Null stream methods for resolvable Pulsar Timing Array signals

Janna Mo Goldstein

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Institute for Gravitational Wave Astronomy
School of Physics and Astronomy
College of Engineering and Physical Sciences
University of Birmingham
21st May 2020

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive
e-theses repository

This unpublished thesis/dissertation is under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) licence.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

Unless otherwise stated, any material in this thesis/dissertation that is cited to a third party source is not included in the terms of this licence. Please refer to the original source(s) for licencing conditions of any quotes, images or other material cited to a third party.

Abstract

Pulsar Timing Arrays (PTAs) are sensitive to nanohertz gravitational waves (GWs), which affect the pulse travel times between pulsars and the Earth. The most promising sources are super massive black hole binaries, some of which are expected to be detected as resolvable signals in the not-too-distant future.

In this work, we develop a Bayesian method for localising these signals using *null streams*. Null streams are combinations of data from different detectors in which the GW signal from a particular direction is cancelled. Applying our method to idealised, simulated data, we assess the localisation capabilities of current PTAs and show the merit of combining data from different collaborations.

For future multi-messenger astronomy with low frequency GWs and electromagnetic observations, it is crucial to identify the galaxy that hosts a detected binary. We develop a method that combines localisation information, a posterior on the signal's amplitude, and galaxy properties to significantly reduce the number of candidate hosts in a galaxy catalogue.

We extend the null-stream method to data with arbitrary sampling times. Exploiting the null-stream property that separates dependencies on the source location and on the GW-model parameters, we construct a likelihood that is independent of the GW model.

Declaration

Chapter 1 and 2 are introductory chapters written by myself. The figures are original work except for Fig. 1.3 which has been made using `http://gwplotter.com/` [1], and Fig. 2.1 which is an artwork by Peter Saville.

Two chapters of this thesis are based on published papers. The text and figures of Chapter 3 are taken from [2], except for Fig. 3.1 which is taken from [3] (I edited symbols and labels in this figure to make it consistent with those in the text). The text and figures of Chapter 4 are taken from [4]. In both chapters, the text and figures have been edited for readability, consistent notation and to make them fit in with the rest of the chapters.

I am first author of both papers and have written the text and carried out the majority of the work myself, with support from my co-authors. For [4], Miguel Holgado investigated the use of galaxy catalogues. The mathematical method was developed in collaboration with Alberto Sesana and John Veitch. The source selection (described in Sec. 4.3.1) was done primarily by Alberto Sesana, who also contributed to the text of the corresponding section as well as the introduction of [4].

Chapter 5 presents unpublished work that I collaborated on with Christopher Moore and Elinore Roebber. It will be submitted for publication in the near future. The application of the Fourier transform to irregularly sampled data (as described in Sec. 5.1.1) has been developed and implemented into PYTHON code in collaboration with Elinore. The text and figures of this chapter are all created by me.

Three papers on which I am co-author have been published during my PhD course [5, 6, 7]. They are not directly related to the rest of this thesis and have not been included in the text.

Throughout the text, the use of "we" is a stylistic choice and does not reflect the level of contribution from the author or collaborators, which has been described above.

Acknowledgements

Even though it was out of necessity rather than choice, I consider myself lucky to have worked so well with three different primary supervisors. So I thank you for your support and relentless positivity: John (and for all the excellent food in Glasgow), Alberto (and for the late evenings with pasta and Amaro), and Chris (and for keeping me sane over zoom chats). I also thank Alberto Vecchio for having his door open for me during all of the past years.

During my PhD I have experienced a growing realisation of how messed up the world we live in is, which has certainly made keeping up the motivation for my academic work more challenging. So I really mean it when I say thank you to all the people around me for your distractions and support. I would especially like to thank Elinore for making work much more fun than it otherwise would have been, Conor for being the drive behind the much loved bouldering sessions and pub club, and Aaron for the open-hearted and honest conversations.

Thank you also to all the friends I have made outside of the university, including everyone from XR Birmingham, for enriching my life and being there for me when needed. A special thank you to Tomas for embracing my weirdness and to Ross, for just embracing me.

Lastly, my parents have been there to support me since long before the start of my PhD. Dus, mams, paps, bedankt dat jullie altijd trots waren op jullie dochter, ook als ik dat zelf niet altijd was.

Contents

List of Figures

List of Tables

Glossary

BBH binary black hole. Two black holes orbiting each other.

BH black hole.

BNS binary neutron star. Two neutron stars orbiting each other.

DFT discrete Fourier transform. A Fourier transform that has been adapted for a discrete quantity.

DM dispersion measure. The difference in delay between high and low frequency photons, due to free electrons in the insterstellar medium. Because this delay scales inversely with the photon frequency squared, it can be measured and used as a proxy for the free electron density between the pulsar and Earth.

DR data release.

EFT extended Fourier transform. The discrete Fourier transform from Eq. 5.8 that has been extended to unevenly sampled data.

EM electromagnetic.

EMRI extreme mass ratio insprial. The inspiral of a stellar mass black hole around a (super) massive black hole.

EPTA European Pulsar Timing Array. A PTA collaboration that uses five radio telescopes at different locations in Europe.

FAST Five-hundred-meter Aperture Spherical radio Telescope. FAST is located in China and has been discovering new pulsars since 2017.

FD frequency domain.

FFT Fast Fourier Transform. An algorith developed by [13] for efficient computation of the discrete Fourier transform of evenly sampled data.

FT Fourier transform.

GR The theory of general relativity, developed by Albert Einstein in the early 20th century.

GW gravitational wave.

IPTA International Pulsar Timing Array. An umbrella consortium of ETPA, NANOGrav and PPTA.

ISM interstellar medium. Any matter not part of a stellar system.

LEAP The Large European Array for Pulsars. Simultaneous observations from the five EPTA telescopes are combined to create one TOA.

LIGO Laser Interferometer Gravitational-Wave Observatory. A gravitational wave observatory with two interferometers, one located in Hanford and one in Livingston, USA.

LISA Laser Inteferometer Space Antenna. A space-based gravitational wave observatory that uses laser interferometry between three spacecraft. Planned to launch in the 2030's.

MSP millisecond pulsar.

NANOGrav North American Nanohertz Observatory for Gravitational Waves. A PTA collaboration that uses two North American radio telescopes (Green Bank and Arecibo).

ngVLA next-generation Very Large Array. A planned upgrade of the Very Large Array (VLA) radio telescope in New Mexico, USA.

null stream Null streams are data combinations in which a signal is cancelled out.

PPTA Parkes Pulsar Timing Array. A PTA collaboration that uses the Parkes radio telescope in Australia.

PTA Pulsar Timing Array. Precision timing of an array of pulsars in the Galaxy.

pulsar timing pulsar timing. Long term observations of a radio pulsar to construct a timing model that can predict future arrival times of pulses.

rms root mean square. The rms of a set of values is computed by taking the mean of the squares of the values, then taking the square root of the result.

S/N signal-to-noise ratio. In equations, the symbol ρ is used.

SB Stochastic background of gravitational wave radiation from unresolved SMBH binaries in the universe.

SKA Square Kilometre Array. A radio telescope planned to be constructed in Australia and South Africa, with a total collecting area of about one square kilometre.

SMBH super massive black hole. A black hole of order $10^5 - 10^{10}\mathrm{M}_{\odot}$, residing in the center of a galaxy.

