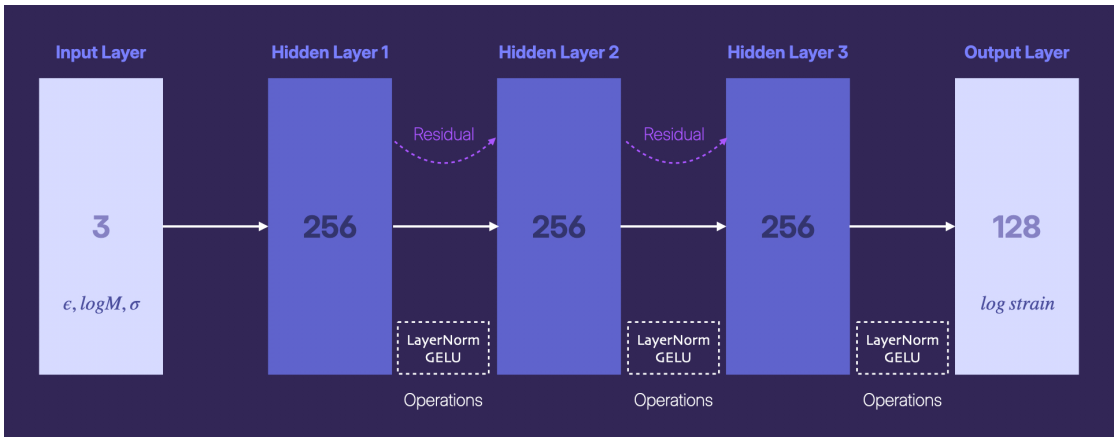

FIGURE 4.1: Architecture of the NN emulator with three input parameters, three uniform hidden layers (256 neurons each) with residual connections, and a 128-dimensional log-strain output. LayerNorm and GELU activations are applied after each hidden layer.

4.3 Training Methodology

The training objective is defined through a combined loss function:

$$\mathcal{L} = \alpha \cdot \mathcal{L}_{\text{MSE}} + (1 - \alpha) \cdot \mathcal{L}_{\text{L1}}, \tag{4.10}$$

where

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N \cdot D} \sum_{i=1}^{N} \sum_{j=1}^{D} \left(y_{ij} - \hat{y}_{ij}\right)^2, \tag{4.11}$$

$$\mathcal{L}_{\text{L1}} = \frac{1}{N \cdot D} \sum_{i=1}^{N} \sum_{j=1}^{D} \left|y_{ij} - \hat{y}_{ij}\right|, \tag{4.12}$$

with the batch size N and the output dimension $D = 128$. The weighting factor $\alpha = 0.7$ assigns increased importance to MSE.

MSE penalizes large errors strongly, which drives the network to minimize outliers. L1 provides robustness and encourages sparse residuals. Since targets are already in log-space (see Section 3.4), minimizing this loss corresponds to minimizing relative errors in the original strain amplitudes.

Parameter optimization is carried out using the Adam optimizer [60], which adapts the learning rate during training to improve convergence:

$$\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1)\mathbf{g}_t, \tag{4.13}$$

$$\mathbf{v}_t = \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2)\mathbf{g}_t^2, \tag{4.14}$$

$$\hat{\mathbf{m}}_t = \mathbf{m}_t / (1 - \beta_1^t), \quad \hat{\mathbf{v}}_t = \mathbf{v}_t / (1 - \beta_2^t), \tag{4.15}$$

$$\theta_{t+1} = \theta_t - \eta \cdot \hat{\mathbf{m}}_t / (\sqrt{\hat{\mathbf{v}}_t} + \epsilon), \tag{4.16}$$

with momentum parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$ for stability. The adaptive rates automatically adjust step sizes for parameters with different gradient magnitudes to navigate the loss landscape efficiently.

To prevent occasional instabilities from large gradients, gradient clipping constrains the maximum norm to 1.0 [87]:

$$\mathbf{g} \leftarrow \mathbf{g} \cdot \min\left(1, \frac{1.0}{\|\mathbf{g}\|}\right). \tag{4.17}$$

After preprocessing, data is split into training (80%) and validation (20%) sets. Training runs for 100 epochs with batch size 32, monitoring validation loss after each epoch. The optimal batch size was selected through systematic hyperparameter comparison across batch sizes ranging from 16 to 128. The model checkpoint with the lowest validation loss is saved and loaded for the

final evaluation. When available, the training uses a single GPU (otherwise CPU). With approximately 167,040 trainable parameters, the training completes within 3 seconds for datasets of 1,000 samples. The experiments were conducted on a local machine with 16GB of memory.

4.4 Architecture Comparison and Validation

The NN architecture design involves trade-offs between model capacity, generalization, and computational efficiency. In the context of GW emulation, these trade-offs are shaped by the smooth dependence of GW spectra on the underlying physical parameters. This motivates architectures that are sufficiently expressive to capture nonlinear dependencies, while avoiding unnecessary complexity that may lead to overfitting or optimization difficulties. To assess the impact of these design choices, a systematic comparison of alternative network configurations is conducted.

The design space explored spans three main dimensions: depth, width and architectural motifs. The depth of the NN depends on its number of hidden layers. Greater network depth allows more abstract representations to be learned through successive transformations. However, deeper networks require careful regularization to avoid overfitting. Additionally, they may suffer from vanishing gradients or optimization plateaus, despite techniques like residual connections. The width is defined through the neurons per layer and determines the capacity of intermediate representations. Wider layers increase capacity and enable richer feature combinations, but computational cost scales quadratically in fully-connected architectures. Architectural motifs describe recurring structural patterns in NN design that determine how information is transformed and passed on across layers. These design choices implicitly encode assumptions about how physical dependencies should be represented and combined [57, 88].

In total nine network configurations were trained on identical data and compared across four metrics: best validation R^2 (maximum coefficient of determination on held-out validation set), final validation loss (combined MSE + L1 at end of training), training time, and parameter count (total trainable weights).

All architectures trained under identical conditions: 100 epochs, batch size 32, Adam optimizer ($\eta = 10^{-3}$, $\lambda = 10^{-4}$), combined MSE + L1 loss (weights 0.7, 0.3), and gradient clipping at norm 1.0. Training and validation datasets contained 800 and 200 samples respectively, drawn from Latin Hypercube Sampling (Section 3.3).

Table 4.1 summarizes the tested architectures, covering variations in depth (2–4 layers), width (64–1024 neurons), and structural motifs.

TABLE 4.1: Architecture configurations tested in the ablation study.

Architecture	**Layers**	**Skip**	**Params**	**Time (s)**
2-Layer Narrow	[128, 64]	No	42k	2.16
2-Layer Standard	[256, 128]	No	99k	2.61
3-Layer Uniform	[256, 256, 256]	No	167k	3.15
3-Layer Uniform+Residual	$[256^3]$	Yes	167k	3.30
3-Layer Tapering	[512, 256, 128]	No	295k	3.72
3-Layer Bottleneck (Shallow)	[256, 128, 256]	Yes	134k	2.98
3-Layer Bottleneck (Deep)	[512, 256, 512]	Yes	374k	4.48
3-Layer Adaptive	[256, 512, 256]	Yes	233k	3.51
4-Layer Deep Bottleneck	[256, 512, 1024, 512]	Yes	741k	5.12

Figure 4.2 shows validation R^2 scores, sorted by performance. All configurations achieve $R^2 > 0.998$. This confirms that the task is tractable across a wide range of designs. However, meaningful differences emerge at high precision. The three-layer uniform architecture with residual connections achieves the highest validation performance ($R^2 = 0.99959$).

Several consistent patterns emerge from the comparison. The uniform-width architectures outperform the tapering and bottleneck designs. This suggests that the smooth mapping from physical parameters to waveforms is best captured when the representational capacity remains consistent across layers. Residual connections further enhance the performance as they facilitate gradient flow and enable additive corrections to intermediate representations. Increasing depth beyond three layers does not yield additional gains and can introduce optimization difficulties, as observed for the four-layer deep bottleneck architecture.

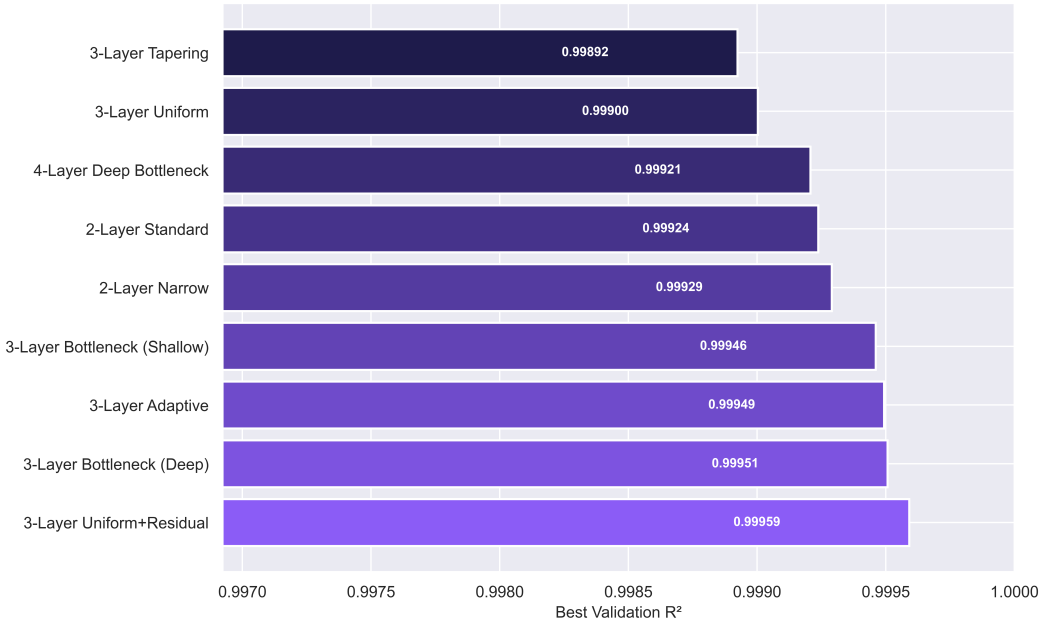

FIGURE 4.2: Validation R^2 comparison across nine NN architectures, demonstrating that the three-layer uniform architecture with residual connections achieves the highest prediction accuracy. Color coding reflects performance ranking, with better architectures shown in lighter shades.

