
DEEP LEARNING FOR GRAVITY-GRADIENT NOISE MITIGATION AT THE EINSTEIN TELESCOPE

David Bertram

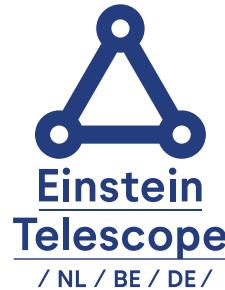
SUPERVISOR:
Prof. Dr. Achim Stahl

submitted to

Physics Institute IIIB
Faculty of Mathematics, Computer Science and Natural Science
RWTH Aachen University

as a partial fulfilment of the requirements for the degree

Master of Science in Physics



DAVID BERTRAM:
*Deep Learning for Gravity-Gradient Noise Mitigation
at the Einstein Telescope*

Master Thesis, RWTH Aachen University

SUPERVISOR:
Prof. Dr. Achim Stahl
Physics Institute IIIB
RWTH Aachen University

CO-SUPERVISOR:
Prof. Dr. Christopher Wiebusch
Physics Institute IIIB
RWTH Aachen University

Abstract

Third-generation gravitational wave observatories like the future Einstein Telescope aim to extend sensitivities to unprecedented dimensions and open a new window to the universe up to cosmological distances. Reaching this goal is inevitably connected with novel techniques to overcome the limitations of the current generation. Specifically in the low-frequency regime of 1 to 10 Hz, seismic gravity-gradient noise is anticipated to become a limiting factor for earthbound detectors.

In this work, a deep learning based gravity-gradient noise mitigation scheme that operates on auxiliary seismic sensors is developed and tested. The tests comprise a stochastic seismic simulation as a foundation for gravity-gradient noise modeling at the Einstein Telescope. Promising cancellation efficiencies by a factor of two on average have been achieved as a first proof of concept. The prospects of substantial improvement in cancellation efficiency with follow-up studies are significant and the technique's potential to overcome shortcomings of current methods is worth further investigation.

CONTENTS

Abstract	i
1 Introduction	1
2 The Einstein Telescope	3
2.1 Gravitational-Wave Astronomy	3
2.2 Einstein Telescope Conceptual Design	8
2.3 Gravity-Gradient Noise	11
3 Deep Neural Networks	17
3.1 Fundamental Concepts	17
3.2 Advanced Architectures	22
3.3 Spatiotemporal Convolutions	25
4 A Gravity-Gradient Noise Model for Data Generation	29
4.1 Theoretical Foundations of the Model	30
4.2 The Seismic Sensor Array	38
4.3 Simulation Algorithm and Parameter Summary	42
5 Deep Learning for Gravity-Gradient Noise Mitigation	49
5.1 The Network Architecture	49
5.2 General Test Scenario	53
5.3 Euregio Test Scenario	59
5.4 Inclination Dependency	63
5.5 Relative Sensor Importance	64
6 Discussion & Outlook	67
Abbreviations & Nomenclature	v
List of Figures	vii
List of Tables	viii
Appendices	ix
Bibliography	xii

INTRODUCTION

A new era of astronomy began with the first detection of Gravitational Waves (GWs) from a binary black hole merger in 2015 [1]. Many decades of research and technical developments have preceded to achieve the measurement of an effect smaller than a per mille of proton radius. A Nobel prize and numerous further detections later, a point is reached where the existing infrastructures of current-generation gravitational wave observatories represent a boundary to achievable sensitivities. Foreseen third-generation observatories are on the verge of pushing these boundaries and reaching sensitivities where no longer the bare detection but the astrophysical observation is of focus [2].

To go through the door that the second generation has opened and start the era of high-precision GW astronomy, the Einstein Telescope (ET) is a European project to host a next-generation observatory. It aims to increase the sensitivity of current detectors by a factor of ten and extend the frequency range down to a few Hz. [2] With these ambitions, the already enormous challenge of noise reduction becomes even more demanding. Noise sources that have not been crucial before can evolve into a limiting factor. Specifically for earthbound detectors like ET, seismic Gravity-Gradient Noise (GGN) is expected to be dominating the sensitivity limitations for the low-frequency observation band of 1 to 10 Hz. [2, 3, 4] This seismic GGN is caused by minuscule density perturbations in the detector's vicinity due to seismic activity¹. Particularly with GW observations in the context of early warning systems for Multi-Messenger Astronomy (MMA), this low-frequency regime is crucial to allow detections in the early phase of the inspiral [5, 6]. The first step toward GGN mitigation at ET is to choose an approx. 250 m deep underground location in order to suppress atmospheric and seismic surface wave noise contributions. Supplementary suppression is still required and foreseen to be conducted based on the data of auxiliary seismic sensors. The essential concept is to predict the GGN based on the seismic measurements and subtract it from the interferometer data. [2]

Although linear Wiener Filters (WFs) are considered the current gold standard for this task, the method comes with specific limitations. A central shortcoming is a dependency on detailed models of sensor-to-interferometer correlations that relies on the detailed knowledge of site-characteristic seismic properties. Some of these properties, for instance, the polarization components of the seismic field, can be difficult to acquire. Furthermore, the instrumental noise of the seismic sensors and the optimal positioning of the sensors is expected to have a significant impact on the cancellation efficiency. [4, 7]

A novel approach of utilizing deep learning as an alternative to WFs for GGN mitigation will be developed and investigated in this thesis. The idea is to use a deep Neural Network (NN) that learns to interpret the spatiotemporal features of a signal from the seismic sensor array and predict the affiliated GGN strain by reconstructing the transfer function. This approach is visualized in fig. 1.1.

¹as discussed in detail in sec. 2.3

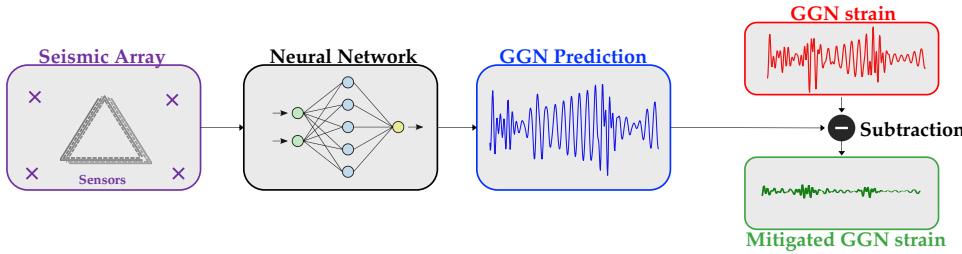


Figure 1.1: The proposed deep learning approach for GGN cancellation.

The strategy is motivated by the fact that a NN is, in theory, able to approximate any real function [8]. Especially the remarkable success of Convolutional Neural-Networks (CNNs) for tasks that include data with spatial and temporal features makes them a promising candidate for this task [9, 10].

A current constraint for training is the availability of the ground truth GGN signal. However, once GGN will be the dominant noise contribution in the relevant frequency band, it should be possible to produce training data with the measurements. Transient GW signals or excess noise sources that would still be present in the data, particularly suspension thermal noise in the given frequency regime [2], would act as small-scaled noise sources on the training ground truth. Techniques for deep learning with noisy ground truth exist and are further studied with the prospect of no degradation in performance [11, 12].

This work presents the transition from theory into praxis with a first proof of concept that a suited network architecture for the proposed strategy exists.

A stochastic seismic simulation as the basis for a GGN model at ET is presented. The data is applied to test a proposed network architecture in a general and a Euregio Meuse-Rhine (EMR) site-specific test scenario. The influence of different levels of seismic sensor instrumental noise is studied as well as the relative importance of sensor positions.

THE EINSTEIN TELESCOPE

This chapter intends to provide the reader with a basic understanding of the scientific context in which this thesis is embedded by briefly introducing the growing field of gravitational-wave astronomy, the role of the proposed Einstein Telescope (ET) in it, and the associated challenges of Gravity-Gradient Noise (GGN) mitigation.

2.1 GRAVITATIONAL-WAVE ASTRONOMY

2.1.1 HISTORICAL CONTEXT

Theory The existence of Gravitational Waves (GWs) has been first postulated by Einstein over a century ago as a direct consequence of his theory of general relativity [13]. Motivated by the unphysical instantaneous propagation of changes in gravitational force in Newton's theory, Einstein developed an advanced theory that describes gravity as an intrinsic feature of spacetime itself. According to the theory, matter defines the curvature of space-time, and vice versa, the curvature of space-time defines how matter moves. A direct consequence is that gravity no longer propagates instantaneously but with the speed of light in the form of small distortions in the curvature of space-time. If these changes are periodic or oscillatory, they are called GWs [14].

First Generation Various experiments proved Einstein's theory, and the existence of GWs was generally considered certain. Since the existence of massive compact objects like neutron stars or black holes was not yet known, irrefutable proof in direct detection seemed impossible however. [15] With the knowledge about the existence of these astrophysical GW sources, the optimism for direct detection increased. Finally, the first detection attempts were initiated in the 1960s using resonance antennas, the so-called first generation of GW detectors. Named after their leading researcher, the so-called *Weber bars* are large cylindrical masses that, in theory, have excited longitudinal modes if a GW with the matching resonance frequency propagates through them. The first detection of GWs was announced shortly after by Weber himself. This raised much attention from the scientific community, and research groups worldwide started reproducing the experiments. None of them were successful, and as misconceptions in Weber's work came to the surface, the scientific consensus did not recognize the detection. [15, 16]

Second Generation Nevertheless, the research field gained significant momentum, and in 1975 the first indirect proof of gravitational radiation was achieved through the observation of the orbital decay of two binary pulsars [17]. Around the same time, interferometry became considered the method of choice for GW detection, and the first prototypes were built [18, 19].

This trend ultimately led to the funding of the Laser Interferometer Gravitational-Wave Observatory (LIGO) [20] and the foundation of an affiliated international collaboration in 1997. The collaboration significantly advanced technical progress, and developments such as vacuum tubes and higher laser power were employed to build two separate interferometers. The second generation of GW detectors ultimately arrived with the first successful locking and start of the first observation run of LIGO in 2002. After years of further technical improvements and detection-less observation runs, a major upgrade was conducted. The resulting advanced LIGO (aLIGO) had an increased sensitivity of factor three, and the final engineering run took place in September 2015. On the 14th of September at 9:50:45 UTC, however, the two detectors nearly simultaneously measured an unknown transient signal as displayed below in [fig. 2.1](#).

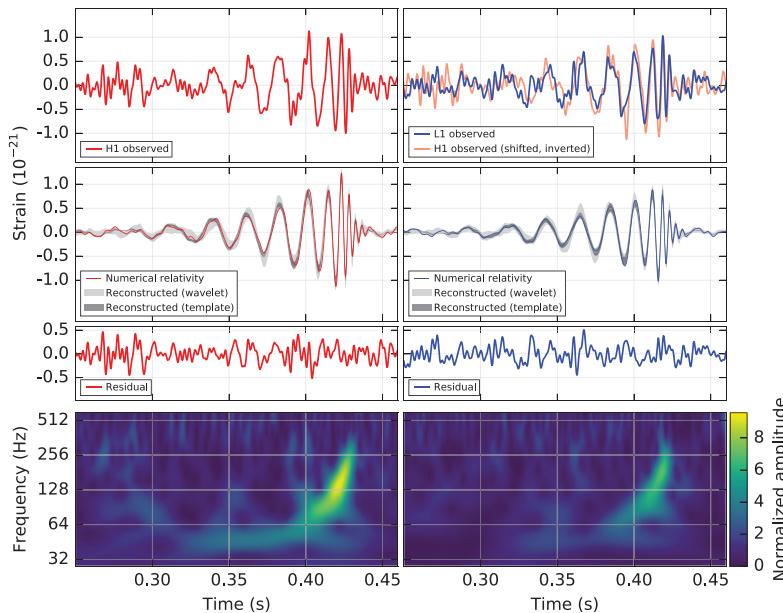


Figure 2.1: The first direct detection of gravitational waves.

For each detector, the measured signal (*top*), the comparison to a general relativity template (*middle*) and the corresponding spectrogram (*bottom*) are shown. An inspiral of increasing frequency peaks in the final merger around 0.42 s. Adapted from [1].

left: The signal in the aLIGO detector in Hanford, Washington
right: The signal in the detector in Livingston, Louisiana.

Months of data analysis and meticulous discussions about the validity of the signal followed. The coincidence in both detectors combined with the significant agreement with general relativity templates supported the hypothesis, and the collaboration officially announced the first direct proof for the existence of GWs in February 2016. The first directly detected GWs were emitted from a system of two black holes with $36^{+5}_{-4} M_{\odot}$ and $29^{+4}_{-4} M_{\odot}$ solar masses that coalesced and finally merged over one billion years ago. This breakthrough was awarded the Nobel prize and marked the beginning of the GW era in astronomy. [1, 15]

A New Era Multiple observations followed, including the first joint detection with the European Virgo [21] detector in 2017. This allowed the observation of a specifically remarkable event called GW170817.

With the information provided by the Virgo detector, it was possible to locate the inspiral of two binary neutron stars sufficiently well to discover the electromagnetic counterpart. The event became the best-observed event in astronomy and provided an essential test for general relativity as well as crucial insights into the long-term puzzle of the nucleosynthesis of heavy elements. Moreover, the event represented a textbook example for Multi-Messenger Astronomy (MMA) with gravitational waves. [6, 22] Despite the success of the current generation of GW detectors, their infrastructure defines a limitation to the maximum feasible sensitivities and bandwidths. Third-generation detectors aim to overcome these limitations and are foreseen to operate by the 2030s[23]. With them, the new window to the universe provided by GWs is planned to be fully exploited, and numerous astrophysical observations providing novel insights are anticipated.

2.1.2 NATURE OF GWs AND DETECTION PRINCIPLE

Modern gravitational wave detectors almost exclusively rely on the principle of interferometry. In order to comprehend the detection principle, the fundamental nature, and most notably, the interaction of GWs with matter or energy has to be understood. In the following, the necessities will be introduced based on [14, 15, 24] unless stated otherwise.

GWs in General Relativity The cornerstone of Einstein’s theory of general relativity is that matter and spacetime are intrinsically connected. This is expressed through the *Einstein field equations* in the four-vector notation as

$$G_\nu^\mu + \Lambda g_\nu^\mu = \kappa T_\nu^\mu \quad (2.1)$$

where G_ν^μ is a tensor that defines the curvature of spacetime and T_ν^μ is a tensor representing matter and energy. The cosmological constant is denoted by Λ and the spacetime metric by g_ν^μ . The essential point here is that the relation between the curvature of spacetime (left side of the equation) and any form of mass and energy (right side of the equation) becomes perceptible. Moreover, the coupling constant $k = \mathcal{O}(10^{-43} \text{ N}^{-1})$ reveals that very massive objects are required to have a significant impact on the curvature of spacetime. Although self-coupling between energy and spacetime makes the general solution to the equations very complicated, appropriately chosen simplifications can demonstrate the wave nature of gravitational radiation. If $h_\nu^\mu \ll 1$ denotes a small perturbation in the spacetime metric the equations can be linearized to

$$\left(-\partial^2 t + \nabla^2 \right) \bar{h}_\nu^\mu = -16\pi T_\nu^\mu \quad (2.2)$$

which can be directly identified as a differential wave equation for the spacetime metric perturbation $\bar{h}_\nu^{\mu 1}$ caused by an external energy source represented by the energy-mass tensor T_ν^μ . If propagation in vacuum ($T_\nu^\mu = 0$) is assumed, the equation further simplifies to

$$\left(-\partial^2 t + \nabla^2 \right) h_\nu^\mu = 0 \quad (2.3)$$

and a solution is a transversally polarized gravitational wave

$$h_\nu^\mu = A_\nu^\mu \exp(i k_\mu x^\nu) \quad (2.4)$$

¹more precisely, the «reverse trace» of h_ν^μ

Polarizations This wave satisfies the *Lorentz* gauge $A_\nu^\mu \cdot k_\nu = 0$ where k_ν is the wave four-vector. The amplitude vector A_ν^μ can either be $+$ or \times -polarized. The two polarization states result from gravitational radiation's dominating mass quadrupole moment. Mass and momentum conservation implies that mass monopole and dipole contributions scale with the distance at least with a factor $\propto 1/r^2$. The mass quadrupole represents the leading order with $\propto 1/r$. Hence another essential condition for gravitational wave sources is demonstrated: not only very massive objects are required, but also a sphere-asymmetrical dynamic. Systems that fulfill these requirements are mainly coalescing binary systems of massive compact objects like neutron stars and black holes. Further candidates are fast-rotating neutron stars (*pulsars*) or supernovae. For a wave propagating in z-direction, the two polarizations are visualized below in fig. 2.2 as the effect on free test masses arranged on circles.

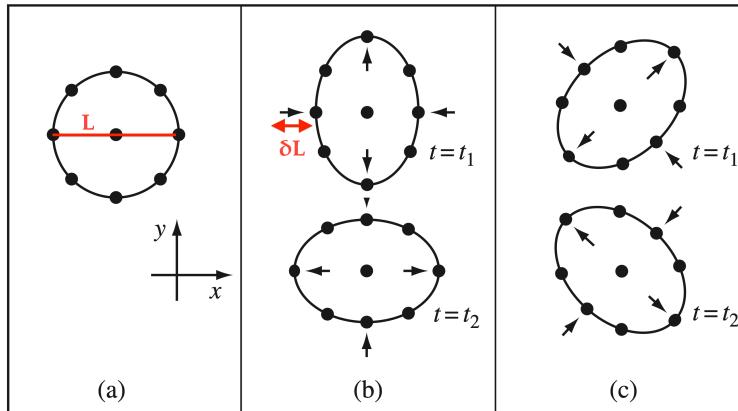


Figure 2.2: The transversal polarization modes of a gravitational wave.

Displayed is the effect of a GW propagating in z-dimension through a circular arrangement of free masses in the xy-plane. The z-axis is perpendicular to the paper plane. Adapted with modifications from [24].
 a): no GW. b): GW with $+$ polarization. c): GW with \times polarization.

Detection Principle This effect on masses is the key to understanding the principle of GW detection with interferometry. The spacetime perturbation due to a GW results in a physical observable called the *strain*. It denotes the relative change of displacement between two test masses

$$h \equiv \frac{\delta L}{L} \quad (2.5)$$

imposed by a GW as indicated in fig. 2.2. This implies that a GW can be directly measured by observing the distance perturbation between two test masses. This would be a fairly simple endeavor if the strain h would not typically be of order $\mathcal{O}(10^{-22})$ due to the small coupling constant in eq.(2.1). Even with test masses hundreds of kilometers apart, the quantity to be measured would be only a fraction of the radius of a proton. Most modern GW detectors address this challenge by utilizing an optical instrument called *Michelson-interferometer*. The instrument was initially developed to measure the speed of light and played a key role in the famous *Michelson-Morley* experiment in 1881. Modern *Michelson-interferometers* are highly sophisticated versions of the original instrument that reached a sensitivity of 10^{-10} . [25]

The fundamental operating principle is visualized in fig. 2.3 using a simplified illustration of the aLIGO setup. The interferometer mirrors represent the test masses probing the spacetime perturbation caused by GWs. Incident GWs with a propagation direction perpendicular to the interferometer plane result in a maximum response function. Generally, the response of the interferometer depends on the polarization as well as the orientation of the incident wave and is expressed by detector-characteristic antenna pattern functions.

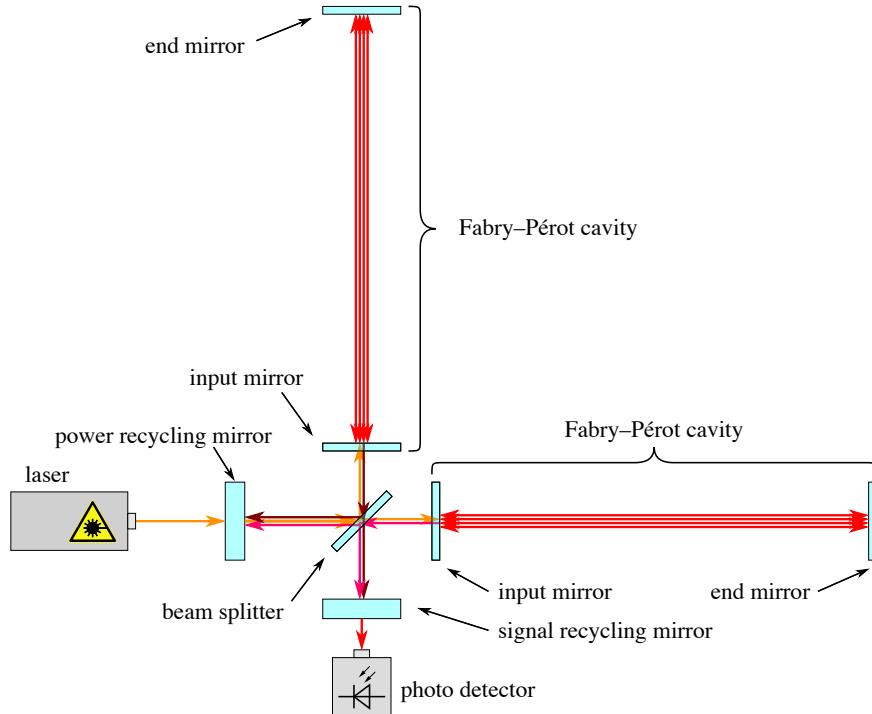


Figure 2.3: **Schematic illustration of a modern Michelson-interferometer.**

Light emitted from a powerful laser (*orange*) enters the interferometer from the left and is partially reflected in the two interferometer arms by a beam splitter. The light is reflected multiple times (*red*) in *Fabry-Pérot* cavities comprised of a fully reflective end mirror and a semi-transparent input mirror. Finally, the light beams from the two arms (*purple & brown*) are reunited for interference. The phase difference and, thus, the relative path difference is observable in the photodetector. Additional mirrors are employed for the purpose of power and signal recycling. The sketch is not to scale, arm lengths of modern interferometers reach several kilometers. Adapted from [26].

Compared to the original optical setup, a major modification are the *Fabry-Pérot* cavities utilized to filter laser wavelengths and amplify the phase differences due to the GW distortion. Additional power and signal recycling mirrors keep the laser light within the system to further intensify the power. Major technical advances like vacuum tubes, mirror coatings, and improved lasers, to name a few, have contributed to the fact that these modern interferometers reach sensitivities that are more than a billion times larger than their historical archetypes. With sensitivities that are eventually sufficient for GW observations, the technical challenges associated with noise suppression are immense. Exemplary noise sources and connected mitigation techniques will be discussed in more detail in the following section.

2.2 EINSTEIN TELESCOPE CONCEPTUAL DESIGN

As part of the third generation of GW detectors, ET aims to overcome the noise limitations of the current generation and increase the sensitivity by one order of magnitude over the entire detection frequency band. The challenges and technological advances that go hand in hand with this ambition will be presented hereafter based on [2] unless stated otherwise.

Optical Layout Similar to aLIGO, the foreseen optical layout for ET is a dual-recycled *Michelson-interferometer* with *Fabry-Pérot* cavities as introduced in section 2.1.2. A measure to increase sensitivity is an increased interferometer arm length 10 km compared to 4 km at aLIGO. The main difference, however, lies in the currently followed «xylophone» design for the Einstein Telescope (ET-D). This design is based on the approach of individual detectors for the low-frequency (LF) and high-frequency (HF) bands. Furthermore, a total amount of three detectors, each containing a LF and HF interferometer, is foreseen to be arranged in a triangular shape as displayed in fig. 2.4. The triangular shape results in an equal sensitivity for both GW polarizations and a more isotropic antenna pattern function compared to a classical 90° interferometer.

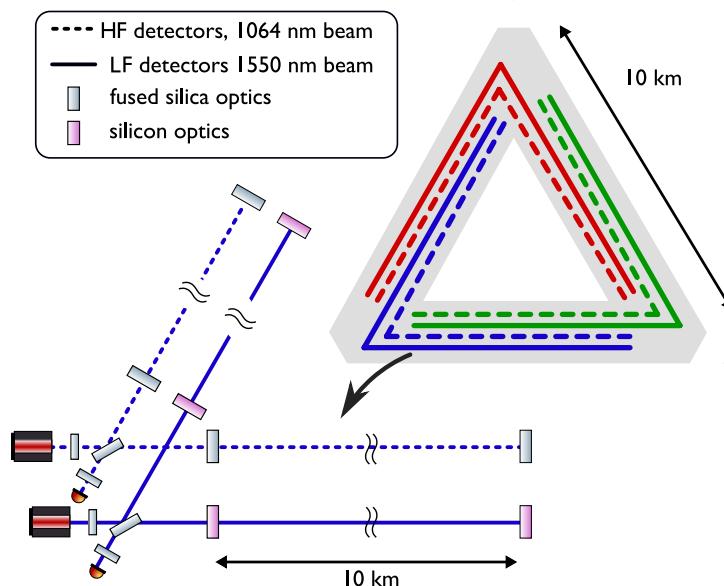


Figure 2.4: The optical layout of ET-D.

Three nested detectors (red, blue & green) are arranged in a triangular shape with 60° rotation. Each contains a low-power, low-frequency (LF) (solid line) and a high power, high-frequency (HF) (dashed line) interferometer of 10 km arm length. Adapted from [27].

A crucial advantage of the xylophone design is that the tradeoff between large sensitivities in both, the LF and the HF band, can be bypassed by meeting the noise requirements of each band in an optimized manner with separate detectors. An illustrative example of the tradeoff is the relation between laser power and the influence of quantum noise. The sources of quantum noise are *Photon Shot Noise*, a direct effect of the counting properties of a photodetector, and *Radiation Pressure Noise*, a consequence of the transmitted photon momentum during reflection from the mirrors.

Photon Shot Noise is dominant for HF and can be reduced by increasing the number of photons, i.e., the laser power. On the other side, *Radiation Pressure Noise* is dominant for LF, and an increased laser power would amplify the transferred momentum and thus the noise. This laser power tradeoff indicates the different technical demands of the LF and HF interferometers. The specifications of both will be discussed in the following.

2.2.1 HIGH-FREQUENCY INTERFEROMETERS

The GW observation band of frequencies between 30 Hz and several kHz will be addressed by the high-frequency interferometers of ET (ET-HF). In this regime, quantum noise is dominated by *photon shot noise* that is foreseen to be mitigated by a high input laser power of 500 W. The infrared laser with 1064 nm wavelength will eventually reach a power of 3 MW in the arm cavities. With this extensive laser power, even the minuscule heat absorption of the fused silica mirrors in the ppm range is sufficient to make a cryogenic operation impossible. The required heat dissipation would spoil the mechanical loss factor of the suspension system. A direct consequence is that ET-HF will be operated at room temperatures, and *Mirror Thermal Noise* is the second predominant fundamental noise source next to quantum shot noise. Thermal noise, or more specifically *Mirror Thermal Noise*, originates from the thermal properties of the atoms and molecules that make up the mirror. As a result of the *Brownian* motion of the atoms, the mean motion summed of the mirror is non-zero and thus introduces a subtle displacement noise. The strength of this effect depends on the material qualities and scales linearly with temperature. How the fundamental noise sources influence the total ET-HF noise budget is displayed in fig. 2.5.

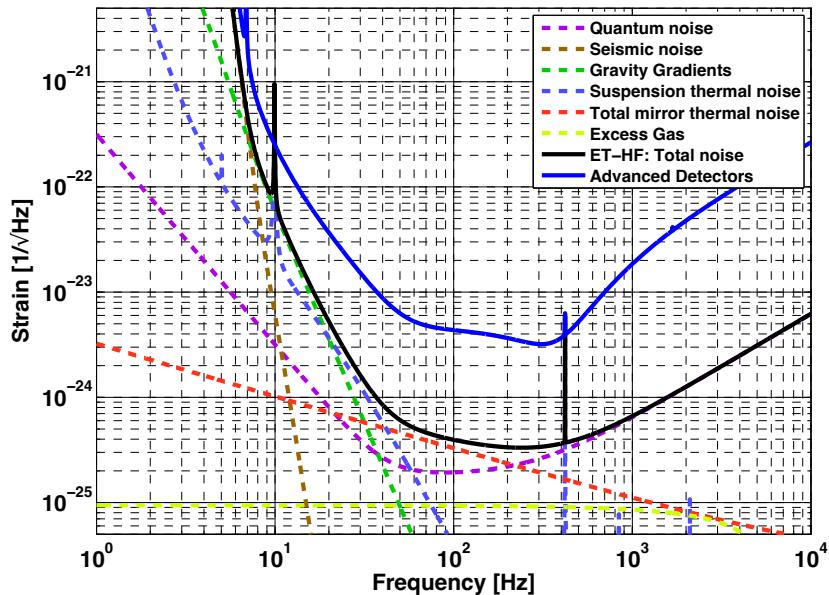


Figure 2.5: **Total noise budget of ET-HF.**

The individual noise contributions (*dashed colored lines*) and the total noise budget of ET-HF (*black line*). The individual sources are added in quadrature and compared to the aLIGO sensitivity curve (*blue line*). Adapted from [28].

In addition to the discussed fundamental noise sources, several technical noise sources must be considered in the noise budget.

An instance of technical noise sources is *Seismic Noise*. Diminutive vibrations in the ground propagating as seismic waves caused by human activities or natural ground eruptions directly couple to the instrumentation and introduce noise as mirror displacements. Seismic isolation of all test masses is crucial to mitigate this noise source and is generally achieved by a combination of active and passive cancellation. The ET approach is orientated on the advanced Virgo superattenuator with a focus on passive inertial isolation achieved by a sophisticated multi-stage pendulum [29]. A total length for the ET-HF attenuator of 8 m is envisioned. Seismic activity also introduces noise via indirect coupling. The consequent *Gravity-Gradient Noise* (GGN) is of particular interest for this thesis and will be discussed in more detail in [section 2.3](#). Residual gas molecules in the laser tubes or on the mirror surfaces inject so-called *Excess Gas Noise*. As a countermeasure, all optical instruments of ET will be operated within an ultra-high vacuum with a maximum tolerance limit of 10^{-8} Pa. The efficient design of such an extensive ultra-high vacuum system for ET is a pending subject of current research. Numerous additional noise sources to the ones presented in the noise budget [fig. 2.5](#) must be considered. Exemplary mitigation techniques range from extreme mirror smoothness for *Scattered Light Noise* to «squeezed» light for *Laser Frequency and Intensity Noise*. A more exhaustive discussion can be found in [2, 28].