SSB solar system barycenter. The midpoint of all mass in the solar system, computed from the measured positions and masses of the sun and planets.

TD time domain.

TM timing model. The modelled arrival times of pulses from a particular pulsar, which can also be used to predict future times of arrival.

TOA Time of arrival. The measured or predicted arrival time of a single pulse of radiation of a pulsar.

Virgo The Virgo gravitational wave observatory. A laser inteferometer located near Pisa, Italy, often used in conjunction with the two LIGO interferometers.

Chapter 1

Introduction to Gravitational Waves

> *"Don't believe everything you read on the internet."*
>
> Albert Einstein

Over a century ago, the existence of gravitational waves (GWs) travelling at the speed of light was predicted by Henri Poincaré [14]. This concept was later formalised by Albert Einstein, on the basis of his theory of general relativity (GR) [15, 16]. Only five years ago, the LIGO detectors made the first direct observation of a GW [17]. It was emitted by the coalescence and subsequent merger of two black holes, each a few tens times the mass of the sun. Now, we have a catalogue of GW observation from the LIGO and Virgo detectors, including GWs sourced by a binary neutron star coalescence.

But these observations from ground-based interferometers will hopefully be just the beginning of GW astronomy. There are plans for a larger scale interferometer, LISA, to be launched into space in the next decade [18]. Because LISA is much larger then the ground-based detectors (and because it is undisturbed by noise from the Earth), it will be sensitive to GWs at lower frequencies. This gives access to new sources such as massive black holes binaries (with component masses around $10^3 - 10^7 \mathrm{M}_{\odot}$) and binary white dwarfs in the Galaxy [18].

Another potential method for GW detection is a Pulsar Timing Array (PTA), which will be the topic of this thesis. A PTA relies on detecting the imprint of GWs on the time signature of millisecond pulsars, observed with radio telescopes [19]. The sensitivity of PTAs lies at much lower frequencies then LIGO+Virgo or LISA, giving access to entirely different sources. It's expected that PTAs will be able to observe supermassive binary black holes, where the black hole masses are of order 10^7 times the mass of the sun or more [e.g. 20].

Before starting on the topic of PTAs as a GW detector, we need an understanding of GWs themselves, which is the aim of this chapter. This begins with the basics of GR – the language that Einstein used to make his prediction a hundred years ago – in Sec. 1.1. Then, a mathematical description of GWs is discussed in Sec. 1.2 before we move onto possible sources in Sec. 1.3. PTAs are discussed in detail in chapter 2, including a section (2.3) on data analysis specifically. This lays the foundation for the main work of this thesis, which concerns a data analysis method that uses so-called *null streams*. Sec. 2.3.3 is devoted to explaining what null streams are.

The following three chapters consist of two published and one unpublished work on PTA data analysis with null streams. In the first, chapter 3, we being developing a method to localise resolvable PTA signals [2], assuming such signals will be detected in the future (see Sec. 2.2). The second, Chapter 4, explores how this localisation information could help identify which galaxies are likely to host the detected source. In these works, several simplifying assumptions are made about PTA data, allowing to explore the workings of the null stream method in principle. In the third, Chapter 5, the (untrue) assumption of evenly sampled data is discarded. This causes some practical problems for null stream analysis, and it is investigated whether Fourier transformations of unevenly sampled data can provide a solution. We also expand the method from Chapter 3 to localise sources without using a model for the GW itself.

A note on notation

In the following, many of the usual conventions for index notation are used, and some notation choices are made. These are:

- $\vec{x}$ is a 3-dimensional (spatial) vector. In index notation, this is written as x_a (or x^a). Any Latin index (here, a) goes from 1 to 3 and labels the spatial coordinates.
- $\boldsymbol{x}$ is a 3+1-dimensional vector, or four-vector. This is written as x^α in index notation. Any Greek index (here, α) goes from 0 to 4. The additional value 0 labels the temporal coordinate.
- x_α is a 3+1-dimensional covariant vector (covector) in index notation.
- $x^{\alpha_1 \cdots \alpha_n}$ is a tensor of rank n. Specifically for matrices (rank-2 tensors), we write $\mathcal{X}$ when not using index notation.
- The Einstein summation convention is that any repeated indices (one upper, one lower index) are implicitly summed over. For example, $x^\alpha y_\alpha$ is shorthand for $\sum_\alpha (x^\alpha y_\alpha)$.

1.1 General relativity

1.1.1 The space-time metric

In general relativity (GR), as opposed to Newtonian physics, gravity is not a force. Instead, the way objects move is directed by the shape of the space and time around them. This introduces the concept of space-time: three spatial dimensions and one temporal dimension are combined into a 3+1-dimensional, all-surrounding expanse. An empty, unaffected space-time is flat, and a test particle would move through it at a constant speed - equivalent to the Newtonian description of no force being put on the particle. In the classical framework, we can then introduce an object with mass to exert a gravitational attraction on the particle. This accelerates the particle towards the massive object, altering its path. In GR, an object with mass changes the shape of the space-time; it introduces curvature. The test particle follows the curvature, such that the resulting path is the same as predicted by classical gravity[1].

To quantify the shape of space-time, a 3+1-dimensional tensor called the *metric* is used. It is usually written as $g_{\mu\nu}$, where the Greek indices go from 0-3. When using a Cartesian coordinate system, the indices label the coordinates $\mathrm{c}t$ (the speed of light × time), x, y and z (the three spatial directions). The metric tensor is always symmetric under transposition, i.e. when swapping the indices, $g_{\mu\nu} = g_{\nu\mu}$. In the special case of a flat space-time, coordinates can be chosen such that the metric is equal to the Minkowski metric $\eta_{\mu\nu}$:

$$g_{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.1}$$

On the contrary, a space-time that is *intrinsically curved* cannot be globally described with the Minkowski metric. It is hard to imagine curvature in three dimensions (or 3+1), but we can think of an example for a two-dimensional space embedded in our three-dimensional world. A flat space would be that of an ant stuck to one side of a piece of paper (even if the paper where bent, this would only create *extrinsic curvature* and the space would still be intrinsically flat). On the other hand, an ant stuck to the surface of a balloon would live in an intrinsically curved space. If the ants were scientists, they could measure this difference in geometry. For example, if both ants add up the angles inside a triangle, the first would get 180° exactly, but the second would get more than 180°. (The balloon surface has positive curvature; in negative curvature, the triangle corners would total to less than 180°.)

With the metric from Eq. 1.1, the infinitesimal path length $\mathrm{d}s$ of a particle moving through space-time is given by:

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -\mathrm{c}^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{1.2}$$

The value of $\mathrm{d}s^2$ can be positive, negative, or zero. In the last case, the path the particle travels on is called *null*. Another name is *lightlike*, because this characterises the path of a particle travelling at the speed of light. In a time $\mathrm{d}t$, it would traverse a physical distance $(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2)^{1/2} = \mathrm{c}\,\mathrm{d}t$, and so $-\mathrm{c}^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 = 0$. For a particle travelling at any subluminal speed $v < \mathrm{c}$, the traversed distance would be $(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2)^{1/2} = v\,\mathrm{d}t < \mathrm{c}\,\mathrm{d}t$, and so $\mathrm{d}s^2 < 0$. These paths are called *timelike* (the separation in time is greater than the separation in space).