TABLE 4.2: Computational efficiency comparison across architectures, showing training time per 100 epochs and total trainable parameters. The selected architecture achieves the best accuracy-to-cost ratio.

Architecture	**Training Time (s)**	**Parameters**	R^2
2-Layer Narrow	2.16	42,304	0.99929
2-Layer Standard	2.61	99,456	0.99924
3-Layer Uniform	3.15	167,040	0.99900
3-Layer Uniform+Residual	3.30	167,040	0.99959
3-Layer Tapering	3.72	295,040	0.99892
3-Layer Bottleneck (Shallow)	2.98	134,272	0.99946
3-Layer Bottleneck (Deep)	4.48	374,144	0.99951
3-Layer Adaptive	3.51	232,576	0.99949
4-Layer Deep Bottleneck	5.12	740,608	0.99921

Interestingly, model capacity alone does not determine performance. Even though architectures with substantially more parameters, such as the deep bottleneck variants, approach similar accuracy, they incur higher computational

cost without proportional benefit. This highlights the importance of architectural inductive biases aligned with the smooth structure of the target function.

From a practical perspective, computational efficiency plays an important role for downstream Bayesian inference. Although the training time is less critical than the inference speed in MCMC applications, the parameter count and architectural simplicity directly impact deployment efficiency. The three-layer uniform architecture with residual connections provides a good compromise, combining high accuracy with a relatively small parameter count of 167,040 and fast inference.

Chapter 5

Parameter Inference Implementation

This chapter describes how the emulator developed in the previous chapter is integrated into a Bayesian parameter inference method based on Markov Chain Monte Carlo (MCMC).

5.1 Bayesian Framework for Parameter Estimation

The general Bayesian framework has already been introduced in Section 2.5. The expressions below are restated here to fix notation for the concrete inference implementation used in this chapter.

Given an observed GW signal $\mathbf{d}$, Bayesian inference characterizes the posterior probability of source parameters $\boldsymbol{\theta}$. The posterior distribution is given by

$$p(\boldsymbol{\theta}|\mathbf{d}) = \frac{p(\mathbf{d}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathbf{d})}. \tag{5.1}$$

From the independence of the evidence $p(\mathbf{d})$ from $\boldsymbol{\theta}$ in parameter estimation problems, it follows that it can be neglected. This results in the unnormalized posterior distribution

$$p(\boldsymbol{\theta}|\mathbf{d}) \propto p(\mathbf{d}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}). \tag{5.2}$$

The three-dimensional parameter space includes the initial orbital eccentricity $e_0 \in [0.0, 0.15]$, defined at a separation of $r = 50\, r_{\mathrm{isco}}$, and the logarithmic total binary mass $\log_{10} M_{\mathrm{tot}} \in [8.0, 10.0]$, expressed in solar masses. In addition,

the stellar velocity dispersion $\sigma \in [100, 400]\,\mathrm{km/s}$ characterizes the galactic environment. The selected parameter ranges correspond to typical values for SMBHB systems contributing to the nanohertz GW band probed by PTAs.

As explained in Section 3.4, the logarithmic parameterization of the total mass provides approximately uniform coverage of signal morphologies, since characteristic GW frequencies scale inversely with mass.

The likelihood describes the agreement between the observed data $\mathbf{d}$ and the model prediction for a given parameter set $\boldsymbol{\theta}$. For independent Gaussian uncertainties (see Section 2.5), the corresponding log-likelihood is

$$\log p(\mathbf{d}|\boldsymbol{\theta}) = -\frac{1}{2}\sum_{i=1}^{N_f}\frac{(d_i - h_i(\boldsymbol{\theta}))^2}{\sigma_i^2} - \frac{1}{2}\sum_{i=1}^{N_f}\log\left(2\pi\sigma_i^2\right). \tag{5.3}$$

Here, d_i denotes the observed characteristic strain in frequency bin i, $h_i(\boldsymbol{\theta})$ is the model prediction, σ_i represents the measurement uncertainty, and N_f is the total number of frequency bins.

GW characteristic strain amplitudes span several orders of magnitude across the frequency band [4, 40]. To obtain approximately equal sensitivity to fractional deviations across frequency, the comparison is performed in logarithmic space:

$$\text{residual}_i = \log_{10}(d_i + \epsilon) - \log_{10}(h_i(\boldsymbol{\theta}) + \epsilon), \tag{5.4}$$

where $\epsilon = 10^{-30}$ prevents numerical issues for very small strain values.

The model prediction $h_i(\boldsymbol{\theta})$ is obtained from the NN GW emulator described in Chapter 4. All normalization steps and architectural decisions are fixed by the trained model and remain unchanged during inference.

Uniform prior distributions are assumed:

$$p(e_0) = \mathcal{U}(0.0, 0.15), \quad p(\log_{10} M_{\mathrm{tot}}) = \mathcal{U}(8.0, 10.0), \quad p(\sigma) = \mathcal{U}(100, 400)\,\mathrm{km/s}. \tag{5.5}$$

Parameter sets outside these intervals are assigned zero prior probability and are therefore excluded from the sampling process.

5.2 MCMC Sampling Strategy

The MCMC analysis uses the affine-invariant ensemble sampler `emcee` [76, 89], as introduced in Section 2.5. This sampler is well suited for correlated parameter spaces and enables efficient parallel exploration.

The so-called stretch move forms the core of the proposal mechanism. In this procedure, new parameter points are generated based on the current positions of other walkers in the ensemble. For walker j, the proposed position $\mathbf{Y}$ is given by

$$\mathbf{Y} = \mathbf{X}_k + Z\left(\mathbf{X}_j - \mathbf{X}_k\right), \tag{5.6}$$

where $\mathbf{X}_j$ is the current position of walker j, $\mathbf{X}_k$ is drawn from a complementary ensemble, and Z is sampled from

$$g(z) \propto \frac{1}{\sqrt{z}}, \quad z \in \left[\frac{1}{a}, a\right], \tag{5.7}$$

with scale parameter $a = 2$.

The Metropolis–Hastings acceptance probability is

$$\alpha = \min\left(1,\, Z^{n-1}\frac{p(\mathbf{Y}|\mathbf{d})}{p(\mathbf{X}_j|\mathbf{d})}\right), \tag{5.8}$$

where $n = 3$ is the dimensionality of the parameter space.

The sampler uses 50 walkers, corresponding to approximately seventeen times the parameter dimension. A burn-in phase of 3,000 steps per walker is discarded. The subsequent production phase consists of 5,000 steps per walker, yielding a total of $50 \times 5{,}000 = 250{,}000$ posterior samples.

Walkers are initialized according to

$$\boldsymbol{\theta}_j^{(0)} = \boldsymbol{\theta}_{\text{init}} + \boldsymbol{\epsilon}_j, \quad \boldsymbol{\epsilon}_j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{init}}), \tag{5.9}$$

with

$$\sigma_{e_0} = 0.01, \tag{5.10}$$

$$\sigma_M = 0.2, \tag{5.11}$$

$$\sigma_\sigma = 20 \text{ km/s}. \tag{5.12}$$

For validation studies with injected signals (Chapter 6), the initial estimate is set to the true parameter values. Initial positions are clipped to remain within the prior bounds.

The sampler is implemented using `emcee`. The trained PyTorch emulator and the associated `scikit-learn` normalization objects are loaded once and reused throughout the sampling run. Listing 5.1 shows the central likelihood evaluation.

```
def log_posterior(theta, f_obs, h_obs,
                  emulator, param_scaler, signal_scaler):
    if not (0.0 <= theta[0] <= 0.15 and
            8.0 <= theta[1] <= 10.0 and
            100 <= theta[2] <= 400):
        return -np.inf

    h_model = emulator_predict(emulator, theta,
                               param_scaler, signal_scaler)

    residuals = np.log10(h_obs + 1e-30) - np.log10(h_model + 1e
    -30)
    chi2 = np.sum((residuals / 0.01)**2)

    return -0.5 * chi2
```

Listing 5.1: Emulator-based likelihood evaluation

5.3 Convergence Diagnostics

Convergence diagnostics are used to verify that the MCMC samples provide a reliable representation of the target posterior distribution. Furthermore, trace plots are inspected for stationarity and mixing.

Acceptance fractions between 0.2 and 0.5 indicate an efficient exploration. The integrated autocorrelation times τ_i are estimated as

$$\tau_i = 1 + 2 \sum_{k=1}^{\infty} \rho_k^{(i)}. \tag{5.13}$$

The effective number of independent samples is

$$N_{\text{eff}} = \frac{N_{\text{walkers}}\, N_{\text{steps}}}{\tau_{\text{max}}}, \tag{5.14}$$

with $\tau_{\text{max}} = \max_i \tau_i$.

The Gelman–Rubin statistic [68, 90] is

$$\hat{R} = \sqrt{\frac{\text{Var}^+}{W}}, \tag{5.15}$$

where W denotes the within-chain variance and Var^+ the total variance estimate. Values $\hat{R} \lesssim 1.1$ indicate convergence.

5.4 Emulator Accuracy Considerations

Using an NN emulator introduces an approximation error in addition to statistical measurement uncertainty. As shown in Chapter 4, the emulator achieves $R^2 > 0.999$ with relative deviations below 1%.

Emulator-induced effects on the inference remain negligible as long as

$$\sigma_{\text{emulator}} \ll \sigma_{\text{measurement}}. \tag{5.16}$$

Validation with injected signals confirms that emulator-related deviations do not introduce relevant systematic effects within the considered parameter range. This is demonstrated by consistently high R^2 scores across the validation set and by direct waveform comparisons, which show no systematic deviations within the considered parameter range. These results indicate that emulator-induced errors remain subdominant compared to measurement uncertainties in the subsequent inference.

The complete inference pipeline is now established. Chapter 6 applies this framework to simulated observations.

Chapter 6

Results and Discussion

This chapter presents the results of the NN emulator and the MCMC parameter estimation based on it from Chapters 4 and 5.

