Bayesian Imaging with Ground-Based Telescopes

Jakob Roth

München 2024

Bayesian Imaging with Ground-Based Telescopes

Jakob Roth

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Jakob Roth
aus Hamburg

München, den 09.09.2024

1st Referee: PD Dr. Torsten Enßlin
2nd Referee: Prof. Dr. Daniel Grün
Date of Doctoral Examination: 05.12.2024

Zusammenfassung

Astronomische Beobachtungen sind nicht nur für die Astrophysik und Kosmologie von großer Bedeutung, sondern auch für die Physik im Allgemeinen. In den letzten Jahrzehnten und Jahrhunderten wurden immer empfindlichere Teleskope entwickelt, um das Universum immer genauer zu erforschen. Heutzutage ist die Astronomie eng mit der Statistik und den Datenwissenschaften verbunden, um Beobachtungsdaten zu analysieren und das volle Potenzial moderner Teleskope auszuschöpfen. In dieser Arbeit werden Algorithmen zur Verbesserung der Rekonstruktion astronomischer Bilder aus den Rohdaten der Beobachtungen vorgestellt.

Die moderne Astronomie misst die Strahlung aus dem Universum im gesamten elektromagnetischen Spektrum. Die Beobachtung von optischem Licht und Radiowellen spielt jedoch nach wie vor eine einzigartige Rolle, da nur in diesen Frequenzen die Erdatmosphäre transparent ist und somit Beobachtungen vom Erdboden aus möglich sind. Die in dieser Arbeit vorgestellten Algorithmen wurden für bodengebundene optische Teleskope und Radioteleskope entwickelt, um den Einfluss atmosphärischer Fluktuationen auf die Bilder zu reduzieren und die Verarbeitung großer Radiointerferometer-Datensätze zu verbessern.

Ohne genaue Korrektur führen atmosphärische Fluktuationen zu unscharfen Teleskopbildern mit reduziertem Kontrast. Der erste Teil der Arbeit befasst sich mit der Rekonstruktion von Bildern in Gegenwart atmosphärischer Störungen. Insbesondere werden zwei Algorithmen für optische und Radioteleskope entwickelt, die aus den Rohdaten der Teleskope Schätzungen der Störungen zusammen mit dem ungestörten Bild rekonstruieren. Die Algorithmen erhöhen die Auflösung und den Kontrast der rekonstruierten Bilder erheblich.

Moderne Radiointerferometer, die in den letzten Jahren gebaut wurden, erzeugen wesentlich größere Messdatensätze als Teleskope früherer Generationen, und die Datenmengen werden bei zukünftigen Instrumenten weiter ansteigen. Diese Zunahme der Datenmenge stellt eine Herausforderung für die Algorithmen zur Bildrekonstruktion dar, insbesondere wenn eine hohe Genauigkeit angestrebt wird. Der zweite Teil der Arbeit befasst sich mit diesen Herausforderungen und stellt den `fast-resolve` Algorithmus vor, der Konzepte des rechnerisch günstigeren, aber weniger genauen `CLEAN` Algorithmus nutzt und dadurch die hochgenaue Rekonstruktion radiointerferometrischer Datensätze erheblich beschleunigt.

Die in dieser Arbeit vorgestellten Algorithmen basieren auf der Bayes'schen Inferenz, die eine Quantifizierung der Unsicherheit der Ergebnisse ermöglicht. Neben der hohen Genauigkeit ist dies ein zusätzlicher Vorteil, da viele andere häufig verwendete bildgebende Verfahren keine zuverlässigen Unsicherheitsabschätzungen erlauben.

Abstract

Astronomical observations have an immense impact not only on astrophysics and cosmology but also on physics in general. Over the past decades and centuries, increasingly sensitive telescopes have been developed to study the universe in ever greater detail. Nowadays, astronomy is closely connected with statistics and data science to analyze observational data and to exploit the full potential of modern telescopes. This thesis contributes algorithms to improve the reconstruction of astronomical images from raw observational data.

Modern astronomy observes radiation originating from the universe across the entire electromagnetic spectrum. However, observations of optical light and radio waves continue to play a unique role, as these are the only frequencies at which the atmosphere of the Earth is transparent, such that ground-based observations are possible. The algorithms presented in this thesis have been developed for ground-based optical and radio telescopes to reduce the influence of atmospheric fluctuations on the images and to improve the processing of large radio interferometer data sets.

Without precise correction techniques, atmospheric fluctuations lead to a blurring of telescope images, reducing the resolution and dynamic range. The first part of the thesis deals with the reconstruction of images in the presence of atmospheric perturbations. In particular, the thesis contributes two algorithms for optical and radio telescopes that reconstruct estimates of the perturbations together with the unperturbed image from the raw telescope data. The algorithms significantly increase the resolution and dynamic range of the recovered images.

Modern radio interferometers constructed in recent years produce significantly larger measurement datasets than previous generation telescopes, and the amount of data will continue to increase with future instruments. This increase in data size imposes challenges for image reconstruction algorithms, especially when aiming at high precision. The second part of the thesis addresses these challenges and introduces the `fast-resolve` algorithm, which uses concepts of the computationally cheaper but less accurate `CLEAN` algorithm to significantly speed up the high-precision reconstruction of radio interferometric datasets.

The algorithms introduced in this thesis are based on Bayesian inference, enabling an uncertainty quantification of the recovered results. Besides the high accuracy of the presented algorithms, this is an additional advantage, as many commonly used imaging methods cannot provide reliable uncertainty estimates.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

For thousands of years, humans have studied the sky. The origins of astronomy are closely tied to religious and mythological beliefs, but already in ancient times, astronomy was used for practical applications such as navigation and the determination of the calendar. Driven by curiosity, people have sought answers to cosmological questions through astronomical observations, leading from geocentric to heliocentric to our modern cosmological models. Initially, astronomical observations were made with the naked eye. Modern astronomy is closely linked to physics and engineering science, relying on telescope facilities for observations. With Karl Jansky's discovery of radio emission originating from the cosmos (Jansky, 1933), astronomy began to expand from observations at visible wavelengths to the entire electromagnetic spectrum. Nowadays the universe is not only observed in the full electromagnetic spectrum ranging from radio to gamma-ray, but also neutrinos, cosmic rays, and gravitational waves are observed. However, observations around the visible spectrum and at radio wavelengths keep playing a unique role in astronomy, as these are the only wavelengths at which ground-based observations of the electromagnetic sky are possible.

Observations with ground-based telescopes are affected by the atmosphere. Even with the naked eye, the impact of the atmosphere is noticeable as twinkling or scintillation of stars. When looking at the night sky the twinkling of the stars is very aesthetic, but for astronomical observation these atmospheric effects impose serious challenges. In optical astronomy, adaptive optics was developed to correct for atmospheric perturbations in real-time. For radio interferometric observations, atmospheric fluctuations are corrected via calibration steps during image synthesis. Nevertheless, the standard atmospheric correction techniques are not perfect, neither for radio interferometers nor for optical telescopes, and residual perturbations degrading image quality remain. The first part of this thesis contributes two novel reconstruction algorithms for optical telescopes and radio interferometers that correct for atmospheric fluctuations. Specifically, a high-contrast imaging algorithm reconstructs residual atmospheric perturbations after adaptive optics corrections of an optical telescope together with the unperturbed image itself. Building on the radio interferometric imaging framework `resolve` originally proposed by Junklewitz et al., 2016, the second algorithm estimates direction and time-dependent corrections for all antennas in a radio interferometer during the image reconstruction. Both algorithms are based on Bayesian inference, enabling uncertainty quantification of the recovered images and estimated atmo-

spheric perturbations.

The big advantage of ground-based telescope facilities over satellites is that the telescopes can be significantly larger with moderate costs. Larger telescopes can achieve higher resolution and better sensitivity, which is the primary reason why most optical and radio telescopes are still ground-based despite the aforementioned challenges of atmospheric perturbations. However, especially for radio interferometers, the construction of larger next-generation arrays leads to a massive increase in the data sizes. The increase in data volume imposes challenges for image formation, particularly for Bayesian and other advanced reconstruction methods, which improve image quality often at the cost of increased computational complexity. The second part of the thesis makes two contributions to address the computational challenges. On the one hand, it contributes a software package `JAXbind` allowing to bind high-performance code into the JAX machine learning framework. This is particularly useful for radio interferometric imaging with Bayesian algorithms as the computational structure of the Bayesian algorithm has many similarities with machine learning methods, but the evaluation of the interferometer response requires high-performance codes. Thus, `JAXbind` allows us to utilize the computational advantages of the modern JAX machine learning frameworks for Bayesian radio interferometric imaging. On the other hand, the second part of the thesis contributes the `fast-resolve` algorithm. `fast-resolve` enables for the first time Bayesian imaging on large radio interferometric datasets, building on fast approximations of the radio interferometric imaging likelihood. `fast-resolve` opens up the possibility of achieving similar improvements in image quality to those previously achieved with Bayesian algorithms on small data sets (Arras et al., 2021a).

This thesis is centered around four articles (Roth et al., 2023a; Roth et al., 2023b; Roth et al., 2024b; Roth et al., 2024a) contributing to enhance the observational capabilities of current and next generation telescopes as briefly outlined above. Before introducing the developments in more detail, the following section highlights some exemplary science cases of modern optical and radio telescopes underlining the need for advanced image reconstruction algorithms as developed in this thesis. Of course, there are numerous other use cases that would also benefit from unlocking the full imaging capabilities of ground-based telescopes through advanced image reconstruction software.

1.1 Science cases

Science cases of modern optical and radio telescopes are extremely diverse, and this section cannot provide a comprehensive overview. However, it highlights a some scientific applications for which the algorithmic improvements presented in the later parts of this thesis are especially relevant.

1.1.1 Optical telescopes

Many exoplanets have been discovered in the last decades. This has led to a versatile science branch of studying the characteristics and formation mechanisms of exoplanets (Perryman, 2018). However, although thousands of exoplanets have been confirmed with indirect methods, only very few could be imaged directly, because direct imaging requires resolv-

ing the extremely high brightness contrasts between the host star and the exoplanet at a very small angular separation. Future telescopes with larger aperture diameters, such as the Extremely Large Telescope (ELT, Gilmozzi et al., 2007), open the possibility of directly imaging a growing number of exoplanets. Nevertheless, adaptive optics and advanced image post-processing algorithms are required to resolve high contrast ratios and achieve near diffraction-limited resolution. This thesis contributes such a post-processing algorithm outlined in chapter 5 and in the publication Roth et al., 2023a.

Imaging exoplanets is by far not the only science case of the ELT requiring near diffraction-limited resolution. Another example is the analysis of the stellar population in a galaxy, which can reveal many aspects of the galaxy's history (van den Bergh, 1975; Renzini, 2006). So far, resolving individual stars was only possible in the Milky Way and nearby galaxies. With larger telescopes such as the ELT, this can also be done for galaxies much further away if atmospheric perturbations degrading the image resolution can be suppressed. Many other science cases exist that require high-resolution images, for example, outlined in Hook, 2009 and Markus, 2010.

1.1.2 Radio interferometers

Radio astronomical observations are possible over a very large frequency range from 10 MHz to nearly 1 THz, and for this reason also the possible science cases are incredibly diverse. MeerKAT (Jonas et al., 2016) is currently the flagship telescope for frequencies around 1 GHz. With currently 64 antennas, MeerKAT has a large collecting area and high sensitivity. Many science cases profit from the high sensitivity, and an overview is given in Booth et al., 2009. One example is the MIGHTEE survey (Jarvis et al., 2016), which observes in a few deep fields the galaxy formation and evolution to redshifts up to $z \approx 7$. As discussed in Heywood et al., 2022, direction-dependent effects in the antenna gains significantly affect the observations and are a limiting factor if not reconstructed. Chapter 6, also published as Roth et al., 2023b, contributes such a reconstruction algorithm.

The high number of antennas of the MeerKAT telescopes not only enables a high sensitivity but also leads to large data volumes, especially for surveys. The large data sizes of MeerKAT observations limit the applicability of novel advanced imaging algorithms as they tend to have higher computational costs compared to the standard method named `CLEAN`. With Roth et al., 2024b; Roth et al., 2024a, this thesis contributes to reducing the computational cost of advanced imaging algorithms. Specifically, Roth et al., 2024a enables Bayesian reconstructions of large MeerKAT datasets at moderate computational costs and is a significant improvement compared to previous Bayesian reconstruction methods, opening new data regimes for Bayesian radio interferometric imaging.

1.2 Overview

The following chapters of this thesis are structured as follows. Chapter 2 gives an overview of the propagation of electromagnetic waves in the atmosphere and introduces the effects perturbing ground-based observations. The measurement principles of optical and radio telescopes are outlined in chapter 3 together with the impact of atmospheric fluctuations.

In chapter 4, I briefly introduce Bayesian inference as the algorithms presented in subsequent chapters all build on Bayesian inference. Originally published as an article (Roth et al., 2023a), chapter 5 introduces a high contrast imaging algorithm for optical telescopes. Chapter 6, published as Roth et al. (2023b), introduces a radio interferometric imaging algorithm with direction-dependent antenna calibration. Chapter 7 shifts the perspective from atmospheric calibration to the computational challenges of imaging with large interferometers. Based on the publication Roth et al. (2024b), chapter 8 introduces the `JAXbind` package for binding custom code into the JAX framework. Building on the JAX framework and `resolve`, chapter 9 introduces the `fast-resolve` method and is also published as a preprint (Roth et al., 2024a) which is currently under review. Chapter 10 gives a brief over other contributions of this thesis. Chapter 11 summarizes this thesis and gives an outlook on potential future developments.

Chapter 2

Atmospheric effects

This chapter introduces the propagation of electromagnetic waves in the Earth's atmosphere. Building on Saha, 2007 focusing on optical astronomy and Thompson et al., 2017; Wilson et al., 2009; Burke et al., 2019 focusing on radio astronomy, this chapter gives a unified description of atmospheric effects relevant to ground-based astronomy.

2.1 Atmospheric composition

The atmosphere is a mixture of many gases. In dry air at ground level, the main components are nitrogen (78%), oxygen (21%), and argon (1%). However, astronomical seeing is also affected by elements with much lower concentrations. Most notable are water vapor ($0 - 7\%$), ozone ($0 - 0.01\%$), and carbon dioxide ($0.01 - 0.1\%$). The atmosphere exhibits a complicated vertical structure in terms of temperature and chemical composition. The vertical structure is often categorized into five layers. The lowest layer is the troposphere, where the temperature drops with altitude. At a height of about 12 km, the stratosphere begins, and temperature increases as a function of altitude. The stratosphere is astronomically relevant as it contains the ozone layer that absorbs UV radiation. Above the stratosphere, there is the mesosphere (50 km to 85 km), and temperatures drop again with height. With the next layer, the thermosphere extending to 600 km, the temperatures rise again and reach 1600° C. The thermosphere is highly relevant for radio astronomy as the free electrons of ionized gas deflect radio waves. Above the thermosphere comes the exosphere, which is gradually transitioning into space. The layers of the atmosphere containing free electrons, thus mostly the thermosphere, are also called ionosphere.

The mixture of many different gases, together with the complex temperature distribution, makes the atmosphere opaque in most parts of the electromagnetic spectrum, making ground-based observations impossible. In figure 2.1, the opacity of the atmosphere is plotted as a function of the wavelength. The atmosphere is fully opaque for the entire spectrum from gamma rays to ultraviolet radiation, and all radiation gets absorbed mostly by ozone, oxygen, and nitrogen. At optical and near-infrared wavelengths ($0.3 - 1.1\,\mu$m), the atmosphere is transparent unless the sky is covered by clouds. For infrared radiation, there are some windows through the atmosphere, in particular in the range between $8 - 14\,\mu$m. Nevertheless, multiple vibrational transitions in water vapor and carbon dioxide absorb infrared

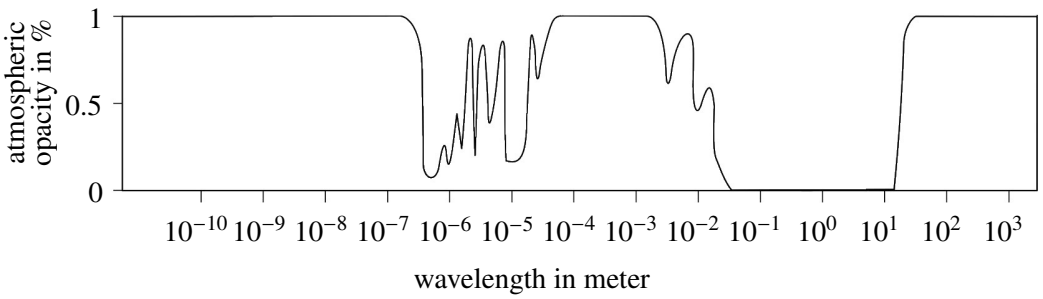

Figure 2.1: Atmospheric opacity as a function of wavelength. The atmosphere is opaque in most parts of the electrodynamic spectrum. All radiation with wavelengths shorter than the optical-near-infrared window is absorbed. At the low frequency end of the spectrum, the radio window permits ground-based observations over a large frequency range. Adapted from Wikimedia Commons, 2023.

radiation. For even lower frequencies with sub-millimeter to millimeter wavelengths, rotational transitions in water vapor absorb most of the radiation. Observations can still be made at places with dry atmosphere, such as the South Pole or the Atacama desert in Chile. At radio wavelengths starting in the millimeter range, there is a broad atmospheric window with only minimal absorption. At the low-frequency end of the radio window, the ionosphere of the atmosphere becomes opaque to radio waves starting from around 30 m and even longer wavelengths.

Although the atmospheric absorption is very low in the optical and radio windows, the atmosphere still significantly perturbs observations even well inside these windows. Even with the naked eye, atmospheric perturbations are visible as twinkling of stars. Without advanced correction techniques, these perturbations severely limit the observational capabilities of large optical and radio telescopes. The following section outlines the physical theory of electromagnetic waves and their propagation to derive a more fundamental description of atmospheric processes.

2.2 Electromagnetic waves

Several theoretical models exist to describe the behavior of light or electromagnetic radiation. The oldest and simplest theory is geometrical optics. Geometrical optics describes the propagation of light in terms of ray tracing and covers effects such as reflection or refraction of light. Building on Maxwell's equations, wave optics describes light as an electromagnetic wave and can additionally explain the phenomena of diffraction and interference. Quantum optics provides the most fundamental description of light in terms of photons and their quantum nature. For astronomy, all three theories are relevant. The spectral absorption lines of water, carbon dioxide, ozone, and other molecules mentioned earlier are a quantum phenomenon. Inside the atmospheric windows, where there are no absorption lines, light propagation is most efficiently described in terms of Maxwell's electrodynamic the-

ory. Therefore we will analyze in the following the behavior of electric ($\boldsymbol{E}$) and magnetic ($\boldsymbol{B}$) fields in the atmosphere.

In the atmosphere it can be assumed that within an macroscopic volume there is no net free charge such than $\varrho_f = 0$. With this assumption Maxwell's equation in CGS units are given by:

$$\begin{aligned} \nabla \cdot \boldsymbol{D} &= 0 & \nabla \times \boldsymbol{E} &= -\frac{1}{c}\dot{\boldsymbol{B}} \\ \nabla \cdot \boldsymbol{B} &= 0 & \nabla \times \boldsymbol{H} &= -\frac{4\pi}{c}\boldsymbol{J} + \frac{1}{c}\dot{\boldsymbol{D}}, \end{aligned} \tag{2.1}$$

with $\boldsymbol{D}$ the electric displacement, and $\boldsymbol{H}$ the magnetic H-field, and c the speed to light. Under the assumption that the atmosphere is a linear medium, such that $\boldsymbol{D} = \epsilon\boldsymbol{E}$ and $\boldsymbol{H} = \frac{1}{\mu}\boldsymbol{B}$, and using Ohms law $\boldsymbol{J} = \sigma\boldsymbol{E}$, the wave equations can be derived

$$\nabla^2\boldsymbol{B} = \frac{\epsilon\mu}{c^2}\ddot{\boldsymbol{B}} + \frac{4\pi\sigma\mu}{c^2}\dot{\boldsymbol{B}} \tag{2.2}$$

$$\nabla^2\boldsymbol{E} = \frac{\epsilon\mu}{c^2}\ddot{\boldsymbol{E}} + \frac{4\pi\sigma\mu}{c^2}\dot{\boldsymbol{E}}, \tag{2.3}$$

with σ being the electric conductivity, ϵ being the electric permittivity, and μ being the magnetic permeability. Both fields follow the same linear differential equation. The equations for $\boldsymbol{E}$ and $\boldsymbol{B}$ are decoupled and identical,[1] and therefore we will for simplicity only consider the electromagnetic field in the following, as the solutions for the magnetic field are completely analogous.

2.2.1 Waves in non-conducting media

Outside of the ionosphere, the atmosphere is non-conducting. In this section, we consider the propagation in the non-conducting part of the atmosphere. In the section 2.2.2, we will deal with propagation effects in the ionosphere. Thus, for now, $\sigma = 0$, and the equation for the electric field (and, of course, also for the magnetic field) simplifies to

$$\nabla^2\boldsymbol{E} - \frac{\epsilon\mu}{c^2}\ddot{\boldsymbol{E}} = 0. \tag{2.4}$$

We will not search for general solutions of the wave equation but instead investigate the behavior of plane wave solutions. For that, we insert a monochromatic plane wave traveling along the z-axis

$$\boldsymbol{E}(z,t) = \boldsymbol{E}_0 \mathrm{e}^{i(kz-\omega t)} \tag{2.5}$$

into the equation 2.4. From this, we get the following relation

$$k^2 = \frac{\epsilon\mu}{c^2}\omega^2 \tag{2.6}$$

[1] In Maxwell's equations, the electric and magnetic fields are coupled. The fields become decoupled only in the wave equations, which can be derived from Maxwell's equations. For this reason, $\boldsymbol{E}$ and $\boldsymbol{B}$ fields solving the Maxwell equations necessarily also solve the wave equations, but the reverse does not hold. Specifically, it can be derived from the Maxwell equations that the $\boldsymbol{E}$ and $\boldsymbol{B}$ fields of an electromagnetic wave must be orthogonal to each other, which is not encoded in the wave equations.

for the wave number k and the angular frequency ω. Furthermore, this implies that points of constant phase move with a velocity of

$$v = \frac{\omega}{k} = \frac{c}{\sqrt{\epsilon\mu}}. \tag{2.7}$$

The values of ϵ and μ depend on the material properties and, therefore, determine the wave's phase velocity. In some materials, the values of ϵ and μ depend on frequency, making the phase velocity frequency dependent. This leads to the phenomenon that a pulse containing a package of waves with multiple frequencies moves through the medium with a velocity different from the phase velocity. The propagation speed of such a pulse is named group velocity and can be shown to be to first order

$$v_g = \frac{d\omega}{dk}. \tag{2.8}$$

If ϵ and μ are frequency-independent, the group velocity equals the phase velocity. The ratio between the phase velocity v and the speed of light c is defined as the index of refraction

$$n = \frac{c}{v} = \sqrt{\epsilon\mu}. \tag{2.9}$$

In vacuum, the permittivity ϵ and the permeability μ are both equal to unity in CGS units, and therefore the refractive index equals unity as well. In the atmosphere, the refractive index slightly deviates from unity. The exact value of the refractive index in the atmosphere depends on temperature, water vapor content, and atmospheric pressure, among other factors. Computing the refractive index of the atmosphere from the first principle is quite involved as it requires a full quantum mechanical treatment of all gas molecules, their electronic transitions, and spectral lines. Instead, extensive studies have been made at on both radio and optical wavelengths to find empirical relations between the refractive index and atmospheric conditions. An overview is given by Rüeger, 2002. It was found that the empirical relation

$$n_{\text{radio}} = 1 + 10^{-6}\left(77.7\frac{p_D}{T} + 71.3\frac{p_V}{T} + 375463\frac{p_V}{T^2}\right) \tag{2.10}$$

describes to a good approximation the refractive index of radio waves in the atmosphere, with p_D and p_V being the partial pressures of dry air and water vapor in units of millibar and T the temperature in Kelvin. The last term in this relation scaling with $1/T^2$ is attributed to originate from infrared resonances of water molecules. At higher frequencies, these resonances do not affect the spectral index. Therefore, an approximate description of the atmospheric spectral index at optical wavelength is given by

$$n_{\text{optical}} = 1 + 10^{-6}\left(77.7\frac{p_D}{T} + 71.3\frac{p_V}{T}\right). \tag{2.11}$$

With these empirical relations, the refractivity of water vapor is at temperatures around 280° K around 20 times higher at radio wavelengths than at optical wavelengths.

Both empirical relations for the refractive index of the non-conducting atmosphere in the optical and the radio windows do not depend on the frequency of the wave. Also in reality,

the refractive index of the non-conducting atmosphere turns out to be almost frequency-independent within these atmospheric windows. For this reason, phase and group velocities are equal, and waves of different frequencies travel at the same speed. As the next section will show, this does not hold for the ionosphere, where the refractive index strongly depends on frequency.

2.2.2 Waves in conducting media

In the previous section, we considered wave propagation in a nonconducting medium. As outlined in section 2.1, parts of the atmosphere, called the ionosphere, are ionized and contain free electrons and ions. In these atmospheric layers, the assumption of $\sigma = 0$ does not hold, and therefore, the wave equation 2.2 does not simplify to equation 2.4.

As in the previous section, we do not attempt to find a general solution to the wave equation but instead analyze plane wave solutions of equation 2.2. By inserting a plane wave equation 2.5 into the general wave equation 2.2, we obtain the dispersion relation for ionized media:

$$k^2 = \frac{\mu\epsilon\omega^2}{c^2}\left(1 + i\frac{4\pi\sigma}{\omega\epsilon}\right). \tag{2.12}$$

In the following, we will relate the ionospheric electron density to the dispersion relation equation 2.12. Therefore, we study the equation of motion of electrons in a tenuous plasma accelerated by the electric field of a wave. Of course, the ions in the plasma are also accelerated, but due to their much higher mass, the resulting velocities are much lower, so that they can be neglected. We assume that the electrons to move freely without collisions and that we are in a thin medium such that $\epsilon \approx \mu \approx 1$. Under these assumptions, the equation of motion of a single electron accelerated by the electric field of a wave is given by

$$m_e\dot{\boldsymbol{v}} = -e\boldsymbol{E}_0\mathrm{e}^{-i\omega t}, \tag{2.13}$$

with m_e, e and v being the electron mass, charge, and velocity. Integrating the equation of motion, we obtain

$$\boldsymbol{v} = \frac{e}{im_e\omega}\boldsymbol{E}_0\mathrm{e}^{-i\omega t} = -i\frac{e}{m_e\omega}\boldsymbol{E}. \tag{2.14}$$

This leads to an electric current density

$$\boldsymbol{J} = -Ne\boldsymbol{v} = i\frac{Ne^2}{m_e\omega}\boldsymbol{E}, \tag{2.15}$$

with N denoting the electron density. By comparison with the material equation $\boldsymbol{J} = \sigma\boldsymbol{E}$ we conclude that the conductivity σ is given by

$$\sigma = i\frac{Ne^2}{m_e\omega}. \tag{2.16}$$

Inserting this expression together with the assumption $\epsilon \approx \mu \approx 1$ into the dispersion relation equation 2.12 we obtain

$$k^2 = \frac{\omega^2}{c^2}\left(1 - \frac{4\pi N e^2}{m_e \omega^2}\right). \tag{2.17}$$

Defining the plasma frequency ω_p as

$$\omega_p := \sqrt{\frac{4\pi N e^2}{m_e}}, \tag{2.18}$$

the dispersion relation simplifies to

$$k^2 = \frac{\omega^2}{c^2}\left(1 - \frac{\omega_p^2}{\omega^2}\right) \tag{2.19}$$

$$= \frac{\omega^2 - \omega_p^2}{c^2}. \tag{2.20}$$

Inserting into the equations for the phase (equation 2.7) and group velocity (equation 2.8), we obtain

$$v = \frac{c}{\sqrt{1 - \frac{\omega_p^2}{\omega^2}}}, \tag{2.21}$$

$$v_g = c\sqrt{1 - \frac{\omega_p^2}{\omega^2}}. \tag{2.22}$$

At frequencies significantly larger than the plasma frequency $\omega \gg \omega_p$, the phase and group velocities are both approximately equal to the speed of light in vacuum $v \approx v_g \approx c$. As the frequency ω approaches the plasma frequency ω_p, the phase velocity v becomes significantly larger than c, while the group velocity v_g goes to zero. The group velocity becomes zero at $\omega = \omega_p$, and the wave can no longer propagate through the plasma. Thus, the plasma frequency is a cutoff below which the plasma becomes opaque to electromagnetic radiation.

The electron density in the ionosphere is not constant, but depends on many factors, such as the time of the day and the solar activity. However, typically, the free electron density does not fall below $10^5 \frac{1}{\mathrm{cm}^3}$. Substituting this density into equation 2.18 gives a plasma frequency of $f_p = \frac{\omega_p}{2\pi} \approx 9\,\mathrm{MHz}$. For this reason, astronomical observations with ground-based telescopes are usually not possible at frequencies below about $f_p \approx 10\,\mathrm{MHz}$. Radio waves at slightly higher frequencies can pass through the ionosphere, but since the refractive index

$$n = \sqrt{1 - \frac{\omega_p^2}{\omega^2}} \tag{2.23}$$

significantly deviates from unity, they get refracted. At frequencies much higher than the plasma frequency, the refractive index is approximately one, and the ionosphere has only minor effects on the observation.

2.3 Refraction

The propagation effects of electromagnetic radiation outlined in section 2.2.1 and section 2.2.2 affect astronomical observations in many ways. The simplest effect is that radiation coming from space with a vacuum refractive index $n_{\text{vac}} = 1$ is refracted as it passes through the atmosphere with a refractive index different from unity. This leads to a change in the angle of arrival at the Earth's surface, and thus a change in the apparent position of the sources.

In a simple model depicted in figure 2.2, the atmosphere can be assumed to consist of plane parallel layers with different refractive indices, leading to a refraction of the wave at the boundary of each layer. Snell's law relates the angle of the incoming wave to the angle of the refracted wave traveling to the next layer. If α_0 is the angle of incidence at the top layer and α_1 the angle towards the normal at the next lower layer than Snell's gives the relation:

$$n_0 \sin(\alpha_0) = n_1 \sin(\alpha_1), \tag{2.24}$$

with n_0 and n_1 being the refractive indices in the layers 0 and 1. If the first layer is the top of the atmosphere, and the zeroth layer is the space, then $n_0 = 1$. Analogously, for the next layer the relation of the angles is:

$$n_1 \sin(\alpha_1) = n_2 \sin(\alpha_2). \tag{2.25}$$

Consequently, the incident angle of the m-th layer can be computed to be $n_m \sin(\alpha_m) = \sin(\alpha_0)$. Thus, in this simple model of plane parallel layers the angle at the bottom layer α_m, which determines the apparent position of the source, does not depend on the vertical structure of the atmosphere, but only on the refractive index at the ground and the original incident angle α_0. This is a consequence of the simplifying assumption of plane parallel layers in the atmosphere, which does not hold in general. For example, in the model presented in Smart, 1977 and depicted in figure 2.3 assuming spherical symmetric layers also the vertical structure of the refractive index distribution is also relevant. Especially for diffraction in high layers of the atmosphere, such as the ionosphere diffracting radio waves, a spherical model is much more accurate than plane parallel layers.

2.4 Dispersion

The refractive index of the ionosphere (equation 2.23) is smaller than 1 and increases with frequency towards unity. In contrast, the refractive index of the non-conducting part of the atmosphere is larger than 1 and is within the optical and radio window nearly frequency independent as seen in the empirical relations equation 2.10 and equation 2.11. The frequency dependence of the ionospheric refractive index causes radio pulses of different frequencies to travel at different speeds, resulting in different arrival times. The propagation time of the pulse is given by

$$t_p = \int_0^L \frac{1}{v_g} dl, \tag{2.26}$$

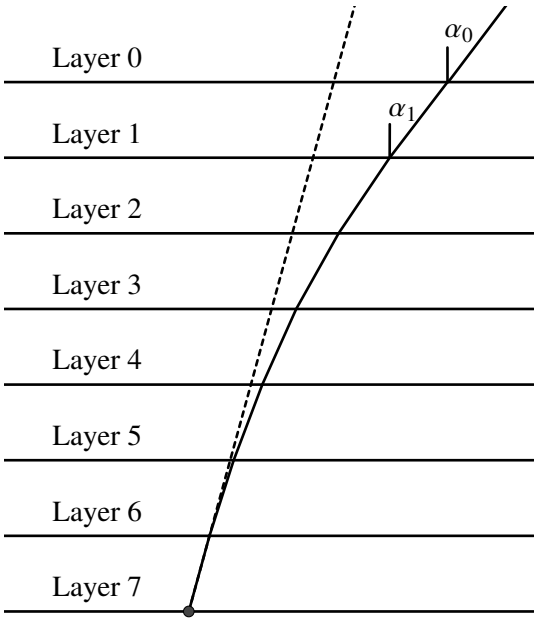

Figure 2.2: Visualization of a plane parallel atmospheric refraction model. The solid line shows the propagation path of the wave from the source to the observer. The waves are refracted at the intersection of each plane layer. The dashed line indicates the apparent direction of the source. The total deflection angle depends only on the difference between the first and last layers.

with L being the physical path length through the ionosphere L. Taylor expanding $\frac{1}{v_g}$ in $\frac{\omega_p}{\omega}$ gives

$$\frac{1}{v_g} \approx \frac{1}{c}\left(1 + \frac{1}{2}\left(\frac{\omega_p}{\omega}\right)^2\right), \tag{2.27}$$

which is an accurate approximation at frequencies significantly larger than the plasma frequency $\omega \gg \omega_p$. With this expansion, the propagation time of a pulse with frequency ω_1 can be expressed as

$$t_p = \int_0^L \frac{1}{v_g(\omega_1)}\, dl \tag{2.28}$$

$$= \frac{L}{c} + \frac{1}{c}\frac{e^2}{2\pi m_e^2}\frac{1}{\omega_1^2}\int_0^L N(l)\, dl, \tag{2.29}$$

where the definition of the plasma frequency was inserted in the step from equation 2.28 to equation 2.29. The integrated electron density is called dispersion measure DM = $\int_0^L N(l)\, dl$. With this definition, the arrival time difference between two pulses with frequencies ω_1 and ω_2 is given by

$$\Delta t_p = \frac{1}{c}\frac{e^2}{2\pi m_e^2}\left(\frac{1}{\omega_1^2} - \frac{1}{\omega_2^2}\right)\mathrm{DM}. \tag{2.30}$$

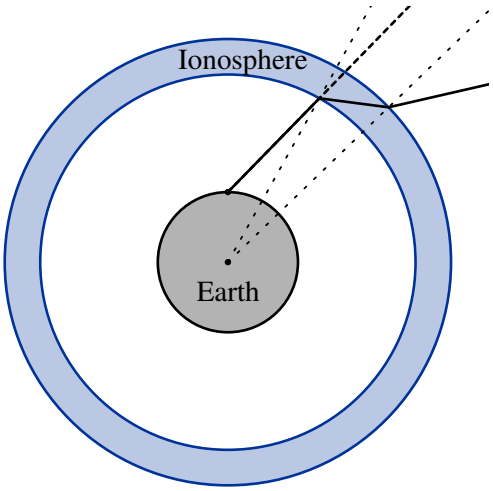

Figure 2.3: Visualization of a spherical atmospheric refraction model. The solid line shows the propagation path of the wave from the source to the observer. The waves are refracted at the intersection of each layer. The dashed line indicates the apparent direction of the source. The thin dotted lines indicate the normal to the layer boundaries. To make the deviation from the plane parallel layer more visible, the Earth and ionosphere are not displayed in corrected proportions. As the ionospheric's refractive index is smaller than one, the wave is refracted away from the normal.

To summarize the equation, pulse travel times delay scales with the inverse of the frequency squared and consequently the high frequencies arrive first.

For time-independent sources, the dispersion of the radiation does not affect astronomical observation. However, for rapidly time-variable sources, such as pulsars, the dispersion leads to the effect that the high-frequency components of the radio pulse reach the telescope before the low-frequency components. Thereby the observations of pulsars are not only affected by the atmospheric dispersion, but also by dispersion in interstellar medium, which can give indications on the distance of the pulsar or the density of the interstellar medium. The algorithms developed in this thesis are primarily targeted to time-independent sources, and for this reason do not consider dispersion.

2.5 Faraday rotation

In the equation of motion for the electrons (equation 2.13) in the derivation of the plasma refractive index, we have assumed that there is no external magnetic field. Nevertheless, due to the magnetic field of the earth, this assumption does not strictly hold in the ionosphere. In the presence of an external magnetic field $\boldsymbol{B}$, the Lorentz force $-\frac{e}{c}\boldsymbol{v} \times \boldsymbol{B}$ acts on the moving electrons in the plasma, such that the equation of motion becomes

$$m_e \dot{\boldsymbol{v}} = -e\left(\boldsymbol{E} + \frac{1}{c}\boldsymbol{v} \times \boldsymbol{B}\right), \tag{2.31}$$

with $\boldsymbol{E}$ being the electric field of the wave. By solving the equation of motion similar to the calculation in section 2.2.2, it can be shown that the external magnetic field leads to a difference in the refractive index for left and right circular magnetic fields. An explicit calculation can, for example, be found in Wilson et al., 2009, with the result that the refractive index is

$$n_{\pm} = \sqrt{1 - \frac{\omega_p}{\omega(\omega \pm \omega_c)}}, \tag{2.32}$$

with $\pm$ standing for right and left circular polarization, ω the frequencies of the wave, ω_p the plasma frequency, and ω_c the gyro or cyclotron frequency defined as

$$\omega_c = \frac{e}{mc} B_{\parallel}. \tag{2.33}$$

Thereby, $B_{\parallel}$ is the strength of the magnetic field along the line of sight or propagation path of the wave. Thus, the refractive index of the plasma is lower for left circular polarized waves than for right circular polarized waves if the magnetic field points along the propagation direction.

When studying only the total intensity of astrophysical sources, as in the example applications of the algorithms proposed in the later chapters, the differences between the refractive index for right and left circular polarization does not matter. Nevertheless, this does matter for polarization imaging. A linear polarized wave can be decomposed into right and left circular polarized waves with equal amplitude and a phase difference determined by the linear polarization direction. When this waves travels through a plasma with a magnetic field component parallel (or antiparallel) to the propagation direction the left and right hand circular components will propagate at different speeds leading to a change in the phase difference and consequently a rotation of the linear polarization direction. This phenomenon is called Faraday rotation. With a series expansion of equation 2.32 and by inserting the definitions of the plasma and gyro frequencies, the rotation angle of the linear polarization direction can be shown to be

$$\Delta\psi = \frac{2\pi e^3}{m^2 c} \frac{1}{\omega^2} \int_0^L B_{\parallel}(l) N(l)\, dl, \tag{2.34}$$

where the integral goes along the path through the plasma. Summarizing equation 2.34, the rotation angle is proportional to the integral of the parallel magnetic field component times the electron density and scales with the inverse of the frequency squared.

Similar to dispersion, Faraday rotation is induced not only by the ionosphere but also by the plasma and magnetic fields in the interstellar medium. Therefore, information about the galactic magnetic fields and electron density can be obtained by studying the rotation measures of polarized sources. Extensive studies of rotation measures have been done, for example by Hutschenreuter et al., 2023. However, the imaging algorithms of this thesis focus on total intensity imaging, and a calibration of the atmospheric Faraday rotation is therefore not required.

2.6 Turbulence

Meteorological processes such as evaporation, condensation, or crystallization constantly change the water vapor content and exchange heat with the surrounding atmosphere, leading to local variations in the refractive index. Similarly, the heating of the sun and the thermal exchange of the atmosphere with the ground or the sea constantly changes the air's temperature and refractive index. The air masses with a different refractive indices are not stationary, but are distributed by winds. The atmospheric air flows are often at least partially turbulent, causing turbulent changes in the refractive index. The situation in the ionosphere is similar in the sense that the electron density, and therefore the refractive index , is not homogeneous. The electron density is strongly influenced by solar activity, but also other factors, such as volcanic activity (Roberts et al., 1982; Molina et al., 2023) and earthquakes (He et al., 2022) also affect the ionosphere.

To summarize, the atmosphere is always affected by turbulent flows and other perturbations leading to a constant partially chaotic change in the refractive index distribution. Observations of all ground-based telescopes are affected by these refractive index fluctuations. Compared to the previously discussed effects of a shift in the apparent position of sources, a frequency-dependent propagation delay, and a frequency-dependent rotation in the polarization direction, refractive index fluctuations are the most complicated effect in the sense that it is the most difficult to correct. The following chapter will outline the exact impact of refractive index fluctuations on observations with optical and radio telescopes.

In 1941 Kolmogorov developed a theory of turbulence (Kolmogorov, 1991 (translation)), describing turbulence as the transfer of kinetic energy from large scales to successively smaller scales until the Kolmogorov scale is reached, at which the fluid becomes viscous and the kinetic energy is dissipated as heat. Tatarski et al., 1961; Tatarski, 1971 applied Kolmogorov's theory of turbulence to atmospheric perturbations. In this model, the refractive index structure function in a turbulent layer is given by

$$D_n(d) = \left\langle (n(x) - n(x-d))^2 \right\rangle \tag{2.35}$$

$$= C_n^2 d^{\frac{2}{3}}, \tag{2.36}$$

with d being the distance to an arbitrary location x, and C_n the strength of the turbulence. With further assumptions, such as that there is only a single turbulent layer and that the waves propagate through the turbulence without bending or refraction, a structure function for the phase of the wave can be derived. Within these assumptions, the amplitude of the wave is not affected. However, the assumptions of these atmospheric models are often only partially satisfied. For example, water vapor, strongly influencing the refractive index, is often poorly mixed with the air in the atmosphere and might, therefore, not accurately trace the turbulence of the air. Also, ionospheric perturbations might not follow Kolmogorov's turbulence model. Furthermore, especially for ionospheric perturbations, the waves will develop amplitude fluctuations in the propagation from the turbulent layer towards the ground.

Correcting for atmospheric turbulence is a major challenge in ground-based astronomy. The chapters 5 and 6 of this thesis contribute novel algorithms to reconstruct sky images from optical and radio telescopes jointly with the atmospheric perturbations. Before introducing these algorithms, chapter 3 outlines the measurement principles of optical and

radio telescopes and how they are affected by atmospheric turbulence. Furthermore, chapter 4 describes the basics of Bayesian inference on which the reconstruction algorithms are based.

Chapter 3

Ground-based telescopes

This chapter introduces the measurement equations of optical and radio telescopes and outlines how they are affected by atmospheric turbulence. The description of optical telescopes is based on Saha, 2007. The description of radio telescopes follows Smirnov, 2011a and Thompson et al., 2017.

3.1 Optical telescopes

A wide range of designs for optical telescopes exist. We will not discuss any of them in detail, but will analyze the image formation and the impact of the turbulent atmosphere from a more abstract mathematical point of view. For this purpose, we describe the telescope as an abstract linear optical system. For the mathematical description, we introduce two coordinate systems for the sky and the image. The first coordinate system are the Cartesian coordinates in a tangential plane centered in the observed field onto which the spherical sky is projected. The Cartesian coordinates of sky locations we denote with $\boldsymbol{x}_0$. With $\boldsymbol{x}_1$, we denote the Cartesian coordinates in the telescope's focal plane. These Cartesian coordinate systems are parallel to each other, such that by choosing proper units, they become basically equivalent. With $E_0(\boldsymbol{x}_0)$ we denote the electric field of the object at the location $\boldsymbol{x}_0$ in the sky plane. Analogously, we denote the electric field of the image at the position $\boldsymbol{x}_1$ in image plane with $E_1(\boldsymbol{x}_1)$. With these definitions, the response of any linear optical system relating the sky and the image electric fields can be written in the form

$$E_1(\boldsymbol{x}_1) = \int K(\boldsymbol{x}_0, \boldsymbol{x}_1) E_0(\boldsymbol{x}_0)\, d\boldsymbol{x}_0, \tag{3.1}$$

with $K(\boldsymbol{x}_0, \boldsymbol{x}_1)$ being the transmission function of the optical system. Typically, optical telescopes are designed such that their transmission function depends only on the difference between $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$, such that $K(\boldsymbol{x}_0, \boldsymbol{x}_1) = K(\boldsymbol{x}_0 - \boldsymbol{x}_1)$. With this assumption, equation 3.1 turns into a convolution of the object's electric field with the telescope transmission function.

For optical wavelengths, only the intensity of the image $I(\boldsymbol{x}_1) = \left\langle E_1(\boldsymbol{x}_1) E_1^*(\boldsymbol{x}_1) \right\rangle$ and not the electric field itself can be measured. Following equation 3.1, the intensity of the image

is given by

$$I(\boldsymbol{x}_1) = \int\int K(\boldsymbol{x}_0 - \boldsymbol{x}_1) K^*(\boldsymbol{x}'_0 - \boldsymbol{x}_1) \left\langle E_0(\boldsymbol{x}_0) E_0^*(\boldsymbol{x}'_0) \right\rangle d\boldsymbol{x}_0 d\boldsymbol{x}'_0 \tag{3.2}$$

$$= \int |K(\boldsymbol{x}_0 - \boldsymbol{x}_1)|^2 \, O(\boldsymbol{x}_0) \, d\boldsymbol{x}_0 \tag{3.3}$$

$$= |K(\boldsymbol{x}_0 - \boldsymbol{x}_1)|^2 * O(\boldsymbol{x}_0), \tag{3.4}$$

assuming in the step from equation 3.2 to equation 3.3 that the object electric field is incoherent such that $\left\langle E_0(\boldsymbol{x}_0) E_0^*(\boldsymbol{x}'_0) \right\rangle = \delta(\boldsymbol{x}_0 - \boldsymbol{x}'_0) O(\boldsymbol{x}_0)$, with $O(\boldsymbol{x}_0)$ being the intensity of the sky. Thus, summarizing equation 3.4, the intensity in the image plane of the telescope is given by the sky intensity convolved with the modulus squared of the transmission function, also called point spread function (PSF)

$$\mathrm{PSF}(\boldsymbol{x}_0 - \boldsymbol{x}_1) = |K(\boldsymbol{x}_0 - \boldsymbol{x}_1)|^2 . \tag{3.5}$$

The exact form of the transmission function, and therefore the PSF, depends on the technical details of the optical setup of the telescope. However, regardless of the technical design, the diffraction pattern created by the telescope pupil always contributes to the overall transmission function. For high-quality optical telescopes aiming at diffraction-limited imaging, the transmission function is well described by the diffraction pattern, and other effects in the telescope optics do not significantly degrade the transmission function. Therefore, the transmission function is given by the Fraunhofer diffraction pattern of the pupil

$$K(\boldsymbol{x}_0 - \boldsymbol{x}_1) = \int P(\boldsymbol{k}) \mathrm{e}^{-i2\pi(\boldsymbol{x}_0 - \boldsymbol{x}_1)\cdot\boldsymbol{k}/\lambda} \, d\boldsymbol{k}, \tag{3.6}$$

with λ being the wavelength and $P(\boldsymbol{k})$ being the pupil function defined as

$$P(\boldsymbol{k}) = \begin{cases} U(\boldsymbol{k}) & \text{inside the pupil,} \\ 0 & \text{outside the pupil.} \end{cases} \tag{3.7}$$

For a perfect telescope without external perturbations $U(\boldsymbol{k}) = 1$.

The resolution of the telescope is determined by the width of the PSF.[1] For a spherical aperture, the diffraction pattern is given by the Airy function. The first minimum of the Airy function occurs at an angle

$$\theta \approx 1.22 \frac{\lambda}{d}, \tag{3.8}$$

with λ being the imaging wavelength and d the diameter of the aperture.

If there are external perturbations, for example, from the atmosphere, $U(k)$ will deviate significantly from 1, distorting the pupil's diffraction pattern. As I will discuss in the next subsection, this can be the main limitation for the resolution of the telescope.

[1] Processing the images with deconvolution algorithms such as the algorithms presented in the chapters 5, 6, and 9 allows to achieve higher resolution than the width of the PSF in regions with good signal-to-noise ratio. Nevertheless, the width of the PSF is a good proxy for the expected resolution.