---

[1] ...in the limit of weak gravitational fields and low velocities compared to the speed of light.

To make sense of the case where $\mathrm{d}s^2 > 0$, consider a space-time interval Δs, rather then an infinitesimal path length $\mathrm{d}s$. If we take A^μ and B^μ to be two events (points in space-time), the interval between them is $\Delta s = A^\mu - B^\mu = (\mathrm{c}\Delta t, \Delta x, \Delta y, \Delta z)$. As long as they are further separated in space than in time, $\Delta s^2 = -\mathrm{c}\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2$ is larger than zero. Such an interval is called *spacelike*. Respecting that no particle can travel faster then the speed of light, spacelike particle paths do not exist.

1.1.2 The geodesic equation

From the path length element in Eq. 1.2, we can find an equation of motion using the principle of least action: A free-falling particle will take the path along which the action is minimised. The total action S for a path is given by the integrated path length, i.e. $S = \int \mathrm{d}s$, which can be minimised by solving for $\mathrm{d}S = 0$. This derivation is not given here, but the result is the geodesic equation:

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu_{\alpha\beta} \frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\beta}{\mathrm{d}\lambda} = 0. \tag{1.3}$$

Here, x^μ are the coordinates of the particle we are describing. The parameter λ can be any affine parameter of the path. For a massive particle, it is often chosen to be the proper time τ - the time the particle itself would measure if it was wearing a watch - because it naturally always increases along a timelike path.

The symbol $\Gamma^\mu_{\alpha\beta}$ is the Christoffel symbol (also known as the affine connection), which depends on derivatives of the metric. The partial derivative with respect to the coordinate labelled by α is written as ∂_α (for example, the partial derivative with respect to z would be ∂_3). In this notation, the Christoffel symbol is:

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2} g^{\nu\mu} \left(\partial_\alpha g_{\nu\beta} + \partial_\beta g_{\nu\alpha} - \partial_\nu g_{\alpha\beta} \right). \tag{1.4}$$

In a flat space-time, the metric $g_{\mu\nu}$ is the Minkowski metric, and so its (partial) derivatives are all zero. This means the Christoffel symbols are zero and the equation of motion simplifies to:

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} = 0, \tag{1.5}$$

analogous to

$$\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} = 0 \tag{1.6}$$

for a free-falling particle in Newtonian mechanics. Actually, an equivalent definition of a geodesic is that a particle's four-velocity, $u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}t}$, does not change along the path (in technical terms, that it is parallel transported). That statement simplifies to Eq. 1.6 in flat space, where the path is straight. In curved space, the path still seems straight from the frame of the particle, but it follows the curvature of space-time. Take the example of the ant living on a balloon, but imagine it is riding around on a frictionless bicycle. From the ant's perspective, it keeps going straight without being accelerated[2], even though it follows the curved surface of the balloon.

1.1.3 The Einstein equations

So far, the geodesic equation has provided us with half a description of gravity: It tells us how a particle moves, given the space-time metric. The other half then, should explain how the shape of space-time comes to be in the first place. In the Newtonian picture, mass is the source of the gravitational force, so in GR too, mass must be the source of space-time curvature. This is where the Einstein equations come in. At the core of GR, this expression relates the space-time metric to the mass and energy that is present in it:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi \mathrm{G}}{\mathrm{c}^4} T_{\mu\nu}. \tag{1.7}$$

The constants here are c, the speed of light and G, the gravitational constant. Λ is the cosmological constant, which is needed to explain the accelerating expansion of the universe.

The Einstein tensor

There are two important tensors in these equations (plural because they hold for any choice of the indices $\mu\nu$). On the left-hand side, there is the Einstein tensor $G_{\mu\nu}$, related to the space-time curvature. On the right-hand side, there is the stress-energy tensor $T_{\mu\nu}$, which quantifies energy and momentum[3]. The first is defined as:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu}. \tag{1.8}$$

The components of this equation are built up from the metric, but there are some definitions needed to get there. Starting from the Christoffel symbol (a function of metric derivatives as in Eq. 1.4), we

---

[2] In our three-dimensional reality, some force (friction) is needed to keep the ant and bicycle stuck to the balloon.

[3] The equations 1.7 and 1.8 are given with lower indices, but the same equations with only upper indices are equivalent.

have the Riemann curvature tensor:

$$R^{\rho}{}_{\sigma\mu\nu} = \partial_{\mu}\Gamma^{\rho}{}_{\nu\sigma} - \partial_{\nu}\Gamma^{\rho}{}_{\mu\sigma} + \Gamma^{\rho}{}_{\mu\lambda}\Gamma^{\lambda}{}_{\nu\sigma} - \Gamma^{\rho}{}_{\nu\lambda}\Gamma^{\lambda}{}_{\mu\sigma}, \tag{1.9}$$

the Ricci tensor, which is a contraction of the Riemann tensor[4]:

$$R_{\alpha\beta} = R^{\rho}{}_{\alpha\rho\beta} = \partial_{\rho}\Gamma^{\rho}{}_{\beta\alpha} - \partial_{\beta}\Gamma^{\rho}{}_{\rho\alpha} + \Gamma^{\rho}{}_{\rho\lambda}\Gamma^{\lambda}{}_{\beta\alpha} - \Gamma^{\rho}{}_{\beta\lambda}\Gamma^{\lambda}{}_{\rho\alpha}, \tag{1.10}$$

and finally the Ricci scalar, which is the trace (with respect to the metric) of the Ricci tensor:

$$R = R^{\mu}_{\mu} = g^{\mu\nu}R_{\mu\nu}. \tag{1.11}$$

Remembering that in flat space-time the Christoffel symbols are always zero, it follows that the Riemann tensor, Ricci tensor and Ricci scalar are all zero as well. In curved space-time, the Riemann tensor always has non-zero components. It is however possible for the Ricci tensor to be zero in curved space-time, resulting in a zero Ricci scalar as well. This turns out to be the case in a vacuum (without a cosmological constant), where $T_{\mu\nu} = 0$.

The Stress-Energy Tensor

The stress-energy tensor $T^{\mu\nu}$, is the second important tensor in Eq. 1.7. It is a GR generalisation of mass in Newtonian physics, the source of gravity. It turns out that in GR, mass is not the only thing that curves space-time, but so do energy, momentum, pressure etc. In general, $T^{\mu\nu}$ is derived from the action associated with a particular state of matter. Here, we provide no such derivation; instead we follow a physical motivation for the stress-energy tensor of point particles.

A very simple model for a universe filled with dust is to describe it as a collection of infinitesimally small point particles that do not interact, extending out in every direction. In their rest frame, the particles can be described with four-velocity $u^{\mu} = (\mathrm{c}, 0, 0, 0)$, and a collective density ρ. A suitable $T^{\mu\nu}$ has to be a tensor of rank 2, which we can create from these quantities by taking the outer product of u^{μ} with itself. This yields $u^{\mu}u^{\nu} = \mathrm{diag}(\mathrm{c}^2, 0, 0, 0)$, a 4×4-matrix.