6.1 Emulator Training and Validation

This section examines the performance of the NN emulator developed in Chapter 4.2. It considers training behavior, generalization across the parameter space, and the accuracy of the reconstructed waveforms with regard to their use in Bayesian inference.

A summary of the key performance indicators is shown in Table 6.1. The three-layer architecture with residual connections achieves high prediction accuracy with low computation time, proving to be well suited for use in MCMC-based inference methods. In particular, the high validation R^2 value and the low validation error demonstrate that the emulator reliably learns the mapping from physical parameters to GW spectra while maintaining good generalization to previously unseen configurations.

Figure 6.1 shows the progression of training and validation losses over the training period. Both curves decrease evenly and converge toward nearly identical final values. This indicates stable optimization and does not suggest significant overfitting.

The fact that training and validation loss do not diverge from each other also confirms the effectiveness of the regularization measures described in Chapter 4. In particular, weight decay regularization, layer normalization, and the

TABLE 6.1: Summary of NN emulator performance across training, validation and inference metrics, demonstrating the system achieves high accuracy with substantial computational speedup.

Metric	**Value**
Architecture	
Configuration	3-Layer Uniform+Residual
Hidden Dimensions	[256, 256, 256]
Skip Connections	Yes
Dropout Rate	0.0
Total Parameters	167,040
Training Performance	
Training Time (100 epochs)	3.3 seconds
Final Training Loss	1.2×10^{-4}
Final Validation Loss	1.3×10^{-4}
Validation R^2	0.99959
Samples with $R^2 > 0.995$	99%
Computational Efficiency	
Inference Time (per sample)	$\sim$1 ms
SMBHBpy Time (per sample)	$\mathcal{O}(10^{-1})$–$\mathcal{O}(10^{1})$ s
Speedup Factor	$\mathcal{O}(10^{2})$–$\mathcal{O}(10^{3})$
MCMC Time with Emulator (250k evals)	4 minutes
MCMC Time without Emulator (250k evals)	2.1 hours
Parameter Recovery (MCMC)	
Gelman–Rubin $\hat{R}$	< 1.01 (all params)
Effective Sample Size	> 1000 (all params)
Parameter Recovery	Within 1σ
Median Relative Error	$< 1.5\%$
Systematic Uncertainties	
Environment Comparison ($\Delta_{\mathrm{Vac-CDM}}$)	$< 5\%$ (all params)
Emulator Error (log-strain)	$< 1\%$
Worst-Case R^2 (1st percentile)	0.995

absence of stochastic dropout during training contribute to the network learning generalizable relationships rather than memorizing individual training examples.

In addition to accuracy, computational efficiency is a key requirement for practical Bayesian inference. Figure 6.2 compares the runtime of direct SMBHBpy simulations with NN-based waveform generation, evaluated both sequentially and in a vectorized setting. Since the reference simulations considered here are computationally lightweight, the reported speedup should be interpreted as a

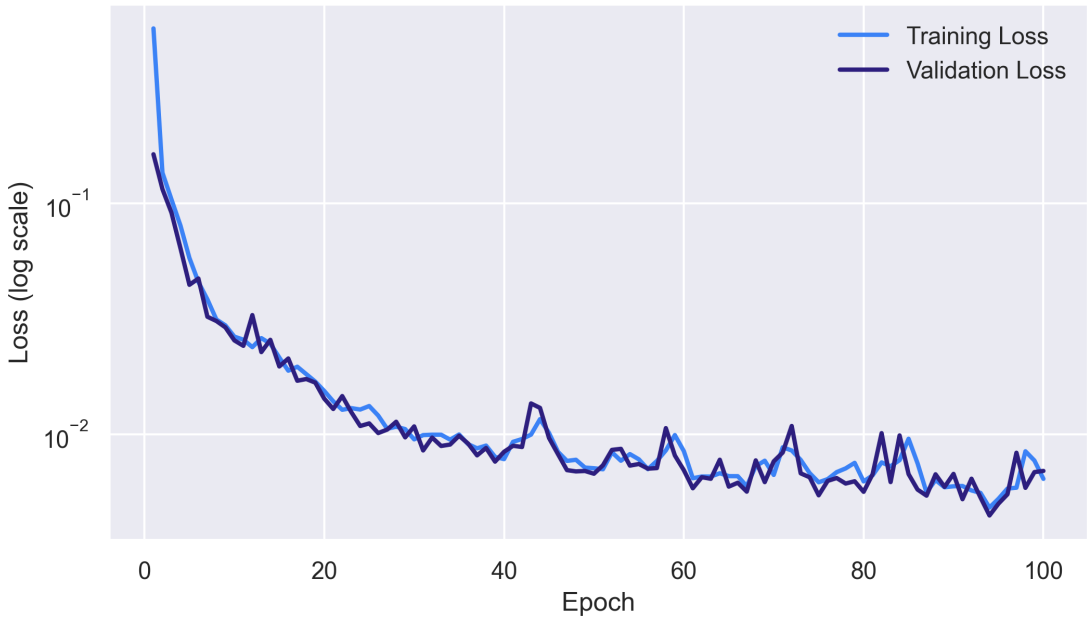

FIGURE 6.1: Training dynamics of the NN emulator showing training and validation loss versus epoch. The smooth convergence of both curves indicates successful learning without overfitting, and the final validation loss corresponds to sub-percent relative errors in log-strain space.

conservative lower bound. In practice, SMBHBpy runtimes vary strongly with numerical accuracy settings and the sampled parameter region. In our tests, individual evolutions ranged from $\mathcal{O}(10^{-1})$ s to several seconds or minutes per waveform, while emulator inference remained at $\sim$1 ms. Hence, for more complex source models, the relative acceleration achieved through NN-based emulation is expected to increase substantially.

Importantly, the inference time of the trained NN emulator is independent of the underlying physical complexity of the reference model, as it depends only on the fixed network architecture. In addition to the loss values, it is also important to verify whether the emulator performs uniformly across the parameter space. Figure 6.3 shows the distribution of per-sample R^2 values evaluated on the independent validation dataset. The distribution is strongly concentrated near unity: the vast majority of samples achieve $R^2 > 0.998$, while fewer than 1% fall below $R^2 = 0.995$. The absence of a pronounced tail toward lower R^2 values indicates that no systematic degradation occurs in specific regions of the parameter space.

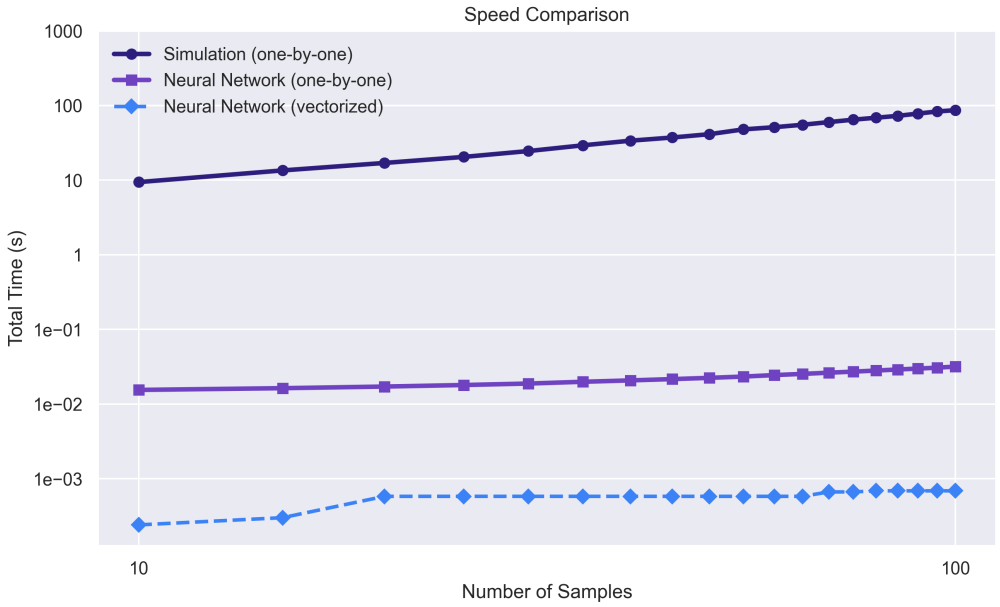

FIGURE 6.2: Runtime comparison between direct SMBHBpy simulations and NN-based waveform generation. The NN emulator achieves substantial speedups, which are further enhanced when evaluating multiple samples in a vectorized manner.

Such uniform reconstruction accuracy is particularly important for Bayesian inference, where the likelihood is evaluated repeatedly at arbitrary parameter combinations within the prior volume. Location-dependent emulator errors could otherwise introduce artificial distortions in the inferred posterior distributions. The narrow R^2 distribution observed here suggests that such effects are negligible within the validated parameter range.

While aggregated performance indicators provide a compact overview of model quality, direct analysis of reconstructed waveforms is essential for evaluating the physical reliability of the emulator. Figure 6.4 presents representative comparisons between characteristic strain spectra generated with SMBHBpy and the corresponding predictions of the trained NN emulator. The examples span different regions of the parameter space and include variations in initial eccentricity, total mass, and velocity dispersion.

In all cases shown, the emulator reproduces both the global amplitude scaling and the detailed frequency-dependent structure of the spectra with high accuracy. In particular, the expected power-law behavior at high frequencies and spectral modifications due to environmental effects are consistently captured.