2.2.2 LOW-FREQUENCY INTERFEROMETERS

The low-frequency interferometers of ET (ET-LF) are foreseen to cover frequency ranges from a few Hz to 30 Hz. In contrast to ET-HF, a laser with lower input power of 3 W is foreseen that reaches a maximum of 18 kW in the cavities. The lower power serves the purpose of quantum *Radiation Pressure Noise* mitigation and further allows the operation in a cryogenic system to suppress thermal noise of the mirrors and suspensions. The ET-LF noise budget is shown in [fig. 2.6](#).

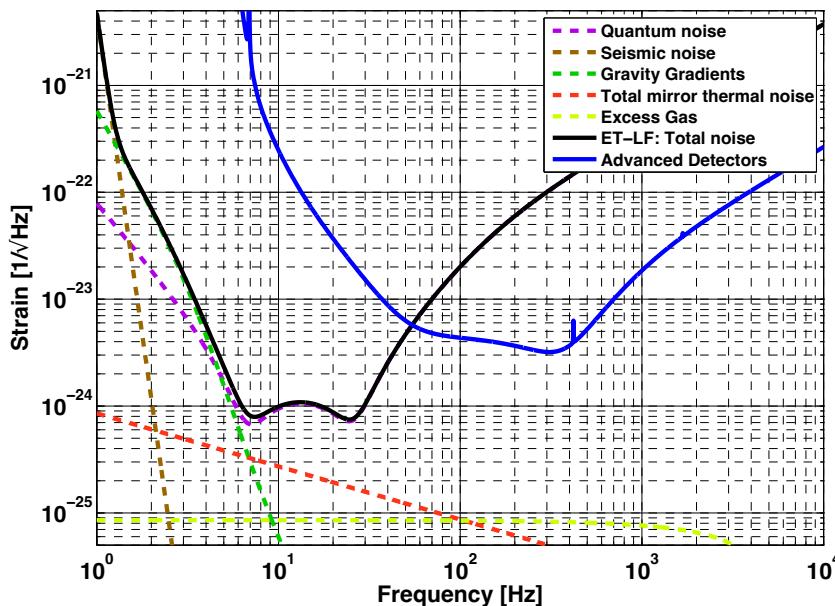


Figure 2.6: **Total noise budget of ET-LF.**

The individual noise contributions (*dashed colored lines*) and the total noise budget of ET-HF (*black line*). The individual sources are added in quadrature and compared to the aLIGO sensitivity curve (*blue line*). Adapted from [28].

The cryogenic system is aimed to reach temperatures between 10 to 20 K in order to suppress the total thermal noise to a subordinate contribution. Since fused silica has poor mechanical properties at cryogenic temperatures, silicon is the material of choice for fabricating the ET-LF mirrors. It meets the mechanical and thermal requirements even at low temperatures and is available in high quality due to its application in the semiconductor industry. However, as silicon is non-transparent for the wavelength of 1064 nm, the ET-LF laser will have an increased wavelength of 1550 nm in the short-wavelength infrared regime.

Quantum noise in the form of *Radiation Pressure Noise* dominates the range of frequencies above 7 Hz. Supplementary to the comparably low laser power of ET-LF, massive mirrors of 211 kg for ET-LF and 200 kg for ET-HF are envisioned to reduce the momentum transferred noise. Analogous to ET-HF, the test mass mirrors of ET-LF will be located in the ultra-high vacuum and suspended from super-attenuators. The higher demands of seismic isolation for lower frequencies will be met with an extended total pendulum length of 17 m. However, seismic isolation is ineffective against the indirect noise coupling of seismic activity. This coupling is particularly relevant for ET-LF and will be discussed in more detail in the following section.

2.3 GRAVITY-GRADIENT NOISE

Gravity-Gradient Noise (GGN), also known as *Newtonian Noise*, is the indirect coupling of density fluctuations into the test-mass mirrors of GW interferometers via the gravitational force. The mechanism and its relevance for ET will be explained in the following based on [2, 30] with a particular focus on the Euregio Meuse-Rhine (EMR) candidate host site [3].

Most general, the density fluctuations that induce GGN can include seismic fields, atmospheric disturbances, and the vibrating infrastructure of the GW observatory. As an instance, the GGN coupling mechanism for density perturbations originating from seismic fields is illustrated in [fig. 2.7](#) on the following page.

Atmospheric GGN The effect of atmospheric GGN follows the same principle, with the only difference being the source of density fluctuations. Since the atmosphere is not static, dynamic fields of pressure or temperature induce density fluctuations with length scales ranging from less than a few meters up to hundreds and thousands of kilometers. These length scales of density perturbation determine the frequency regime in which the corresponding GGN occurs. Specifically for GW detection, the relevant fluctuations have characteristic dimensions below a few kilometers with connected GGN in the range of a few Hz. The associated atmospheric processes are pressure fluctuations in the first meters above the earth's surface caused by sound, wind, or the interaction of both. It has been assessed that atmospheric GGN contributions need to be suppressed by one order of magnitude to reach the ET-D design sensitivity. For this purpose, an underground location of 200 to 300 m below the surface is foreseen for ET. Although follow-up research confirms that this depth provides sufficient shielding to mitigate atmospheric GGN below the aimed sensitivity, the contributions could still have significance and are recommended to be reevaluated with site-specific data [31].

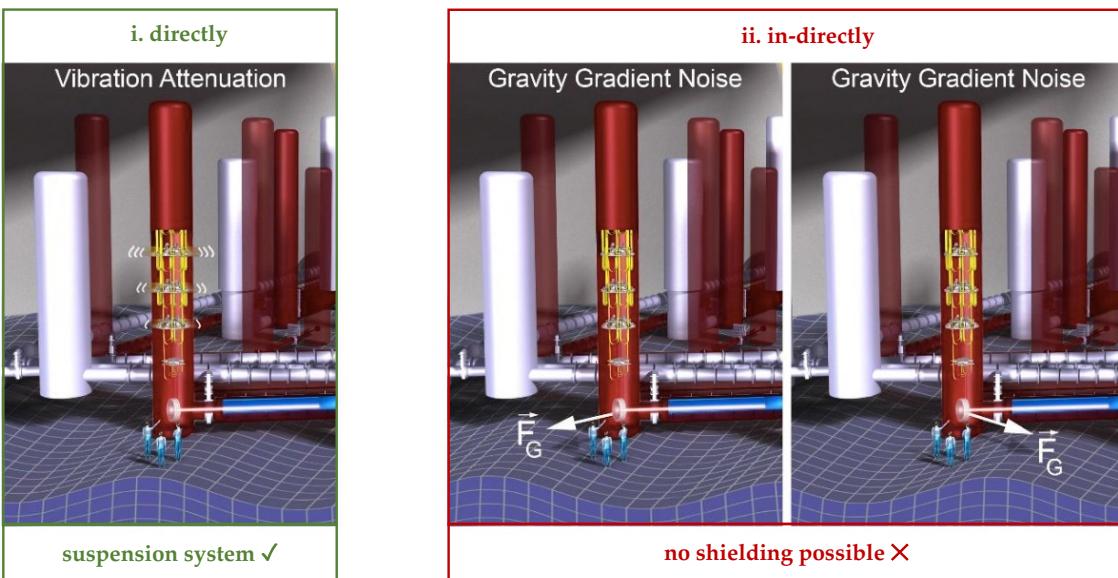


Figure 2.7: **Coupling between seismic noise and interferometer test-masses.**

Shown is an illustration of the interferometer mirrors suspended from the multistage super-attenuators. The blue surface represents seismic density perturbations. Adapted with modifications from [32].

left: Ground vibrations couple directly to the instruments. The suspension system achieves strong suppression of this effect.

right: GGN as the indirect coupling via gravitational forces between the test masses and the high-density regimes. Shielding from this gravitational force is not possible.

Another primary motivation for the underground location is the reduction of seismic activity in general and specifically the mitigation of surface contributions to the seismic GGN. Although the first principle to seismic GGN mitigation at ET is to choose a quiet host site, natural and anthropogenic seismic noise sources can never be mitigated entirely, and supplementary isolation is crucial.

Surface Wave GGN Seismic activity arising from these sources propagates either as surface or body modes of seismic waves. The surface waves mainly consist of so-called *Rayleigh* or *Love* modes. They are characterized by a vertical displacement of the medium's surface and present the predominant GGN limitation for aboveground GW detectors. By going underground, the contribution of surface wave GGN can be strongly suppressed and limited to effects arising from the walls of the hosting cavities. Local geological properties determine how efficient the suppression with depth is. The general rule is that the shorter the seismic wavelength of the surrounding medium, the stronger the isolation. Therefore, detailed knowledge of a candidate host site's seismic characteristics is required to estimate the residual contributions of surface waves at a certain depth. To address this question specifically for the EMR candidate site, a passive and active seismic campaign has been executed in Terziet, South Limburg[3, 32].

Over the extent of two years, surface seismometer grids and sensors in two deep boreholes were used for seismic characterization. Furthermore, the local geology was reconstructed by an active seismic inversion study. The results are summarized as a relatively soft soil layer with excellent attenuation of anthropogenic noise followed by a deeper layer of hard rock that is suited for constructing the ET host cavities.

Euregio GGN Based on the geological properties of the EMR candidate site along with the seismic characteristics, the expected spectral density of seismic GGN at a depth of 250 m has been accessed by the campaign. Their results in terms of surface and body wave contributions are displayed in fig. 2.8.

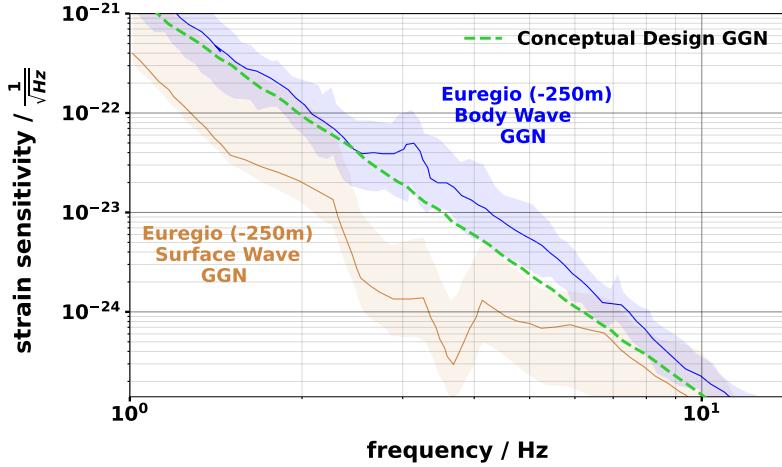


Figure 2.8: **Modulated seismic surface and body wave GGN in the Euregio.**

Seismic characterization results of EMR site-specific GGN spectral density from surface waves (*solid beige line*) and body waves (*solid blue line*) at 250 m depth under surface. The shaded bands enclose the 10th and 90th percentile. Data source [3]. A comparison to the considered GGN in the conceptual design report [2] (*dashed green line*) is shown.

The main results underline that the geology in the EMR is beneficial to attenuate the surface wave contributions and the residual seismic GGN is dominated by body waves. Although the location is comparatively quiet and has good attenuation qualities, the expected total seismic GGN spectral density lies slightly above the spectral density considered for the conceptual design report update [2]. The total seismic GGN compared to the ET-D design sensitivity is shown in fig. 2.9.

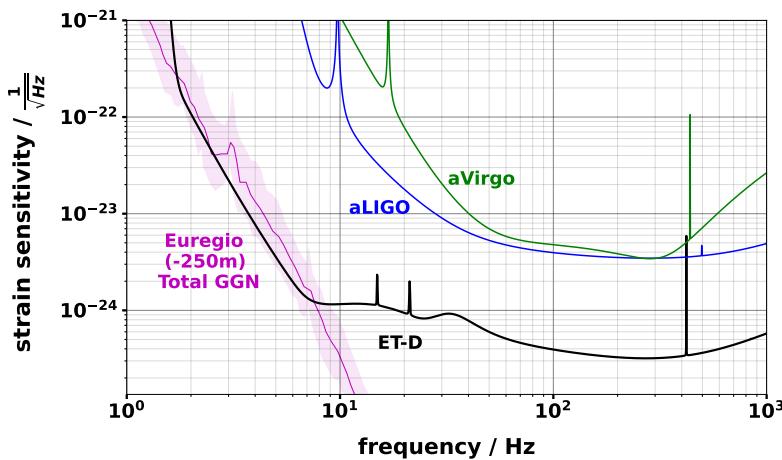


Figure 2.9: **Modulated total seismic GGN in the Euregio.**

Seismic characterization results of EMR site-specific total GGN spectral density at 250 m depth under surface. (*magenta*) The shaded band enclose the 10th and 90th percentile. Data source [3]. A comparison to the ET design sensitivity [2] (*black*) and the sensitivities of the second generation [33] (*blue & green*) is shown.

It can be concluded that the seismic GGN in the EMR will be dominated by body waves, and active cancellation by a factor of two to three is required to reach the ET-D design sensitivity in the low-frequency regime up to 10 Hz. The comparison between the total seismic GGN and the design sensitivities of the advanced second-generation detectors underlines why GGN cancellation is a topic that especially gains significance with the anticipated third generation of terrestrial GW observatories. However, as the detectors of the second generation are entirely exposed, the levels of surface wave GGN are significantly higher, and active cancellation has started to be investigated [34, 35].

Due to the requirement for precise knowledge about the dynamic seismic field and environmental properties, the exact modeling of seismic GGN is very complicated. Consequently, the general strategy for active GGN cancellation relies on the measurement of auxiliary seismic sensors. Based on the signal of these sensors, a GGN estimate should be produced and subsequently subtracted from the noisy data. [2, 36]

2.3.1 COHERENT CANCELLATION WITH WIENER FILTERS

The current «gold standard» for the purpose of active cancellation is to exploit coherent correlations between the seismic sensor signal and the affiliated GGN with linear filters. Assuming all signals are stationary, the Wiener Filter (WF) is the optimal filter for this purpose. Its characteristics will be introduced in the following based on [2, 30, 34].

The WF is generally defined in the frequency domain as the linear mapping of the signal from a seismic sensor array $\vec{s}(f)$ to the GGN prediction $\hat{n}(f)$

$$\hat{n}(f) = \vec{w}_F^\dagger(f) \cdot \vec{s}(f) \quad (2.6)$$

Estimation of the correct filter utilizes the cross-spectral correlation density of the auxiliary seismic sensors and the specific GW detector. It can be calculated via

$$\vec{w}_F(f) = \mathbf{C}_{\text{SS}}(f) \cdot \vec{\mathbf{C}}_{\text{SN}}(f) \quad (2.7)$$

where \mathbf{C}_{SS} is the correlation matrix between all seismic sensors and $\vec{\mathbf{C}}_{\text{SN}}$ is the correlation vector between the auxiliary seismic sensors and the GW detector. Although the correlation between the seismic sensors can be measured once installed, the WF is model-dependent and requires the calculation of the correlation between the seismic sensors and the GW detector. Furthermore, the relative cancellation efficiency of such a filter is determined by the sensor correlations and can be evaluated as

$$\epsilon_{\text{WF}}(f) = \sqrt{1 - \frac{\vec{\mathbf{C}}_{\text{SN}}^\dagger(f) \cdot \mathbf{C}_{\text{SS}}(f) \cdot \vec{\mathbf{C}}_{\text{SN}}(f)}{C_N(f)}} \quad (2.8)$$

where $C_N(f)$ is the spectral GGN density. This demonstrates an important limitation for WF cancellation: the optimal positioning of an array of seismic sensors is vital for good cancellation efficiency. The coupling of the sensor array to the GGN signal $\vec{\mathbf{C}}_{\text{SN}}$ should be maximized, whereas the correlation among the seismic sensors \mathbf{C}_{SS} must be kept as small as possible. Realizing such an optimized positioning can result in considerable expenses due to potentially extensive boreholes.

In addition to this geometrical limitation, a further performance limitation is opposed by the instrumental noise of the seismic sensors. The latter enters the efficiency via the diagonal entries of the sensor correlation matrix C_{SS} .

Moreover, accurate modeling of the sensor-to-detector cross-correlation \vec{C}_{SN} imposes the biggest challenge for active GGN cancellation with WFs. It requires adequate knowledge about hardly accessible information of the seismic field, such as the composition of longitudinal and transversal propagation modes. Local measurement techniques are not feasible at this point, and a potential resolution would require hundreds of seismometers in boreholes [4].

2.3.2 CANCELLATION WITH DEEP LEARNING

As previously discussed, challenges of the current active GGN cancellation approach with WFs are the model dependency on the one hand and fundamental limitations due to seismic sensor noise and sensor positioning on the other.

A novel approach that aims to overcome these challenges, is investigated in this thesis. The idea is to replace the linear filter with a highly non-linear deep neural network that learns to replicate the transfer function from the seismic sensor signal to the GGN estimate.² The network requirements connected to this regression task are that it needs to operate on seismic field data containing both spatial and temporal features. Architectures proven efficient for such tasks exist and are under further development [37, 38, 39].

A significant advantage of this approach is that the Neural Network (NN) is not model-dependent. Theoretically, it should learn some state of latent abstract representation for the underlying correlation matrices between seismic sensors and GW detector and does not require external correlation information. The only limitation is the availability of the true GGN signal for the training. However, once GGN dominates the detector noise, a potential training with real data should be feasible, as discussed in ch. 1. Preliminary engineering and commissioning tests to optimize suited network architectures can be performed using adequate simulations. In theory, the deep learning approach has the potential to reduce the WF limitations by superior instrumental noise robustness or more flexibility in sensor positioning. At this point, however, this is speculative and needs thorough investigation.

A proof of concept for the proposed deep learning approach for active GGN cancellation is pending. This proof and the demonstration of a suited network architecture are the main objectives of this work.

²see fig. 1.1

DEEP NEURAL NETWORKS

This chapter is intended to provide the reader with a basic understanding of Deep Learning (DL). The goal is not an exhaustive resume but an introduction to the necessities and techniques required to build the deep Neural Network (NN) architecture as used in this work.

3.1 FUNDAMENTAL CONCEPTS

DL is the subdomain of machine learning that includes the application of deep NNs for artificial intelligence tasks. More precisely, it is the approach of utilizing multi-layered NNs to find increasingly abstract representations of the input features. Depending on the given task, these learned latent representations are then employed for affine mappings. State-of-the-art applications for natural language processing, computer vision, and a generally large field of artificial intelligence applications achieved significant progress due to DL. [40, 41] What is referred to as simply NNs are actually *artificial* NNs that were initially designed to model the biological archetypes as found in brains¹ [42]. The field of application rapidly increased as NNs have proven to be efficient universal approximators for any real function up to arbitrary precision given a sufficiently large network parameter space [8, 43]. The central concepts will be presented hereafter based on [40, 41, 44].

3.1.1 MULTILAYER PERCEPTRON

Artificial Neurons The elemental building blocks of NNs are artificial neurons, also called *perceptrons*. They are designed based on the biological neuron as depicted in fig. 3.1. These biological neurons receive electric input signals via the *dendrites*. If the total sum of signals in the *nucleus* exceeds a neuron-specific threshold value, it will emit an output signal via the *axon*. This in turn is passed on to other neurons, which together form a network. [42]

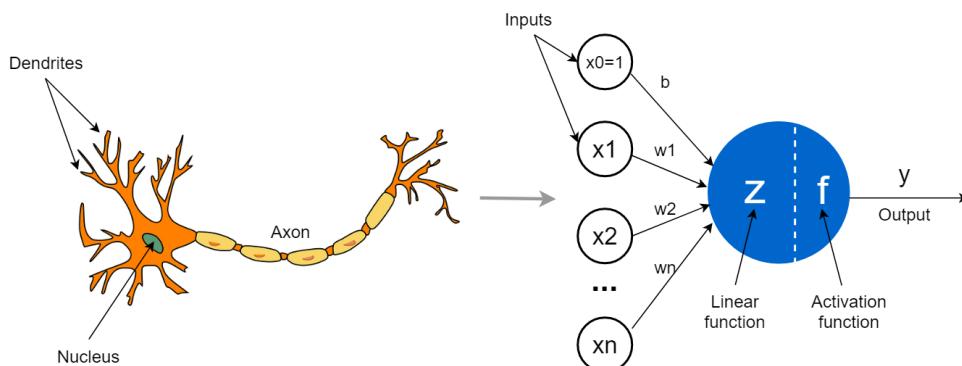


Figure 3.1: The biological neuron as archetype for artificial neurons.

The structure of biological neuron (*left*) compared to the artificial neuron (*right*). Both are discussed in detail in the text. Adapted from [45].

¹In this work, the term NN will refer to the artificial NNs unless explicitly stated otherwise.

In contrast to the biological neuron, the artificial neuron receives numerical inputs $x_i \in \vec{x}_n$. Each input originating from a connected neuron i is evaluated with a specific weight w_i to produce a linear mapping

$$\vec{z}_m = \mathbf{W}_{m \times n} \cdot \vec{x}_n + \vec{b}_n \quad (3.1)$$

where m and n are the dimensionalities of outputs and inputs, respectively. The weight matrix $\mathbf{w}_i \in \mathbf{W}_{m \times n}$ and bias vector $b_i \in \mathbf{b}_n$ are intrinsic numerical features of a specific neuron that play a crucial role in the adaptation to a specific task. The output of a neuron is additionally transformed via a non-linear activation function

$$\vec{y}_m = f_\sigma \left(\mathbf{W}_{m \times n} \cdot \vec{x}_n + \vec{b}_n \right) \quad (3.2)$$

in order to achieve non-linear mappings through a network of artificial neurons. Such a network typically consists of multiple *fully connected layers* that each host a specific number of individual neurons. These layers are also called *hidden layers* if they are used to map features from an *input layer* to a distinctive *output layer* as displayed in [fig. 3.2](#).

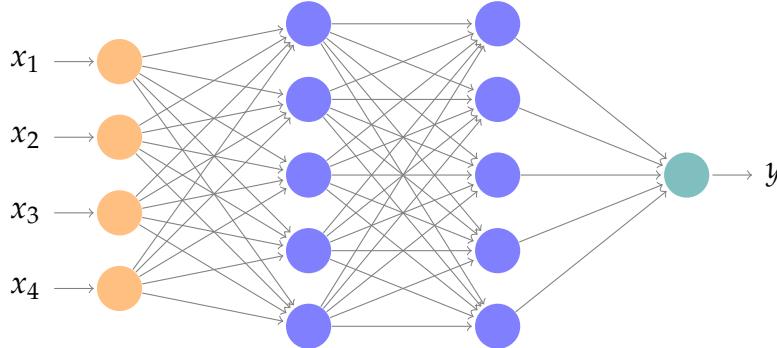


Figure 3.2: Architecture of a shallow multilayer perceptron.

The artificial neurons are organized in two hidden layers (*blue*) to produce a non-linear mapping of the four-dimensional input (*orange*) to a single output (*green*). Weights, biases, and activation functions are applied in between but are not displayed.

3.1.2 GRADIENT DESCENT AND BACKPROPAGATION

Building a NN requires the initialization of its parameters $\theta \equiv \{\mathbf{W}, \vec{b}\}$. The exact method thereby depends on the specific use case and will be discussed in more detail in [sec. 3.2.2](#). At this stage, however, the network is impracticable for the intended task. A process called *training* is needed to optimize the set of parameters θ . This is also referred to as the network *learning* a specific task. In a *supervised* learning scenario, the network is trained with a set of inputs $x_i \in \vec{x}$ and the corresponding true labels $\hat{y}_i \in \hat{y}$. For this purpose, a measure of performance is provided by the *objective function*, also called the *loss*. The standard loss in regression tasks is the Mean Squared Error (MSE) defined as

$$\mathcal{L}_{\text{MSE}}(\vec{x}, \theta) = \frac{1}{N} \sum_{i=0}^N [\hat{y}_i - y_i(x_i, \theta)]^2 \quad (3.3)$$

where $y_i(x_i, \theta)$ is the networks output given the parameters θ and inputs $\vec{x} \in \mathbb{R}^N$.

For each input x_i the network prediction is compared with the true label \hat{y}_i to evaluate the mean over the N dimensional dataset. Note that, in general, the individual samples from the dataset are not restricted to scalars but can also be higher dimensional. For this work, a time series regression is relevant, implying that each training sample is a vector with the same dimensionality as the number of time steps.

Gradient Descent The loss quantifies the performance of the network during training and is the central point of reference for the training. A set of parameters that minimizes the loss function if evaluated over the training data is aimed for

$$\hat{\theta} = \arg \min_{\theta} [\mathcal{L}(\vec{x}, \theta)] \quad (3.4)$$

This purpose is served by a method called *gradient descent*. It can be expressed as

$$\theta \rightarrow \theta - \eta \nabla_{\theta} \mathcal{L}(\vec{x}, \theta) \quad (3.5)$$

and describes an iterative process of parameter optimization. With every step, the parameter set θ is updated by a step of size η , which is also referred to as *learning rate*. The direction of this step is determined by the gradient of the loss function with respect to the parameter space. For the basic network shown in [fig. 3.2](#), this parameter space would already have a dimensionality of 61, but much higher dimensions on the order of thousands to millions are typical. The optimization process can then be understood as the search for a global minimum in the surface of the loss function in the hyper-dimensional parameter space. On this surface, the negative gradient points to the direction of the steepest descent and determines the step direction of each parameter update. This is visualized in [fig. 3.3](#).

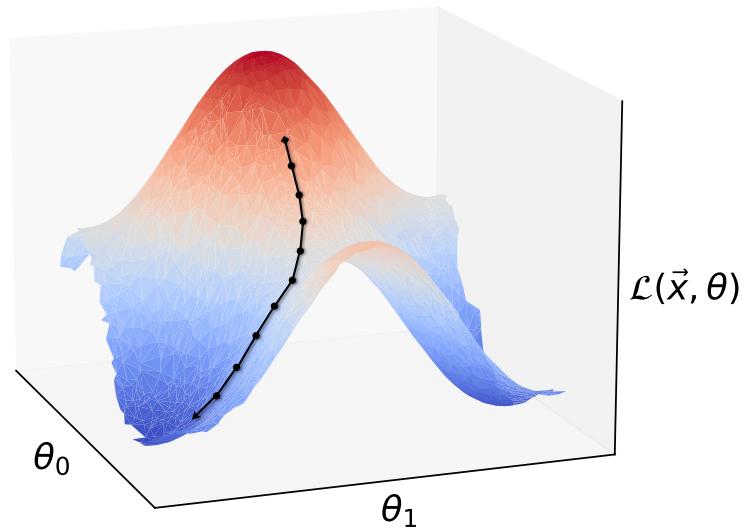


Figure 3.3: Optimization of network parameters with gradient descent.

A simplified illustration of the loss topology as the surface in a two-dimensional network parameter space. Starting from a randomly initialized point $\{\theta_0, \theta_1\}$, the network parameters get iteratively updated in the direction of the steepest gradient. The goal is to ultimately arrive at a global minimum $\{\hat{\theta}_0, \hat{\theta}_1\}$ to minimize $\mathcal{L}(\vec{x}, \theta)$. Typical parameter spaces are extremely high-dimensional, with often multiple local minima.

Backpropagation Computation of the gradient is a challenge that is commonly addressed with an algorithm called *backpropagation*. It is initiated by a *forward pass* through the entire network for each training instance. This means that the input is propagated through the network by applying each individual numerical operation connected to the neurons in one layer. The results are stored and propagation continues. Finally arrived at the output layer, the network prediction error is quantified with the specific loss function and each output connection contribution is evaluated analytically via the *chain rule*. This step is repeated interactively for every previous layer in a so-called *backward pass*. With this backward pass, a measurement of the gradient with respect to each neuron's parameter contribution is possible. Due to the analytical computability, the *backpropagation* algorithm is computationally efficient and considered the groundbreaking achievement for the successful training of NNs. [44, 46]

3.1.3 TRAINING AND MONITORING

Applying gradient descent and backpropagation to train a neural network is often practiced by splitting the dataset into *mini-batches* of training instances. This is referred to as Stochastic Gradient Descent (SGD) since the gradient is no longer computed over the whole data but estimated for smaller populations which introduces a certain stochasticity. For each mini-batch, backpropagation and a parameter update are performed and a complete iteration over all batches in the dataset is called an *epoch*. This method's obvious advantage is the computational cost reduction per parameter update. In most cases, SGD is even inevitable as the entire dataset would extend the accessible memory. Apart from this computational advantage, the stochastical fluctuation of the gradient can further help to escape shallow local minima and ensure a more homogenous performance of the network. In combination with SGD, a technique called *momentum* is applied to speed up convergence. With the use of gradients of previous iterations, the momentum of the gradient descent is estimated and applied to the current step to compensate partially for the statistical fluctuations. This is displayed in fig. 3.4.

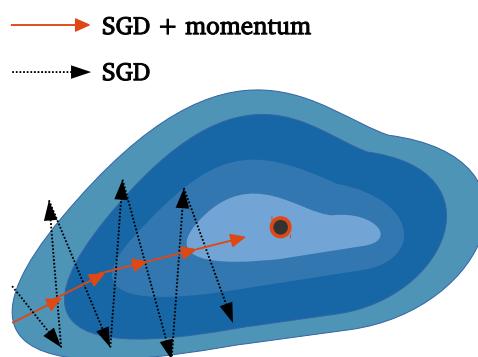


Figure 3.4: **Stochastic gradient descent with momentum.**

The loss topology is visualized in blue with a central minimum. SGD with momentum (*solid red arrows*) helps to reduce stochastic fluctuations and speed up convergence compared to the raw SGD (*dashed black arrows*). Adapted from [47].

Momentum is implemented with algorithms called *optimizers*. In this work, the *Nadam* optimizer is chosen due to its robustness and efficiency. It comprises *Nesterov-accelerated* momentum with an adaptive learning rate. [44, 48].

Monitoring & Generalization The decrease in loss can monitor training progress as a function of epochs. A loss that does not increase with epochs implies that the network is not able to learn, and in general, high loss values usually indicate *underfitting*. This means that either the network architecture or the training algorithm is deficient in comprehending the data's underlying pattern. On the other hand, scenarios are possible where the network might be too complex for the task and simply learns the presented training data «by heart». This is called *overfitting* and cannot be independently monitored by the training loss. In order to detect such behavior, a supplementary *validation* dataset is utilized. An additional *test* dataset is employed once as a final performance evaluation. The validation data is not used for the parameter updates but exclusively to test the network's *generalization* abilities as demonstrated in fig. 3.5.