3.1.1 Impact of turbulence on optical telescopes

Phase and amplitude distortions induced by atmospheric turbulence can be encoded in the pupil function. Due to temporal variations of the atmosphere, the pupil function $U(\boldsymbol{k}, t)$, and consequently the telescope PSF becomes time-dependent. For a given time t, equation 3.7 determines the instantaneous PSF of the telescope at that timestamp. The image obtained in a long exposure is not be determined by the instantaneous PSF, but by the average PSF during the exposure. Incorporating the average PSF into the image formation equation 3.4 leads to

$$I_1(\boldsymbol{x}_1) = \langle I_1(\boldsymbol{x}_1, t)\rangle \tag{3.9}$$

$$= \langle \mathrm{PSF}(\boldsymbol{x}_0 - \boldsymbol{x}_1, t)\rangle * O(\boldsymbol{x}_0), \tag{3.10}$$

with $\langle \mathrm{PSF}(\boldsymbol{x}_0 - \boldsymbol{x}_1, t)\rangle$ being the time-averaged PSF and assuming that the object does not dependent on time during the exposure.

The effect of atmospheric turbulence on the retrieved image depends strongly on the meteorological conditions and the telescope used for the observation. Typical values for the temporal coherence time of $U(\boldsymbol{k}, t)$ are around 20 msec at optical wavelengths and 100 msec in the near-infrared. The spatial phase coherence length at optical wavelengths is typically around 10 cm, and the amplitude coherence length is usually significantly longer. According to equation 3.8 the resolution of a telescope with a 10 cm diameter is around 1 arcsec at a wavelength of 400 nm. This is the typical resolution limit for an optical telescope without using atmospheric correction techniques. Observations with optical telescopes with aperture diameters larger than 10 cm will be strongly affected by the atmospheric perturbations, and their resolution will not increase with aperture diameter as expected by equation 3.8, but will remain at approximately 1 arcsec.

The Hubble Space Telescope[2] (HST) and James Webb Space Telescopes[3] (JWST) have both a resolution around 0.1 arcsec. Achieving similar or better resolution with ground-based optical telescopes is only possible by correcting atmospheric perturbations. The most well-known atmospheric correction technique is adaptive optics. Adaptive optics measures the wavefront deformations and counteracts them by deforming mirrors. The mirror deformations create differences in the optical path length, that compensate for phase differences induced by the atmosphere. The adaptive optics system corrects the wavefront on millisecond time scales as this is the typical atmospheric temporal correlation. Adaptive optics can significantly improve the image resolution of large telescopes. For example, the SPHERE instrument (Beuzit et al., 2019) of the Very Large Telescope (VLT) archives with adaptive optics a resolution of around 0.02 arcsec.

Nevertheless, adaptive optics does not fully correct for atmospheric perturbations. Especially the dynamic range of the acquired images is reduced, limiting the detectability of fainter companions in the vicinity of bright sources. This, for example, severely limits the ability to image exoplanets as they are always close to their host star, which is much brighter

[2] HST reaches the highest resolution with the Wide Field Camera 3. Technical specifications and the exact resolution can be found in the handbook: `https://hst-docs.stsci.edu/wfc3ihb` (last visited 07-08-24).

[3] JWST achieves the highest resolution with the near infrared camera. Technical specifications and the exact resolution are listed here: `https://jwst-docs.stsci.edu/jwst-near-infrared-camera/nircam-observing-modes/nircam-imaging` (last visited 07-08-24).

than the planet itself, which only reflects the starlight. Fast cadence imaging was developed to further enhance the image contrast of ground-based telescopes. In fast cadence imaging, image sequences are acquired where the exposure times of the individual frames are on millisecond timescales such that the atmospheric perturbations are essentially static during the individual exposures. These frame sequences are then post-processed after the observation to retrieve a joint sky image with increased contrast. This thesis contributes a novel post-processing algorithm for such frame sequences in chapter 5.

3.2 Radio telescopes

There are two main classes of telescopes used in radio astronomy, single dish telescopes and interferometers. Single dish telescopes work similarly to optical telescopes, focusing the incoming radiation onto a receiver. Therefore, as for an optical telescope, the resolution of a single dish radio telescope is approximately determined by equation 3.8. Because radio waves have much longer wavelengths than optical light, the resulting resolution is significantly lower. For example, for a large single dish telescope with a diameter of 100 m and a frequency of 10 GHz the resolution according to equation 3.8 would be only 1.26 arcmin. While the shifts in the apparent positions of sources, dispersion, and Faraday rotation can be observed with single dish radio telescopes, atmospheric turbulence can be neglected due to the relatively low resolution.[4] Nevertheless, the correction of atmospheric turbulence is crucial for radio interferometers, archiving orders of magnitude higher resolution. The following subsections introduce the measurement principles of radio interferometers (section 3.2.1) and outline calibration approaches (section 3.2.2).

3.2.1 Radio interferometers

A radio interferometer is an array of multiple telescopes distributed over an area. Each individual telescope, also called an antenna, measures the phase and amplitude of the electric field at its location. The electric field at the telescopes is a superposition of many different sources. For simplicity, however, we assume for now that there is only a single source in the sky. Following the derivation of Smirnov, 2011a, we denote the electric field of the radio wave of the source with the vector

$$\boldsymbol{e} = \begin{pmatrix} e_x \\ e_y \end{pmatrix}. \tag{3.11}$$

The components of the vector correspond to the polarization modes of the wave. Thus, in a linear polarization basis, the first component is the phase and amplitude of the wave along some direction x, and the second component describes the wave along the orthogonal direction y. Alternatively $\boldsymbol{e}$ could also be expressed in a circular polarization basis. Under the assumption of a linear propagation from the source to the antennas, the electric field at

[4] As described in the main text, atmospheric turbulence perturbing the phase and amplitude of the radio wave can be neglected for single dish radio telescopes. Nevertheless, phase perturbations resulting from deformations of the telescope reflector under its own weight can have significant effects and are sometimes corrected by "active surfaces" as used for example by the Robert C. Byrd Green Bank Telescope (Lacasse, 1998).

antenna p can be written as

$$\boldsymbol{e}_p = \boldsymbol{J}_p \boldsymbol{e}, \tag{3.12}$$

with $\boldsymbol{J}_p$ being the Jones matrix encoding the propagation effects from the source to the antenna p. The electric fields measured by a pair of antennas are then correlated with each other by the correlator. For the pair of antennas p and q, the correlator computes the visibility

$$V_{pq} = 2\left\langle \boldsymbol{e}_p \boldsymbol{e}_q^\dagger \right\rangle, \tag{3.13}$$

with $\langle\rangle$ denoting component wise averaging over a small time interval and † standing for transposition and complex conjugation. The factor 2 in the computation is a convention in radio astronomy. Assuming that the Jones matrices are constant over the averaging time, the visibility V_{pq} can be rewritten as

$$V_{pq} = 2\boldsymbol{J}_p \left\langle \boldsymbol{e}\boldsymbol{e}^\dagger \right\rangle \boldsymbol{J}_q^\dagger \tag{3.14}$$

$$= \boldsymbol{J}_p \boldsymbol{B} \boldsymbol{J}_q^\dagger, \tag{3.15}$$

with $\boldsymbol{B} = 2\left\langle \boldsymbol{e}\boldsymbol{e}^\dagger \right\rangle$ being the brightness of the source. In a linear polarization basis the relation between the brightness matrix and the Stokes parameters can be shown to be

$$\boldsymbol{B} = 2\left\langle \boldsymbol{e}\boldsymbol{e}^\dagger \right\rangle = 2\begin{pmatrix} \langle e_x e_x^* \rangle & \langle e_x e_y^* \rangle \\ \langle e_y e_x^* \rangle & \langle e_y e_y^* \rangle \end{pmatrix} = \begin{pmatrix} I+Q & U+iV \\ U-iV & I-Q \end{pmatrix}. \tag{3.16}$$

While the absolute phases of the electric fields cancel out in equation 3.14, the phase difference between the antennas matters. As visualized in figure 3.1, the path length from the source to the individual antennas is different.

These path length differences create phase offsets between the electric fields at the antenna locations, imprinting on the visibilities. The path length difference ΔP_{pq} is given by

$$\Delta P_{pq} = \boldsymbol{D}_{pq} \cdot \boldsymbol{s}, \tag{3.17}$$

with $\boldsymbol{D}_{pq}$ being the vector from antenna p to antenna q and $\boldsymbol{s}$ pointing in the direction of the source and being normalized to unit length. This geometric path length difference leads to a phase difference in the electric fields at the antennas given by

$$\Delta\phi_{pq} = \frac{2\pi\Delta P_{pq}}{\lambda}, \tag{3.18}$$

with λ being the wavelength of the radio wave. In the idealized case that the radio waves travel undisturbed from the source to the antennas, the Jones matrices will only phase shift the radio wave according to the path length such that

$$V_{pq} = \boldsymbol{J}_p \boldsymbol{B} \boldsymbol{J}_q^\dagger \tag{3.19}$$

$$= \boldsymbol{B} \mathrm{e}^{-i\Delta\phi_{qp}} \tag{3.20}$$

$$= \boldsymbol{B} \mathrm{e}^{-i\frac{2\pi}{\lambda}\boldsymbol{D}_{pq}\cdot\boldsymbol{s}}. \tag{3.21}$$

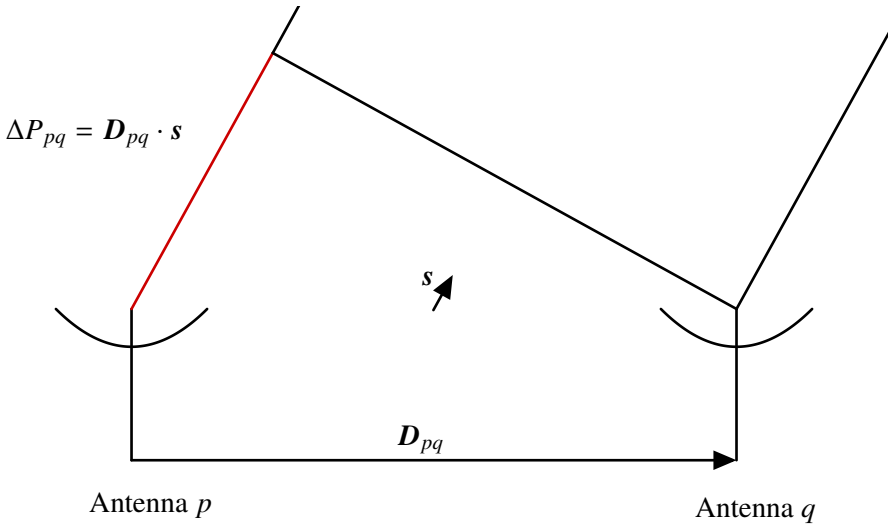

Figure 3.1: The red segment indicates the path length difference between the two antennas. $\boldsymbol{D}_{pq}$ denotes the baseline vector from antenna p to antenna q, and the unit vector $\boldsymbol{s}$ denotes the direction to the source. With these definitions, the path length difference, thus the length of the red segment is given by $\boldsymbol{D}_{pq} \cdot \boldsymbol{s}$.

In reality, the measurement of the electric fields at the antennas is affected by atmospheric effects (chapter 2) and the antenna electronics. The atmospheric and antenna effects can perturb the phase and amplitude of the measured electric field. For this reason, the Jones matrices contain for each antenna one unknown complex valued factor called antenna gain G. Therefore, the measurement equation of the visibility (equation 3.21) is modified to

$$V_{pq} = \boldsymbol{G}_p \boldsymbol{B} \boldsymbol{G}_q^* \mathrm{e}^{-i\frac{2\pi}{\lambda}\boldsymbol{D}_{pq}\cdot\boldsymbol{s}}. \tag{3.22}$$

So far, we have assumed that the sky consists of a single radio source. If we drop this assumption, the electric fields measured by the antennas are a superposition of the fields from many sources. Since the electric fields from different sources are incoherent, their intensities are additive. By integrating over all directions with possible source the measurement equation for the visibilities (equation 3.22) can be generalized to multiple sources

$$V_{pq} = \int \boldsymbol{G}_p(\boldsymbol{s}, t) \boldsymbol{B}(\boldsymbol{s}) \boldsymbol{G}_q^*(\boldsymbol{s}, t) \mathrm{e}^{-i\frac{2\pi}{\lambda}\boldsymbol{D}_{pq}\cdot\boldsymbol{s}} d\Omega. \tag{3.23}$$

Here, $\boldsymbol{G}_p(\boldsymbol{s}, t)$ is the gain of antenna p which can depend on time t and direction $\boldsymbol{s}$, and $\boldsymbol{B}(\boldsymbol{s})$ is the direction-dependent sky brightness. The integration goes over all possible directions of sources influencing the electric field measurement of the antenna, which is, in principle, half of the unit sphere (since the other half of the Celestial sphere is below the horizon).

Instead of the spherical coordinate system of equation 3.23, usually a Cartesian coordinate system is used in radio interferometry because it is easier to handle, and usually the antennas are only sensitive within a field of view much smaller than the full sky. By convention, the antenna locations and baseline coordinates are represented by the Cartesian

coordinates u, v, w oriented such that the w-axis points towards the center of the observation. In the Cartesian sky coordinate system, the l and m coordinates are aligned with the u and w axis of the antenna coordinates. The sky coordinate n is aligned with the w axis. Since l, m, n are parametrizing directions on the unit sphere, they are not independent but need to fulfill the relation

$$l^2 + m^2 + n^2 = 1. \tag{3.24}$$

Therefore, n can be computed from l and m via $n = \sqrt{1 - l^2 - m^2}$. In this Cartesian coordinate system, the phase factor is given by

$$\boldsymbol{D}_{pq} \cdot \boldsymbol{s} = u_{pq} l + v_{pq} m + w_{pq} \sqrt{1 - l^2 - m^2}. \tag{3.25}$$

By an additional convention, the phase offset corresponding to the observation center, thus the direction $\boldsymbol{s}_0 = (l = 0, m = 0, n = 1)$, is usually removed from the visibilities during correlation. Thus, the remaining phase factor is given by

$$\boldsymbol{D}_{pq} \cdot \tilde{\boldsymbol{s}} = u_{pq} l + v_{pq} m + w_{pq} (\sqrt{1 - l^2 - m^2} - 1), \tag{3.26}$$

with $\tilde{\boldsymbol{s}} = \boldsymbol{s} - \boldsymbol{s}_0$. Following all these conventions, the relation between the visibilities, the sky brightness and the antenna gains is given by

$$V_{pqt} = \int\int \boldsymbol{G}_p(l, m, t) \boldsymbol{B}(l, m) \boldsymbol{G}_q^\dagger(l, m, t) \mathrm{e}^{-i \frac{2\pi}{\lambda} (u_{pq} l + v_{pq} m + w_{pq} (\sqrt{1 - l^2 - m^2} - 1))} \frac{dl\, dm}{\sqrt{1 - l^2 - m^2}}. \tag{3.27}$$

with $\frac{1}{\sqrt{1 - l^2 - m^2}}$ being the Jacobian determinant originating from the coordinate transform from the spherical to Cartesian sky coordinates.

The radio interferometric measurement equation 3.27 is very similar to a Fourier transformation from the spatial coordinates l, m to the Fourier coordinates u, v, w. For small fields of view, such that $l, m \ll 1$, $\sqrt{1 - l^2 - m^2} \approx 1$ and all antenna gains being equal to one, the measurement equation actually becomes a Fourier transformation from l, m to u, v. Therefore, a radio interferometer measures the Fourier transform of the sky brightness modulated by the antenna gains and wide field effects where $\sqrt{1 - l^2 - m^2}$ cannot be approximated by one. Nevertheless, the interferometer does not measure at all l, m, w locations since it only has a finite number of antennas. For this reason, radio interferometric imaging essentially requires recovering the sky brightness $\boldsymbol{B}$ from the partially sampled Fourier transform. Mathematically, this means that the measurement equation 3.27 has no unique inverse. Therefore, any algorithm reconstructing the sky brightness from the visibility data needs to impose additional constraints.

A probabilistic approach to such reconstructions on which all algorithms of this thesis are based will be discussed in chapter 4. The computational challenges of imaging will be discussed in chapter 7. Subsequently, chapter 8 introduces a package allowing to bind high-performance code for evaluating equation 3.27 into the machine learning framework JAX, enabling radio interferometric imaging in a modern machine learning framework. Especially for radio interferometers with many antennas, as required for science cases such like

the one presented in chapter 1.1.2, radio interferometric imaging is computationally very demanding, in especially in combination with the statistical techniques introduced in chapter 4. Chapter 9 addresses the these computational challenges and introduces a new algorithm combining statistical accuracy with numerical efficiency, enabling accurate reconstructions of large datasets.

3.2.2 Calibration of radio interferometers

Radio interferometric imaging is further complicated since the measurement involves the unknown antenna gains G, which need to be determined for a reconstruction of the sky brightness. Some factors influencing the gains, such as the shape and size of the antenna reflectors, are known; other factors, such as atmospheric perturbations, are unknown. Therefore, the exact antenna gains need to be measured and cannot be computed from first principle. The gains are time-dependent, as many factors influencing them, such as the atmospheric perturbations, are time-dependent. For this reason, the gains need to be measured shortly before or during the actual science observation. Determining the antenna gain values from measurement data is called calibration.

This subsection briefly outlines strategies to calibrate the antenna gains. A more comprehensive overview of different calibration strategies is given by Smirnov, 2011b; Smirnov, 2011c. The classical approach to calibration is to alternately observe the science target and a calibrator source, of which the brightness distribution is assumed to be known. Furthermore, in the classical approach, the direction-dependence of the antenna gains is neglected $\boldsymbol{G}(l, m, t) = \boldsymbol{G}(t)$. Neglecting the direction-dependence allows to pull the gains out of the integral in equation 3.27 such that the measurement equation is given by

$$V_{pqt} = \boldsymbol{G}_p(t)\left(\int\int \boldsymbol{B}(l,m)\mathrm{e}^{-i\frac{2\pi}{\lambda}(u_{pq}l+v_{pq}m+w_{pq}(\sqrt{1-l^2-m^2}-1))}\,\frac{dl\,dm}{\sqrt{1-l^2-m^2}}\right)\boldsymbol{G}_q^{\dagger}(t). \quad (3.28)$$

The integral in equation 3.28 can be evaluated by substituting the assumed brightness distribution of the calibrator, allowing to fit the antenna gains to the calibration observation. Afterwards, approximate antenna gains for the science target can be determined by extrapolating the obtained gains from the calibrator to the time intervals of the science observation.

Because the antenna gains are extrapolated from the time intervals of the calibrator observation to the science observation and direction-dependent effects are neglected, the accuracy of the obtained gain solutions is limited. To increase the accuracy, the obtained antenna gains are often used only as a starting point for an initial image reconstruction step of the science target. This initial reconstruction is inserted in equation 3.28, and residual fluctuations in the antenna gains are fitted. This scheme, called self-calibration, then iterates between improving the current science target reconstruction and improving the antenna gain calibration. Some approaches, such as Arras et al., 2019b, also combine calibration and imaging in a joint optimization problem.

While self-calibration overcomes the limitations due to the temporal extrapolation of the antenna gain solution, it still neglects direction-dependent effects. Incorporating direction-dependent effects in the antenna gains makes a fast numerical evaluation of equation 3.27 significantly more challenging as the antenna gains can no longer be pulled out of the integral as in equation 3.28. Thus, in the presence of direction-dependent gains, the integral has

to be recomputed for every antenna pair and time step. Besides the computational aspects, incorporating direction-dependent gains is also challenging from a statistical point of view. Already with fixed gains, the radio interferometric imaging problem is underdetermined in the sense that the response cannot be inverted. This problem becomes even more challenging when the direction-dependent effects of the gains need to be reconstructed, adding many more a priori unknown parameters. For this reason, most radio interferometric calibration and imaging frameworks neglect direction-dependent effects of the antenna gains. Nevertheless, this can often be a limiting factor, as outlined in the science case in the introductory section 1.1.2.

This thesis contributes a direction-dependent calibration and imaging algorithm presented in chapter 6, building on the inference techniques presented in the next chapter to overcome the challenge of reconstructing an underdetermined problem.

Chapter 4

Inverse problems and Bayesian inference

Chapter 2 introduced the physics behind atmospheric electromagnetic wave propagation and the effects perturbing observations. The previous chapter introduced the measurement equation of optical and radio telescopes. Before developing image reconstruction methods for such telescopes in the following chapters 5 and 6, this chapter takes a more abstract perspective on inverse problems and Bayesian inference and introduces the fundamental concepts on which the algorithms in the following chapters build on. This chapter does not give a complete overview of Bayesian inference or inverse problems but only outlines some fundamental ideas relevant for the following chapters. For an extensive introduction, see for example, MacKay, 2003.

4.1 Inverse problems

We start with considering from a mathematical perspective the reconstruction of a quantity of interest, called signal s, from a set of measurement data d. The measurement instrument determines the relation between the signal s and the measurement data d. In chapter 3, we derived the measurement relations between the sky brightness and the telescope data for optical and radio telescopes. In this chapter, we assume a general measurement response mapping from the signal s to the data d and denote it with R. Thus, a perfect measurement without measurement errors can mathematically be written as

$$d = R(s). \tag{4.1}$$

In realistic measurements, there will always be some noise. To cover the effect of noise in the measurement, we can modify equation 4.1 to a more general measurement equation

$$d = R(s) + n, \tag{4.2}$$

with n representing some (additive) measurement noise. Although the actual noise value n is unknown, the statistic of n is usually known, at least for a well-characterized instrument.

While equation 4.2 defines how to simulate for a given signal s potential measurement data d, the inversion thus going from some data back to the underlying signal is often more challenging. In the idealized noise-free case of equation 4.1, the signal could be recovered from the data as

$$s = R^{-1}(d), \tag{4.3}$$

if the instrument response R is invertible. Nevertheless, when the response R is not invertible, or the noise is not negligible, a simple inversion to recover the signal is no longer possible. Non-invertible response functions are commonly encountered in physical measurements, such as the response of a radio interferometer discussed in the previous chapter. Thus, in the presence of significant noise and a potentially non-invertible response R, multiple potential signals s exist that are consistent with the data. This leads to the question of how to find compatible signals s and how to discriminate between them or weight them. This setting of a well-defined forward relation from the signal to the data without a unique inverse, is commonly called inverse problem.

4.2 Bayesian inference

Probability theory provides a natural way to reason about unknown quantities. The probabilistic answer to the inverse problem equation 4.2 is given by the probability of the signal conditioned to the measurement data $P(s|d)$. Even with significant noise and a non-invertible response, the conditional probability of the signal is conceptually well-defined. Bayes' theorem

$$P(s|d) = \frac{P(d|s)P(s)}{P(d)} \tag{4.4}$$

relates the conditional probability of the signal, called posterior, to the likelihood $P(s|d)$, the prior $P(s)$ and the evidence $P(d)$. With the measurement equation 4.2, the likelihood can easily be evaluated via $P(d|s) = P(n = d - R(s)|s)$. Thus, evaluating the likelihood essentially boils down to evaluating the instrument response. The prior $P(s)$ encodes knowledge and assumptions about the signal before conducting the measurement and is, therefore, very much problem specific. Specific choices of prior models will be discussed in the chapters 5 and 6. The evidence $P(d)$ acts as a normalization constant independent of the signal and does not need to be evaluated in many applications.

For example, to compute the posterior maximum, the evidence can be ignored, and only $P(d|s)P(s)$ needs to be maximized. Statistically more robust techniques, such as posterior mean $\langle s \rangle_{p(s|d)}$ estimates and many other summary statistics, require integration over the posterior distribution. Many Bayesian inference algorithms have been developed for that. The most widely used classes of algorithms are Markov Chain Monte-Carlo (MCMC, Brooks et al., 2011) and variational inference (VI, Blei et al., 2017) methods. While MCMC-based methods come with the theoretical guarantee of recovering the true distribution within the limit of infinite runtime, VI methods are usually faster and scale better with the number of dimensions of the posterior. The inference problems in this thesis all deal with the reconstruction of telescope images, where essentially every image pixel is a parameter of the

posterior. Therefore, the inference problems are all high dimensional and, for this reason, are all based on variational inference methods.

The basic idea of VI is to approximate the true underlying posterior with a simpler distribution. This approximation is found by minimizing the information theoretical distance between the true posterior probability distribution and the approximation. The information theoretical distance measurement is the Kullback-Leibler divergence (KL, Kullback et al., 1951), defined as

$$\mathrm{KL}(Q, P) = \int Q(x) \ln\left(\frac{Q(x)}{P(x)}\right) dx. \tag{4.5}$$

In variational inference, P is the true underlying distribution defined via Bayes theorem, and Q the approximating distribution.[1]

The variational inference algorithms used in the applications of this thesis are Metric Gaussian Variational Inference (MGVI, Knollmüller et al., 2019) and geometric Variational Inference (geoVI, Frank et al., 2021). Both variational inference algorithms ultimately approximate the true posterior distribution P with a Gaussian distribution Q, although in different coordinate systems. Since the approximation error is smaller when the posterior distribution does not deviate strongly from a Gaussian, we do not directly infer the distribution of s, the quantity of interest, but a reparametrized version of the signal s. This reparametrization ξ is constructed such that the prior on ξ is a Gaussian and $s = s(\xi)$ is distributed according to the desired prior distribution. This reparametrization was introduced in the MGVI context by Knollmüller et al., 2018, but was well known in the statistics community under other names before, such as inverse transform sampling (Devroye, 1986) or reparametrization trick (Kingma et al., 2013).

The Gaussian posterior approximation is optimized in MGVI and geoVI by minimizing the KL (equation 4.5) with respect to the mean of the Gaussian. The approximation is not explicitly optimized with respect to the covariance of the approximating Gaussian. Instead, the covariance is parameterized by the inverse Fisher Metric, also called the inverse metric tensor of the true distribution. The Fisher Metric of the posterior distribution $P(\xi|d)$ is given by

$$M(\xi) = \left\langle \frac{\partial H(\xi|d)}{\partial \xi} \frac{\partial H(\xi|d)}{\partial \xi}^{\dagger} \right\rangle_{P(d|\xi)}, \tag{4.6}$$

with H denoting the negative logarithm of P thus $H(\xi|d) = -\ln P(\xi|d)$. Via a short calculation, it can be verified that in case the posterior is a Gaussian distribution, the Fisher Metric is the inverse covariance. Thus, in the Gaussian case the MGVI and geoVI techniques can recover the posterior distribution without approximation error. For non-Gaussian posteriors, MGVI is affected by an approximation error, as the approximating distribution is Gaussian.

[1] The KL is not symmetric in its arguments. As derived in Leike et al. (2017), the correct choice would be that the first argument of the KL is the true distribution P and the approximation Q the second argument. However, in (almost) all variational inference algorithms, the arguments of the KL are swapped, because in the correct order, the evaluation and optimization of the KL involves an integration over the true distribution. If it were be possible to integrate over the true distribution, there would be no need to use variational inference in the fist place. Therefore, variational inference involves not only the approximation of the true posterior with a simpler distribution but also an approximation due to neglecting the asymmetry of the KL.

GeoVI can partially also cover non-Gaussian posteriors by constructing a coordinate transformation in which the posterior becomes nearly Gaussian and performing the Gaussian approximation in this new coordinate system. A detailed analysis of the use of the Fisher information metric as a covariance of the approximating distribution is given in the original publications Knollmüller et al., 2019 and Frank et al., 2021.

The following chapters show applications of Bayesian inference to reconstruct the sky brightness or atmospheric perturbations. Of course, for numerical computations, the sky or atmospheric perturbations are discretized into pixels, although in principle, these quantities are field-like objects and assign a specific value to each direction on the sky or point in time. Conceptually, the sky brightness and other field-like quantities can be thought of as infinite-dimensional vectors. Mathematically, this discretization is not necessary for the applicability of Bayes theorem. Information field theory (IFT; Enßlin et al., 2009; Enßlin, 2019) formalizes Bayesian inference on field-like quantities without discretization.

The software package NIFTy (Numerical Information Field theorY; Edenhofer et al., 2024a; Steininger et al., 2017; Selig et al., 2013b) implements tools for the inference of pixelated field-like quantities, such as an image of the sky. Utilizing the scalable variational inference algorithms described above, NIFTy allows the inference of high-resolution images with millions of pixels. Following the spirit of information field theory, NIFTy is constructed such that the inference models are (nearly) independent from the exact pixelation of the images. The software package `resolve` provides additional functionality to NIFTy tailored to radio interferometry.

To summarize this chapter, Bayesian inference is a conceptually clean way to solve inverse problems. The posterior distribution encodes all knowledge about the signal that can be obtained from the data, including uncertainty information. Via variational inference algorithms, this posterior distribution can be accessed with some approximations even in high dimensions, such as high-resolution images. The algorithms of the following chapters are based on the NIFTy and `resolve` frameworks.

Chapter 5

Fast-cadence high-contrast imaging with information field theory

I have published this chapter as first author in the Astronomical Journal as an article (Roth et al., 2023a). This project originated from a part of my master thesis. My master thesis already included the basic idea of modeling the atmospheric phase perturbations as a phase screen in the telescope aperture. Additionally, my master thesis already included a prototype implementation of this basic model. In my PhD thesis, I followed up on this work and refined the modeling of the detector, developed the phase jump mitigation technique, implemented the final version of the algorithm, executed the benchmarking of the final algorithm, and wrote this publication. Gianluca Li Causi and Vicenzo Testa contributed via many discussions about the SHARK-VIS instrument and high-contrast imaging in general. Gianluca Li Causi additionally executed the comparison methods in section 5.6.1. Philipp Arras contributed via general help with the NIFTy software package. Torsten Enßlin contributed with many discussions about Bayesian inference and modeling of the instrument. All coauthors read, commented on, and approved the text. The text and figures were slightly adapted for this thesis.

5.1 Abstract

Although many exoplanets have been indirectly detected over the last years, direct imaging of them with ground-based telescopes remains challenging. In the presence of atmospheric fluctuations, it is ambitious to resolve the high brightness contrasts at the small angular separation between the star and its potential partners. Post-processing of telescope images has become an essential tool to improve the resolvable contrast ratios. This chapter contributes a post-processing algorithm for fast-cadence imaging, which deconvolves sequences of telescopes images. The algorithm infers a Bayesian estimate of the astronomical object as well as the atmospheric optical path length, including its spatial and temporal structures. For this, we utilize physics-inspired models for the object, the atmosphere, and the telescope. The algorithm is computationally expensive but allows to resolve high contrast ratios despite short observation times and no field rotation. We test the performance of the algorithm with point-like companions synthetically injected into a real data set acquired with the SHARK-

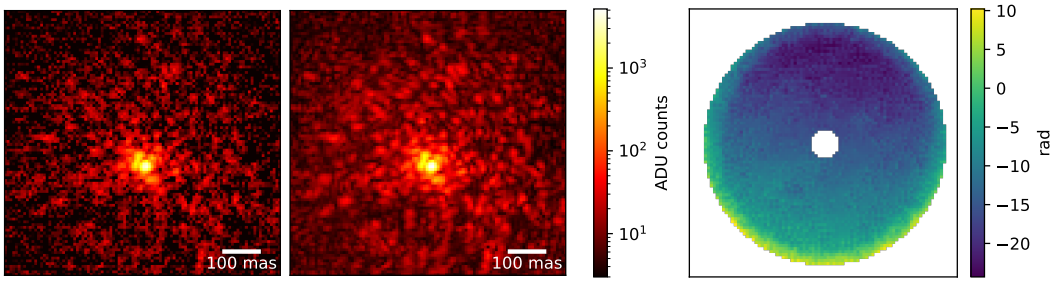

Figure 5.1: Left panel: Instantaneous PSF in the focal plane (rebinned to 100x100 pixels) of the 5.5 mag star Gliese 777, acquired with the SHARK-VIS Forerunner instrument at the LBT in a 1 ms exposure. The interference pattern shaped by the atmospheric fluctuations ("speckles") are well visible all around the star. Central panel: HCBI reconstruction of the PSF convolved star for the corresponding frame, shown with the same logarithmic colormap. Right panel: The reconstruction in the pupil plane of the optical path difference field for that frame of the Gliese 777 sequence, which determines the reconstructed PSF shown in the central panel.

VIS pathfinder instrument at the LBT telescope. Sources with brightness ratios down to $6 \cdot 10^{-4}$ to the star are detected at 185 mas separation with a short observation time of 0.6 s.

5.2 Introduction

The direct imaging of exoplanets is challenging due to the high brightness contrasts between a star and its potential exoplanets at angular separations of tiny fractions of an arcsecond, a context called "high-contrast imaging" (HCI; see Marois et al. (2006)). The atmospheric turbulence (section 2.6) that corrupts the incoming wavefront before entering the telescope, thus producing a distorted point spread function (PSF) at the focal plane, makes it difficult to resolve these high brightness contrasts. This corruption, especially pronounced for large telescopes, is nowadays mostly compensated by the extreme adaptive optics (ExAO) technology (Esposito et al., 2010). However, the correction is not perfect, and the post-AO PSF still shows high spatio-temporal variations, with typical times of a few ms and scales of a few mas (see Stangalini et al. (2017)). This results in a spread of the star light into a complex evolving pattern of spots, called "speckles", surrounding the star (see figure 5.1, left panel) and hiding the much fainter companions.

To further improve the achievable contrast ratios, data post-processing techniques have become an essential part of HCI. Mathematically, such data post-processing is a reconstruction problem in which a subtle signal shall be retrieved that is affected by three disturbances: i) a strong disturbing signal, i.e., the star's PSF with its photon noise, ii) its strong spatio-temporal variability, called "speckle noise", and iii) the signature and read-out noise of the detector. Many post-processing algorithms have been developed for HCI, like the classical Angular Differential Imaging (ADI; Marois et al., 2006), the Speckle Free ADI (SFADI; Li Causi et al., 2017), the Local Combination of Images (LOCI; Lafrenière et al., 2007),

the Stochastic Speckle Discrimination (SSD; Walter et al., 2019), the Principal Component Analysis ADI (PCA-ADI; Soummer et al., 2012, Amara et al., 2012), as well as the recent NMF data imputation method (Ren et al., 2020). All of them estimate the time-dependent PSF leveraging on the large field rotation in a sequence of pupil-stabilized images across a long temporal interval. This model is then used to subtract the central star contribution from the data.

Here we present a different approach, which works on fast-cadence frames sequences to jointly reconstruct the static true brightness distribution of the source (i.e. star and companions) and the temporal evolution of the PSF. We use an explicit forward data modeling that simulates the natural information flow from the signal to the detector (as recently done by Hope et al., 2022 and, in a different way, by the MAYO (Pairet et al., 2021) and REXPACO (Flasseur et al., 2021) methods). Thereby we exploit the statistical properties of the fast cadence PSFs whose exposure time is comparable to the speckles timescale. In particular, we do this by using information field theory (IFT; Enßlin (2019))

As outlined in chapter 4, IFT uses Bayesian probability theory and methods from statistical field theories to infer fields, e.g., source distribution and PSF, from noisy data, allowing to incorporate prior knowledge from physical constraints. In astronomy it has already been used in a number of contexts e.g. in radio astronomy (Arras et al. 2021a, Arras et al. 2019b, Arras et al. 2022), galactic dust tomography (Leike et al., 2020), or Faraday sky imaging (Hutschenreuter et al., 2020). This is the first time IFT is applied to HCI and hereafter we demonstrate its ability to reveal faint companions at one-tenth of an arcsecond separation from the host star in a very short acquisition time and without the need for field rotation. Since, at its core, our method rests on Bayesian inference, we name our algorithm High Contrast Bayesian Imaging (HCBI). Our code is open source and publicly available at: `https://gitlab.mpcdf.mpg.de/ift/public/hcbi`.

This chapter is organized as follows. In section 5.3 we briefly recapitulate the general concepts and methods of Bayesian inference and IFT introduced in the previous chapter and apply them to the HCI problem. In section 5.4 we develop a physically inspired prior model for short exposure imaging. Section 5.5 outlines the symmetries in the likelihood. Section 5.6 presents the results on real data, also in comparison with other methods. Finally section 5.7 discusses the performance and planned improvements.

5.3 Information field theory and its application to HCI

IFT is a theoretical framework for the reconstruction of fields, i.e. continuous quantities having a value at any location, from measurement data. In the HCI context, examples for fields are the static source brightness distribution, which we call *object*, and the time-variable PSF, which are both continuous in the sky or detector plane, respectively.

As fields are continuous quantities, they have an infinite number of degrees of freedom (d.o.f.) unless we restrict them to be of a parametric from. Thus, reconstructing them non parametrically from finite measurement data is impossible without additional information. IFT provides the theoretical foundation to incorporate such necessary pieces of additional information into field reconstruction algorithms building on Bayesian inference.

Many physical fields vary mostly smoothly in space and time as the physical processes

that govern them typically erase discontinuities quickly. Thus, values of the field at nearby locations are correlated, and below a certain distance, no significant differences are expected. This effectively reduces the infinite number of d.o.f. to a finite number, enabling the inference of the field from finite data sets. Hence the key concept for inferring infinite-dimensional quantities is their correlation. Should the correlation structure be unknown a priori, it can be inferred simultaneously with the field as described in section 5.4.2.3. On a computer, these fields are sampled into pixels or voxels, requiring that the sampling is fine enough to resolve all relevant structures. The open-source Python Bayesian field inference package NIFTy[1] (Arras et al., 2019a, Steininger et al. (2017), Selig et al., 2013a) is a numerical implementation of IFT, on which the method presented in this chapter as well as the other methods of this thesis built on.

From a mathematical point of view, HCI is described by the imaging equation derived in chapter 3 and valid within an isoplanatic patch of the sky, for which the data vector $d = (d_1, d_2, ...)$, i.e. the pixel values of all acquired frames, are a convolution of the object s with the PSF p to which some noise n is added

$$d_i(x) = (p * s)(x, t_i) + n_i, \tag{5.1}$$

where x is the detector grid coordinate and t_i the acquisition time of frame i. The PSF $p(x, t)$ includes the atmosphere and the optics response and depends on the coordinate x and time t, while the object $s(x)$ is assumed to be static in time.

Our method computes from d a Bayesian estimate of s and p, i.e. it infers the posterior probability distribution $P(s, p|d)$ of s and p given the data d, following Bayes' Theorem (equation 4.4):

$$P(s, p|d) = \frac{P(d|s, p)P(s, p)}{P(d)}. \tag{5.2}$$

Thereby the likelihood $P(d|s, p)$ is the probability of the data d for a given realization of the object s and the PSF p; $P(s, p)$ is the prior, which encodes all previous knowledge about s and p, like the correlation of s and p, and physical constraints, like the strict positivity of those fields. Finally, the term $P(d)$ is the evidence, which only acts as a normalization constant that is not used in our algorithm.

In our case, we usually deal with frames sequences containing thousands of images, each having in the order of 10^4 to 10^5 spatio-temporal pixels, so that s, p, and d are all very high dimensional quantities, and equation 5.2 contains too many variables to be computed directly. Thus a fast and scalable algorithm is needed to infer an approximate posterior $P(s, p|d)$ and get a solution for s and p.

5.3.1 Metric Gaussian Variational Inference

The simplest approximation of the posterior is to only compute its maximum so that the solution for s and p corresponds to the highest probability density. This method is commonly referred to as maximum a posteriori (MAP) probability estimation.
Several more accurate algorithms for approximating the posterior distribution exist. A suitable iterative algorithm for this work, capable of dealing with a very high dimensional parameters space, is Metric Gaussian Variational Inference (MGVI; Knollmüller et al., 2019

[1] `https://gitlab.mpcdf.mpg.de/ift/nifty`

and chapter 4). The concept of MGVI is to iteratively approximate the true posterior with a multi-dimensional Gaussian distribution. This Gaussian approximation is not directly performed on the parameters of interest, e.g. p and s, but on some model parameters Θ. The mappings $p(\Theta)$ and $s(\Theta)$ can be built according to physical considerations encoding prior knowledge on s and p. In the next sections, these mappings are outlined in detail. Furthermore, performing the Gaussian approximation in a latent parameters space instead of directly on p and s enables the algorithm to capture the non-Gaussian statistics of s and p. This more accurate approximation provides multiple advantages compared to the MAP estimator.

First, similar to MAP, MGVI has a near to linear scaling with the problem size, which makes it suitable for large problems with many free parameters.

Second, MGVI probes the phase space volume of the posterior, while MAP does not. This ensures that differences in the prior volume are more properly reflected in the inference with respect to MAP. This is of particular importance for our simultaneous object and PSF reconstruction since we are dealing with a degenerate inference problem in which several quantities are able to explain the same data feature. As only prior information (including its phase space volume) can break many of these degeneracies, its more proper treatment by MGVI is a clear advantage over MAP.

To summarize, IFT is a theoretical framework for Bayesian reconstructions of field-like quantities. MGVI is a specific numerical algorithm for obtaining an approximation of the actual posterior distribution for a given prior and likelihood model.

To apply MGVI to equation 5.2 we need a model of the prior $P(s, p)$ and a model of the likelihood $P(d|s, p)$, that, in order to be meaningful, need to be derived using physically motivated arguments, as we are going to describe in section 5.4 and section 5.5 hereafter.

5.4 Prior model

The MGVI algorithm requires the object prior and the measurement likelihood in the form of a generative model to draw realizations of $s(\Theta)$ and $p(\Theta)$ out of their probability distributions, starting from standard Gaussian distributed random variables Θ. In the following, we build up such generative models for the object, the atmosphere, the telescope, and the measurement instrument by following physical considerations.

5.4.1 Object model

Bayes' Theorem quantifies how to update the prior knowledge with newly acquired data. Therefore our prior probability model should include all knowledge about the imaged object before evaluating the actual image data.

This means that the best generative model for the prior distribution is tailored to the source under study and will be different if our target is e.g. a star with a companion, a young stellar object with an accretion disk, a resolved object, or something else.

As in this work we focus our investigation on the detection of exoplanets or low mass stellar companions next to a central star, we need a point source model for both the star

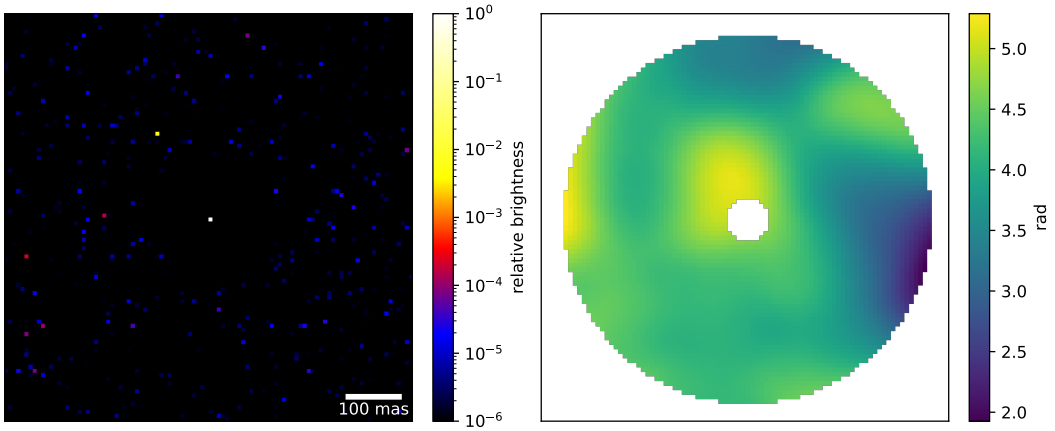

Figure 5.2: Samples drawn from the prior generative model. Left: One prior sample for the object, showing a random realization of the central star and the inverse gamma intensity distribution in all other pixels. In the close vicinity of the star it is enforced that the pixels have zero flux. This improves the quality of the reconstructed star PSF at the price of not being able to detect planets there. Right: One prior sample for the wavefront at telescope's pupil, generated from a Gaussian process with a spatial covariance structure given by the Matérn kernel of equation 5.6.

and its potential partners (which for the sake of simplicity we will call "planets" in the following, disregarding to their physical nature). We also know that the brightness of these sources and their locations in the sky do not vary across the short observation interval; therefore, our model object is time independent.

We focus our present work on the most common case where the star's apparent diameter is smaller than the pixel size so that the prior for the central star is a fixed point source at the center of the field of view. The only free stellar parameter of this model is the star intensity since telescope jittering and pointing errors that shift the star's apparent position in the image will be included in the PSF model. The intensity of the star must be positive and can vary on logarithmic scales, so we use a log-normal prior for it, which also ensures strict positivity in the reconstruction.

In contrast to the star, the locations of the planets are unknown, so we model them by inferring the intensity of every sky pixel, adopting for simplicity the same pixel grid of the data. We also know that there are no planets at most locations. Therefore the reconstructed intensity of most pixels should be very close to zero, given that the sky background has been subtracted from the data. To account for this statistics, we adopt independent inverse gamma priors for the brightness of any pixel. Consequently, the planet fluxes in the different pixels are a priori uncorrelated with each other, and it is much more probable for any pixel to belong to background than to host a planet. A prior sample for the object model is displayed in figure 5.2, left panel.

5.4.2 PSF model

Two separate effects contribute to a physical model for the evolving PSF. The first effect is the fluctuation in the atmospheric path length (section 5.4.2.1), while the second is the telescope optics (section 5.4.2.2). Because a typical high-contrast imaging observation has a field of view of the order of a few arcseconds, or even smaller, the assumption of isoplanatism of equation 5.1 is valid, so we consider both contributions constant across the image.

5.4.2.1 Atmospheric fluctuations model

The light from a pointlike astronomical source approaches the Earth as a plane wavefront but reaches the telescope's aperture as a distorted wavefront after corruption by the rapidly changing density fluctuations of the atmosphere. These distortions can be described by the optical path difference $\phi_{\text{atm}}(t, u)$, between the actual wavefront and the ideal plane wave. It depends on time t and position u in the aperture plane. When expressed in units of wavelength, u corresponds to spatial frequencies in the image plane. The instantaneous wavefront ψ entering the aperture is thus given by:

$$\psi_{\text{atm}}(t, u) = \mathrm{e}^{i\phi_{\text{atm}}(t,u)} \tag{5.3}$$

5.4.2.2 Telescope model

The telescope's adaptive optics introduces opposite phase shifts to partly remove the path differences caused by the atmosphere. These phase shifts can be modeled in the same way as the atmosphere. Hence we have only one path length field ϕ_{eff} describing the combined effects of atmosphere and adaptive optics.

Due to wind and mechanical vibrations, the telescope can jitter, introducing a shift in the detector's field of view. A shift in the detector space is equivalent to a pointwise multiplication of the Fourier transformed image by a phase gradient in the aperture space, with the slope of the gradient being proportional to the shift length. Therefore we add for both spatial directions a linear gradient $\phi_{\text{jit}}(t)$ with a time-dependent slope to the optical path difference.

The geometry of the telescope aperture blocks parts of this incoming wavefront, selecting the spatial frequencies that reach the focal plane. In our equation, we can model this with a multiplication factor $A(u)$, being zero for locations outside the aperture.

Finally, variations in the adaptive optics response and lensing effects in high layers of the atmosphere can change not only the phase but also the intensity of the wave. To also model such effects, we allow for a scalar time-dependent intensity factor, which modulates the amplitude in the aperture plane, which we write as an exponential parameter $e^{\alpha(t)}$ to guarantee the physical constraint of strict positivity.

To summarize, the resulting wavefront model in the optical pupil of the instrument can be written as:

$$\psi(t, u) = A(u)\mathrm{e}^{i\phi_{\text{eff}}(t,u)+i\phi_{\text{jit}}(t)+\alpha(t)} \tag{5.4}$$

The final instantaneous image of the PSF model point source produced at the focal plane is the squared amplitude of the Fourier transform $\mathcal{F}$ of the wave in the optical pupil as derived in chapter 3:

$$p(t, x) = \left|\mathcal{F}_u^x \psi(t, u)\right|^2 = \left|\mathcal{F}_u^x A(u) \mathrm{e}^{i\phi_{\mathrm{eff}}(t,u) + i\phi_{\mathrm{jit}}(t) + \alpha(t)}\right|^2 . \tag{5.5}$$

To this equation, we could add the dependence on λ, since the effects of atmosphere, adaptive optics, and telescope are wavelength-dependent. In this work, though, we do not make this explicit because we apply our method to narrow-band observations.

5.4.2.3 Correlation prior

In every space-time voxel, $\phi_{\mathrm{eff}}(t, u)$, $\phi_{\mathrm{jit}}(t)$, and $\alpha(t)$ can have different values. Nevertheless, because of the physical mechanisms involved, not all values are equally likely. For example, the path length field $\phi_{\mathrm{eff}}(t, u)$ will evolve smoothly in time and space since also the real atmosphere evolves smoothly. Similar, $\phi_{\mathrm{jit}}(t)$ and $\alpha(t)$ will be correlated in time.

Modeling the correlations and using them as prior information substantially improves the reconstructions. Especially for the path length $\phi_{\mathrm{eff}}(t, u)$, the correlations are important to deal with the degeneracy between object and PSF in the imaging equation (5.1). In section 5.6, we demonstrate the benefits of accurately reconstructing these correlations.