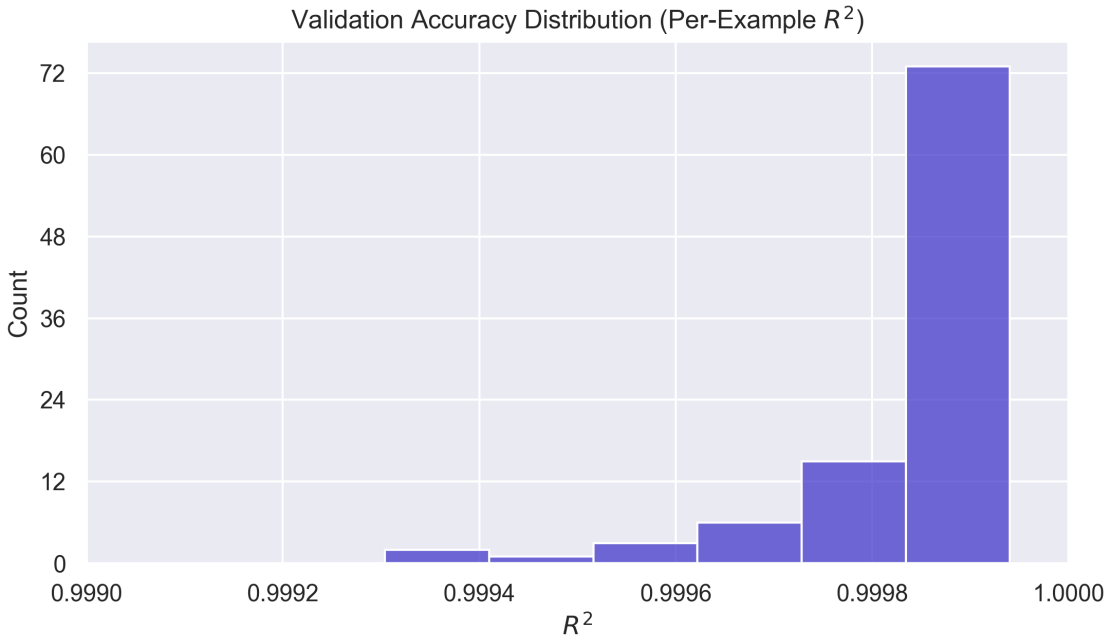

FIGURE 6.3: Distribution of per-sample validation R^2 values demonstrating robust and consistent waveform reconstruction accuracy across the validation dataset, with fewer than 1% of samples achieving $R^2 < 0.995$.

The remaining deviations are small and show no systematic frequency dependence, suggesting that the emulator reliably approximates the underlying physical mapping.

Taken together, the results demonstrate that the NN emulator combines high reconstruction accuracy with a substantial reduction in computational cost, without introducing systematic errors across the validated parameter space. It fulfills the necessary requirements for its use as a surrogate model in Bayesian parameter inference, which is examined in the following section.

6.2 MCMC Parameter Recovery

To validate the full Bayesian inference pipeline, injection–recovery tests are performed using synthetic GW signals generated with `SMBHBpy`. The injected signal is noise-free and chosen from the interior of the prior volume to avoid boundary effects, with fiducial parameters $e_0 = 0.08$, $\log_{10}(M_{\rm tot}/M_\odot) = 9.0$, and $\sigma = 250\,\mathrm{km/s}$.

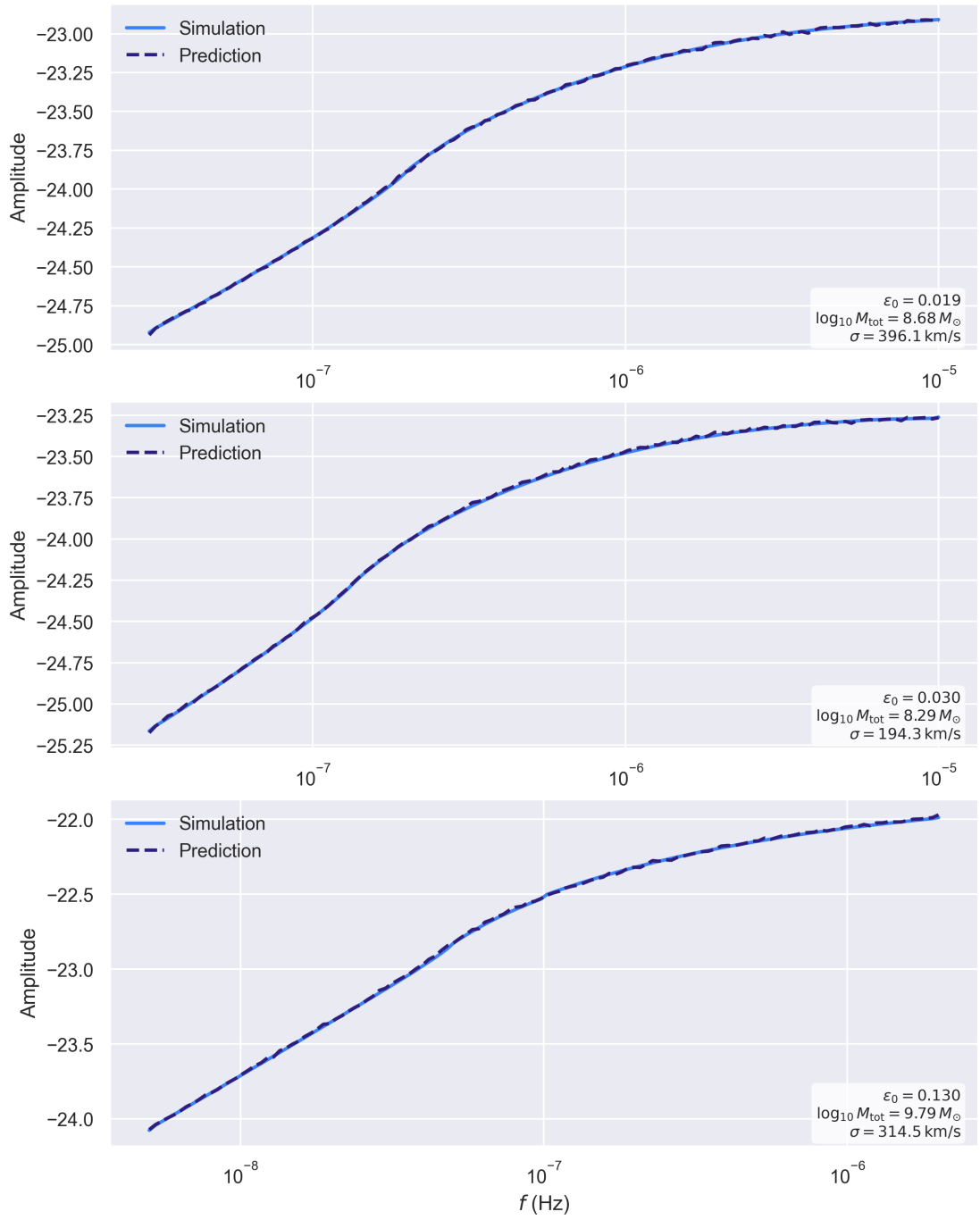

FIGURE 6.4: Comparison between SMBHBpy-generated characteristic strain spectra and NN predictions for validation samples spanning the parameter space.

Bayesian parameter inference is carried out using an MCMC sampler. The analysis is performed both with direct waveform evaluations using `SMBHBpy`

and with waveform predictions obtained from the trained NN emulator. In both cases, identical sampler settings, priors, and likelihood definitions are employed.

Within numerical precision, the resulting posterior distributions obtained with and without the emulator are indistinguishable. For this reason, only a single corner plot is shown for each environment, as separate visualizations would be redundant. This direct agreement demonstrates that replacing the physical waveform model with the NN emulator does not alter the inferred posterior distributions.

The sampling behavior is stable, with acceptance fractions within the expected range for the affine-invariant ensemble sampler and convergence diagnostics indicating well-mixed chains. For all inferred parameters, the posterior medians are consistent with the injected values within 1σ, with a median relative error below 1.5%.

In the vacuum scenario, environmental dynamical friction is absent. Consequently, the velocity dispersion parameter σ is only weakly constrained and effectively acts as a dummy variable. The absence of spurious correlations confirms that the emulator does not introduce artificial sensitivity to environmental parameters when the underlying physical model implies no such dependence.

The accurate recovery of all injected parameters in both environments confirms that the NN emulator can be used as a drop-in replacement for direct waveform generation in MCMC-based parameter inference. Within the explored parameter regime, emulator-induced errors remain negligible compared to statistical uncertainties, and no systematic biases are introduced by the surrogate model.

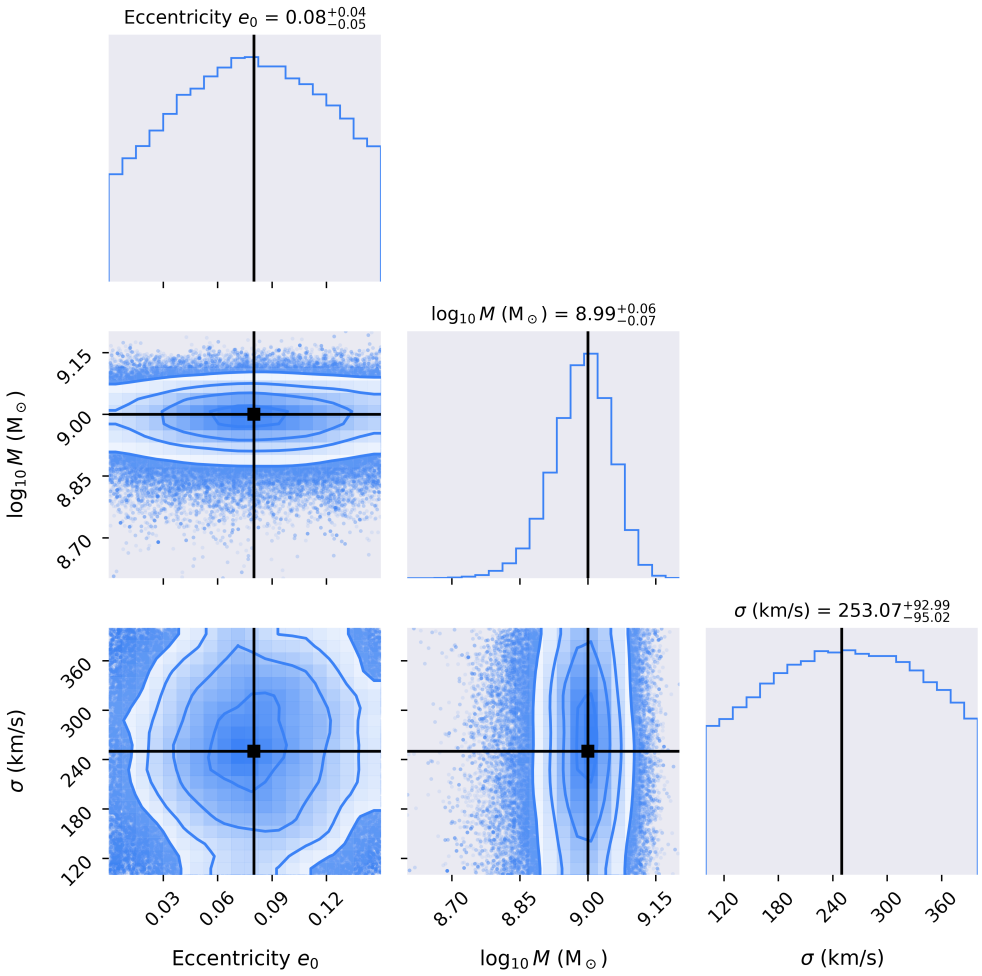

FIGURE 6.5: Corner plot showing posterior distributions from the MCMC analysis in a vacuum environment. The injected true parameter values are indicated by black lines. Identical posterior distributions are obtained with direct waveform evaluation and with the NN emulator, and therefore only a single plot is shown.