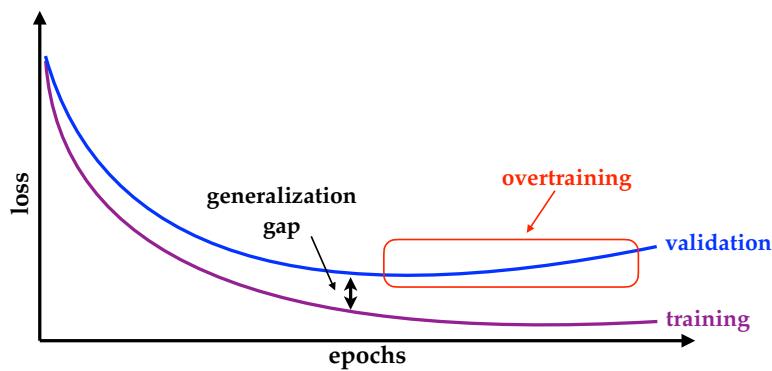


Figure 3.5: Training and validation loss for performance monitoring.

The training (purple line) and validation loss (blue line) as a function of training epochs. The validation loss is used to monitor the network's generalization capabilities (black) and potential overtraining (red).

A validation loss that starts to increase while the training loss still decreases or stagnates is an indication of *overtraining*. The network is learning specific features of the training data and the generalization gap broadens. An aspired network performance includes an overall low loss with a small generalization gap. The common way to achieve this are deep² network architectures in combination with *regularization* techniques to prevent overtraining. Specific regularization techniques, alongside further advanced methods, will be discussed in the following section.

²The term «deep» is not clearly defined. Originally, networks with more than two hidden layers were already considered deep in the 1990s. Nowadays, deep neural networks usually have dozens or hundreds of stacked layers. [44]

3.2 ADVANCED ARCHITECTURES

With the success of deep NNs for a large field of artificial intelligence tasks, advanced techniques were elaborated to meet the requirements of specific use cases and optimize performance in general. Some of these techniques play a crucial role in this work and will be presented hereafter.

3.2.1 REGULARIZATION

Deeper and more complex network architectures raise the requirement for more robust regularization techniques. The most straightforward approach is *early stopping*. If the validation loss stagnates or starts to increase over a specified number of epochs, early stopping interrupts the training and avoids overtraining. Additional *norm penalties* can help to mitigate overtraining by penalizing unreasonable large weight or bias values in particular neurons.

Although these measures are practical to circumvent overtraining, they cannot directly improve the intrinsical generalization capability of a network. [41, 44]

Dropout One of the most common regularization methods for deep network architectures is *dropout*. As the name suggests, the principle of the algorithm is to «drop out», i.e., set to zero, a random fraction of neurons during each training step. Except for the neurons in the output layer, every neuron has a predefined probability of being entirely omitted during the backpropagation. The predefined probability p is a hyperparameter that needs to be adjusted before the training has typical values of 0.1 to 0.5. Dropped-out neurons are only set to zero during training steps and integrated entirely during inference. Implying that, on average, each neuron has virtually more connections during inference than training. To account for this, the weights are scaled by the *keep probability* $1 - p$. Dropout has proven to be very efficient even for already optimized state-of-the-art networks. An explanation for the effectiveness is that individual neurons become less exposed to connected dominant neurons and, thus, more independent. The sensitivity to slight changes in inputs decreases, ultimately resulting in a more robust network with improved generalization. Altogether the better generalization comes at the cost of slower convergence during training. [44, 49]

More general regularization techniques include the use of more training data. *Data augmentation* can help to artificially increase the amount of data if the availability is limited. Furthermore, the injection of stochastic noise into the input data can benefit the generalization. In fact, this can be seen as a form of data augmentation. The network learns a certain robustness to deviations in the data and is less prone to overfitting. This could play an interesting role for seismometer noise in the context of GGN denoising with deep learning and will be discussed in sections 5.2 and 5.3. [41, 50]

3.2.2 NONSATURATING ACTIVATION FUNCTIONS

Activation functions are needed to reconstruct non-linear mappings with the otherwise linear function described by NNs. As introduced in section 3.1.1, the activation function is applied to the output of each neuron. The backpropagation algorithm implies the requirement of differentiability. Activation functions in the output layer are either linear, or their range is orientated on the possible output space, for instance 0 to 1 for classification probabilities.

Initially, the *Heaviside* step function was used to mimic the firing threshold of biological neurons [42]. Although this may be an appropriate option for logic-related tasks, more monotonic functions such as *sigmoid* or later *tanh* became the means of choice. Nowadays, the Rectified Linear Unit (ReLU) function and its variants are the most commonly used activation functions due to their computational efficiency and nonsaturating properties. [44, 51] The activation function as used in this work is leakyReLU [52], a variant of ReLU that is defined as

$$\text{leakyReLU}_\beta(x) = \max(\beta x, x) \quad , \quad \beta \in \mathbb{R} \quad (3.6)$$

The discussed non-linear activation functions are presented in fig. 3.6 alongside their (piecewise) derivatives.

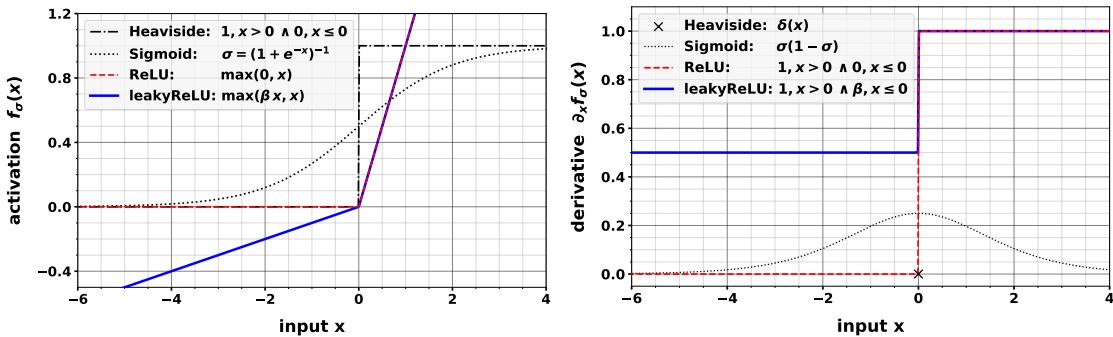


Figure 3.6: Important activation functions and their derivatives.

The activation functions and their definitions are on the *left*. The corresponding derivatives play a crucial role during backpropagation and are displayed on the *right*. LeakyReLU (blue solid line) is the function of choice for this work.

Vanishing & Exploding Gradients A problematic effect that can prevent the successful training of deep network architectures are vanishing or exploding gradients. Since the backpropagation algorithm includes a layer-wise multiplication of gradients (chain-rule), already small or large gradients get amplified and get even smaller or larger. This can result in stagnating training updates or that effectively only the lowest layers of a network get trained. For exploding gradients, a large fluctuation of the loss function, which finally leads to divergence, can be observed. Logistic activation functions like sigmoid or tanh are especially prone to vanishing gradients because of their saturating properties. The reason can be observed in fig. 3.6: If the input values of a layer are not close to zero, a logistic activation like sigmoid saturates towards small or large values. The derivatives become increasingly smaller, leaving virtually no gradient to propagate. The *Glorot-initialization* [53] is an approach to overcome this issue. In order to break the symmetry of a network, i.e., increasing the non-uniformity of neuron weights and biases during training, the values get initialized to random numbers. Glorot-initialization draws the values for each layer from an independent density function that considers the number of connected neurons with the intention of keeping the values closer to zero to avoid saturation. [44, 53]

Dying ReLUs Rectified activation functions like ReLU are nonsaturating and thus less prone to vanishing gradients. However, they come at the risk of *dying ReLUs*. This is an empirically discovered phenomenon where individual neurons with ReLU activation «die», i.e. they only output zero. This can occur if the weighted sum over all connected neurons $W \cdot \vec{x}$ for an individual neuron is negative for all training instances. Note that negative input data does not necessarily increase the risk of dying ReLUs since the weights can take negative values as well. *He-initialization* [10] follows a similar approach to Glorot-initialization with the aim of minimizing the risk of dying ReLUs through sampling from an appropriate density function. An effective solution to avoid dying ReLUs is the use of He-initialization in combination with the leakyReLU variant as defined in eq. (3.6). The slope for negative values ensures that the activation function does not stagnate for negative inputs. Other variants like *pReLU*, *ELU* or *SELU* exist but are generally more computationally extensive. [10, 44, 52]

Batch Normalization While He-normalization in combination with leakyReLU is efficient in bypassing dying ReLUs, additional measures might be required to prevent vanishing or exploding gradients in deep architectures. *Batch normalization* serves this purpose by normalizing the inputs before or after the activation function in each hidden layer. The idea is to zero-center the inputs with a normalized variance. Implementation is achieved with the batch-wise evaluation of mean and standard deviation. The network then essentially learns two parameters to shift and scale the inputs in order to achieve normalized gradients that do not vanish or explode. A further benefit of the normalized gradients is a certain regularization as the weights are adapted more homogeneously. The extra normalization step consequences in a longer runtime per epoch which the overall faster convergence mostly compensates for. [44, 54]

3.2.3 RESIDUAL UNITS

Cutting-edge networks tend to utilize increasingly deep architectures to improve performance. With a certain depth, techniques like batch normalization and proper initialization can become insufficient to prevent vanishing or exploding gradients. To ensure trainability even at significant depths, so-called *residual units* have proven to be very efficient. The idea is to add skip connections to the hidden neuron layers as sketched in fig. 3.7. An input x propagates through the layers producing the residual $\mathcal{F}(x)$. The skip connection leads to an identity mapping which is finally concatenated with the residual. In between and subsequently, activation functions f_σ are applied. [55] Optionally, a learnable zero-initialized parameter α can be multiplied to the residual to form a *ReZero* unit. [56] An important aspect of the effectiveness of these residual units is that the network no longer has to learn a transfer function $h(x)$ but only the modifications to the input via $f(x) = h(x) - x$. This can speed up the convergence significantly. [44, 55, 56]

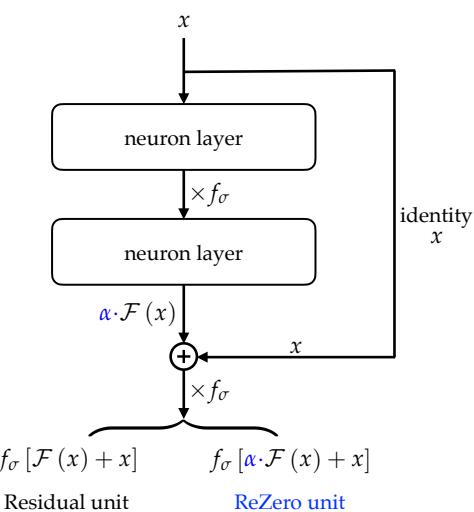


Figure 3.7: Sketch of a residual unit.

Since network parameters are commonly initialized with random values around zero, the initial predictions before training are also close to zero. With residual units, however, the initial prediction is closer to the network's input, which can benefit the divergence. For instance, with time series regression of signals that oscillate around zero, an initial prediction close to zero is very disadvantageous since it can be understood as a local minimum or saddle point of the MSE loss topology. Another crucial benefit of residual units is the improved gradient flow during backpropagation. Even if certain layers have small or vanishing gradients, the skip connections still allow a gradient flow and shallow network layers incorporated more efficiently. [41, 44]

3.3 SPATIOTEMPORAL CONVOLUTIONS

Although network architectures consisting of multiple stacked layers of perceptrons can approximate any real function in theory [8], the implementation in a real-life application is challenging. Computational resources and the optimization of the parameters are limiting factors. Especially in computer vision applications, the vast parameter space aggravates the training process, and symmetry considerations become necessary. For this purpose, Convolutional Neural-Networks (CNNs) have proven to be the most powerful tool.

They are based on findings about the biological visual cortex: Each visual neuron has a limited receptive field and mostly reacts to specific patterns only. It has been shown that specific neurons in a cat's brain only react to horizontal patterns, while others exclusively react to vertical patterns. These patterns can have strongly varying orders of complexity, which makes the ensemble in the visual cortex very efficient³ and powerful. [58, 59]

CNNs aim to mimic this methodology and have achieved considerable success for computer vision tasks, including image search engines and self-driving cars. For specific tasks like image classification, deep CNNs have even achieved super-human performance [10]. The main advantage of CNNs is that they can extract spatial patterns in image-like data with an efficient number of parameters. However, their application is not limited to spatial data but can be suitably extended to extract temporal patterns as well. The technical details that enable these capabilities are discussed in the following sections.

3.3.1 SPATIAL CONVOLUTIONS

The building blocks of CNNs are *convolutional layers*. *Spatial convolutions* particularly address the extraction of patterns from data with underlying spatial distributions. An instance is image-like data, where the position of each pixel is essential for the resulting picture. More abstract data structures like a sensor arrangement with different positions can also be understood as «image-like» data in this context. A convolutional operation is displayed exemplarily for a grayscale image in fig. 3.8. The approach is that the input image gets scanned by multiple, usually significantly smaller, *filters* or *kernels*. These filters consist of individual neurons with learnable weights and biases. Each neuron processes the input of a single pixel and together, a filter can learn to focus on a particular pattern.

³In fact, most biological networks have evolved to be very efficient with a minimal number of neurons. A large field of studies, e.g. [57], investigates reasons for their efficiency to employ them in artificial networks ultimately.

For the instance shown in fig. 3.8, the weights of the filter have adapted to be sensitive to corners and ignore most other shapes. Since the filter slides over the image, the features are translational invariant, i.e., they can be processed independently of their position in the image.

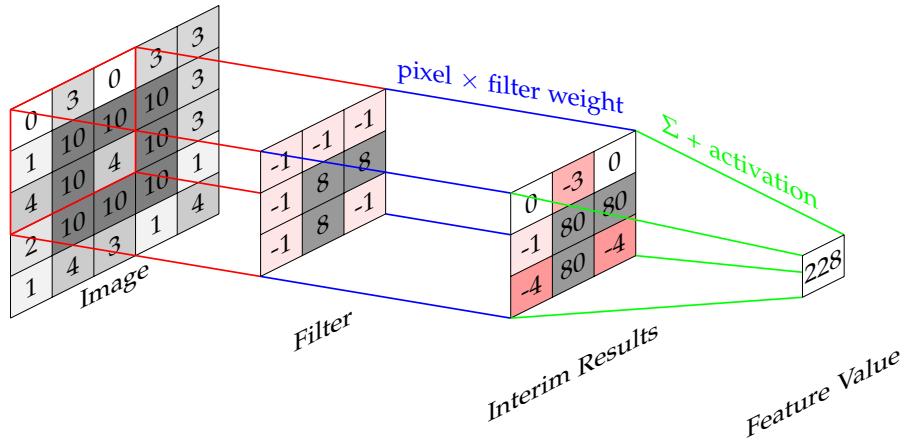


Figure 3.8: Spatial two-dimensional convolution for an image.

On the left is an image of a dark square consisting of 5×5 grayscale pixels. A 3×3 filter is applied to the pixels bordered in red. Filter weights and pixel values are multiplied and produce the interim results. A final summation and activation function of choice finally results in the feature value on the right. Subsequently, the filter slides over the image, and new feature values are computed with the same filter weights but different pixel regions. The dimension of the feature map depends on the predefined filter positions. Adapted from [60].

Sliding the filter over multiple regions of an input image results in multiple feature values. They again form an image, the so-called *feature map* of a specific filter. In each convolutional layer, several filters can be applied simultaneously, resulting in a corresponding number of feature maps that act as input images for the following layer. Hence, the hyperparameters defining a convolutional layer are the total number of filters, the kernel size of each filter, and the activation function. Stacking multiple convolutional layers can favor a *hierarchical learning* process, i.e., the degree of feature abstractness increases with depth. A large variety of increasingly abstract features usually favors the generalization capability of a CNN. Moreover, since each filter applies the same weights to different regions of the image («*weight sharing*»), CNNs can reach a significant level of noise-robustness and are generally less prone to overfitting. [9, 44]

Padding Another hyperparameter for the implementation of a convolutional layer are the regions where the filter should be applied to the input image. For the example shown in fig. 3.8, there are 3×3 possibilities to position the filter over the image if overlapping regions are allowed. However, if the corner of the square would coincidentally lay in the corner of the image, the filter would not be able to detect it. To avoid such artifacts, a technique called *padding* can be applied. It comprises a padded frame of artificial pixels padded around the input image. These pixels are usually filled with zeros and can be used to keep the dimensionality of the data constant through the convolutional layers. [41]

Striding Depending on the ratio of filter size to image size, it can be beneficial to apply a given filter, not to every possible pixel region but only every second, for instance. This is called *striding* and can be understood as lowering the «sampling rate» of the sliding filter to increase the receptive field of each filter. [41]

Pooling Although CNNs are very efficient in processing high-dimensional inputs with a comparably small number of network parameters, the training can get memory extensive due to the backpropagation of multiple filters. *Pooling* is a common method to reduce the dimensionality within the network's layers to minimize the computational costs. Furthermore, a reduced dimensionality can force the network to find more abstract feature representations, which in turn can have a positive influence on the overall performance and especially the generalization. The pooling algorithm is not very different from a convolutional operation with the essential difference that no weights get multiplied, and no parameters are learned. A pooling kernel of the desired size is slid over the image, and the pixel values are aggregated. An operation is performed to reduce the aggregated values in a kernel to a single pooled value. The most common operations are *average pooling* and *max pooling*, where either the average or maximum value of the ensemble is applied respectively. By sliding the pooling kernel over the input image, a pooling map is created in analogy to the convolutional feature map. Stacking several convolutional layers followed by a max pooling operation usually serves the purpose of filtering dominant features out of less important features for the specific task. [44]

Spatial Dropout Dropout in CNNs can also be applied in the form of *spatial dropout*. The method helps with regularization in the same way as the regular dropout as discussed in sec. 3.2.1. The main difference is that not only individual neurons are dropped out, but entire feature maps. It is reasonable to apply a spatial dropout layer immediately before a max pooling layer to prevent overfitting to distinct features only and allow subdominant features to play a role. [61]

Graph Convolutions The discussed spatial convolutions were introduced for image-like data on regular grids. Although this is their largest field of application, it is possible to extend the techniques to use cases with irregularly arranged data. Examples include curved detector geometries in particle physics [62] or irregularly arranged seismic sensors [37]. The approach is to encode the spatial information in mathematical structures called *graphs*. With this structure, a spatial convolutional approach, that is very similar to the regular case, or a spectral convolution approach can be followed. The spectral approach has its mathematical foundation in the spectral theory of graphs [63], and a large field of spectral convolution techniques has developed on this basis. Altogether the interest and research invested in graph convolutional networks is increasing, and graph networks are designed that equal the performance of regular CNNs and extend their field to irregular data structures. [64] In the context of this work, they can become a powerful tool for the adaption to an irregularly arranged seismic sensor array for active GGN cancellation at ET.

3.3.2 TEMPORAL CONVOLUTIONS

The application of convolutional layers is not limited to spatial patterns but can also adequately be applied to exploit temporal patterns in time series data. Use cases where *temporal convolutions* have achieved major success are particularly natural language processing tasks like voice recognition or text-to-speech generation. [44, 65] The operation of temporal convolutions is analogous to the spatial case, with the only difference being that the data is the time domain, i.e., one-dimensional only. The convolutional layers now no longer operate on pixels but on time steps. Methods like padding, striding, and pooling can be applied in the same manner.

Dilated Temporal Convolutions A concept that is particularly useful for processing large time series with hundreds to thousands of time steps are *dilated temporal convolutions*. The idea is that the temporal filter is not continuous but atrous. This means that the kernel does not include neighbored time steps only but can leave gaps and only include every d^{th} time step in the convolution. The *dilation rate* d can be set for every layer independently which allows for different levels of temporal patterns. A successful approach is to increase the dilation rate exponentially with network depth as demonstrated in fig. 3.9.

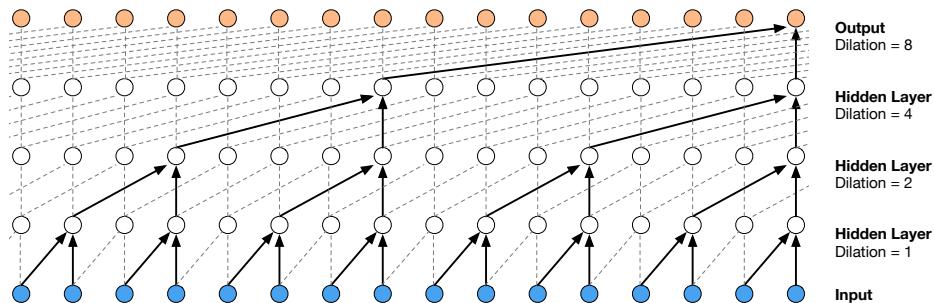


Figure 3.9: One-dimensional dilated temporal convolutions for time series.

Each circle represents a convolutional neuron corresponding to a distinct time step. Each layer performs a dilated convolutional operation with an increasing dilation factor of 2^i where i is the layer index. The receptive field of one convolutional neuron in the output is 16.

Adapted from [65].

This results in an exponentially increasing receptive field which allows very efficient processing of even extremely large time series. Lower layers will exploit more short-term temporal patterns, whereas the deeper layers make the connection to long-term patterns. [44, 65]

Recurrent Cells A commonly used alternative to temporal convolutions for processing sequential data are *recurrent cells* like LSTMs or GRUs. [44, 41] Their concept can be roughly summarized as unrolling the network through time, i.e., the input gets not only propagated through the layer but also through every time step. Even though this method can work well for relatively short time series, it suffers from vanishing gradients and a limited short-term memory for larger time series with hundreds of time steps as a direct consequence of the *backpropagation through time*. Since the gradients and inputs get propagated through time, early time steps can get significantly neglected. [44] For the purpose of this work, dilated temporal convolutions have superior properties and will be chosen over recurrent architectures.

A GRAVITY-GRADIENT NOISE MODEL FOR DATA GENERATION

In order to evaluate the potential application of neural networks in terms of GGN mitigation at ET, a suited simulation is needed to generate appropriate datasets. Due to the fundamental principles of machine learning, a crucial requirement for the successful training of a neural network is a reasonable large sample size. Therefore, the simulation has to meet the criteria of a tradeoff between, on the one hand, high computational efficiency and, on the other, a degree of physical realism that ensures the significance of the results.

This tradeoff primarily depends on the seismic simulation's complexity, which constructs the foundation of any GGN model. Seismic simulations addressing the potentially complicated geologies and seismic characteristics¹ of a possible host site in realistic detail, are of high complexity. Effects like scattering at inhomogeneities in the surrounding medium or contributions arising from the site-specific infrastructure have to be considered. The potential complexity is underlined by the fact that even noise sources such as one of the fans of the ventilation system at LIGO or wind turbines located several kilometers away can have a non-negligible influence. [66, 67]

Based on the above considerations, it becomes legitimate that state-of-the-art simulations addressing realistic topologies often depend on elaborate numerical methods. For instance, in the context of GGN modeling for ET, a simulation study based on seismic characterization measurements at a potential host site in the EMR is presented in [3]. The geology at the site is identified as a horizontally layered structure of several soil types, and the simulation relies on numerical wave equation solvers. The effort involved in such a simulation is underlined by the fact that it has been developed in an interdisciplinary science and business collaboration [68, 69].

In summary of the above-listed arguments, it becomes apparent that the complexity of a seismic simulation can be arbitrarily high and scales strongly with the degree of detail. However, as the intention of this study is a first proof-of-concept and the development of a suited pipeline to evaluate a potential GGN cancellation scheme at ET, several assumptions are made to simplify the problem and set up a fast simulation tool that is based on analytical calculations. These assumptions, alongside the theoretical foundations and physics implementations, are described in the following sections.

¹e.g., the propagation modes, dispersion relation, and angular distribution of seismic activity

4.1 THEORETICAL FOUNDATIONS OF THE MODEL

According to Newton's theory of gravity, the GGN acceleration imposed by a density perturbation $\delta\rho(\vec{r}, t)$ on a test mass located at \vec{r}_m is given by

$$\delta\vec{a}(\vec{r}_m, t) = G \int \delta\rho(\vec{r}, t) \frac{\vec{r} - \vec{r}_m}{|\vec{r} - \vec{r}_m|^3} d^3\vec{r} \quad (4.1)$$

where $G = 6.674(10) \times 10^{-11} \text{ m}^3/\text{kg s}^2$ [70] is the gravitational constant. [30] The cause of density perturbations is not limited to seismic activity but can also include atmospheric turbulences such as strong winds or fluctuations caused by fields of pressure, temperature, or humidity [2, p. 170]. As a countermeasure, it is foreseen to locate ET in an underground location in order to suppress atmospheric contributions to a negligible order of magnitude [2]. The validity of this assumption is a question of current research, but first studies suggest that the suppressed atmospheric GGN lies below the ET-D design sensitivity [31].

The remaining dominant contributions are of seismic nature and can be described by the continuity equation

$$\delta\rho(\vec{r}, t) = -\nabla \cdot [\rho(\vec{r}, t) \vec{\xi}(\vec{r}, t)] \quad (4.2)$$

where $\vec{\xi}(\vec{r}, t)$ denotes the seismic displacement field relative to the equilibrium, and $\rho(\vec{r}, t)$ is the density of the medium. [30] As the knowledge of $\rho(\vec{r}, t)$ and $\vec{\xi}(\vec{r}, t)$ will never be complete, models are needed to describe these quantities in an adequate manner. The seismic displacement field is generally composed of different seismic wave modes and is, in first-order, categorized into seismic body and surface waves. The latter cause vertical displacements on the surface and are the dominant GGN source for aboveground GW detectors. At the foreseen underground location for ET in the EMR, however, it has been assessed that the attenuation due to the wave dispersion is suppressing the surface wave GGN contributions to a subdominant order of magnitude [3]. This implies that seismic body waves are the dominant source of GGN at ET; thus, the simulation is restricted to their contributions.

The ground-laying approach to the simulation can be described as an elemental *Monte-Carlo* method where the seismic wave field $\vec{\xi}(\vec{r}, t)$ is constructed through the stochastic sampling of individual seismic waves and their properties. This wave field is the basis to calculate the interferometers mirror dislocation from equilibrium and finally correlate the mirror dislocations with the ET geometry to obtain a GGN strain estimate. In the following sections, the analytical details of the aforementioned simulation steps are listed in more detail.

4.1.1 SEISMIC BODY WAVES

The standard in seismology for an analytical description of seismic waves is based on the theory of continuum mechanics in a linear elastic solid medium [71]. The homogenous equation of motion in such a linear elastic medium is given by

$$\rho \ddot{\vec{\xi}}(\vec{r}, t) = (\lambda + 2\mu) \nabla \cdot [\nabla \cdot \vec{\xi}(\vec{r}, t)] - [\mu \nabla \times \nabla \times \vec{\xi}(\vec{r}, t)] \quad (4.3)$$

where $\vec{\xi}(\vec{r}, t)$ is the displacement field relative to the equilibrium position.

Further, ρ is the mass density and λ, μ are the material specific Lamé constants.[71] Following the Helmholtz theorem of vector fields [72], the seismic displacement field can be equally expressed through the sum of a rotation-free seismic scalar potential $\nabla \phi_p(\vec{r}, t) = 0$ and a divergence-free seismic vector potential $\nabla \times \vec{\psi}_s(\vec{r}, t) = 0$ as

$$\vec{\xi}(\vec{r}, t) = \nabla \phi_p(\vec{r}, t) + \nabla \times \vec{\psi}_s(\vec{r}, t) \quad (4.4)$$

and the equation of motion eq. (4.3) can be written as

$$\begin{aligned} & \nabla \left[(\lambda + 2\mu) \Delta \phi_p(\vec{r}, t) - \rho \ddot{\phi}_p(\vec{r}, t) \right] \\ & + \nabla \times \left[\mu \Delta \cdot \vec{\psi}_s(\vec{r}, t) - \rho \ddot{\vec{\psi}}_s(\vec{r}, t) \right] = 0 \end{aligned} \quad (4.5)$$

This formulation shows that the equation of motion is solved if each seismic potential satisfies an independent harmonic wave equation:

$$\begin{aligned} \Delta \phi_p(\vec{r}, t) - \frac{1}{c_p^2} \ddot{\phi}_p(\vec{r}, t) &= 0 \quad \text{where } c_p \equiv \sqrt{\frac{\lambda + 2\mu}{\rho}} \\ \Delta \vec{\psi}_s(\vec{r}, t) - \frac{1}{c_s^2} \ddot{\vec{\psi}}_s(\vec{r}, t) &= 0 \quad \text{where } c_s \equiv \sqrt{\frac{\mu}{\rho}} \end{aligned} \quad (4.6)$$

This implies that seismic body waves can be classified into two main types, a longitudinal pressure wave, represented by a scalar potential and propagation velocity c_p , and a transversal shear wave, represented by a vector potential and propagation velocity c_s . Based on the larger propagation velocity $c_p > c_s$, pressure waves are also called *primary* or *p-waves* whereas the shear waves are called *secondary* or *s-waves*. In the presence of a seismic source like an earthquake or anthropogenic activities, the equation of motion and seismic wave equations become in-homogenous and exemplary; the p-wave equation now reads

$$\square^p \phi_p(\vec{r}, t) = -4\pi \mathcal{F}(t) \delta(r_e) \quad (4.7)$$

where $\square^p \equiv \Delta - c_p^{-2} \partial^2 / \partial t^2$ can be summarized as the *d'Alembertian* operator, $\mathcal{F}(t)$ is the source function of the seismic activity and r_e is the *elastic radius* of the source which accounts for nonlinearities of the medium in the vicinity of the seismic source [71]. The general solution to this equation is a spherical wave propagating outward from the source origin.

In the limit of distant sources, however, the curvature of the spherical wave front becomes sufficiently small, and the wave can be adequately described as a plane wave. This has the benefit of further simulation steps becoming analytically solvable [30]. Such a plane wave with source origin \vec{r}_0 at time t_0 and positive propagation direction solves eq. (4.6) and can be expressed as

$$\phi_p(\vec{r}, t) = \phi_0 \exp \left\{ i \omega (t - t_0) - i \vec{k} \cdot (\vec{r} - \vec{r}_0) + \varphi_0 \right\} \quad (4.8)$$

where ϕ_0 is the maximum amplitude, and φ_0 is the initial wave phase. If the medium is assumed to be dispersion free, the wave vector \vec{k} is related to the angular wave frequency ω via $\vec{k} = \omega/c_p \hat{e}_k$ where \hat{e}_k is the propagation direction of the wave.