We model all correlated fields as zero-centered Gaussian processes. The correlation structures of $\phi_{\mathrm{eff}}(t, u)$, i.e. its power spectrum, might be different in spatial and temporal directions. Because we assume the spatial correlations not to change significantly during an observation, we factorize the Gaussian process into a temporal and a spatial axis and write the total correlation kernel as the outer product of spatial and temporal kernels. For both axes we employ a Matérn kernel (Genton, 2002, Matérn (1986)). The Matérn covariance is translation invariant and therefore diagonal in Fourier space. The Matérn kernel can be parametrized in Fourier space

$$\mathcal{P}(k) = \frac{a^2}{(1 + (k/b)^2)^c}, \tag{5.6}$$

where $\mathcal{P}(k)$ indicates the power spectrum with k as frequency.

The parameters a, b, and c are a priori unknown and are simultaneously reconstructed with the path length field. The parameter a determines the amplitude of the fluctuations, b is a characteristic length scale, and c is the spectral slope of the spatial or temporal fluctuations. For the temporal and spatial axis, the parameters of the kernel can be different.

For $\phi_{\mathrm{jit}}(t)$ and $\alpha(t)$ a similar, but non-parametric, correlation model is used, whose details can be found in Arras et al., 2022.

Figure 5.2, right panel depicts one realization of the wavefront field $\phi(t, u)$ drawn from the described generative model, showing its spatial correlations.

5.5 Likelihood model

For a given object s and evolving PSF p, the model described above determines the time dependent intensity field in the focal plane. The detector's readout and the photon count

statistics are accounted for in equation 5.2 by the likelihood term $P(d|s, p)$ by using Poisson and Gaussian statistics, respectively.

For simplicity, the details of the detector signature like pixel-to-pixel bias non-uniformities, frame-to-frame column bias variations, and pixel-to-pixel non-uniform responses are currently not included in the forward model but will be incorporated in the future. For the current work, the images are pre-corrected for these effects before the actual reconstruction (see section 5.6).

5.5.1 Symmetries of the likelihood

A given PSF $p(t, x)$ does not uniquely determine the wavefront $\psi(t, u)$, because several distinct path length functions will result in the same PSF. In other words, the PSF is symmetric under specific transformations of the wavefront. A detailed review of these symmetries can be found in Paxman et al., 2019, here we shortly outline them:

- Global phase offset symmetry: A global additive offset to the path length does not change the PSF, since the absolute value in equation 5.5 remains constant.

- 2π ambiguity: One can locally add multiples of 2π to the path length field without changing the PSF, since $e^{i2\pi n} = 1$ for all $n \in \mathbb{Z}$.

- Sign ambiguity of the symmetric part of the wavefront (for symmetric apertures): If one writes the path length field $\phi(t, u)$ as the sum of a symmetric part $\phi_s(t, u) = \phi(t, u) + \phi(t, -u)$ and an anti-symmetric one $\phi_a(t, u) = \phi(t, u) - \phi(t, -u)$, then $\phi_+(t, u) = \phi_a(t, u) + \phi_s(t, u)$ and $\phi_-(t, u) = \phi_a(t, u) - \phi_s(t, u)$ both correspond to the same PSF.

Since the PSF model has these three symmetries also the likelihood has these symmetries because the probability of a given frame in the data is the same for any realization of these symmetries in the wavefront.

The symmetries are divided into two categories, continuous symmetries and discrete symmetries. The first symmetry is continuous, while the second and third are discrete. For the continuous symmetry, one solution of the phase screen can be transformed into another solution by a series of local changes without ever changing the resulting PSF or the likelihood. This is not possible for the two discrete symmetries. For example, to transform the path length function $\phi_+(t, u)$ into $\phi_-(t, u)$ without changing the PSF, one has to entirely change the sign of the symmetric part.

MGVI, our inference scheme, iteratively improves the current estimate by making local changes. After performing a sequence of local optimization steps, it can happen that parts of the path length function converged to $\phi_+(t, u)$ and other parts to $\phi_-(t, u)$. Similarly, parts of the path length function can have phase offsets of multiples of 2π. This can be suboptimal for exploiting the correlations of the path length function since 2π jumps and jumps between $\phi_+(t, u)$ and $\phi_-(t, u)$ make the path length function appear less correlated. Even in the presence of some artefacts, there are correlations in the path length function, which can help to improve the reconstruction of the sky brightness.

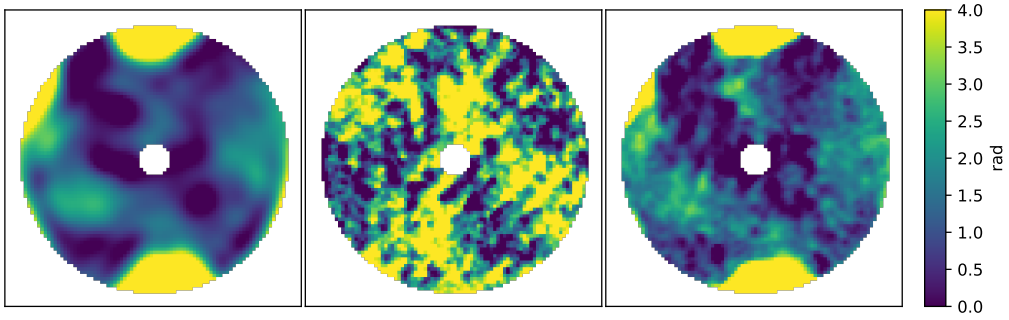

Figure 5.3: Path length functions in the pupil plane at different stages of the optimization. The left panel shows the primary path length function before the auxiliary path length function was removed. As described in the main text it only contains large scale structure explaining the central peak of the PSF. The central panel depicts the auxiliary path length function at the same stage of the optimization. In the right panel the final result of the path length reconstruction is plotted. Many small scale structures form the auxiliary path length function have been transferred to the primary path length function without introducing phase wrapping effects.

5.5.2 Double path length function model

As outlined in the previous section, the phase wrap artefacts caused by the discrete symmetries of the likelihood are only hardly removed by the MGVI optimization. To reduce these artefacts, we alter the model for the initial iterations of the optimization in order to separate the smooth structures from the artefacts. We add an auxiliary path length function for the initial iterations and superimpose the resulting image in the focal plane with the image determined by the primary path length function. Thus each feature in the data can be explained by either of the two path length functions.

For the auxiliary path length function, we tune the hyper-parameters of the Matérn kernel towards favoring flatter power spectrum slopes, thus allowing less correlation for this function. When starting the inference, the additional path length function will quickly accumulate many small-scale structures at the cost of having multiple symmetry artefacts. In contrast, the primary path length function will be smooth, mainly explaining the central peak of the PSF without modeling the small-scale features. Thus the primary path length function reconstruction will be physically plausible, having only little artefacts, but lacking small-scale structures. In figure 5.3 (left and central panel) the primary and auxiliary path length functions are depicted for an intermediate stage of the optimization.

In subsequent iterations, we reduce the weight of the image resulting from the auxiliary path length function in the superposition with the primary image. This forces the reconstruction to explain an increasing amount of small-scale features by the primary path length function, thus gradually transferring small-scale features back to the main path length function without creating phase wrap artifacts. In the right panel of figure 5.3 the path length function is plotted after the transfer of the small scale features from the auxiliary path length function.

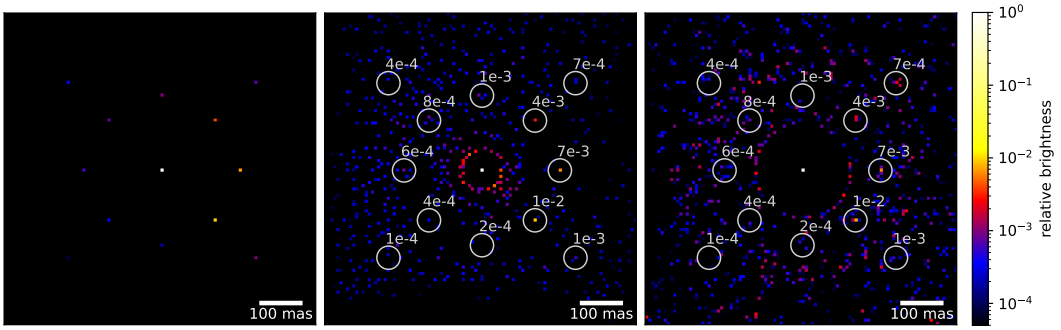

Figure 5.4: Left: Ground truth object: A central star surrounded by 12 injected planets with contrasts ranging from 10^{-2} to 10^{-4}, shown in the field of view of the frames. Center: HCBI object reconstruction from 0.6 s of Gliese 777 data from 600 sequential frames. Right: HCBI object reconstruction from using only 0.1 s of the data. The 0.1 s reconstruction can only detect the brightest 3 companions. Furthermore the 0.6 s reconstruction has a significantly lower background residual.

5.6 Application to a real data

We applied the described algorithm to real on-sky frames sequence acquired with the SHARK-VIS pathfinder instrument, called Forerunner (Pedichini et al., 2017). SHARK-VIS (Mattioli et al., 2018) is the upcoming visible band fast imaging camera for high-contrast imaging at the 8-meter class Large Binocular Telescope (LBT).

The data were acquired on the 5.5 mag star Gliese 777, imaged at a frame rate of 1 KHz through a 40 nm FWHM filter centered at 630 nm, and recorded by an Andor Zyla sCMOS camera. The low read noise of this camera, of less than 2 electrons per pixel, and its small pixels of 6.5μm allows to use an image scale of 3.73 mas/pix, and a frame exposure of 1 ms, that was chosen to oversample the typical spatial and temporal scales of the speckles in the visible band, as measured in Stangalini et al. (2017).

During the observation, the seeing was around 1 arcsec FWHM, and the LBT adaptive optics system (Esposito et al., 2010) was correcting 500 modes in closed loop. An exemplary frame is shown in figure 5.1, left panel.

Fast cadence imaging allows for numerical re-centering of the frames and a frame selection discarding the few frames whose peak is fragmented by strong turbulence. This potentially yields two important advantages. First, it improves the planet's detection due to the increased flux concentration, and second, it increases spatial resolution.
As described in section 5.4 the presented method reconstructs the evolving instantaneous PSF, which makes it suitable to utilize the full potential of such fast-cadence observations.

The whole Gliese 777 data set consists of $1.2 \cdot 10^6$ sequential frames for an interval of 20 minutes, acquired without using the telescope de-rotator (pupil-stabilized imaging) to get the necessary field rotation needed for the application of the standard angular differential techniques mentioned in the introduction. These raw frames were pre-processed before the HCBI reconstruction to reduce detector artifacts, like the frame-to-frame column bias

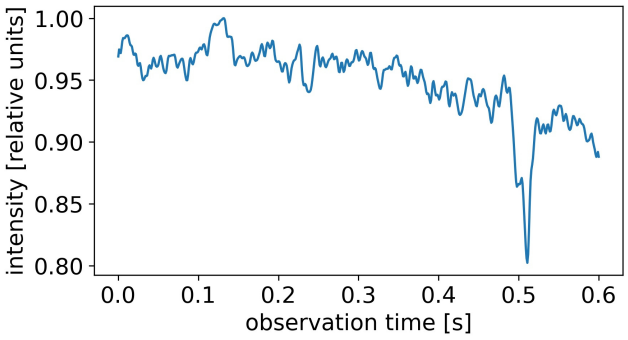

Figure 5.5: Reconstructed amplitude variation of the PSF $\exp(\alpha(t))$ as a function of observation time.

variations and the pixel-to-pixel non-uniform response.

For testing the HCBI reconstruction we only use 600 sequential frames, corresponding to the first 0.6 s of acquisition, in which we injected 12 synthetic planets between 185 mas and 315 mas from the star and with contrasts ranging from 10^{-2} down to 10^{-4} (figure 5.4, left panel).

The mock planets are produced by shifting and scaling each frame to the position and intensity of the planets and adding the corresponding Poisson random counts to the data. These frames have then been binned to a plate scale of 7.46 mas/pix in order to work with 100×100 pixel frames and reduce the high computational workload of the processing.

The MGVI inference process for the current algorithm starts by generating a set of random prior samples from all input variables of the generative model (section 5.4), e.g. the object, the optical path length, the telescope jittering, and the scale of the wavefront amplitude. Afterwards the algorithm computes a model of the frames by a convolution of the thereby generated object samples with the PSF samples and evaluates the likelihood (section 5.5).

Then the MGVI algorithm optimizes the Gaussian posterior approximation with these samples. In the following, this procedure is iterated until convergence, using samples from the current estimate of the posterior instead of prior samples. During these iterations, the weight of the auxiliary PSF is reduced until only the primary PSF is left.

Finally, the sample average can be computed for all quantities in the generative model (section 5.4). Thus HCBI does not only reconstruct the astronomical object and the optical path length but simultaneously calibrates for all the modeled effects, like e.g. the total flux variation $\exp(\alpha(t))$ due to atmospheric scintillation, as shown in figure 5.5.

Figure 5.1, in central and right panels, shows the model frame and the model wavefront for one data frame. All prominent data features are accurately reconstructed, indicating that the model is sufficiently flexible to describe the data. Smaller structures are attributed to noise and are not picked up in the reconstruction thanks to the prior favoring smooth path length functions. A strong correlation is noticeable in the wavefront, as expected from our physical considerations in section 5.4, and also following the prior model.

Figure 5.4, central panel, depicts the reconstructed object after convergence of the algorithm. It is composed of the pointlike star at the center, the correctly reconstructed planets,

and a background residual mainly consisting of isolated points between $5 \cdot 10^{-3}$ and 10^{-4} intensity with respect to the star. The planets are all correctly reconstructed as pointlike objects in the right locations down to a contrast of $6 \cdot 10^{-4}$.

The residual background is not Gaussian, as we expect from the inverse gamma prior that we adopted for the source field.

This result demonstrates that a $0.6s$ acquisition is enough for the HCBI to achieve a detection limit of $6 \cdot 10^{-4}$ on a 5.5 mag star without the need for field rotation. The measured intensities of the reconstructed planets are correct within an error of 5% for the brightest and 17% for the faintest with respect to the ground truth, while the identified planets less or equal to 10^{-3} are comparable with the false positives of the same level in the residual background.

It is worth noting that with fast-cadence imaging, the problem of quasi-static speckles that can mimic a planet signal in long exposure frames is much less severe. Instead, our false positives mostly come from noise.

5.6.1 Comparison with other methods

In order to directly compare the HCBI result with other methods, we convolved the object reconstruction with a Gaussian of the same FWHM of the instrumental PSF. This produces the image in figure 5.6, panel d.

Since field rotation is negligible on a 0.6 s short sequence, the common angular differential methods cannot be used to process these frames. So in figure 5.6 we made a comparison with i) the simple average of the frames after co-registration (panel a), ii) the simple star removal obtained by subtraction of average radial profile and local median (panel b), and iii) the result of a multi-threshold Speckle Free Imaging (SFI; Li Causi et al. (2017), Mattioli et al. (2019)), shown in panel c.

We see that the reconstruction from HCBI (panel d) unambiguously reveals all planets down to $6 \cdot 10^{-4}$ (indicated for reference in panel f) against a fainter background residual than the SFI method, in which they are not distinguishable from the residual spots. Anyway, these planets are present in the SFI result, which suggests that a combination of the two results could even increase the detectability, since planets are correlated while residuals are not, as shown by their geometric average (panel e).

Such comparison, however, is only possible for sufficiently bright stars, because the SFI method is only efficient when speckles are well detectable in a single frame (Li Causi et al. (2017)), while HCBI is less affected by the star being dim because it exploits the entire information across the whole frames sequence.

It is worth noticing that for applying any ADI-based technique, like e.g., the PCA-ADI, to these fast cadence data, at least 10^5 frames would be needed for having a sufficient field rotation to avoid planet self-subtraction.

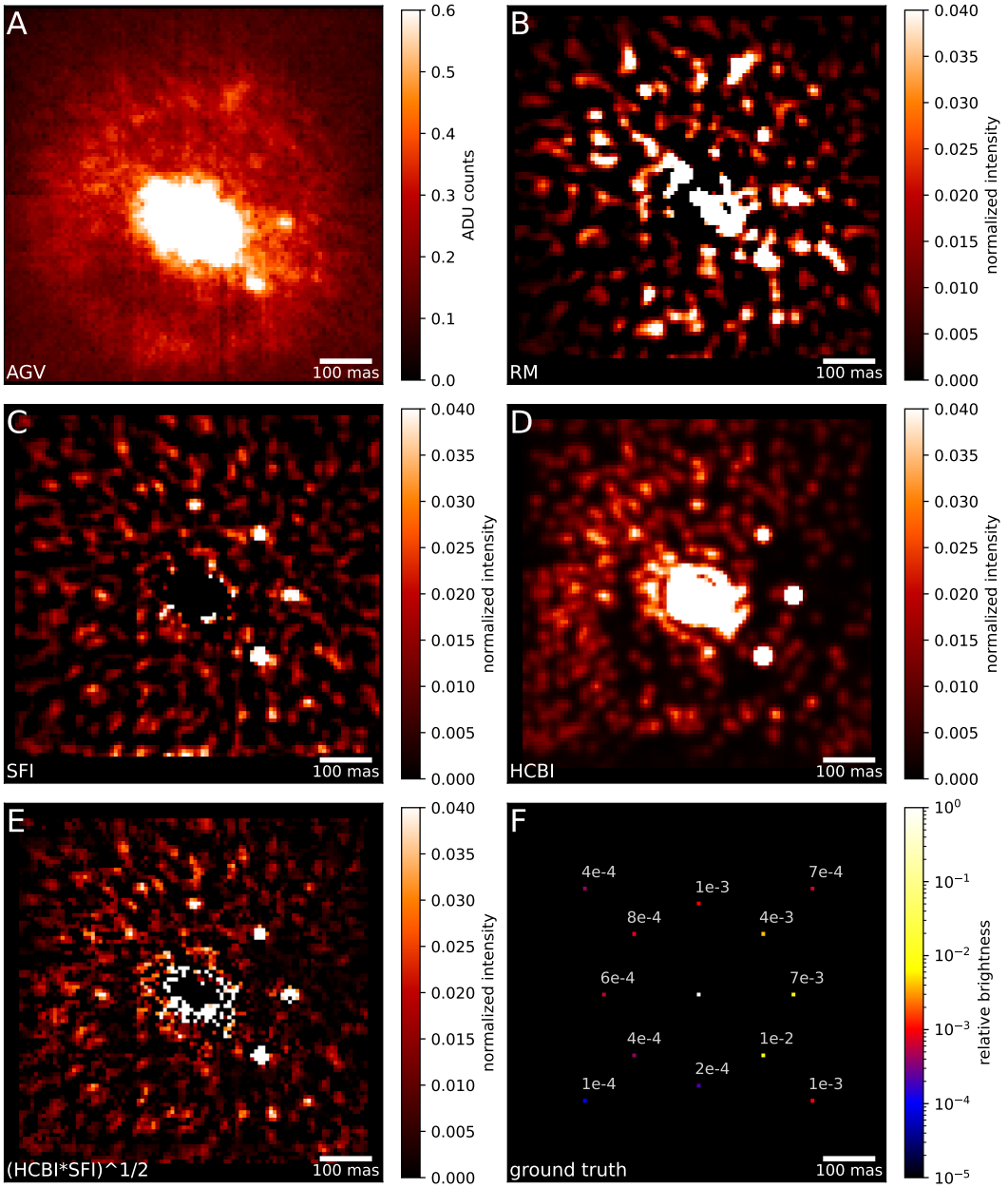

Figure 5.6: Performances of different methods for the same 600 frames: a) simple average after aligning; b) average of the original frames after subtraction of a radial profile and a local median; c) multi-threshold SFI; d) HCBI reconstruction (namely, central panel of figure 5.4) convolved with a Gaussian of the same FWHM of the instrumental PSF; e) geometric average of results from SFI and HCBI; f) ground truth inserted into the data set. All the panels from b to e are shown with the same linear color scale in order to facilitate the comparison of the residuals.

5.7 Discussion

This chapter describes a novel data modeling algorithm for obtaining direct high-contrast images of extrasolar planets, or faint stellar companions, at sub-arcsecond distance from their host star. The algorithm is based on the mathematical methods of information field theory (IFT) and uses Bayesian inference to simultaneously reconstruct the true object and the PSF evolution from a kHz-rate frames sequence acquisition. For the reconstruction, it uses a set of prior information reflecting the physics of the object, the atmospheric perturbations, the telescope and detector response. The prior model resembles the observational effects present at fast-cadence imaging instruments such as SHARK-VIS, allowing the method to unravel the full potential of fast-cadence frame sequences. Special care is taken to deal with symmetries in the wavefront that are observationally indistinguishable but may affect the inference. The wavefront is modeled with a Gaussian process whose correlation power spectrum is learned from data as part of the inference process. Metric Gaussian Variational Inference is adopted at the core of the present algorithm to get robust convergence to an optimal Gaussian approximation of the posterior distribution, from which the final reconstruction is computed.

The performance of HCBI is tested on real on-sky data with artificially injected faint planets, and its detection limits are discussed in comparison with other methods. We show that the HCBI algorithm outperforms the Speckle-Free Imaging (SFI), producing a smaller residual and correspondingly increasing the planets detection limit. Moreover, in contrast to SFI, which needs to detect speckles in each frame, the HCBI method is not limited to bright stars because it uses the whole data cube at once. Finally it does not need field rotation, which is needed by the ADI-based methods to avoid self-subtraction of the planet.

To study the sensitivity increase of the HCBI method as more frames are processed, we also reconstructed a subset of the data consisting of only 100 frames corresponding to 0.1 s of observation. The result of this 100 frames reconstruction is depicted in figure 5.4 right panel. As expected, the sensitivity increases as more frames are processed. Hence processing thousands of frames should improve detection limits further, with reduced background residuals and less false positives. In practice however, we are currently limited to work with a few hundreds of frames, due to the high computational cost of the MGVI reconstruction, which took about 2 weeks to converge to the 600 frames result of figure 5.4, using two 2.4GHz cores of an Intel Xeon Gold 6148 processor with a 20 GB RAM. To face this problem, our plan for the continuation of this work is to study the possibility of splitting large datasets into a number of smaller chunks, to be treated independently and processed in parallel by adopting cloud computing, whose reconstructed information could be joined incrementally. This is physically justified because the whole speckles pattern becomes fully uncorrelated after $\sim 70 - 90$ ms, as measured by Stangalini et al. (2017).

We expect to improve the performance of the method by including the detector signature into the generative data model, which in this work has been corrected by a preprocessing step. Incorporating the signature into the generative model would allow for a joint reconstruction of the object, the PSF evolution, and the detector fluctuations. Also, we will introduce more priors, specific for the cases of resolved stars or extended sources like circumstellar planet-forming disks, in order to enlarge the applicability of the method to all the most common high-contrast imaging situations.

Finally, we plan to ascertain the possibility of modeling long exposure frames, which requires a completely different PSF prior model. Possible models are under investigation but go beyond the scope of the present work.

Chapter 6

Bayesian radio interferometric imaging with direction-dependent calibration

I have published this chapter as first author in the journal Astronomy & Astrophysics as an article (Roth et al., 2023b). Philipp Arras contributed via numerous discussions and helped in the C++ implementation of the image domain gridder. With his enormous experience in C++ development also Martin Reinecke helped in the implementation and optimization of the C++ image domain gridder. Torsten Enßlin contributed via many in-depth discussions on Bayesian inference. Richard Perly helped with the physical interpretation of the obtained calibration solutions. The text was read, commented and approved by all coauthors. For this thesis, the text and figures were slightly adapted.

6.1 Abstract

Radio interferometers measure frequency components of the sky brightness, modulated by the gains of the individual radio antennas. Due to atmospheric turbulence and variations in the operational conditions of the antennas, these gains fluctuate. Thereby the gains do not only depend on time, but also on the spatial direction on the sky. To recover high-quality radio maps, an accurate reconstruction of the direction and time-dependent individual antenna gains is required. This chapter aims to improve the reconstruction of radio images, by introducing a novel joint imaging and calibration algorithm including direction-dependent antenna gains. Building on the `resolve` framework, we designed a Bayesian imaging and calibration algorithm utilizing the image domain gridding method for numerically efficient application of direction-dependent antenna gains. Furthermore, by approximating the posterior probability distribution with variational inference, our algorithm can provide reliable uncertainty maps. We demonstrate the ability of the algorithm to recover high resolution high dynamic range radio maps from VLA data of the radio galaxy Cygnus A. We compare the quality of the recovered images with previous work relying on classically calibrated data. Furthermore, we compare the results with a compressed sensing algorithm also incor-

porating direction-dependent gains. Including direction-dependent effects in the calibration model significantly improves the dynamic range of the reconstructed images compared to reconstructions from classically calibrated data. Compared to the compressed sensing reconstruction, the resulting sky images have a higher resolution and show fewer artifacts. For utilizing the full potential of radio interferometric data, it is essential to consider the direction dependence of the antenna gains.

6.2 Introduction

Over the last decades, many modern radio interferometers such as the Very Large Array (VLA), the Atacama Large Millimeter Array (ALMA), the Low-Frequency Array (LOFAR), the Australian Square Kilometer Array Pathfinder (ASKAP), and MeerKAT have been constructed and are now providing high-quality data. With LOFAR 2.0, MeerKAT+, and the Square Kilometre Array (SKA), the future will bring even more instrumental progress to radio astronomy. Utilizing the full potential of these instruments has become a major challenge, not only because of the enormous data size but also because of the complexity and variety of processes affecting the measurement.

This makes the accurate calibration of the antenna gains an important aspect for unraveling all of the information from the measurement data, as introduced in section 3.2.1. The antenna gains are complex-valued functions, describing the sensitivity of the individual antennas of the interferometer, and depend on time and spatial direction on the sky. Traditionally the direction-dependent effects (DDEs) were neglected, and only direction-independent effects (DIEs) were considered, rendering the antenna gains only time-dependent.

Clearly, DIEs are the dominant contribution to the antenna gains; nevertheless, DDEs exist, and neglecting them in the calibration procedure limits the image fidelity and dynamic range of the resulting sky reconstructions. From the instruments mentioned above, especially LOFAR measurements are affected by DDEs in the antenna gains, making a DDE calibration essential. However, correcting for DDEs is also beneficial for other instruments.

Historically, the first DDE considered in radio interferometry was the so called w-effect resulting from non-coplanar baselines of the interferometer. Many algorithms have been developed to account for the w-effect (as for example Cornwell et al., 1992, Perley, 1999, Cornwell et al., 2008, Offringa et al., 2014, Pratley et al., 2019, Ye et al., 2021, Arras et al., 2021b). Since the w-effect is purely geometrical, it can be computed form the antenna configuration and, in contrast to DDEs arising form atmospheric fluctuations and operational conditions of the antennas, it does not need to be reconstructed from observational data. For the rest of this chapter, we focus on DDEs that are not known a priori but need to be recovered from data, and that might also be time-dependent.

Calibration algorithms correcting for such unknown and potentially time-dependent DDEs have been developed over the last years, initially building on such w-effect correction techniques as for example proposed by Bhatnagar et al., 2008. Nowadays, DDE correction is most commonly achieved via partitioning the field of view into small facets, assuming piecewise constant DDEs. Within each facet, a DIE calibration is performed. This faceting based approach was developed in Smirnov et al., 2015, Weeren et al., 2016, and Tasse et al., 2018. Albert et al., 2020 extended this approach by proposing to infer a smooth phase

screen within each facet.

Repetti et al., 2017 introduced a joint DDE calibration and imaging algorithm going beyond faceting and the assumption of piecewise constant DDEs. For regularization, smoothness of the DDEs is enforced via modeling them as band-limited Fourier kernels. Additionally, the sky brightness is regularized via compressed sensing, assuming sparsity in an orthonormal basis. This approach was generalized in Thouvenin et al., 2018, Dabbech et al., 2019, and Birdi et al., 2020, adding the time dependence to the DDE correction and also accounting for polarization. Building on this algorithm, Dabbech et al., 2021 demonstrate the benefit of DDE calibration on a Cygnus A VLA dataset.

In Arras et al., 2019b, a joint Bayesian imaging and DIE calibration method is presented. Here we extend this approach to a joint Bayesian imaging and DIE as well as DDE calibration method. Similar to Repetti et al., 2017, we a priori assume smoothness of the DDEs. Nevertheless, we did not fix the degree of smoothness via a band-limited kernel but used self-adaptive correlations priors described in section 6.4.1.3. Also for the DIE gain solutions and the sky brightness we used self-adaptive correlation priors. As in Dabbech et al., 2021, we also demonstrate the method presented in this chapter on Cygnus A VLA data. This allows for insightful comparisons and validations of the two algorithms.

For computational efficiency, we utilized the recently developed image domain gridding algorithm (Tol et al., 2018). Image domain gridding allows for an efficient evaluation of the radio interferometric measurement equation in the presence of direction-dependent antenna gains. Therefore, image domain gridding is a key component for correcting DDEs that do not have band-limited kernels.

The remainder of the chapter is structured as follows. In section 6.3 we outline the radio imaging and calibration problem, already introduced in section 3.2.2, in a Bayesian setting. Section 6.4 describes our reconstruction algorithm. In section 6.5 we demonstrate the algorithms performance and compare the image quality to existing work. Section 6.6 gives a conclusion and an outlook for future work.

6.3 The inference problem

The radio interferometric measurement equation, derived in Smirnov, 2011a and chapter 3, relates a given sky brightness and antenna gains to model visibilities. To be more precise, the model visibility $\tilde{V}_{pqt}$ of antennas p and q at time t is given by

$$\tilde{V}_{pqt} = \int C(\boldsymbol{l}, w_{pqt}) I(\boldsymbol{l}) G_p(\boldsymbol{l}, t) G_q^*(\boldsymbol{l}, t) \mathrm{e}^{-2\pi i (\boldsymbol{k}_{pqt} \cdot \boldsymbol{l})} d\boldsymbol{l}, \tag{6.1}$$

with $\boldsymbol{l} = (l, m)$ being the sky coordinates, t the time coordinate, $\boldsymbol{k}_{pqt} = (u_{pqt}, v_{pqt})$ the baseline uv-coordinates in units of the imaging wavelength, $C(\boldsymbol{l}, w_{pqt}) = \exp\left(-2\pi i w_{pqt}\left(\sqrt{1-\boldsymbol{l}^2}-1\right)\right)/\sqrt{1-\boldsymbol{l}^2}$ the w-effect, $I(\boldsymbol{l})$ the sky brightness, and $G_p(\boldsymbol{l}, t)$ the gain of antenna p. Thus, model visibilities are Fourier components of the sky brightness I modulated by the antenna gains G_p and G_q as well as by C, encoding the w-effect. The visibilities V_{pqt} actually measured by the interferometer are the noise less visibility $\tilde{V}_{pqt}$ plus

some additive noise n_{qpt}

$$V_{pqt} = \tilde{V}_{pqt} + n_{qpt}. \tag{6.2}$$

While the w-effect C is known, the sky brightness I as well as the antenna gains G have to be reconstructed. Recovering the underlying sky brightness I and antenna gains G from the noisy measurement data is an inverse problem, as discussed in chapter 4.

As outlined in chapter 4 the inverse problems of the algorithms presented in this thesis are addressed from a probabilistic perspective. Thus, we seek for an estimate of the posterior distribution $P(I, G|V)$, which is the probability distribution of the sky I and antenna gains G conditioned on the measured visibilities V. Via Bayes' Theorem (equation 4.4)

$$P(I, G|V) = \frac{P(V|I, G)P(I, G)}{P(V)}, \tag{6.3}$$

this posterior distribution can be expressed in terms of the likelihood $P(V|I, G)$, the prior $P(I, G)$, and the evidence $P(V)$. While the likelihood and the prior actually depend on I and G, the evidence only acts as a normalization constant. The prior $P(I, G)$ encodes general knowledge and assumptions about the sky brightness I and the antenna gains G before considering the data. Based on these assumptions, the prior assigns a probability density to each possible configuration of I and G. We discuss the prior assumptions of this algorithm in detail in section 6.4.1.

For a given I and G, the model visibilities $\tilde{V}$ can be computed via the measurement equation 6.1. The likelihood $P(V|I, G) = P(V|\tilde{V})$ is determined by the statistics of the measurement noise in equation 6.2. We discuss the likelihood and the noise model in section 6.4.3.

6.4 The algorithm

6.4.1 Prior model

The sky brightness $I(\boldsymbol{l})$ is a largely continuous function of the spatial position $\boldsymbol{l}$ on the sky, except for locations of point sources. Similarly, the antenna gains $G_{pt}(\boldsymbol{l}, t)$ are largely continuous functions of position $\boldsymbol{l}$ and time t. Unless some parametric form is considered, continuous quantities have an infinite number of degrees of freedom. Thus, reconstructing them from finite measurement data is an ill-posed problem, since the measurement data does not uniquely determine the solutions for $I(\boldsymbol{l})$ and $G_{pt}(\boldsymbol{l}, t)$. To circumvent this problem, some form of regularization is required.

Information field theory (Enßlin, 2019) gives the theoretical foundation for expanding Bayesian inference to continuous quantities. In Bayesian inference, classical regularizers do not exist. Instead, the prior allows one to discriminate between all solutions compatible with the data. We make some mild assumptions about the involved quantities to derive a reasonable prior distribution. The key prior assumptions are smoothness and positivity, which we detail in sections 6.4.1.1 and 6.4.1.2, respectively.

We encode the prior probability distributions in the form of standardized generative models (Knollmüller et al., 2018). Standardized generative models transform a set of inde-

pendently distributed Gaussian random variables $\{\xi_i\}$ via a nonlinear mapping to the desired target distribution.

6.4.1.1 Smoothness

Almost all continuous physical quantities, called fields hereafter, are smooth functions of their respective coordinates. Thus, the values of a field at nearby locations are not independent of each other but are correlated. Exploiting such correlations enables the inference of fields from finite measurement data.

We model all correlations via Gaussian processes. A priori we assume statistical homogeneity and isotropy of the Gaussian process, meaning that a priori there is no special position or direction. According to the Wiener-Khintchine theorem (Khintchine, 1934) the correlation kernels of homogeneous and isotropic processes are diagonal in Fourier space, with the power spectrum on the diagonal. We reconstruct the power spectrum of the field jointly with the field values themselves. For the power spectrum, we utilize a nonparametric model, encoding mild prior assumptions, such as, it is falling and is itself a smooth function of the Fourier frequency. The details of the correlation model can be found in Arras et al., 2022.

As we already pointed out in section 6.4.1, we encode our prior distributions in the form of generative models. Also, this correlation model has the form of a generative model. Thus, it maps uncorrelated Gaussian random numbers ξ to a correlated Gaussian process $\psi(\xi)$.

6.4.1.2 Sky prior model

In the current version of the algorithm, we ignore polarization and only consider Stokes I imaging. Since there is no negative flux, the sky brightness is constrained to be strictly positive $I(\boldsymbol{l}) > 0$. In our model, $I(\boldsymbol{l}) = I_d(\boldsymbol{l}) + I_p(\boldsymbol{l})$ consists of two components, with $I_d(\boldsymbol{l})$ modeling the diffuse sky emission and $I_p(\boldsymbol{l})$ modeling the point source emission. Both have to obey the positivity constraint $I_{d/p}(\boldsymbol{l}) > 0$ individually.

The characterizing property of diffuse sky emission is that it is smoothly varying as a function of sky direction. The brightness of the diffuse sky emission varies over several orders of magnitude and is strictly positive. For these reasons, we model the logarithm of the diffuse sky emission with a two-dimensional generative model for correlated Gaussian process $\psi_{sky}(\xi_{sky})$. After exponentiation we get the generative model for $I(\boldsymbol{l})$

$$I_d = \mathrm{e}^{\psi_{sky}(\xi_{sky})}, \tag{6.4}$$

encoding a correlated log-normal distribution. This satisfies the requirements of a strict positivity and smoothness.

Also, the point source model should satisfy the positivity constraint. Furthermore, the point source model needs to be flexible enough to capture extreme variations in the brightness of point sources. Therefore, we choose to model point sources with pixel-wise inverse gamma-distributed brightness values, as already done in Arras et al., 2021a and Selig et al., 2015. We insert two such pixels with inverse gamma-distributed point source fluxes at the locations of the two sources near the core of Cygnus A.

6.4.1.3 Antenna gain prior model

At the current stage of the algorithm, we neglect polarization leakage. Therefore, the gain of antenna p can be described as a diagonal matrix

$$G_p(\boldsymbol{l}, t) = \begin{pmatrix} G_p^{rr}(\boldsymbol{l}, t) & 0 \\ 0 & G_p^{ll}(\boldsymbol{l}, t) \end{pmatrix}, \tag{6.5}$$

with $G_p^{rr}(\boldsymbol{l}, t)$ and $G_p^{ll}(\boldsymbol{l}, t)$ being the calibrations of the two antenna feeds $\langle ll \rangle$ and $\langle rr \rangle$. The physical origin of the $\boldsymbol{l}$, t and jointly $(\boldsymbol{l}, t)$-dependent variations of the antenna gains can be different. For example, the purely time-dependent variations arise from fluctuations in the amplification of the antenna signals, as well as from atmospheric effects with angular scales larger than the antenna beam. Only direction-dependent variation might arise from deviations between the actual antenna power pattern and the assumed model. Finally, pointing errors of the antennas and atmospheric effects with angular scales smaller than the antenna beam cause joint time and direction-dependent gain variations.

Thus, the physical origins of the time and/or direction dependent variations are different, and therefore, the statistics of the gain variations along these axes are expected to be different. For this reasons, we separate the purely time and direction-dependent terms from the joint direction and time-dependent term. Thus, we split $G_{pt}^{ll/rr}(\boldsymbol{l}, t)$ into

$$G_p^{ll/rr}(\boldsymbol{l}, t) = g_p^{ll/rr}(t) \cdot g_p^{ll/rr}(\boldsymbol{l}) \cdot g_p^{ll/rr}(\boldsymbol{l}, t), \tag{6.6}$$

with ll/rr standing for the two cases $< ll >$ or $< rr >$. Splitting the antenna gain into these three factors has the advantage that it allows one to encode different prior knowledge into the generative models of the individual terms. For example, we can encode that the a priori expected fluctuations of $g_p^{ll/rr}(\boldsymbol{l})$ and $g_p^{ll/rr}(\boldsymbol{l}, t)$ are much smaller than that for $g_p^{ll/rr}(t)$.

In order to avoid degeneracies between the three antenna gain factors in equation 6.6, we suppress the only time t and only direction $\boldsymbol{l}$ dependent contributions in the generative model for $g_p^{ll/rr}(\boldsymbol{l}, t)$ by following a similar idea presented in Guardiani et al., 2022 which we outline below.

Each antenna gain term is a complex-valued function. Because the variations of phase and amplitude of the antenna gains have different origins, the typical temporal and spatial correlation length is expected to be different. This makes it a natural choice to have distinct models for phase and amplitudes. We model the logarithm of the gain amplitude and the phase with two distinct Gaussian processes. Thus, we have the following forward model

$$g_p^{ll/rr}(\cdot) = \mathrm{e}^{\psi_{p,\mathrm{logamp}}^{ll/rr}(\cdot) + i\psi_{p,\mathrm{phase}}^{ll/rr}(\cdot)}, \tag{6.7}$$

with $(\cdot)$ representing either (t) or $(\boldsymbol{l})$.

As already mentioned we suppress the contribution of only time t or only direction $\boldsymbol{l}$ dependent parts in the generative model for $g_p^{ll/rr}(\boldsymbol{l}, t)$ as they should be captured by $g_p^{ll/rr}(t)$ and $g_p^{ll/rr}(\boldsymbol{l})$ respectively. To do so, we alternate the phase and amplitude models in equation 6.7 for the case of the joint $(\boldsymbol{l}, t)$-dependence. Instead of directly using the output of the Gaussian processes $\psi_{\mathrm{logamp}}^{ll/rr}(\boldsymbol{l}, t)$ and $\psi_{\mathrm{phase}}^{ll/rr}(\boldsymbol{l}, t)$ as a model for the jointly $(\boldsymbol{l}, t)$-dependent

amplitudes and phases, we first subtract the integrals over $\boldsymbol{l}$ and t. Thus, we use $\tilde{\psi}^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t)$ defined as

$$\begin{aligned}\tilde{\psi}^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t) = \psi^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t) &- (\mathrm{vol}_{\boldsymbol{l}})^{-1} \cdot \int \psi^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t)\, d\boldsymbol{l} \\ &- (\mathrm{vol}_t)^{-1} \cdot \int \psi^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t)\, dt, \end{aligned} \tag{6.8}$$

for the amplitude and phase models, with $(\cdot)$ standing for (logamp) or (phase) respectively. Thereby $\mathrm{vol}_{\boldsymbol{l}} = \int 1\, d\boldsymbol{l}$ and $\mathrm{vol}_t = \int 1\, dt$ are the volumes of the direction and time spaces. Thus, the jointly $(\boldsymbol{l}, t)$-dependent calibration factor can be summarized as

$$g^{ll/rr}_p(\boldsymbol{l}, t) = \mathrm{e}^{\tilde{\psi}^{ll/rr}_{\mathrm{logamp}}(\boldsymbol{l},t) + i\tilde{\psi}^{ll/rr}_{\mathrm{phase}}(\boldsymbol{l},t)}, \tag{6.9}$$

with $\tilde{\psi}^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t)$ as defined in equation 6.8.

This definition of $\tilde{\psi}^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t)$ reduces the contributions of only t or only $\boldsymbol{l}$ dependent parts. To explicitly verify this, we can, as a test, add an only $\boldsymbol{l}$-dependent contribution $\psi'^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t) = \psi^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t) + \phi(\boldsymbol{l})$. Then, the resulting $\tilde{\psi}'^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t)$ of the modified $\psi'^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t)$ is:

$$\begin{aligned}\tilde{\psi}'^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t) &= \psi^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t) + \phi(\boldsymbol{l}) - (\mathrm{vol}_{\boldsymbol{l}})^{-1} \cdot \int \left(\psi^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t) + \phi(\boldsymbol{l})\right) d\boldsymbol{l} \\ &\quad - (\mathrm{vol}_t)^{-1} \cdot \int \left(\psi^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t) + \phi(\boldsymbol{l})\right) dt \\ &= \tilde{\psi}^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t) + c. \end{aligned} \tag{6.10}$$

Thus, only $\boldsymbol{l}$-dependent contributions, and similarly also only t-dependent contributions, in $\psi^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t)$ imprint on $\tilde{\psi}^{ll/rr}_{(\cdot)}(\boldsymbol{l}, t)$ only as a constant offset c. This guides the reconstruction to capture only t or only $\boldsymbol{l}$-dependent variations of the gain factor by $g^{ll/rr}_p(t)$ and $g^{ll/rr}_p(\boldsymbol{l})$, respectively.

6.4.2 Forward model

Besides the science target, calibration sources are also observed. We combine the science observation with the calibration observations into a joint forward mode as described in the next three subsections.

6.4.2.1 Science target

For the science target observation, we use the antenna gain and sky models as described above and compute the corresponding model visibilities via the measurement equation 6.1. For the numeric evaluation of the measurement equation we use the image domain gridding algorithm (Tol et al., 2018) as outlined in section 6.4.2.4.

6.4.2.2 Flux calibrator

For the flux calibration observation, we do not reconstruct the sky brightness but assume it to be a point source with known flux in the center of the field of view. Thereby, we assume the fluxes of the calibration sources reported in Perley et al., 2013.

We use only the direction-independent factors in the antenna gain model, since direction-dependent calibration cannot be performed with a single point source. Furthermore, the flux calibration antenna gains only share the amplitudes with the other observations while having a separate phase model. Thus, the flux calibration antenna gains share the Gaussian process for the logarithm of the amplitude $\psi_{\text{logamp}}(t)$ but not the phase process $\psi_{\text{phase}}(t)$.

Additionally, we allow the flux calibration antenna amplitudes to have a small offset to the amplitudes of the science target. We assume this offset to be a piecewise constant function of time t with the breaking points at the time steps where the telescope switches the target. To summarize, the antenna gain for the flux calibration is given by

$$g_p^{ll/rr}(t)_{\text{Fcal}} = \mathrm{e}^{\psi_{\text{logamp}}^{ll/rr}(t)+\alpha_{\text{Fcal}}^{ll/rr}(t)+i\psi_{\text{phase}_{Fcal}}^{ll/rr}(t)}. \tag{6.11}$$

Here $\psi_{\text{logamp}}^{ll/rr}(t)$ is the only time-dependent Gaussian process modeling the logarithm of the amplitude. This part of the gain model is shared with the other observations. $\alpha_{Fcal}^{ll/rr}(t)$ is the piecewise constant amplitude offset, and $\psi_{\text{phase}_{Fcal}}^{ll/rr}(t)$ is the phase of the amplitude calibration gain. We use a zero-centered Gaussian as a prior model for the amplitude offsets $\alpha_{Fcal}^{ll/rr}(t)$. We do not assume a fixed variance of this Gaussian but simultaneously reconstruct it from the calibration data. Thereby, we use an inverse gamma prior on variance. Via the heavy tails of the inverse gamma distribution, we encode the prior knowledge that the offset is probably small but can sometimes significantly deviate from zero.

6.4.2.3 Phase calibrator

Similar to the flux calibration observation, we assume the phase calibrator to be a point source in the center of the field of view. Nevertheless, we do not set a fixed brightness of this point source but reconstruct it from the data.

Again we only use the direction-independent factors of the antenna gain model, since direction-dependent calibration can not be done with a single point source. In contrast to the flux calibration, for the phase calibration not only the amplitude model but also the phase model of the antenna gains is shared with the other observations.

As already introduced for the flux calibration model, we allow for small discontinuities in the calibration solution at the timesteps where the telescope switches the observation target. Since for the phase calibration observation, not only the amplitude model but also the phase model is shared with the science target, we also allow for an offset in phases between the two observations. Expressed as a formula, the antenna gain model for the phase calibration is defined by:

$$g_p^{ll/rr}(t)_{\text{Phcal}} = \mathrm{e}^{\psi_{\text{logamp}}^{ll/rr}(t)+\alpha_{\text{Phcal}}^{ll/rr}(t)+i\psi_{\text{phase}}^{ll/rr}(t)+i\beta_{\text{Phcal}}^{ll/rr}(t)}. \tag{6.12}$$

Here, $\psi_{\text{logamp}}(t)$ and $\psi_{\text{phase}}(t)$ are the Gaussian processes modeling the phase and the logarithm of the antenna gains, which are shared with the other observations. $\alpha_{\text{Phcal}}^{ll/rr}(t)$ and $\beta_{\text{Phcal}}^{ll/rr}(t)$ are the piecewise constant offset functions. As a prior for the offsets $\alpha_{\text{Phcal}}^{ll/rr}(t)$ and $\beta_{\text{Phcal}}^{ll/rr}(t)$, we again use a zero-centered Gaussian with an inverse gamma distributed variance, as already described for the flux calibrator offset function.

6.4.2.4 Image domain gridding

The radio response $R(I, G)$ maps a given sky brightness and antenna gains to model visibilities as defined by the measurement equation 6.1. Essentially the radio response R is a nonuniform Fourier transformation of the sky brightness modulated by the w-effect and the antenna gains. These modulations by the antenna gains are different for every baseline and timestep. They make classical algorithms for computing the radio response numerically inefficient if one includes the direction-dependence of the gains. The novel image domain gridding algorithm by Tol et al., 2018 overcomes this challenge and enables a fast evaluation of R even in the presence of direction-dependent gains. We use image domain gridding to compute the radio response R.

For a fast optimization of the joint posterior estimate of I and G (see section 6.4.4), we not only need to evaluate $R(I, G)$ but also its derivative. More specifically, we implemented the image domain gridding in such a way that we can also apply the Jacobian of R at the position (I_0, G_0) to perturbations of the sky dI and the gains dG. Thus, we can compute:

$$DR_{I_0,G_0}(dI, dG) = \left(\frac{\partial R}{\partial I}\Big|_{I_0,G_0} \quad \frac{\partial R}{\partial G}\Big|_{I_0,G_0} \right) \begin{pmatrix} dI \\ dG \end{pmatrix}. \tag{6.13}$$

Furthermore, we also need to apply the adjoint Jacobian to perturbed visibilities dV. Thus, we can evaluate:

$$\left(DR_{I_0,G_0}\right)^\dagger (dV) = \begin{pmatrix} \frac{\partial R}{\partial I}\Big|^*_{I_0,G_0} \\ \frac{\partial R}{\partial G}\Big|^*_{I_0,G_0} \end{pmatrix} \Big(dV\Big). \tag{6.14}$$

The Jacobian and its adjoint are implicitly evaluated, thus they are neither stored nor applied as a full matrix.