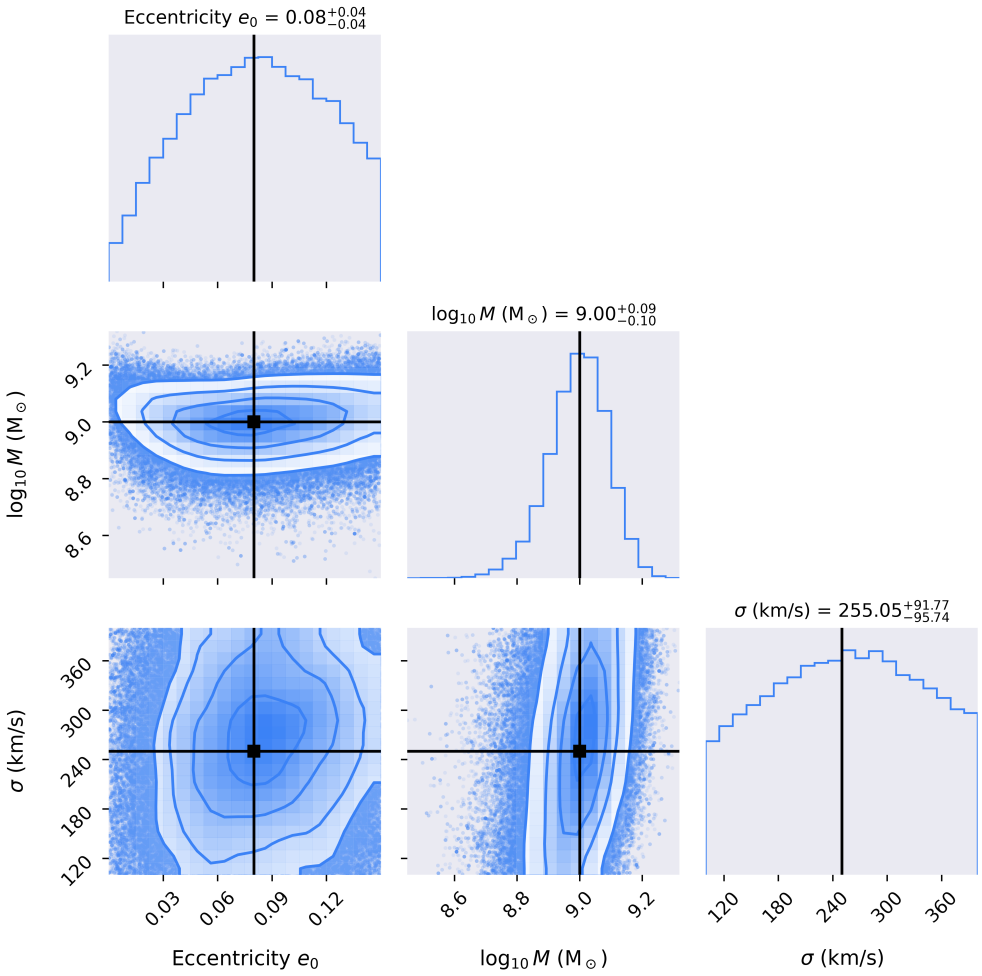

FIGURE 6.6: Corner plot showing posterior distributions from the MCMC analysis in a CDM environment. The injected true parameter values are indicated by black lines. As in the vacuum case, emulator-based and direct waveform inference yield indistinguishable posteriors, and only a single plot is shown.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

This work demonstrates that NNs as surrogate models represent a practical and powerful method for accelerating GW signal modeling for SMBHB systems in dense astrophysical environments. By training on physically motivated reference simulations from SMBHBpy, the developed emulator is able to reproduce characteristic GW strain spectra with a high accuracy across a multidimensional parameter space that includes the initial eccentricity, total mass of the binary system, and stellar velocity dispersion.

The achieved reconstruction accuracy, evidenced by validation R^2 values close to one and direct waveform comparisons in the frequency domain, shows that the emulator reliably learns the relevant physical dependencies. In particular, the absence of systematic accuracy losses within the validated parameter range is an essential prerequisite for the use of the model in inference applications.

A key contribution of this work is the substantial reduction in computational effort through NN-based emulation. While direct simulations with SMBHBpy require seconds to minutes per waveform due to the underlying physical modeling, corresponding predictions can be generated within milliseconds using the trained emulator. This acceleration eliminates a central bottleneck in Bayesian GW parameter inference and enables analyses that would be practically impossible with purely simulation-based approaches.

The integration of the emulator into an MCMC-based inference pipeline further demonstrates the practical applicability of the approach. The injection

recovery tests show that NN-generated waveforms can be used in parameter inference in a manner consistent with direct SMBHBpy simulations, without introducing significant distortions, and that the resulting posteriors remain physically plausible. Although the inference study presented here does not constitute a complete astrophysical analysis, it demonstrates the consistency and robustness of the method within established GW analysis workflows.

In summary, this work bridges the gap between physically detailed GW source simulations and the computational requirements of Bayesian parameter inference. While SMBHBs serve as a concrete application case, the methodology developed here is not restricted to this specific source class. By combining physically motivated modeling with a substantial reduction in waveform evaluation time, the framework enables systematic studies of extended source models and environmental effects that would be computationally prohibitive with purely simulation-based approaches. This makes it well suited for exploring deviations from standard assumptions in GW source modeling within a consistent inference framework.

7.2 Future Work

While our work focuses on demonstrating the feasibility and evaluating the performance of NN-based waveform emulation for SMBHB systems in CDM environments, several obvious extensions arise. An important extension would be the explicit propagation of emulator-induced uncertainties into the Bayesian likelihood, for example through probabilistic surrogate models, in order to avoid potential biases in parameter inference [91].

An immediate next step is Bayesian model selection between different environmental assumptions, such as vacuum, CDM, and SIDM scenarios. By comparing Bayesian evidence, it would be possible to quantify the extent to which GW observations can distinguish between different DM models. Previous studies indicate that environmental effects predominantly influence the low-frequency inspiral regime, which implies that high signal-to-noise ratio observations may be sensitive to such deviations from vacuum evolution [19, 36].

For realistic applications, it is also necessary to incorporate detector-specific noise models. While the validation studies in this work are deliberately limited

to noise-free signals, intrinsic pulsar noise, correlated noise processes, and network geometries would have to be taken into account for PTA analyses. For space-based detectors such as LISA, instrumental and confusion noise sources also play a central role.

Another promising area of application lies in population-based analyses, particularly in connection with the SGWB in the nanohertz range. The efficient generation of large waveform ensembles is a decisive bottleneck here, which NN-based surrogates can significantly alleviate.

Furthermore, the approach can be extended to space-based observations with LISA [10]. The long-duration SMBHB signals with high frequency resolution expected there place high demands on simulation efficiency, especially when environmental influences are taken into account.

Finally, the expansion of the parameter space, for example, to include mass ratios, spins, orbital inclinations, or alternative environment models, offers great potential for further investigations. Even though this requires larger training datasets and careful validation, the developed modular framework is well suited to accommodate such extensions.

Overall, this work demonstrates that NN-based waveform emulation is a scalable and physically consistent approach to accelerating GW data analysis. With increasing complexity of source models and growing data volumes from future observatories, surrogate models will play a central role in the systematic exploration of the GW source population.

Acknowledgements

I would like to begin by thanking my professor Laura and my supervisor Alessio for their openness, trust and support throughout this thesis. I am deeply grateful for the opportunity to pursue an interdisciplinary collaboration that allowed me to explore my own ideas and connect different scientific perspectives as well as for the constructive discussions and guidance along the way. I would also like to thank Laura for encouraging me to engage with the broader research community, which led to the collaboration with Alessio that was shaped by a shared enthusiasm for the project and strongly influenced this thesis.

The members of the DMGW research group provided an inspiring academic environment and many valuable discussions. I am also thankful to all fellow physics students and colleagues I have met over the years, as the many conversations, shared challenges and moments of mutual curiosity have been an important part of my academic journey.

I would like to express my sincere gratitude to my parents for their continuous support and for enabling me to pursue my education. From an early age, they fostered my curiosity and interest in physics and supported me to follow this path. My mother encouraged me to approach challenges creatively and to search for solutions beyond conventional ones, while my father profoundly shaped the way I think about questions and ideas, teaching me to reason carefully, deeply, and with intellectual honesty. His influence continues to shape how I see the world, even though he is no longer here.

Finally, my deepest gratitude goes to my fiancé, Gabriel. His patience, encouragement, and steadfast confidence in me have meant more to me than I can put into words. He supported me through difficult times and shared in moments of success along the way. His presence has been a constant source of strength along this journey, and I am fortunate to share it with him.

Bibliography

[1] The LIGO Scientific Collaboration and the Virgo Collaboration. "Observation of Gravitational Waves from a Binary Black Hole Merger". In: *Physical Review Letters* 116.6 (Feb. 2016). arXiv:1602.03837 [gr-qc]. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.116.061102. URL: http://arxiv.org/abs/1602.03837 (visited on 07/18/2025).