The corresponding seismic displacement field also describes a plane wave propagation and can be derived from eq. (4.4) as the real part of

$$\vec{\xi}(\vec{r}, t) = \nabla \phi_p(\vec{r}, t) = i \phi_0 \vec{k} \exp \left\{ i \omega(t - t_0) - i \vec{k} \cdot (\vec{r} - \vec{r}_0) + \varphi_0 \right\} \quad (4.9)$$

In the case of s-waves, the plane wave displacement field can be obtained analogously, with the only difference being the polarization with respect to the wave vector \vec{k} . The two wave types are parametrized as

$$\vec{\xi}_{p,s}(\vec{r}, t) = \xi^0 k_{p,s} \hat{e}_{p,s} \cos \left\{ \omega(t - t_0) - \vec{k}_{p,s} \cdot (\vec{r} - \vec{r}_0) + \varphi_0 \right\} \quad (4.10)$$

where $\vec{k}_{p,s} = k_{p,s} \hat{e}_k$ describes the propagation of the wave whereby the wavenumber $k_{p,s} = \omega/c_{p,s}$ is defined via the propagation frequency and speed. The maximum amplitude is denoted as ξ^0 and for p-waves, the polarization is longitudinal $\hat{e}_p \cdot \hat{e}_k = 1$ and transversal $\hat{e}_s \cdot \hat{e}_k = 0$ in the case of s-waves.

As realistic seismic sources have finite radiation durations, the waves are modeled as gaussian wave packets [35], and the seismic displacement field gains an additional factor

$$\exp \left\{ \frac{-\tau^2}{(2\Delta T)^2} \right\} \quad \text{with} \quad \tau \equiv t - t_0 - \frac{\vec{r} - \vec{r}_0}{c_p} \quad (4.11)$$

where τ is the retarded time relative to the wave reaching the observer and ΔT is the width of the wave packet.

Another property that has to be considered for wave propagation is attenuation due to energy dissipation. This effect is generally very challenging to model as it strongly depends on geological properties. A fundamental property, however, is that the attenuation due to absorption and scattering decreases for low seismic frequencies. For the relevant frequencies of 1 to 10 Hz, the absorption attenuation is very small and dominated by the geometrical attenuation. [71] As the geometric attenuation scales with the distance to the source as $\propto 1/r^2$, for plane waves originating from distant sources, the geometrical attenuation has an insignificantly small impact on the length scale of the interferometer mirrors of $\mathcal{O}(10 \text{ km})$.

4.1.2 GRAVITY-GRADIENT ACCELERATION

The GGN acceleration caused by seismic activity can be better understood by performing the differential operation on the continuity equation eq. (4.12):

$$\begin{aligned} \delta\rho(\vec{r}, t) &= -\nabla \cdot [\rho(\vec{r}, t) \vec{\xi}(\vec{r}, t)] \\ &= -\underbrace{\nabla \rho(\vec{r}, t) \cdot \vec{\xi}(\vec{r}, t)}_{\text{surface}} - \underbrace{\rho(\vec{r}, t) [\nabla \cdot \vec{\xi}(\vec{r}, t)]}_{\text{bulk}} \end{aligned} \quad (4.12)$$

In this form, it becomes visible that the integration in eq. (4.1) is executed over two distinct contribution terms. The first term describes contributions arising from the surface motion of the medium due to the seismic displacement field (the so-called *Rayleigh* or *Love* modes), and the second term accounts for contributions due to density changes in the bulk medium.

Inserting eq. (4.4) into the bulk term

$$\begin{aligned}\delta\rho^{\text{bulk}}(\vec{r}, t) &= -\rho(\vec{r}, t) \nabla \cdot [\nabla\phi(\vec{r}, t) + \nabla \times \vec{\psi}(\vec{r}, t)] \\ &= -\rho(\vec{r}, t) \Delta\phi(\vec{r}, t) + \rho(\vec{r}, t) \underbrace{\nabla \cdot [\nabla \times \vec{\psi}(\vec{r}, t)]}_{\text{div}[\text{rot}(\vec{x})]=0 \forall \vec{x}} \\ &= -\rho_0 \Delta\phi(\vec{r}, t)\end{aligned}\quad (4.13)$$

shows that only p-waves with scalar potential cause density changes in the bulk medium, and s-waves only enter the GGN via the surface term, i.e., contributions arising from the cavities.

In the case of plane seismic waves, it is possible to analytically solve eq. (4.1) under the simplifying assumption of test masses located in the center of spherical cavities of radii a surrounded by a homogenous medium of constant mass density ρ_0 . This results in [30]:

$$\delta\vec{a}^p(\vec{r}_m, t) = 8\pi G\rho_0 \vec{\xi}^p(\vec{r}_m, t) \frac{j_1(k^p a)}{k^p a} \quad (4.14)$$

where the spherical *Bessel* functions of the first kind $j_n(x)$ are defined as [73, pp. 622-632]:

$$\begin{aligned}j_n(x) &= 2^n x^n \sum_{i=0}^{\infty} \frac{(-1)^i (i+n)!}{i! (2i+2n+1)!} x^{2i} \\ &= (-1)^n x^n \left(\frac{d}{x dx} \right)^n \frac{\sin(x)}{x}\end{aligned}\quad (4.15)$$

$$\text{and } j_1(x) = \frac{\sin(x)}{x^2} - \frac{\cos(x)}{x}$$

The contribution of plane s-waves can be found through the same integration approach as [30]:

$$\delta\vec{a}^s(\vec{r}_m, t) = -4\pi G\rho_0 \vec{\xi}^s(\vec{r}_m, t) \frac{j_1(k^s a)}{k^s a} \quad (4.16)$$

And it is noteworthy that the dependency on $k^s a$ is the same as for the p-wave contribution, even though the s-waves only enter via the surface term of eq. (4.12). This can be understood by the perpendicular wave polarization that merely introduces a differing geometric coupling term of $-1/2$. A plot of the geometrical coupling factors for p-waves and s-waves is shown in fig. 4.1.

It must be noted that the solutions in eqs. (4.14) and (4.16) do not include scattering effects of the incident waves from the cavities. However, in [30], the gravity perturbations originating from scattered incident p-waves and s-waves are analyzed with the result that the contributions have a negligible impact on the GGN acceleration as long as the cavity radius is much smaller than a seismic wavelength. For cavity radii on the order of $\mathcal{O}(10^1 - 10^2 \text{ m})$ this criterion is met as the expected seismic wavelengths are of order $\mathcal{O}(10^3 \text{ m})$ [3].

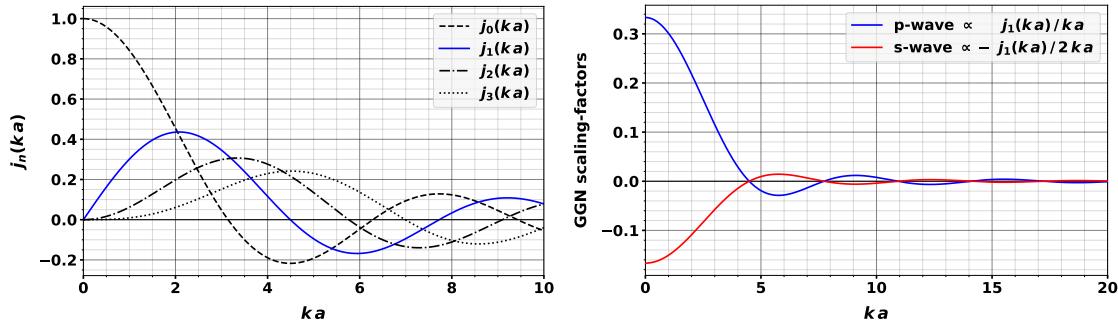


Figure 4.1: Geometrical scaling factors of GGN cavity contributions.

left: The first four spherical *Bessel* functions of the first kind.

right: The coupling factors for p-waves (blue) and s-waves (red) as a function of cavity radius.

To further estimate the impact of different cavity shapes, the *Bessel* factor can be expanded for small cavities as

$$\frac{j_1(ka)}{ka} = \frac{1}{3} \left[1 - \frac{(ka)^2}{10} - \frac{(ka)^4}{280} + \mathcal{O}\left((ka)^6\right) \right] \xrightarrow{a \rightarrow 0} \frac{1}{3} \quad (4.17)$$

and for vanishing cavities, the numerical coupling factor between seismic displacement field and GGN acceleration becomes $4\pi/3$. In this limit, it is possible to compare the assumption of spherical cavities to different cavern shapes and test-mass locations. The summary in [table 4.1](#) includes analytical and numerical solutions for alternative cavity models and shows that, at least in the small-cavity limit, the assumption of spherical cavities results in the most extensive coupling. This motivates the choice of spherical caverns for the simulation, which is in agreement with recent simulation studies [3, 74].

Table 4.1: GGN coupling factors for different cavity models.

The spherical cavity model in comparison to different cavity models in the small-cavity limit. Adapted from [4].

Cavity Model	Test-Mass Position	Coupling Factor
Sphere	center	$4\pi/3$
Cube	center	$4\pi/3$
Cube	center on floor	2.8
Half-sphere	center on floor	2.1
Cuboid (2:1:1)	center on floor	1.3

For the simulation in this thesis, the GGN acceleration imposed by the seismic field on a test mass is computed accordingly to [eqs. \(4.14\)](#) and [\(4.16\)](#) without any small-cavity approximation.

4.1.3 MODELING THE MIRROR RESPONSE

Based on the GGN acceleration, the following step is to model the mirror response in terms of dislocation from equilibrium to derive the GGN strain subsequently. To mitigate the direct seismic noise that couples via ground vibrations into the interferometer, the mirrors will be suspended via a multistage pendulum based on the model of the Virgo super-attenuator [2, 29]. The advantage of such a multistage pendulum is that the transmissibility, defined as the ratio of mirror response movement to external movement, decreases very quickly above the maximum resonance frequency. To further increase the sensitivity in the low-frequency band (1-10 Hz), the total length of the attenuator will be expanded to either 12 m or 17 m [2, p. 163]. The transmissibilities of the proposed multistage pendulum for ET are shown for different total pendulum lengths in fig. 4.2.

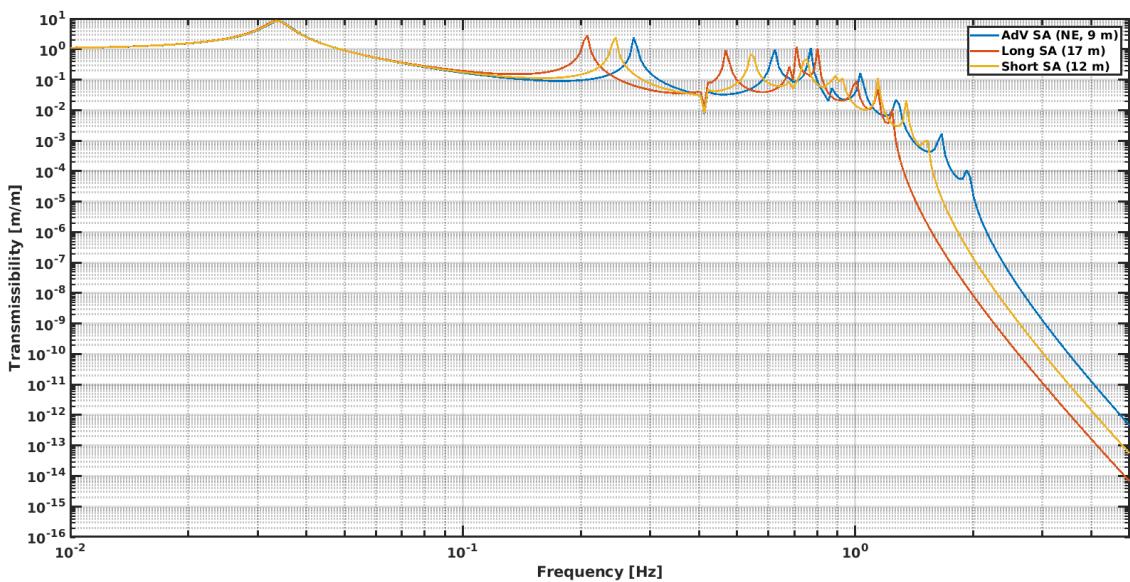


Figure 4.2: Transmissibility curves of the foreseen super-attenuator for ET.

The blue line corresponds to a total pendulum length of 9 m as used in the advanced Virgo experiment. The red and orange lines correspond to the same-spaced pendulum with an increased total length of 17 m and 12 m respectively. Adapted from [2, p. 163].

Multiple resonance frequencies are present as a result of the multistaged pendulum. Above 1 to 2 Hz, the transmissibility decreases quickly by several orders of magnitude. The exact modeling of this behavior can get significantly complex and extensive.

For the purpose of this study, however, it is sufficient to model the mirror response in a simplified manner. For this sake, it is assumed that the suspension acts as a pendulum with a single resonant frequency

$$\omega_0 = \sqrt{\frac{g}{l}} \approx 0.904 \text{ Hz} \quad (4.18)$$

where $g = 9.806\,65 \text{ m/s}^2$ [70] is the standard value of gravitational acceleration and a total pendulum length of $l = 12 \text{ m}$ is assumed.

In this case, the mirror can be described as a damped forced oscillator and the equation of motion reads

$$\ddot{\delta\vec{r}_m}(t) + 2\gamma\dot{\delta\vec{r}_m}(t) + \omega_0^2\delta\vec{r}_m(t) = \delta\vec{a}(\vec{r}_m, t) \quad (4.19)$$

where γ describes the damping constant of the system and $\delta\vec{r}_m(t)$ is the dislocation of the mirror from the equilibrium position \vec{r}_m . This ordinary linear differential equation is nested since the absolute mirror position is time-dependent due to the GGN perturbation. However, as the expected mirror dislocation is very small in comparison to the absolute mirror position and the seismic wavelengths, the time dependency of the mirror position in the evaluation of the seismic fields can be neglected

$$\vec{r}_m(t) = \vec{r}_m + \delta\vec{r}_m(t) \approx \vec{r}_m \quad (4.20)$$

With this approach, it is straightforward to solve eq. (4.19) by applying the separation Ansatz

$$\delta\vec{r}_m(t) = \delta\vec{r}_m^0 e^{i\omega t} e^{i\alpha(\omega)} \quad (4.21)$$

which is based on the consideration that the mirror should follow the external frequency ω after a certain settling time. A possible phase shift between external force and mirror movement is expressed as $\alpha(\omega)$. Inserting this Ansatz in eq. (4.19) results in

$$\delta\vec{r}_m(t) = \mathcal{H}(\omega) e^{i\alpha(\omega)} \delta\vec{a}(\vec{r}_m, t) \quad (4.22)$$

where the mechanical transfer function is derived as

$$\mathcal{H}(\omega) = \frac{1}{\sqrt{(\omega_0^2 - \omega^2)^2 + (2\gamma\omega)^2}} \quad (4.23)$$

The damping factor γ is expected to be very small compared to the resonant frequency and will have a negligible impact. The phase shift can be calculated as

$$\alpha(\omega) = \arctan\left(\frac{-2\gamma\omega}{\omega_0^2 - \omega^2}\right) - \pi \quad (4.24)$$

The transmissibility of the described pendulum is shown in fig. 4.3.

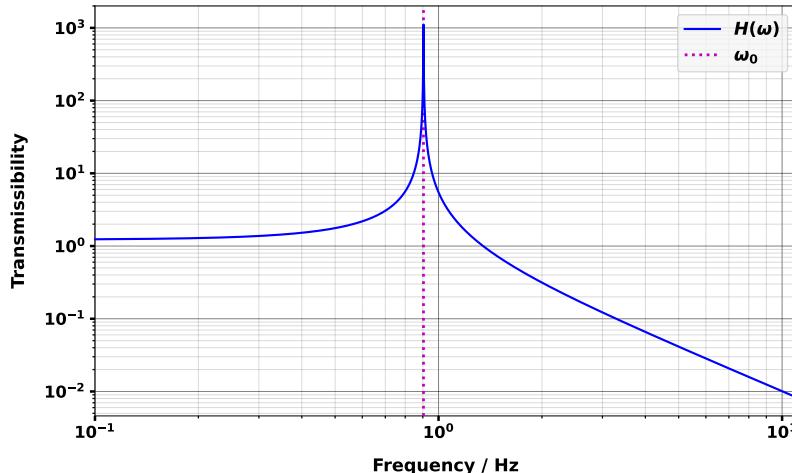


Figure 4.3: Transmissibility of the simplified model of the super-attenuator.
The dashed magenta line represents the resonance frequency.

It must be noted that the aforementioned simplification steps should only have a minor impact on the network performance. The modeling of a more realistic attenuator will be possible with complete knowledge of all involved resonance frequencies and damping factors.

4.1.4 STRAIN NOISE TRANSFER

Finally, the perturbations of each mirror are combined under consideration of the foreseen geometry of the three ET-LF to obtain the GGN strain defined as

$$h_G^j(t) = \frac{\delta L_G^j(t)}{L} , \quad j = 1, 2, 3 \quad (4.25)$$

where $\delta L_G^j(t) = \delta L_G^{j,2}(t) - \delta L_G^{j,1}(t)$ is the differential length change between the two interferometer arms of the j_{th} interferometer due to the GGN perturbation. The length change of each interferometer arm is connected to the mirror displacements via

$$\delta L_G^{j,i}(t) = [\delta \vec{r}_{bs}^j(t) - \delta \vec{r}_{em}^{j,i}(t)] \cdot \hat{e}_{bs \rightarrow em}^{j,i} , \quad i = 1, 2 \quad (4.26)$$

where bs represents the beam-splitter mirror of each interferometer and em represents the end-mirror of the i_{th} arm. The unit vectors $\hat{e}_{bs \rightarrow em}^{j,i}$ project the mirror displacements in the direction of the arms. The complete mirror arrangement is shown in fig. 4.4.

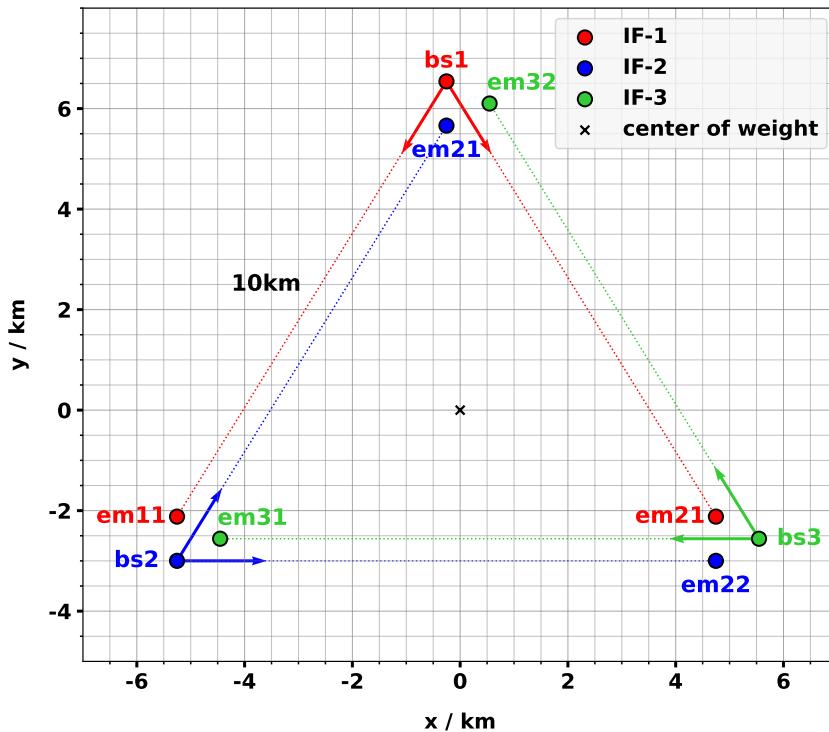


Figure 4.4: The simulated detector geometry of ET-LF.

A coordinate system is chosen where the detector plane corresponds to the xy -plane, and the center of weight of all mirror positions is in the xy -origin. The unit vectors $\hat{e}_{bs \rightarrow em}^{j,i}$ are represented as arrows, and the parallel and orthogonal displacements of the interferometers are upscaled for visibility.

4.2 THE SEISMIC SENSOR ARRAY

The second central component of the model covers the simulation of the seismic sensors. For that purpose, the question of the potential sensor positions around ET has to be addressed. However, as ET is still in the planning phase and especially the field of GGN mitigation only gained momentum over the last years, the answer to this question is highly speculative. Further, a method of considering the sensor's intrinsic noise sources is needed for an adequate model. These two points will be discussed in the following subsections.

4.2.1 POSITIONING OF THE SEISMIC SENSORS

It is expected that surface seismometers alone are not sufficient and underground deployment is inevitable [4]. From a financial point of view, an underground seismometer arrangement oriented on the ET host cavities would be beneficial as the borehole expenses' are the main driving cost factor. This aspect would imply seismic sensors that are positioned on an irregular grid. Which, in a DL application, would impose a challenge as most DL methods, especially CNNs, have mainly been developed for regular input data, i.e., image-like data. A way of overcoming this hurdle is the growing field of Graph Neural-Networks (GNNs) that made significant progress in recent years [64].

Regardless of the circumstance that there is *a priori* no specific stand-out approach for an optimized irregular sensor positioning, it is sensible to choose a regular array for several reasons. Mainly a regular grid offers the advantage of a vast repertoire of, on the one hand, computationally optimized and, on the other well-studied methods that can be incorporated to build a suited network architecture. Moreover, the architecture can be designed modularly, making it uncomplicated to adapt to an irregular input space in case a specific approach for an irregular sensor positioning becomes concrete.

Based on these considerations, a regular 2-dimensional grid of 8×8 seismic sensors in the plane of the interferometers is chosen with a spacing of 3 km in both directions. (see fig. 4.5) Additionally, a 3-dimensional case with spherically arranged sensors around the ET vertices is implemented that can be used for follow-up studies.

4.2.2 SENSOR NOISE IMPLEMENTATION

A seismic sensor is exposed to internal instrumentation noise sources, including primarily electronic and *Brownian* noise [71]. The noise level of a particular sensor compared to the measured signal is expressed as the Signal-to-Noise Ratio (SNR) and varies with the quality of the sensor. In agreement with the recent EMR seismic characterization study, tri-axial Kinematics Streckeisen STS-5A [75] state-of-the-art borehole seismometers are taken into account to model the sensor self-noise.

The modeling is based on measurements within the framework of an evaluation study of the abovementioned sensors [76]. Starting from the measured self-noise Amplitude Spectral-Density (ASD), colored noise is generated by multiplying gaussian white noise with the ASD in the frequency domain.

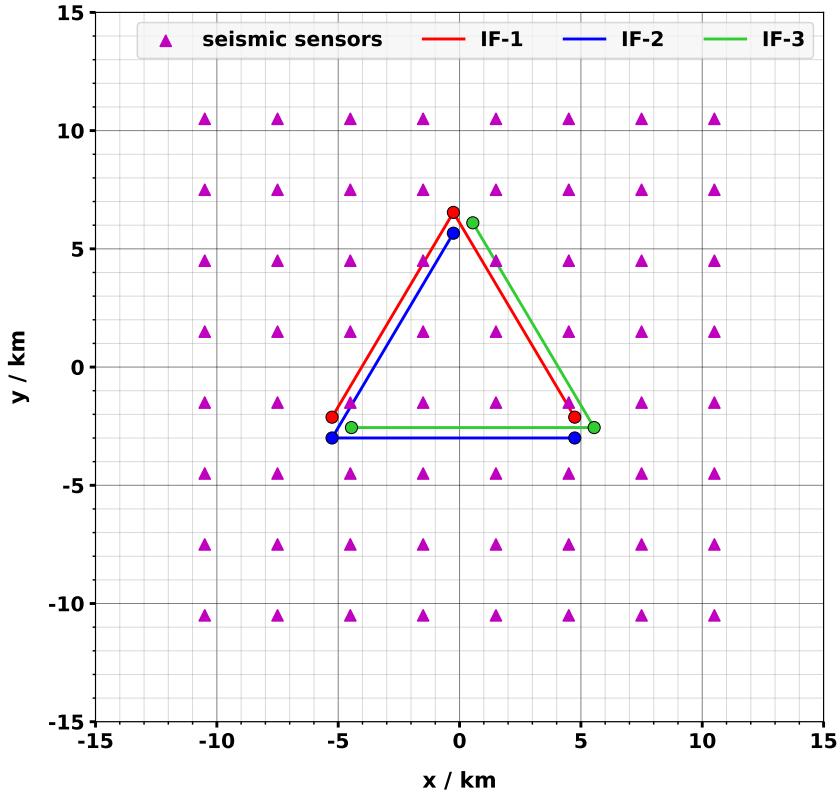


Figure 4.5: The arrangement of the simulated seismic sensor array.

The 8×8 seismic sensors (magenta triangles) are arranged on a regular 2-dimensional grid in the interferometer plane with a spacing of 3 km in each direction.

In detail, the white noise in the frequency domain is generated via

$$n_w(f) = a(f) + i b(f) \quad (4.27)$$

where $a(f)$ and $b(f)$ are sampled from a *Gaussian* distribution

$$p(f) = \frac{1}{\sqrt{2\pi}\sigma_f} \exp \frac{f^2}{2\sigma_f^2} \quad \text{with} \quad \sigma_f = \frac{1}{2\sqrt{\delta_f}} \quad \text{and} \quad \delta_f = \frac{f_s}{N_t} \quad (4.28)$$

where f_s and N_t are the sampling frequency and the total number of time steps, respectively. This scaling serves the purpose of gaussian white noise with a mean amplitude of one, which can be colored in the frequency domain via

$$n_c(f) = n_w(f) \times ASD(f) \quad (4.29)$$

The resulting colored noise, together with the sensor self-noise ASD, are shown in fig. 4.6. It is visible that the noise spectrum is dominated by low-frequency components (< 0.1 Hz) outside the relevant GGN band (1-10 Hz). A band-pass filter is applied to exclude these contributions, which narrows the noise frequency band down to (0.5-12 Hz). With this, a tolerance interval that is orientated on the resulting simulated frequency bandwidth is considered.²

²the simulation bandwidth is discussed in sec. 4.3.3

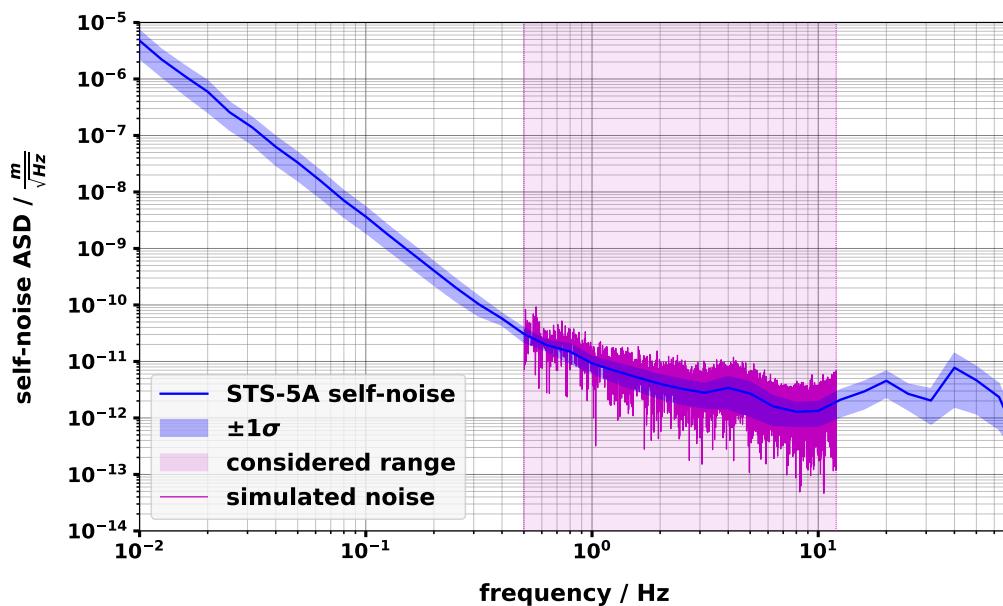


Figure 4.6: The ASD colored simulated seismic sensor noise.

The blue line is the measured self-noise ASD within the 1σ band [76]. In magenta, the generated noise in the relevant frequency region is shown.

Finally, the time domain noise is obtained by an inverse *Fourier* transformation

$$n_c(t) = \mathcal{F}^{-1}\{n_c(f)\} \quad (4.30)$$

and an instance of simulated seismic sensor signal with the corresponding self-noise of SNR = 4 is shown in fig. 4.7.

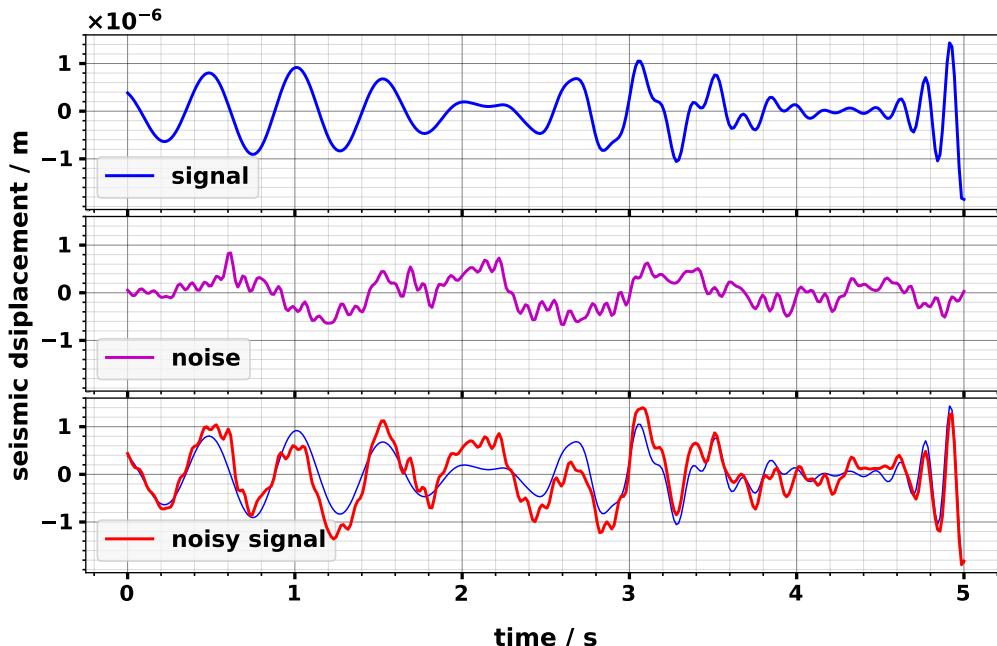


Figure 4.7: Example of a simulated seismic sensor signal with noise.