6.4.3 Likelihood model

With the above forward model and the image domain gridding algorithm for the measurement equation 6.1 the model visibilities can be computed. The statistics of the noise in equation 6.2 then determines the likelihood. We assume Gaussian measurement noise statistics. Nevertheless, we do not assume a fixed noise covariance, but allow for an outliers correction by introducing a noise correction factor per data point. During the inference we jointly reconstruct these correction factors with the calibration and sky brightness. For each correction factor we use an inverse gamma distribution as a prior, having a strong peak at 1. Thereby, we can ensure that if the data is preferentially explained by the model, and only if this is not possible is classified as an outlier. A similar procedure was introduced in Arras et al., 2021a and was named "automatic weighting". In essence, this automatic weighting is equal to natural weighting, where the weights get updated during the reconstruction. More specifically, if a visibility gets detected to be an outlier its weight gets decreased.

For all targets we compute the likelihood $P(V|I, G)$ as described above. The overall likelihood is then the product of the individual likelihoods for calibration and science targets:

$$P(V|I, G) = \prod_i P(V_i|I, G). \tag{6.15}$$

Because we include temporal smoothness into our model (see section 6.4.1.3) the likelihoods contribute information on the calibration solutions not only for the exact time intervals of their observation, but also for other observations. Thereby, the temporal correlation kernels, reconstructed for the gain factors, determine the extrapolation of the gain solution of the calibration observations to the science observation (and vice versa). Therefore, jointly inferring the calibration solution and the science image is beneficial.

6.4.4 Posterior

In the sections above, we have developed a likelihood model $P(V|I,G)$ and encoded the prior distribution $P(I,G)$ via the generative form of I and G. To actually obtain an image I and antenna gains G from the observed visibilities V, it is necessary to compute, or approximate, the posterior distribution $P(I,G|V)$. Directly computing the posterior distribution via equation 6.3 is impossible since the model involves too many parameters and is highly nonlinear. Instead, we approximate the posterior distribution via a two-stage algorithm.

In the first stage, we compute the maximum a posteriori solution (MAP) for the sky I and the antenna gains G by maximizing the joint probability $P(V,I,G) = P(V|I,G) \cdot P(I,G)$ of visibilities, sky brightness, and antenna gains. This gives us already a decent reconstruction of the sky brightness and the calibration solutions; nevertheless, it lacks a reliable uncertainty quantification.

Therefore, in the second stage, we compute a more accurate approximation of the posterior probability distribution of the sky brightness using the antenna gain solutions obtained in the first stage of the inference. Thus, expressed as a formula, we approximate $P(I|G,V)$. We compute this posterior approximation using geometric variational inference (geoVI), a technique developed by Frank et al., 2021. geoVI, as outlined in chapter 4, is an accurate variational inference technique capable of capturing non Gaussian posterior statistics and intercorrelation between the posterior parameters. The output of the geoVI inference is a set of posterior samples of the sky brightness. The mean sky brightness and its standard deviation can be computed from these samples.

Of course, ideally, not only the sky I but also the antenna gains G should be inferred via geoVI to include calibration uncertainties into the uncertainty quantification of the recovered sky. However, the computational demand of the geoVI inference is significantly higher than computing the MAP solution. This renders it at the current stage infeasible to have a joint geoVI estimate of the sky and the antenna gains. Therefore, in this work, the uncertainties on the sky brightness are only relative to the antenna gain calibration computed by MAP. Nevertheless, we believe these uncertainty maps to be significantly more accurate than noise estimates obtained from residual flux images of `clean`-like algorithms, also because residual images are always conditional to the assumed calibration solutions.

6.5 Demonstration on Cygnus A VLA data

We demonstrate the performance of our algorithm on a VLA Cygnus A observation. We use the S band data at 2.052 GHz in all four VLA configurations. In recent years the Cygnus A VLA data has been widely used for presenting and benchmarking novel imaging and cali-

bration algorithm (see, for example, Arras et al. 2021a, Dabbech et al. 2021, Sbokolodi et al. 2020, Dabbech et al. 2018). This allows for insightful comparison and validation of the imaging algorithms. In particular, we compare the results of this work with the `resolve` reconstruction from Arras et al., 2021a and the compressed sensing reconstruction from Dabbech et al., 2021. In the following two subsections we briefly summarize these two algorithms. For completeness we also did a reconstruction using the algorithm of this work but only included direction-independent calibration terms. The results of this reconstruction are discussed in the appendix 6.8.1. The code underlying the shown reconstructions can be found here.[1] The results of the sky reconstruction are available in fits format at.[2]

We choose a FoV of 0.05×0.05 degrees and a grid of 1000×1000 pixels. Thus the individual pixels have a side length of $0,00005$ degrees or 0.18 arcsec. The pixel of the DIE antenna gains have a size of 20 s for all four VLA configurations. We use 32×32 spatial pixels for DDE gains, which get Fourier interpolated on the sky grid by the image domain gridding algorithm Tol et al., 2018. For the temporal resolution of the jointly direction-dependent antenna gains we choose 400 s. For this work we used the uncalibrated Cygnus A data together with observations of the flux and phase calibrators to make a joint reconstruction.

6.5.1 `resolve` with traditional calibration

The `resolve` reconstruction of Arras et al., 2021a uses the same underlying framework and sky model as this work. Nevertheless, in contrast to this work, it uses traditionally calibrated data as an input, ignoring direction-dependent effects. This comparison allows studying the impact of direction-dependent calibration on Bayesian imaging.

In Arras et al., 2021a, the old `resolve` reconstruction without direction-dependent calibration was compared to single and multi-scale `clean` images, showing that the images reconstructed with `resolve` had a higher resolution than the single and multi-scale `clean` images. By comparing with higher frequency data, the small scale emission additionally recovered by the `resolve` algorithm could be confirmed to be real.

6.5.2 Compressed sensing with direction-dependent calibration

Furthermore, we compare with the algorithm of Dabbech et al., 2021. Similar to our method, Dabbech et al., 2021 also performs a direction-dependent calibration avoiding faceting. In contrast to our algorithm this method takes the already classically calibrated visibilities as an input, and performs the direction-dependent calibration and imaging on top. Also the methodology of this method is quite different as it uses compressed sensing for the regularization of the sky brightness and band limited Fourier kernels for the direction-dependent gains. Since the methodology of this method is very different is it well suited as a confirmation of the results of this work.

[1] `https://gitlab.mpcdf.mpg.de/ift/public/cygnus_a_dde_cal`
[2] `https://doi.org/10.5281/zenodo.8178695`

6.5.3 Comparison of Cygnus A images

In figure 6.1, the Cygnus A image reconstructions of this algorithm and the two studies mentioned above are displayed. All three algorithms deliver similar images of Cygnus A. Especially the two Bayesian `resolve` reconstructions have only minimal artifacts in the recovered images. The compressed sensing algorithm produces some artifacts at the hot spots and the core of Cygnus A.

An important quantity for comparing the quality of radio maps is their resolution, thus the smallest length scale at which features can still be imaged. While for classical `CLEAN` images, this length scale is usually set by the size of the `CLEAN` bean, more advanced imaging algorithms such as `resolve` and compressed sensing techniques can achieve super-resolution (Arras et al. 2021a, Dabbech et al. 2018). Thus, for sufficiently high signal-to-noise, these algorithms can recover features smaller than the intrinsic resolution of the interferometer.

In high surface brightness regions, the resolution of both Bayesian images is slightly higher. Already in Arras et al., 2021a, the very high resolution of the Bayesian reconstruction compared with `clean` was discussed, and its fidelity was confirmed by comparison with higher frequency data. In order to depict the difference in resolution more clearly, the right column of figure 6.2 shows a zoom-in on the western hot spot. Also, in the zoom on the hot spots, the three algorithms yield very compatible reconstructions.

Figure 6.2 left column depicts the core, the jet, and regions of the lobes with lower surface brightness. In low surface brightness regions, the resolution of the `resolve` reconstruction from Arras et al., 2021a using classically calibrated data is significantly lower than the resolution of the reconstruction of this work and also lower than the resolution of the compressed sensing reconstruction of Dabbech et al., 2021. Both algorithms utilizing direction-dependent calibration depict the sharp structures of the jet much more clearly than the algorithm of Arras et al., 2021a building on classical calibration. That direction-dependent calibration enhances the jet was already found in Dabbech et al., 2021 compared to other methods building on classical calibration.

Besides the resolution, the sensitivity, specifying the dimmest recovered emission, is the second important quantity determining the quality of a radio map. This sets the dynamic range of a radio map as the ratio between the faintest to the brightness emission. By comparison of three images in figure 6.1, it is evident that the two algorithms performing direction-dependent calibration provide radio maps with high dynamic range, as they recover structures with much lower surface brightness. The `resolve` reconstruction of Arras et al., 2021a without direction-dependent calibration has a significantly higher residual background brightness. The dynamic range gain also manifests in figure 6.3, which depicts as a histogram the fraction of the images with a given surface brightness. For a dynamic range of about 2.5 orders of magnitude, all three reconstructions predict the same fraction for the given surface brightness. For surface brightnesses lower than $50\,\mathrm{mJy/arcsec^2}$, only the two algorithms, including direction-dependent calibration, deliver consistent results. In contrast, the `resolve` reconstruction of Arras et al., 2021a utilizing classically calibrated data deviates significantly from the other two methods. For surface brightnesses of less than $1\,\mathrm{mJy/arcsec^2}$ also the two methods, including direction-dependent calibration, start to deviate. These low surface brightness regions lie outside of Cygnus A, and no significant

sources are detected there. The deviation in background intensity is due to the different priors used for the sky brightness. For interpreting the off-source residuals, the plot of the relative uncertainty of the `resolve` reconstruction in figure 6.5 is also helpful. As visible in figure 6.5, the relative uncertainty of the `resolve` reconstruction is larger than 1 in regions outside of Cygnus A, meaning that the reconstructed background intensity is statistically compatible with 0 in these regions.

The accurate agreement of the `resolve` reconstruction of this work and the compressed sensing reconstruction of Dabbech et al., 2021 can also be confirmed by the contour plots in figure 6.4. The contour lines, spanning a surface brightness range of 4.3 orders of magnitude, are in close agreement between the two reconstructions. In contrast, the flux contours of the `resolve` reconstruction from Arras et al., 2021a show significant deviations in the low surface brightness regions.

Thus, to conclude, ignoring direction-dependent effects in the calibration solutions limits the dynamic range of the recovered radio maps. The Bayesian `resolve` reconstructions utilizing direction-dependent calibration significantly improve the dynamic range, and the results are compatible with other work.

The geoVI algorithm used in the `resolve` framework outputs a set of posterior samples of the sky brightness (see section 6.4.4). The posterior mean sky brightness can be computed from these samples, as shown in the previous plots. Besides the mean, also other summary statistics of the sky brightness posterior distribution can be computed, such as the uncertainty. The left column of figure 6.5 depicts the relative uncertainty. In the regions of the hot spots of Cygnus A, the relative uncertainty is on the order of a few percent. Toward the fainter ends of the lobes, the uncertainty rises to 10% or even higher in the faintest regions. Outside the contours of Cygnus A, the relative uncertainty is significantly larger and exceeds 100%. This is expected since the surface brightness is negligibly small in these regions. Also, the `resolve` reconstruction of Arras et al., 2021a provides uncertainty estimates, which are compatible with the uncertainty estimates of this work. The compressed sensing method of Dabbech et al., 2021 cannot estimate the uncertainty.

The posterior distribution can also be analyzed based on individual samples. A set of four representative posterior samples is shown in figure 6.5 in the right column. While the samples show only tiny variations in high surface brightness regions, they are more diverse in regions with lower brightness, indicating higher posterior uncertainty.

In the appendix 6.8.2, we display the dirty image of the residuals between our reconstruction and the data. Overall the residuals are very small, and only minor structures of the source are left in the residuals, indicating a good fit to the data.

6.5.4 Calibration solutions

In this section, we briefly discuss the antenna gain solutions. As our antenna gain model (section 6.4.2) factorizes the antenna gains into a purely time-dependent term $g_p(t)$, an only direction-dependent $g_p(\boldsymbol{l})$, and a jointly direction and time-dependent term $g_p(\boldsymbol{l}, t)$, we can analyze the reconstruction of these terms separately. As it will turn out in the following sections, the statistical properties of the three factors are very different. As we model each of the factors with separate Gaussian processes, inferring their power spectra, the algorithm

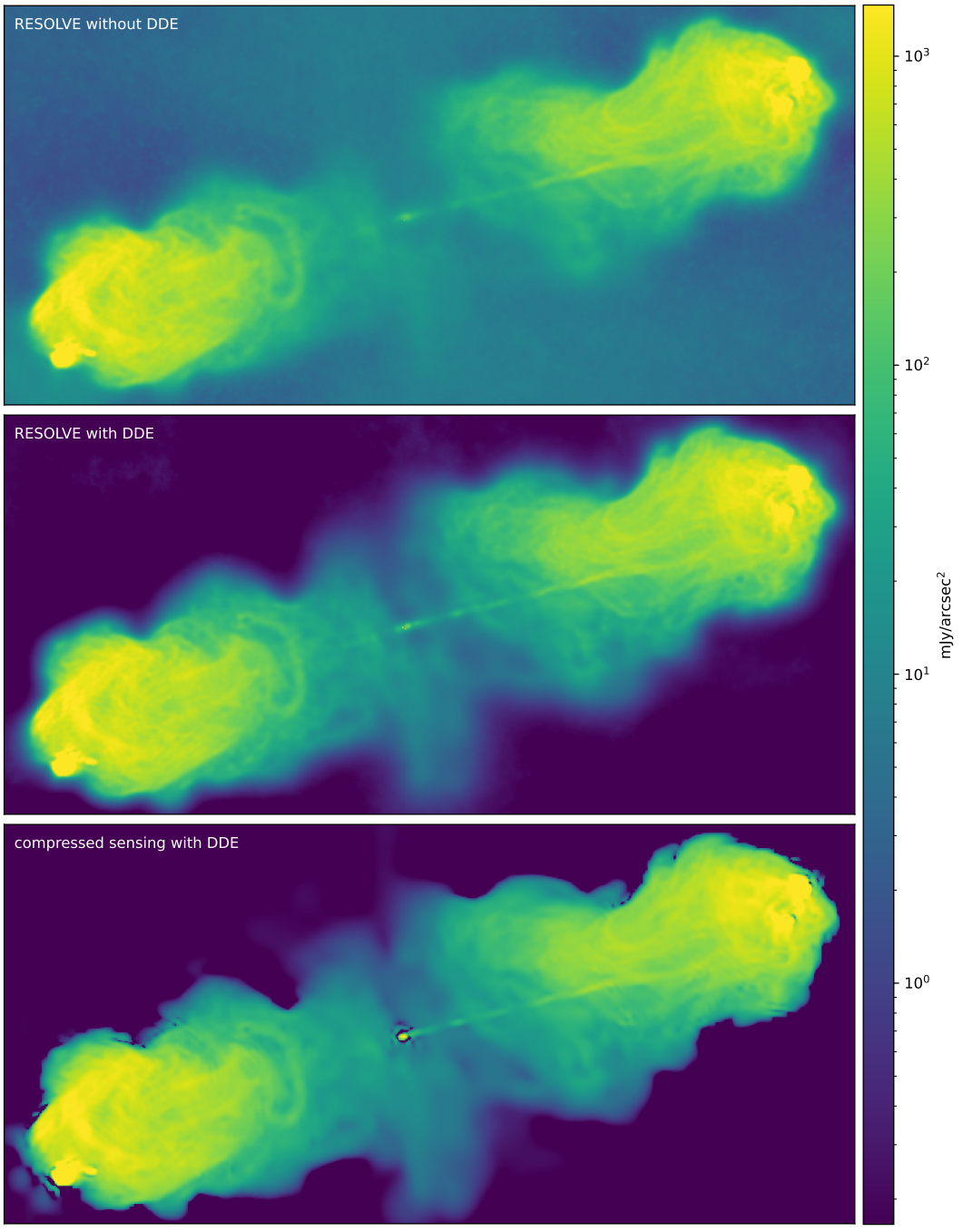

Figure 6.1: Comparison of Cygnus A images with three algorithms. Top panel: Cygnus A image from Arras et al., 2021a obtained with `resolve` without direction-dependent calibration. Central panel: The Cygnus A image reconstructed with the algorithm of this work. Bottom panel: Cygnus A image from Dabbech et al., 2021 recovered by a compressed sensing algorithm including direction-dependent calibration. The colorbar is clipped for better visibility.

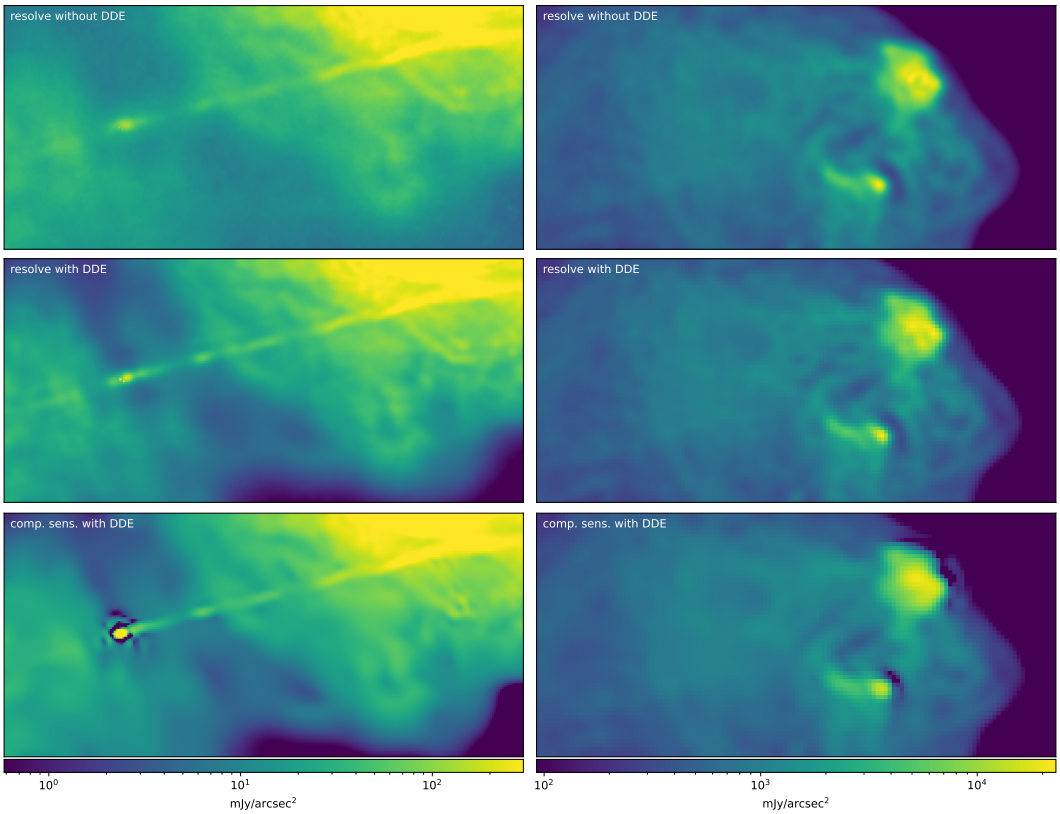

Figure 6.2: Zooms into high and low flux regions of the reconstructions. Left column: Zoom on the core and jet. Top panel: Arras et al., 2021a; Central panel: This work; Bottom panel: Dabbech et al., 2021. Right column: Zoom on the hot spots of western lobe. Top panel: Arras et al., 2021a; Central panel: This work; Bottom panel: Dabbech et al., 2021. The colorbars of the left and right columns are clipped and adapted to the respective surface brightness range.

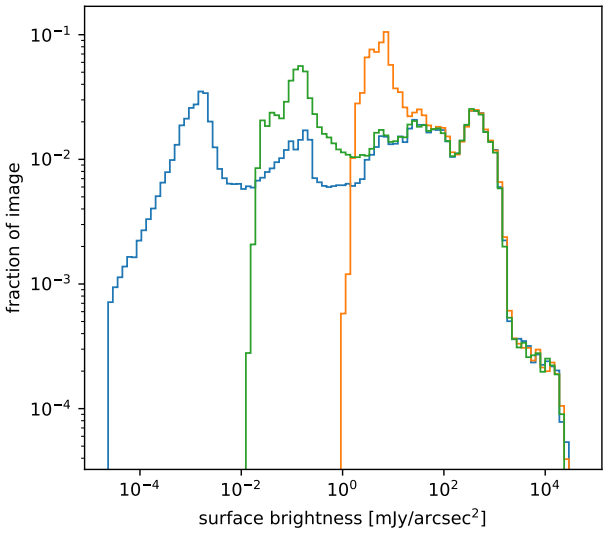

Figure 6.3: Histogram of surface brightness values of the `resolve` reconstruction with direction-dependent calibration (green), the `resolve` reconstruction of Arras et al., 2021a with classical calibration (orange), and the compressed sensing reconstruction from Dabbech et al., 2021 also with direction-dependent calibration (blue). The histogram is computed from the images displayed in figure 6.1 showing on the y-axis the fraction of the image with this surface brightness.

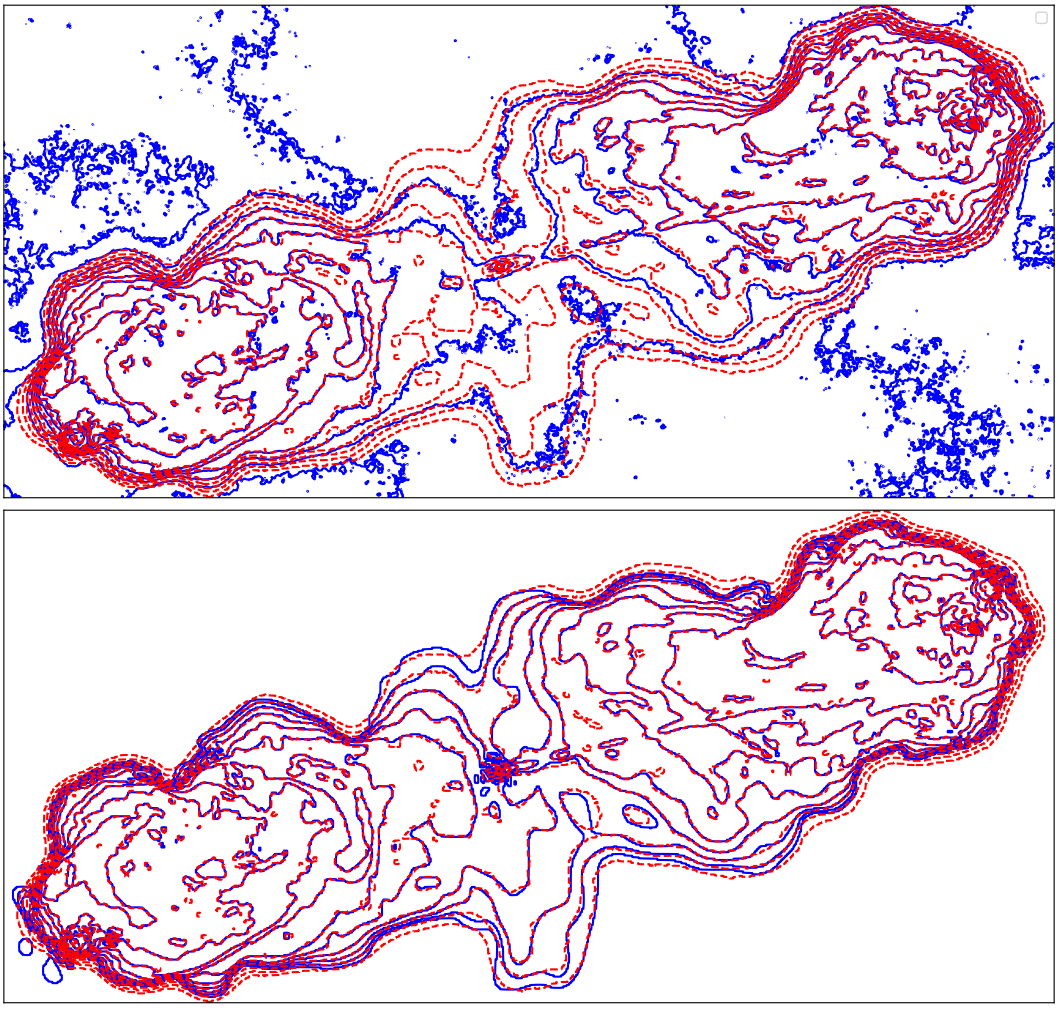

Figure 6.4: Flux contours of the Cygnus A reconstructions. The surface brightness levels of the contours are logarithmically spaced and range from $1\,\mathrm{mJy/arcsec^2}$ to $4.25 \cdot 10^4\,\mathrm{mJy/arcsec^2}$. Top panel: In solid blue `RESOLVE` reconstruction from Arras et al., 2021a without direction-dependent calibration. In dashed red `RESOLVE` reconstruction of this work with direction-dependent calibration. Bottom panel: In solid blue compressed sensing reconstruct of Dabbech et al., 2021 with direction-dependent calibration. In dashed red again `RESOLVE` reconstruction of this work with direction-dependent calibration.

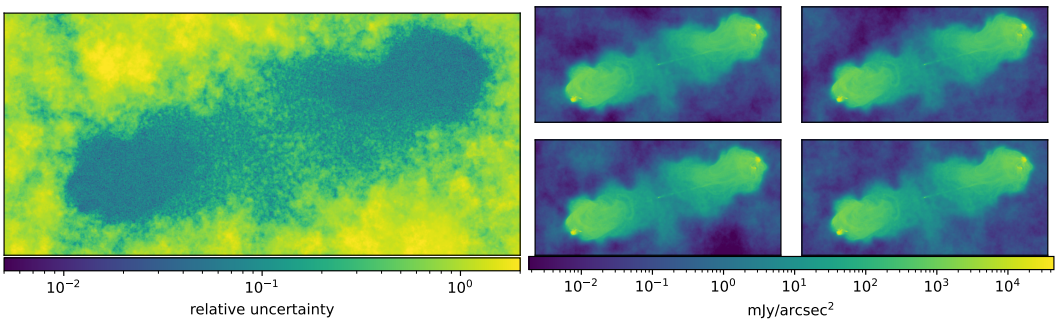

Figure 6.5: Uncertainty quantification of the `resolve` reconstruction. Left column: Relative uncertainty of the reconstruct sky brightness computed from posterior samples. Two right columns: Four representative posterior samples of the geoVI reconstructions.

can capture the different statistics and exploit them to further constrain the reconstructions.

6.5.4.1 Only time-dependent gains: $g_p(t)$

Figure 6.6 displays only time-dependent calibration solutions. Both the phase and the amplitude are largely stable over time, with some small-scale fluctuations. After approximately one-third of the observation, the amplitude of several antenna gains suddenly decreases before returning to the old value. The phases of the gains constantly undergo some variations.

In figure 6.7, the absolute value of the only time-dependent part of the calibration solution is depicted for one baseline, thus in formulas $\mathrm{abs}(g_p(t)g_q^*(t))$. To visualize the jumps in the calibration solutions discussed in the previous section, we display the absolute value of the visibilities of the flux calibrator as well as the science target Cygnus A projected to the calibration space. More precisely, we divide the visibilities of the respective observation by the model visibilities computed from the sky model without the only time-dependent calibration factor $g_p(t)g_q^*(t)$. Thus, as a formula,

$$V_{pqt}^{\mathrm{projected}} = \frac{V_{pqt}}{\int C(\boldsymbol{l}, w_{pqt}) I(\boldsymbol{l}) \widetilde{G}_{pt}(l,t) \widetilde{G}_{qt}^*(l,t) \mathrm{e}^{-2\pi i (\boldsymbol{k}_{pqt} \cdot \boldsymbol{l})} d\boldsymbol{l}}, \tag{6.16}$$

with V_{pqt} being the visibilities of the observation, $C(\boldsymbol{l})$ the w-term correction factor, $I(\boldsymbol{l})$ the reconstructed sky model, and

$$\widetilde{G}_{pt}(l,t)\widetilde{G}_{qt}^*(l,t) = g_p(\boldsymbol{l},t)g_q^*(\boldsymbol{l},t)g_p(\boldsymbol{l})g_q^*(\boldsymbol{l}) \tag{6.17}$$

the calibration solution excluding the only time-dependent part $g_p(t)g_q^*(t)$. The offset between the visibilities from the flux calibrator and the visibilities of the Cygnus A observation is clearly visible in figure 6.7, and is on the order of a few percent. The discontinuities in the calibration solution, as described in section 6.4.2, were introduced to model these offsets.

6.5.4.2 Only direction-dependent gains: $g_p(\boldsymbol{l})$

Figure 6.8 depicts the direction-dependent but time independent gain $g_p^{ll/rr}(\boldsymbol{l})$ of one antenna. The phase and amplitude variations are both on the level of a few percent. The interferometer is only sensitive to variations of the gain solutions in directions with significant sky surface brightness since the antenna gains are a multiplicative effect on the sky. Therefore, to indicate which regions of the reconstructed gains solutions actually affect the likelihood, we superimpose the flux contours of Cygnus A on the plot of the calibration solution.

6.5.4.3 Jointly time and direction-dependent gains: $g_p(\boldsymbol{l}, t)$

The absolute value of the jointly time and direction-dependent antenna gains $g_p^{ll/rr}(\boldsymbol{l}, t)$ from which the only time and only direction-dependent variations are excluded is displayed in figure 6.9. Similar to figure 6.8, the flux contours of Cygnus A are superimposed to indicate which regions actually have a significant effect. The variations of the jointly direction and time-dependent gains are less than a percent and therefore smaller than the only time and only direction dependent effects. The spatial correlation length is, compared to the only direction-dependent gain solutions, significantly larger. This result underlines that a split of the antenna gains into an only direction and a jointly direction and time-dependent part can be beneficial because the generative model for $g_p^{ll/rr}(\boldsymbol{l}, t)$ can exploit this longer correlation length.

A direct quantitative comparison with the work of Dabbech et al., 2021, as done for the Cygnus A sky images, is not possible for the calibration solutions, as the calibration model in Dabbech et al., 2021 is different and the results are only published in the form of images. Nevertheless, the scale of the amplitude and phase variations are comparable. Also, the correlation length of the jointly direction and time-dependent gain solutions is similar to the correlation length of the direction and time-dependent calibration solutions in Dabbech et al., 2021. However, in the work of Dabbech et al., 2021, the spatial and temporal correlation lengths are set somewhat externally by specifying the support size of the corresponding Fourier kernels. In contrast, in this work, the correlation structures are not fixed but reconstructed from the data.

Pointing errors are expected to be an important source of jointly time and direction-dependent gains. The antenna beam is falling off with increasing distance to the center of the field of view. This makes the sky to appear brighter in the direction of the pointing offset and dimmer in the opposite direction, leading to a direction-dependent antenna gain. Mathematically the direction-dependent antenna gain can be computed by

$$g_p(l, t) = \frac{B_p(\delta x_t, \delta y_t)}{B_p(0, 0)}, \tag{6.18}$$

with $B_p(\delta x_t)(\delta y_t)$ being the beam of antenna p with an pointing offset of $(\delta x_t, \delta y_t)$, and $B_p(0, 0)$ the beam of the nominal pointing. Using equation 6.18, we fit pointing offsets to the absolute value of the direction-dependent gains, resulting in pointing offsets of up to 30 arcsec as depicted in figure 6.10. As visible in figure 6.10, the gains resulting from pointing offsets relatively closely follow the reconstructed gain. Nevertheless, some deviations

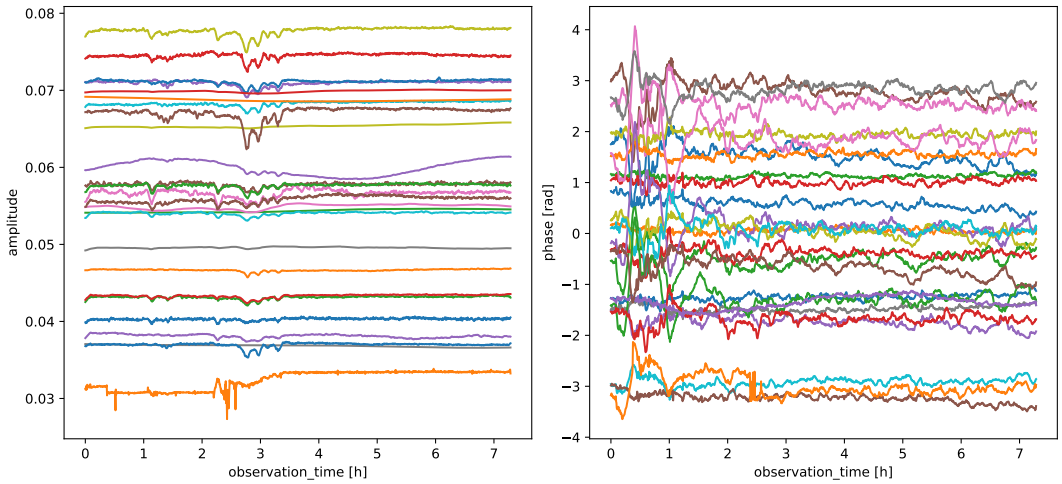

Figure 6.6: DIE amplitude (left panel) and phase (right panel) of the calibration solutions for all 26 antenna in VLA configuration A.

remain as pointing errors are not the only source for time and direction-dependent gains.

6.6 Conclusions

This chapter introduces a novel Bayesian radio interferometric imaging and calibration algorithm based on the `resolve` framework. Calibration and imaging are performed jointly. The calibration includes corrections for direction-dependent effects. To this goal, the antenna gain solutions are split into three factors, a purely time-dependent term, a purely direction-dependent term, and a jointly direction and time-dependent term. Having a separate model for each factor of the gain solutions allows one to exploit the different statistics of the individual factors.

We demonstrate the method on VLA's Cygnus A observation, a widely used target for benchmarking novel algorithms. In comparison with previous work (Arras et al., 2021a), imaging Cygnus A also with the `resolve` framework but classically calibrated data, the dynamic range is significantly increased. At the same time, in high and medium surface brightness regions, the two reconstructions are closely consistent. When comparing with Dabbech et al., 2021, a compressed sensing method including direction-dependent calibration, the results not only agree in high and medium surface brightness regions but also in areas with very low flux. As the reconstruction algorithms of the sky radio maps of this work and Dabbech et al., 2021 are based on very different methodologies, namely Bayesian inference and compressed sensing, their close agreement is a strong indication of their reliability even in low surface brightness regions. Furthermore, this underlines the necessity of an accurate calibration model for obtaining high-fidelity radio maps with a large dynamic range. An accurate calibration model will become a critical component for exploiting the full potential of upcoming radio interferometers.

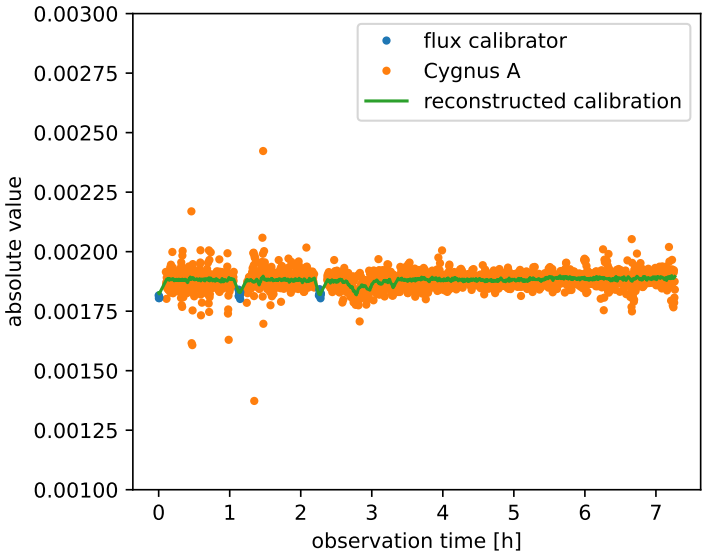

Figure 6.7: Green: Absolute value of the reconstructed time-dependent, but direction-independent, calibration for one baseline. The blue and orange dots are the absolute values of the visibilities of the flux calibrator and target observations, projected to the calibration space by dividing out their respective sky models. More specifically, blue dots: Flux calibrator visibilities divided by the model visibilities computed from the flux calibrator sky model without time-dependent calibration factors. Orange dots: Cygnus A visibilities divided by the model visibilities computed from the Cygnus A sky model, including the direction-dependent calibration but excluding the purely time-dependent calibration. The offset between the calibration data points and the target data points makes it necessary to allow for possible discontinuities in the antenna gain reconstructions when switching the observation target. Partly this offset might also be due to a oversimplified flux calibrator model.

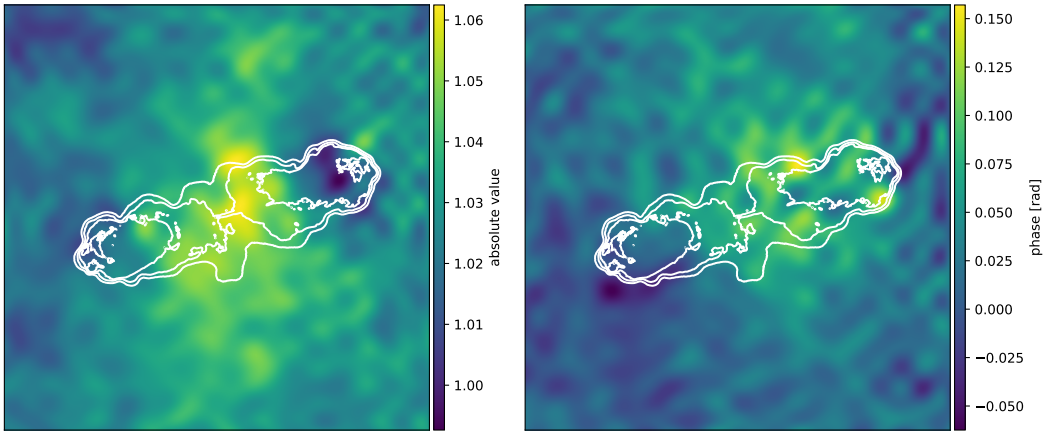

Figure 6.8: Phase and amplitude of the direction-dependent, time-independent gain factor for one antenna. On top the flux contours of Cygnus A are indicated in white, since the interferometer is only sensitive to direction-dependent effects in regions with significant surface brightness. Thus, gain values outside the contours are neither accurate nor relevant.

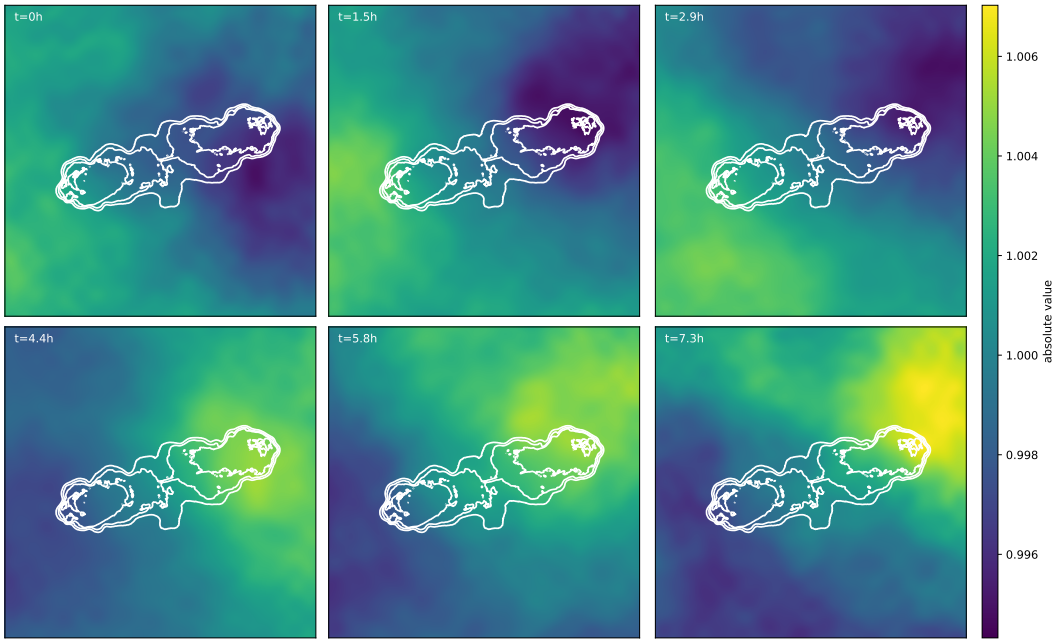

Figure 6.9: Absolute value of jointly time and direction-dependent gain of antenna 0, depicted at six timesteps of the observation. On top, the contours of Cygnus A are shown since the interferometer is only sensitive to the direction-dependent gain variations in regions with significant sky brightness. Gain values outside the contours are neither accurate nor relevant.

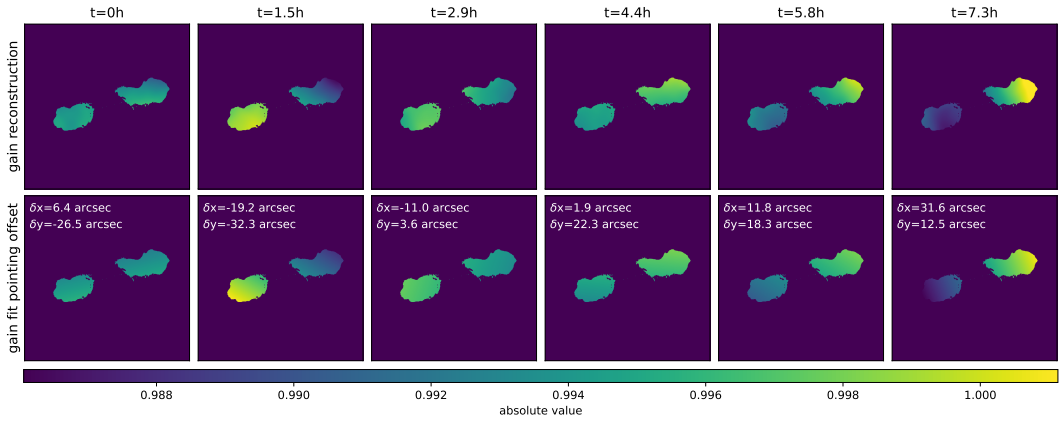

Figure 6.10: Top row: Absolute value of jointly time and direction-dependent gain of antenna 0 (same as figure 6.9) restricted to sky region with significant flux. Bottom: Fit of the antenna gain resulting from pointing errors to the reconstructed gain in the top row. Except for minor differences, the gains resulting from pointing errors can reproduce the reconstructed gains. The offset fit to the nominal pointing direction is specified as δx and δy. As Cygnus A is less extended in y-direction than in x-direction, the gains are much less sensitive to pointing offsets in y-direction.

The computational complexity of this method in its current version and the `resolve` framework, in general, is significantly higher compared with the `clean` algorithm. The higher computational costs partially arise from our inference algorithm (section 6.4.4) computing not only a single estimate of the sky brightness but providing samples of the posterior distribution allowing for statistically robust uncertainty quantification. Additionally to the computational overhead of the posterior inference algorithm, it is in the current version of `resolve` necessary to always degridd to the visibility space to evaluate the likelihood. Despite fast gridding algorithms, this can, depending on the interferometer setup, lead to significantly higher computational costs compared with the clean algorithm's major/ minor cycle scheme. We plan to also introduce a major/ minor cycle scheme to the `resolve` framework for speeding up the likelihood evaluations. Incorporating the major/ minor cycle scheme into a Bayesian imaging algorithm is nontrivial and is present in chapter 9 as the `fast-resolve` algorithm.

6.7 Acknowledgements

Jakob Roth and Philipp Arras acknowledge financial support by the German Federal Ministry of Education and Research (BMBF) under grant 05A20W01 (Verbundprojekt D-MeerKAT).

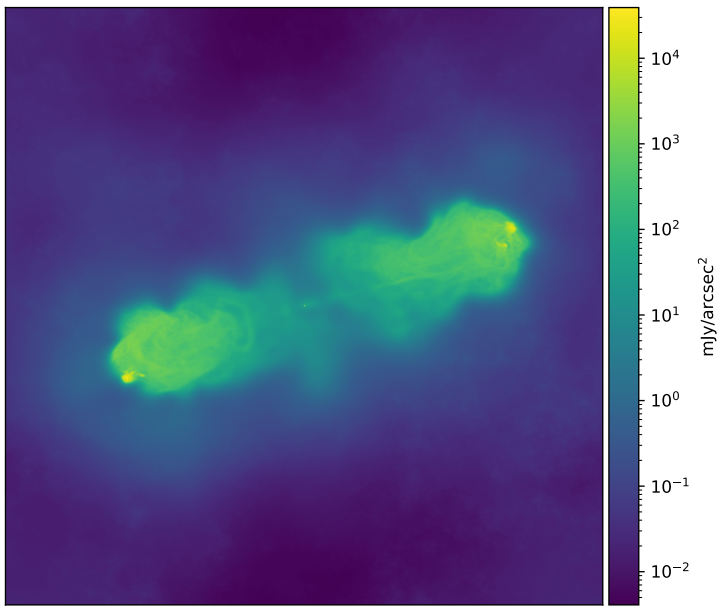

Figure 6.11: Cygnus A reconstruction with `resolve` using DIE calibration only.

6.8 Appendix

6.8.1 Only DIE calibration reconstruction

For completeness, we also reconstructed Cygnus A using the same imaging and calibration setup described above, except that we removed the direction-dependent calibration terms. Thus starting from the uncalibrated data, we jointly reconstructed Cygnus A and the calibration solutions, considering only direction-independent calibration. The result of this reconstruction is depicted in figure 6.11. Overall, the result of this reconstruction is similar to the reconstruction of Arras et al., 2021a (displayed in figure 6.1) building on classically calibrated data.

6.8.2 Residual image

We computed the dirty image of the residuals between the data and our reconstruction with direction-dependent calibration. As visible in figure 6.12, the amplitude of the residuals is very small compared with the brightness of Cygnus A. Nearly no structures of the source have imprinted on the residual image, indicating that the reconstructed sky model and antenna gains model the data very well. The only structures of Cygnus A of which some remnants can be found in the residual image are the hot spots of the lobes and the jet. Interestingly remnants of the jet are not only visible at the actual position of the jet but also at various shifted positions. We believe this originates from the nature of our prior model for diffuse emission. As outlined in section 6.4.1.1, we model the diffuse emission with a Gaussian process represented in Fourier space. The Fourier mode corresponding to the width and orientation of the jet should be strongly exited, leading to an oscillation with this

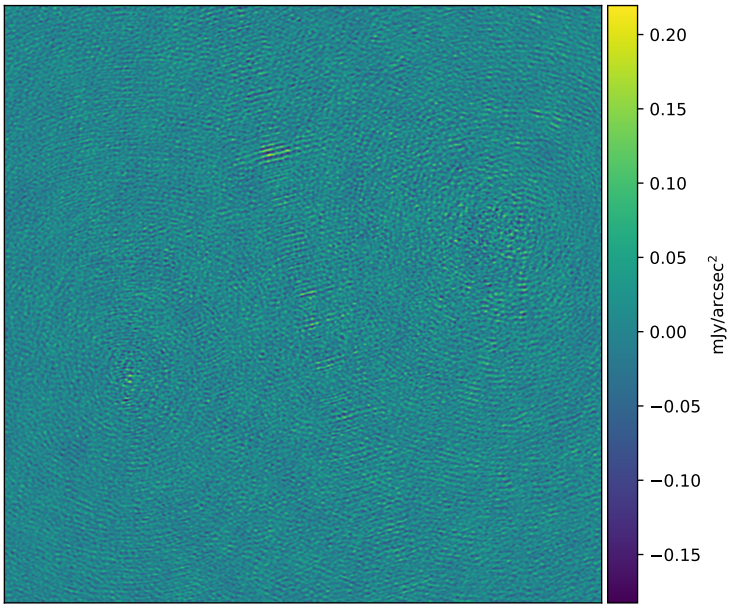

Figure 6.12: Dirty image of the residuals between our reconstruction and the actual data.

width and orientation in the entire field of view. Nevertheless, the amplitude of this effect is very small compared to the brightness of Cygnus A and its jet. Therefore, this oscillatory effect cannot be seen in the reconstructions (figure 6.1).

Dabbech et al., 2021 reported the detection of three additional sources in the residual image further away from Cygnus A. We cannot confirm these sources as they are not found in our residual image.

Chapter 7

Computational aspects of radio interferometric imaging

7.1 Bayesian imaging with machine learning frameworks

Computationally and conceptually, the imaging algorithms presented so far in this thesis (chapters 5 and 6) have many similarities with classic machine learning methods. A detailed comparison can be found in Enßlin, 2022. Computationally, both machine learning methods and Bayesian image inference minimize a scalar-valued cost function. In classical machine learning, the minimized objective function is typically called "loss function", while in the Bayesian inference method presented here, the Kullback-Leibler divergence (see chapter 4) between the true posterior and the approximation is minimized. Furthermore, in both cases, the input domain of the cost function is high dimensional, and the evaluation of the cost function requires a large number of floating point operations.