$T^{\mu\nu}$ needs to be chosen such that the units on the right-hand side in Eq. 1.7 match the units on the left-hand side. Because the coordinate derivative in the Christoffel symbol (Eq. 1.4) and the second derivate (or multiplication) in the Riemann tensor (Eq. 1.9) each add a dimension $[\text{length}]^{-1}$,

---

[4] It is unfortunate that both use the letter R in the symbol, but the Riemann tensor always has four indices (it is rank four), whereas the Ricci tensor always has two (it is rank two).

the Einstein tensor[5] has dimensions [length]$^{-2}$. Considering that $u^0u^0 = \mathrm{c}^2$ and multiplying with the constants G/c^4, the r.h.s. of Eq. 1.7 currently has dimensions [length]/[mass]. The equation can be completed by including the density ρ in $T^{\mu\nu}$, which has the required dimensions [mass]/[length]^3. A more intuitive explanation is that some form of mass or density had to be included, because that is the source of gravity of the particles in the first place.

The resulting stress-energy is $T^{\mu\nu} = \rho u^\mu u^\nu$, which gives the correct result for particles at rest. It is possible the particles have some non-zero velocity v, or equivalently, that we are observing them from a different frame than their rest-frame. The value of ρ is affected by relativistic length contraction as $\rho = \gamma^2\rho_0$, where ρ_0 is the rest-frame denstiy, and γ is the Lorentz factor:

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{\mathrm{c}^2}}}. \tag{1.12}$$

However, $T^{\mu\nu}$ has to be invariant under a Lorentz transformation, so instead of the changing ρ, the unchanging ρ_0 should be used to compute it.

It is also possible to write $T^{\mu\nu}$ in terms of energy density, using the equivalence between mass and energy:

$$E = m\mathrm{c}^2 = \gamma m_0 \mathrm{c}^2, \tag{1.13}$$

where m_0 is the rest-mass. The (mass) density can be replaced with energy density (ρ_E, energy per volume) and a factor 1/c^2. The updated stress-energy tensor for dust – which is the correct one – is:

$$T^{\mu\nu} = \rho_0 u^\mu u^\nu = \frac{\rho}{\gamma^2} u^\mu u^\nu = \frac{\rho_E}{\gamma^2 \mathrm{c}^2} u^\mu u^\nu. \tag{1.14}$$

For more complex states of matter, this equation would be extended. For example, if the particles were charged, the energy of the electric and magnetic fields would be included. If the particles were interacting, $T^{\mu\nu}$ would include terms for the pressure p:

$$T^{\mu\nu} = \left(\rho_0 + \frac{p}{\mathrm{c}^2}\right) u^\mu u^\nu + g^{\mu\nu} p. \tag{1.15}$$

1.2 The plane wave solution

Einstein realised that it was possible to find a wave solution to his equations (Eq. 1.7) [15, 16]. Starting with a vacuum and a flat space-time, a small perturbation in the metric was found to propagate as a plane wave at the speed of light: a gravitational wave (GW). In this section, we follow the derivation of the plane wave solution and obtain the mathematical description of GWs (adapted from [21]).

---

[5] Tensor contractions in Eqs. 1.10 and 1.11 do not change the dimensions.

1.2.1 Linearised gravity

GWs are assumed to be a small perturbation in an otherwise flat space-time. The background space-time is described by the Minkowski metric $\eta_{\mu\nu}$, and so the perturbed metric is:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1, \tag{1.16}$$

where we have chosen to label the perturbation with $h_{\mu\nu}$. Because the metric is always symmetrical under an exchange of the indices, so must the perturbation be; $h_{\mu\nu} = h_{\nu\mu}$. Given this assumption, the non-linear equations from Sec. 1.1 can be rewritten in *linearised* form; all terms of second order in $|h|$ and higher can be disregarded. Starting with the Christoffel symbol from Eq. 1.4, which becomes:

$$\begin{aligned}\Gamma^{\mu}{}_{\alpha\beta} &= \frac{1}{2}(\eta^{\mu\nu} + h^{\mu\nu})\left(\cancelto{0}{\partial_\alpha \eta_{\nu\beta}} + \partial_\beta h_{\nu\alpha} + \cancelto{0}{\partial_\beta \eta_{\nu\alpha}} + \partial_\beta h_{\nu\alpha} - \cancelto{0}{\partial_\nu \eta_{\alpha\beta}} - \partial_\nu h_{\alpha\beta}\right)\\ &= \frac{1}{2}\eta^{\mu\nu}\left(\partial_\alpha h_{\nu\beta} + \partial_\beta h_{\nu\alpha} - \partial_\nu h_{\alpha\beta}\right).\end{aligned} \tag{1.17}$$

In the second step, all terms of $\mathcal{O}(h^2)$ and higher order have been left out. To get the linearised form of the Riemann tensor from Eq. 1.9, first all terms with two Christoffel symbols can disregarded, since they are all of at least $\mathcal{O}(h^2)$. From the other terms, the linearised Riemann tensor is:

$$\begin{aligned}R^{\rho}_{\sigma\mu\nu} &= \partial_\mu \Gamma^{\rho}_{\nu\sigma} - \partial_\nu \Gamma^{\rho}_{\mu\sigma} + \mathcal{O}(h^2)\\ &= \frac{1}{2}\eta^{\lambda\rho}\partial_\mu\left(\cancel{\partial_\nu h_{\lambda\sigma}} + \partial_\sigma h_{\lambda\nu} - \partial_\lambda h_{\nu\sigma}\right) - \frac{1}{2}\eta^{\lambda\rho}\partial_\nu\left(\cancel{\partial_\mu h_{\lambda\sigma}} + \partial_\sigma h_{\lambda\mu} - \partial_\lambda h_{\mu\sigma}\right)\\ &= \frac{1}{2}\eta^{\lambda\rho}\left(\partial_{\mu\sigma} h_{\lambda\nu} + \partial_{\nu\lambda} h_{\mu\sigma} - \partial_{\mu\lambda} h_{\nu\sigma} - \partial_{\nu\sigma} h_{\lambda\mu}\right).\end{aligned} \tag{1.18}$$

The notation $\partial_{\mu\nu}$ is used as shorthand for $\partial_\mu \partial_\nu$.

In general, the metric tensor is used to raise or lower indices, as in $a^\mu = g^{\mu\nu} a_\nu$, for any vector a^μ. Under the weak-field assumption of Eq. 1.16, raising and lowering indices of small quantities (of $\mathcal{O}(h)$) can be done with the Minkowski metric only, when working to linear order in h:

$$h^{\mu}_{\nu} = g^{\mu\rho} h_{\rho\nu} = (\eta^{\mu\rho} + h^{\mu\rho}) h_{\rho\nu} = \eta^{\mu\rho} h_{\rho\nu} + \mathcal{O}(h^2). \tag{1.19}$$

Using this method, the Riemann tensor from Eq. 1.18 can also be written as:

$$R^{\rho}_{\sigma\mu\nu} = \frac{1}{2}\left(\partial_{\mu\sigma} h^{\rho}_{\nu} + \partial^{\rho}_{\nu} h_{\mu\sigma} - \partial^{\rho}_{\mu} h_{\nu\sigma} - \partial_{\nu\sigma} h^{\rho}_{\mu}\right). \tag{1.20}$$