[2] The LIGO Scientific Collaboration et al. "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run". In: *Physical Review X* 13.4 (Dec. 2023). arXiv:2111.03606 [gr-qc]. ISSN: 2160-3308. DOI: 10.1103/PhysRevX.13.041039. URL: http://arxiv.org/abs/2111.03606 (visited on 07/18/2025).

[3] The LIGO Scientific Collaboration and The Virgo Collaboration. "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral". In: *Physical Review Letters* 119.16 (Oct. 2017). arXiv:1710.05832 [gr-qc], p. 161101. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.119.161101. URL: http://arxiv.org/abs/1710.05832 (visited on 09/14/2025).

[4] Michele Maggiore. *Gravitational Waves: Volume 1: Theory and Experiments*. en. 1st ed. Oxford University PressOxford, Oct. 2007. ISBN: 978-0-19-857074-5 978-0-19-171766-6. DOI: 10.1093/acprof:oso/9780198570745.001.0001. URL: https://academic.oup.com/book/41655 (visited on 09/21/2025).

[5] M. Bailes et al. "Gravitational-wave physics and astronomy in the 2020s and 2030s". en. In: *Nature Reviews Physics* 3.5 (May 2021), pp. 344–366. ISSN: 2522-5820. DOI: 10.1038/s42254-021-00303-8. URL: https://www.nature.com/articles/s42254-021-00303-8 (visited on 12/13/2025).

[6] Peter Couvares et al. *Gravitational Wave Data Analysis: Computing Challenges in the 3G Era*. arXiv:2111.06987 [gr-qc]. Nov. 2021. DOI: 10.48550/

arXiv.2111.06987. URL: http://arxiv.org/abs/2111.06987 (visited on 09/21/2025).

[7] Alberto Sesana. "The promise of multi-band gravitational wave astronomy". In: *Physical Review Letters* 116.23 (June 2016). arXiv:1602.06951 [gr-qc], p. 231102. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.116.231102. URL: http://arxiv.org/abs/1602.06951 (visited on 09/21/2025).

[8] Nelson Christensen. "Stochastic Gravitational Wave Backgrounds". In: *Reports on Progress in Physics* 82.1 (Jan. 2019). arXiv:1811.08797 [gr-qc], p. 016903. ISSN: 0034-4885, 1361-6633. DOI: 10.1088/1361-6633/aae6b5. URL: http://arxiv.org/abs/1811.08797 (visited on 09/21/2025).

[9] ESA (European Space Agency). *The spectrum of gravitational waves*. Sept. 2021. URL: https://www.esa.int/ESA_Multimedia/Images/2021/09/The_spectrum_of_gravitational_waves.

[10] Pau Amaro-Seoane et al. *Laser Interferometer Space Antenna*. arXiv:1702.00786 [astro-ph]. Feb. 2017. DOI: 10.48550/arXiv.1702.00786. URL: http://arxiv.org/abs/1702.00786 (visited on 09/15/2025).

[11] Michele Maggiore et al. "Science Case for the Einstein Telescope". In: *Journal of Cosmology and Astroparticle Physics* 2020.03 (Mar. 2020). arXiv:1912.02622 [astro-ph], pp. 050–050. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2020/03/050. URL: http://arxiv.org/abs/1912.02622 (visited on 09/15/2025).

[12] Matthew Evans et al. *A Horizon Study for Cosmic Explorer: Science, Observatories, and Community*. arXiv:2109.09882 [astro-ph]. Oct. 2021. DOI: 10.48550/arXiv.2109.09882. URL: http://arxiv.org/abs/2109.09882 (visited on 09/15/2025).

[13] B. P. Abbott et al. "Exploring the Sensitivity of Next Generation Gravitational Wave Detectors". In: (2016). DOI: 10.48550/ARXIV.1607.08697. URL: https://arxiv.org/abs/1607.08697 (visited on 12/13/2025).

[14] Gabriella Agazie et al. "The NANOGrav 15-year Data Set: Evidence for a Gravitational-Wave Background". In: *The Astrophysical Journal Letters* 951.1 (July 2023). arXiv:2306.16213 [astro-ph], p. L8. ISSN: 2041-8205, 2041-8213. DOI: 10.3847/2041-8213/acdac6. URL: http://arxiv.org/abs/2306.16213 (visited on 07/15/2025).

[15] J. Antoniadis et al. "The second data release from the European Pulsar Timing Array III. Search for gravitational wave signals". In: *Astronomy & Astrophysics* 678 (Oct. 2023). arXiv:2306.16214 [astro-ph], A50. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202346844. URL: http://arxiv.org/abs/2306.16214 (visited on 09/21/2025).

[16] Daniel J. Reardon et al. "Search for an isotropic gravitational-wave background with the Parkes Pulsar Timing Array". In: *The Astrophysical Journal Letters* 951.1 (July 2023). arXiv:2306.16215 [astro-ph], p. L6. ISSN: 2041-8205, 2041-8213. DOI: 10.3847/2041-8213/acdd02. URL: http://arxiv.org/abs/2306.16215 (visited on 07/15/2025).

[17] Planck Collaboration et al. "Planck 2018 results. VI. Cosmological parameters". In: *Astronomy & Astrophysics* 641 (Sept. 2020). arXiv:1807.06209 [astro-ph], A6. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201833910. URL: http://arxiv.org/abs/1807.06209 (visited on 09/24/2025).

[18] Paolo Gondolo and Joseph Silk. "Dark matter annihilation at the galactic center". In: *Physical Review Letters* 83.9 (Aug. 1999). arXiv:astro-ph/9906391, pp. 1719–1722. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.83.1719. URL: http://arxiv.org/abs/astro-ph/9906391 (visited on 09/24/2025).

[19] Bradley J. Kavanagh et al. "Detecting dark matter around black holes with gravitational waves: Effects of dark-matter dynamics on the gravitational waveform". In: *Physical Review D* 102.8 (Oct. 2020). arXiv:2002.12811 [gr-qc], p. 083006. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.102.083006. URL: http://arxiv.org/abs/2002.12811 (visited on 11/02/2025).

[20] Eric Thrane and Colm Talbot. "An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models". In: *Publications of the Astronomical Society of Australia* 36 (2019). arXiv:1809.02293 [astro-ph], e010. ISSN: 1323-3580, 1448-6083. DOI: 10.1017/pasa.2019.2. URL: http://arxiv.org/abs/1809.02293 (visited on 09/23/2025).

[21] Gregory Ashton and Colm Talbot. "Bilby-MCMC: An MCMC sampler for gravitational-wave inference". In: *Monthly Notices of the Royal Astronomical Society* 507.2 (Aug. 2021). arXiv:2106.08730 [gr-qc], pp. 2037–2051.

ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stab2236. URL: http://arxiv.org/abs/2106.08730 (visited on 09/23/2025).

[22] Geraint Pratten et al. "Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes". en. In: *Physical Review D* 103.10 (May 2021), p. 104056. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.103.104056. URL: https://link.aps.org/doi/10.1103/PhysRevD.103.104056 (visited on 09/23/2025).

[23] Lorenz Zwick et al. *A novel category of environmental effect in gravitational waves from binaries perturbed by periodic forces*. arXiv:2405.05698 [gr-qc]. Oct. 2024. DOI: 10.48550/arXiv.2405.05698. URL: http://arxiv.org/abs/2405.05698 (visited on 09/23/2025).

[24] Albert Einstein. "Näherungsweise Integration der Feldgleichungen der Gravitation". In: *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* (Jan. 1916). ADS Bibcode: 1916SPAW.....688E, pp. 688–696. URL: https://ui.adsabs.harvard.edu/abs/1916SPAW.....688E (visited on 10/31/2025).

[25] Luc Blanchet et al. "Gravitational Radiation from Inspiralling Compact Binaries Completed at the Third Post-Newtonian Order". en. In: *Physical Review Letters* 93.9 (Aug. 2004), p. 091101. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.93.091101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.93.091101 (visited on 10/31/2025).

[26] Curt Cutler and Éanna E. Flanagan. "Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform?" en. In: *Physical Review D* 49.6 (Mar. 1994), pp. 2658–2697. ISSN: 0556-2821. DOI: 10.1103/PhysRevD.49.2658. URL: https://link.aps.org/doi/10.1103/PhysRevD.49.2658 (visited on 10/31/2025).

[27] B. S. Sathyaprakash and B. F. Schutz. "Physics, Astrophysics and Cosmology with Gravitational Waves". In: *Living Reviews in Relativity* 12.1 (Dec. 2009). arXiv:0903.0338 [gr-qc], p. 2. ISSN: 2367-3613, 1433-8351. DOI: 10.12942/lrr-2009-2. URL: http://arxiv.org/abs/0903.0338 (visited on 10/31/2025).

[28] P. C. Peters and J. Mathews. "Gravitational Radiation from Point Masses in a Keplerian Orbit". en. In: *Physical Review* 131.1 (July 1963), pp. 435–

440. ISSN: 0031-899X. DOI: 10.1103/PhysRev.131.435. URL: https://link.aps.org/doi/10.1103/PhysRev.131.435 (visited on 10/31/2025).

[29] P. C. Peters. "Gravitational Radiation and the Motion of Two Point Masses". en. In: *Physical Review* 136.4B (Nov. 1964), B1224–B1232. ISSN: 0031-899X. DOI: 10.1103/PhysRev.136.B1224. URL: https://link.aps.org/doi/10.1103/PhysRev.136.B1224 (visited on 10/31/2025).

[30] Heng Xu et al. "Searching for the nano-Hertz stochastic gravitational wave background with the Chinese Pulsar Timing Array Data Release I". In: *Research in Astronomy and Astrophysics* 23.7 (July 2023). arXiv:2306.16216 [astro-ph], p. 075024. ISSN: 1674-4527. DOI: 10.1088/1674-4527/acdfa5. URL: http://arxiv.org/abs/2306.16216 (visited on 07/15/2025).

[31] R. W. Hellings and G. S. Downs. "Upper limits on the isotropic gravitational radiation background from pulsar timing analysis". en. In: *The Astrophysical Journal* 265 (Feb. 1983), p. L39. ISSN: 0004-637X, 1538-4357. DOI: 10.1086/183954. URL: http://adsabs.harvard.edu/doi/10.1086/183954 (visited on 10/31/2025).