To address these computational challenges, dedicated programming frameworks, such as TensorFlow, PyTorch, or JAX, have been developed for machine learning. While being easy to use, these frameworks can automatically differentiate the cost function, allowing the use of efficient gradient-based minimizers for the optimization. In addition, these frameworks allow to seamlessly run the code on GPUs for a fast execution of a large number of floating point operations.

To take advantage of modern machine learning programming frameworks for Bayesian inference, NIFTy has been reimplemented in JAX. An overview of the reimplemented NIFTy package is given in Edenhofer et al., 2024a. While most Bayesian inference algorithms based on NIFTy can easily switch to the JAX-based version, this is more challenging for radio interferometry and the `resolve` framework because the numerical evaluation of the radio interferometric measurement equation (chapter 3, equation 3.27) is challenging. Many high-performance C++ and CUDA codes, such as Arras et al., 2021b; Tol et al., 2018 or the implementation for direction-dependent calibration in chapter 6, have been developed for an efficient evaluation of equation 3.27. A native reimplementation of these codes in a machine learning framework like JAX is not easy, if not impossible, as JAX does not allow

manual low-level optimization. For this reason, we developed the JAXbind package[1] (Roth et al., 2024b) allowing to easily bind high-performance codes into the JAX framework and connecting them with the automatic differentiation engine of JAX. In chapter 8, the package is presented. Building on JAXbind, it is possible to connect the `resolve` frameworks to the JAX version of NIFTy. JAXbind has also been used in other contexts, such as binding spherical harmonic transforms of the DUCC library (Reinecke, 2024) to JAX and binding the PICARD multi-grid solver (Kissmann, 2014) to JAX.

7.2 Imaging with large arrays

Section 3.2.1 introduced radio interferometers as a technique to overcome the resolution limits of single-dish radio telescopes. The largest distance between a pair of antennas determines the smallest angular scale to which the interferometer is sensitive. Similarly, the shortest baseline determines the largest angular scale that the interferometer can probe. Many baselines with intermediate lengths are required to image all angular scales in between. The interferometer's sensitivity, thus its ability to detect faint emission, scales with the total collecting area of the antennas. Thus, both resolution and sensitivity ultimately require a large number of antennas.

The number of baselines N_B, and therefore the data size, scales quadratically with the number of antennas N_A and is given by

$$N_B = \sum_{i=1}^{N_A-1} i = \frac{N_A(N_A-1)}{2}. \tag{7.1}$$

The Very Large Array (VLA), constructed in the 1970s, consists of 27 antennas and consequently forms 351 baselines. Newer arrays often have more antennas for the reasons mentioned above. For example, the Australian Square Kilometre Array Pathfinder (ASKAP) consists of 36 antennas, and the Atacama Large Millimeter/Submillimeter Array (ALMA) has 66 antennas. MeerKAT, the most advanced array at GHz frequencies, currently has 64 (and soon 80) antennas, resulting in 2016 and, in the future, 3160 baselines. Next generation instruments such as the Square Kilometre Array (SKA) and the next generation Very Large Array (ng-VLA) will have hundreds of antennas and well over 10000 baselines. The DSA-2000 will have, if it is built in its full extension, 2000 antennas leading to enormous 1999000 baselines. In addition to the increase in the number of antennas, advancements in receiver technology, which increase the number of frequency channels, also lead to growing data sizes.

The development of ever larger arrays starts a new era in radio astronomy, vastly increasing the observational capabilities of radio interferometers. Besides the need to deploy and operate many antennas, this new era will dramatically increase the size of the measurement data. The increase in data size is a challenge for all stages of the data processing. Many high-accuracy imaging algorithms, such as the `resolve` framework, need to evaluate the measurement response (equation 3.27) much more often than the standard imaging

[1] `https://github.com/NIFTy-PPL/JAXbind`

method named `CLEAN`. Since the measurement response maps to the data space, the evaluations become computationally increasingly demanding as the number of baselines grows. Therefore, many advanced imaging algorithms are particularly affected by the increase in data volume. With the `resolve` framework, imaging VLA data as presented in chapter 6 or Arras et al., 2021a already requires significant computational resources. Consequently, imaging MeerKAT data over a broad bandwidth with many frequency channels is computationally infeasible for the `resolve` framework, and imaging data from next-generation interferometers is even further out of reach with `resolve`.

The `resolve` framework not only increases the computational costs but also significantly increases the retrieved images' accuracy. The previous chapter demonstrated the high accuracy of `resolve` with direction-dependent calibration compared to another advanced imaging algorithm. Compared to the standard algorithm `CLEAN`, the advantages of `resolve` in terms of accuracy are even larger, as presented in Arras et al., 2021a. Figure 7.1 shows a zoom on the eastern hot spot of Cygnus A from the `resolve` and `CLEAN` reconstructions of Arras et al., 2021a at different frequencies. As clearly visible, the resolution of the `resolve` reconstructions is significantly higher. Thus, the `CLEAN` algorithm computationally outperforms many imaging algorithms, but the accuracy of `CLEAN` results is fairly suboptimal. For this reason, the design of scalable and accurate imaging algorithms is a high priority in radio interferometric algorithm development.

The following two chapters outline developments contributing to fast accurate imaging in radio interferometry. Chapter 8 presents the `JAXbind` package, allowing to bind a high-performance implementation of the radio interferometry response into the JAX framework, as already discussed in the previous section. Chapter 9 introduces the `fast-resolve` algorithm. `fast-resolve` is a redesign of the classic `resolve` algorithm and aims to combine the computational efficiency of `CLEAN` with the accuracy of `resolve`. As presented in chapter 9, `fast-resolve` enabled a Bayesian reconstruction of a large MeerKAT dataset. To the best of my knowledge so far, no other Bayesian radio interferometric imaging algorithm ever reconstructed a similar-sized dataset.

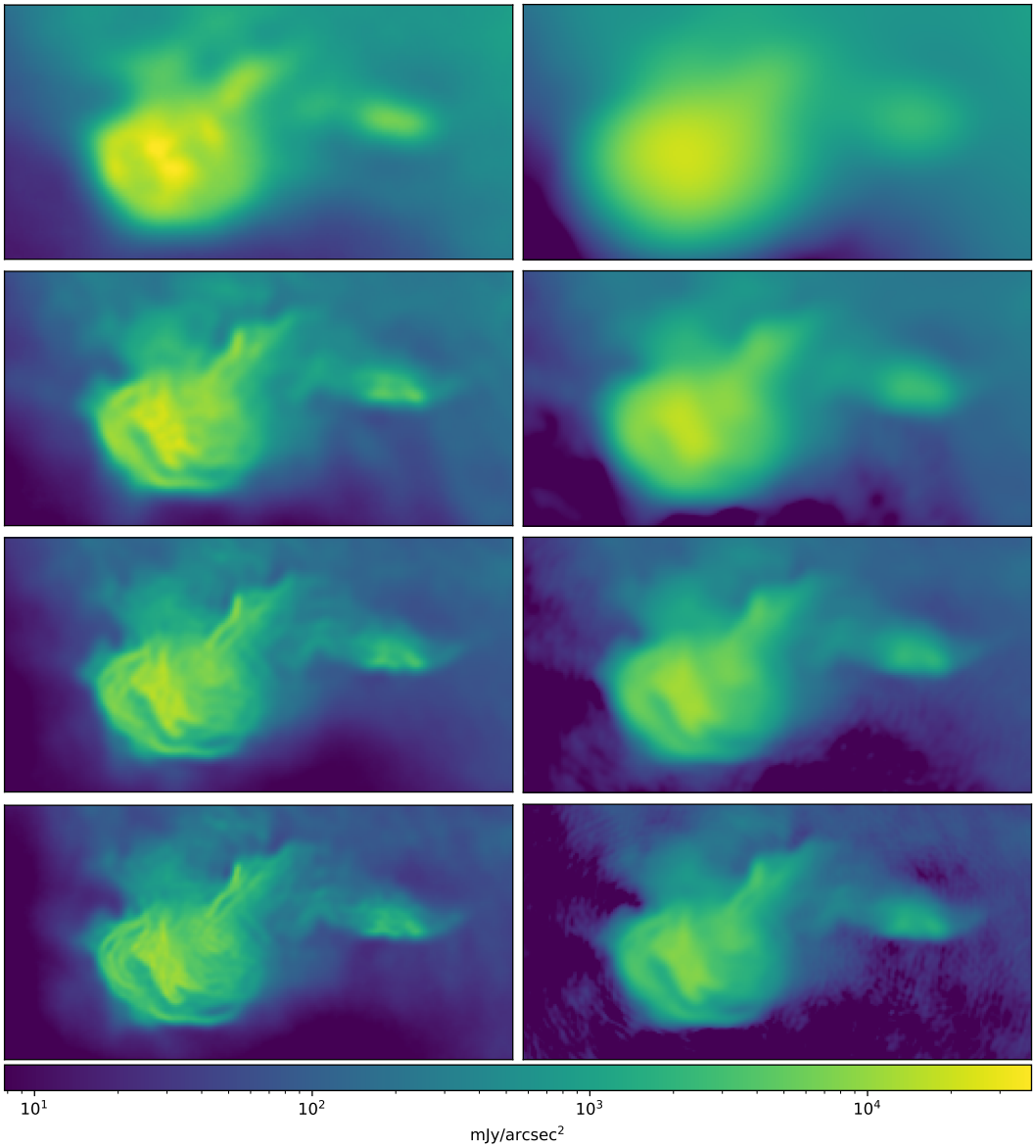

Figure 7.1: Reconstruction of the eastern hot spot of Cygnus A with `resolve` and `CLEAN` at different frequencies from Arras et al., 2021a. The left column shows `resolve`, and the right column shows `CLEAN`. The imaging frequencies are from top to bottom 2052, 4811, 8427, and 13360 MHz. As clearly visible, `resolve` can recover at a given frequency a higher resolution image than `CLEAN`.

Chapter 8

JAXbind: bind any function to JAX

I have published this chapter in the Journal of Open Source Software as an article (Roth et al., 2024b). This article is published with equal author contributions. Specifically, I coded most of the interfacing with the JAX automatic differentiation engine and an initial version of the binding of the C++ code to the JAX XLA backend. Gordian Edenhofer coded most of the functionality for mapping over the input axis. Gordian Edenhofer and I jointly designed the scheme for handling functions with an arbitrary number of arguments. Martin Reinecke contributed most of the C++ code. Gordian Edenhofer wrote an initial version of the Statement of Need section. I wrote the initial version of the rest of the paper. Afterwards, all coauthors contributed to improving the text. For this thesis, the text was slightly adapted.

8.1 Summary

JAX is widely used in machine learning and scientific computing, the latter of which often relies on existing high-performance code that we would ideally like to incorporate into JAX. Reimplementing the existing code in JAX is often impractical and the existing interface in JAX for binding custom code either limits the user to a single Jacobian product or requires deep knowledge of JAX and its C++ backend for general Jacobian products. With `JAXbind` we drastically reduce the effort required to bind custom functions implemented in other programming languages with full support for Jacobian-vector products and vector-Jacobian products to JAX. Specifically, `JAXbind` provides an easy-to-use Python interface for defining custom, so-called JAX primitives. Via `JAXbind`, any function callable from Python can be exposed as a JAX primitive. `JAXbind` allows a user to interface the JAX function transformation engine with custom derivatives and batching rules, enabling all JAX transformations for the custom primitive.

8.2 Statement of need

The use of JAX (Bradbury et al., 2018) is widespread in the natural sciences. Of particular interest is JAX's powerful transformation system. It enables a user to retrieve arbitrary derivatives of functions, batch computations, and just-in-time compile code for additional

performance. Its transformation system requires that all components of the computation are written in JAX.

A plethora of high-performance code is not written in JAX and thus not accessible from within JAX. Rewriting these codes is often infeasible and/or inefficient. Ideally, we would like to mix existing high-performance code with JAX code. However, connecting code to JAX requires knowledge of the internals of JAX and its C++ backend.

In this paper, we present `JAXbind`,[1] a package for bridging any function to JAX without in-depth knowledge of JAX's transformation system. The interface is accessible from Python without requiring any development in C++. The package is able to register any function and its partial derivatives and their transpose functions as a JAX native call, a so-called primitive.

We believe `JAXbind` to be highly useful in scientific computing. We intend to use this package to connect the Hartley transform and the spherical harmonic transform from DUCC (Reinecke, 2024) to the probabilistic programming package NIFTy (Edenhofer et al., 2024a) as well as the radio interferometry response from DUCC with the radio astronomy package `resolve`. Furthermore, we intend to connect the non-uniform FFT from DUCC with JAX for applications in strong-lensing astrophysics. We envision many further applications within and outside of astrophysics.

The functionality of `JAXbind` extends the external callback functionality in JAX. Currently, `JAXbind`, akin to the external callback functions in JAX, briefly requires Python's global interpreter lock (GIL) to call the user-specified Python function. In contrast to JAX's external callback functions, `JAXbind` allows for both a custom Jacobian-vector product and vector-Jacobian product. To the best of our knowledge no other code currently exists for easily binding generic functions and both of their Jacobian products to JAX, without the need for C++ or LLVM. The package that comes the closest is Enzyme-JAX (Moses et al., 2024), which allows one to bind arbitrary LLVM/MLIR, including C++, with automatically-generated (Moses et al., 2020; Moses et al., 2021; Moses et al., 2022) or manually-defined derivatives to JAX.

PyTorch (Ansel et al., 2024) and TensorFlow (Abadi et al., 2015) also provide interfaces for custom extensions. PyTorch has an extensively documented Python interface[2] for wrapping custom Python functions as PyTorch functions. This interface connects the custom function to PyTorch's automatic differentiation engine, allowing for custom Jacobian and Jacobian transposed applications, similar to what is possible with JAXbind. Additionally, PyTorch allows a user to interface its C++ backend with custom C++ or CUDA extensions.[3] JAXbind, in contrast, currently only supports functions executed on the CPU, although the JAX built-in C++ interface also allows for custom GPU kernels. TensorFlow includes a C++ interface[4] for custom functions that can be executed on the CPU or GPU. Custom gradients can be added to these functions.

[1] `https://github.com/NIFTy-PPL/JAXbind`

[2] `https://pytorch.org/docs/stable/notes/extending.html` (last visited 05-09-24)

[3] `https://pytorch.org/tutorials/advanced/cpp_extension.html` (last visited 05-09-24)

[4] `https://www.tensorflow.org/guide/create_op` (last visited 05-09-24)

8.3 Automatic differentiation and code example

Automatic differentiation is a core feature of JAX and often one of the main reasons for using it. Thus, it is essential that custom functions registered with JAX support automatic differentiation. In the following, we will outline which functions our package requires to enable automatic differentiation via JAX. For simplicity, we assume that we want to connect the nonlinear function $f(x_1, x_2) = x_1 x_2^2$ to JAX. The `JAXbind` package expects the Python function for f to take three positional arguments. The first argument, `out`, is a `tuple` into which the results are written. The second argument is also a `tuple` containing the input to the function, in our case, x_1 and x_2. Via `kwargs_dump`, any keyword arguments given to the registered JAX primitive can be forwarded to f in a serialized form.

```
import jaxbind

def f(out, args, kwargs_dump):
    kwargs = jaxbind.load_kwargs(kwargs_dump)
    x1, x2 = args
    out[0][()] = x1 * x2**2
```

JAX's automatic differentiation engine can compute the Jacobian-vector product `jvp` and vector-Jacobian product `vjp` of JAX primitives. The Jacobian-vector product in JAX is a function applying the Jacobian of f at a position x to a tangent vector. In mathematical nomenclature this operation is called the pushforward of f and can be denoted as $\partial f(x) : T_x X \mapsto T_{f(x)} Y$, with $T_x X$ and $T_{f(x)} Y$ being the tangent spaces of X and Y at the positions x and $f(x)$. As the implementation of f is not JAX native, JAX cannot automatically compute the `jvp`. Instead, an implementation of the pushforward has to be provided, which `JAXbind` will register as the `jvp` of the JAX primitive of f. For our example, this Jacobian-vector-product function is given by $\partial f(x_1, x_2)(dx_1, dx_2) = x_2^2 dx_1 + 2x_1 x_2 dx_2$.

```
def f_jvp(out, args, kwargs_dump):
    kwargs = jaxbind.load_kwargs(kwargs_dump)
    x1, x2, dx1, dx2 = args
    out[0][()] = x2**2 * dx1 + 2 * x1 * x2 * dx2
```

The vector-Jacobian product `vjp` in JAX is the linear transpose of the Jacobian-vector product. In mathematical nomenclature this is the pullback $(\partial f(x))^T : T_{f(x)} Y \mapsto T_x X$ of f. Analogously to the `jvp`, the user has to implement this function as JAX cannot automatically construct it. For our example function, the vector-Jacobian product is $(\partial f(x_1, x_2))^T (dy) = (x_2^2 dy, 2x_1 x_2 dy)$.

```
def f_vjp(out, args, kwargs_dump):
    kwargs = jaxbind.load_kwargs(kwargs_dump)
    x1, x2, dy = args
    out[0][()] = x2**2 * dy
    out[1][()] = 2 * x1 * x2 * dy
```

To just-in-time compile the function, JAX needs to abstractly evaluate the code, i.e., it needs to be able to infer the shape and dtype of the output of the function given only

the shape and dtype of the input. We have to provide these abstract evaluation functions returning the output shape and dtype given an input shape and dtype for f as well as for the `vjp` application. The output shape of the `jvp` is identical to the output shape of f itself and does not need to be specified again. The abstract evaluation functions take normal positional and keyword arguments.

```
def f_abstract(*args, **kwargs):
    assert args[0].shape == args[1].shape
    return ((args[0].shape, args[0].dtype),)

def f_abstract_T(*args, **kwargs):
    return (
        (args[0].shape, args[0].dtype),
        (args[0].shape, args[0].dtype),
    )
```

We have now defined all ingredients necessary to register a JAX primitive for our function f using the `JAXbind` package.

```
f_jax = jaxbind.get_nonlinear_call(
    f, (f_jvp, f_vjp), f_abstract, f_abstract_T
)
```

`f_jax` is a JAX primitive registered via the `JAXbind` package supporting all JAX transformations. We can now compute the `jvp` and `vjp` of the new JAX primitive and even jit-compile and batch it.

```
import jax
import jax.numpy as jnp

inp = (jnp.full((4,3), 4.), jnp.full((4,3), 2.))
tan = (jnp.full((4,3), 1.), jnp.full((4,3), 1.))
res, res_tan = jax.jvp(f_jax, inp, tan)

cotan = [jnp.full((4,3), 6.)]
res, f_vjp = jax.vjp(f_jax, *inp)
res_cotan = f_vjp(cotan)

f_jax_jit = jax.jit(f_jax)
res = f_jax_jit(*inp)
```

8.4 Higher order derivatives and linear functions

JAX supports higher order derivatives and can differentiate a `jvp` or `vjp` with respect to the position at which the Jacobian was taken. Similar to first derivatives, JAX can not automatically compute higher derivatives of a general function f that is not natively implemented in JAX. Higher order derivatives would again need to be provided by the user. For many

algorithms, first derivatives are sufficient, and higher order derivatives are often not implemented by high-performance codes. Therefore, the current interface of `JAXbind` is, for simplicity, restricted to first derivatives. In the future, the interface could be easily expanded if specific use cases require higher order derivatives.

In scientific computing, linear functions such as, e.g., spherical harmonic transforms are widespread. If the function f is linear, differentiation becomes trivial. Specifically for a linear function f, the pushforward or `jvp` of f is identical to f itself and independent of the position at which it is computed. Expressed in formulas, $\partial f(x)(dx) = f(dx)$ if f is linear in x. Analogously, the pullback or `vjp` becomes independent of the initial position and is given by the linear transpose of f, thus $(\partial f(x))^T(dy) = f^T(dy)$. Also, all higher order derivatives can be expressed in terms of f and its transpose. To make use of these simplifications, `JAXbind` provides a special interface for linear functions, supporting higher order derivatives, only requiring an implementation of the function and its transpose.

8.5 Platforms

Currently, `JAXbind` only supports primitives that act on CPU memory. In the future, GPU support could be added, which should work analogously to the CPU support in most respects. The automatic differentiation in JAX is backend agnostic and would thus not require any additional bindings to work on the GPU.

8.6 Acknowledgements

We would like to thank Dan Foreman-Mackey for his detailed guide (https://dfm.io/posts/extending-jax/) on connecting C++ code to JAX. Jakob Roth acknowledges financial support from the German Federal Ministry of Education and Research (BMBF) under grant 05A23WO1 (Verbundprojekt D-MeerKAT III). Gordian Edenhofer acknowledges support from the German Academic Scholarship Foundation in the form of a PhD scholarship ("Promotionsstipendium der Studienstiftung des Deutschen Volkes").

Chapter 9

`fast-resolve`: fast bayesian radio interferometric imaging

I have submitted this chapter as first author to the journal Astronomy & Astrophysics as an article. Currently, this article is under review. A preprint is available on arXiv (Roth et al., 2024a). Philipp Frank contributed via numerous discussions on the response and noise approximations. Herzog L. Bester contributed via many discussions and his profound knowledge of the MeerKAT instrument and the MeerKAT data we use for benchmarking. Oleg Smirnov contributed by calibrating the MeerKAT data. Torsten Enßlin contributed via many in-depth discussions on Bayesian inference. The text was read, commented and approved by all coauthors. For this thesis, the text and figures were slightly adapted.

9.1 Abstract

Interferometric imaging is algorithmically and computationally challenging as there is no unique inversion from the measurement data back to the sky maps, and the datasets can be very large. Many imaging methods already exist, but most of them focus either on the accuracy or the computational aspect. This chapter aims to reduce the computational complexity of the Bayesian imaging algorithm `resolve`, enabling the application of Bayesian imaging for larger datasets. By combining computational shortcuts of the `CLEAN` algorithm with the Bayesian imaging algorithm `resolve` we developed an accurate and fast imaging algorithm which we name `fast-resolve`. We validate the accuracy of the presented `fast-resolve` algorithm by comparing it with results from `resolve` on VLA Cygnus A data. Furthermore, we demonstrate the computational advantages of `fast-resolve` on a large MeerKAT ESO 137-006 dataset which is computationally out of reach for `resolve`. The presented algorithm is significantly faster than previous Bayesian imaging algorithms, broadening the applicability of Bayesian interferometric imaging. Specifically for the single channel VLA Cygnus A datasets `fast-resolve` is about 144 times faster than `resolve`. For the MeerKAT dataset with multiple channels the computational speedup of `fast-resolve` is even larger.

9.2 Introduction

Interferometric imaging is a versatile technique in astronomy that allows us to achieve enormous sensitivity and resolution by combining multiple telescopes. The effective resolution is roughly equivalent to that of a single telescope with a diameter equal to the largest distance between individual stations of the interferometer, which can be thousands of kilometers. For the upcoming Square Kilometer Array (Labate et al., 2022; Swart et al., 2022), the total collecting area might eventually approach one square kilometer, resulting in superior sensitivity compared to any single-dish telescope. Overcoming the resolution and sensitivity limitations of single telescopes comes at the cost of making it more difficult to retrieve images from the observational data.

Mathematically, the recovery of the sky images from interferometric data can be formulated as an inverse problem. The forward relation that computes the corresponding measurement data from a given sky image is known as the radio interferometric measurement equation (Smirnov, 2011a). As discussed in detail in chapter 3 and recapitulated in section 9.3, the measured data points are essentially noisy and undersampled Fourier modes of the sky image. This makes direct inversion of the measurement equation to obtain the sky image impossible, turning radio interferometric imaging into an ill-posed inverse problem.

Solving the inverse problem of radio interferometric imaging requires sophisticated algorithms that impose additional constraints on the sky brightness and regularize possible solutions to it. Historically, CLEAN (Högbom, 1974; Schwab et al., 1983) has been by far the most widely used imaging algorithm because it is computationally efficient, simple, and easy to use. Over the last decades, CLEAN-based imaging algorithms have been significantly improved, especially for diffuse emission imaging, spectral imaging, and wide-field imaging (Bhatnagar et al., 2004; Cornwell et al., 2008; Rau et al., 2011; Offringa et al., 2014; Offringa et al., 2017). However, CLEAN-based algorithms have several drawbacks, such as limited image fidelity, suboptimal resolution of recovered images, and lack of uncertainty quantification (see eg. Arras et al., 2021a for a detailed discussion). To improve these limitations, many other imaging algorithms have been developed.

A large class of new imaging algorithms builds on applying compressed sensing techniques to astronomical data (Wiaux et al., 2009). Several incarnations of such algorithms have shown significant improvements in terms of image fidelity and resolution over CLEAN based reconstructions. Recent examples are Dabbech et al., 2018; Abdulaziz et al., 2019; Dabbech et al., 2021 utilizing sparsity based regularizers in combination with convex optimization. Birdi et al., 2018; Birdi et al., 2020 extended these approaches to full polarization imaging. In Repetti et al., 2019 and Liaudat et al., 2023, some form of uncertainty quantification was added. Dabbech et al., 2022; Thouvenin et al., 2023 parallelized the image regularizer to distribute it over multiple CPUs. Terris et al., 2023a; Dabbech et al., 2022 presented a plug-and-play algorithm combining convex optimization with a deep-learning regularizer. Terris et al., 2023b further explored neural network regularizers and studied the influence of the training dataset on the regularizer.

Bayesian imaging algorithms are another important class of imaging algorithms addressing the uncertainty quantification problem. Early Bayesian imaging approaches such as Cornwell et al., 1985 built on the maximum entropy principle. Other Bayesian imaging

algorithms such as Sutton et al., 2006; Sutter et al., 2014; Cai et al., 2018; Tiede, 2022 rely on posterior sampling techniques. The Bayesian imaging framework `resolve`[1] originally proposed by Junklewitz et al., 2016 builds on variational inference instead of sampling based techniques for posterior approximation, reducing the computational costs. `resolve` has already successfully been applied to VLA, EHT, VLBA, GRAVITY, and ALMA data, providing high-quality radio maps with superior resolution compared to CLEAN, as well as an uncertainty map. Recent examples are Arras et al., 2019b joining Bayesian imaging with calibration. In Arras et al., 2021a `resolve` is compared with CLEAN, while in Roth et al., 2023b (chapter 6), direction-dependent calibration is added and compared with sparsity based imaging results of Dabbech et al., 2021. In GRAVITY Collaboration et al., 2022 `resolve` is applied to the optical interferometer GRAVITY. In Arras et al., 2022 `resolve` is applied to EHT data.

The improvements in imaging algorithms mentioned above come at the cost of increased computational complexity. For Very Long Baseline Interferometer (VLBI) observations, data sizes are typically small and the higher computational cost of advanced imaging methods is unproblematic. However, for large arrays such as MeerKAT (Jonas et al., 2016), this limits the applicability of many of these algorithms. For this reason, only very few advanced imaging algorithms have so far been successfully applied to data sets the size of a typical MeerKAT observation. Dabbech et al., 2022 presents an application of imaging algorithms using sparsity based and neural network regularizers to the MeerKAT ESO 137-006 observation. To handle the enormous computational complexity, Dabbech et al., 2022 massively parallelize their imaging algorithm and distribute it across a high-performance computing system, requiring hundreds to thousands of CPU hours to converge. In Thouvenin et al., 2023, this approach of parallelizing the regularizer is extended to the spectral domain.

To the best of our knowledge, no Bayesian radio interferometric imaging algorithm has so far been applied to a similar sized dataset. The goal of the `resolve` framework has always been to enable Bayesian imaging for a wide range of radio telescopes, not just VLBI observations. However, for instruments such as MeerKAT, imaging with `resolve` becomes computationally prohibitively expensive. In this chapter, we present the algorithm `fast-resolve`, which significantly reduces the computational complexity of classical `resolve` and enables Bayesian image reconstruction for large data sets. Greiner et al., 2016 previously already named an imaging method fast`RESOLVE`. The goal of the fast`RESOLVE` algorithm of Greiner et al., 2016 is the same as the new incarnation of `fast-resolve` presented in this chapter. However, although the new and the old variants of `fast-resolve` share some ideas, there are significant differences. In section 9.4.4, we will highlight the common concepts between the old the and the new `fast-resolve`. Note that the framework behind both `fast-resolve` and `resolve` has evolved significantly since the time of Greiner et al., 2016.

Machine learning-based imaging algorithms are a third relatively new class of imaging algorithms. Once the underlying machine learning models are trained, the computational cost for imaging is often lower than for other algorithms. Nevertheless, assessing image fidelity is challenging in such a framework given the lack of interpretability of machine learning models. Recent examples of such algorithms are Aghabiglou et al., 2024; Aghabiglou

[1] https://gitlab.mpcdf.mpg.de/ift/resolve

et al., 2023; Connor et al., 2022; Schmidt et al., 2022.

The remainder of the chapter is organized as follows. In section 9.3, we discuss the radio interferometric measurement equation and the imaging inverse problem in detail. In section 9.4, we briefly review the existing resolve framework in its current form and outline the CLEAN algorithm. Building on some of the computational shortcuts of the CLEAN algorithm, we derive fast-resolve, and in section 9.5, we show several applications of it. In particular, in section 9.5.1 we compare imaging results of fast-resolve with the classical resolve framework on VLA (Perley et al., 2011) data to validate the fidelity of the resulting sky images, and in section 9.5.2 we present a fast-resolve reconstruction of the ESO 137-006 MeerKAT observation to demonstrate the computational speedup.

9.3 The inverse problem

The radio interferometric measurement equation, derived, for example, in Smirnov, 2011a, and outlined in chapter 3, relates the radio sky brightness to the datapoints, often called visibilities. More specifically, via the measurement equation, model visibilities $\tilde{V}$ can be computed for an assumed sky brightness I and antenna sensitivity G. Under the assumption of scalar antenna gains G, the model visibility $\tilde{V}_{pqt}$ of antennas p and q at time t is given as:

$$\tilde{V}_{pqt} = \int C(\boldsymbol{l}, w_{pqt}) I(\boldsymbol{l}) G_p(t, \boldsymbol{l}) G_q^*(t, \boldsymbol{l}) \mathrm{e}^{-2\pi i (\boldsymbol{k}_{pqt} \cdot \boldsymbol{l})} d\boldsymbol{l}, \tag{9.1}$$

where

- $\boldsymbol{l} = (l, m)$ are the sky coordinates,
- t is the time coordinate,
- $\boldsymbol{k}_{pqt} = (u_{pqt}, v_{pqt})$ are the baseline uv-coordinates in units of the imaging wavelength,
- $C(\boldsymbol{l}, w_{pqt}) = \exp\left(-2\pi i w_{pqt}\left(\sqrt{1-\boldsymbol{l}^2} - 1\right)\right) / \sqrt{1-\boldsymbol{l}^2}$ is the w- or non-coplanar baselines effect,
- $I(\boldsymbol{l})$ is the sky brightness,
- $G_p(t, \boldsymbol{l})$ is the gain of antenna p depending on time and potentially also direction.

The model visibilities are therefore Fourier-like components of the sky brightness I modulated by the antenna gains G_p and G_q as well as the w-effect C. The visibilities actually recorded in the measurement are related to the model visibilities according to

$$V_{pqt} = \tilde{V}_{pqt} + n = R(I) + n, \tag{9.2}$$

with n representing some unknown noise in the measurement, and R the mapping from sky brightness to visibilities defined in equation 9.1. With equation 9.1 and equation 9.2, it is straightforward to compute simulated visibilities for an assumed sky brightness I, antenna gain G, and noise statistics. Inverting this relation is not possible without additional assumptions, since first, the measured visibilities V are corrupted by the measurement noise

n, and second, equation 9.1 is generally not invertible since in practical applications not all Fourier components are measured.

The non-uniqueness of solutions to equation 9.2 means that additional regularisation needs to be imposed to discriminate between all possible sky images compatible with the data. This additional information is either an explicit prior or regularization term, or is implicitly encoded in the structure of the imaging algorithm.

9.4 Methods

In this section, we derive the `fast-resolve` algorithm (section 9.4.3). Since `fast-resolve` builds on the already existing Bayesian imaging framework `resolve`, we start with a brief review of the classic `resolve` imaging method (section 9.4.1). As some of the computational speedups of `fast-resolve` are inspired by the `CLEAN` imaging algorithm, we also outline the basic concepts behind `CLEAN` (section 9.4.2).

9.4.1 `resolve`

The `resolve` imaging algorithm addresses the imaging inverse problem from a probabilistic perspective. Thus, instead of reconstructing a single estimate of the sky brightness, it infers the posterior probability distribution $P(I|V)$ of possible sky images given the measured visibilities. Bayes' Theorem

$$P(I|V) = \frac{P(V|I)P(I)}{P(V)} \tag{9.3}$$

expresses the posterior probability $P(I|V)$ in terms of the likelihood $P(V|I)$, the prior $P(I)$, and the evidence $P(V)$. `resolve` provides models for the likelihood $P(V|I)$ and the prior $P(I)$. The posterior distribution can be inferred for a given prior and likelihood model building on the functionality of the Bayesian inference package NIFTy[2] (Selig et al., 2013b; Steininger et al., 2017; Edenhofer et al., 2024a). In the next three subsections we briefly outline the likelihood and prior of `resolve` as well as the variational inference algorithms of NIFTy.

9.4.1.1 `resolve` prior

The `resolve` framework provides predefined prior models for the two types of radio emission, point sources and extended diffuse emission. Both priors encode that the brightness must be positive, since there is no negative flux. Furthermore, both priors are very flexible and allow for brightness variations over several orders of magnitude.

For the point source prior, pixels are independently modeled with an inverse gamma prior for their intensity. The inverse gamma distribution is strictly positive and has a wide tail, allowing extremely bright sources. In the example in section 9.5.1, such a prior is applied for the two bright point sources in the core of Cygnus A. While the brightness of

[2] https://github.com/NIFTy-PPL/NIFTy

point sources is reconstructed from the data, the locations currently need to be manually set. This limits the applicability of the current point source prior, as discussed in the application to MeerKAT ESO 137-006 data (section 9.5.2).

Besides positivity and possible variations over several orders of magnitude, the diffuse emission prior also encodes correlations of the brightness of nearby pixels, which is essentially the defining property of diffuse emission. The correlation of nearby pixels in the diffuse emission prior is modeled by Gaussian processes. The result of the Gaussian process is exponentiated to ensure positivity. Detailed explanations of the `resolve` prior models can be found in Arras et al., 2021a. An important aspect of all the prior models in `resolve` is that they are fast and scalable to very large numbers of pixels. For example, in Arras et al., 2021a a 4096×2048 pixel Gaussian process-based diffuse emission prior is used to image Cygnus A at various frequencies.

9.4.1.2 `resolve` likelihood

The likelihood is evaluated in `resolve` using equation 9.1. More specifically, using equation 9.1, we can write the likelihood as $P(V|I) = P(V|R(I)) = P(V|\tilde{V})$. The noise statistics in equation 9.2 then determines the likelihood. In `resolve`, Gaussian noise statistics are assumed. For numerical reasons, `resolve` works with the negative logarithm of the likelihood, named the likelihood Hamiltonian, instead of the likelihood itself. With the Gaussian noise assumption, the likelihood Hamiltonian is given by

$$H(V|I) = \frac{1}{2} (V - R(I))^{\dagger} N^{-1} (V - R(I)) + \frac{1}{2} \ln |2\pi N|, \tag{9.4}$$

with $\dagger$ denoting complex conjugate transpose and $\frac{1}{2} \ln |2\pi N|$ coming form the normalization of the Gaussian, which can be ignored in many applications. Different possibilities are implemented for the covariance of the noise N. In the simplest case, the weights of the visibilities are used as the inverse noise covariance. Alternatively, the noise covariance can also be estimated during the image reconstruction.

The computationally important aspect of the likelihood in the classic `resolve` framework is that for every update step where the likelihood is evaluated, also $R(I)$ and thus equation 9.1 need to be computed. This can be computationally expensive, especially for datasets with many visibilities, as we will discuss in detail later.

9.4.1.3 `resolve` posterior inference

`resolve` is built on the probabilistic programming package NIFTy, which provides variational inference methods (Knollmüller et al., 2019; Frank et al., 2021) to approximate the posterior distribution for a given prior and likelihood function. The advantage of variational inference techniques over sampling techniques such as MCMC or HMC is that they scale better with the number of parameters, which is the number of pixels in the imaging context. For example, the variational inference method of NIFTy has recently been used in a 3D reconstruction with 607 million voxels (Edenhofer et al., 2024b). While the variational inference algorithms scale very well with the number of parameters, they still need to evaluate

the likelihood very often. This means variational inference is fast as long as the evaluation of the likelihood and prior is fast.

As discussed in section 9.4.1.2, evaluating the likelihood in `resolve` boils down to evaluating the radio interferometric measurement equation (equation 9.1). To do so `resolve` relies on the parallelizable `wgridder` (Arras et al., 2021b) implemented in the DUCC[3] library. Nevertheless, evaluating the measurement equation becomes computationally intensive for data sets with many visibilities. For example, for the MeerKAT data set considered in section 9.5.2, evaluating equation 9.1 on eight threads takes about 45 seconds. This becomes prohibitive for algorithms which require many thousands of likelihood evaluations.

9.4.2 CLEAN

In this section, we briefly outline some of the concepts behind the `CLEAN` algorithm (Högbom, 1974; Schwab et al., 1983), as the computational speedups of `fast-resolve` over classic `resolve` are partly inspired by `CLEAN`. We will refrain from delving into the details behind the numerous improvements that have been made to the `CLEAN` algorithm over the last decades (see Rau et al., 2009 for example) and focus only on the aspects which are relevant to speeding up `resolve`. We refer the reader to Arras et al., 2021a for a detailed comparison between `CLEAN` and `resolve`.

`CLEAN` is an iterative optimization algorithm minimizing the weighted square residuals between the measured visibilities and the model visibilities computed from the sky brightness model. Expressed as a formula, the objective function minimized by `CLEAN` is identical to equation 9.4, the likelihood Hamiltonian of `resolve`. Minimizing equation 9.4 with respect to the sky brightness I is equivalent to solving

$$R^\dagger N^{-1} R I = R^\dagger N^{-1} V, \tag{9.5}$$

with $R^\dagger$ being the adjoint operation of R, thus mapping from visibilities to the sky brightness. Neglecting wide-field effects originating from non-coplanar baselines, the operation $R^\dagger N^{-1} R$ is equivalent to a convolution with the effective point spread function of the interferometer I^{PSF}. $R^\dagger N^{-1} V$ is the back projection of the noise-weighted data into the sky domain, called dirty image I^{D}. This can be expressed with the formula

$$I^{\mathrm{PSF}} * I \approx R^\dagger N^{-1} V = I^{\mathrm{D}}, \tag{9.6}$$

with I^{PSF} being the PSF of the interferometer and $*$ denoting convolution. Thus, the dirty image $I^{\mathrm{D}} = R^\dagger N^{-1} V$ is approximately the true sky brightness I convolved with the PSF I^{PSF} of the interferometer. Radio interferometric imaging is, therefore, nearly equivalent to deconvolving the dirty image.

As discussed in section 9.3, no unique solution to the imaging inverse problems exists. The absence of a unique solution manifests as $R^\dagger N^{-1} R$ and $I^{\mathrm{PSF}}*$ not being invertible operations. While in `resolve`, additional regularization was provided via an explicit prior, regularization of the sky images in `CLEAN` is implicitly encoded into the structure of the algorithm. More specifically, `CLEAN` starts with an empty sky model as an initial estimate

[3] https://gitlab.mpcdf.mpg.de/mtr/ducc

I_0^m for the true sky brightness and iteratively adds components, in the simplest form point sources, to this estimate until a stopping criterion is met. This introduces the implicit prior that the sky is sparsely represented by a finite number of CLEAN components.

In the following, we briefly summarize the algorithmic structure of CLEAN as relevant for fast-resolve. The iterative procedure of adding components to the current model image I^m is split into major and minor cycles. In the major cycles, a current residual image $I^{\mathrm{RES}} = R^\dagger N^{-1}(V - RI_i^m)$ is computed, with I_i^m being the current model image. In the minor cycle, additional components are added to the model, and their PSFs are removed from the residual image. In the subsequent major cycle, the residual image is recomputed for the updated model image, and in the following new minor cycle, more components get added to the model. This scheme is iterated until a global stopping criterion is met.

From a computational perspective, the important aspect of the major/minor scheme is that R and $R^\dagger$ only need to be evaluated in major cycles. In minor cycles, the PSF of the added CLEAN components is subtracted from the residual image, but for this, no evaluations of R and $R^\dagger$ are needed since the PSF can be precomputed once. Most of the computations of the CLEAN algorithm are performed in image space, and only major cycles go back to the visibility space. In contrast, resolve needs to map from image to data space by applying R for every likelihood evaluation.

9.4.3 fast-resolve

As outlined in the previous section, evaluating the radio interferometric instrument response (equation 9.1) contributes a substantial fraction to the overall runtime of a resolve image reconstruction. Therefore, reducing the number of necessary evaluations of equation 9.1 has the potential for significant speedups. The basic idea of fast-resolve is to perform most of the computations in image space, similar to CLEAN, and only evaluate the radio interferometric instrument response once in a while.

9.4.3.1 fast-resolve measurement equation

The radio interferometric measurement equation in the form of equation 9.2 leads to the likelihood Hamiltonian equation 9.4 of the classic resolve framework, which involves an evaluation of the interferometric instrument response R. To get a likelihood Hamiltonian that does not include evaluating R, one has to transform the measurement equation such that all involved quantities live in image space, not data space. To project all involved quantities to image space, we apply $R^\dagger N^{-1}$ from the left to equation 9.2, and get:

$$R^\dagger N^{-1} V = R^\dagger N^{-1} RI + R^\dagger N^{-1} n \tag{9.7}$$

$$I^{\mathrm{D}} = R' I + n', \tag{9.8}$$

with $R' = R^\dagger N^{-1} R$, and $n' = R^\dagger N^{-1} n$. The quantities of the new measurement equation are I^{D}, I, and n' and are all defined in image space and not in data space. The statistics of the new noise n' remains Gaussian as $R^\dagger N^{-1}$ is a linear transformation. The corresponding

likelihood Hamiltonian is given by:

$$H(I^{\mathrm{D}}|I) = \frac{1}{2}(I^{\mathrm{D}} - R'I)^{\dagger}N'^{-1}(I^{\mathrm{D}} - R'I) + \frac{1}{2}\ln|2\pi N'|. \tag{9.9}$$

The covariance of the transformed noise is given by

$$N' = \left\langle n'n'^{\dagger}\right\rangle = R^{\dagger}N^{-1}\left\langle nn^{\dagger}\right\rangle N^{-1}R = R^{\dagger}N^{-1}R. \tag{9.10}$$

This new noise covariance N' is not invertible as $R^{\dagger}N^{-1}R$ is not invertible, which is problematic as the likelihood Hamiltonian contains the inverse noise covariance N'^{-1}. To mitigate the problem of a singular noise covariance, we modify the measurement equation once more by adding uncorrelated Gaussian noise with a small amplitude. This leads to a noise covariance

$$N' = R^{\dagger}N^{-1}R + \epsilon\mathbb{I}, \tag{9.11}$$

with $\mathbb{I}$ being the unit matrix and ϵ a small number, which is the variance of the additional noise. With this artificially introduced additional noise, the full noise covariance N' is invertible, and the new likelihood Hamiltonian equation 9.9 becomes well defined.

9.4.3.2 `fast-resolve` response R'

How large the speedup of using the new likelihood Hamiltonian equation 9.9 is compared to the old Hamiltonian equation 9.4 depends on how much faster the new Hamiltonian can be evaluated. The idea of `fast-resolve` is to approximate $R' = R^{\dagger}N^{-1}R$ with a PSF convolution $R' = R^{\dagger}N^{-1}R \approx I^{\mathrm{PSF}}*$, as it is done in a similar way in the `CLEAN` algorithm. Approximating $R' = R^{\dagger}N^{-1}R$ by a convolution with the PSF is only exact for coplanar arrays. Corrections for the inaccuracy of the approximation are applied in some major cycles in analogy to the `CLEAN` algorithm, as we describe in section 9.4.3.4.

The convolution with the PSF $I^{\mathrm{PSF}}*$ can be efficiently applied via an FFT convolution. For data sets with many visibilities, this FFT-based convolution with the PSF I^{PSF} has the potential to be significantly faster than an evaluation of the interferometer response R (equation 9.1). In section 9.5, we will compare the computational speedup of the new likelihood for different data sets.

As the sky brightness I and the PSF I^{PSF} are non-periodic, some padding of the sky is needed for an FFT-based convolution. More specifically, to exactly evaluate $I^{\mathrm{PSF}} * I$, the PSF with which we convolve needs to be twice as big as the field of view we want to image since some emission in the sky I could be at the edge of the field we are imaging. The sky image needs to be padded with zeros to the same size as the PSF. As a formula, this can be noted as

$$\begin{aligned} R'I &= R^{\dagger}N^{-1}RI && (9.12)\\ &\approx I^{\mathrm{PSF}} * I && (9.13)\\ &= P^{\dagger}\mathrm{FFT}^{-1}\left[\mathrm{FFT}[I^{\mathrm{PSF}}]\cdot\mathrm{FFT}[PI]\right], && (9.14) \end{aligned}$$

with P denoting the padding operation, and $P^\dagger$ for slicing out the region not padded. By neglecting PSF sidelobes and reducing its size, the necessary amount of zero padding can be reduced. This, however, reduces the accuracy of the approximation $R' \approx I^{\mathrm{PSF}}*$, which might make it necessary to perform more major cycles in the image reconstruction.

9.4.3.3 fast-resolve noise model

To evaluate the likelihood Hamiltonian of fast-resolve, we need to apply both $R' = R^\dagger N^{-1} R$ and $N'^{-1} = (R^\dagger N^{-1} R + \epsilon \mathbb{I})^{-1}$. Similar to R', we need to approximate N'^{-1} such that we can apply it without having to evaluate R and $R^\dagger$ every time. The basic idea of the approximation of N'^{-1} is the same as for R', thus replacing the exact operation with an FFT convolution. Expressed as a formula, we compute the application of N' to some input x via

$$N'(x) \approx \mathrm{FFT}^{-1}\left[K \cdot \mathrm{FFT}(x)\right], \tag{9.15}$$

with some appropriate Kernel K. The approximate noise covariance still needs to fulfill the mathematical properties of covariances, because for our Bayesian model, the likelihood is a probability and not just an arbitrary cost function. Therefore, the approximated noise covariance matrix needs to be Hermitian and positive definite. Specifically, for the convolution approximation, this implies that all entries of the convolution kernel K must be real and strictly positive. To fulfill these constraints, we parameterize K as

$$K_\xi = \exp \xi + \epsilon, \tag{9.16}$$

with ϵ as defined in equation 9.11 and ξ being an implicitly defined real-valued vector. We set this vector ξ such that it minimizes the square residual between the true and the approximate noise covariance when applied to a test image with a point source in the center. Thus ξ is set to

$$\bar{\xi} = \mathrm{argmin}_\xi \left(N'(I_\delta) - \mathrm{FFT}^{-1}\left[(\exp \xi + \epsilon) \cdot \mathrm{FFT}(I_\delta)\right]\right)^2 \tag{9.17}$$

with I_δ having a point source or delta peak in the center of the field of view and otherwise being zeros. In words, we approximate the noise covariance N' with an appropriate Kernel K that yields a proper covariance by construction, is easy to invert, and minimizes the squared distance to the effect that N' has on a point source.

To evaluate the likelihood (equation 9.9) we need to apply N'^{-1}. We do this by convolving with the inverse kernel:

$$N'^{-1}(x) \approx \mathrm{FFT}^{-1}\left[K_{\bar{\xi}}^{-1} \cdot \mathrm{FFT}(x)\right]. \tag{9.18}$$

This involves a second approximation, as we here implicitly assume periodic boundary conditions. As long as the main lobe of the PSF is much smaller than the field of view, this assumption does not create large errors.

9.4.3.4 fast-resolve inference scheme

The previous sections derived the approximate likelihood for fast-resolve inspired by the CLEAN algorithm. In the same spirit, we adapt the major/minor cycle scheme of CLEAN

to utilize it with Bayesian inference for a probabilistic sky brightness reconstruction. In the minor cycles, we optimize the current estimate of the posterior distribution $P(I|V)$ for the sky brightness I using the above approximations for a fast likelihood evaluation. In the major cycles, we apply, similar to `CLEAN`, the exact response operations to correct for the approximation error.