From here, a contraction of the first and last indices gives the Ricci tensor (Eq. 1.10):

$$\begin{aligned}R_{\sigma\mu} = R^{\rho}_{\sigma\mu\rho} &= \frac{1}{2}\left(\partial_{\mu\sigma} h^{\rho}_{\rho} + \partial^{\rho}_{\rho} h_{\mu\sigma} - \partial^{\rho}_{\mu} h_{\rho\sigma} - \partial_{\rho\sigma} h^{\rho}_{\mu}\right)\\ &= \frac{1}{2}\left(\partial_{\mu\sigma} h + \Box h_{\mu\sigma} - \partial^{\rho}_{\mu} h_{\rho\sigma} - \partial_{\rho\sigma} h^{\rho}_{\mu}\right),\end{aligned} \tag{1.21}$$

where $h = h^\rho_\rho$ is the trace of the metric perturbation, and $\Box = \partial^\rho_\rho$ is the d'Alembertian operator. The contraction of the linearised Ricci tensor gives the linearised Ricci scalar (Eq. 1.11):

$$\begin{aligned} R = R^\sigma_\sigma = \eta^{\sigma\mu} R_{\sigma\mu} &= \frac{1}{2}\left(\partial^\sigma_\sigma h + \Box h^\sigma_\sigma - \partial^{\sigma\rho} h_{\rho\sigma} - \partial_{\rho\sigma} h^{\rho\sigma}\right) \\ &= \frac{1}{2}\left(\Box h + \Box h - \partial_{\sigma\rho} h^{\rho\sigma} - \partial_{\sigma\rho} h^{\rho\sigma}\right) \\ &= \Box h - \partial_{\sigma\rho} h^{\rho\sigma}. \end{aligned} \tag{1.22}$$

Then, given the Ricci tensor and scalar, the linearised Einstein tensor (Eq. 1.7) becomes:

$$\begin{aligned} G_{\mu\nu} &= R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R \eta_{\mu\nu} + \mathcal{O}(h^2) \\ &= \frac{1}{2}\big(\Box h_{\mu\nu} + \partial_{\mu\nu} h - \partial_{\nu\rho} h^\rho_\mu - \partial_{\rho\mu} h^\rho_\nu - \Box h + \partial_{\sigma\rho} h^{\rho\sigma}\big). \end{aligned} \tag{1.23}$$

To simplify Eq. 1.23 somewhat, it is useful to introduce the *trace-reversed* metric tensor, which is defined as:

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h. \tag{1.24}$$

It can be checked that its trace is $\bar{h} = -h$ (hence the name), and so $h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\bar{h}$. By substituting this in 1.23, and only keeping terms up to linear order in $|h|$, the resulting Einstein equations are:

$$\frac{1}{2}\big(\Box \bar{h}_{\mu\nu} - \partial_{\nu\rho}\bar{h}^\rho_\mu - \partial_{\rho\mu}\bar{h}^\rho_\nu + \eta_{\mu\nu}\partial_{\sigma\rho}\bar{h}^{\rho\sigma}\big) = \frac{8\pi \mathrm{G}}{\mathrm{c}^4} T_{\mu\nu}. \tag{1.25}$$

We have implicitly assumed in Eq. 1.16 that the background space-time is flat, and so without a cosmological constant Λ. Hence it is not included in the Einstein equations here.

1.2.2 Gauge freedom

By choosing the Minkowski metric in the form of Eq. 1.1, we have implicitly chosen to work in Cartesian coordinates in a particular inertial frame. In principle, the weak-field assumption of Eq. 1.16 could be chosen in any coordinates. Regarding the frame, it turns out if the assumption holds in one inertial frame, it holds in any that are connected by a Lorentz transformation. I.e. if we consider a different frame that moves with speed v with respect to the original frame, the metric still takes the form of Eq. 1.16, albeit with a transformed $h'_{\mu\nu}$.

Another coordinate transformation that preserves the form of the metric is any infinitesimal change in the coordinates x^μ that is of the same order as h (very small):

$$x'^\mu = x^\mu + \xi^\mu(x), \quad |\xi^\mu| \ll 1. \tag{1.26}$$

From there, the following useful expressions can be derived:

$$\frac{\partial x'^{\mu}}{\partial x^{\nu}} = \delta^{\mu}_{\nu} + \partial_{\nu}\xi^{\mu} \tag{1.27}$$

$$\frac{\partial x^{\mu}}{\partial x'^{\nu}} = \delta^{\mu}_{\nu} - \partial_{\nu}\xi^{\mu}. \tag{1.28}$$

The second expression holds up to linear order, terms of $\mathcal{O}(\xi^2) = \mathcal{O}(h^2)$ are disregarded. The metric, being a rank 2 tensor, transforms with two copies of Eq. 1.28 (again up to linear order in ξ and h):

$$\begin{aligned} g'_{\mu\nu} &= \frac{\partial x^{\alpha}}{\partial x'^{\mu}}\frac{\partial x^{\beta}}{\partial x'^{\nu}} g_{\alpha\beta} \\ &= \left(\delta^{\alpha}_{\mu} - \partial_{\mu}\xi^{\alpha}\right)\left(\delta^{\beta}_{\nu} - \partial_{\nu}\xi^{\beta}\right)\left(\eta_{\alpha\beta} + h_{\alpha\beta}\right) \\ &= \eta_{\mu\nu} + h_{\mu\nu} - \partial_{\mu}\xi_{\nu} - \partial_{\nu}\xi_{\mu} \\ &\equiv \eta_{\mu\nu} + h'_{\mu\nu}. \end{aligned} \tag{1.29}$$

So the metric tensor still takes the form of Eq. 1.16, with the perturbation now being $h'_{\mu\nu}$. An alternative perspective to applying a coordinate transformation is that $h_{\mu\nu}$ is a tensor field for which we are free to choose an alternative *gauge*, taking the form of $h'_{\mu\nu}$. This *gauge freedom* can be exploited to simplify the problem at hand.

Lorenz gauge

The current form of the Einstein equations are written in terms of the trace-reversed tensor $\bar{h}_{\mu\nu}$, in Eq. 1.25. Under the coordinate transformation (or gauge change) just discussed, $\bar{h}_{\mu\nu}$ transforms as:

$$\begin{aligned} \bar{h}'_{\mu\nu} &= h'_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h' \\ &= h_{\mu\nu} - \partial_{\mu}\xi_{\nu} - \partial_{\nu}\xi_{\mu} - \frac{1}{2}\eta_{\mu\nu}\left(h^{\sigma}_{\sigma} - \partial_{\sigma}\xi^{\sigma} - \partial_{\sigma}\xi^{\sigma}\right) \\ &= \bar{h}_{\mu\nu} - \partial_{\mu}\xi_{\nu} - \partial_{\nu}\xi_{\mu} + \eta_{\mu\nu}\partial_{\sigma}\xi^{\sigma}. \end{aligned} \tag{1.30}$$

This can be used to set all the terms in the Einstein equation containing $\partial_{\rho}\bar{h}^{\rho}_{\mu}$ to zero. This expression, after applying the transformation, is:

$$\begin{aligned} \partial_{\rho}\bar{h}'^{\rho}_{\mu} &= \partial_{\rho}\left(\bar{h}^{\rho}_{\mu} - \partial^{\rho}\xi_{\mu} - \partial_{\mu}\xi^{\rho} + \eta^{\rho}_{\mu}\partial_{\sigma}\xi^{\sigma}\right) \\ &= \partial_{\rho}\bar{h}^{\rho}_{\mu} - \Box\xi_{\mu}. \end{aligned} \tag{1.31}$$

So, by choosing a gauge transformation where $\Box\xi_{\mu} = \partial_{\rho}\bar{h}^{\rho}_{\mu}$, these terms become zero. Adapting this commonly named *Lorenz gauge*, the Einstein equations simplify to:

$$\frac{1}{2}\Box\bar{h}_{\mu\nu} = \frac{8\pi \mathrm{G}}{\mathrm{c}^4}T_{\mu\nu}. \tag{1.32}$$

Here, $\bar{h}_{\mu\nu}$ was previously written with a prime (the transformed tensor). From here instead, we assume that whatever $\bar{h}_{\mu\nu}$ is used, it satisfies the Lorenz gauge condition:

$$\partial_\mu \bar{h}^{\mu\nu} = 0, \tag{1.33}$$

knowing that if it did not, we could always choose to use the transformed $\bar{h}'_{\mu\nu}$ that does satisfy it instead.