The blue line shows a simulated sensor signal without noise, and the red line corresponds to the same signal with added noise (magenta) of SNR = 4.

A remaining open question concerning the sensor noise implementation is the simulated SNR. An approach to answering this question is to estimate the SNR based on available measurements, i.e., the measured seismic activity [3] and the measured noise level [76]. The SNR is given as the ratio of the average signal power to average noise power and can be expressed in terms of Root Mean Square (RMS) as

$$\text{SNR} \equiv \frac{\bar{P}_{\text{signal}}}{\bar{P}_{\text{noise}}} = \left(\frac{\text{RMS}_{\text{signal}}}{\text{RMS}_{\text{noise}}} \right)^2 \quad (4.31)$$

where the RMS can be estimated based on the ASD in a given frequency band Δf as [71]

$$\text{RMS}_{(\Delta f)} \approx \sqrt{2 \cdot \text{ASD}_{(\Delta f)}^2 \cdot \Delta f} \quad (4.32)$$

Applying eqs. (4.31) and (4.32) to data from [3, 76] results in an estimated achievable SNR of

$$\text{SNR}_{(1-10 \text{ Hz})} \approx \mathcal{O}(10^3) \quad (4.33)$$

which implies that state-of-the-art seismometers should be able to measure the seismic activity at the candidate site in the EMR with very high accuracy in the relevant frequency band.

However, for the sake of testing the tolerance of the network in case of worse seismometer accuracy³, an additional scenario with $\text{SNR} = \mathcal{O}(10)$ is implemented in the simulation model. This is still slightly more pessimistic compared to a simulation study that investigated GGN cancellation using WFs and assumed a constant SNR of 15 [34].

³Either due to the sensor quality itself or less loud seismic signals.

4.3 SIMULATION ALGORITHM AND PARAMETER SUMMARY

The previously discussed concepts of the GGN model are combined in an elemental *Monte-Carlo* manner, i.e., by sampling the relevant parameters to construct a seismic wave field and the corresponding GGN strain and sensor signals. In order to give a brief understanding of how exactly that is achieved, a strongly simplified overview of the underlying algorithm is sketched below.

Algorithm 1: Simplified sketch of the simulation algorithm.

```

Initialize: ET-LF mirror postions  $\vec{r}_m^{j,i}$ , sensor positions  $\vec{r}_s^k$ 
for event in  $N_{events}$  do
    seismic displacement  $\vec{\xi}(\vec{r}, t) = 0$ 
    mirror acceleration  $\delta\vec{a}(\vec{r}_m^{j,i}, t) = 0$ 
    mirror displacement  $\delta\vec{r}_m^{j,i}(t) = 0$ 
    for wave in  $N_{waves}$  do
        sample wave parameters ;
        → seismic displacement  $\vec{\xi}(\vec{r}, t) += eq. (4.10)$ 
        → mirror acceleration  $\delta\vec{a}(\vec{r}_m^{j,i}, t) += eqs. (4.14) and (4.16)$ 
        → mirror displacement  $\delta\vec{r}_m^{j,i}(t) += eq. (4.22)$ 
    end
    for j in (1,2,3) do
        get mirror displacement  $\delta\vec{r}_m^{j,i}(\vec{r}, t)$  ;
        strain noise  $h_G^j(t) = eq. (4.25) \leftarrow \delta\vec{r}_m^{j,i}(\vec{r}, t)$ 
    end
    for k in  $N_{sensors}$  do
        get seismic displacement  $\vec{\xi}(\vec{r}_s^k, t)$  ;
        seismic sensor signals  $S_k(t) = \vec{\xi}(\vec{r}_s^k, t)$ 
    end
end
Return: strain noise  $h_G^j(t)$  ; seismic sensor signals  $S_k(t)$ 

```

The fundamental idea is to construct a seismic field as the sum of N_{Waves} individually sampled seismic wave packets and subsequently derive the GGN and seismic sensor signal as previously defined. This procedure is repeated N_{Events} times to create the training and test datasets for the NN. All relevant parameters, either fixed ones or sampled ones, will be summarized in the following.

4.3.1 FIXED PARAMETERS

The fixed parameters are grouped into either parameters addressing the simulated environment or hyperparameters of the simulation. The latter are chosen mainly under computational aspects.

For instance, the total number of time steps N_t of the GGN strain and sensor signal time-series directly scales the input-space dimension and thus the whole parameter space of the NN.

Therefore, for a desired total time series length of, e.g., 5 s, the sampling frequency f_s has to be chosen small enough to keep the dimension in check and large enough to provide sufficient resolution and avoid aliasing effects. Any f_s that is larger than $f_{\text{Nyquist}} = 2f_{\text{max}}$ should give a non-aliased resolution according to the *Shannon-Nyquist theorem*. [77]. The number of waves per event N_{Waves} corresponds to one wave packet injected for every second of the time-series, which is adapted from [35]. An overview of the simulation hyperparameters is presented in [table 4.2](#).

Table 4.2: Summary of simulation hyperparameters.

Parameter	Value
Sampling frequency	$f_s = 70 \text{ Hz} \approx 3f_{\text{Nyquist}}$
Time steps	$N_t = 350 (\cong 5 \text{ s})$
Waves per event	$N_{\text{Waves}} = 5$

The environmental parameters are, as far as feasible, adjusted to the scenario of the current design concept for ET-LF [2] located in the EMR. An overview is presented in [table 4.3](#)

Table 4.3: Summary of parameters of the simulated environment.

The values for ET-LF are based on [2] if not stated otherwise and geology values are based on [3]. The cavity radius is a rough estimate based on the assumption that two super-attenuators (ET-LF and ET-HF) will be placed in the same cavity.

	Parameter	Value
ET-LF	IF arm length	$L = 10 \times 10^3 \text{ m}$
	depth under surface	$z = -250 \text{ m}$
	parallel displacement of IFs	$d_p = 4 \text{ m}$
	orthogonal displacement of IFs	$d_0 = 400 \text{ m}$
	mirror resonance frequency	$\omega_0 = 0.9 \text{ Hz}$ [sec. 4.1.3]
	spherical cavity radii	$a = 20 \text{ m}$
Seismic Sensors	total number	8×8
	axis	tri-axial
	spacing	$d_x = d_y = 3 \times 10^3 \text{ m}$
	depth under surface	$z = -250 \text{ m}$
Geology	mass density	$\rho_0 = 2.8 \times 10^3 \text{ kg/m}^3$
	p-wave speed	$c_p = 5 \times 10^3 \text{ m s}^{-1}$
	s-wave speed	$c_s = 3 \times 10^3 \text{ m s}^{-1}$

4.3.2 SAMPLING PARAMETERS

Parameters relevant to constructing the underlying seismic wave field are all appearing in eqs. (4.10) and (4.11) and are sampled stochastically. The sampling ranges and distributions aim to mimic the seismic properties as expected in the EMR. An overview is provided in table 4.4.

Table 4.4: Summary of sampling parameter ranges and distributions.

Parameter	Range	Distribution
Frequency f	1 – 10 Hz	uniform
Amplitude ξ^0	$2 \times 10^{-6} – 3 \times 10^{-6} \text{ m}^2$	uniform or psd-scaled
Arrival time t_0	0 – 5 s	uniform
P-wave speed c_p	$(5 \pm 0.5) \times 10^3 \text{ m s}^{-1}$	gaussian
S-wave speed c_s	$(3 \pm 0.3) \times 10^3 \text{ m s}^{-1}$	gaussian
Phase φ_0	0 – 2π	uniform
Origin \vec{r}_0	unit sphere ($i \leq 45^\circ$)	uniform
Width ΔT	$0.04/f – 2.00/f$	uniform

Frequencies and Phases The wave frequencies f are drawn in a range corresponding to the frequency band where GGN is expected to be a limiting factor for ET. Two underlying sampling distributions are implemented, each representing a different test scenario. On the one hand, a uniform frequency distribution for a general test, and on the other, a distribution that is scaled by the characteristic seismic Power Spectral-Density (PSD) as measured in [3] to create a site-specific test scenario in the EMR. Wave phases φ_0 are uniformly distributed between 0 and 2π .

Arrival Times Waves arriving from sources of varying distances are taken into account by sampling the arrival time of each wave packet t_0 uniformly within the total time-series duration. Therefore the number of waves present in each second of time-series varies from event to event.

Amplitudes and Source Durations Maximum wave amplitudes ξ^0 and widths of the gaussian wave packets ΔT are determined indirectly in such a way that the GGN output matches with the predictions made in the EMR site characterization study [3]. By allowing the wave packets to exceed the boundaries of the simulated time-series, the data replicates snippets from real data streams. To further keep the number of wavelengths per wave packet roughly equal over the frequency range, the wave packet width is scaled by $1/f$. The consequent source durations of the gaussian wave packets can be estimated based on their widths. The resulting values between 0.024 – 6 s illustrate that the simulation covers cases ranging from short burst-like seismic sources over more extended transient sources up to quasi-ambient signals exceeding the range of the observed time-series.

Polarizations and Propagation Velocities For each sampled wave packet, the polarization $\hat{e}_{p,s}$ is chosen as longitudinal $\hat{e}_p \cdot \hat{e}_k = 1$ with a probability of 70% or transversal $\hat{e}_s \cdot \hat{e}_k = 0$ otherwise. This is motivated by the assumption that the main contribution to the seismic field should originate from p-waves and s-waves are mainly created through scattering processes of incident p-waves [71]. The corresponding propagation velocities $c_{p,s}$ are drawn from a gaussian distribution where the mean values are equal to the values presented in [table 4.3](#). Though this approach is likely to be an overestimate, it can be seen as a «brute-force method» to include wave scattering effects [35].

Source Origins and Propagation Directions The source origins \vec{r}_0 are uniformly distributed on a unit sphere around the center of weight⁴ of the ET-LF geometry and the propagation direction \hat{e}_k points from \vec{r}_0 inwards to the sphere origin in order to ensure that the sampled waves reach the simulated interferometers. A graphical representation of the described circumstance is shown in [fig. 4.8](#).

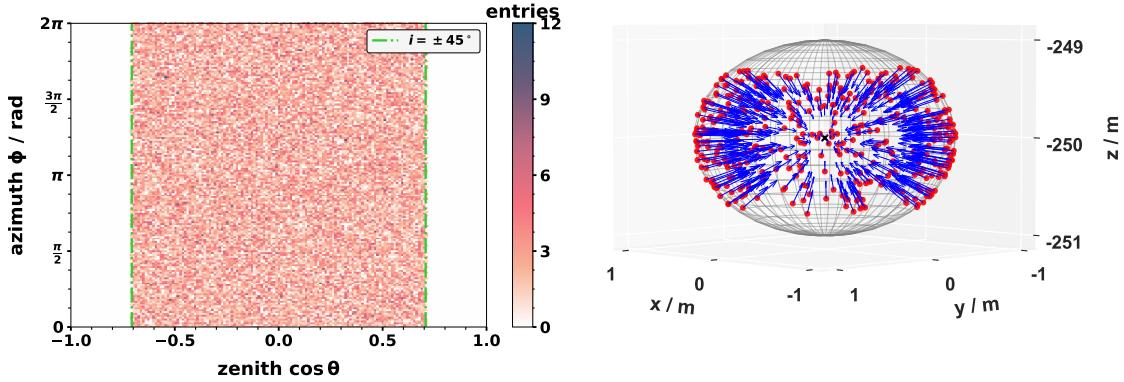


Figure 4.8: Sampling of the seismic wave origins and propagation directions.
left: The uniform angular distribution of the seismic source incident angles. The cutoff values are $i = \pi/2 - \theta \leq \pm\pi/4$.
right: The origins of seismic sources on the unit sphere in red; their propagation directions in blue; and the center of the simulated ET-LF geometry as black cross.

The angle i represents the inclination angle of seismic waves relative to the plane containing the ET-LF interferometers and the seismic sensors. The simulation is restricted to values between $\pm 45^\circ$ to improve the training convergence by reducing the dataset's sparsity [44]. This is based on considering that the wave field's main contributions are longitudinal, and therefore the strain projection in the detector plane decreases as the inclination angle increases. Hence, excluding large inclination angles reduces the number of sparse events containing little spatiotemporal information. This should not impose a loss of generality, and for verification, a test of network performance as a function of inclination angle will be conducted.⁵

⁴as shown in [fig. 4.4](#).

⁵see [sec. 5.4](#)

4.3.3 SIMULATION RESULTS

To conclude, an arbitrarily chosen simulated event, consisting of the seismic sensor signal⁶ and the related GGN strains, is presented in fig. 4.9. It is visible that the GGN strain signals in the three interferometers are correlated, and their sum, the so-called *null-stream* [78] is not vanishing.

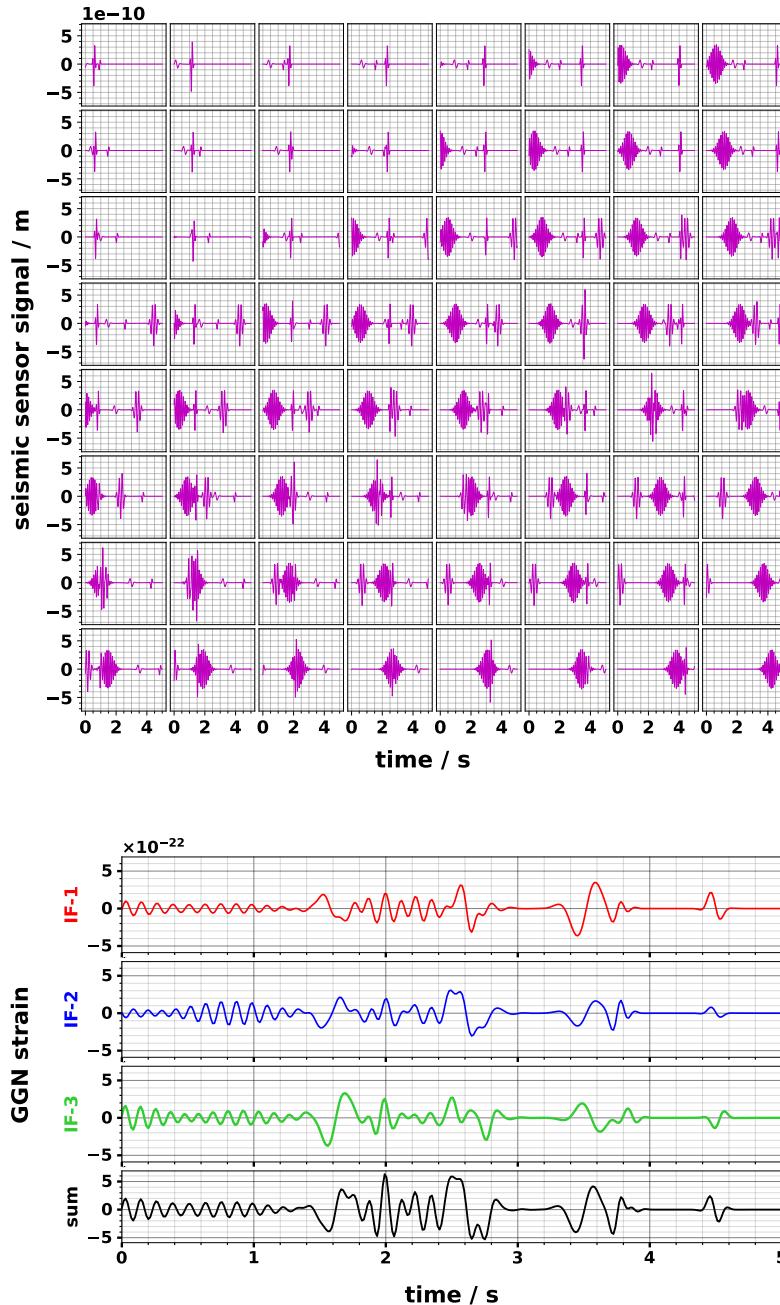


Figure 4.9: Exemplary display of a simulated event.

top: The measured signals of the 8×8 seismic sensors. Shown are the x-axis measurements. The remaining axes are displayed in fig. A1 in the appendix.

bottom: The corresponding strain signals in the three interferometers and the null-stream.

⁶Realtime data from one of the Terziet borehole sensors is available at:

<http://www.orfeus-eu.org/data/odc/realtime/?network=NL&station=TERZ>

In addition to the representation in the time domain, further insights are gained by a frequency-domain representation of the simulated GGN. For this purpose, the Amplitude Spectral-Density (ASD) of the strain is a helpful quantity. It describes the amplitude power distribution of a time-series signal $s(t)$ in relation to the frequency content and is defined as [79]

$$\text{ASD}_s(f) \equiv \sqrt{\tilde{s}^*(f) \cdot \tilde{s}(f)} \quad (4.34)$$

where $\tilde{s}(f) = \mathcal{F}\{s(t)\}$ is the *Fourier* transformation of the signal and $*$ denotes the complex conjugation. The ASD output of the simulation is computed following *Welch's* method [80]. This method divides a long time-series signal into several shorter snippets, performs the ASD calculation, and subsequently averages over all snippets to estimate the total ASD. To avoid spectral leakage arising from the time-series boundaries, a *Hanning* windowing function is applied to each snippet [79]. Applying to the simulation data means that one snippet corresponds to one out of 5000 test events. As each event has a duration of 5 s, the resulting ASD is computed over approximately 7 h of GGN data.

The resulting ASD for the general test scenario with uniformly distributed frequencies in the relevant GGN band of 1 to 10 Hz, is presented in [fig. 4.10](#).

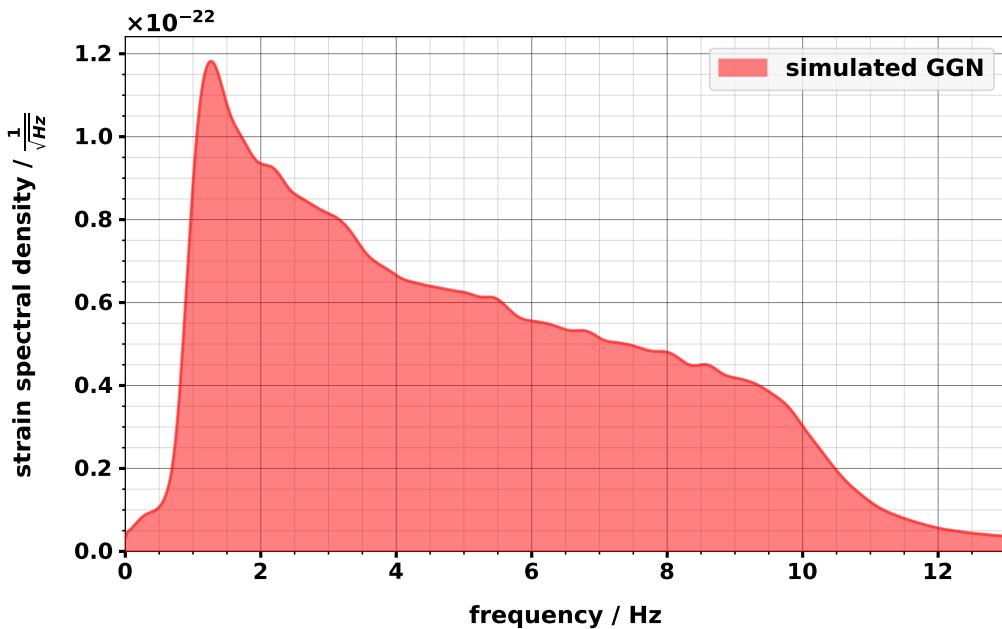


Figure 4.10: Strain amplitude spectral density of the simulated gravity-gradient noise in the general test scenario.

It can be observed that the absolute bandwidth slightly extends the 1 to 10 Hz range because the wave packets add small tails towards higher and lower frequencies. The resulting bandwidth is considered for the seismic sensor noise implementation as mentioned in [sec. 4.2.2](#).

Further, an increasing density towards lower frequencies can be observed as a result of the frequency-dependent wave packet width. As discussed in [sec. 4.3.2](#), lower frequencies have larger wave packet widths, making the relative contribution to the ASD more significant.

The ASD of the scenario addressing the EMR site-specific simulation setup is displayed below in fig. 4.11.

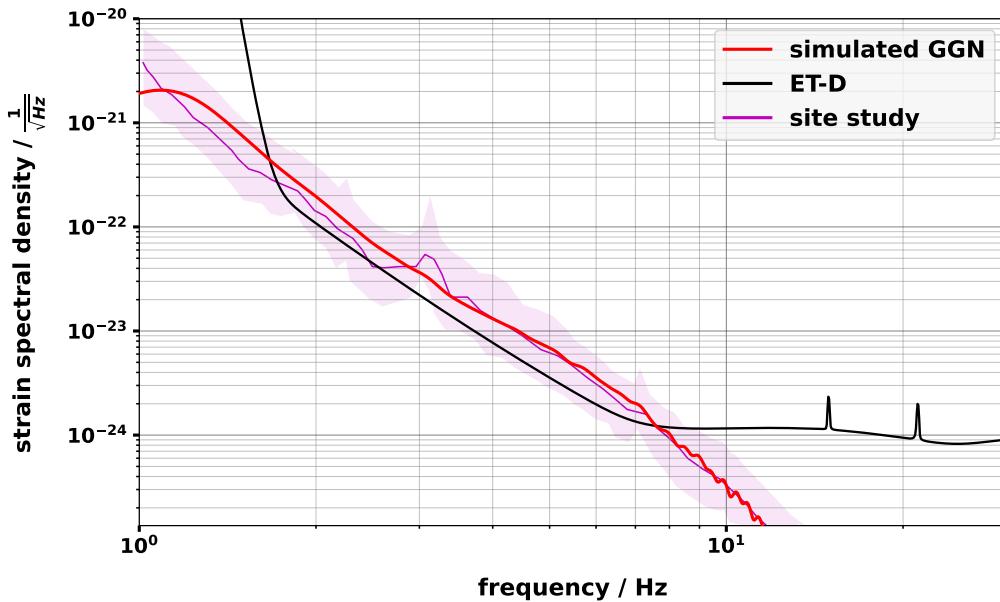


Figure 4.11: Strain amplitude spectral density of the simulated gravity-gradient noise in the Euregio specific test scenario.

The output of the simulation (*red*) compared to the GGN estimate from the EMR site characterization study [3] (*magenta*) and the ET-D design sensitivity (*black*). The shaded band encloses the 10th and 90th percentile.

The simulated ASD from this thesis is compared to the GGN prediction from the EMR site characterization study that includes an extensive seismic measurement campaign as the basis for detailed modeling of seismic properties [3].

The curves are in very good agreement within the 10th and 90th percentile of the site study over the whole frequency band.

DEEP LEARNING FOR GRAVITY-GRADIENT NOISE MITIGATION

Based on the previously presented GGN model for data generation, a cancellation scheme using NNs will be developed and tested in this chapter.

The scheme follows the proposal as made in the ET technical design report update: First, a GGN prediction is made based on the seismic sensors, and secondly, the prediction is subtracted from the noisy data [2, p. 172].¹

For this purpose, the main challenge is to develop a suited network architecture that can sufficiently approximate the transfer function that maps the high-dimensional seismic sensor input onto the GGN strain prediction. In other words, the network has to solve a many-to-one sequence-to-sequence regression task with spatiotemporal input data. Comparable dimensions would occur in a video-to-caption task.

The test scenarios will restrict to the GGN strain of the interferometer IF1 without any loss of generality as the other interferometers are spatially rotated variations in the same plane as the test IF and can be treated equally.

In this chapter, the proposed network architecture is introduced and tested on two scenarios; a general and a EMR site-specific test case. Further, the network resolution is evaluated as a function of the source inclination angles. Finally, an analysis of the sensor's relative importance as a function of sensor position is presented as a first step toward an optimized sensor positioning.

5.1 THE NETWORK ARCHITECTURE

Constructing a network architecture that is suited for a given task is the main challenge of any DL application. The building blocks that make up the network must be matched precisely, starting from the initialization and ending with the final output layer. Moreover, a large set of parameters must be adjusted just right to allow the network to learn.

In the following, the proposed network architecture for GGN mitigation at ET will be presented together with all settings and adjustments necessary for the training.

5.1.1 DATA PREPROCESSING

Before building the actual network, the correct preprocessing of input data has to be considered. As the input tensors will propagate through a long chain of tensor operations, properly scaling the input and target data is essential to avoid any numerical issues.

With a target strain on the order of $\mathcal{O}(10^{-22})$, for instance, the network is prone to a numerical error called «underflow» [41].

¹the scheme is shown in [fig. 1.1](#).

Underflow describes the issue of values being rounded to zero if they are too small to get adequately represented by the underlying bit pattern. In order to avoid underflow, the label and target data will get rescaled by a constant value so that the resulting values are of order $\mathcal{O}(1 - 10)$.

To further increase the robustness of the network against different scaled data and to support the backpropagation by normalizing the gradient distributions, the input data will additionally be «standardized» [44]. This means that the input data will be scaled to zero mean and unit variance by applying

$$x'_i = \frac{x_i - \bar{x}}{\sigma_x} \quad (5.1)$$

where x'_i is the standardized value of a data sample x_i of the dataset \mathbf{x} with mean value \bar{x} and standard deviation σ_x .

A final preprocessing step is to exclude the z-axis of the seismic sensor signals and only consider the projection that lies in the ET-LF detector plane, i.e., the xy-plane. This measure is based on experience gained during the commissioning of the network: with the entire input space, the network is unable to learn sufficiently, and removing the z-axis has a crucial impact. A potential explanation for this behavior could be an overload of less-relevant information that aggravates the filtering of relevant features for the network. Beneficial aspects that accompany this measure are, on the one hand, financially, as the cost of seismic sensors could potentially be reduced, and on the other, computationally, as the input space and thus the whole parameter space of the networks gets decreased.

A potential preprocessing step that suggests itself with the application of CNNs is using image-like spectrograms as input data by transforming into the frequency domain. However, as this approach would significantly increase the input dimensions and further, due to the additional computational step, make the network impractical for a fast GGN mitigation², this approach is voided.

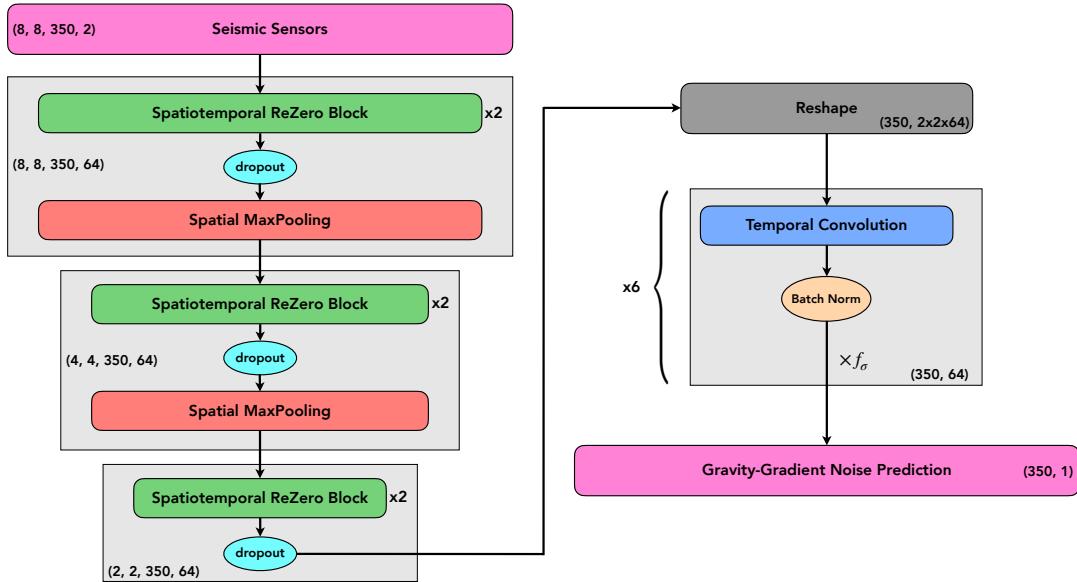
5.1.2 NETWORK DESIGN

The proposed network architecture for GGN mitigation at ET is constructed from the building blocks that have been introduced in ch. 3.

The essential concept is to extract the spatiotemporal features of a seismic event with two-dimensional spatial and one-dimensional temporal convolutions arranged in *ReZero* blocks. Subsequently, dimension reduction in the form of spatial *MaxPooling* is conducted twice to enforce more abstract latent representations of the captured features. A potential third pooling layer is omitted and substituted by a *Reshape* layer that flattens the four-dimensional spatiotemporal feature space into a two-dimensional multivariate time-series. The advantage of the *Reshape* layer at this point is that the complete set of extracted latent features is retained [37] for the follow-up temporal convolution layers that translate the multivariate feature time-series into the GGN prediction.

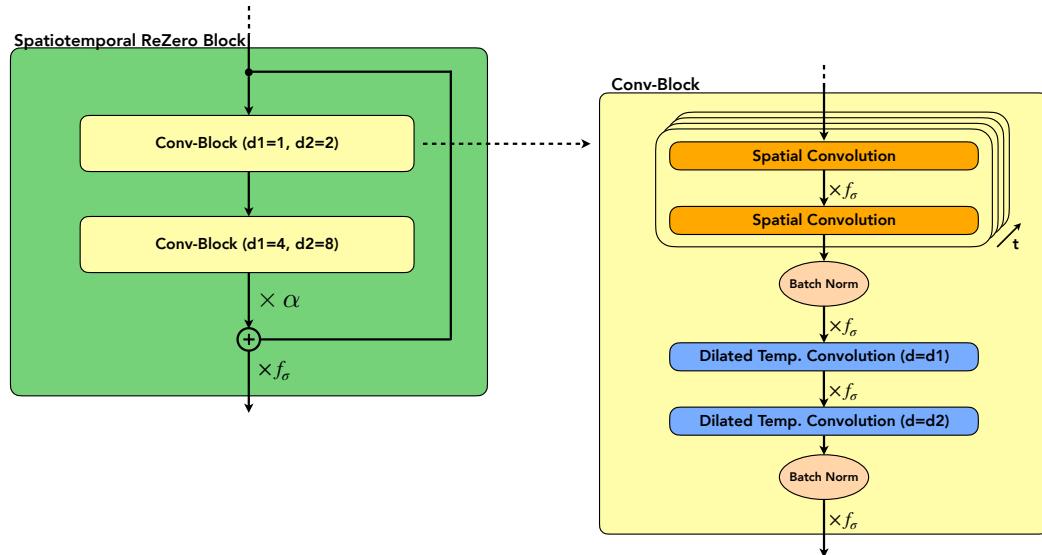
Additional *spatial dropout*, *batch-normalization* and *activation* layers are inserted between the discussed layers and the resulting network architecture is presented in fig. 5.1.