The algorithm starts by initializing the dirty image $I^{\mathrm{D}} = R^{\dagger}N^{-1}V$ from the visibilities V using the exact response function R. The dirty image is the input data $d_0 = I^{\mathrm{D}}$ for the first minor cycle. The first minor cycle computes an initial estimate of the posterior distribution P_0 of the sky brightness using the measurement equation

$$d_0 = I^{\mathrm{PSF}} * I + n', \tag{9.19}$$

with the approximations discussed in section 9.4.3.2 and section 9.4.3.3. We will discuss the exact algorithm used to estimate the posterior distribution in section 9.4.3.5. From this initial estimate of the posterior distribution P_0, we compute the posterior mean of the sky brightness $I_0 = \langle I \rangle_{P_0}$. The posterior mean I_0 is the output of the first minor cycle and the input to the first major cycle. The first major cycle computes the residual d_1 between the dirty image I^{D} and the posterior mean of the first minor cycle passed through the response $R^{\dagger}N^{-1}R$ of the fast resolve measurement equation:

$$d_1 = I^{\mathrm{D}} - R^{\dagger}N^{-1}RI_0. \tag{9.20}$$

In contrast to the minor cycle, $R^{\dagger}N^{-1}R$ is here computed exactly and not approximated with $I^{\mathrm{PSF}}*$. The residual d_1 is the output of the major cycle and the input to the second minor cycle. In the second minor cycle, the posterior P_1 of I for the measurement equation

$$d_1 = I^{\mathrm{PSF}} * (I - I_0) + n', \tag{9.21}$$

is approximated. This approximation can be done efficiently by starting with the posterior estimate of the previous minor cycle and refining this estimate. The output of the second minor cycle is the updated posterior mean $I_1 = \langle I \rangle_{P_1}$ which is the input to the next major cycle computing the new residual to the dirty image using the exact response:

$$d_2 = I^{\mathrm{D}} - R^{\dagger}N^{-1}RI_1. \tag{9.22}$$

The new residual data d_2 is the input to the next minor cycle. This scheme is iterated until converged, thus until no significant structures are left in the residual d_n and the posterior estimates I_n are not changing anymore. In algorithm 1 the full inference scheme is summarized as a pseudocode algorithm.

The overall structure of the `fast-resolve` major/minor scheme is very similar to the major/minor scheme of `CLEAN`. The main difference is that `fast-resolve` updates in the minor cycles a probability distribution for possible sky images instead of a simple model image, as is the case for `CLEAN`. The algorithm for inferring the posterior distribution is outlined in the following subsection 9.4.3.5.

9.4.3.5 `fast-resolve` minor cycles

In the minor cycles, we optimize the approximation of the posterior distribution $P(I|V)$ using the scalable variational inference algorithms (Knollmüller et al., 2019; Frank et al., 2021) of

Algorithm 1 `fast-resolve` inference scheme.

```
# compute dirty image
I^D ← R^† N^{-1} V

# initialize input with dirty image
d_0 ← I^D

# reconstruct n major cycles
for i in 0 ... n − 1 do
    # minor cycle
    estimate P_i(I) from d_i = I^PSF ∗ I + n'
    # compute posterior mean
    I_i ← ⟨I⟩_{P_i}

    # major cycle updating input data
    d_{i+1} ← I^D − R^† N^{-1} R I_i

# final result: P_{n−1}(I)
```

the NIFTy package already employed in the classic `resolve` algorithm. These variational inference algorithms account for correlations between parameters of the model. While the method Metric Gaussian Variational Inference (MGVI) of Knollmüller et al., 2019 relies on a Gaussian approximation of the posterior distribution, the algorithm geometric Variational Inference (geoVI) of Frank et al., 2021 can also capture non-Gaussian posterior distributions. Although `fast-resolve` relies for the minor cycles on the same inference algorithms as `resolve`, updating the posterior distribution is much faster in `fast-resolve` since the approximate likelihood described above is used instead of evaluating the exact measurement equation.

Furthermore, the NIFTy package has also undergone a major rewrite, switching from a NumPy based (Harris et al., 2020) implementation to a JAX[4] (Bradbury et al., 2018) based backend, allowing for GPU-accelerated computing. While `resolve` was so far building on the old NumPy-based NIFTy, `fast-resolve` makes use of the JAX accelerated NIFTy version, also named NIFTy.re (Edenhofer et al., 2024a). Especially the FFT convolutions in `fast-resolve` have the potential for significant GPU acceleration. In section 9.5, we compare the runtime of `resolve` and `fast-resolve` using the CPU and the GPU backend.

In the future, we also plan to port the classic `resolve` framework to JAX. Porting `resolve` is more involved than porting `fast-resolve` because it requires binding a high-performance implementation of the radio interferometric measurement operator to JAX. As a preparatory work, we have developed the `JAXbind`[5] package (Roth et al., 2024b, and chapter 8) which allows to bind custom functions to JAX. However, since the `wgridder`

[4] https://github.com/google/jax

[5] https://github.com/NIFTy-PPL/JAXbind

from DUCC[6] is used to evaluate the radio interferometric measurement equation, the JAX version of `resolve` will also be limited to run on the CPU.

9.4.4 Previous fastRESOLVE of Greiner et al., 2016

The algorithm named fastRESOLVE introduced by Greiner et al., 2016 was an earlier attempt to speed up `resolve`. Similar to the `fast-resolve` of this work, Greiner et al., 2016 used approximations of the likelihood avoiding applications of R in every step. More specifically, Greiner et al., 2016 derived a maximum a posteriori estimator of the sky brightness using the gridded weights of the visibilities as a noise covariance. Since fastRESOLVE had some limitations, such as it could not account for the w-term and could not provide uncertainty estimates, this approach was not followed up in subsequent developments of `resolve` in Arras et al., 2019b; Arras et al., 2021a, or Roth et al., 2023b (chapter 6).

9.5 Applications

For verification, we demonstrate the proposed `fast-resolve` method on different data sets. First, we reconstruct the Cygnus A VLA observation at four different frequency bands, and second, we image ESO 137-006 using a MeerKAT observation. The Cygnus A data is suitable to validate the accuracy as the `fast-resolve` images can be compared to previous results from Arras et al., 2021a and Roth et al., 2023b (chapter 6) obtained on the same dataset. The Cygnus A reconstruction has a relatively small field of view, and wide field effects should be negligible. The MeerKAT observation is significantly larger than the VLA observation, and imaging with `resolve` is computationally out of reach, allowing to demonstrate the computational advantages of `fast-resolve` over `resolve`. Furthermore, the field of view of the ESO 137-006 observations is also significantly larger such that the major cycles of `fast-resolve` correcting for wide field effects become more important.

9.5.1 Application to VLA Cygnus A data

9.5.1.1 Data and algorithm setup

For imaging Cygnus A, we use the exact same data already imaged in Arras et al., 2021a with `resolve` and `CLEAN`. The data contains single frequency channels at 2052 MHz (S-band), 4811 MHz (C-band), 8427 MHz (X-band), and 13360 MHz (Ku-band). Also, in Roth et al., 2023b, an uncalibrated version of the S-band data was jointly calibrated and imaged with `resolve`. In all observing bands, all four VLA configurations were used. Computationally relevant details of the observations, such as the number of visibilities and the grid size of the reconstructed images, are listed in table 9.1. For details on the calibration of the data, we refer to Sebokolodi et al., 2020.

As a prior model, we use the combination of the diffuse emission prior and the point source prior described in section 9.4.1.1. This prior model is identical to the prior model already used in the existing `resolve` framework. Specifically, in Arras et al., 2021a and

[6] https://gitlab.mpcdf.mpg.de/mtr/ducc

Table 9.1: VLA Cygnus A observations. All observations are single channel and used all four VLA configurations. For all frequencies the same field of view (Fov) is imaged.

Freq [MHz]	N_{rows}	image size	Fov [deg]
2052	1281930	512×1024	0.025×0.05
4811	1281934	1024×2048	0.025×0.05
8427	2523954	2048×4096	0.025×0.05
13360	2544102	2048×4096	0.025×0.05

Roth et al., 2023b (chapter 6), this prior model was already used for Cygnus A reconstructions. For completeness, we summarize the main aspects of the prior model here. We refer to Arras et al., 2021a for details on the prior model. We model the diffuse emission with an exponentiated Gaussian Process spanning the entire field of view. For the two point sources in the nucleus of Cygnus A, we insert two separate point source models at the locations of these sources. For the brightness of these sources we use an inverse gamma prior. The exact hyper-parameters for the Gaussian process and the inverse gamma prior are listed in appendix 9.9.1. The number of pixels used to model the diffuse emission is listed in table 9.1 for the different frequencies.

In Arras et al., 2021a, the MGVI algorithm was used for inferring the posterior of the sky brightness map. For direct comparability between the results of this work and the previous `resolve` maps, we also use the MGVI in the minor cycles (section 9.4.3.5) for the posterior approximation. For the imaging of the S-band data in Roth et al., 2023b, the geoVI algorithm was used. However, we do not expect significant differences in the resulting sky maps.

To summarize, Arras et al., 2021a and Roth et al., 2023b used the same prior model setup as we use for imaging Cygnus A with `fast-resolve`. Furthermore, the posterior inference algorithm is expected to produce similar results for a given likelihood. Therefore, the previous `resolve` results of Arras et al., 2021a and Roth et al., 2023b are ideally suited to validate the `fast-resovle` likelihood approximation in a direct comparison on a dataset with negligible wide field effects. Furthermore, we compare with the multi-scale `CLEAN` reconstruction of the S-band data from Arras et al., 2021a. For a more detailed comparison with `CLEAN`, we refer to Arras et al., 2021a, where `resolve` was extensively compared to single and multi-scale `CLEAN` for all frequency bands.

9.5.1.2 Comparison with previous `resolve` results

In the following, we compare the `fast-resolve` reconstructions with previous results to validate the accuracy of `fast-resolve`. Figure 9.1 displays the `resolve` reconstructions of Arras et al., 2021a and Roth et al., 2023b as well as the S-band data multi-scale `CLEAN` reconstruction in comparison with the `fast-resolve` reconstructions of this work. The overall quality of the `fast-resolve` maps is on par with the `resolve` results. The multi-scale `CLEAN` reconstruction has a lower resolution in bright regions of the lobes than the `resolve` based reconstructions. All bright emission features are consistently reconstructed by all algorithms in all frequency bands. The `fast-resolve` and `CLEAN` maps have a higher dynamic range than the results of Arras et al., 2021a. Nevertheless, this is not

due to a conceptional problem of the resolve algorithm, but rather the reconstructions of Arras et al., 2021a were not fully converged due to the very high computational cost of resolve. resolve reconstructs the brightest emission of the field before modeling fainter features. Thus, faint features are missing in the radio map when a reconstruction is stopped before convergence. For fast-resolve, where the overall runtime of the algorithm is much shorter, it is easier to ensure that the reconstruction is fully converged. In section 9.5.1.3 the convergence of resolve and fast-resolve is analyzed in detail. The resolve reconstruction of Roth et al., 2023b has a similar dynamic range compared to the fast-resolve reconstructions.

In the comparison between CLEAN and resolve in Arras et al., 2021a, resolve produced significantly higher resolved maps than CLEAN in regions with high surface brightness. A zoom into such a region, the eastern hot spot of Cygnus A, is depicted in figure 9.2. In this region, the results are consistent, and fast-resolve achieves the same resolution as the resolve based reconstructions.[7] Furthermore, fast-resolve also shows minimal imaging artifacts around the hot spot, as does resolve. The multi-scale CLEAN map from Arras et al., 2021a is added again for comparison. As already discussed in Arras et al., 2021a, the CLEAN reconstruction has a significantly lower resolution than the resolve reconstruction. These super-resolution capabilities were validated by comparing the morphological features of the resolve reconstructions with higher frequency observations under the assumption of spectral smoothness.

Figure 9.3 zooms on the nucleus and jet of Cygnus A. In the zooms on the nucleus the pixels modeling the two point sources are clearly visible. The fast-resolve S-band reconstruction and the resolve reconstruction of Roth et al., 2023b show consistent results depicting the nucleus and jet. In the S-band reconstruction of Arras et al., 2021a, the jet is also visible, although due to the smaller dynamic range of the map, it is less pronounced. In the higher frequency bands, the shape of the core of Cygnus A is consistent between the reconstructions of resolve and fast-resolve. As the old resolve reconstructions are not fully converged, their dynamic range is lower, and the fainter emission of the jet is barely visible at higher frequencies. At the highest frequencies 13360 MHz, the jet is also barely visible in the fast-resolve reconstruction. The CLEAN image of the jet is consistent with the fast-resolve reconstruction and the resolve reconstruction from Roth et al., 2023b.

resolve as well as fast-resolve provide posterior samples of the sky brightness distribution. From these samples, not only the posterior mean but also other summary statistics can be computed. As an example, we show in figure 9.4 the pixel-wise relative uncertainty of the S-band data reconstruction of fast-resolve in comparison with the corresponding uncertainty map of Arras et al., 2021a. The estimated uncertainty of resolve tends to be slightly higher than the uncertainty estimate of fast-resolve. This difference might come from the fact that the resolve reconstruction of Arras et al., 2021a was not fully converged since the uncertainty estimate of resolve usually becomes smaller when the algorithm converges. Nevertheless, we cannot exclude that this is related to the approximations of fast-resolve.

[7] The pixelizations of the multi-scale CLEAN, the previous resolve, and the fast-resolve reconstructions differ as they originate from different works.

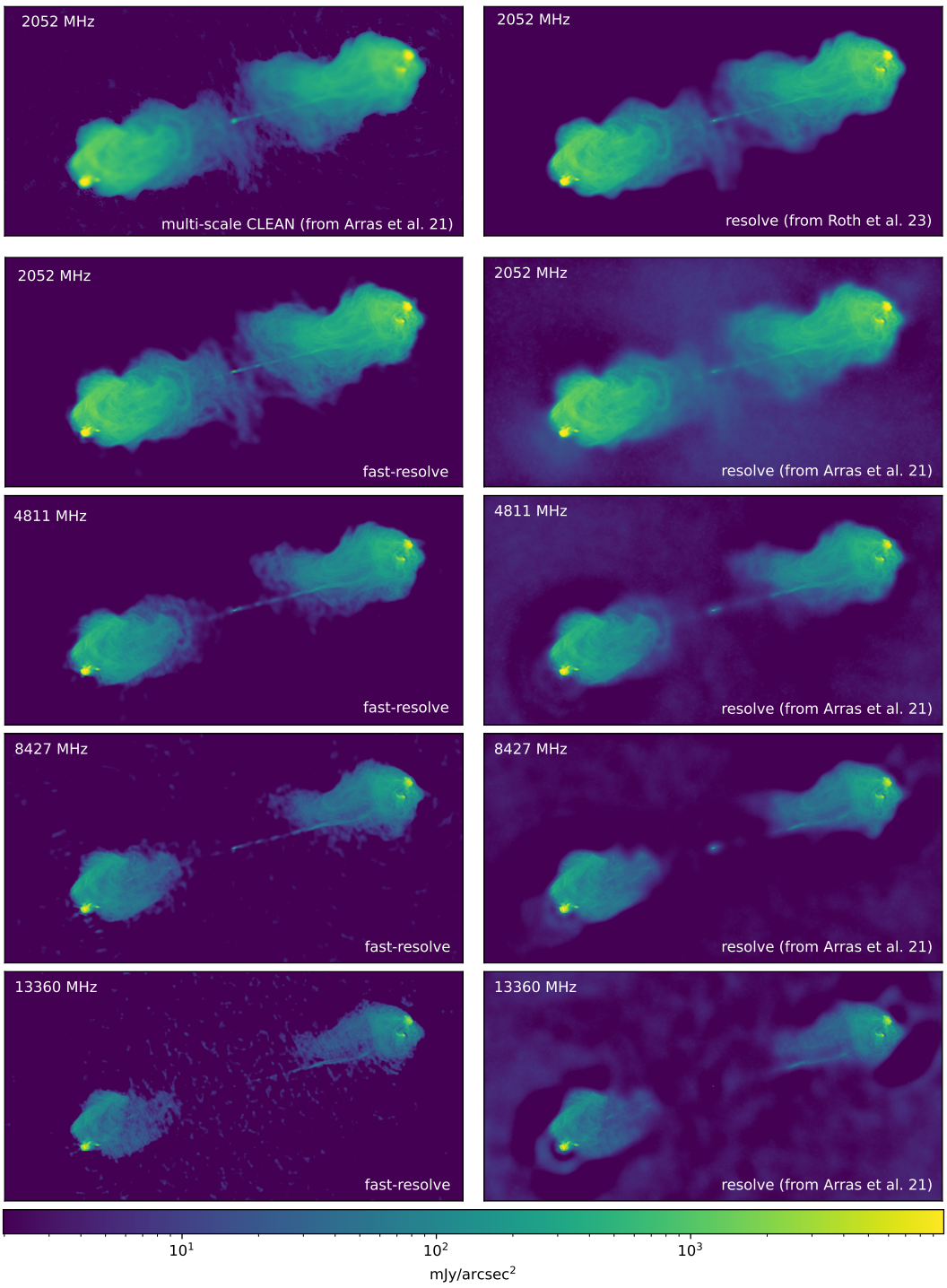

Figure 9.1: Comparison of `fast-resolve` Cygnus A reconstructions with `resolve` and `CLEAN` reconstructions. The top left panel shows the multi-scale `CLEAN` map of the S-band data form Arras et al., 2021a. The left column below the `CLEAN` map shows `fast-resolve` reconstructions at four different frequencies indicated in the top left corner of each panel. The right column shows `resolve` reconstructions of Roth et al., 2023b (chapter 6) and Arras et al., 2021a using the same data. The dynamic range of the `fast-resolve` reconstructions is higher than in some of the previous `resolve` reconstructions. This is because the old `resolve` reconstructions were not fully converged due to their high computational cost.

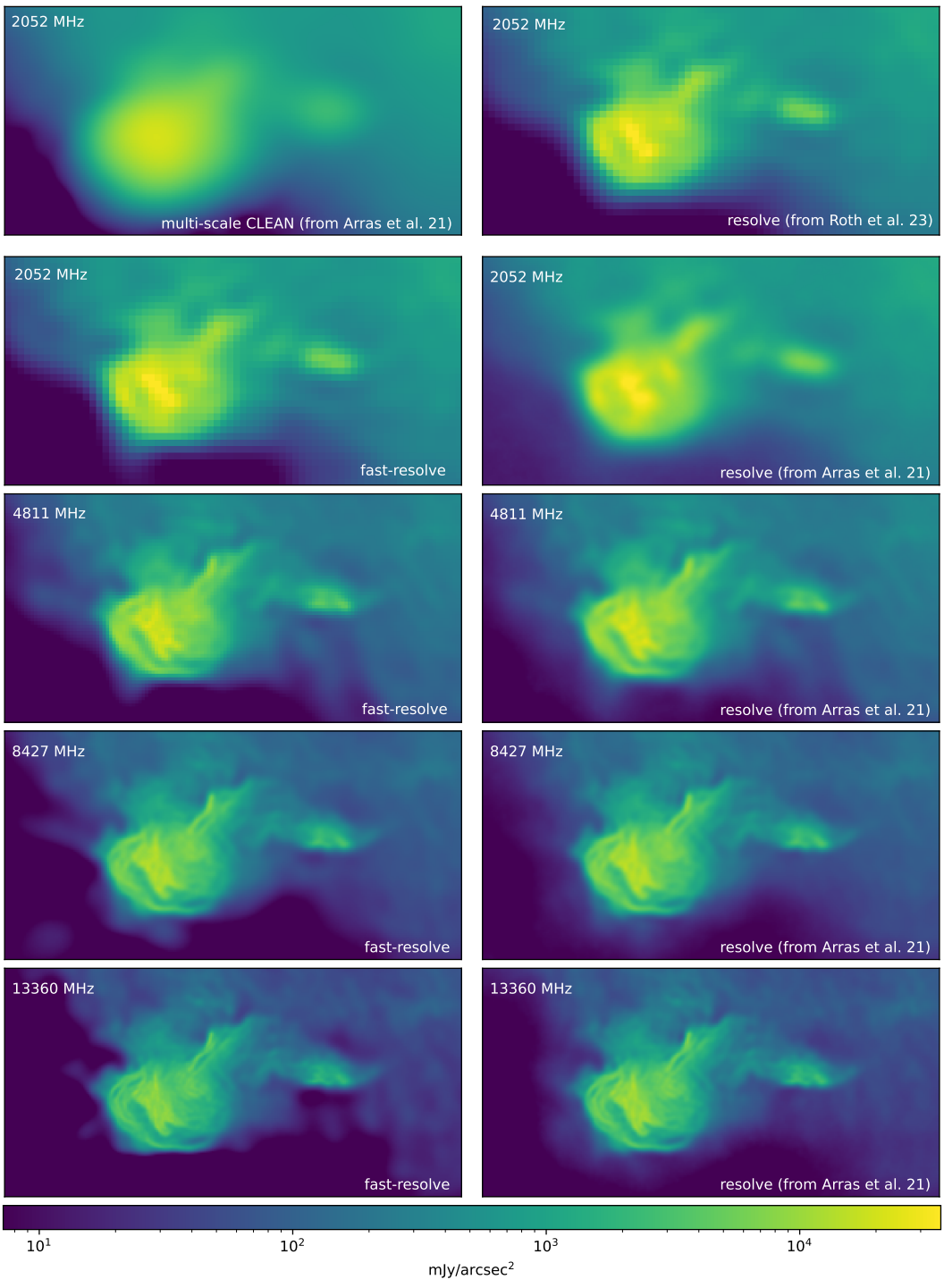

Figure 9.2: Zoom on the eastern hot spot of the Cygnus A reconstructions from figure 9.1. The resolution of the `fast-resolve` maps in the left column is on par with the `resolve` maps in the right column. The resolution of the `CLEAN` map is significantly lower than the in the other maps.

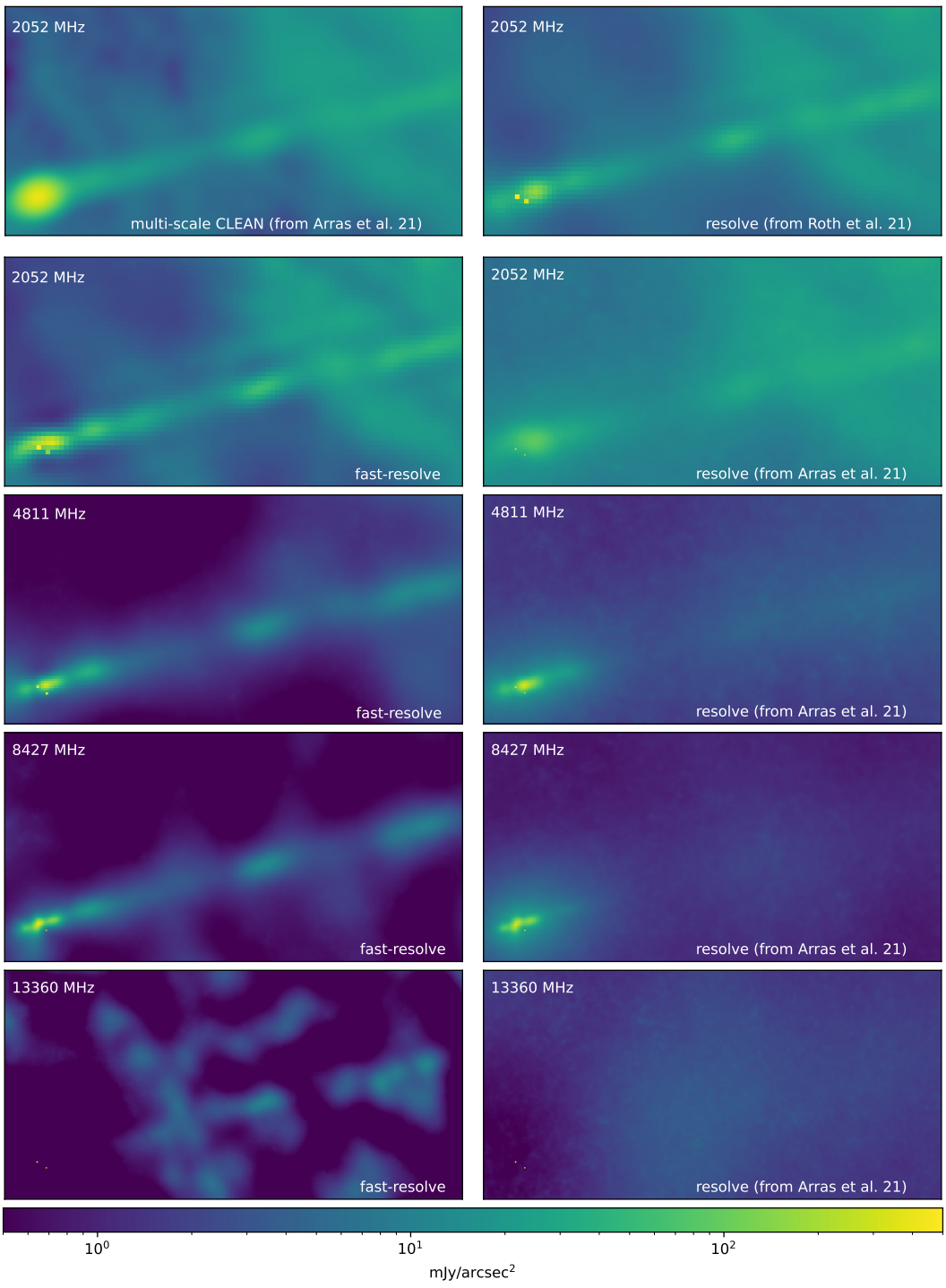

Figure 9.3: Zoom on the nucleus and jet of the Cygnus A reconstructions form figure 9.1. In all `resolve` and `fast-resolve` reconstructions, the two point sources in the nucleus are modeled with two pixels that are uncorrelated with the brightness of the neighboring pixels. Since the `resolve` of Arras et al., 2021a are not fully converged due to their high computational cost, their dynamic range is lower, and the jet of Cygnus A is not visible. The multi-scale `CLEAN` map from Arras et al., 2021a is consistent with the `fast-resolve` result and the `resolve` reconstruction from Roth et al., 2023b. At the highest frequency the jet is hardly visible also with `fast-resolve`, with background artifacts having a similar brightness.

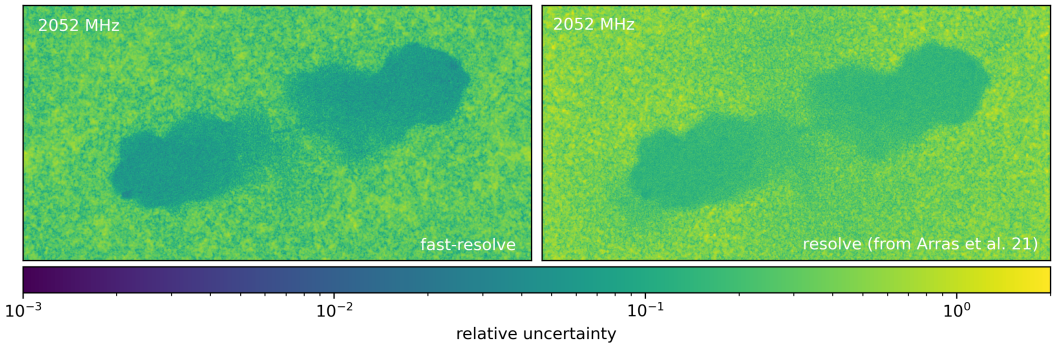

Figure 9.4: Pixel-wise relative uncertainty map of `fast-resolve` S-band reconstruction compared to the `resolve` uncertainty map of Arras et al., 2021a. The relative uncertainty is lower for both reconstructions in regions of high surface brightness. In the hot spot, the relative uncertainty is around 10^{-3} to 10^{-2} and increases towards the lower surface brightness regions. Outside of Cygnus A, the relative uncertainty fluctuates around 1, indicating that the reconstructed intensity is statistically consistent with zero surface brightness. The estimated uncertainty of `resolve` tends to be slightly higher than the uncertainty estimate of `fast-resolve`. This might be due to the `resolve` reconstruction of Arras et al., 2021a not being fully converged.

9.5.1.3 Computational time of `fast-resolve`

In the previous subsection, we compared the results of `fast-resolve` with `resolve`, confirming that `fast-resolve` can deliver the same high quality radio maps as `resolve`. In this subsection, we want to analyze the computational speed of `fast-resolve` and compare it to the existing `resolve` framework. To have a direct comparison between the runtimes, we imaged the S-band data with exactly the same hyperparameters for the prior model with `fast-resolve` and `resolve` and saved snapshots of the reconstructions at several points during the optimization. While `resolve` can only be executed on a CPU, `fast-resolve` can also run on a GPU.

Figure 9.5 depicts the direct comparison between `resolve` and `fast-resolve`, with both algorithms being executed on the same CPU. Specifically, the algorithms were run on an Intel Xeon W-1270P CPU with 3.80 GHz clock speed and 8 cores. For the `resolve` reconstruction, the MGVI algorithm inferring the posterior distribution was parallelized with 4 MPI tasks, each using 2 cores for evaluating the radio interferometric measurement equation (equation 9.2), thus utilizing 8 cores in total. The `fast-resolve` reconstruction ran on the single python process without manual threading, relying on JAX to parallelize the computations over the cores. Thus, in addition to the algorithmic difference, differences in the parallelization strategy and programming language might impact the reconstruction runtime. Snapshots of the reconstructions after 10, 60, 300, 600, and 1440 min are displayed. After 10 min, the hot spot and parts of the lobes are visible in the reconstructions of both algorithms. Nevertheless, the core and the jet are missing, and both reconstructions have significant background artifacts. After 60 min runtime, fainter parts of the lobes become

visible, and the fast-resolve map shows significantly more details in the lower surface brightness regions. After 300 min the fast-resolve map also depicts the very low surface brightness regions with high resolution, while in the resolve map, the outflow is still partially missing, and the lower surface brightness regions of the lobes are reconstructed with low resolution. After 600 min the fast-resolve map has only slightly changed compared to the 300 min snapshot. At this point, we assume the fast-resolve algorithm to be fully converged. In total we performed 20 fast-resolve major cycles. The resolve reconstruction is also after 600 min not yet fully converged. The final snapshot of the resolve reconstruction is after 1440 min. At this stage also the resolve reconstruction is nearly converged. Only in the very low surface brightness regions of the lobes, the resolution is still lower than in the fast-resolve reconstruction.

While figure 9.5 analyzes the convergence of fast-resolve on the CPU, fast-resolve is mainly designed for GPUs. The performance of fast-resolve on the S-band data using GPUs is shown in figure 9.6. Specifically, figure 9.6 displays snapshots of the fast-resolve reconstruction after 1, 5, 10, 22, and 46 min using an NVIDIA A100 high-performance computing GPU compared with a reconstruction on a consumer-level GPU, an NVIDIA GeForce RTX 3090. When comparing the results, one should consider that the A100 GPU is on the order of 10 times more expensive than the RTX 3090. On both GPUs, we executed the exact same fast-resolve reconstruction we ran on a CPU in figure 9.5. Similar to the CPU run, the high surface brightness regions are reconstructed first before the method picks up the low surface brightness flux. On both GPUs, the algorithm is much faster than on the CPU. While on the CPU, the reconstruction was finished after 600 min, the same number of major and minor cycles were finished on the RTX 3090 GPU in 46 min and on the A100 GPU in 22 min. Thus, the fast-resolve reconstruction was 13 times faster on the RTX 3090 GPU and 27 times faster on the A100 GPU than the CPU reconstruction. In comparison, the resolve reconstruction on the CPU was not fully converged even after 1416 min.

To quantify the convergence rate of the reconstructions, we computed the mean square residual between the logarithmic brightness of the final fast-resolve iteration on the NVIDIA A100 GPU and earlier iterations of resolve and fast-resolve reconstructions. We reran the fast-resolve reconstructions with different random seeds to be independent of the random seed used for the NVIDIA A100 reconstruction with respect to which the residuals are computed. Figure 9.7 shows the mean square residuals for resolve and the three fast-resolve reconstructions as a function of wall time. The fast-resolve reconstructions converge within the displayed time, and their curves of the mean square residual flatten. The reconstruction's final mean square residuals are not exactly the same since the final maps are not numerically identical because of the different hardware, JAX, and CUDA versions. The resolve reconstruction does not converge in the displayed time interval and the residual error keeps falling until the end after 1416 min (1 day). Of course, using the residual of the logarithmic brightness is an arbitrary metric for quantifying convergence speed. Nevertheless, it can roughly quantify the speedups of fast-resolve. After 1416 min, the mean square residual of the resolve reconstruction is around 10^{-3}. In comparison, the fast-resolve reconstructions reach the same mean squared residual after around 10 min, 20 min, and 200 min for runs on the NVIDIA A100 GPU, the NVIDIA RTX 3090 GPU and the Intel Xeon CPU, respectively. Thus, the indicative speedup of

`fast-resolve` over `resolve` on the Cygnus A S-band data is a factor of 144 for the A100 GPU, a factor of 72 for the RTX 3090 GPU and a factor of 7.2 for the CPU run. The timings reported above do not include the time needed to precompute the convolution kernels for the response and noise of `fast-resolve`. Nevertheless, these are small compared to the time needed for imaging. For the S-band data, for example, the computation of the kernels takes only 0.6 min, which is much shorter than the imaging runtime, even on the GPU.

As indicated in table 9.1, the Cygnus A data has only a single frequency channel. For datasets with more baselines and frequency channels, such as the MeerKAT datasets considered in the next section (see table 9.2), the algorithmic advantage of `fast-resolve` of not having to compute the radio response in each evaluation of the likelihood is much larger. Thus, for such datasets, the speedup of `fast-resolve` will be significantly larger than for the Cygnus A single channel imaging. Indeed for the MeerKAT dataset of the next section (section 9.5.2) imaging with `resolve` is computationally out of scope, while with `fast-resolve` images can still be reconstructed with moderate computational costs.

The same comparison as in figure 9.6 but for the C-band data is displayed in figure 9.8. As indicated in table 9.1, the grid size we used for the C-band sky map is a factor of two larger along both spatial axes. Thus, in total, we have four times more pixels, increasing the computational cost of the algorithm. On the A100 GPU, the `fast-resolve` reconstruction was finished after 56 min, while on the RTX 3090 GPU, the reconstruction took 165 min. We believe that the larger difference between the two GPUs for the C-band data compared to the S-band data might be because, for the smaller grid size of the S-band reconstruction, the A100 GPUs were not fully utilized.

For completeness, we also display snapshots of the X and Ku-band reconstructions in figure 9.9. These reconstructions were only carried out on the A100 GPU. After 132 min both reconstructions were finished.

9.5.2 Application to MeerKAT data

In this section, we present an application of `fast-resolve` to an L-band (856 – 1712 MHz) MeerKAT (Jonas et al., 2016) observation of the radio galaxy ESO 137-006. Originally, this observation was presented in Ramatsoku et al., 2020. The observation utilized all 64 MeerKAT antennas and the 4k mode of the SKARAB correlator. The total on target observation time is 14 h in full polarization with 4096 channels. For the VLA Cygnus A observations above, only a single frequency channel was used for each band. Here, we use two sub-bands with each about 200 channels (after averaging) and a bandwidth of approximately 200 MHz. Consequently the MeerKAT datasets are more than 400 times larger than the VLA datasets of the previous section, making an evaluation of the radio interferometer response (equation 9.1) significantly more expensive. A `resolve` reconstruction, where the exact response needs to be computed for each evaluation of the likelihood (equation 9.4.1.2), is computationally unfeasible for such a dataset.

Ramatsoku et al., 2020 detected hitherto unknown collimated synchrotron threads linking the lobes of the radio galaxy ESO 137-006 in this observation. Since then, this data set has also been used by Dabbech et al., 2022 for demonstrating convex optimization based imaging algorithms. Dabbech et al., 2022 also published the results as `FITS` files, allowing

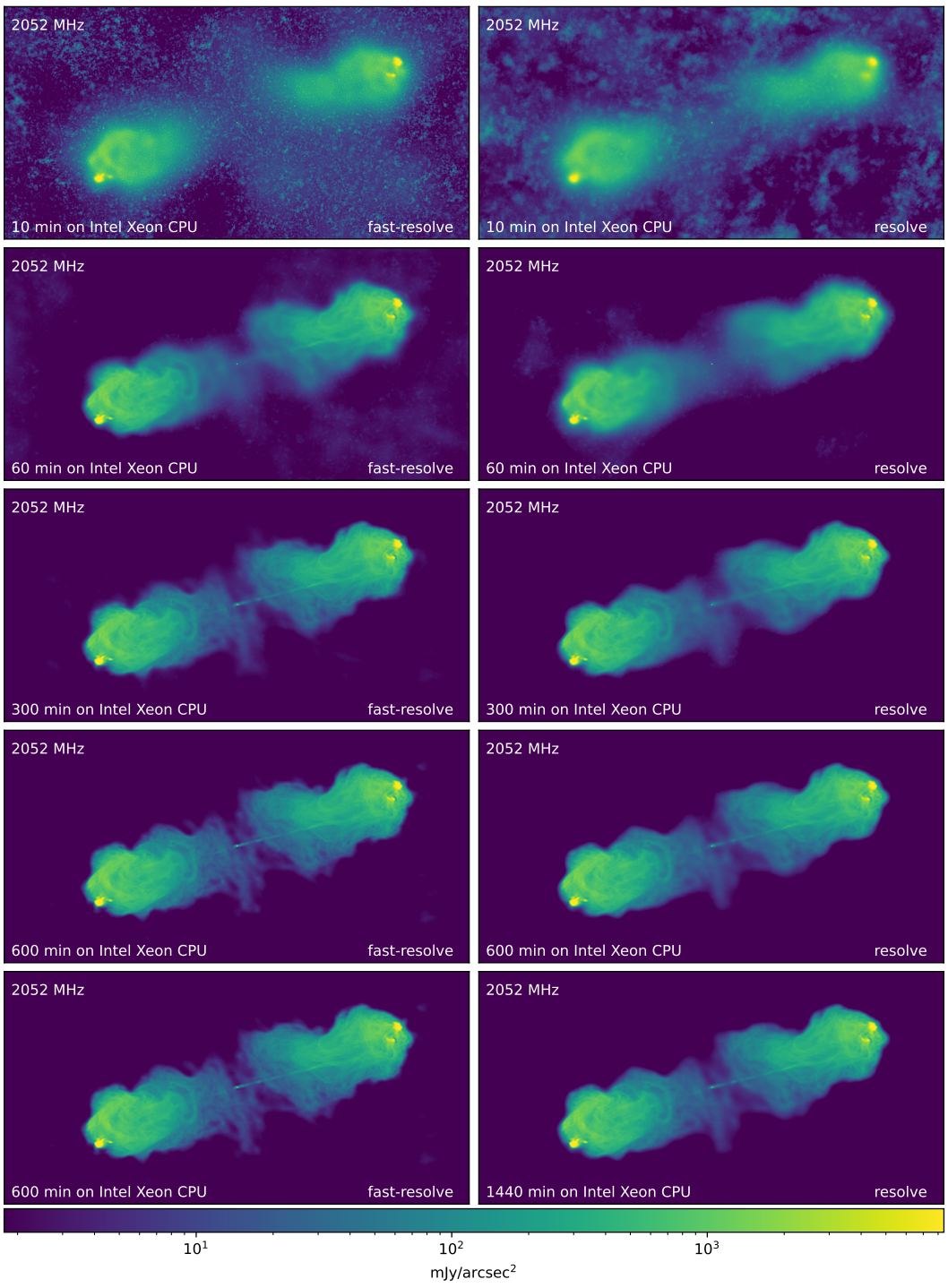

Figure 9.5: `fast-resolve` and `resolve` reconstructions of the 2052 MHz Cygnus A data at different stages of the reconstruction. The left column shows `fast-resolve` the right column `resolve`. The rows display snapshots of the reconstruction after different amounts of wall time indicated in the lower left of each panel. The `fast-resolve` reconstruction was performed on two cores of an Intel Xeon CPU. The `resolve` reconstruction used eight cores of the same CPU. The `fast-resolve` reconstructions in the last two rows are at identical time snapshots, since the `fast-resolve` reconstruction is considered to be converged after 600 min. The `resolve` reconstruction is not fully converged even after 1416 min.

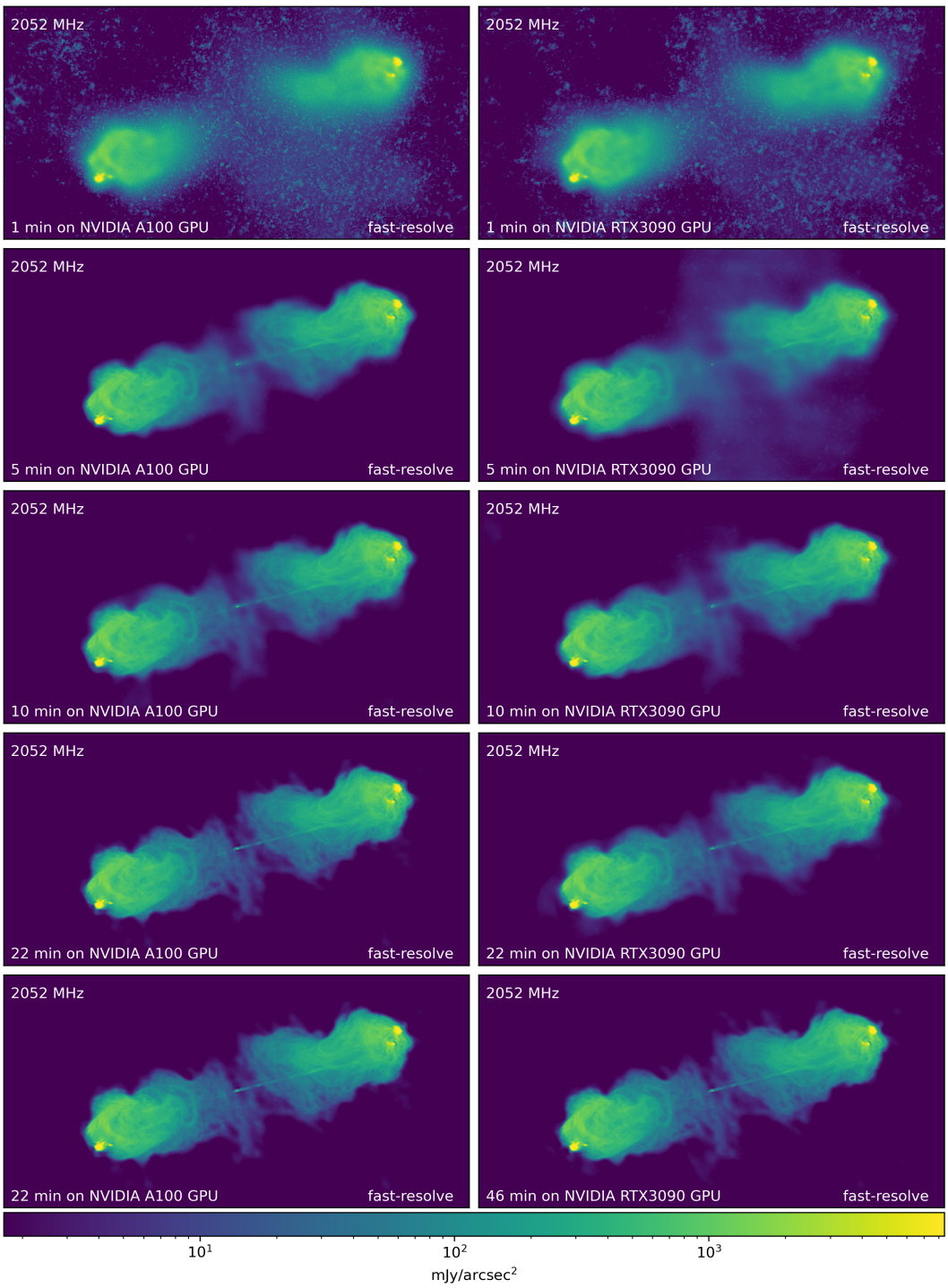

Figure 9.6: `fast-resolve` reconstructions of the 2052 MHz Cygnus A data. The left column displays snapshots of a `fast-resolve` reconstruction on an NVIDIA A100 GPU. The right column shows snapshots of a `fast-resolve` reconstruction on an NVIDIA RTX 390 GPU. Both reconstructions ran the same number of major and minor iterations. Since the reconstruction on the A100 GPU was finished after 22 min, the bottom left panel is identical to the row above. On the A100 GPU `fast-resolve` is approximately twice as fast as on a RTX 390 GPU, and 25 times faster than on the CPU displayed in figure 9.5.

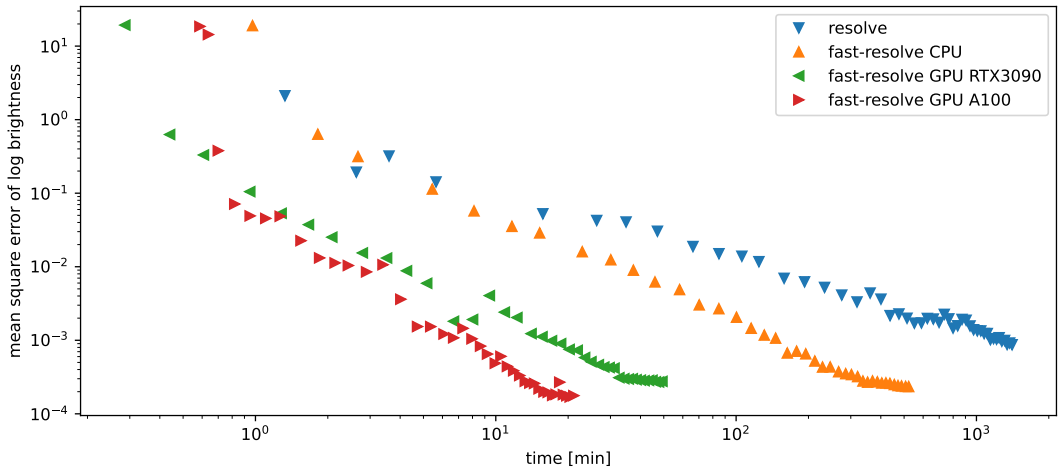

Figure 9.7: Mean square residual of the log brightness of `resolve` and `fast-resolve` reconstruction as a function of time for the S-band data. The residual is computed to the final iteration of `fast-resolve` on the NVIDIA A100 GPU depicted in figure 9.6, which we believe to be converged. The `fast-resolve` reconstructions are recomputed with a different random seed. The mean square residual of the `fast-resolve` reconstructions falls until their curves flatten when they converge. Due to the different hardware, JAX, and CUDA versions, the final results are numerically not exactly identical. The `resolve` reconstruction (blue curve) does not fully converge in the displayed time range.

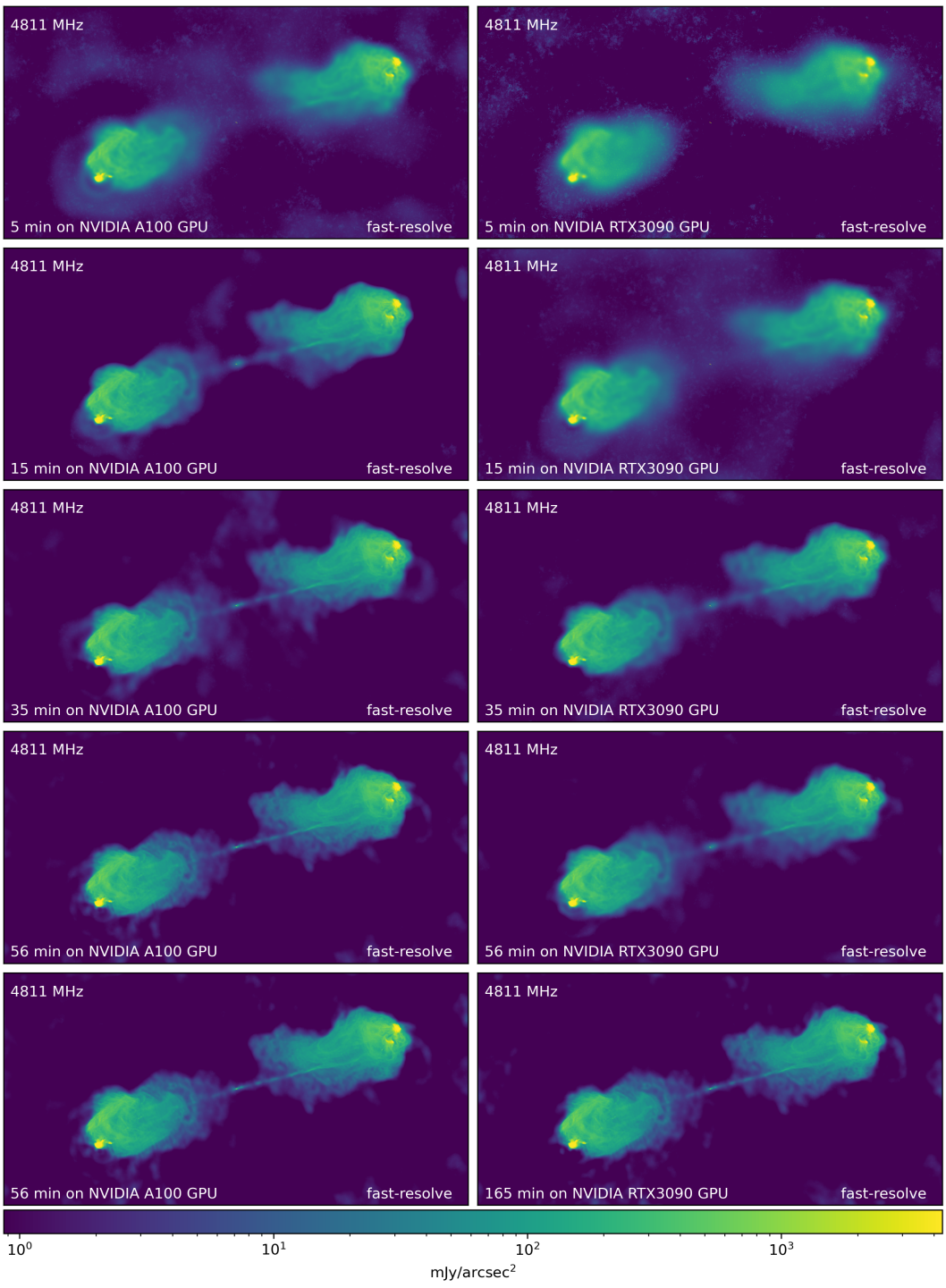

Figure 9.8: Same as figure 9.6 but for the 4811 MHz data. On an NVIDIA A100 GPU `fast-resolve` is 2 – 3 times faster than on an NVIDIA RTX 3090 GPU. The same number of major and minor iterations are performed on both GPUs. Since the reconstruction on the A100 GPU is finished after 56 min, the bottom left panel is identical to the row above.