1.2.3 Gravitational waves in vacuum

At the start of this section, it was mentioned that GWs were found assuming a small perturbation on a flat space-time in vacuum. The former leads to the linearised Einstein equations. To implement the latter, set $T_{\mu\nu} = 0$, as there is no form of matter, energy or anything present. Now the equations read:

$$\Box \bar{h}_{\mu\nu} = 0. \tag{1.34}$$

This is the form of a wave equation, with a familiar solution:

$$\bar{h}_{\mu\nu}(x) = A_{\mu\nu} \exp(ik_\rho x^\rho), \tag{1.35}$$

where $A_{\mu\nu}$ is the amplitude of the wave (each component of this rank 2 tensor is constant), k_ρ is the four-wavevector, and x^ρ is the space-time coordinate. The four-wavevector has components $(\omega/\mathrm{c}, \vec{k})$, where ω is the angular frequency, and $\vec{k}$ is the spatial wavevector in three dimensions. The frequency ω and the propagation speed v are related via the associated wavenumber $|\vec{k}| = k = \frac{2\pi}{\lambda} = \frac{\omega}{v}$.

To verify the GW solution, it is substituted back into the wave equations 1.34. It can be checked that the solution holds, under the condition that

$$k_\alpha k^\alpha = 0. \tag{1.36}$$

In other words, the four-wavevector has to be null. From the definition of k^μ, it follows that the propagation speed v of GWs has to equal the speed of light. Apart from the Einstein equations, the Lorenz gauge condition (Eq. 1.33) has to be satisfied as well. This results in the following additional constrained on k^μ:

$$k_\mu A^{\mu\nu} = 0. \tag{1.37}$$

Any plane wave of the form of Eq. 1.35, provided that the wave vector is null and that it obeys Eq. 1.37, is a valid solution to the linearised Einstein equations in vacuum. Because the equations are linear, any superposition of these plane waves is a solution as well.

The transverse-traceless gauge

In order to illuminate the physical properties of GWs, here we investigate the amplitude tensor $A^{\mu\nu}$ and the constraints on its components. Firstly, the symmetry of the metric tensor and of $h^{\mu\nu}$ means that $A^{\mu\nu} = A^{\nu\mu}$ must also be symmetric. To simplify the following arguments, consider only a single plane wave propagating in the z-direction. It's four-wavevector is $k^\mu = (k, 0, 0, k)$, with wavenumber $k = \omega/\mathrm{c}$. The second constraint on $A^{\mu\nu}$ follows from the Lorenz gauge condition (Eq. 1.37): $A^{\mu 0} = A^{\mu 3}$.

This still leaves $A^{\mu\nu}$ with six independent components. This number can be further reduced by introducing another gauge transformation:

$$x'^\mu = x^\mu + \zeta^\mu(x), \quad \zeta^\mu(x) = \epsilon^\mu \exp(ik_\rho x^\rho). \tag{1.38}$$

This new gauge does not break the previously set gauge conditions: it can be verified that $\Box\eta^\mu = -k_\rho k_\rho \zeta^\mu = 0$. This does not depend on the choice of ϵ^μ (assuming they are constants), so we are free to specify them. The following choice turns out to greatly simplify the amplitude tensor:

$$\begin{aligned}
\epsilon^0 &= -\frac{i}{4k}(2A^{00} + A^{11} + A^{22}),\\
\epsilon^1 &= -\frac{i}{k}A^{01},\\
\epsilon^2 &= -\frac{i}{k}A^{02},\\
\epsilon^3 &= -\frac{i}{4k}(2A^{00} - A^{11} - A^{22}).
\end{aligned} \tag{1.39}$$

The algebra is not shown here, but it results in the additional constraints $A'^{00} = A'^{01} = A'^{02} = 0$ and $A'^{11} = -A'^{22}$. The collective constraints on $A'^{\mu\nu}$ mean that it only has two non-zero, independent components, which are usually denoted $A'^{11} = h^+$ and $A'^{12} = h^\times$.

Adopting this *transverse-traceless* gauge and re-labelling the transformed GW $h^{\mu\nu}_{\mathrm{TT}}$, the result is:

$$h^{\mu\nu}_{\mathrm{TT}} = A^{\mu\nu}_{\mathrm{TT}} \exp(i(-kx^0 + kx^3)) = A^{\mu\nu}_{\mathrm{TT}} \exp(i(kz - \omega t)), \tag{1.40}$$

where we have defined:

$$A^{\mu\nu}_{\mathrm{TT}} \equiv \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h^{+} & h^{\times} & 0 \\ 0 & h^{\times} & -h^{+} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1.41}$$

This wave is transverse: its propagation direction (z) is perpendicular to its amplitude, which acts in the x and y directions. It is called traceless, because its trace is zero: $\mathrm{Tr}(h_{\mathrm{TT}}) \propto h^{+} - h^{+} = 0$.

Two wave polarisations

GWs are a perturbation of spacetime, changing the metric as they pass through. The observable effect this has is a change in physical distance between two objects, lengthening or shortening it. This is often described as a "stretching and squashing" of space-time. To explore this statement, first consider a single test particle on a geodesic. If it is initially at rest, its four-velocity is:

$$u^{\mu} = \frac{\mathrm{d}x^{\mu}}{\mathrm{d}\tau} = c(1,0,0,0). \tag{1.42}$$

Substituting this into the geodesic equation (Eq. 1.3), we have:

$$\frac{\mathrm{d}u^{\mu}}{\mathrm{d}\tau} + \mathrm{c}^2\Gamma^{\mu}_{00} = 0. \tag{1.43}$$

Computing the Christoffel symbol gives $\Gamma^{\mu}_{00} = 0$, as all components of $h^{\mathrm{TT}}_{\mu\nu}$ with at least one zero index are zero (Eq. 1.41). Therefore, the four-velocity is constant, $\frac{\mathrm{d}u^{\mu}}{\mathrm{d}\tau} = 0$, and the particle stays at rest with respect to its coordinates.

Evidently, we need to consider multiple particles to observe any GWs. The following provides an example using a very simple setup of a Pulsar Timing Array (PTA). Given a ring of completely regularly spinning pulsars in the xy-plane, around the solar system barycenter (SSB), we measure[6] radio pulses from each of them at a constant frequency ν. A GW propagating in the z-direction passing through does not move the pulsars out of the plane (as was just shown). It does, however, affect the measured pulsar frequencies.