²which would be particularly interesting in a MMA context.



(a) Low-level overview representation of the network architecture.

The workflow is as described in the text. Two spatiotemporal *ReZero* blocks are followed by a *Dropout* layer in front of a *MaxPooling* layer to increase the focus on sub-dominant features. After two dimension reductions, the data is reshaped and translated into the GGN predictions via six temporal convolutions in combination with *Batch-Normalization*. The data dimensions at each building block are displayed in brackets where the input is equal to (sensors^x, sensors^y, timesteps, sensoraxis).



(b) High-level representation of the spatiotemporal ReZero blocks.

The *ReZero* blocks with trainable parameter α consist of two spatiotemporal convolution blocks with exponentially increasing dilation factors d_i . The spatial convolutions are regular two-dimensional convolutions that, due to the modular approach, can be replaced straightforwardly by graph convolutions in the case of irregularly arranged seismic sensors. As indicated by the labeled arrow, the spatial convolutions are time-distributed, i.e., the same convolution is applied to every time step of the data to extract more general spatial features.

Figure 5.1: The network architecture as a high- and low-level representation.

The activation functions are $f_\sigma = \text{leakyReLU}$ and all spatial or temporal convolutions have a filter size of $N_F = 64$.

A detailed summary of the network hyperparameters is given below in [table 5.1](#).

Table 5.1: Summary of network architecture hyperparameters.

Parameter	Value
Activation function f_σ	leakyReLU, $\beta = 0.5$
Kernel initialization	<i>He-normalization</i> [10]
Bias initialization	zeros
Dropout rate	0.4
Convolutional filters N_F	64
Convolutional kernel sizes k_F^{spat} & k_F^{temp}	3×3 & 3
Dilation rates d_i	2^i , $i \leq 4$
Spatial pooling kernel	2×2

5.1.3 TRAINING SPECIFICATIONS

The network is implemented and trained using Keras [81] with Tensorflow [82] as GPU backend for tensor operations. The wall time per training epoch is approximately 14 min if trained on a single Nvidia A40 GPU unit with CUDA architecture. [83, 84] Relevant hyperparameters for the training process are listed below in [table 5.2](#).

Table 5.2: Summary of network training hyperparameters.

Parameter	Value
Epochs	75
Batch size	8
Learning rate	1×10^{-3}
Loss	MSE
Number of events	5×10^4
Training split	0.8
Validation + Test split	0.1 each
Optimizer	<i>Nadam</i> [48]
Dropout rate	0.4

All previously listed hyperparameters from [sec. 5.1.2](#) and [5.1.3](#), addressing the architecture as well as the training, are determined and tuned based on rational considerations and the experience gained in the process. It is expected that a supplementary optimization technique, e.g., utilizing *Bayesian* optimization algorithms like Hyperopt [85], can further improve the final network performance.

5.2 GENERAL TEST SCENARIO

As an initial test, the network is applied to the general dataset with uniformly distributed frequencies. The primary intent is to get an unbiased impression of how well the network can perform in the relevant frequency band and evaluate any frequency dependencies if present.

Moreover, based on the discussion in sec. 4.2.2, two seismic sensor self-noise scenarios will be compared: a «low-noise» scenario, in which the case of a potentially achievable SNR of $\mathcal{O}(10^3)$ is considered, and a «high-noise» scenario, in which a more pessimistic case of $\mathcal{O}(10)$ is addressed. For the sake of completeness, it should be mentioned that it is waived to implement noise in the low-noise scenario because the resulting noise would be of an order of magnitude that has a negligible influence on the data and further would needlessly harm the computational efficiency.

5.2.1 LEARNING CURVES

For each scenario, the network is trained on a dataset with the setup as described in sec. 5.1.3 and subsequently evaluated on the test dataset. The resulting learning curves are presented in fig. 5.2.

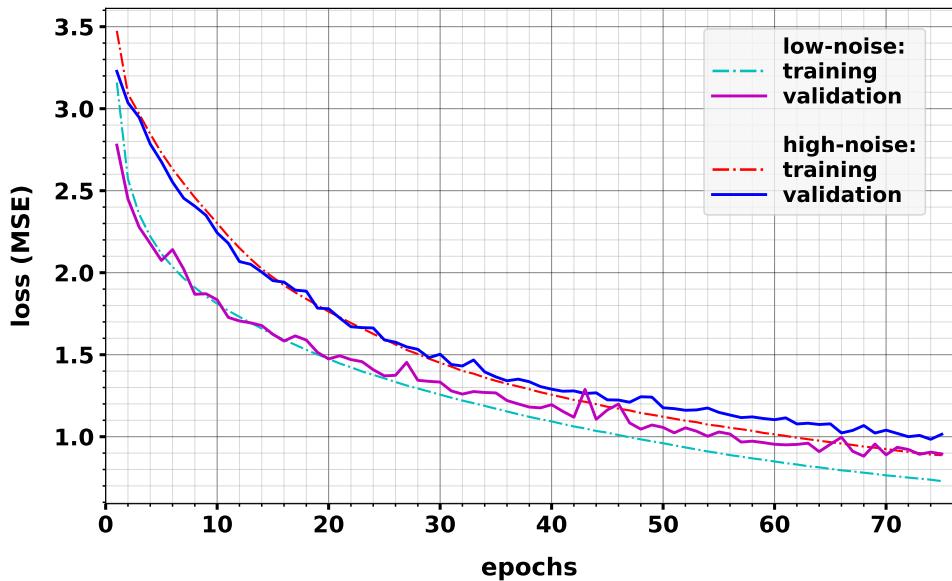


Figure 5.2: Learning curves for the generic test scenario.

The training and validation losses as a function of training epochs are compared for the low-noise and high-noise scenarios.

As might be expected naively, the loss curve in the low-noise scenario is overall smaller. However, both scenarios show significantly similar training slopes, and the distance between the curves is relatively small if taking into account that the noise differs over two orders of magnitude.

A crucial difference between the two scenarios manifests itself in the generalization gap: the high-noise scenario shows a slightly better generalization behavior that results in final validation losses that differ by just 11%. In fact, noise injection is a well-known regularization technique as it can help the network to prevent overfitting to the data [41, 50]. Based on this circumstance, it can be presumed that with NNs, superior noise robustness might be achievable compared to WFs.

The saturation of the validation loss curves towards the end of the training indicates that a total training loop ranging over 75 epochs is a reasonable choice. Nonetheless, a slight downward trend indicates that the network's full potential is not yet fully exploited and can be further optimized. The potentially gained performance, however, is not crucial for the sake of this study.

A detail that seems unusual at first glance is that the validation loss lies beneath the training loss within roughly the first decade of epochs. This can be explained as an artifact of the training process in which the training loss is averaged over the whole epoch, and the validation loss is only computed once at the end of each epoch.

5.2.2 EVALUATION METRICS

Further evaluation metrics will be applied to assess the network's performance over the event scope. Although the MSE is an appropriate measure of the absolute deviation between the ground truth and the predicted signal, it lacks sensitivity to the signal phase. For the latter purpose, additionally, the linear correlation between the two time-series is analyzed via *Pearson's* correlation coefficient [86] defined as

$$r_{xy} \equiv \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \in [-1, 1] \quad (5.2)$$

where x_i represents a sample of the population x with mean \bar{x} . In the use case, the populations x and y represent the timesteps of a ground truth and a predicted time-series respectively. Positive values indicate a correlation between the two waveforms up to an optimal value of $r = 1$ in the case of identical populations. Vice versa, negative values indicate disagreement with a maximal anti-correlation at $r = -1$.

Since it cannot be generally ruled out that the network reconstructs the waveform correctly, except for a constant phase shift, the *Pearson* correlation can additionally be applied in a sliding-window manner to obtain the *cross-correlation* [p. 46-47][87]. The cross-correlation is sensitive to any temporal offset between the two waveforms, and examples of the *Pearson* correlation and cross-correlation for an arbitrarily chosen event are displayed in fig. 5.3.

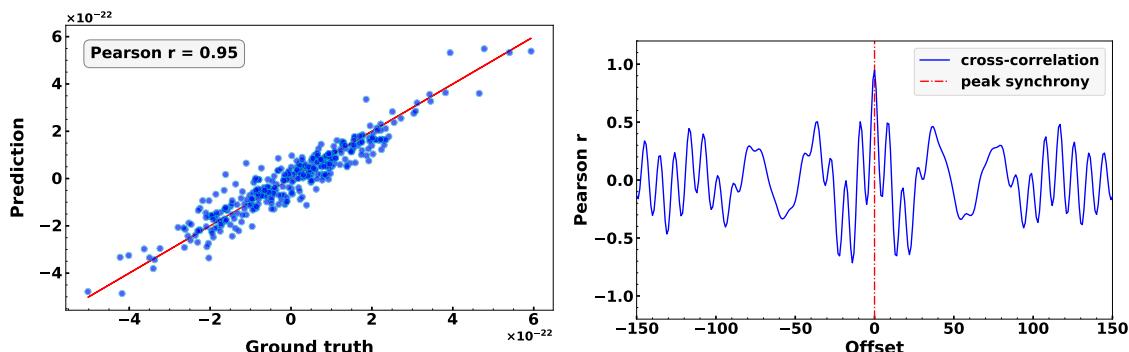


Figure 5.3: Exemplary display of the correlation metric between waveforms.

left: The *Pearson* coefficient indicates a strong linear correlation.

right: The cross-correlation validates optimal temporal synchrony as the peak value is positioned at zero-offset. Secondary peaks appear due to the periodicity of the waveforms.

In order to quantify the magnitude of GGN cancellation, a parameter ϵ_{RMS} is defined as the relative cancellation efficiency in terms of the RMS values of the simulated and the residual GGN waveform after the subtraction of the prediction:

$$\epsilon_{\text{RMS}} \equiv \frac{\text{RMS}_{\text{sim}} - \text{RMS}_{\text{res}}}{\text{RMS}_{\text{sim}}} \quad (5.3)$$

5.2.3 PERFORMANCE ANALYSIS

The network performance is analyzed over the whole test dataset. A first impression is shown in [fig. 5.4](#) with the plot of a reconstructed waveform from the high-noise scenario compared to the corresponding simulated ground truth.

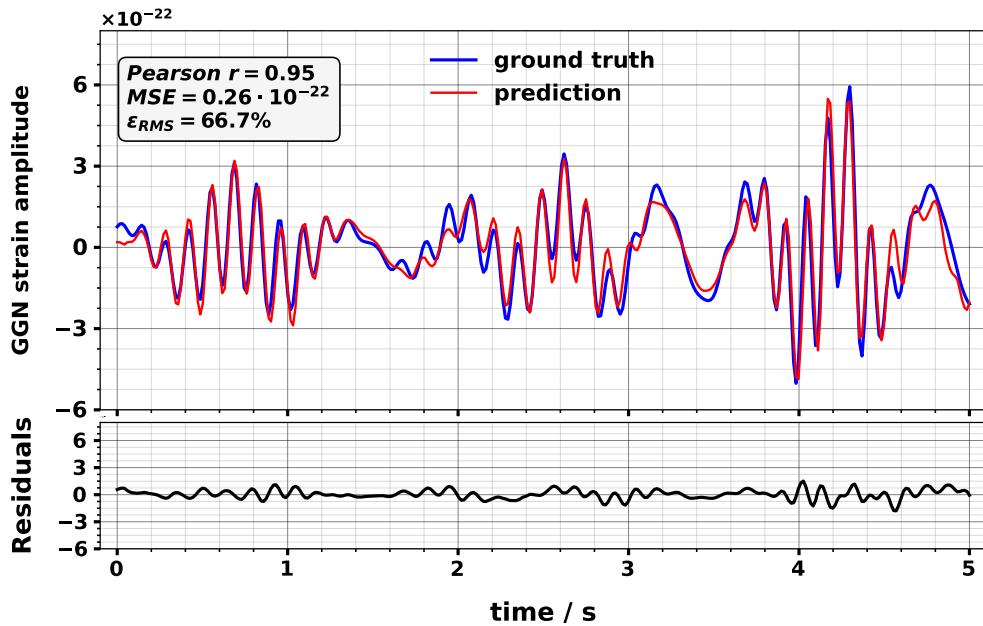


Figure 5.4: Exemplary display of a predicted waveform and GGN cancellation.
top: The network prediction (red) compared to the simulation ground truth (blue).
bottom: The residual (black) as the subtraction of the network prediction from the ground truth.

It is demonstrated that the network is indeed able to learn the complex spatiotemporal features of the seismic sensor signal and henceforth translate them into the GGN strain. Furthermore, the subtraction of the predicted GGN waveform represents an actual noise reduction as indicated by the residuals. The metrics, as shown in the legend, underline this statement. The relative noise reduction efficiency implies mitigation by a factor of three.

A closer look at the waveform reveals that the network virtually recognizes the attributes of the waveform entirely. Due to the underlying loss function (MSE), however, sections of lower amplitude are generally reconstructed with slightly worse resolution since the absolute reward for the network is less. Choosing an alternative loss function that is sensitive to the relative deviation is a potential way of improving this behavior.

An overview of the general performance is presented as histograms of the relevant metrics in [fig. 5.5](#).

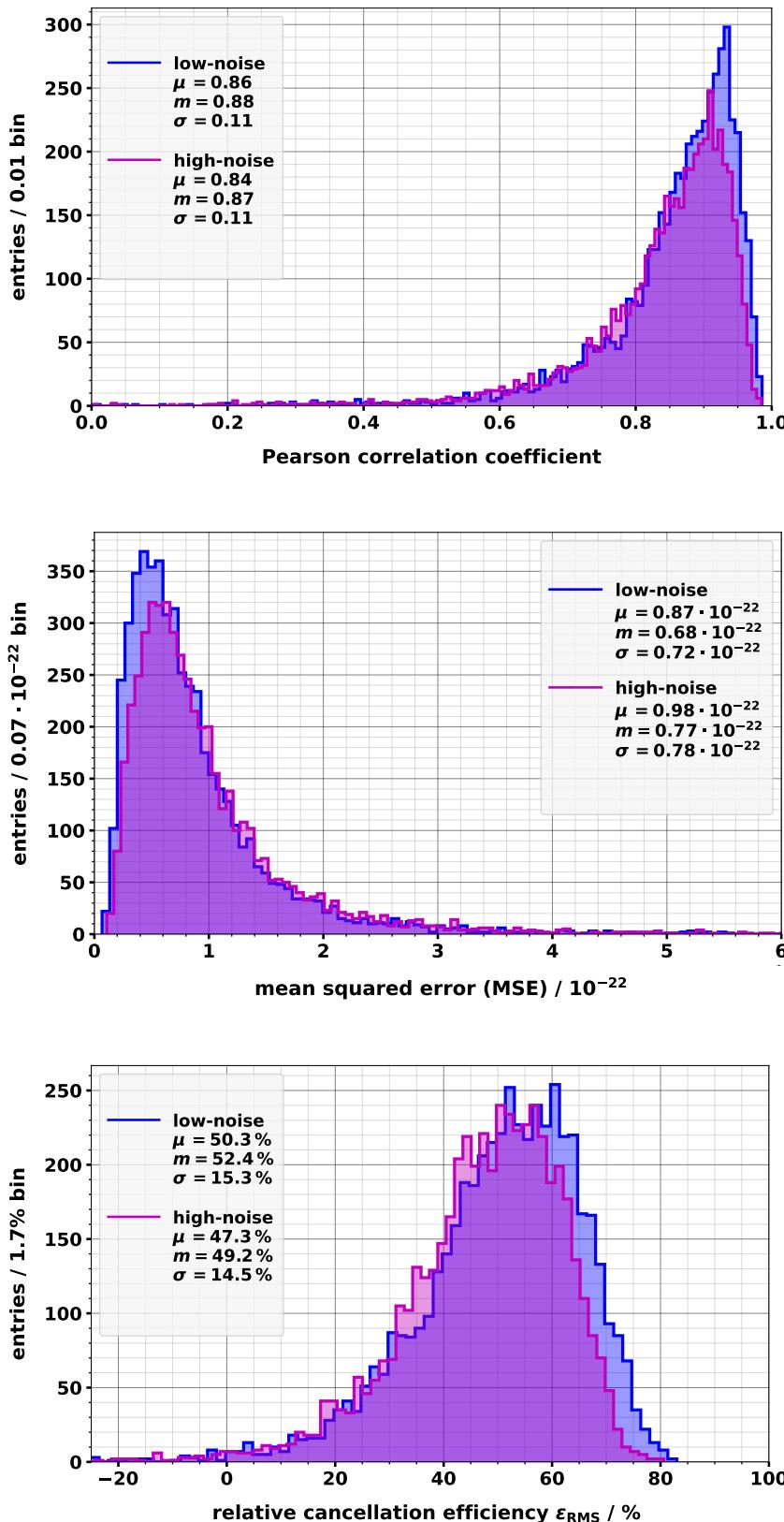


Figure 5.5: Evaluation metric histograms of the generic test case.

The blue and magenta histograms belong to the low-noise and high-noise scenarios, respectively. The mean (μ), median (m), and standard deviation (σ) values are given in the legends.

top: The distribution of linear correlation.

top: The distribution of linear correlation.

middle: The absolute deviation expressed as MSE.

bottom: The efficiency of the mitigation based on the RMS values.

The distributions of evaluation metrics, as shown in [fig. 5.5](#), portray the network performance evaluated over 5000 waveforms. All three metrics exhibit skewed distributions. In the following, the mean value is used as an estimator of the central values. Additionally, the median values are listed in the legend of [fig. 5.5](#) to give an impression of the skewness of each distribution. Further, for the sake of simplicity, the standard deviation is chosen to describe the (one-sided) broadness of the distributions.

As indicated by the learning curves, the performance in the low-noise scenario is slightly superior in all three metrics. However, the deviation of the mean values is not significantly large and the distributions share almost identical widths. This underlines the claim made in [sec. 5.2.1](#) that a neural network might be more robust to sensor noise than, for instance, a WF.

Mean linear correlations of 0.86 and 0.84 for the low-noise and high-noise scenarios respectively, demonstrate the strong linear correlation of predicted and true waveform. The spread of the distributions expressed as standard deviations is 0.11 and equal to the second decimal place for both noise cases. The evaluation of the cross-correlation results in a zero-offset for all test cases.

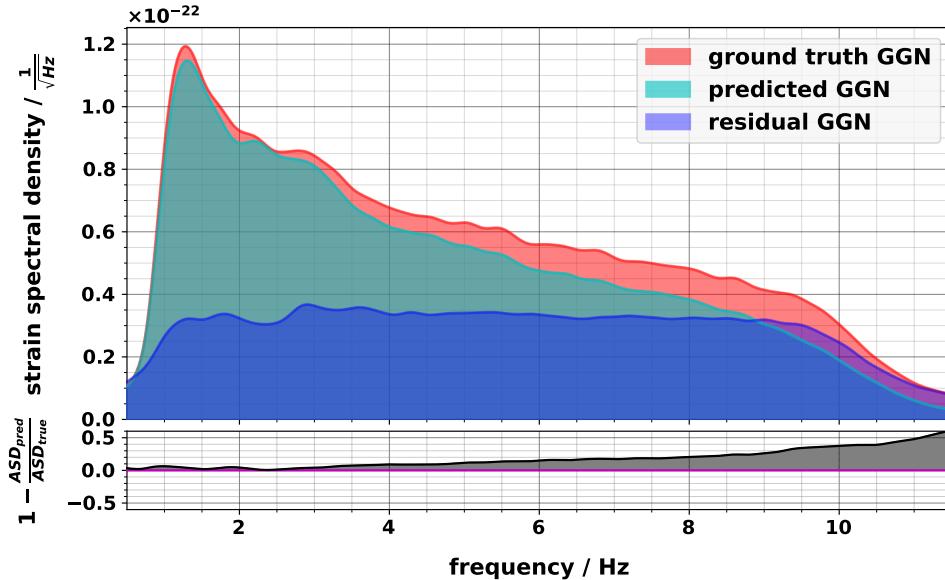
For the absolute deviation as expressed in MSE, the distributions show similar behavior with mean values 0.87×10^{-22} and 0.98×10^{-22} for the low-noise and high-noise scenario, respectively. The broadness of both skewed distributions is similar as expressed by the standard deviations of 0.72×10^{-22} for low-noise and 0.78×10^{-22} for high-noise.

An average noise suppression by a factor of two in both noise scenarios is indicated by the cancellation efficiency histograms. The mean values and standard deviations are 50.4 % and 15.3 % for the low-noise case and 47.3 % and 14.5 % for the high-noise case. The fact that negative efficiency values appear in the histogram indicates an important detail. In a few exceptional cases, the RMS of the residual strain is larger than the RMS of the initial strain before the subtraction. These cases share the same pattern: Fractions of the waveform get reconstructed out of phase, which leads to «positive interference» in the subtraction of the two waveforms. However, the prediction is mostly in phase for the remaining waveform length, where the GGN is suppressed. An example is given in [fig. A2](#) in the appendix. These cases can be identified by a weak linear correlation and a potential way to mitigate this effect is to include the correlation coefficient proficiently in the loss function of the training. It is important to note that these cases comprise only 0.64 % of the test cases in the low-noise scenario and 0.70 % in the high-noise scenario.

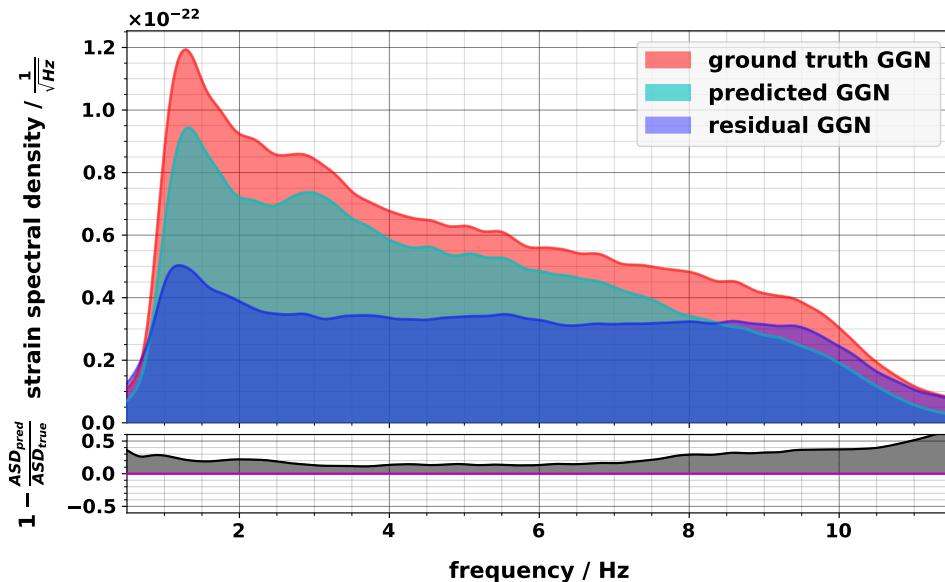
5.2.4 SPECTRAL ANALYSIS

An evaluation of the spectral performance is conducted accordingly to the in [sec. 4.3.3](#) described method of ASD analysis. The method is additionally applied to all test dataset predictions and residuals after the GGN cancellation. Results of the low-noise and high-noise scenarios are presented in [fig. 5.6](#).

Both cases exhibit similar spectral features. On average, the GGN is suppressed by a factor of two, and the residuals exhibit a flat spectral density for the most part. This implies that after the noise cancellation, the GGN has no specific frequency dependency anymore, similar to white noise. A factor adding to this is that the network tends to focus on the dominant frequencies in the training datasets. For high frequencies, particularly above 10 Hz, the network has fewer statistics available and thus develops a bias to predict them less often.



(a) Low sensor noise scenario.



(b) High sensor noise scenario.

Figure 5.6: Spectral performance analysis for the generic test case.

top frame in (a) and (b): The ASDs of the ground truth (red), predicted (green) and residual (blue) GGN strains for the low-noise (a) and high-noise (b) scenario.

bottom frame in (a) and (b): The relative deviation of the prediction from the ground truth normalized to zero. Positive values indicate underprediction and negative values indicate overprediction.

Specific features like a small bump around 2 Hz in the residuals of the low-noise scenario or a peak around 1 Hz in the residuals of the high-noise scenario can be hard to interpret and are likely an artifact due to the individual training process. Repetition of the training and analysis steps under the same conditions can test this assumption. A beneficial aspect that can be observed is that the network tends to rather underpredict than overpredict. An overpredicted GGN strain would introduce artificial noise if subtracted from the strain data.

5.3 EUREGIO TEST SCENARIO

The network is additionally trained and tested on EMR site-specific data by adjusting the simulation as described in [sec. 5.2](#). In this scenario, low-frequency content contributes dominantly to the GGN strain as a result of the site's characteristic seismic activity. A central objective of this analysis is to test the influence of specific characteristics of seismic data on the network's learning process. Further, the prospects of sufficient noise mitigation to suppress the expected GGN at the EMR candidate site below the ET-D design sensitivity and potential challenges are investigated.

In analogy to the methodology in [sec. 5.2](#), two noise scenarios will be compared: One representing negligible seismic sensor self-noise and another one representing self-noise with SNR of order $\mathcal{O}(10)$.

5.3.1 LEARNING CURVES

The learning curves of the two noise scenarios are displayed below in [fig. 5.7](#).

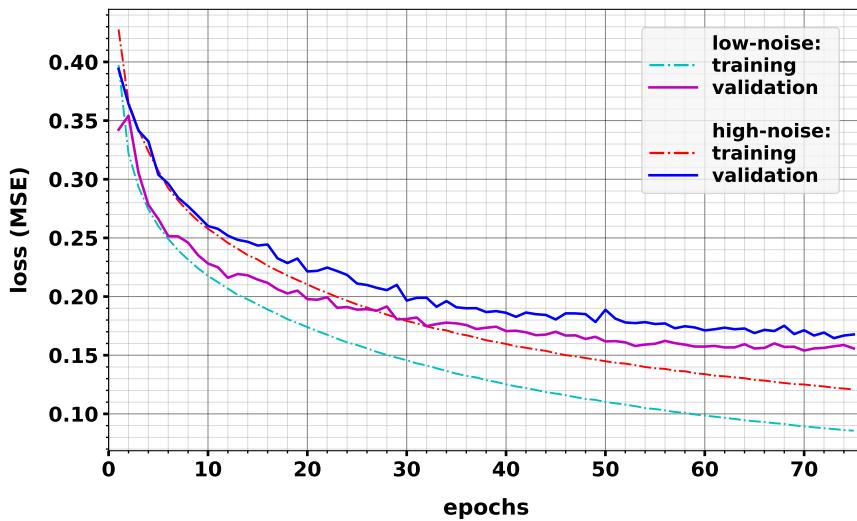


Figure 5.7: Learning curves for the EMR specific test scenario.

The training and validation losses as a function of training epochs are compared for the low-noise and high-noise scenarios.

In comparison to the general test case ([sec. 5.2](#)), the loss curves are lower in general. However, the generalization gap between training and validation loss is noticeably more prominent for both noise scenarios. As a countermeasure, the training is conducted with an increased dropout rate of 0.6 compared to an initial value of 0.4 in the general case training. A potential explanation for the worse generalization behavior might lie in the dominance of low-frequency seismic waves. With a lower frequency, the number of wavelengths present in each waveform of one event decreases, and the overlap of wave-packets increases. Hence, the network's challenge of extracting spatiotemporal features from the seismic sensor signals grows, which in turn negatively impacts generalization capability.

Similar to the general test, the generalization gap of the network is slightly smaller for the high-noise scenario, whereas the total loss is less for the low-noise scenario. The relative difference in final validation loss between both noise scenarios is 7.6 %.

5.3.2 PERFORMANCE ANALYSIS

To evaluate the network's performance for each test event, the in sec. 5.2 introduced metrics are utilized. The resulting histograms are presented in fig. 5.8.

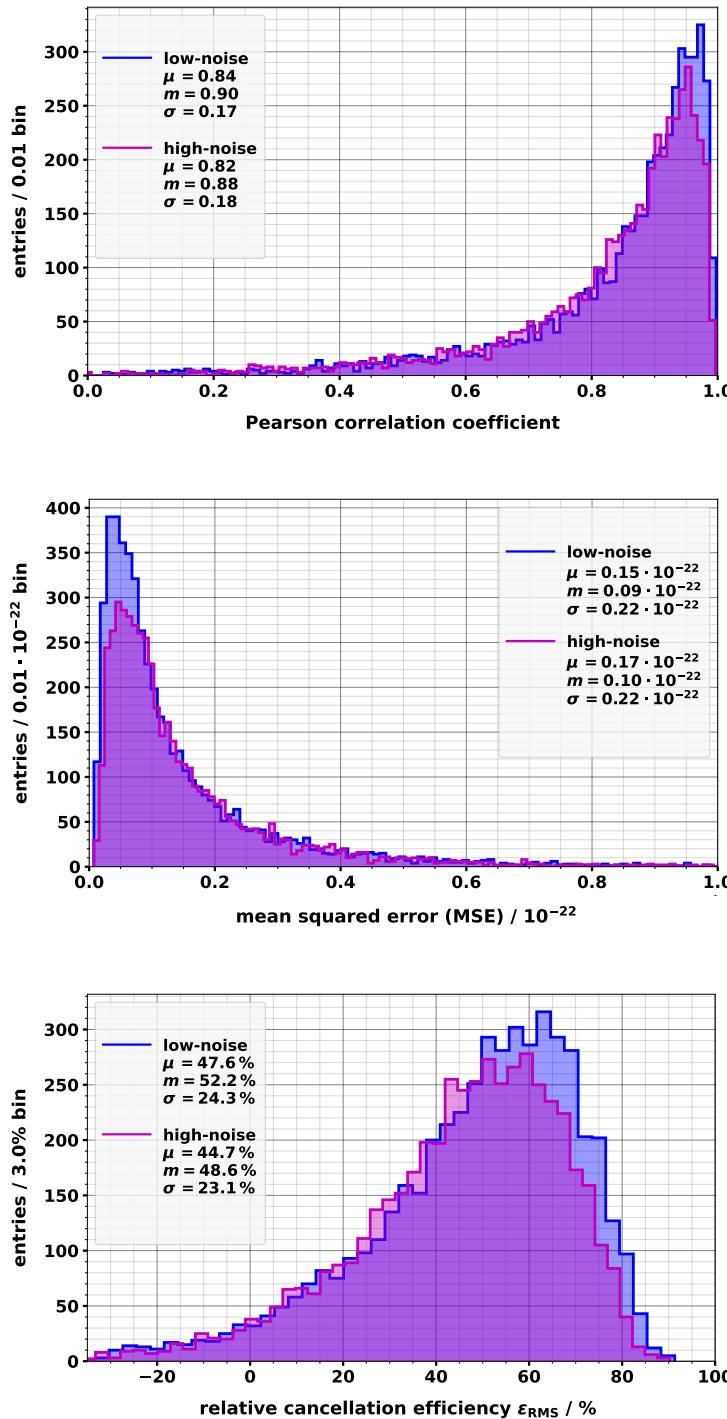


Figure 5.8: Evaluation metric histograms of the EMR specific test case.