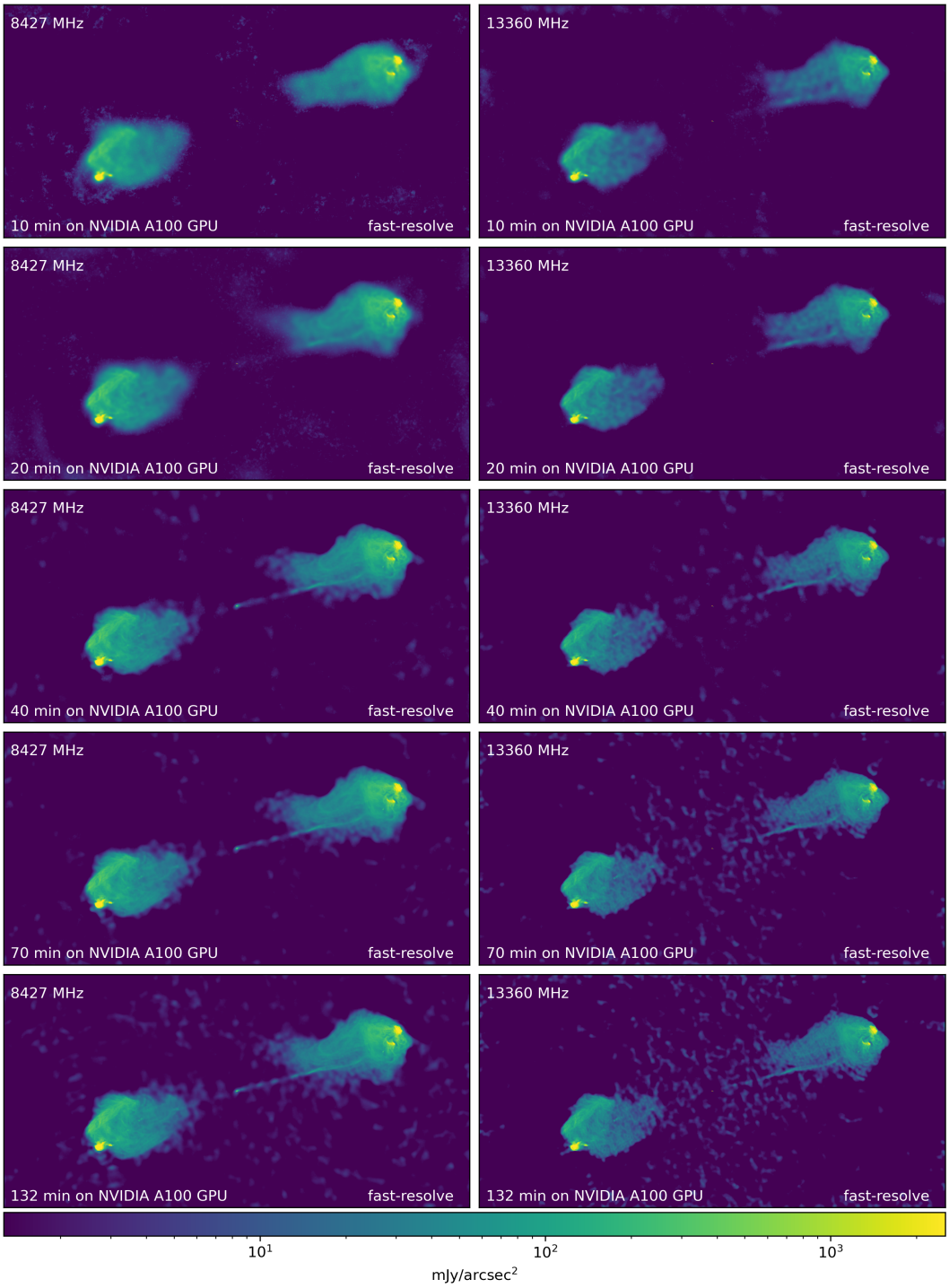

Figure 9.9: Snapshots of the `fast-resolve` Cygnus A reconstruction of the 8427 MHz and 13360 MHz data on a NIFTy A100 GPU. The wall time after which the snapshot was taken is indicated in the bottom left of each panel.

Table 9.2: MeerKAT ESO 137-006 observations. For both frequency bands the same field of view (Fov) is imaged.

Freq [MHz]	N_{rows}	N_{chan}	image size	Fov [deg]
961-1145	4598424	220	3600×3600	2×2
1295-1503	4598424	248	3600×3600	2×2

us to use them as a reference for validating the `fast-resolve` algorithm for a MeerKAT-sized reconstruction.

Technical details relating to the initial flagging and transfer calibration of the ESO 137-006 data using the CARACAL pipeline (Józsa et al., 2020) are given in Ramatsoku et al., 2020. As in Ramatsoku et al., 2020, the data is averaged from 4096 to 1024 frequency channels and split into two sub-bands spanning 961 – 1145 MHz and 1295 – 1503 MHz that are relative free from radio frequency intereference, known as the LO and HI bands respectively. These subbands are then phase self-calibrated using WSClean multi-scale CLEAN (Offringa et al., 2017) for imaging and CubiCal (Kenyon et al., 2018) for calibration. We then image the data independently in the LO and HI band with `fast-resolve`. Computational relevant details of the calibrated data in the two bands are listed in table 9.2.

9.5.2.1 `fast-resolve` prior model

As for the Cygnus A observation, we built our prior model around the exponentiated Gaussian process model described in section 9.5.1.1 and in detail in Arras et al., 2021a. Nevertheless, for the MeerKAT observation, there is, besides the main source, ESO 137-006, also a second galaxy, ESO 137-007, in the field of view. Due to the nature of the generative prior models in `resolve`, the full prior model can easily be composed of multiple components. We model each of these sources with a separate exponentiated Gaussian process to decouple the prior models. Furthermore, due to the high sensitivity of MeerKAT, many compact background sources are detected. For the Cygnus A reconstruction, we placed two point source models at the locations of the two point sources in the core of Cygnus A. However, for the MeerKAT observation, the number of background sources is far too large to manually place point source models at their locations. Therefore, we use a third Gaussian process model to represent all the background sources outside of the models for ESO 137-006 and ESO 137-007. A sketch of the layout of the three models is shown in figure 9.10. The exact parameters for all Gaussian process models are listed in appendix 9.9.2.

9.5.2.2 Imaging of Dabbech et al., 2022

In Dabbech et al., 2022, ESO 137-006 is imaged with WSClean (Offringa et al., 2014; Offringa et al., 2017) and two convex optimization algorithms in the same bands 961 – 1145 MHz and 1295 – 1503 MHz using the same data as we do. The two convex optimization algorithms of Dabbech et al., 2022 utilize two different regularizers. One regularizer is named uSARA, promoting sparsity in a wavelet basis. The other regularizer, AIRI, originally presented in Terris et al., 2023a, is based on a neural network denoiser. The resulting images for both regularizers are compared with results from WSClean, showing significant

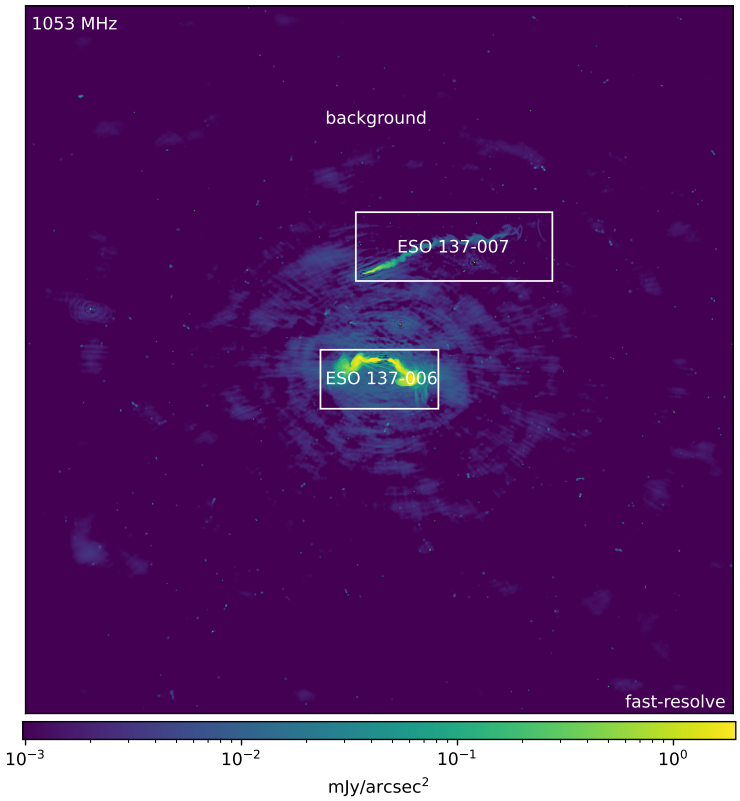

Figure 9.10: Reconstruction of the ESO 137-006 MeerKAT observation in the LO subband (961 − 1145 MHz) with `fast-resolve`. The radio galaxies ESO 137-006 and ESO 137-007, as well as the background sources, are modeled with independent Gaussian processes. White boxes indicate the spatial extent of each Gaussian process. Zooms on the reconstructions of ESO 137-006 and ESO 137-007 for both subbands are shown in figure 9.11, 9.12, 9.13, and 9.14.

improvements in the image quality. The exact computational costs of all three imaging algorithms are reported in Dabbech et al., 2022. Imaging with the sparsity promoting regularizer, uSARA needed about 1500 to 3000 CPU hours per band to converge. With the neural network-based denoiser AIRI, around 900 to 1600 CPU hours plus around 5 GPU hours were needed.

9.5.2.3 Imaging results

Figure 9.10 depicts the results of the `fast-resolve` reconstruction for the LO band data. 25 and 28 `fast-resolve` major cycles were executed for the LO and the HI band reconstructions. Zooms on the two radio galaxies are shown in figure 9.11 and figure 9.12 for the LO band and in figure 9.13 and figure 9.14 for the HI band. Thereby, figure 9.11 and figure 9.13 display the ESO 137-006 radio galaxy while figure 9.12 and 9.14 show the radio galaxy ESO 137-007 north of it. These figures also include the convex optimization reconstruction results of Dabbech et al., 2022 for comparison and validation. The imaging results of `fast-resolve` and Dabbech et al., 2022 are consistent. The `fast-resolve` maps have higher background artifacts than the convex optimization maps. In Ramatsoku et al., 2020, collimated synchrotron threads between the two lobes of the galaxy ESO 137-006 were detected. Besides these already known threads, the `fast-resolve` map shows additional threads north of the core of the galaxy. These additional threads are probably artifacts in the image. With a lower intensity, similar features are also found in the maps of Dabbech et al., 2022 and are mostly considered artifacts.

We believe these features are imaging artifacts and originate from suboptimal calibration. Due to the very flexible prior model, both `fast-resolve` and `resolve` are very sensitive the data. In the case of suboptimal calibration, this leads to imaging artifacts. At present, self-calibration with `fast-resolve` is not yet possible. For the future, the integration of a self-calibration routine into `fast-resolve` is planned, which could potentially mitigate such artifacts. The appendix 9.9.3 presents an application of `fast-resolve` to synthetic data with the same uv-coverage as the MeerKAT LO band data. The synthetic data has no calibration inaccuracies, and the reconstructions do not show similar artifacts.

Additionally, a dedicated prior model for compact sources could improve the `fast-resolve` reconstructions. At present, point sources can either be modeled by manually placing inverse gamma distributed pixels at their locations (see section 9.4.1.1) or by including them in the Gaussian process model. For the VLA Cygnus A reconstruction, the point sources in the nucleus were modeled by manually placing inverse gamma distributed pixels at their locations. Thereby, imaging artifacts around the two-point source could be avoided, despite them being very bright. Due to the high number of background sources, this is impractical for the ESO 137-006 reconstruction. Compact sources are therefore also modelled using an exponentiated Gaussian process prior. The Gaussian process based prior is very sensitive to diffuse emission but is less sensitive to point sources. For this reason, many fainter point sources are not picked up by the model but left in the `fast-resolve` residual data, which is shown in the appendix 9.9.4. A dedicated prior model that is also applicable to observations with a very high number of background sources could improve their reconstruction. Ideally, such a point source model should also be capable of reconstructing the exact positions of the sources and not rely on error-prone manual placement

of sources as in the current point source prior version. However, the development of an improved point source model is left for future work.

Furthermore, in a close examination of the results, minor artifacts at the edges of the prior models of the sources ESO137-006 and ESO137-007 are visible. These artifacts could be avoided by an overlap between source models and the background model and interpolating between them.

9.5.2.4 Computational costs

Despite the high number of visibilities and frequency channels, the computational costs of fast-resolve are still moderate since fast-resolve only needs to evaluate the full radio interferometric measurement equation (equation 9.1) in the major cycles. As outlined above, 25 and 28 major cycles were executed for the LO and HI band. The fast-resolve reconstructions of both frequency bands were done on an NVIDIA A100 GPU. Before the actual fast-resolve imaging, the convolution kernels for the response and the noise were constructed on a CPU. For both frequency bands, less than two hours were needed on a single core of an Intel Xeon CPU to compute the convolution kernel. For imaging, 24 hours using an NVIDIA A100 GPU and eight cores of an Intel Xeon CPU were needed. Since most of the fast-resovle operators, except for the major update steps, are done on the GPU, the CPU was mostly idle during imaging. A limitation of the current implementation of the fast-resolve algorithm is the high GPU memory consumption, preventing applications with significantly larger grid sizes than those presented in this paper. We believe these constraints can be softened with future improvements in the implementation.

A direct one-to-one comparison of the computational cost to the reconstructions of Dabbech et al., 2022 is not possible, as the reconstructions of Dabbech et al., 2022 were executed on different hardware and also the algorithms are implemented in different programming languages. Nevertheless, the computational costs of the Dabbech et al., 2022 reconstructions tend to be significantly higher as they range between 900 and 3000 CPU hours per band and algorithm. Dabbech et al., 2022 for comparison also presented a WSClean reconstruction. For the WSClean reconstruction Dabbech et al., 2022 report a total computational cost of 132 and 236 CPU hours for the two imaging bands. Although CPU and GPU hours are not directly comparable, and hardware and programming languages differ, this underlines the computational performance and applicability of fast-resolve to imaging setups with massive datasets.

9.6 Conclusion

This chapter introduces the fast Bayesian imaging algorithm fast-resolve. fast-resolve combines the accuracy of the Bayesian imaging framework resovle with computational shortcuts of the CLEAN algorithm. This significantly broadens the applicability of Bayesian radio interferometric imaging. fast-resolve transforms the likelihood of the Bayesian radio interferometric imaging problem into a likelihood of a deconvolution problem, which is much faster to evaluate. Using the major/minor cycle scheme of CLEAN, fast-resolve corrects for inaccuracies of the transformed likelihood

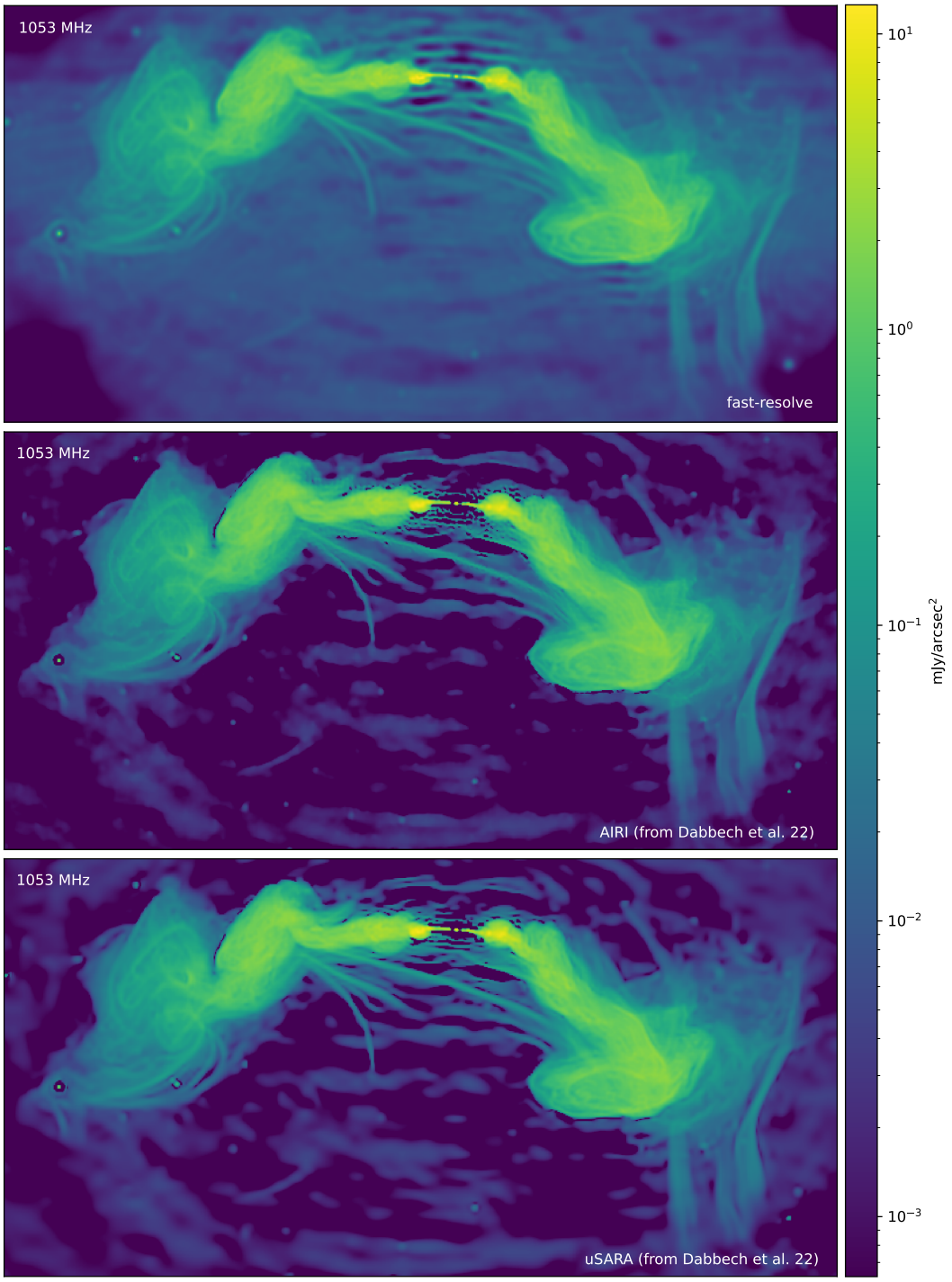

Figure 9.11: Radio Galaxy ESO 137-006 in the phase center of the observation. The top panel shows the `fast-resolve` reconstruction. For comparison the lower two panels display the reconstructions form Dabbech et al., 2022 with the AIRI and uSARA regularizers. All reconstructions, but especially `fast-resovle`, are affected by the suboptimal calibration.

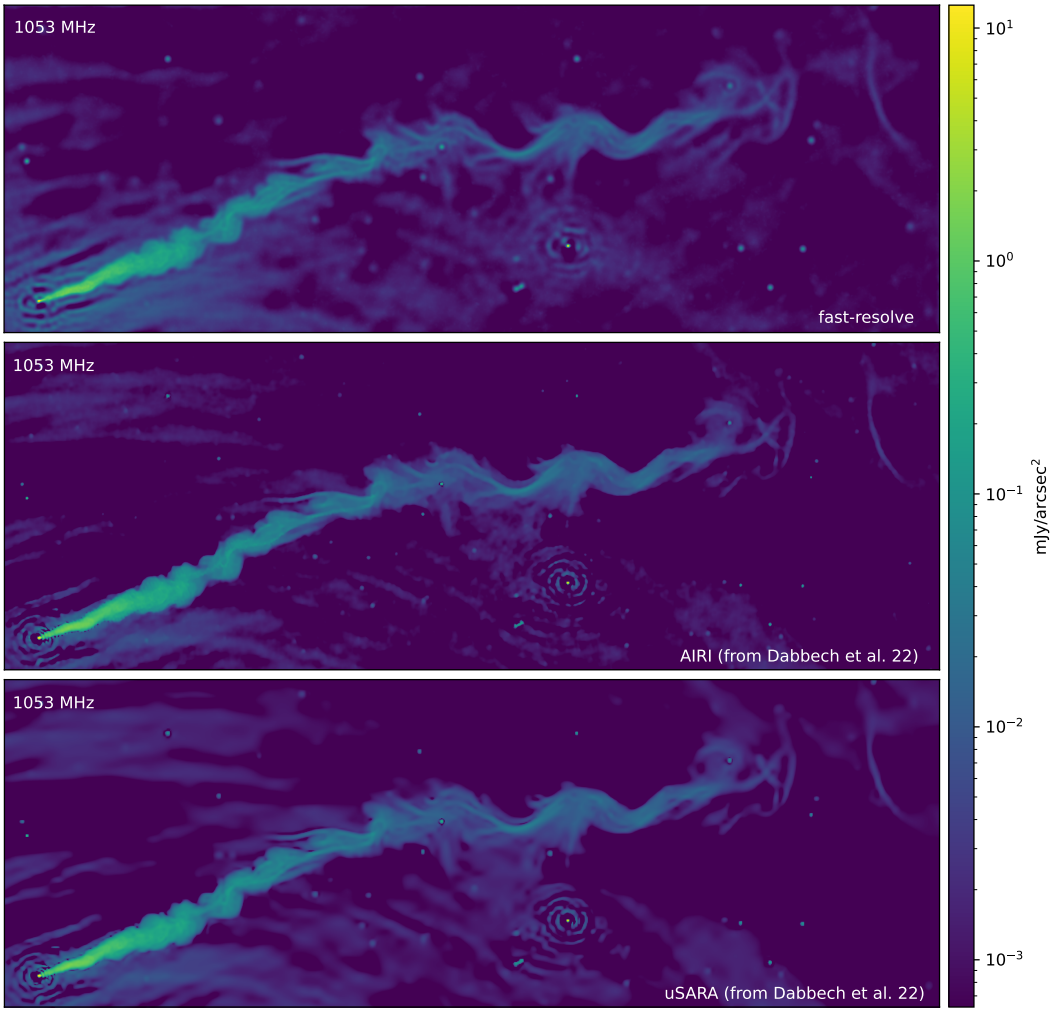

Figure 9.12: Radio Galaxy ESO 137-007 north of the phase center of the observation. The top panel shows the `fast-resolve` reconstruction. For comparison the lower two panels display the reconstructions form Dabbech et al., 2022 with the AIRI and uSARA regularizers. All reconstructions, but especially `fast-resovle`, are affected by the suboptimal calibration.

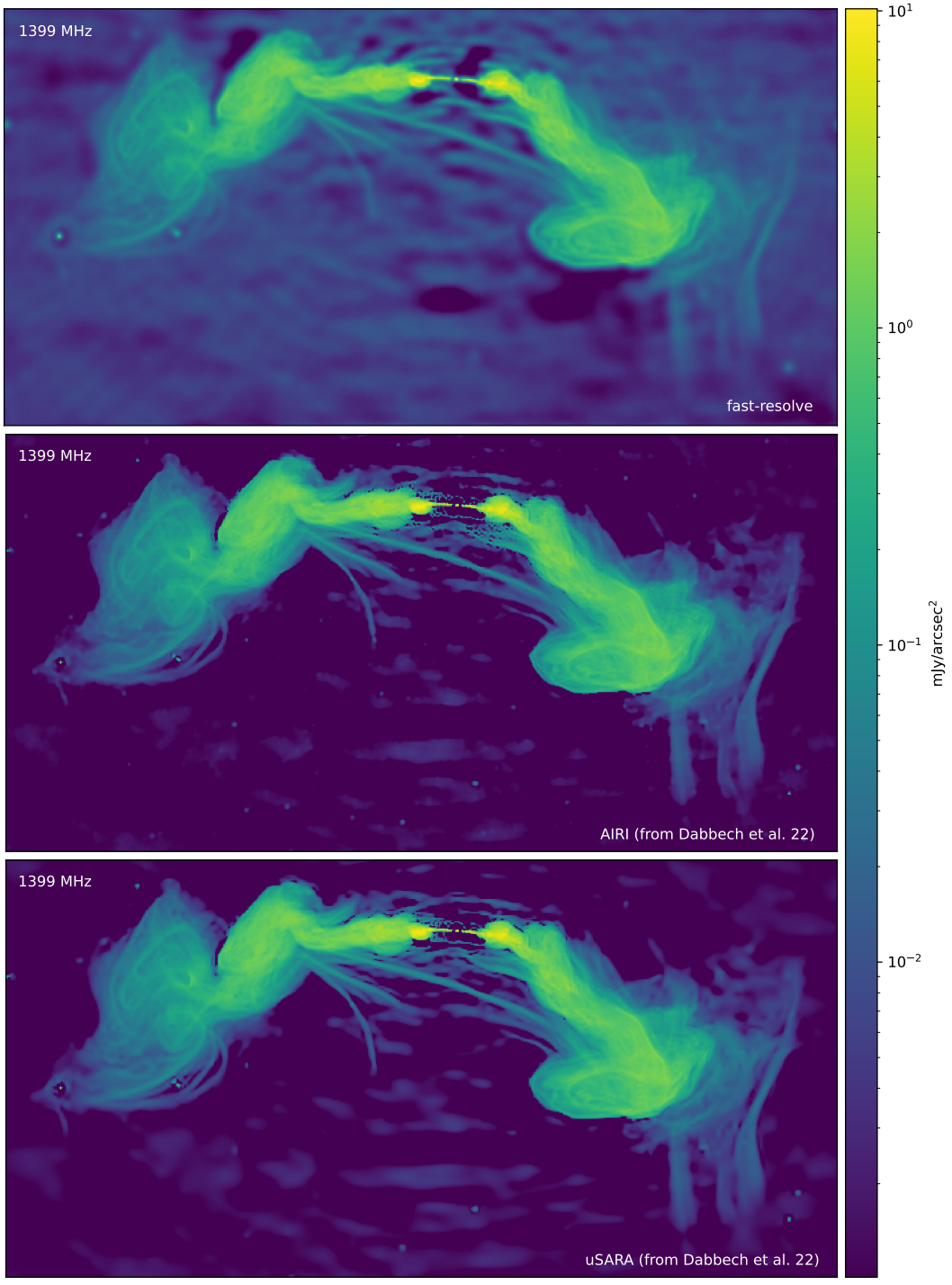

Figure 9.13: Like figure 9.11 but at 1399 MHz.

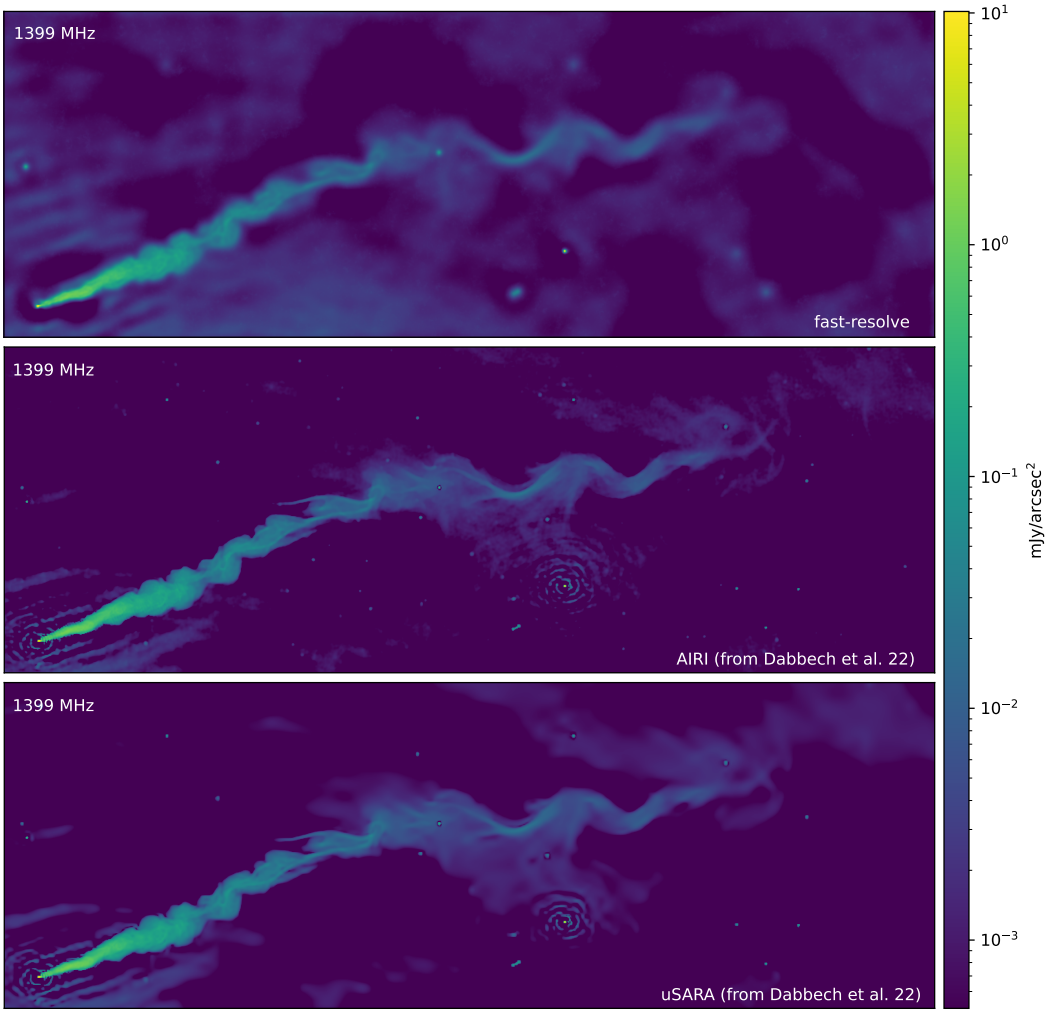

Figure 9.14: Like figure 9.12 but at 1399 MHz.

and accounts for the w-effect. The accuracy of `fast-resolve` is validated on Cygnus A VLA data by comparing with previous `resolve` and multi-scale `CLEAN` reconstructions. The comparison shows that `fast-resolve` achieves the same resolution as `resolve`. Likewise, the imaging artifacts are comparable and on a very low level.

Furthermore, the computational speed of `fast-resolve` is analyzed and compared to `resolve`, showing significant speedups for the VLA Cygnus A data. As `fast-resolve` is implemented in JAX, it can also be executed on a GPU, accelerating the reconstruction compared to the CPU by more than an order of magnitude. For the single channel Cygnus A VLA dataset, `fast-resolve` is at least 140 times faster than `resolve` when executed on a GPU. For datasets with more frequency channels, the computational advantages of `fast-resolve` can be even larger.

Additionally, we present a Bayesian image reconstruction of the radio galaxies ESO 137-006 and ESO 137-007 from MeerKAT data with `fast-resolve` and compare the results for validation with Dabbech et al., 2022. The MeerKAT dataset is significantly larger than the VLA datasets, but the computational costs of `fast-resolve` remain moderate due to the major/minor cycle scheme. A reconstruction of these sources with the classic `resolve` algorithm using the same amount of data would be computationally out of scope. To the best of our knowledge, no other Bayesian radio interferometric imaging algorithm has been successfully applied to a dataset of similar size before.

9.7 Data availability

The raw data of the Cygnus A observation is publicly available in the NRAO data archive[8] under project ID 14B-336. The raw data for the ESO137-006 observation is publicly available via the SARAO archive.[9] (project ID SCI-20190418-SM-01) The `fast-resolve` reconstruction results are archived on zenodo.[10] The implementation of the `fast-resolve` algorithm will be integrated into the `resolve` algorithm.[11]

9.8 Acknowledgements

The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. J. R. acknowledges financial support from the German Federal Ministry of Education and Research (BMBF) under grant 05A23WO1 (Verbundprojekt D-MeerKAT III). P. F. acknowledges funding through the German Federal Ministry of Education and Research for the project “ErUM-IFT: Informationsfeldtheorie für Experimente an Großforschungsanlagen” (Förderkennzeichen: 05D23EO1). O. M. S.’s research is supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation (grant No. 81737).

[8] `https://data.nrao.edu/portal/`
[9] `https://archive.sarao.ac.za`
[10] `https://doi.org/10.5281/zenodo.11549302`
[11] `https://gitlab.mpcdf.mpg.de/ift/resolve`

Table 9.3: Prior parameters for Cygnus A reconstructions.

Cygnus A, band:	S	C	X	Ku
zero mode offset	18	21	20	20
zero mode mean	1	1	1	1
zero mode stddev	0.1	0.1	0.1	0.1
fluctuations mean	5	5	5	5
fluctuations stddev	1	1	1	1
loglogavgslope mean	-2	-2	-2.2	-2.2
loglogavgslope stddev	0.2	0.2	0.2	0.2
flexibility mean	1.2	1.2	1.2	1.2
flexibility stddev	0.4	0.4	0.4	0.4
asperity mean	0.2	0.2	0.2	0.2
asperity stddev	0.2	0.2	0.2	0.2

Table 9.4: Prior parameters for ESO 137 reconstructions. For the two frequency bands, the LO ($961 - 1145\,\mathrm{MHz}$) and the HI ($1295 - 1503\,\mathrm{MHz}$) band we used the same hyper parameters.

ESO 137	006	007	background
zero mode offset	13	13	13
zero mode mean	1	1	1
zero mode stddev	0.1	0.1	0.1
fluctuations mean	3	4	4
fluctuations stddev	1	1	1
loglogavgslope mean	-2.4	-2.4	-2.0
loglogavgslope stddev	0.1	0.1	0.1
flexibility mean	0.1	0.1	1.2
flexibility stddev	0.01	0.1	0.4
asperity mean	0.01	0.01	0.2
asperity stddev	0.001	0.001	0.2

9.9 Appendix

9.9.1 Prior parameters for Cygnus A imaging

In table 9.3 we list the hyper parameters of the Gaussian process model for the diffuse emission in the Cygnus A reconstructions. The exact definition of these parameters is explained in detail in Arras et al., 2021a.

9.9.2 Prior parameters for ESO 137 imaging

In table 9.4 we list the hyper parameters of the Gaussian process model for the diffuse emission in the ESO 137 reconstructions. The exact definition of these parameters is explained in detail in Arras et al., 2021a.

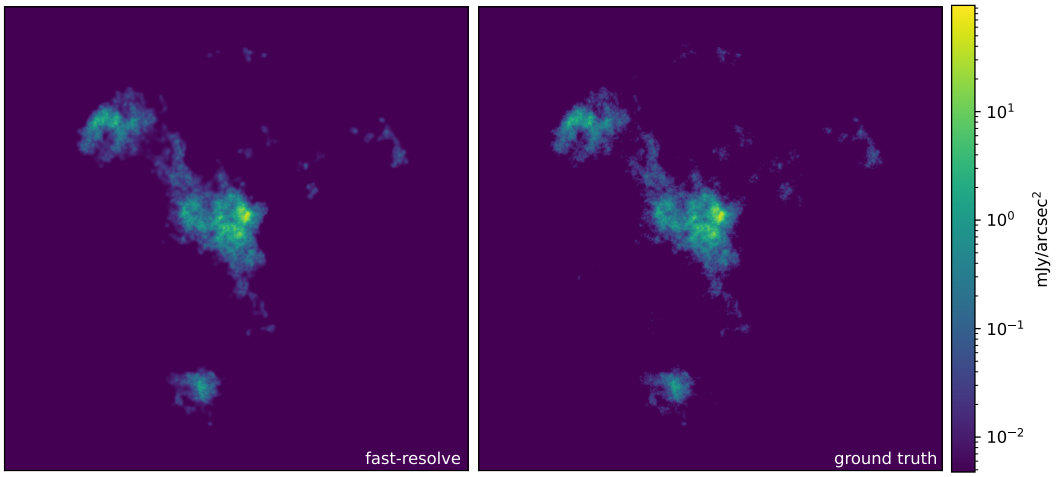

Figure 9.15: The left panel shows the `fast-resolve` reconstruction of the synthetic dataset. The right panel displays the underlying ground truth sky brightness. The field of view (2×2 deg) and the uvw-coverage are the same as in the MeerKAT LO band observation (see table 9.2).

9.9.3 Application to synthetic data

For further validation, we applied `fast-resolve` to synthetic data for which we know the underlying ground truth sky brightness distribution. The synthetic data mimics the setup of the MeerKAT ESO137-006 observation in the LO band. As in the MeerKAT example, a field of view of 2×2 deg is chosen. Furthermore, the uvw-coverage and frequency channels of the MeerKAT ESO137-006 LO band data (see table 9.2) are used. A prior sample from the diffuse emission sky brightness model was drawn to the generate sky brightness ground truth. The prior sample was multiplied with a taper function exponentially falling towards zero at the edges of the field of view to ensure no bright emission is at the edges. The resulting sky brightness distribution used as ground truth is displayed in figure 9.15. Model visibilities were computed from this ground truth. Gaussian noise was added to the mode visibilities using the weights of the real ESO137-006 observation for the inverse noise covariance. The resulting dataset was reconstructed with `fast-resolve`.

The reconstruction results are displayed in figure 9.15. All bright emission features are accurately reconstructed. Also, fainter features could successfully be reconstructed, although with a lower resolution. Some very faint features could not be recovered. Overall, no significant artifacts are visible in the reconstruction. Specifically, no artifacts similar to the spurious synchrotron threads or the ripple around bright sources as in the real MeerKAT data reconstruction are visible. As discussed in the main text, we assume that a suboptimal calibration creates these artifacts in the MeerKAT data reconstruction. The absence of similar artifacts in the synthetic data supports this assumption, as no calibration errors were added to the synthetic data.

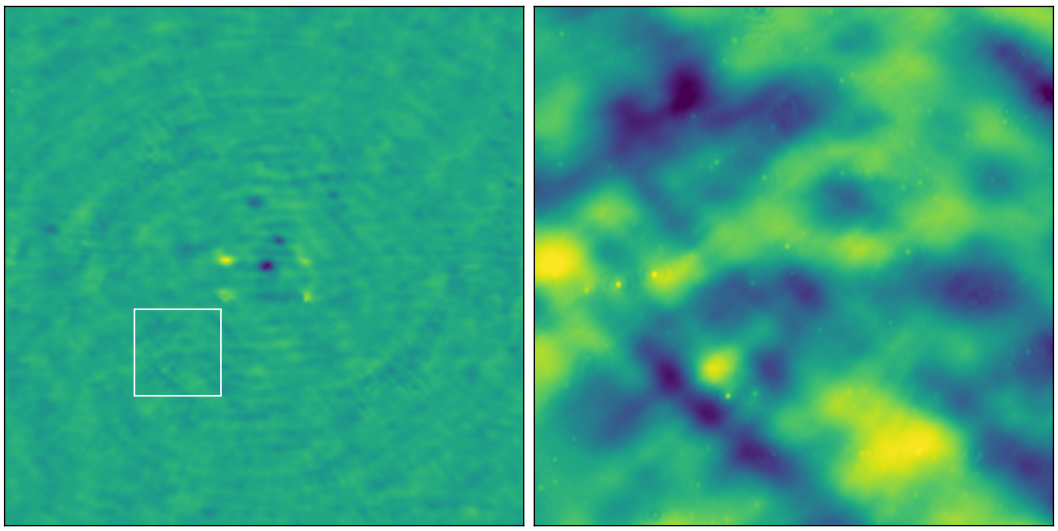

Figure 9.16: The left panel shows the fast-resolve residual data of the last minor cycle in arbitrary units. On large scales, no significant structures of the source are left in the residual, except for some small structures at the corners of the box containing the source ESO137-006. The right panel shows a zoom into a representative region marked by the box on the left. The zoomed image shows many point sources not fitted by the fast-resolve reconstruction but left in the residual data.

9.9.4 Residual data

In figure 9.16, the fast-resolve residual data of the last major cycle of the MeerKAT ESO137-006 LO band reconstruction is depicted. The residual data of the HI band reconstruction looks similar. The left shows the residual for the full field of view. On large scales, the edges of the box containing the discussed emission prior for the source ESO137-006 (see figure 9.10) are visible. Otherwise, only very little structure of the primary sources remained in the residual.

As discussed in the main text for the MeerKAT observations, only the diffuse emission prior was used, which also modeled the flux of the point sources. The diffuse emission prior is not tailored to point sources, reducing sensitivity. For this reason, many point sources are not captured by the reconstruction but are left in the residual. The right panel of figure 9.16 shows a zoom into a representative region of the residual data image. As expected, many faint point sources that were not (fully) captured by the reconstruction are visible in the zoomed image. For observations like the ESO137-006 MeerKAT data, a dedicated point source model for resolve and fast-resolve, capable of handling many sources, would be beneficial but is left for future work.

Chapter 10

Other contributions

The previous chapters presented the results of my primary PhD projects. This chapter outlines contributions I made to other projects. The connecting element of all projects is their methodology, building on Bayesian inference.

10.1 VLBI calibration

In the development of the radio interferometric imaging and calibration methods described in the chapters 6 and 9, I concentrated on the application to GHz data of the single-site radio interferometers VLA and MeerKAT. Mathematically, very long baseline interferometry (VLBI) with antennas distributed over much larger distances is equivalent to interferometry with nearby antennas. The only difference is that the uvw-coordinates in the measurement equation 3.27 are significantly larger. For this reason, very similar methods can be used for the calibration and imaging of VLBI data. I contributed to the Bayesian calibration and imaging of Very Long Baseline Array (VLBA) 43 Ghz data of the source M87, led by Jong-Seo Kim and described in Kim et al., 2024. In contrast to the algorithm in chapter 6, this work neglected the direction-dependent effects of the antenna gains. Nevertheless, a higher resolution and an improved reconstruction of the jet and counter-jet of M87 could be achieved compared to the standard imaging algorithm `CLEAN`. By an analysis of simulated data, the accuracy of the Bayesian calibration and imaging method was validated.

In the future, we plan to extend the algorithm for full polarization imaging and especially incorporate a calibration of the polarization leakage in the antenna receivers. However, this work is not yet published.

10.2 Bayesian imaging with ALMA

In addition to applying Bayesian imaging to VLBI data, I also contributed to the development of new imaging methods for the Atacama Large Millimeter/submillimeter Array (ALMA). ALMA is a single-site radio interferometer similar to VLA and MeerKAT, measuring at significantly higher frequencies. This project aims at developing novel imaging techniques for ALMA and builds besides other methods on the `resolve` framework. The

project is led by Fabrizia Guglielmetti and an overview is given in Guglielmetti et al., 2023. The part of the method development building on `resolve` is currently led by Vishal Johnson, aiming at obtaining spectrally resolved images from ALMA data.

10.3 Gravitational lensing

I was also involved in gravitational lens reconstruction projects. Gravitational lens reconstructions are a challenging inverse problem, similar to radio interferometric calibration and imaging. In radio interferometric data reconstructions, there is an ambiguity between the antenna gains and the sky brightness. In gravitational lens reconstructions, there is an ambiguity between the brightness distribution of the lensed source and the mass distribution in the lens. Like all inverse problems, also gravitational lensing can be understood as a Bayesian inference problem as outlined in chapter 4.

I contributed to two Bayesian gravitational lens reconstruction projects. Specifically, I contributed to the development of the `LensCharm` framework led by Julian Rüstig and Mattheo Guardiani. `LenseCharm` provides parametric and nonparametric lens and source models and can reconstruct strong-lensing systems using the variational inference methods discussed in chapter 4. In the publication Rüstig et al., 2024 `LensCharm` was introduced and validated on simulated data as well as real JWST observations. Furthermore, I contributed to the work of Aymeric Galan, combining the `HercuLens` (Galan et al., 2022) package with Bayesian inference. `HercuLens` combined with Bayesian inference was successfully applied in Galan et al., 2024 to JWST data of a galaxy cluster lense.

10.4 Nested sampling

Besides the astrophysics projects mentioned above, I also contributed to a statistics project improving the estimation of Bayesian evidence $P(d)$ (see equation 4.4) from nested sampling data. While the evidence does not need to be computed for many Bayesian inference applications, such as the ones presented in this thesis, accurate evidence estimates are required for model selection tasks. The evidence estimate of nested sampling relies on stochastic estimates of the prior volumes enclosed by likelihood contours. The prior volume estimation itself can be turned into a Bayesian inference problem. By incorporating additional information as a prior into this Bayesian inference problem, the uncertainty on the relation between likelihood and prior volume can be reduced. Specifically, the prior information we incorporate is that in many applications the likelihood prior volume relation is a smooth function. Ultimately, this increases the accuracy of the nested sampling evidence estimate. This project was led by Margret Westerkamp, and progress on this idea has been presented in Westerkamp et al., 2023 and Westerkamp et al., 2024.

10.5 NIFTy

Furthermore, I contributed to the development of the Bayesian inference package NIFTy (Selig et al., 2013b; Steininger et al., 2017; Arras et al., 2019a), on which the Bayesian

inference applications of this thesis build. The original NIFTy implementation was built on NumPy for numerical computations and included a custom automatic differentiation engine in later versions to enable efficient gradient-based optimization. The use of our NumPy-based automatic differentiation engine became unfavorable with the development of machine learning frameworks such as JAX that natively support automatic differentiation and GPU-accelerated computing. For these reasons, NIFTy was reimplemented and modernized in JAX and named NIFTy.re (Edenhofer et al., 2024a). The algorithms of the chapters 5 and 6 still build on the old NIFTy implementation, while the `fast-resolve` algorithm (chapter 9) utilizes the JAX accelerated NIFTy.re. I contributed to both the original NumPy-based implementation and the JAX implementation NIFTy.re.

Chapter 11

Conclusion

11.1 Summary

The Earth's atmosphere is opaque to radiation for most of the electromagnetic spectrum. However, the atmosphere is transparent to optical light and radio waves, allowing us to study the cosmos at these wavelengths with ground-based telescopes. This thesis contributes software and algorithms to improve the observational capabilities of ground-based telescopes. The algorithms are based on Bayesian inference to reconstruct images of the sky with improved resolution and dynamic range from raw telescope data. The probabilistic approach to the imaging problem enables reliable uncertainty quantification for the reconstructed images.

The first part of the thesis addresses the reconstruction of images in the presence of atmospheric perturbations. Specifically, chapter 5 contributes a high-contrast imaging algorithm for optical telescopes. This algorithm can reconstruct faint companions, such as exoplanets, at a sub-arcsecond distance from the host star. As shown in chapter 5, this algorithm outperforms other reconstruction algorithms on short high-cadence frame sequences. The direction-dependent calibration and imaging algorithm for radio interferometry presented in chapter 6 goes in a similar direction. It can reconstruct direction and time-dependent variations of the antenna gains simultaneously with the radio sky image. Incorporating the direction-dependent effects leads to an increased dynamic range and resolution of the recovered images, also in comparison with another direction-dependent calibration algorithm building on compressed sensing. The accurate calibration of direction-dependent effects is important, especially in imaging modalities where very high dynamic ranges are required, such as for very bright sources as Cygnus A or the deep field observations as presented in the science case in the introduction (section 1.1.2).