Consider one of the pulsars in the ring, at an angle α from the x-axis and distance R, i.e. at a position $(R\cos(\alpha), R\sin(\alpha), 0)$. The photons sent out from this pulsar travel on a geodesic x^{μ}, parametrised by affine parameter λ. The ones that reach the SSB, in the absence of a GW, must have

---

[6] Clearly, we do not have radio telescopes at the SSB. In reality, the pulsars are observed on Earth, and the observations are corrected for the difference in timing between the Earth and SSB.

a four-momentum in the right direction[7]:

$$q^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \propto \nu\,(1, -\cos(\alpha), -\sin(\alpha), 0)\,. \tag{1.44}$$

The photons move at the speed of light, and so their four-momentum is null: $\eta_{\mu\nu}q^\mu q^\nu = 0$. By perturbing the metric with $h^{\mathrm{TT}}_{\mu\nu}$, photons from a slightly different path reach the SSB. The perturbed four-momentum $q'^\mu = q^\mu + \delta q^\mu$ has to be null with respect to the perturbed metric:

$$g_{\mu\nu}q'^\mu q'^\nu = (\eta_{\mu\nu} + h^{\mathrm{TT}}_{\mu\nu})(q^\mu + \delta q^\mu)(q^\nu + \delta q^\nu) = 0. \tag{1.45}$$

Using the nullity of the unperturbed momentum, and discarding terms with higher than linear order in $|\delta q|$ or $|h|$, the following solution can be found:

$$q'^\mu = q^\mu - \frac{1}{2}\eta^{\mu\nu}h^{\mathrm{TT}}_{\nu\rho}q^\rho \tag{1.46}$$

$$\propto \nu \begin{pmatrix} 1 \\ -\cos(\alpha)\left(1 - \frac{1}{2}h^+\cos(\phi_{\mathrm{GW}})\right) + \sin(\alpha)\left(\frac{1}{2}h^\times\cos(\phi_{\mathrm{GW}})\right) \\ -\sin(\alpha)\left(1 - \frac{1}{2}h^+\cos(\phi_{\mathrm{GW}})\right) + \cos(\alpha)\left(\frac{1}{2}h^\times\cos(\phi_{\mathrm{GW}})\right) \\ 0 \end{pmatrix}. \tag{1.47}$$

The explicit form of the second line is obtained by substituting the unperturbed four-momentum (Eq. 1.44), and $h^{\mathrm{TT}}_{\mu\nu}$ from Eq. 1.40 and Eq. 1.41. As we are concerned with an observable effect, we have taken the real part of $h^{\mathrm{TT}}_{\mu\nu}$ only. The GW phase is defined as $\phi_{\mathrm{GW}} \equiv kz - \omega t$ (which is equal to $-\omega t$ in the xy-plane).

The geodesic equation for the four-momentum can be used to derive a relation between the pulsar frequency ν and the GW components h^+ and $h^\times$. Its time-component reads[8]:

$$\frac{dq'^0}{d\lambda} + \Gamma^0_{\alpha\beta}q'^\alpha q'^\beta = 0. \tag{1.48}$$

In the TT-gauge, $\Gamma^0_{\alpha\beta} = \frac{1}{2}\frac{\partial}{\partial t}h^{\mathrm{TT}}_{\alpha\beta}$. Substituting this and Eq. 1.47 into the geodesic equation yields:

$$\frac{d\nu}{d\lambda} \propto \nu^2\Big(-\frac{1}{2}\frac{\partial}{\partial t}\big(h^+\cos(\phi_{\mathrm{GW}})\big)\big(\cos(\alpha)^2 - \sin(\alpha)^2\big) + \frac{\partial}{\partial t}\big(h^\times\cos(\phi_{\mathrm{GW}})\big)\sin(\alpha)\cos(\alpha)\Big). \tag{1.49}$$

The derivatives with respect to the affine parameter λ, and the coordinates t and z, can be related to

---

[7] Because the Doppler shift is a ratio of frequencies, we can use a proportionality rather then filling in the correct constants in this equation.

[8] The photons travel on a null geodesic, which makes their proper time not well-defined. Therefore a generic affine parameter λ has to be used.

each other with:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda} &= \frac{\partial t}{\partial \lambda}\frac{\partial}{\partial t} + \frac{\partial z}{\partial \lambda}\frac{\partial}{\partial z} \\ &= q'^0 \frac{\partial}{\partial t} + q'^3 \frac{\partial}{\partial z} \\ &\propto \nu \frac{\partial}{\partial t} + 0. \end{aligned} \tag{1.50}$$

Using this turns Eq. 1.49 into a differential equation for the pulsar frequency:

$$\frac{1}{\nu}\frac{\mathrm{d}\nu}{\mathrm{d}t} \propto \Big(-\frac{1}{2}\frac{\partial}{\partial t}\big(h^{+}\cos(\phi_{\mathrm{GW}})\big)\big(\cos(\alpha)^2 - \sin(\alpha)^2\big) + \frac{\partial}{\partial t}\big(h^{\times}\cos(\phi_{\mathrm{GW}})\big)\sin(\alpha)\cos(\alpha)\Big). \tag{1.51}$$

It can be solved by integrating over the time t, from the start time at the pulsar P to the end time at the SSB:

$$\frac{\nu_{\mathrm{SSB}}}{\nu_{\mathrm{P}}} = \exp\Big(-\frac{1}{2}\Delta h^{+}\cos(\phi_{\mathrm{GW}})\big(\cos(\alpha)^2 - \sin(\alpha)^2\big) + \Delta h^{\times}\cos(\phi_{\mathrm{GW}})\sin(\alpha)\cos(\alpha)\Big), \tag{1.52}$$

where we have defined $\Delta h^{+/\times} = h^{+/\times}_{\mathrm{SSB}} - h^{+/\times}_{\mathrm{P}}$. The final Doppler shift z can be found by expanding the exponential to first order in $|h|$:

$$z \equiv \frac{\nu_{\mathrm{P}} - \nu_{\mathrm{SSB}}}{\nu_{\mathrm{P}}} = \Big(\frac{1}{2}\Delta h^{+}\cos(\phi_{\mathrm{GW}})\big(\cos(\alpha)^2 - \sin(\alpha)^2\big) - \Delta h^{\times}\cos(\phi_{\mathrm{GW}})\sin(\alpha)\cos(\alpha)\Big). \tag{1.53}$$

This Doppler shift can be nicely visualised in the two *polarisation* states of the GW. A "plus"-polarised GW has a non-zero h^{+}, and $h^{\times} = 0$. When this passes through the ring PTA, the pulsar Doppler shifts show up along the legs of the + symbol, as in Fig. 1.1. Conversely a "cross"-polarised GW has non-zero $h^{\times}$ and $h^{+} = 0$. It induces a Doppler shift along the legs of the × symbol, rotated 45° with respect to the +-polarisation, as in Fig. 1.2. In general, a GW is a superposition of these two polarisations.