The blue and magenta histograms belong to the low-noise and high-noise scenarios, respectively. The mean (μ), median (m), and standard deviation (σ) values are given in the legends.

top: The distribution of linear correlation.

middle: The absolute deviation expressed as MSE.

bottom: The efficiency of the mitigation based on the RMS values.

Overall, the distributions indicate broader relative spreads and more significant skewness in all three metrics compared to the general test case. The slightly worse generalization capability of the network, as indicated in [fig. 5.7](#), increases the variance of prediction performance for the evaluation on test data.

Mean linear correlation coefficients of 0.84 and 0.82 for the low-noise and high-noise scenarios, respectively, indicate a still strong but slightly weaker linear correlation compared to the general test case. The broadness of the distributions, as measured by the standard deviation, is with 0.17 (low-noise) and 0.18 (high-noise), almost identical for both noise scenarios but wider and more skewed compared to the general test case.

The mean MSE values are with 0.15×10^{-22} for low-noise and 0.17×10^{-22} for high-noise, nearly a factor of six smaller compared to the general test case. Again, both cases display a broader and more skewed distribution with equivalent standard deviations of 0.22×10^{-22} .

The average noise suppression is slightly less than a factor of two. The mean values are 47.6 % for the low-noise case and 44.7 % for the high-noise case. More significant skewness and broadness compared to the general test are indicated by standard deviations of 24.3 % for the low-noise case and 23.1 % for the high-noise case. As a result, the fraction of reconstructed GGN waveforms with phase-shifted parts, manifested by a negative ϵ_{RMS} value, increases to 4.4 % in both sensor noise scenarios.

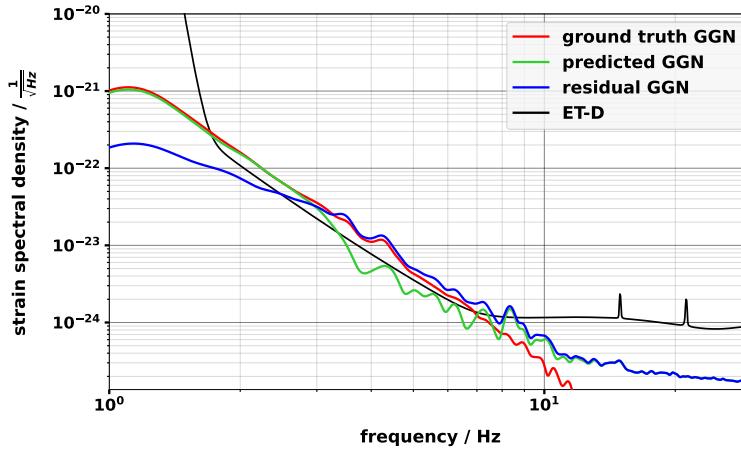
In summary of the presented metric analysis, it can be concluded that it might be crucial to adjust the training process to data with specific characteristics. More precisely, optimized regularization techniques might be required in the case of dominant low-frequency contributions to the seismic field. Based on the observation that an overall lower MSE loss does not necessarily improve the correlation metric nor the GGN cancellation efficiency, an adapted loss function that incorporates the correlation coefficient appears profitable.

Nonetheless, an average GGN cancellation by almost a factor of two is promising, especially with the prospect of significantly improved performance after additional network optimization.

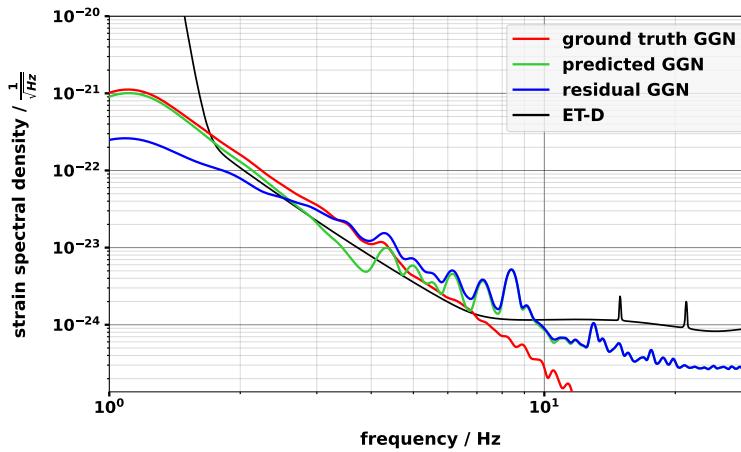
5.3.3 SPECTRAL PERFORMANCE ANALYSIS

A spectral performance analysis is performed in analogy to [sec. 5.2.4](#). The ASD of the simulated GGN strain reflects the expected GGN level at the EMR candidate site [3]. Based on the network predictions, GGN cancellation is implemented by subtraction from the simulated data. The ASD of the consequential residual GGN strain is compared to the ET-D design sensitivity [2]. The results for the low-noise and high-noise scenarios are presented in [fig. 5.9](#).

A pattern shows in both cases: the GGN cancellation is very effective for low frequencies ($f \lesssim 2.5$ Hz) with almost one order of magnitude suppression, which is well sufficient to reach the ET-D sensitivity. Moreover, the difference between the low and high seismic sensor noise cases is not significantly large. This demonstrates once again that the network has a certain input noise robustness and is able to filter the seismic features relevant for the corresponding GGN strain. However, the network predictions get significantly smaller than the ground truth for larger frequencies. Around 5 Hz in the high-noise and 7 Hz in the low-noise case, the predictions start to saturate with small amplitudes that are larger than the ground truth. This has the effect of artificial noise in the subtracted residuals.



(a) Low sensor noise scenario.



(b) High sensor noise scenario.

Figure 5.9: Spectral performance analysis for the EMR specific test case.

The ASDs of the ground truth (red), predicted (green) and residual (blue) GGN strains for the low-noise (a) and high-noise (b) scenario.

The design sensitivity (black) of ET-D as stated in [2].

In reflection of the properties of the training dataset, this behavior becomes plausible. If a neural network, or more generally any data-driven machine learning algorithm, is trained on biased data, the algorithm will internalize this bias. [88] In the present use case, the seismic field and the consequent GGN strain are dominated by low frequencies. Hence the network focuses on predicting them with high accuracy. In relation, the higher frequency contributions become unimportant due to the statistically lower significance and the lower gain in the loss for an accurate prediction. So if the network is applied to the less recurring higher frequency components in the seismic data, it has not learned to predict the corresponding GGN strain correctly and is, more or less, forced to guess.

Therefore, the main challenge is to train a non-biased network that performs equally well over the relevant frequency bandwidth with seismic data that has potentially strong frequency dependencies and varies over several orders of magnitude. It is an open question whether or not it is sufficient to increase the sheer volume of data and optimize the network architecture as well as the training process to the point where less frequent features in the data gain importance. Alternatively, artificial data augmentation to balance the training data or supplementary techniques may become inevitable.

5.4 INCLINATION DEPENDENCY

As discussed in [sec. 4.3.2](#), the inclination angles of seismic waves relative to the plane of the ET-LF interferometers and the seismic sensors are restricted to values between $\pm 45^\circ$ to reduce the sparsity of the datasets. In this section, the network's performance as a function of inclination angle is evaluated with the primary intent of investigating any potential inclination bias due to the planar seismic sensor arrangement and the restricted inclination angle.

For this purpose, the network is trained on a dataset containing events of one incident seismic wave of known inclination angle. The training does not include sensor noise and the learning curves can be found in [fig. A3](#) in the appendix.

As an analysis, the loss of each individual event from the test dataset is determined and plotted against the corresponding inclination angle. In order to have uniformly distributed angles in the histogram, the cosine of the zenith angle relative to the detection plane is plotted. The transformation is $\theta_{\text{zenith}} = \pi/2 - i$ and the resulting inclination angle dependency is displayed below in [fig. 5.10](#).

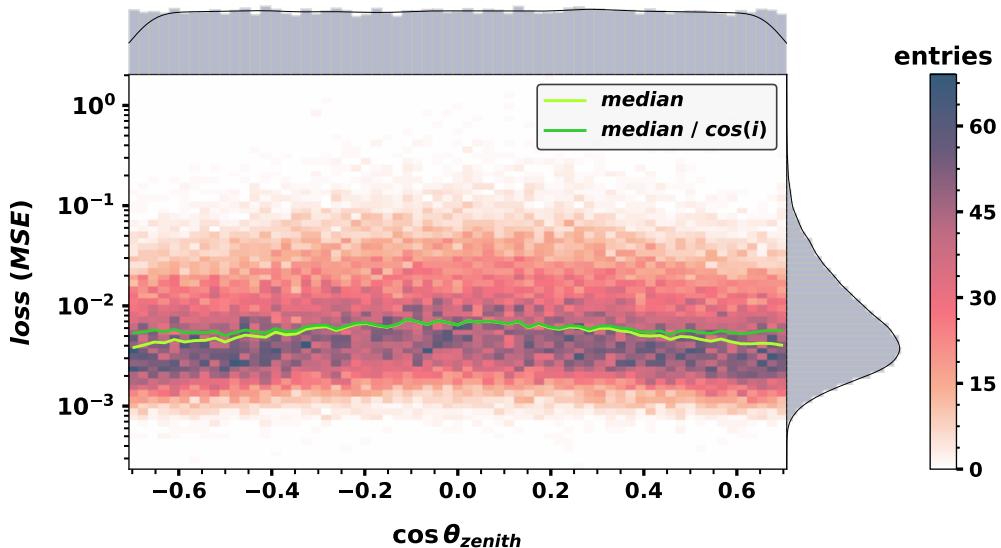


Figure 5.10: Evaluation of the network performance as a function of seismic wave inclination angle relative to the seismic sensor plane.

The MSE loss as a function of inclination angle is displayed as a two-dimensional histogram. Inclination angles are expressed in terms of the uniformly distributed cosine of the zenith angle (equal to $\sin i$). Median (lightgreen) and the median corrected for the inclination angle (green) are shown as horizontal lines. The axes histograms with probability density function display the distributions of angles and loss values.

A nearly homogenous distribution over the range of inclination angles can be observed. The numerical spread of loss values for a fixed inclination angle is relatively large as the MSE strongly scales with the absolute magnitude of the signal. Nevertheless, it can be excluded that the network develops a bias towards seismic waves in or close to the seismic sensor plane. Quite on the contrary, the loss values slightly decrease towards the boundaries of $\pm 45^\circ$ as indicated by the median³ of the distributions.

³a median representation is chosen due to the robustness against outliers.

However, the slight curvature is most likely explained by the geometrical effect that has already been mentioned in sec. 4.3.2: As the wave field is mainly longitudinally polarized, the projection of seismic displacement into the relevant interferometer plane decreases as the inclination angle increases and so does the MSE loss. A loss function sensitive to the relative and not the absolute deviation should not exhibit this trend. As compensation for this geometrical effect, the median divided by the cosine of the inclination angle is included in the representation in fig. 5.10. Alternatively, a representation of loss relative to the respective order of magnitude of GGN strain can be found in fig. A4 in the appendix. Both representations exhibit an almost constant reconstruction performance as a function of the inclination angle.

In summary of the above-made observations, it can be concluded that the inclination angle of seismic waves relative to the seismic sensor and ET-LF interferometer plane has a negligible impact on the network's ability to reconstruct the affiliated GGN strain.

5.5 RELATIVE SENSOR IMPORTANCE

As a first step toward an optimized seismic sensor array, an analysis is performed to evaluate the relative importance of each seismic sensor for the network. The idea is to investigate how the positioning of each seismic sensor influences the sensor's contribution to the network's GGN prediction.

For this sake, a method called *Saliency Mapping* [89] is followed. Initially, this method was developed to investigate the importance of distinct features in image-like data for classification tasks with CNNs. However, it can straightforwardly be adapted to a regression task as present in this thesis. The fundamental concept is that important pixels, or in this case seismic sensors, are indicated by large gradients during the training. To analyze the gradients of a given prediction, backpropagation through the network is performed analogously to the training process. However, the backpropagation is extended to the input layer, and the result is a *Saliency* map with the exact dimensions of the network's input data. For the GGN cancellation network, the *Saliency* maps have the dimensions (events, sensors^x, sensors^y, timesteps, sensoraxis). For a generalized analysis, the *Saliency* maps of 10 000 events are computed and averaged. Further, the mean average over the time axis is performed to get a time-independent representation of the importance of each seismic sensor.

The resulting *Saliency* maps are characteristics of each training. Exemplary, an instance of the high-noise general test case in sec. 5.2 is presented in fig. 5.11 together with an instance of the low-noise EMR test case in sec. 5.3.

A general pattern is that seismic sensors in the immediate vicinity of test-mass mirrors are of higher relevance due to the more significant spatial correlation of the signals with the GGN strain. Especially the sensors measuring the y-axis of the seismic field show a strong dependency on the seismic sensors close to the two end-mirrors of the interferometer. Moreover, since the gradients can be both negative and positive, symmetry patterns become visual in the *Saliency* maps. Generally, a symmetry of sensor importance around the interferometer's symmetry axis at $x = 0$ can be observed. For seismic sensors measuring the seismic field's x-axis, this symmetry is almost perfectly mirror-symmetrical. In contrast, the symmetry for the sensor y-axis shows an anti-correlated pattern.

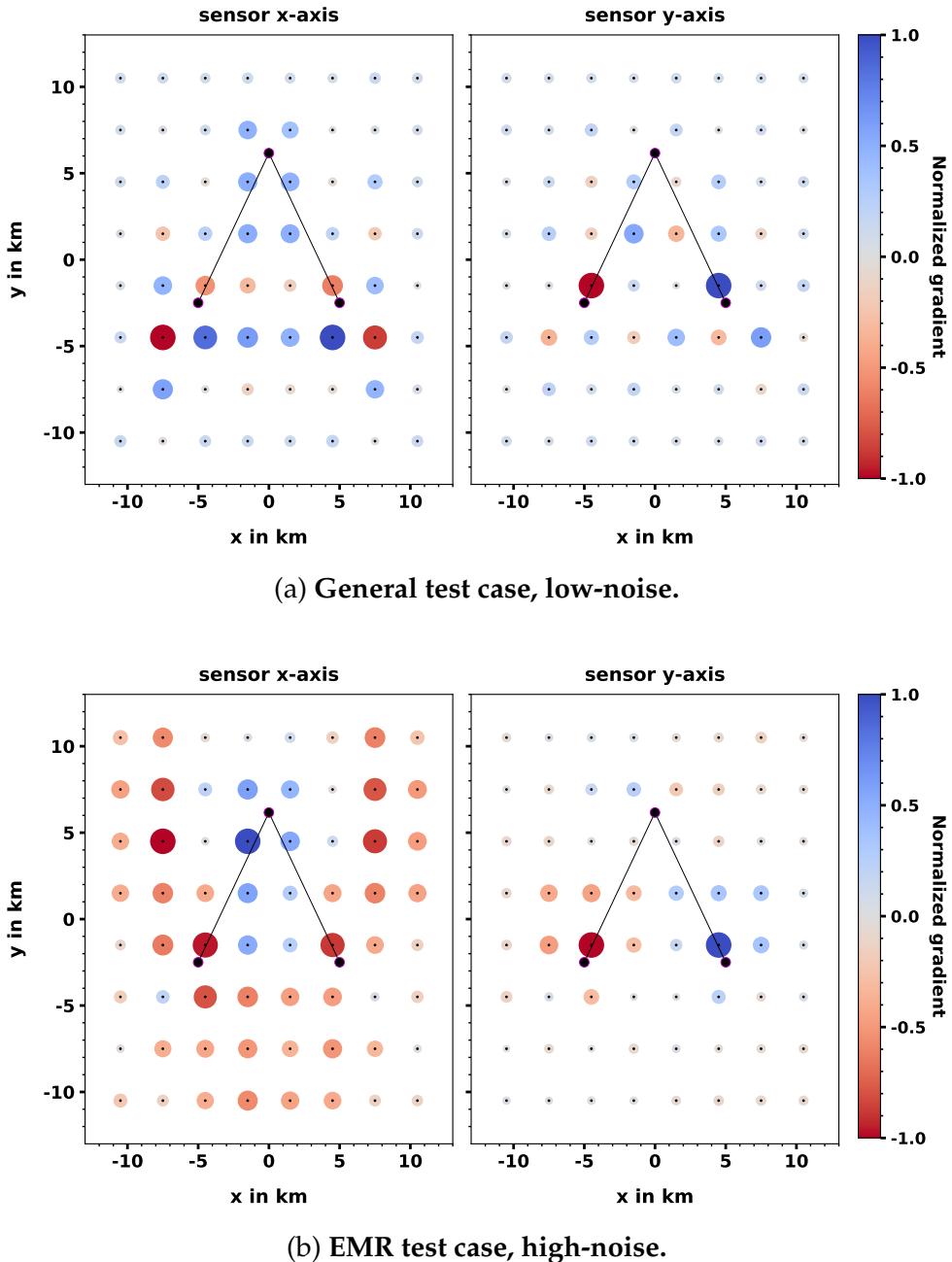


Figure 5.11: **Relative importance of seismic sensors for the network prediction.**

The *Saliency* maps representing the relative importance of each seismic sensor for the low-noise general test case (a) and the high noise EMR test case (b). Each scatter-point represents a seismic sensor at the respective grid position. The gradient is encoded in the color and size of each point. For better visualization, the gradients are normalized between $[-1, +1]$. The mirrors of the relevant ET interferometer are displayed in black.

left/right in (a) and (b): The seismic sensor x-axis/y-axis.

Those symmetries shed light upon the network's functionality and show that the network internalizes the geometrical properties of the working principle of the interferometer. The insights into important sensor positions and potentially exploitable symmetries can help to build the bridge toward a seismic sensor arrangement with optimized efficiency.

DISCUSSION & OUTLOOK

Einstein Telescope (ET) is a third-generation gravitational wave observatory aiming to increase the sensitivity of existing detectors by a factor of ten over the whole frequency band. Given these unprecedented dimensions, the challenge of noise suppression grows even more demanding. In the low-frequency regime of 1 to 10 Hz particularly, Gravity-Gradient Noise (GGN) is expected to be a limiting environmental noise factor for the design sensitivity and mitigation techniques will become crucial. [2, 3, 4]

In this work, a novel approach of utilizing deep learning methods for GGN cancellation was developed and investigated. For this purpose, a GGN model, including auxiliary seismic sensors, has been devised and implemented with a stochastic simulation of seismic fields as foundation. A neural network architecture was designed and tested for the task of predicting GGN strains based on the measurements of a seismic sensor array.

A first general test demonstrated that the proposed network architecture is capable of reconstructing the GGN strain given the corresponding signal from the seismic sensor array. The successful proof of concept shows that the network can capture complex spatiotemporal features and learn the transfer function to translate them into the affiliated GGN strain. Two sensor noise scenarios were compared. One case addressed a potentially achievable sensor accuracy with negligible instrumental noise, and another treated instrumental sensor noise with a signal-to-noise ratio of order $\mathcal{O}(10)$. In both cases, the GGN could be mitigated by a factor of two on average. The attribute that the network performance has turned out to be only slightly affected by the sensor noise represents a potential advantage over the current gold standard of Wiener Filters (WFs). The latter's efficiency can strongly reduce in the presence of instrumental noise [7].

An additional site-specific test was performed with data of modified frequency content that reproduces the seismic characteristics of a candidate site in the EMR. The resulting simulation output is in good agreement with the GGN strain spectral density as predicted from a seismic drilling campaign [3].

Altogether the results of the general case could be reproduced with a slightly worse overall performance of the network. Due to the more significant generalization gap, an average cancellation factor of slightly less than two has been achieved. Nonetheless, a spectral analysis demonstrated that the cancellation has proven to be sufficient to suppress the GGN below the ET design sensitivity for the dominant frequency range of $f \lesssim 2.5$ Hz. Training a network that should perform homogeneously well in the frequency domain with strains that vary over several orders of magnitude has been identified as the major challenge in this context. Further research should address the issue of mitigating the bias that the network learns from the underlying seismic data.

So far, the presented tests have been performed for seismic sensors arranged on a regular two-dimensional grid. An optimized positioning of seismic sensors should be targeted in order to reduce borehole costs. Thus, the efficiency of the sensor array should be improved by exploiting existing symmetry patterns and essential sensor positions. It has been demonstrated that neural networks can be utilized for this purpose in the form of a relative sensor importance analysis. For follow-up studies that address advanced seismic sensor arrangements, the proposed network architecture can, as a consequence of the modular approach, straightforwardly be adapted to graph convolutions [64].

Additional optimization steps like Bayesian parameter estimation [85] and deeper network architectures have the potential to significantly increase the GGN cancellation efficiency. To further improve the significance of the presented results and quantify the achievable GGN suppression, the GGN model should be refined to include site-specific geology and accurate observatory infrastructure. Numerical techniques such as finite element methods seem inevitable as the analytical modeling of realistic geologies is limited [4].

The associated challenge for deep learning applications is the required computational resources to produce reasonably large datasets. Fast seismic simulations aided by deep learning techniques are currently being investigated. They are in an early stage but show promising potential. [90]

With the prospect of third-generation detectors, however, interest in GGN mitigation will increase, as are scientific and computational resources. As computational resources increase, a potential online or forecasting application can become a topic of interest in the context of multi-messenger astronomy. Future studies should investigate the feasibility in terms of inference time and its optimization.

Overall the application of deep learning has demonstrated the potential to become a powerful tool for GGN cancellation at third-generation earthbound gravitational wave observatories like ET. Probable advantages over existing methods like WFs include superior instrumental noise robustness and model independence as the network autonomously learns the correlation between seismic sensors and detectors.

ABBREVIATIONS & NOMENCLATURE

LIST OF ABBREVIATIONS

- **aLIGO** advanced LIGO
- **ASD** Amplitude Spectral-Density
- **CNN** Convolutional Neural-Network
- **DL** Deep Learning
- **EMR** Euregio Meuse-Rhine
- **ET** Einstein Telescope
- **GGN** Gravity-Gradient Noise
- **GNN** Graph Neural-Network
- **GPU** Graphics Processing Unit
- **GW** Gravitational Wave
- **HF** high-frequency
- **IF** Interferometer
- **LF** low-frequency
- **LIGO** Laser Interferometer Gravitational-Wave Observatory
- **MMA** Multi-Messenger Astronomy
- **MSE** Mean Squared Error
- **NN** Neural Network
- **PSD** Power Spectral-Density
- **ReLU** Rectified Linear Unit
- **RMS** Root Mean Square
- **SGD** Stochastic Gradient Descent
- **SNR** Signal-to-Noise Ratio
- **WF** Wiener Filter

LIST OF NOMENCLATURE

· d_i	dilation rate of the i_{th} temporal convolution layer
· $\delta \vec{a}(\vec{r}, t)$	GGN acceleration perturbation
· $\vec{\xi}(\vec{r}, t)$	seismic displacement field
· $\delta \vec{r}_m(\vec{r}, t)$	mirror perturbation due to GGN
· ϵ_{RMS}	relative GGN cancellation efficiency in terms of RMS
· ET-D	«xylophone» design for the Einstein Telescope
· ET-HF	high-frequency interferometers of ET
· ET-LF	low-frequency interferometers of ET
· f_σ	activation function layer
· i	inclination angle relative to the detector plane
· k_F	kernel size of a convolution layer
· μ	mean value of a population
· m	median value of a population
· N_F	number of filters for a convolution layer
· r	<i>Pearson</i> coefficient of linear correlation
· $j_n(x)$	spherical <i>Bessel</i> function of n_{th} order
· σ	standard deviation of a population

LIST OF FIGURES

1.1	The proposed deep learning approach for GGN cancellation.	2
2.1	The first direct detection of gravitational waves.	4
2.2	The transversal polarization modes of a gravitational wave.	6
2.3	Schematic illustration of a modern Michelson-interferometer.	7
2.4	The optical layout of ET-D.	8
2.5	Total noise budget of ET-HF.	9
2.6	Total noise budget of ET-LF.	10
2.7	Coupling between seismic noise and interferometer test-masses. . .	12
2.8	Modulated seismic surface and body wave GGN in the Euregio. . .	13
2.9	Modulated total seismic GGN in the Euregio.	13
3.1	The biological neuron as archetype for artificial neurons.	17
3.2	Architecture of a shallow multilayer perceptron.	18
3.3	Optimization of network parameters with gradient descent.	19
3.4	Stochastic gradient descent with momentum.	20
3.5	Training and validation loss for performance monitoring.	21
3.6	Important activation functions and their derivatives.	23
3.7	Illustration of Residual and ReZero units.	24
3.8	Spatial two-dimensional convolution for an image.	26
3.9	One-dimensional dilated temporal convolutions for time series. .	28
4.1	Geometrical scaling factors of GGN cavity contributions.	34
4.2	Transmissibility curves of the foreseen super-attenuator for ET. .	35
4.3	Transmissibility of the simplified model of the super-attenuator. .	36
4.4	The simulated detector geometry of ET-LF.	37
4.5	The arrangement of the simulated seismic sensor array.	39
4.6	The ASD colored simulated seismic sensor noise.	40
4.7	Example of a simulated seismic sensor signal with noise.	40
4.8	Sampling of the seismic wave origins and propagation directions. .	45
4.9	Exemplary display of a simulated event.	46
4.10	ASD of the simulated GGN in the general test scenario.	47
4.11	ASD of the simulated GGN in the EMR specific test scenario. . .	48
5.1	The network architecture as a high- and low-level representation. .	51
5.2	Learning curves for the generic test scenario.	53
5.3	Exemplary display of the correlation metric between waveforms. .	54
5.4	Exemplary display of a predicted waveform and GGN cancellation.	55
5.5	Evaluation metric histograms of the generic test case.	56
5.6	Spectral performance analysis for the generic test case.	58
5.7	Learning curves for the EMR specific test scenario.	59
5.8	Evaluation metric histograms of the EMR specific test case. . . .	60
5.9	Spectral performance analysis for the EMR specific test case. . .	62
5.10	Evaluation of loss as a function of inclination angle.	63
5.11	Relative importance of seismic sensors for the network prediction.	65

A1	Y-axis and Z-axis seismic sensor signals of the exemplary event.	ix
A2	Example illustration of a partially phase-shifted GGN reconstruction.	x
A3	Learning curves for the inclination dependency test.	x
A4	Alternative representation of the inclination dependency analysis.	xi

LIST OF TABLES

4.1	GGN coupling factors for different cavity models.	34
4.2	Summary of simulation hyperparameters.	43
4.3	Summary of parameters of the simulated environment.	43
4.4	Summary of sampling parameter ranges and distributions.	44
5.1	Summary of network architecture hyperparameters.	52
5.2	Summary of network training hyperparameters.	52

APPENDICES

CHAPTER 4:

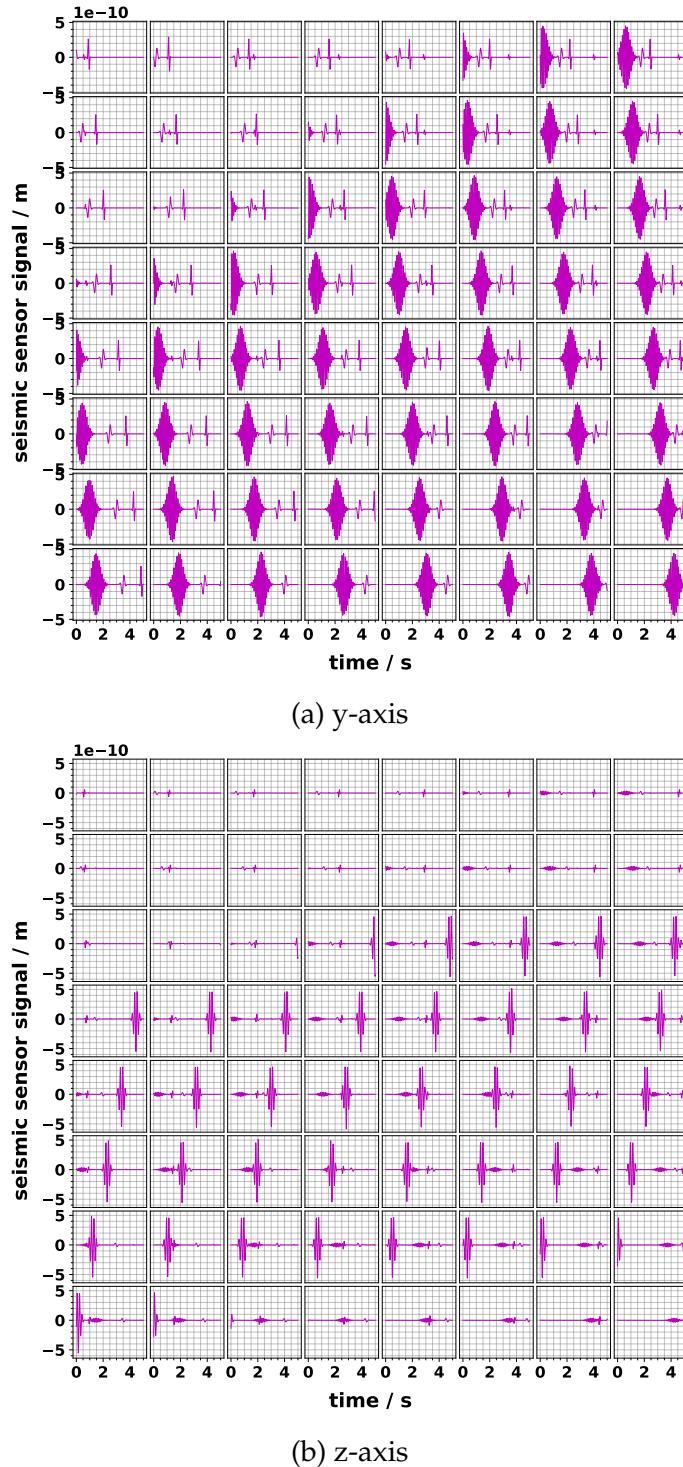


Figure A1: Y-axis and Z-axis seismic sensor signals of the exemplary event.