The second part of the thesis focuses on the challenges of image reconstruction in the era of modern radio interferometry with large data volumes. Chapter 8 presents the software package `JAXbind` to easily bind external code into the machine learning framework JAX. While `JAXbind` is useful for many projects, the use case for radio interferometry is to bind high-performance implementations of the instrument response into the JAX framework. This allows us to utilize the advantages of the JAX programming framework for `resolve`. Additionally, `JAXbind` has already been used for binding spherical harmonic transformation

of the DUCC package (Reinecke, 2024) to JAX, as well as binding the multi-grid solver PICARD (Kissmann, 2014) for cosmic ray propagation to JAX.

The traditional `CLEAN` imaging algorithm has many limitations, such as, suboptimal resolution of the recovered images. Many advanced imaging algorithms, such as the `resolve` framework, have been developed in recent years to improve radio interferometric imaging capabilities. While `resolve`, and in particular the method presented in chapter 6, have significantly improved the quality of the recovered images, these improvements come at an increased computational cost compared to the standard imaging method `CLEAN`. This increase in computational cost becomes prohibitively expensive for current flagship interferometers such as MeerKAT or future arrays. Similarly, many other advanced imaging algorithms are computationally significantly more expensive than `CLEAN`. For this reason, none of the novel imaging algorithms has yet been widely adopted to replace the traditional `CLEAN` algorithm. `fast-resolve` presented in chapter 9 translates some of the computational shortcuts of the `CLEAN` algorithm into a Bayesian reconstruction algorithm. While the accuracy of the `fast-resolve` reconstructions is on par with `resolve`, the runtime of the reconstructions is reduced by orders of magnitude. Therefore, `fast-resolve` has the potential to overcome the computational limitations of advanced imaging algorithms and make Bayesian imaging applicable to a much larger range of datasets. As a demonstration, chapter 9 presented a `fast-resolve` reconstruction of a large MeerKAT dataset. To my knowledge, no other Bayesian imaging algorithm has previously reconstructed a similar-sized radio interferometric dataset.

11.2 Conclusion and Outlook

The developments of this thesis demonstrate the impact of high-accuracy reconstruction algorithms on the observational capabilities of telescopes. Such developments are critical, as improvements in telescope hardware alone are not sufficient for many science cases, as outlined in the examples in the introduction (section 1.1). Among the algorithms in this thesis, `fast-resolve` probably has the largest potential to significantly impact the future of imaging with a ground-based telescope. `fast-resolve`, or an algorithm based on its methodology, has the potential to become a widely applicable radio interferometric imaging algorithm, overcoming the limitations of the current `CLEAN`-based imaging, and thereby significantly improving the observational capabilities. This is a major result of this thesis.

Chapter 9 demonstrated the accuracy and scalability of the `fast-resolve` concept. Additional capabilities are required to turn `fast-resolve` into a widely applicable imaging algorithm. These include polarization and multi-frequency imaging. Furthermore, antenna gain calibration methods should be added to `fast-resolve`, ideally incorporating direction-dependent effects as presented in chapter 6 for the classic `resolve` framework. These developments are left for the future, and I am curious to see which of them will become reality, and how the algorithms for which this thesis has laid the foundations will evolve.

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: https://www.tensorflow.org/.

Abdulaziz, A., Dabbech, A., and Wiaux, Y. (Oct. 2019). "Wideband super-resolution imaging in Radio Interferometry via low rankness and joint average sparsity models (Hyper-SARA)". In: *Monthly Notices of the Royal Astronomical Society* 489.1, pp. 1230–1248. DOI: 10.1093/mnras/stz2117.

Aghabiglou, A., Chu, C. S., Jackson, A., Dabbech, A., and Wiaux, Y. (2023). "Ultra-fast high-dynamic range imaging of Cygnus A with the R2D2 deep neural network series". In: arXiv: 2309.03291 [astro-ph.IM].

Aghabiglou, A., San Chu, C., Dabbech, A., and Wiaux, Y. (Mar. 2024). "The R2D2 deep neural network series paradigm for fast precision imaging in radio astronomy". In: *arXiv e-prints*, arXiv:2403.05452, arXiv:2403.05452. DOI: 10.48550/arXiv.2403.05452.

Albert, J. G., Weeren, R. J. van, Intema, H. T., and Röttgering, H. J. A. (2020). "Probabilistic direction-dependent ionospheric calibration for LOFAR-HBA". In: *Astronomy & Astrophysics* 635, A147. DOI: 10.1051/0004-6361/201937424.

Amara, A. and Quanz, S. P. (Nov. 2012). "pynpoint: an image processing package for finding exoplanets". In: *Monthly Notices of the Royal Astronomical Society* 427.2, pp. 948–955. DOI: 10.1111/j.1365-2966.2012.21918.x.

Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., Bao, B., Bell, P., Berard, D., Burovski, E., et al. (Apr. 2024). "PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation". In: *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*. ACM. DOI: 10.1145/3620665.3640366.

Arras, P., Baltac, M., Ensslin, T. A., Frank, P., Hutschenreuter, S., Knollmueller, J., Leike, R., Newrzella, M.-N., Platz, L., Reinecke, M., et al. (Mar. 2019a). *NIFTy5: Numerical Information Field Theory v5*. Astrophysics Source Code Library, record ascl:1903.008. ascl: 1903.008.

Arras, P., Bester, H. L., Perley, R. A., Leike, R., Smirnov, O., Westermann, R., and Enßlin, T. A. (Feb. 2021a). "Comparison of classical and Bayesian imaging in radio interferometry. Cygnus A with CLEAN and resolve". In: *Astronomy & Astrophysics* 646, A84, A84. DOI: 10.1051/0004-6361/202039258.

Arras, P., Frank, P., Haim, P., Knollmüller, J., Leike, R., Reinecke, M., and Enßlin, T. (Jan. 2022). "Variable structures in M87* from space, time and frequency resolved interferometry". In: *Nature Astronomy* 6, pp. 259–269. DOI: 10.1038/s41550-021-01548-0.

Arras, P., Frank, P., Leike, R., Westermann, R., and Enßlin, T. A. (July 2019b). "Unified radio interferometric calibration and imaging with joint uncertainty quantification". In: *Astronomy & Astrophysics* 627, A134, A134. DOI: 10.1051/0004-6361/201935555.

Arras, P., Reinecke, M., Westermann, R., and Enßlin, T. A. (Feb. 2021b). "Efficient wide-field radio interferometry response". In: *Astronomy & Astrophysics* 646, A58, A58. DOI: 10.1051/0004-6361/202039723.

Beuzit, J. L., Vigan, A., Mouillet, D., Dohlen, K., Gratton, R., Boccaletti, A., Sauvage, J. .-., Schmid, H. M., Langlois, M., Petit, C., et al. (Nov. 2019). "SPHERE: the exoplanet imager for the Very Large Telescope". In: *Astronomy & Astrophysics* 631, A155, A155. DOI: 10.1051/0004-6361/201935251.

Bhatnagar, S. and Cornwell, T. J. (Nov. 2004). "Scale sensitive deconvolution of interferometric images. I. Adaptive Scale Pixel (Asp) decomposition". In: *Astronomy & Astrophysics* 426, pp. 747–754. DOI: 10.1051/0004-6361:20040354.

Bhatnagar, S., Cornwell, T. J., Golap, K., and Uson, J. M. (2008). "Correcting direction-dependent gains in the deconvolution of radio interferometric images". In: *Astronomy & Astrophysics* 487.1, pp. 419–429. DOI: 10.1051/0004-6361:20079284.

Birdi, J., Repetti, A., and Wiaux, Y. (Aug. 2018). "Sparse interferometric Stokes imaging under the polarization constraint (Polarized SARA)". In: *Monthly Notices of the Royal Astronomical Society* 478.4, pp. 4442–4463. DOI: 10.1093/mnras/sty1182.

Birdi, J., Repetti, A., and Wiaux, Y. (Mar. 2020). "Polca SARA - full polarization, direction-dependent calibration, and sparse imaging for radio interferometry". In: *Monthly Notices of the Royal Astronomical Society* 492.3, pp. 3509–3528. DOI: 10.1093/mnras/stz3555.

Blei, D. M., Kucukelbir, A., and D., M. J. (2017). "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518, pp. 859–877. DOI: 10.1080/01621459.2017.1285773.

Booth, R. S., de Blok, W. J. G., Jonas, J. L., and Fanaroff, B. (Oct. 2009). "MeerKAT Key Project Science, Specifications, and Proposals". In: *arXiv e-prints*, arXiv:0910.2935, arXiv:0910.2935. DOI: 10.48550/arXiv.0910.2935.

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., et al. (2018). *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. URL: http://github.com/google/jax.

Brooks, S., Gelman, A., Jones, G., and Meng, X. L. (2011). *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC. DOI: 10.1201/b10905.

Burke, B. F., Graham-Smith, F., and Wilkinson, P. N. (2019). *An Introduction to Radio Astronomy*. 4th ed. Cambridge University Press. DOI: 10.1017/9781316987506.

Cai, X., Pereyra, M., and McEwen, J. D. (Nov. 2018). "Uncertainty quantification for radio interferometric imaging - I. Proximal MCMC methods". In: *Monthly Notices of the Royal Astronomical Society* 480.3, pp. 4154–4169. DOI: 10.1093/mnras/sty2004.

Connor, L., Bouman, K. L., Ravi, V., and Hallinan, G. (Aug. 2022). "Deep radio-interferometric imaging with POLISH: DSA-2000 and weak lensing". In: *Monthly No-*

tices of the Royal Astronomical Society 514.2, pp. 2614–2626. DOI: 10.1093/mnras/stac1329.

Cornwell, T. J. and Evans, K. F. (Feb. 1985). "A simple maximum entropy deconvolution algorithm". In: *Astronomy & Astrophysics* 143.1, pp. 77–83.

Cornwell, T. J. and Perley, R. A. (July 1992). "Radio-interferometric imaging of very large fields. The problem of non-coplanar arrays." In: *Astronomy & Astrophysics* 261, pp. 353–364.

Cornwell, T. J., Golap, K., and Bhatnagar, S. (2008). "The Noncoplanar Baselines Effect in Radio Interferometry: The W-Projection Algorithm". In: *IEEE Journal of Selected Topics in Signal Processing* 2.5, pp. 647–657. DOI: 10.1109/JSTSP.2008.2005290.

Dabbech, A., Onose, A., Abdulaziz, A., Perley, R. A., Smirnov, O. M., and Wiaux, Y. (Feb. 2018). "Cygnus A super-resolved via convex optimization from VLA data". In: *Monthly Notices of the Royal Astronomical Society* 476.3, pp. 2853–2866. ISSN: 0035-8711. DOI: 10.1093/mnras/sty372.

Dabbech, A., Repetti, A., Perley, R. A., Smirnov, O. M., and Wiaux, Y. (July 2021). "Cygnus A jointly calibrated and imaged via non-convex optimization from VLA data". In: *Monthly Notices of the Royal Astronomical Society* 506.4, pp. 4855–4876. ISSN: 0035-8711. DOI: 10.1093/mnras/stab1903.

Dabbech, A., Terris, M., Jackson, A., Ramatsoku, M., Smirnov, O. M., and Wiaux, Y. (Nov. 2022). "First AI for Deep Super-resolution Wide-field Imaging in Radio Astronomy: Unveiling Structure in ESO 137-006". In: *The Astrophysical Journal Letters* 939.1, L4, p. L4. DOI: 10.3847/2041-8213/ac98af.

Dabbech, A., Repetti, A., and Wiaux, Y. (2019). "Self Direction-Dependent Calibration For Wideband Radio-Interferometric Imaging". In: *International BASP Frontiers workshop 2019*, p. 23. URL: http://archive.baspfrontiers.org/archive/2019/proceedings.php.html.

Devroye, L. (1986). "Sample-based non-uniform random variate generation". In: *Proceedings of the 18th Conference on Winter Simulation*. WSC '86. Washington, D.C., USA: Association for Computing Machinery, pp. 260–265. ISBN: 0911801111. DOI: 10.1145/318242.318443.

Edenhofer, G., Frank, P., Roth, J., Leike, R. H., Guerdi, M., Scheel-Platz, L. I., Guardiani, M., Eberle, V., Westerkamp, M., and Enßlin, T. A. (2024a). "Re-Envisioning Numerical Information Field Theory (NIFTy.re): A Library for Gaussian Processes and Variational Inference". In: *Journal of Open Source Software* 9.98, p. 6593. DOI: 10.21105/joss.06593.

Edenhofer, G., Zucker, C., Frank, P., Saydjari, A. K., Speagle, J. S., Finkbeiner, D., and Enßlin, T. A. (May 2024b). "A parsec-scale Galactic 3D dust map out to 1.25 kpc from the Sun". In: *Astronomy & Astrophysics* 685, A82, A82. DOI: 10.1051/0004-6361/202347628.

Enßlin, T. (2022). "Information Field Theory and Artificial Intelligence". In: *Entropy* 24.3. ISSN: 1099-4300. DOI: 10.3390/e24030374.

Enßlin, T. A. (2019). "Information Theory for Fields". In: *Annalen der Physik* 531.3, p. 1800127. DOI: https://doi.org/10.1002/andp.201800127.

Enßlin, T. A., Frommert, M., and Kitaura, F. S. (Nov. 2009). "Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis". In: *Physical Review D* 80.10, 105005, p. 105005. DOI: 10.1103/PhysRevD.80.105005.

Esposito, S., Riccardi, A., Fini, L., Puglisi, A. T., Pinna, E., Xompero, M., Briguglio, R., Quirós-Pacheco, F., Stefanini, P., Guerra, J. C., et al. (July 2010). "First light AO (FLAO) system for LBT: final integration, acceptance test in Europe, and preliminary on-sky commissioning results". In: *Adaptive Optics Systems II*. Ed. by B. L. Ellerbroek, M. Hart, N. Hubin, and P. L. Wizinowich. Vol. 7736. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 773609, p. 773609. DOI: 10.1117/12.858194.

Flasseur, O., Thé, S., Denis, L., Thiébaut, É., and Langlois, M. (July 2021). "REXPACO: An algorithm for high contrast reconstruction of the circumstellar environment by angular differential imaging". In: *Astronomy & Astrophysics* 651, A62, A62. DOI: 10.1051/0004-6361/202038957.

Frank, P., Leike, R., and Enßlin, T. A. (2021). "Geometric Variational Inference". In: *Entropy* 23.7. ISSN: 1099-4300. DOI: 10.3390/e23070853.

Galan, A., Caminha, G. B., Knollmüller, J., Roth, J., and Suyu, S. H. (Feb. 2024). "El Gordo needs El Anzuelo: Probing the structure of cluster members with multi-band extended arcs in JWST data". In: *arXiv e-prints*, arXiv:2402.18636, arXiv:2402.18636. DOI: 10.48550/arXiv.2402.18636.

Galan, A., Vernardos, G., Peel, A., Courbin, F., and Starck, J. .-. (Dec. 2022). "Using wavelets to capture deviations from smoothness in galaxy-scale strong lenses". In: *Astronomy & Astrophysics* 668, A155, A155. DOI: 10.1051/0004-6361/202244464.

Genton, M. G. (Mar. 2002). "Classes of Kernels for Machine Learning: A Statistics Perspective". In: *J. Mach. Learn. Res.* 2, pp. 299–312. ISSN: 1532-4435.

Gilmozzi, R. and Spyromilio, J. (2007). "The European extremely large telescope (E-ELT)". In: *The Messenger* 127.11, p. 3.

GRAVITY Collaboration, Abuter, R., Aimar, N., Amorim, A., Arras, P., Bauböck, M., Berger, J. P., Bonnet, H., Brandner, W., Bourdarot, G., et al. (Jan. 2022). "Deep images of the Galactic center with GRAVITY". In: *Astronomy & Astrophysics* 657, A82, A82. DOI: 10.1051/0004-6361/202142459.

Greiner, M., Vacca, V., Junklewitz, H., and Enßlin, T. A. (May 2016). "fastRESOLVE: fast Bayesian imaging for aperture synthesis in radio astronomy". In: *arXiv e-prints*, arXiv:1605.04317, arXiv:1605.04317. DOI: 10.48550/arXiv.1605.04317.

Guardiani, M., Frank, P., Kostić, A., Edenhofer, G., Roth, J., Uhlmann, B., and Enßlin, T. (Oct. 2022). "Causal, Bayesian, & non-parametric modeling of the SARS-CoV-2 viral load distribution vs. patient's age". In: *PLOS ONE* 17.10. Ed. by C. Strumann, e0275011. DOI: 10.1371/journal.pone.0275011.

Guglielmetti, F., Delli Veneri, M., Baronchelli, I., Blanco, C., Dosi, A., Enßlin, T., Johnson, V., Longo, G., Roth, J., Stoehr, F., et al. (2023). "A BRAIN Study to Tackle Image Analysis with Artificial Intelligence in the ALMA 2030 Era". In: *Physical Sciences Forum* 9.1. ISSN: 2673-9984. DOI: 10.3390/psf2023009018.

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., et al. (Sept. 2020). "Array programming

with NumPy". In: *Nature* 585.7825, pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

He, Y., Zhao, X., Yang, D., Wu, Y., and Li, Q. (2022). "A study to investigate the relationship between ionospheric disturbance and seismic activity based on Swarm satellite data". In: *Physics of the Earth and Planetary Interiors* 323, p. 106826. ISSN: 0031-9201. DOI: https://doi.org/10.1016/j.pepi.2021.106826.

Heywood, I., Jarvis, M. J., Hale, C. L., Whittam, I. H., Bester, H. L., Hugo, B., Kenyon, J. S., Prescott, M., Smirnov, O. M., Tasse, C., et al. (Jan. 2022). "MIGHTEE: total intensity radio continuum imaging and the COSMOS/XMM-LSS Early Science fields". In: *Monthly Notices of the Royal Astronomical Society* 509.2, pp. 2150–2168. DOI: 10.1093/mnras/stab3021.

Högbom, J. A. (June 1974). "Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines". In: *Astronomy and Astrophysics Supplement* 15, p. 417.

Hook, I. (2009). "The Science Case for the European ELT". In: *Science with the VLT in the ELT Era*. Ed. by A. Moorwood. Dordrecht: Springer Netherlands, pp. 225–232. ISBN: 978-1-4020-9190-2.

Hope, D. A., Jefferies, S. M., Li Causi, G., Landoni, M., Stangalini, M., Pedichini, F., and Antoniucci, S. (Feb. 2022). "Post-AO High-resolution Imaging Using the Kraken Multiframe Blind Deconvolution Algorithm". In: *The Astrophysical Journal* 926.1, 88, p. 88. DOI: 10.3847/1538-4357/ac2df3.

Hutschenreuter, S. and Enßlin, T. A. (Jan. 2020). "The Galactic Faraday depth sky revisited". In: *Astronomy & Astrophysics* 633, A150, A150. DOI: 10.1051/0004-6361/201935479.

Hutschenreuter, S., Haverkorn, M., Frank, P., Raycheva, N. C., and Enßlin, T. A. (Apr. 2023). "Disentangling the Faraday rotation sky". In: *arXiv e-prints*, arXiv:2304.12350, arXiv:2304.12350. DOI: 10.48550/arXiv.2304.12350.

Jansky, K. G. (July 1933). "Radio Waves from Outside the Solar System". In: *Nature* 132.3323, p. 66. DOI: 10.1038/132066a0.

Jarvis, M., Taylor, R., Agudo, I., Allison, J. R., Deane, R. P., Frank, B., Gupta, N., Heywood, I., Maddox, N., McAlpine, K., et al. (Jan. 2016). "The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey". In: *MeerKAT Science: On the Pathway to the SKA*, 6, p. 6. DOI: 10.22323/1.277.0006.

Jonas, J. and MeerKAT Team (Jan. 2016). "The MeerKAT Radio Telescope". In: *MeerKAT Science: On the Pathway to the SKA*, 1, p. 1. DOI: 10.22323/1.277.0001.

Józsa, G. I. G., White, S. V., Thorat, K., Smirnov, O. M., Serra, P., Ramatsoku, M., Ramaila, A. J. T., Perkins, S. J., Maccagni, F. M., Makhathini, S., et al. (Jan. 2020). "MeerKATHI - an End-to-End Data Reduction Pipeline for MeerKAT and Other Radio Telescopes". In: *Astronomical Data Analysis Software and Systems XXIX*. Ed. by R. Pizzo, E. R. Deul, J. D. Mol, J. de Plaa, and H. Verkouter. Vol. 527. Astronomical Society of the Pacific Conference Series, p. 635. DOI: 10.48550/arXiv.2006.02955.

Junklewitz, H., Bell, M. R., Selig, M., and Enßlin, T. A. (Feb. 2016). "RESOLVE: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy". In: *Astronomy & Astrophysics* 586, A76, A76. DOI: 10.1051/0004-6361/201323094.

Kenyon, J. S., Smirnov, O. M., Grobler, T. L., and Perkins, S. J. (Aug. 2018). "CUBICAL - fast radio interferometric calibration suite exploiting complex optimization". In:

Monthly Notices of the Royal Astronomical Society 478.2, pp. 2399–2415. DOI: 10.1093/mnras/sty1221.

Khintchine, A. (Dec. 1934). "Korrelationstheorie der stationären stochastischen Prozesse". In: *Mathematische Annalen* 109.1, pp. 604–615. DOI: 10.1007/bf01449156.

Kim, J.-S., Nikonov, A. S., Roth, J., Ensslin, T. A., Janssen, M., Arras, P., Mueller, H., and Lobanov, A. P. (July 2024). "Bayesian self-calibration and imaging in Very Long Baseline Interferometry". In: *arXiv e-prints*, arXiv:2407.14873, arXiv:2407.14873. DOI: 10.48550/arXiv.2407.14873.

Kingma, D. P. and Welling, M. (Dec. 2013). "Auto-Encoding Variational Bayes". In: *arXiv e-prints*, arXiv:1312.6114, arXiv:1312.6114. DOI: 10.48550/arXiv.1312.6114.

Kissmann, R. (2014). "PICARD: A novel code for the Galactic Cosmic Ray propagation problem". In: *Astroparticle Physics* 55, pp. 37–50. DOI: https://doi.org/10.1016/j.astropartphys.2014.02.002.

Knollmüller, J. and Enßlin, T. A. (Dec. 2018). "Encoding prior knowledge in the structure of the likelihood". In: arXiv: 1812.04403 [stat.ML].

Knollmüller, J. and Enßlin, T. A. (Jan. 2019). "Metric Gaussian Variational Inference". In: arXiv: 1901.11033 [stat.ML].

Kolmogorov, A. N. (1991). "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers". In: *Proceedings of the Royal Society A* 434.1890, pp. 9–13. DOI: 10.1098/rspa.1991.0075.

Kullback, S. and Leibler, R. A. (1951). "On Information and Sufficiency". In: *The Annals of Mathematical Statistics* 22.1, pp. 79–86. DOI: 10.1214/aoms/1177729694.

Labate, M. G., Waterson, M., Alachkar, B., Hendre, A., Lewis, P., Bartolini, M., and Dewdney, P. (Jan. 2022). "Highlights of the Square Kilometre Array Low Frequency (SKA-LOW) Telescope". In: *Journal of Astronomical Telescopes, Instruments, and Systems* 8, 011024, p. 011024. DOI: 10.1117/1.JATIS.8.1.011024.

Lacasse, R. J. (May 1998). "Green Bank Telescope active surface system". In: *Telescope Control Systems III*. Ed. by H. Lewis. Vol. 3351. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pp. 310–319. DOI: 10.1117/12.308802.

Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., and Artigau, É. (May 2007). "A New Algorithm for Point-Spread Function Subtraction in High-Contrast Imaging: A Demonstration with Angular Differential Imaging". In: *The Astrophysical Journal* 660.1, pp. 770–780. DOI: 10.1086/513180.

Leike, R. H., Glatzle, M., and Enßlin, T. A. (July 2020). "Resolving nearby dust clouds". In: *Astronomy & Astrophysics* 639, A138, A138. DOI: 10.1051/0004-6361/202038169.

Leike, R. H. and Enßlin, T. A. (2017). "Optimal Belief Approximation". In: *Entropy* 19.8. ISSN: 1099-4300. DOI: 10.3390/e19080402.

Li Causi, G., Stangalini, M., Antoniucci, S., Pedichini, F., Mattioli, M., and Testa, V. (Nov. 2017). "SFADI: The Speckle-free Angular Differential Imaging Method". In: *The Astrophysical Journal* 849.2, 85, p. 85. DOI: 10.3847/1538-4357/aa8e98.

Liaudat, T. I., Mars, M., Price, M. A., Pereyra, M., Betcke, M. M., and McEwen, J. D. (Nov. 2023). "Scalable Bayesian uncertainty quantification with data-driven priors for radio interferometric imaging". In: *arXiv e-prints*, arXiv:2312.00125, arXiv:2312.00125. DOI: 10.48550/arXiv.2312.00125.

MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press.

Markus, K.-P. (Jan. 2010). "Overall science goals and top level AO requirements for the E-ELT". In: *Adaptative Optics for Extremely Large Telescopes*, 01001, p. 01001. DOI: 10.1051/ao4elt/201001001.

Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., and Nadeau, D. (Apr. 2006). "Angular Differential Imaging: A Powerful High-Contrast Imaging Technique". In: *The Astrophysical Journal* 641.1, pp. 556–564. DOI: 10.1086/500401.

Matérn, B. (1986). *Spatial Variation*. Springer New York. DOI: 10.1007/978-1-4615-7892-5.

Mattioli, M., Pedichini, F., Antoniucci, S., Causi, G. L., Piazzesi, R., Stangalini, M., and Testa, V. (2018). "SHARK-VIS the LBT high contrast imager at visible wavelengths". In: *Ground-based and Airborne Instrumentation for Astronomy VII*. Ed. by C. J. Evans, L. Simard, and H. Takami. Vol. 10702. International Society for Optics and Photonics. SPIE, 107024F. DOI: 10.1117/12.2312591.

Mattioli, M., Pedichini, F., Antoniucci, S., Li Causi, G., Piazzesi, R., Stangalini, M., Testa, V., Vaz, A., Pinna, E., Puglisi, A., et al. (Jan. 2019). "First Direct Imaging Detection of the Secondary Component of α Andromedae with the LBT/SHARK-VIS Pathfinder Experiment". In: *Research Notes of the American Astronomical Society* 3.1, 20, p. 20. DOI: 10.3847/2515-5172/ab0111.

Molina, C., Boudriki Semlali, B.-E., González-Casado, G., Park, H., and Camps, A. (2023). "The 2021 La Palma volcanic eruption and its impact on ionospheric scintillation as measured from GNSS reference stations, GNSS-R and GNSS-RO". In: *Natural Hazards and Earth System Sciences* 23.12, pp. 3671–3684. DOI: 10.5194/nhess-23-3671-2023.

Moses, W. and Churavy, V. (2020). "Instead of Rewriting Foreign Code for Machine Learning, Automatically Synthesize Fast Gradients". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin. Vol. 33. Curran Associates, Inc., pp. 12472–12485. URL: https://proceedings.neurips.cc/paper/2020/file/9332c513ef44b682e9347822c2e457ac-Paper.pdf.

Moses, W. S., Churavy, V., Paehler, L., Hückelheim, J., Narayanan, S. H. K., Schanen, M., and Doerfert, J. (2021). "Reverse-Mode Automatic Differentiation and Optimization of GPU Kernels via Enzyme". In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. SC '21. St. Louis, Missouri: Association for Computing Machinery. ISBN: 9781450384421. DOI: 10.1145/3458817.3476165.

Moses, W. S., Narayanan, S. H. K., Paehler, L., Churavy, V., Schanen, M., Hückelheim, J., Doerfert, J., and Hovland, P. (2022). "Scalable Automatic Differentiation of Multiple Parallel Paradigms through Compiler Augmentation". In: *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*. SC '22. Dallas, Texas: IEEE Press. ISBN: 9784665454445. DOI: 10.1109/SC41404.2022.00065.

Moses, W. S. and Zinenko, O. (2024). *Enzyme-JAX*. Version 0.0.6. URL: https://github.com/EnzymeAD/Enzyme-JAX.

Offringa, A. R., McKinley, B., Hurley-Walker, N., Briggs, F. H., Wayth, R. B., Kaplan, D. L., Bell, M. E., Feng, L., Neben, A. R., Hughes, J. D., et al. (Aug. 2014). "wsclean: an implementation of a fast, generic wide-field imager for radio astronomy". In: *Monthly Notices of the Royal Astronomical Society* 444.1, pp. 606–619. ISSN: 0035-8711. DOI: 10.1093/mnras/stu1368.

Offringa, A. R. and Smirnov, O. (Oct. 2017). "An optimized algorithm for multiscale wide-band deconvolution of radio astronomical images". In: *Monthly Notices of the Royal Astronomical Society* 471.1, pp. 301–316. DOI: 10.1093/mnras/stx1547.

Pairet, B., Cantalloube, F., and Jacques, L. (Mar. 2021). "MAYONNAISE: a morphological components analysis pipeline for circumstellar discs and exoplanets imaging in the near-infrared". In: *Monthly Notices of the Royal Astronomical Society* 503.3, pp. 3724–3742. ISSN: 0035-8711. DOI: 10.1093/mnras/stab607.

Paxman, R. G., Carrara, D. A., Miller, J. J., Gleichman, K. W., Rucci, M. A., and Karch, B. K. (2019). "Estimation ambiguities encountered when imaging through turbulence". In: *Unconventional and Indirect Imaging, Image Reconstruction, and Wavefront Sensing 2019*. Ed. by J. J. Dolne, M. F. Spencer, and M. E. Testorf. Vol. 11135. International Society for Optics and Photonics. SPIE, 111350G. DOI: 10.1117/12.2533979.

Pedichini, F., Stangalini, M., Ambrosino, F., Puglisi, A., Pinna, E., Bailey, V., Carbonaro, L., Centrone, M., Christou, J., Esposito, S., et al. (Aug. 2017). "High Contrast Imaging in the Visible: First Experimental Results at the Large Binocular Telescope". In: *The Astronomical Journal* 154.2, 74, p. 74. DOI: 10.3847/1538-3881/aa7ff3.

Perley, R. A. and Butler, B. J. (Jan. 2013). "AN ACCURATE FLUX DENSITY SCALE FROM 1 TO 50 GHz". In: *The Astrophysical Journal Supplement Series* 204.2, p. 19. DOI: 10.1088/0067-0049/204/2/19.

Perley, R. A., Chandler, C. J., Butler, B. J., and Wrobel, J. M. (Sept. 2011). "The Expanded Very Large Array: A New Telescope for New Science". In: *The Astrophysical Journal Letters* 739.1, L1, p. L1. DOI: 10.1088/2041-8205/739/1/L1.

Perley, R. A. (Jan. 1999). "Imaging with Non-Coplanar Arrays". In: *Synthesis Imaging in Radio Astronomy II*. Vol. 180. Astronomical Society of the Pacific Conference Series, p. 383. URL: https://www.aspbooks.org/a/volumes/article_details/?paper_id=17953.

Perryman, M. (2018). *The Exoplanet Handbook*. Cambridge University Press. DOI: 10.1017/9781108304160.

Pratley, L., Johnston-Hollitt, M., and McEwen, J. D. (Apr. 2019). "A Fast and Exact w-stacking and w-projection Hybrid Algorithm for Wide-field Interferometric Imaging". In: *The Astrophysical Journal* 874.2, p. 174. DOI: 10.3847/1538-4357/ab0a05.

Ramatsoku, M., Murgia, M., Vacca, V., Serra, P., Makhathini, S., Govoni, F., Smirnov, O., Andati, L. A. L., de Blok, E., Józsa, G. I. G., et al. (Apr. 2020). "Collimated synchrotron threads linking the radio lobes of ESO 137-006". In: *Astronomy & Astrophysics* 636, L1, p. L1. DOI: 10.1051/0004-6361/202037800.

Rau, U. and Cornwell, T. J. (Aug. 2011). "A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry". In: *Astronomy & Astrophysics* 532, A71, A71. DOI: 10.1051/0004-6361/201117104.

Rau, U., Bhatnagar, S., Voronkov, M. A., and Cornwell, T. J. (Aug. 2009). "Advances in Calibration and Imaging Techniques in Radio Interferometry". In: *IEEE Proceedings* 97.8, pp. 1472–1481. DOI: 10.1109/JPROC.2009.2014853.

Reinecke, M. (2024). *DUCC: Distinctly Useful Code Collection*. Version 0.33.0. URL: https://gitlab.mpcdf.mpg.de/mtr/ducc.

Ren, B., Pueyo, L., Chen, C., Choquet, É., Debes, J. H., Duchêne, G., Ménard, F., and Perrin, M. D. (Mar. 2020). "Using Data Imputation for Signal Separation in High-contrast Imaging". In: *The Astrophysical Journal* 892.2, p. 74. DOI: 10.3847/1538-4357/ab7024.

Renzini, A. (Sept. 2006). "Stellar Population Diagnostics of Elliptical Galaxy Formation". In: *Annual Review of Astronomy and Astrophysics* 44.1, pp. 141–192. DOI: 10.1146/annurev.astro.44.051905.092450.

Repetti, A., Birdi, J., Dabbech, A., and Wiaux, Y. (May 2017). "Non-convex optimization for self-calibration of direction-dependent effects in radio interferometric imaging". In: *Monthly Notices of the Royal Astronomical Society* 470.4, pp. 3981–4006. ISSN: 0035-8711. DOI: 10.1093/mnras/stx1267.

Repetti, A., Pereyra, M., and Wiaux, Y. (2019). "Scalable Bayesian Uncertainty Quantification in Imaging Inverse Problems via Convex Optimization". In: *SIAM Journal on Imaging Sciences* 12.1, pp. 87–118. DOI: 10.1137/18M1173629.

Roberts, D. H., Rogers, A. E. E., Allen, B. R., Bennett, C. L., Burke, B. F., Greenfield, P. E., Lawrence, C. R., and Clark, T. A. (Aug. 1982). "Radio interferometric detection of a traveling ionospheric disturbance excited by the explosion of Mount St. Helens". In: *Journal of Geophysical Research* 87.A8, pp. 6302–6306. DOI: 10.1029/JA087iA08p06302.

Roth, J., Li Causi, G., Testa, V., Arras, P., and Ensslin, T. A. (Mar. 2023a). "Fast-cadence High-contrast Imaging with Information Field Theory". In: *The Astronomical Journal* 165.3, 86, p. 86. DOI: 10.3847/1538-3881/acabc1.

Roth, J., Arras, P., Reinecke, M., Perley, R. A., Westermann, R., and Enßlin, T. A. (Oct. 2023b). "Bayesian radio interferometric imaging with direction-dependent calibration". In: *Astronomy & Astrophysics* 678, A177, A177. DOI: 10.1051/0004-6361/202346851.

Roth, J., Frank, P., Bester, H. L., Smirnov, O. M., Westermann, R., and Enßlin, T. A. (June 2024a). "fast-resolve: Fast Bayesian Radio Interferometric Imaging". In: *arXiv e-prints*, arXiv:2406.09144, arXiv:2406.09144. DOI: 10.48550/arXiv.2406.09144.

Roth, J., Reinecke, M., and Edenhofer, G. (June 2024b). "JAXbind: Bind any function to JAX". In: *Journal of Open Source Software* 9.98, p. 6532. DOI: 10.21105/joss.06532.

Rüeger, J. M. (2002). *Refractive Indices of Light, Infrared, and Radio Waves in the Atmosphere*. Unisurv Report S-68, School of Surveying and Spatial Information Systems, University of New South Wales, Sydney, Australia.

Rüstig, J., Guardiani, M., Roth, J., Frank, P., and Enßlin, T. (Feb. 2024). "Introducing LensCharm. A charming Bayesian strong lensing reconstruction framework". In: *Astronomy & Astrophysics* 682, A146, A146. DOI: 10.1051/0004-6361/202348256.

Saha, S. K. (2007). *Diffraction-limited imaging with large and moderate telescopes*. World Scientific. DOI: 10.1142/6492.

Sbokolodi, M. L. L., Perley, R., Eilek, J., Carilli, C., Smirnov, O., Laing, R., Greisen, E. W., and Wise, M. (Oct. 2020). "A Wideband Polarization Study of Cygnus A with the Jansky Very Large Array. I. The Observations and Data". In: *The Astrophysical Journal* 903.1, p. 36. DOI: 10.3847/1538-4357/abb80e.

Schmidt, K., Geyer, F., Fröse, S., Blomenkamp, P. .-., Brüggen, M., de Gasperin, F., Elsässer, D., and Rhode, W. (Aug. 2022). "Deep learning-based imaging in radio interferometry". In: *Astronomy & Astrophysics* 664, A134, A134. DOI: 10.1051/0004-6361/202142113.

Schwab, F. R. and Cotton, W. D. (May 1983). "Global fringe search techniques for VLBI". In: *Astronomical Journal* 88, pp. 688–694. DOI: 10.1086/113360.

Sebokolodi, M. L. L., Perley, R., Eilek, J., Carilli, C., Smirnov, O., Laing, R., Greisen, E. W., and Wise, M. (Nov. 2020). "A Wideband Polarization Study of Cygnus A with the Jansky Very Large Array. I. The Observations and Data". In: *The Astrophysical Journal* 903.1, 36, p. 36. DOI: 10.3847/1538-4357/abb80e.

Selig, M., Bell, M. R., Junklewitz, H., Oppermann, N., Reinecke, M., Greiner, M., Pachajoa, C., and Enßlin, T. A. (June 2013a). "NIFTY - Numerical Information Field Theory. A versatile PYTHON library for signal inference". In: *Astronomy & Astrophysics* 554, A26, A26. DOI: 10.1051/0004-6361/201321236.

Selig, M., Bell, M. R., Junklewitz, H., Oppermann, N., Reinecke, M., Greiner, M., Pachajoa, C., and Ensslin, T. A. (June 2013b). "NIFTY - Numerical Information Field Theory. A versatile PYTHON library for signal inference". In: *Astronomy & Astrophysics* 554, A26, A26. DOI: 10.1051/0004-6361/201321236.

Selig, M., Vacca, V., Oppermann, N., and Enßlin, T. A. (Sept. 2015). "The denoised, deconvolved, and decomposed Fermi γ-ray sky. An application of the D^{3}PO algorithm". In: *Astronomy & Astrophysics* 581, A126, A126. DOI: 10.1051/0004-6361/201425172.

Smart, W. M. (1977). *Textbook on Spherical Astronomy*. Ed. by R. M. Green. 6th ed. Cambridge University Press. DOI: 10.1017/CBO9781139167574.

Smirnov, O. M. (2011a). "Revisiting the radio interferometer measurement equation - I. A full-sky Jones formalism". In: *Astronomy & Astrophysics* 527, A106. DOI: 10.1051/0004-6361/201016082.

Smirnov, O. M. (Mar. 2011b). "Revisiting the radio interferometer measurement equation. II. Calibration and direction-dependent effects". In: *Astronomy & Astrophysics* 527, A107, A107. DOI: 10.1051/0004-6361/201116434.

Smirnov, O. M. (Mar. 2011c). "Revisiting the radio interferometer measurement equation. III. Addressing direction-dependent effects in 21 cm WSRT observations of 3C 147". In: *Astronomy & Astrophysics* 527, A108, A108. DOI: 10.1051/0004-6361/201116435.

Smirnov, O. M. and Tasse, C. (Apr. 2015). "Radio interferometric gain calibration as a complex optimization problem". In: *Monthly Notices of the Royal Astronomical Society* 449.3, pp. 2668–2684. ISSN: 0035-8711. DOI: 10.1093/mnras/stv418.

Soummer, R., Pueyo, L., and Larkin, J. (Aug. 2012). "Detection and Characterization of Exoplanets and Disks Using Projections on Karhunen-Loève Eigenimages". In: *The Astrophysical Journal Letters* 755.2, L28, p. L28. DOI: 10.1088/2041-8205/755/2/L28.

Stangalini, M., Pedichini, F., Pinna, E., Christou, J., Hill, J., Puglisi, A., Bailey, V., Centrone, M., Del Moro, D., Esposito, S., et al. (Apr. 2017). "Speckle statistics in adaptive optics

images at visible wavelengths". In: *Journal of Astronomical Telescopes, Instruments, and Systems* 3, 025001, p. 025001. DOI: 10.1117/1.JATIS.3.2.025001.

Steininger, T., Dixit, J., Frank, P., Greiner, M., Hutschenreuter, S., Knollm"uller, J., Leike, R., Porqueres, N., Pumpe, D., Reinecke, M., et al. (Aug. 2017). "NIFTy 3 - Numerical Information Field Theory - A Python framework for multicomponent signal inference on HPC clusters". In: *ArXiv e-prints*. arXiv: 1708.01073 [astro-ph.IM].

Sutter, P. M., Wandelt, B. D., McEwen, J. D., Bunn, E. F., Karakci, A., Korotkov, A., Timbie, P., Tucker, G. S., and Zhang, L. (Feb. 2014). "Probabilistic image reconstruction for radio interferometers". In: *Monthly Notices of the Royal Astronomical Society* 438.1, pp. 768–778. DOI: 10.1093/mnras/stt2244.

Sutton, E. C. and Wandelt, B. D. (Feb. 2006). "Optimal Image Reconstruction in Radio Interferometry". In: *The Astrophysical Journal Supplement Series* 162.2, pp. 401–416. DOI: 10.1086/498571.

Swart, G. P., Dewdney, P. E., and Cremonini, A. (Jan. 2022). "Highlights of the SKA1-Mid telescope architecture". In: *Journal of Astronomical Telescopes, Instruments, and Systems* 8, 011021, p. 011021. DOI: 10.1117/1.JATIS.8.1.011021.

Tasse, C., Hugo, B., Mirmont, M., Smirnov, O., Atemkeng, M., Bester, L., Hardcastle, M. J., Lakhoo, R., Perkins, S., and Shimwell, T. (2018). "Faceting for direction-dependent spectral deconvolution". In: *Astronomy & Astrophysics* 611, A87. DOI: 10.1051/0004-6361/201731474.

Tatarski, V. I. (1971). *The effects of the turbulent atmosphere on wave propagation*. Israel Program for Scientific Translations.

Tatarski, V. I., Silverman, R. A., and Chako, N. (Jan. 1961). "Wave Propagation in a Turbulent Medium". In: *Physics Today* 14.12, p. 46. DOI: 10.1063/1.3057286.

Terris, M., Dabbech, A., Tang, C., and Wiaux, Y. (Jan. 2023a). "Image reconstruction algorithms in radio interferometry: From handcrafted to learned regularization denoisers". In: *Monthly Notices of the Royal Astronomical Society* 518.1, pp. 604–622. DOI: 10.1093/mnras/stac2672.

Terris, M., Tang, C., Jackson, A., and Wiaux, Y. (Dec. 2023b). "Plug-and-play imaging with model uncertainty quantification in radio astronomy". In: *arXiv e-prints*, arXiv:2312.07137, arXiv:2312.07137. DOI: 10.48550/arXiv.2312.07137.

Thompson, A. R., Moran, J. M., and Swenson Jr., G. W. (2017). *Interferometry and Synthesis in Radio Astronomy*. Springer Open. DOI: 10.1007/978-3-319-44431-4.

Thouvenin, P.-A., Abdulaziz, A., Dabbech, A., Repetti, A., and Wiaux, Y. (May 2023). "Parallel faceted imaging in radio interferometry via proximal splitting (Faceted HyperSARA): I. Algorithm and simulations". In: *Monthly Notices of the Royal Astronomical Society* 521.1, pp. 1–19. DOI: 10.1093/mnras/stac1521.

Thouvenin, P.-A., Repetti, A., Dabbech, A., and Wiaux, Y. (2018). "Time-Regularized Blind Deconvolution Approach for Radio Interferometry". In: *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 475–479. DOI: 10.1109/SAM.2018.8448891.

Tiede, P. (Aug. 2022). "Comrade: Composable Modeling of Radio Emission". In: *The Journal of Open Source Software* 7.76, 4457, p. 4457. DOI: 10.21105/joss.04457.

Tol, S. van der, Veenboer, B., and Offringa, A. R. (2018). "Image Domain Gridding: a fast method for convolutional resampling of visibilities". In: *Astronomy & Astrophysics* 616, A27. DOI: 10.1051/0004-6361/201832858.

van den Bergh, S. (Jan. 1975). "Stellar populations in galaxies." In: *Annual Review of Astronomy and Astrophysics* 13, pp. 217–255. DOI: 10.1146/annurev.aa.13.090175.001245.

Walter, A. B., Bockstiegel, C., Brandt, T. D., and Mazin, B. A. (Nov. 2019). "Stochastic Speckle Discrimination with Time-tagged Photon Lists: Digging below the Speckle Noise Floor". In: *Publications of the Astronomical Society of the Pacific* 131.1005, p. 114506. DOI: 10.1088/1538-3873/ab389a.

Weeren, R. J. van, Williams, W. L., Hardcastle, M. J., Shimwell, T. W., Rafferty, D. A., Sabater, J., Heald, G., Sridhar, S. S., Dijkema, T. J., Brunetti, G., et al. (Mar. 2016). "LOFAR FACET CALIBRATION". In: *The Astrophysical Journal Supplement Series* 223.1, p. 2. DOI: 10.3847/0067-0049/223/1/2.

Westerkamp, M., Roth, J., Frank, P., Handley, W., and Enßlin, T. (2023). "Inferring Evidence from Nested Sampling Data via Information Field Theory". In: *Physical Sciences Forum* 9.1. ISSN: 2673-9984. DOI: 10.3390/psf2023009019.

Westerkamp, M., Roth, J., Frank, P., Handley, W., and Enßlin, T. (Aug. 2024). "Towards a Field Based Bayesian Evidence Inference from Nested Sampling Data". In: *arXiv e-prints*, arXiv:2408.09889, arXiv:2408.09889. DOI: 10.48550/arXiv.2408.09889.

Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M. M., and Vandergheynst, P. (May 2009). "Compressed sensing imaging techniques for radio interferometry". In: *Monthly Notices of the Royal Astronomical Society* 395.3, pp. 1733–1742. DOI: 10.1111/j.1365-2966.2009.14665.x.

Wikimedia Commons (2023). *Atmospheric electromagnetic opacity*. URL: https://commons.wikimedia.org/wiki/File:Atmospheric_electromagnetic_opacity.svg.

Wilson, T. L., Rohlfs, K., and Hüttemeister, S. (2009). *Tools of Radio Astronomy*. Springer. DOI: 10.1007/978-3-540-85122-6.

Ye, H., Gull, S. F., Tan, S. M., and Nikolic, B. (Dec. 2021). "High accuracy wide-field imaging method in radio interferometry". In: *Monthly Notices of the Royal Astronomical Society* 510.3, pp. 4110–4125. ISSN: 0035-8711. DOI: 10.1093/mnras/stab3548.

Acknowledgements

Many people have been directly and indirectly involved in this thesis and I am thankful to all of them. I am particularly grateful to my supervisor Torsten Enßlin for all his support throughout the course of my PhD. I would like to thank you for all the time you spent discussing new ideas and giving feedback on all sorts of manuscripts, while giving me the freedom to choose what to work on. I really enjoyed being part of your group. I would also like to thank Rüdiger Westerman for handling all the formalities of the D-MeerKAT collaboration and for his openness to astrophysics.

Philipp Arras deserves a big thank you for introducing me to radio interferometry and for developing the `resolve` framework on which many of the algorithms in this thesis are based. I would also like to thank Philipp Frank for insightful discussions on various aspects of Bayesian inference. It was a really great time working with you. Special thanks go to Martin Reinecke, who helped me immensely with his C++ skills. I would also like to thank the whole group at MPA and especially the other PhD students Vincent Eberle, Gordian Edenhofer, Matteo Guardiani, Johannes Harth-Kitzerow, Vishal Johnson, Viktoria Kainz, Margret Westerkamp and Hanieh Zandinejad for the productive discussions and the friendly atmosphere. You all were truly great colleagues. A special thanks goes to Gordian Edenhofer, my office mate, for our long scientific discussions and most of all for being a really good friend.

Besides my colleagues at MPA, I am also grateful to my international collaborators Landman Bester, Oleg Smirnov and Rick Perley. I would especially like to thank Landman Bester for our regular discussions and for welcoming me to Cape Town. I would also like to thank Rick Perley for giving me an extensive tour of the VLA.

Furthermore would I like to thank my close friends Johannes Bürger, Gordian Edenhofer, Haochun Ma, Lukas Nakamura, and Paul Ockenfuß. You were great study mates since the first semester and it has been a fantastic time with you in Munich.

I would also like to thank my family, especially my parents. Your support laid the foundation for this whole study. Finally, I would like to say a special thank you to Tabea Gleiter for always standing by my side and supporting me.

I acknowledge the use of DeepL and Grammarly in the preparation of this thesis for checking spelling and grammar as well as improving wording.