[32] A. Sesana. "Insights on the astrophysics of supermassive black hole binaries from pulsar timing observations". In: *Classical and Quantum Gravity* 30.22 (Nov. 2013). arXiv:1307.2600 [astro-ph], p. 224014. ISSN: 0264-9381, 1361-6382. DOI: 10.1088/0264-9381/30/22/224014. URL: http://arxiv.org/abs/1307.2600 (visited on 07/18/2025).

[33] Sarah Burke-Spolaor et al. "The astrophysics of nanohertz gravitational waves". en. In: *The Astronomy and Astrophysics Review* 27.1 (Dec. 2019), p. 5. ISSN: 0935-4956, 1432-0754. DOI: 10.1007/s00159-019-0115-7. URL: http://link.springer.com/10.1007/s00159-019-0115-7 (visited on 10/31/2025).

[34] M. C. Begelman, R. D. Blandford, and M. J. Rees. "Massive black hole binaries in active galactic nuclei". en. In: *Nature* 287.5780 (Sept. 1980), pp. 307–309. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/287307a0. URL: https://www.nature.com/articles/287307a0 (visited on 10/31/2025).

[35] Miloš Milosavljević and David Merritt. "Long-Term Evolution of Massive Black Hole Binaries". en. In: *The Astrophysical Journal* 596.2 (Oct. 2003), pp. 860–878. ISSN: 0004-637X, 1538-4357. DOI: 10.1086/378086. URL: https://iopscience.iop.org/article/10.1086/378086 (visited on 10/31/2025).

[36] A. Sesana. "Systematic investigation of the expected gravitational wave signal from supermassive black hole binaries in the pulsar timing band". en. In: *Monthly Notices of the Royal Astronomical Society: Letters* 433.1 (July 2013), pp. L1–L5. ISSN: 1745-3933, 1745-3925. DOI: 10.1093/mnrasl/slt034. URL: https://academic.oup.com/mnrasl/article/433/1/L1/954113 (visited on 10/31/2025).

[37] Luke Zoltan Kelley, Laura Blecha, and Lars Hernquist. "Massive black hole binary mergers in dynamical galactic environments". en. In: *Monthly Notices of the Royal Astronomical Society* 464.3 (Jan. 2017), pp. 3131–3157. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stw2452. URL: https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stw2452 (visited on 10/31/2025).

[38] Pablo A. Rosado, Alberto Sesana, and Jonathan Gair. "Expected properties of the first gravitational wave signal detected with pulsar timing arrays". In: *Monthly Notices of the Royal Astronomical Society* 451.3 (June 2015). arXiv:1503.04803 [astro-ph], pp. 2417–2433. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stv1098. URL: http://arxiv.org/abs/1503.04803 (visited on 10/31/2025).

[39] H Middleton et al. "Massive black hole binary systems and the NANOGrav 12.5 yr results". en. In: *Monthly Notices of the Royal Astronomical Society: Letters* 502.1 (Feb. 2021), pp. L99–L103. ISSN: 1745-3925, 1745-3933. DOI: 10.1093/mnrasl/slab008. URL: https://academic.oup.com/mnrasl/article/502/1/L99/6123732 (visited on 10/31/2025).

[40] E. S. Phinney. *A Practical Theorem on Gravitational Wave Backgrounds*. arXiv:astro-ph/0108028. Aug. 2001. DOI: 10.48550/arXiv.astro-ph/0108028. URL: http://arxiv.org/abs/astro-ph/0108028 (visited on 10/31/2025).

[41] Gianfranco Bertone, Dan Hooper, and Joseph Silk. "Particle dark matter: evidence, candidates and constraints". en. In: *Physics Reports* 405.5-6 (Jan. 2005), pp. 279–390. ISSN: 03701573. DOI: 10.1016/j.physrep.2004.08.031. URL: https://linkinghub.elsevier.com/retrieve/pii/S0370157304003515 (visited on 12/13/2025).

[42] Douglas Clowe et al. "A Direct Empirical Proof of the Existence of Dark Matter". en. In: *The Astrophysical Journal* 648.2 (Sept. 2006), pp. L109–L113. ISSN: 0004-637X, 1538-4357. DOI: 10.1086/508162. URL: https://iopscience.iop.org/article/10.1086/508162 (visited on 12/13/2025).

[43] James S. Bullock and Michael Boylan-Kolchin. "Small-Scale Challenges to the CDM Paradigm". en. In: *Annual Review of Astronomy and Astrophysics* 55.1 (Aug. 2017), pp. 343–387. ISSN: 0066-4146, 1545-4282. DOI: 10.1146/annurev-astro-091916-055313. URL: https://www.annualreviews.org/doi/10.1146/annurev-astro-091916-055313 (visited on 12/13/2025).

[44] Ben Moore. "Evidence against dissipation-less dark matter from observations of galaxy haloes". en. In: *Nature* 370.6491 (Aug. 1994), pp. 629–631. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/370629a0. URL: https://www.nature.com/articles/370629a0 (visited on 01/18/2026).

[45] David N. Spergel and Paul J. Steinhardt. "Observational Evidence for Self-Interacting Cold Dark Matter". en. In: *Physical Review Letters* 84.17 (Apr. 2000), pp. 3760–3763. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.84.3760. URL: https://link.aps.org/doi/10.1103/PhysRevLett.84.3760 (visited on 12/13/2025).

[46] Manoj Kaplinghat, Sean Tulin, and Hai-Bo Yu. "Dark Matter Halos as Particle Colliders: Unified Solution to Small-Scale Structure Puzzles from Dwarfs to Clusters". en. In: *Physical Review Letters* 116.4 (Jan. 2016), p. 041302. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.116.041302. URL: https://link.aps.org/doi/10.1103/PhysRevLett.116.041302 (visited on 12/13/2025).

[47] Milos Milosavljevic and David Merritt. "The Final Parsec Problem". In: *AIP Conference Proceedings*. Vol. 686. arXiv:astro-ph/0212270. 2003, pp. 201–210. DOI: 10.1063/1.1629432. URL: http://arxiv.org/abs/astro-ph/0212270 (visited on 11/02/2025).

[48] Piero Ullio, HongSheng Zhao, and Marc Kamionkowski. "A Dark-Matter Spike at the Galactic Center?" In: *Physical Review D* 64.4 (July 2001). arXiv:astro-ph/0101481, p. 043504. ISSN: 0556-2821, 1089-4918. DOI: 10.1103/PhysRevD.64.043504. URL: http://arxiv.org/abs/astro-ph/0101481 (visited on 11/02/2025).

[49] Stuart L. Shapiro and Jessie Shelton. "Weak annihilation cusp inside the dark matter spike about a black hole". In: *Physical Review D* 93.12 (June 2016). arXiv:1606.01248 [astro-ph], p. 123510. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.93.123510. URL: http://arxiv.org/abs/1606.01248 (visited on 11/02/2025).

[50] David Merritt. "Evolution of the Dark Matter Distribution at the Galactic Center". In: *Physical Review Letters* 92.20 (May 2004). arXiv:astro-ph/0311594, p. 201304. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.92.201304. URL: http://arxiv.org/abs/astro-ph/0311594 (visited on 11/02/2025).

[51] Kazunari Eda et al. "Gravitational waves as a probe of dark matter mini-spikes". In: *Physical Review D* 91.4 (Feb. 2015). arXiv:1408.3534 [gr-qc], p. 044045. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.91.044045. URL: http://arxiv.org/abs/1408.3534 (visited on 11/02/2025).

[52] Zhao-Qiang Shen et al. "Dark Matter Spike surrounding Supermassive Black Holes Binary and the Nanohertz Stochastic Gravitational Wave Background". In: *Physics of the Dark Universe* 49 (Sept. 2025). arXiv:2306.17143 [astro-ph], p. 102004. ISSN: 2212-6864. DOI: 10.1016/j.dark.2025.102004. URL: http://arxiv.org/abs/2306.17143 (visited on 07/18/2025).

[53] Monica Colpi. "Massive binary black holes in galactic nuclei and their path to coalescence". In: *Space Science Reviews* 183.1-4 (Sept. 2014). arXiv:1407.3102 [astro-ph], pp. 189–221. ISSN: 0038-6308, 1572-9672. DOI: 10.1007/s11214-014-0067-1. URL: http://arxiv.org/abs/1407.3102 (visited on 11/02/2025).

[54] Lucas Lombriser and Andy Taylor. "Breaking a Dark Degeneracy with Gravitational Waves". In: *Journal of Cosmology and Astroparticle Physics* 2016.03 (Mar. 2016). arXiv:1509.08458 [astro-ph], pp. 031–031. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2016/03/031. URL: http://arxiv.org/abs/1509.08458 (visited on 11/16/2025).

[55] Stephen R. Taylor, Joseph Simon, and Laura Sampson. "Constraints On The Dynamical Environments Of Supermassive Black-hole Binaries Using Pulsar-timing Arrays". In: *Physical Review Letters* 118.18 (May 2017). arXiv:1612.02817 [astro-ph]. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.118.181102. URL: http://arxiv.org/abs/1612.02817 (visited on 07/18/2025).

[56] F. Rosenblatt. "The perceptron: A probabilistic model for information storage and organization in the brain." en. In: *Psychological Review* 65.6 (1958), pp. 386–408. ISSN: 1939-1471, 0033-295X. DOI: 10.1037/h0042519. URL: https://doi.apa.org/doi/10.1037/h0042519 (visited on 11/16/2025).

[57] Ian Goodfellow and Yoshua Bengio and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL: http://www.deeplearningbook.org (visited on 10/15/2025).

[58] Christopher M. Bishop. *Pattern recognition and machine learning*. eng. Information science and statistics. New York: Springer, 2006. ISBN: 978-0-387-31073-2.

[59] Kurt Hornik. "Approximation capabilities of multilayer feedforward networks". en. In: *Neural Networks* 4.2 (1991), pp. 251–257. ISSN: 08936080. DOI: 10.1016/0893-6080(91)90009-T. URL: https://linkinghub.elsevier.com/retrieve/pii/089360809190009T (visited on 11/16/2025).