CHAPTER 5:

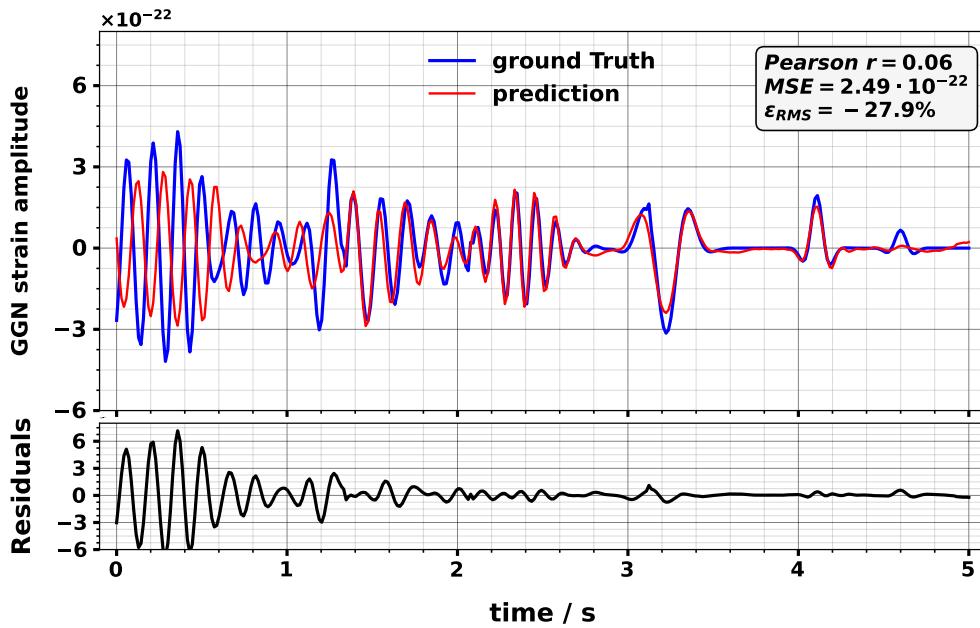


Figure A2: Example illustration of a partially phase-shifted GGN reconstruction.

top: The network prediction (red) compared to the simulation ground truth (blue).

bottom: The residual (black) as the subtraction of the network prediction from the ground truth.

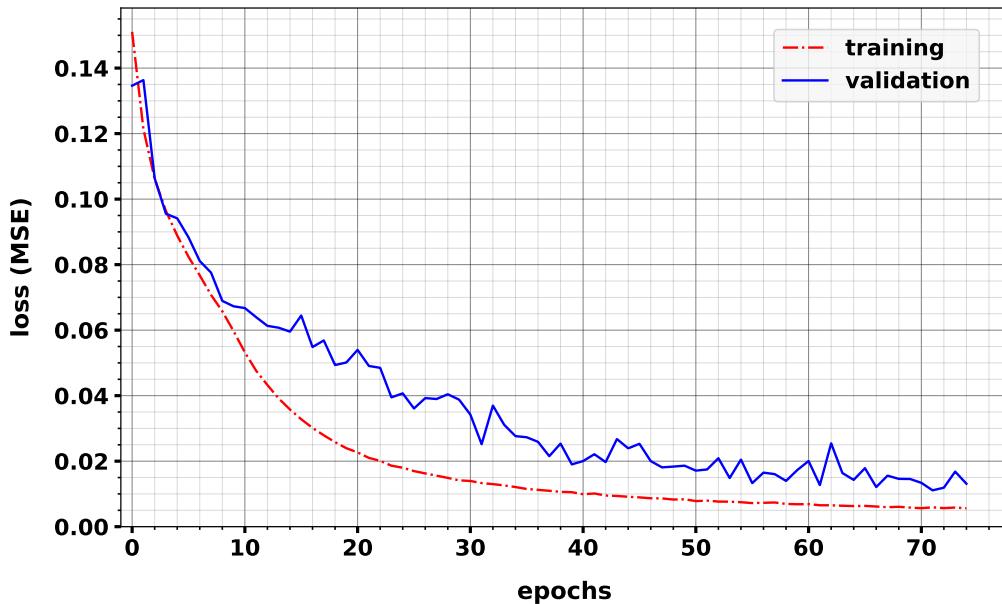


Figure A3: Learning curves for the inclination dependency test.

The training (blue) and validation (red) loss as a function of training epochs.

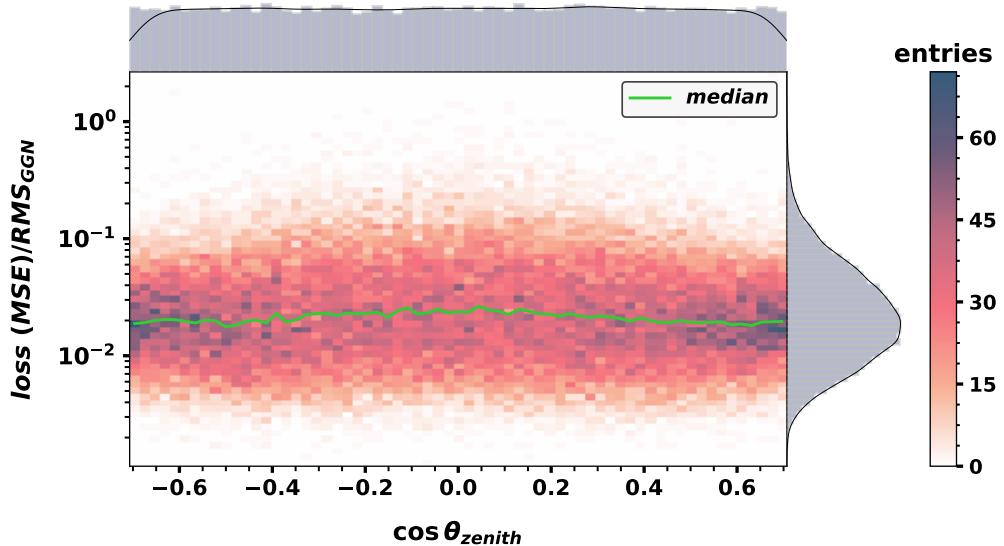


Figure A4: Alternative representation of the evaluation of the network performance as a function of seismic wave inclination angle relative to the seismic sensor plane.

The MSE loss as a function of inclination angle is displayed as a two-dimensional histogram. Inclination angles are expressed in terms of the uniformly distributed cosine of the zenith angle (equal to $\sin i$). The loss values are normalized by the RMS of the GGN strain and the median is displayed in green.

BIBLIOGRAPHY

- [1] R. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration). "Observation of Gravitational Waves from a Binary Black Hole Merger". *Physical Review Letters*, Vol. 116, 2016. doi:[10.1103/PhysRevLett.116.061102](https://doi.org/10.1103/PhysRevLett.116.061102).
- [2] ET Steering Committee Editorial Team. "Design Report Update for the Einstein Telescope": ET-0007B-20, 2020. Available from: <https://apps.et-gw.eu/tds/ql/?c=15418>.
- [3] M Bader, S Koley, J Van Den Brand, et al. "Newtonian-noise characterization at Terziet in Limburg—the Euregio Meuse–Rhine candidate site for Einstein Telescope". *Classical and Quantum Gravity*, Vol. 39, 2022. doi:[10.1088/1361-6382/AC1BE4](https://doi.org/10.1088/1361-6382/AC1BE4).
- [4] J Harms, L Naticchioni, E Calloni, et al. "A lower limit for Newtonian-noise models of the Einstein Telescope". *The European Physical Journal Plus*, Vol. 137, 2022. doi:<https://doi.org/10.1140/epjp/s13360-022-02851-z>.
- [5] R M Winslow, C Scolini, N Lugaz, et al. "Multi-messenger Astronomy: Gravitational Waves, Neutrinos, Photons, and Cosmic Rays". *Journal of Physics: Conference Series*, Vol. 718, 2016. doi:[10.1088/1742-6596/718/2/022004](https://doi.org/10.1088/1742-6596/718/2/022004).
- [6] K Cannon, R Cariou, A Chapman, et al. "Toward Early-Warning Detection of Gravitational Waves from Compact Binary Coalescence". *The Astrophysical Journal*, Vol. 748, 2012. doi:[10.1088/0004-637X/748/2/136](https://doi.org/10.1088/0004-637X/748/2/136).
- [7] M Coughlin, N Mukund, J Harms, et al. "Towards a first design of a Newtonian-noise cancellation system for Advanced LIGO". *Classical and Quantum Gravity*, Vol. 33, 2016. doi:[10.1088/0264-9381/33/24/244001](https://doi.org/10.1088/0264-9381/33/24/244001).
- [8] Kurt Hornik. "Approximation Capabilities of Multilayer Feedforward Networks". *Neural Networks*, Vol. 4, 1991. doi:[10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T).
- [9] Keiron O'Shea and Ryan Nash. "An Introduction to Convolutional Neural Networks". 2015. doi:[10.48550/arxiv.1511.08458](https://arxiv.org/abs/1511.08458).
- [10] Kaiming He, Xiangyu Zhang, and Shaoqing Ren. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification". *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015. doi:[10.1109/ICCV.2015.123](https://doi.org/10.1109/ICCV.2015.123).
- [11] H Song, M Kim, D Park, et al. "Learning from Noisy Labels with Deep Neural Networks: A Survey". *IEEE Transactions on Neural Networks and Learning Systems*, 2020. doi:[10.1109/TNNLS.2022.3152527](https://doi.org/10.1109/TNNLS.2022.3152527).
- [12] Lu Jiang, Di Huang, Mason Liu, and Weilong Yang. "Beyond Synthetic Noise: Deep Learning on Controlled Noisy Labels". *Proceedings of the 37th International Conference on Machine Learning, ICML*, 2020. doi:[10.48550/arxiv.1911.09781](https://arxiv.org/abs/1911.09781).

BIBLIOGRAPHY

- [13] Albert Einstein. "Die Grundlagen der Allgemeinen Relativitätstheorie". *Annalen der Physik*, Vol. 49, 1916. doi:[10.1002/ANDP.19163540702](https://doi.org/10.1002/ANDP.19163540702).
- [14] Michele Maggiore. "Gravitational Waves: Theory and Experiments". *Oxford University Press*, 2007. Available from: ISBN:[0198570740](#).
- [15] Hartmut Grote. "Gravitationswellen Geschichte einer Jahrhundertentdeckung". *Verlag C.H. Beck*, 2018. Available from: ISBN:[3406719414](#).
- [16] Joseph Weber. "Gravitational Radiation". *Physical Review Letters*, Vol. 18, 1967. doi:[10.1103/PhysRevLett.18.498](https://doi.org/10.1103/PhysRevLett.18.498).
- [17] Russel A. Hulse and Joseph H. Taylor. "Discovery of a Pulsar in a Binary System". *Astrophysical Journal*, Vol. 195, 1975. doi:[10.1086/181708](https://doi.org/10.1086/181708).
- [18] Walter Winkler. "Ein Laser-Interferometer als Gravitationswellendetektor". *Physik in unserer Zeit*, Vol. 16, 1985. doi:[10.1002/PIUZ.19850160502](https://doi.org/10.1002/PIUZ.19850160502).
- [19] Peter R. Saulson. "Interferometric Gravitational Wave Detection: Accomplishing the Impossible". *Classical and Quantum Gravity*, Vol. 17, 2000. doi:[10.1088/0264-9381/17/12/315](https://doi.org/10.1088/0264-9381/17/12/315).
- [20] R. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration). "LIGO: the Laser Interferometer Gravitational-Wave Observatory". *Reports on Progress in Physics*, Vol. 72, 2009. doi:[10.1088/0034-4885/72/7/076901](https://doi.org/10.1088/0034-4885/72/7/076901).
- [21] F. Acernese, M. Agathos, K. Agatsuma, et al. "Advanced Virgo: a second-generation Interferometric Gravitational Wave Detector". *Classical and Quantum Gravity*, Vol. 32, 2014. doi:[10.1088/0264-9381/32/2/024001](https://doi.org/10.1088/0264-9381/32/2/024001).
- [22] R. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration). "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral". *Physical Review Letters*, Vol. 119, 2017. doi:[10.1103/PhysRevLett.119.161101](https://doi.org/10.1103/PhysRevLett.119.161101).
- [23] M. Punturo, M. Abernathy, F. Acernese, et al. "The Third Generation of Gravitational Wave Observatories and their Science Reach". *Classical and Quantum Gravity*, Vol. 27, 2010. doi:[10.1088/0264-9381/27/8/084007](https://doi.org/10.1088/0264-9381/27/8/084007).
- [24] Bernard Schutz. "A First Course in General Relativity". *Cambridge University Press*, 2009. doi:[10.1017/9781108610865](https://doi.org/10.1017/9781108610865).
- [25] Albert A. Michelson and Max von Laue. "Die Relativbewegung der Erde gegen den Lichtaether". *Naturwissenschaften*, Vol. 19, 1931. doi:[10.1007/BF01528662](https://doi.org/10.1007/BF01528662).
- [26] Wikimedia Commons. "LIGO's Interferometer, simplified", Retrieved: 5/11/2022. Available from: https://commons.wikimedia.org/wiki/File:LIGO_simplified.svg.
- [27] S Rowlinson, A Dmitriev, A Jones, et al. "Feasibility Study of Beam-Expanding Telescopes in the Interferometer Arms for the Einstein Telescope". *Physical Review D*, Vol. 103, 2020. doi:[10.1103/PhysRevD.103.023004](https://doi.org/10.1103/PhysRevD.103.023004).

- [28] S Hild, S Chelkowski, A Freise, et al. "A Xylophone Configuration for a third Generation Gravitational Wave Detector". *Classical and Quantum Gravity*, Vol. 27, 2009. doi:10.1088/0264-9381/27/1/015003.
- [29] M G Beker, M Blom, J Van Den Brand, et al. "Seismic Attenuation Technology for the Advanced Virgo Gravitational Wave Detector". *Physics Procedia*, Vol. 37, 2012. doi:10.1016/j.phpro.2012.03.741.
- [30] Jan Harms. "Terrestrial Gravity Fluctuations". *Living Reviews in Relativity*, Vol. 22, 2019. doi:10.1007/S41114-019-0022-2.
- [31] D Brundu, M Cadoni, M Oi, et al. "Atmospheric Newtonian Noise Modeling for Third-Generation Gravitational Wave Detectors". *Phys. Rev. D*, Vol. 106, 2022. doi:10.1103/PhysRevD.106.064040.
- [32] S Koley, M Bader, J Van Den Brand, et al. "Seismic Charaterization of the Euregio Meuse-Rhine in View of the Einstein Telescope: First results of seismic studies of the Belgian-Dutch-German site for Einstein Telescope", 2019, Accessed: 30/10/2022. Available from: <https://www.nikhef.nl/wp-content/uploads/2019/10/Terziet-Drilling-Campaign-Final-NoC.pdf>.
- [33] The LIGO Collaboration. "Sensitivity Curves for aLIGO, Kagra, Virgo, Voyager, Cosmic Explorer, and Einstein Telescope", 2021, Accessed: 4/11/2022. Available from: <https://dcc.ligo.org/LIGO-T1500293/public>.
- [34] Francesca Badaracco and Jan Harms. "Optimization of Seismometer Arrays for the Cancellation of Newtonian Noise from Seismic Body Waves". *Classical and Quantum Gravity*, Vol. 36, 2019. doi:10.1088/1361-6382/ab28c1.
- [35] Jennifer C. Driggers, Jan Harms, and Rana X. Adhikari. "Subtraction of Newtonian Noise using Optimized Sensor Arrays". *Physical Review D - Particles, Fields, Gravitation and Cosmology*, Vol. 86, 2012. doi:10.1103/PhysRevD.86.102001.
- [36] M Punturo, M Abernathy, F Acernese, et al. "The Einstein Telescope: a Third-Generation Gravitational Wave Observatory". *Classical and Quantum Gravity*, Vol. 27, 2010. doi:10.1088/0264-9381/27/19/194002.
- [37] S Bloemheuvel, J van den Hoogen, D Jozinović, et al. "Graph Neural Networks for Multivariate Time Series Regression with Application to Seismic Data". *International Journal of Data Science and Analytics*, 2022. doi:10.1007/s41060-022-00349-6.
- [38] F Amato, F Guignard, S Robert, and Mikhail Kanevski. "A Novel Framework for Spatio-Temporal Prediction of Environmental Data using Deep Learning". *Scientific Reports*, Vol. 10, 2020. doi:10.1038/s41598-020-79148-7.
- [39] Matthew F Dixon, Nicholas G Polson, and Vadim O Sokolov. "Deep Learning for Spatio-Temporal Modeling: Dynamic Traffic Flows and High Frequency Trading". *Applied Stochastic Models In Business and Industry*, Vol. 35, 2018. doi:10.1002/asmb.2399.
- [40] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep Learning". *Nature*, Vol. 521, 2015. doi:10.1038/nature14539.

BIBLIOGRAPHY

- [41] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. "Deep Learning (Adaptive Computation and Machine Learning Series)". *The MIT Press*, 2016. Available from: <http://www.deeplearningbook.org>.
- [42] Warren S. McCulloch and Walter Pitts. "A Logical Calculus of the Ideas Immanent in Nervous Activity". *The Bulletin of Mathematical Biophysics*, Vol. 5, 1943. doi:10.1007/BF02478259.
- [43] George Cybenko. "Approximation by Superpositions of a Sigmoidal Function". *Mathematics of Control, Signals and Systems*, Vol. 2, 1989. doi:10.1007/BF02551274.
- [44] Aurélien Géron. "Hands-On Machine Learning with Scikit-Learn and TensorFlow". *O'Reilly Media, Inc*, 2nd Edition, 2019. Available from: ISBN: 9781492032649.
- [45] Rukshan Pramoditha. "The Concept of Artificial Neurons (Perceptrons) in Neural Networks". *Towards Data Science - Neural Networks and Deep Learning Course*, 2021, Retrieved: 7/11/2022. Available from: shorturl.at/yBUXZ.
- [46] David E. Rumelhart and James L. McClelland. "Learning Internal Representations by Error Propagation". *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, pages 318–362, 1987, The MIT Press. Available from: <https://ieeexplore.ieee.org/document/6302929>.
- [47] M Erdmann, J Glombitza, G Kasieczka, and U Klemradt. "Deep Learning for Physics Research". 2021. doi:10.1142/12294.
- [48] Sebastian Ruder. "An Overview of Gradient Descent Optimization Algorithms". 2016. doi:10.48550/arxiv.1609.04747.
- [49] G E. Hinton, N Srivastava, A Krizhevsky, et al. "Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors". 2012. doi:10.48550/arxiv.1207.0580.
- [50] R M. Zur, Y Jiang, L L. Pesce, and Karen Drukker. "Noise Injection for Training Artificial Neural Networks: A Comparison with Weight Decay and Early Stopping". *Medical Physics*, Vol. 36, 2009. doi:10.1118/1.3213517.
- [51] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. "Deep Sparse Rectifier Neural Networks". *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011. Available from: <https://proceedings.mlr.press/v15/glorot11a.html>.
- [52] B Xu, N Wang, H Kong, et al. "Empirical Evaluation of Rectified Activations in Convolutional Network". 2015. doi:10.48550/arxiv.1505.00853.
- [53] Xavier Glorot and Y Bengio. "Understanding the Difficulty of Training Deep Feedforward Neural Networks". *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Vol. 9, 2010. Available from: <https://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf>.
- [54] Sergey Ioffe and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". *32nd International Conference on Machine Learning (ICML)*, Vol. 1, 2015. doi:10.48550/arxiv.1502.03167.

- [55] K He, X Zhang, S Ren, and Jian Sun. "Deep Residual Learning for Image Recognition". *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015. doi:10.48550/arxiv.1512.03385.
- [56] T Bachlechner, B P Majumder, H Mao, et al. "ReZero is All You Need: Fast Convergence at Large Depth". *37th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2020. doi:10.48550/arxiv.2003.04887.
- [57] Yuchen Liang, Chaitanya K. Ryali, Benjamin Hoover, et al. "Can a Fruit Fly Learn Word Embeddings?". 2021. doi:10.48550/arxiv.2101.06887.
- [58] D. H. Hubel and T. N. Wiesel. "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex". *The Journal of Physiology*, Vol. 160, 1962. doi:10.1113/JPHYSIOL.1962.SP006837.
- [59] Kunihiko Fukushima. "Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition unaffected by Shift in Position". *Biological Cybernetics*, Vol. 36:193–202, 1980. doi:10.1007/BF00344251.
- [60] Thilo Birkenfeld. "Electron-Positron Discrimination in the Jiangmen Under-ground Neutrino Observatory with Deep Neural Networks". Master Thesis, RWTH Aachen University, 2018, Accessed: 7/11/2022. Available from: https://www.institut3b.physik.rwth-aachen.de/global/show_document.asp?id=aaaaaaaaabbjnpa.
- [61] Jonathan Tompson, Yann LeCun, and Christoph Bregler. "Efficient Object Localization Using Convolutional Networks". *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 7, 2014. doi:10.48550/arxiv.1411.4280.
- [62] M. Bachlechner, T. Birkenfeld, P. Soldin, et al. "Partition Pooling for Convolutional Graph Network Applications in Particle Physics". *Journal of Instrumentation*, Vol. 17, 2022. doi:10.1088/1748-0221/17/10/P10004.
- [63] David K. Hammond, Pierre Vandergheynst, and Rémi Gribonval. "Wavelets on graphs via spectral graph theory". *Applied and Computational Harmonic Analysis*, Vol. 30, 2011. doi:10.1016/J.ACHA.2010.04.005.
- [64] J Zhou, G Cui, S Hu, et al. "Graph Neural Networks: A Review of Methods and Applications". *AI Open*, Vol. 1, 2020. doi:10.1016/J.AIOPEN.2021.01.001.
- [65] A van den Oord, S Dieleman, H Zen, et al. "WaveNet: A Generative Model for Raw Audio". 2016. doi:10.48550/arXiv.1609.03499.
- [66] Jan Harms and Jennifer C. Driggers (LIGO Collaboration). "Results of Phase 1 Newtonian Noise Measurements at the LIGO Sites", 2011, Accessed: 28/10/2022. Available from: <https://dcc.ligo.org/LIGO-T1100237/public>.
- [67] L Cauchie et al. G Saccorotti, Piccinini. "Seismic Noise by Wind Farms: A Case Study from the Virgo Gravitational Wave Observatory, Italy". *Bulletin of the Seismological Society of America*, Vol. 101, 2011. doi:10.1785/0120100203.

BIBLIOGRAPHY

- [68] Soumen Koley. "Sensor Networks to Measure Environmental Noise at Gravitational Wave Detector Sites", PhD Thesis, Vrije Universiteit Amsterdam, 2020, Accessed: 14/10/2022. Available from: <https://inspirehep.net/literature/1797599>.
- [69] Maria K. M. Bader. "Seismic and Newtonian Noise Modeling for Advanced Virgo and Einstein Telescope", PhD Thesis, Vrije Universiteit Amsterdam, 2021, Accessed: 14/10/2022. Available from: <https://inspirehep.net/literature/1852716>.
- [70] E Tiesinga, P J. Mohr, D B. Newell, and B N. Taylor. "Codata Recommended Values of the Fundamental Physical Constants: 2018". *Reviews of Modern Physics*, Vol. 93, 2021. doi:<https://doi.org/10.1063/5.0064853>.
- [71] C J. Ammon, A A. Velasco, T Lay, and T C. Wallace. "Foundations of Modern Global Seismology". Elsevier, 2020. doi:[10.1016/B978-0-12-815679-7.00002-1](https://doi.org/10.1016/B978-0-12-815679-7.00002-1).
- [72] H. Helmholtz. "Über Integrale der hydrodynamischen Gleichungen, welche den Wirbelbewegungen entsprechen". *Journal fur die Reine und Angewandte Mathematik*, 1858. doi:<https://doi.org/10.1515/crll.1858.55.25>.
- [73] George Arfken. "Spherical Bessel Functions". *Mathematical Methods for Physicists*, pages 622–635, Academic Press, 1985. Available from: ISBN: 9781483288062.
- [74] A. Allocca, A. Berbellini, L. Boschi, et al. "Seismic Glitchness at Sos Enattos Site: Impact on Intermediate Black Hole Binaries Detection Efficiency". *The European Physical Journal Plus*, Vol. 136, 2021. doi:[10.1140/EPJP/S13360-021-01450-8](https://doi.org/10.1140/EPJP/S13360-021-01450-8).
- [75] Kinematics. "Streckeisen STS-5A Datasheet", Accessed: 01/10/2022. Available from: https://kinematics.com/post_products/sts-5a/.
- [76] Bion John Merchant and George William Slad for the Sandia National Laboratories (SNL). "Next Generation Qualification: Kinematics STS-5A Seismometer Evaluation". 2017, Accessed: 01/10/2022. doi:[10.2172/1405269](https://doi.org/10.2172/1405269).
- [77] Claude E. Shannon. "Communication in the Presence of Noise". *Proceedings of the IRE*, Vol. 37, 1949. doi:[10.1109/JRPROC.1949.232969](https://doi.org/10.1109/JRPROC.1949.232969).
- [78] Boris Goncharov, Alexander H. Nitz, and Jan Harms. "Utilizing the Null Stream of the Einstein Telescope". *Physical Review D*, Vol. 105, 2022. doi:[10.1103/PhysRevD.105.122007](https://doi.org/10.1103/PhysRevD.105.122007).
- [79] Alan V Oppenheim, Ronald W Schafer, and John R Buck. "Discrete-Time Signal Processing". *Prentice-hall Englewood Cliffs*, Prentice-hall Englewood Cliffs, 2nd edition, 1999. Available from: ISBN: 978-0137549207.
- [80] Peter D. Welch. "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms". *IEEE Transactions on Audio and Electroacoustics*, Vol. 15, 1967. doi:[10.1109/TAU.1967.1161901](https://doi.org/10.1109/TAU.1967.1161901).

- [81] Francois Chollet et al. "Keras", 2015, Accessed: 02/10/2022. Available from: <https://github.com/fchollet/keras>.
- [82] A Martin, A Ashish, B Paul, et al. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems", 2015. Software available from tensorflow.org. doi:10.5281/zenodo.4724125.
- [83] NVIDIA. "A40 Graphic Processing Unit", 2022, Accessed: 02/10/2022. Available from: <https://www.nvidia.com/de-de/data-center/a40/>.
- [84] P Vingelmann and F.H.P Fitzek for NVIDIA. "CUDA", 2020, Accessed: 02/10/2022. Available from: <https://developer.nvidia.com/cuda-toolkit>.
- [85] J. Bergstra, D. Yamins, and D. D. Cox. "Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures". 2013. Available from: <https://github.com/hyperopt/hyperopt>.
- [86] David Freedman, Robert Pisani, and Roger Purves. "statistics (international student edition)", 4th edition. WW Norton & Company, New York, pages 37–40, 2007. Available from: ISBN:978-0-393-93043-6.
- [87] Ronald Bracewell. "Pentagram Notation for Cross Correlation". *The Fourier Transform and Its Applications.*, pages 46–47, McGraw-Hill, 1999. Available from: ISBN:978-0073039381.
- [88] E Ntoutsi, P Fafalios, U Gadiraju, et al. "Bias in Data-Driven Artificial Intelligence Systems: An Introductory Survey". *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 10, 2020. doi:10.1002/WIDM.1356.
- [89] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps". *2nd International Conference on Learning Representations (ICLR)*, 2013. doi:10.48550/arxiv.1312.6034.
- [90] Ben Moseley, Tarje Nissen-Meyer, and Andrew Markham. "Deep Learning for Fast Simulation of Seismic Waves in Complex Media". *Solid Earth*, Vol. 11, 2020. doi:10.5194/SE-11-1527-2020.

DANKSAGUNG

An dieser Stelle möchte ich die Gelegenheit nutzen und mich bei denjenigen Menschen zu bedanken, ohne die diese Arbeit nicht möglich möglich gewesen wäre.

Zunächst gilt mein Dank Prof. Achim Stahl für die Möglichkeit meine Abschlussarbeit über ein Thema zu verfassen, dass zwei mich besonders faszinierende Forschungsgebiete; einerseits die Gravitationswellen-Astronomie und andererseits das maschinelle Lernen, auf neue Art und Weise kombiniert. Weiter bin ich dankbar die Chance gehabt zu haben meine Arbeit auf einer internationalen Konferenz vorstellen zu dürfen. Außerdem möchte ich Prof. Christopher Wiebusch für das Übernehmen der Rolle als Zweitprüfer danken.

Weiterhin möchte ich mich bei Markus Bachlechner für die engagierte Betreuung bedanken. Unsere Diskussionen haben mir sehr geholfen klarere Gedanken zu fassen und größere sowie kleinere Probleme zu lösen.

Vielen Dank auch an die ET Aachen und Deep-Learning Arbeitsgruppe im Institut. Das aufrichtige Interesse und die konstruktiven Ratschläge in unseren Meetings waren sehr hilfreich.

Zu guter Letzt gehört mein Dank meiner Familie, meinen Freunden und Dine. Ohne euch wäre mein gesamtes Studium nicht möglich gewesen.

DECLARATION IN LIEAU OF AN OATH

I hereby declare in lieu of an oath that I have completed the present Master thesis independently and without illegitimate assistance from third parties (such as academic ghostwriters).

I have used no other than the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical.

DATE: _____ **SIGNATURE:** _____