[60] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. arXiv:1412.6980 [cs]. Jan. 2017. DOI: 10.48550/arXiv.1412.6980. URL: http://arxiv.org/abs/1412.6980 (visited on 01/07/2026).

[61] Scott E. Field et al. "Fast prediction and evaluation of gravitational waveforms using surrogate models". In: *Physical Review X* 4.3 (July 2014). arXiv:1308.3565 [gr-qc], p. 031006. ISSN: 2160-3308. DOI: 10.1103/PhysRevX.4.031006. URL: http://arxiv.org/abs/1308.3565 (visited on 11/16/2025).

[62] Manuel Tiglio and Aarón Villanueva. "Reduced Order and Surrogate Models for Gravitational Waves". In: *Living Reviews in Relativity* 25.1 (Dec. 2022). arXiv:2101.11608 [gr-qc], p. 2. ISSN: 2367-3613, 1433-8351. DOI: 10.1007/s41114-022-00035-w. URL: http://arxiv.org/abs/2101.11608 (visited on 11/16/2025).

[63] Alvin J. K. Chua, Chad R. Galley, and Michele Vallisneri. "Reduced-order modeling with artificial neurons for gravitational-wave inference". In: *Physical Review Letters* 122.21 (May 2019). arXiv:1811.05491 [astro-ph], p. 211101. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.122.211101. URL: http://arxiv.org/abs/1811.05491 (visited on 11/16/2025).

[64] Elena Cuoco et al. "Enhancing Gravitational-Wave Science with Machine Learning". In: *Machine Learning: Science and Technology* 2.1 (Dec. 2020). arXiv:2005.03745 [astro-ph], p. 011002. ISSN: 2632-2153. DOI: 10.1088/2632-2153/abb93a. URL: http://arxiv.org/abs/2005.03745 (visited on 11/16/2025).

[65] Hunter Gabbard et al. "Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy". In: *Nature*

Physics 18.1 (Jan. 2022). arXiv:1909.06296 [astro-ph], pp. 112–117. ISSN: 1745-2473, 1745-2481. DOI: 10.1038/s41567-021-01425-7. URL: http://arxiv.org/abs/1909.06296 (visited on 11/16/2025).

[66] Michael Pürrer. "Frequency domain reduced order model of aligned-spin effective-one-body waveforms with generic mass-ratios and spins". In: *Physical Review D* 93.6 (Mar. 2016). arXiv:1512.02248 [gr-qc], p. 064041. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.93.064041. URL: http://arxiv.org/abs/1512.02248 (visited on 11/16/2025).

[67] Deviderjit S. Sivia and John Skilling. *Data analysis: a Bayesian tutorial; [for scientists and engineers]*. eng. 2. ed., repr. Oxford science publications. Oxford: Oxford Univ. Press, 2011. ISBN: 978-0-19-856831-5 978-0-19-856832-2.

[68] Andrew Gelman et al. *Bayesian data analysis*. eng. Third edition. Texts in statistical science series. Boca Raton London New York: CRC Press, Taylor & Francis Group, 2014. ISBN: 978-1-4398-4095-5.

[69] David J. C. MacKay. *Information theory, inference, and learning algorithms*. eng. 22nd printing. Cambridge: Cambridge University Press, 2019. ISBN: 978-0-521-64298-9.

[70] Harold Jeffreys. *Theory of probability*. eng. 3. ed., repr. Oxford classic texts in the physical sciences. Oxford: Clarendon Pr, 2003. ISBN: 978-0-19-850368-2.

[71] David W. Hogg, Jo Bovy, and Dustin Lang. *Data analysis recipes: Fitting a model to data*. Version Number: 1. 2010. DOI: 10.48550/ARXIV.1008.4686. URL: https://arxiv.org/abs/1008.4686 (visited on 06/27/2025).

[72] Yudi Pawitan. *In all likelihood: statistical modelling and inference using likelihood*. eng. Paperback edition. Oxford: Oxford University Press, 2013. ISBN: 978-0-19-850765-9 978-0-19-165057-4.

[73] Nicholas Metropolis et al. "Equation of State Calculations by Fast Computing Machines". en. In: *The Journal of Chemical Physics* 21.6 (June 1953), pp. 1087–1092. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.1699114. URL: https://pubs.aip.org/jcp/article/21/6/1087/202680/Equation-of-State-Calculations-by-Fast-Computing (visited on 11/16/2025).

[74] W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications". en. In: *Biometrika* 57.1 (Apr. 1970), pp. 97–109. ISSN: 1464-3510, 0006-3444. DOI: 10.1093/biomet/57.1.97. URL: https:

//academic.oup.com/biomet/article/57/1/97/284580 (visited on 11/16/2025).

[75] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. New York, NY: Springer New York, 2004. ISBN: 978-1-4419-1939-7 978-1-4757-4145-2. DOI: 10.1007/978-1-4757-4145-2. URL: http://link.springer.com/10.1007/978-1-4757-4145-2 (visited on 11/16/2025).

[76] Daniel Foreman-Mackey et al. "emcee : The MCMC Hammer". en. In: *Publications of the Astronomical Society of the Pacific* 125.925 (Mar. 2013), pp. 306–312. ISSN: 00046280, 15383873. DOI: 10.1086/670067. URL: http://iopscience.iop.org/article/10.1086/670067 (visited on 11/16/2025).

[77] Steve Brooks et al. *Handbook of Markov Chain Monte Carlo*. en. 1st ed. New York: Chapman and Hall/CRC, May 2011. ISBN: 978-0-429-13850-8. DOI: 10.1201/b10905. URL: https://www.taylorfrancis.com/books/9780429138508 (visited on 11/16/2025).

[78] Adam Paszke et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. arXiv:1912.01703 [cs]. Dec. 2019. DOI: 10.48550/arXiv.1912.01703. URL: http://arxiv.org/abs/1912.01703 (visited on 01/22/2026).

[79] Matthias Daniel. "Constraining dark matter with gravitational waves from supermassive black hole binaries using Cassini data". PhD thesis. Goethe Universität Frankfurt am Main, 2023. DOI: 10.21248/gups.79949. URL: https://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/79949 (visited on 07/18/2025).

[80] Niklas Becker. *Dancing above the abyss: Environmental effects and dark matter signatures in inspirals into massive black holes*. arXiv:2404.02808 [gr-qc]. Apr. 2024. DOI: 10.48550/arXiv.2404.02808. URL: http://arxiv.org/abs/2404.02808 (visited on 07/11/2025).

[81] Yann A. LeCun et al. "Efficient BackProp". en. In: *Neural Networks: Tricks of the Trade*. Ed. by Grégoire Montavon, Geneviève B. Orr, and Klaus-Robert Müller. Vol. 7700. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 9–48. ISBN: 978-3-642-35288-1 978-3-642-35289-8. DOI: 10.1007/978-3-642-35289-8_3. URL: http://link.springer.com/10.1007/978-3-642-35289-8_3 (visited on 11/16/2025).

[82] M. D. McKay, R. J. Beckman, and W. J. Conover. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code". In: *Technometrics* 21.2 (May 1979), p. 239. ISSN: 00401706. DOI: 10.2307/1268522. URL: https://www.jstor.org/stable/1268522?origin=crossref (visited on 11/16/2025).

[83] Dan Hendrycks and Kevin Gimpel. *Gaussian Error Linear Units (GELUs)*. arXiv:1606.08415 [cs]. June 2023. DOI: 10.48550/arXiv.1606.08415. URL: http://arxiv.org/abs/1606.08415 (visited on 01/22/2026).

[84] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. *Layer Normalization*. arXiv:1607.06450 [stat]. July 2016. DOI: 10.48550/arXiv.1607.06450. URL: http://arxiv.org/abs/1607.06450 (visited on 01/18/2026).

[85] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. arXiv:1502.03167 [cs]. Mar. 2015. DOI: 10.48550/arXiv.1502.03167. URL: http://arxiv.org/abs/1502.03167 (visited on 01/18/2026).

[86] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15 (2014), pp. 1929–1958. URL: https://www.jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf?utm_content=buffer79b4.

[87] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. *On the difficulty of training Recurrent Neural Networks*. Version Number: 2. 2012. DOI: 10.48550/ARXIV.1211.5063. URL: https://arxiv.org/abs/1211.5063 (visited on 01/18/2026).

[88] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, June 2016, pp. 770–778. ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.90. URL: http://ieeexplore.ieee.org/document/7780459/ (visited on 01/18/2026).

[89] Jonathan Goodman and Jonathan Weare. "Ensemble samplers with affine invariance". en. In: *Communications in Applied Mathematics and Computational Science* 5.1 (Jan. 2010), pp. 65–80. ISSN: 2157-5452, 1559-3940. DOI: 10.2140/camcos.2010.5.65. URL: http://msp.org/camcos/2010/5-1/p04.xhtml (visited on 06/27/2025).

[90] Dootika Vats and Christina Knudson. *Revisiting the Gelman-Rubin Diagnostic*. arXiv:1812.09384 [stat]. Sept. 2020. DOI: 10.48550/arXiv.1812.09384. URL: http://arxiv.org/abs/1812.09384 (visited on 01/22/2026).

[91] Carl Edward Rasmussen. "Gaussian Processes in Machine Learning". In: *Advanced Lectures on Machine Learning*. Ed. by Olivier Bousquet, Ulrike Von Luxburg, and Gunnar Rätsch. Vol. 3176. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 63–71. ISBN: 978-3-540-23122-6 978-3-540-28650-9. DOI: 10.1007/978-3-540-28650-9_4. URL: http://link.springer.com/10.1007/978-3-540-28650-9_4 (visited on 01/03/2026).

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel verfasst habe. Alle Stellen der Arbeit, die wörtlich oder sinngemäß aus Veröffentlichungen oder aus anderen fremden Texten entnommen wurden, sind von mir als solche kentlich gemacht worden. Ferner erkläre ich, dass die Arbeit nicht - auch nicht auszugsweise - für eine andere Prüfung verwendet wurde.

Hamburg, den 23. Januar 2026:

Jasmin Hartmann