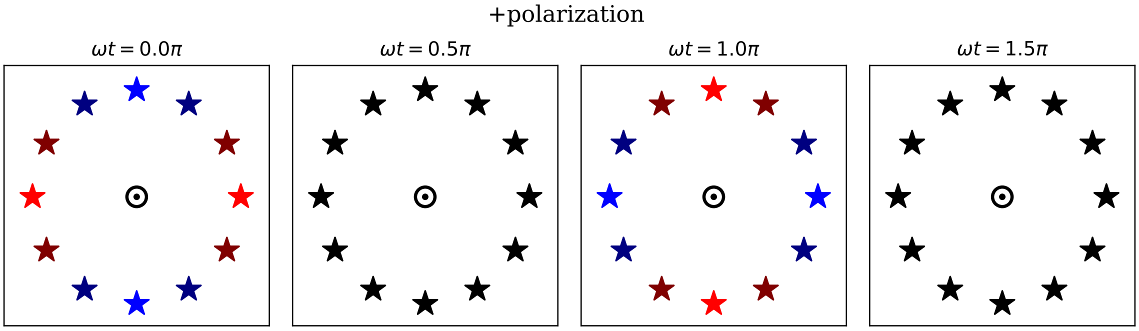

Figure 1.1: Doppler shifts in the ring PTA (see text) with a +-polarised GW passing through. The example pulsars are marked with stars, the SSB is at the centre. Red stars are pulsars whose frequency is redshifted, blue ones are blueshifted, and black ones are unaffected. The four plots show subsequent snapshots at different points in the GW phase, ωt.

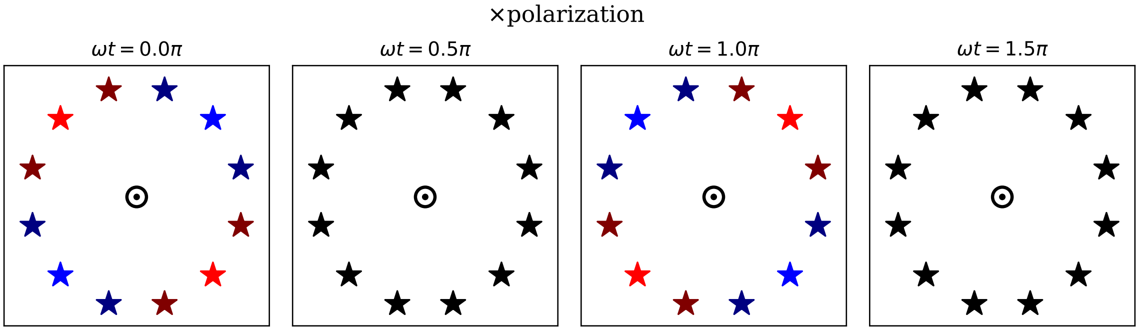

Figure 1.2: Same as Fig. 1.1, but with a ×-polarised GW passing through the ring PTA.

1.3 The gravitational wave universe

1.3.1 Sources of gravitational waves

In the previous section, we have found that a GW is a solution to the linearised Einstein equations in vacuum (Eq. 1.34). However, in the linear theory, the vacuum only allows the propagation of GWs, not their generation. To find what sources a GW, we need to go a step back to Eq. 1.32, which includes the stress-energy tensor:

$$\frac{1}{2}\Box\bar{h}_{\mu\nu} = \frac{8\pi \mathrm{G}}{\mathrm{c}^4}T_{\mu\nu}. \tag{1.54}$$

It is not shown here (the derivation uses standard techniques such as a Green's function, see for example [21]), but a general solution can be found at any point $\boldsymbol{x} = (\mathrm{c}t, \vec{x})$ given a source localised at points $\vec{y}$:

$$\bar{h}^{\mu\nu}(\mathrm{c}t, \vec{x}) = -\frac{4\mathrm{G}}{\mathrm{c}^4} \int \frac{T^{\mu\nu}(\mathrm{c}t_\mathrm{r}, \vec{y})}{|\vec{x} - \vec{y}|} \mathrm{d}^3\vec{y}. \tag{1.55}$$

Here, $t_\mathrm{r} = t - |\vec{x} - \vec{y}|/\mathrm{c}$ is the *retarded time*, as it takes the information from a point $\vec{y}$ a time $|\vec{x} - \vec{y}|/\mathrm{c}$ to travel to $\vec{x}$ and have an effect. This solution is found in the Lorenz gauge, and so the condition 1.33 has to be satisfied. It can be checked that this is the case when respecting conservation of the stress-energy tensor, $\partial_\mu T^{\mu\nu} = 0$.

A powerful tool for dealing with Eq. 1.55, is to expand the distance to the source point $|\vec{x} - \vec{y}|$ in terms of the distance to the origin $r \equiv |\vec{x}|$, which we have chosen at (or in) the source. This leads to the multipole expansion:

$$\bar{h}^{\mu\nu}(\mathrm{c}t, \vec{x}) = -\frac{4\mathrm{G}}{\mathrm{c}^4} \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} M^{\mu\nu i_1 i_2 \ldots i_l}(\mathrm{c}t_\mathrm{r})\, \partial_{i_1}\partial_{i_2} \ldots \partial_{i_l}(1/r), \tag{1.56}$$

where the multipole moments of the source are defined as:

$$M^{\mu\nu i_1 i_2 \ldots i_l}(\mathrm{c}t) \equiv \int T^{\mu\nu}(\mathrm{c}t, \vec{y}) y^{i_1} y^{i_2} \ldots y^{i_l} \mathrm{d}^3\vec{y}. \tag{1.57}$$

This expansion is especially useful if we only consider field points $\boldsymbol{x}$ outside the source, and far enough away such that the distance to the source is relatively large compared to its size, i.e. the source is compact. In this *compact source approximation*, only the first term of Eq. 1.56 is needed, and we can use that $\mathrm{c}t_\mathrm{r} \approx \mathrm{c}t - r$:

$$\bar{h}^{\mu\nu}(\mathrm{c}t, \vec{x}) = -\frac{4\mathrm{G}}{\mathrm{c}^4} \frac{1}{r} \int T^{\mu\nu}(\mathrm{c}t - r, \vec{y}) \mathrm{d}^3\vec{y}. \tag{1.58}$$

A physical interpretation can be associated with the components of $T^{\mu\nu}$ in this equation. The 00-term represents the total mass M (and energy) in the source, $\int T^{00} \mathrm{d}^3\vec{y} = M\mathrm{c}^2$. The $0i$- and $i0$-terms are related to the total momentum P^i of the source particles, $\int T^{0i} \mathrm{d}^3\vec{y} = P^i\mathrm{c}$. Both these quantities are constant for an isolated source (no mass or momentum can be added or taken away). The other components ij represent the internal stresses in the source, which can be related to the quadrupole moment (given in the following). First, changing coordinates to the centre of momentum frame allows to set $P^i = 0$, so that the first components of Eq. 1.58 become:

$$\bar{h}^{00} = -\frac{4\mathrm{G}}{\mathrm{c}^2} \frac{1}{r} M. \tag{1.59}$$

$$\bar{h}^{0i} = \bar{h}^{i0} = 0. \tag{1.60}$$

Some manipulation of the integral over the spatial components allows $\bar{h}^{ij}$ to be expressed in terms of the quadrupole moment tensor I^{ij}:

$$\bar{h}^{ij}(\mathrm{c}t, \vec{x}) = -\frac{2\mathrm{G}}{\mathrm{c}^6 r}\left[\frac{\mathrm{d}^2 I^{ij}(\mathrm{c}t')}{\mathrm{d}t'^2}\right]_r, \tag{1.61}$$

$$I^{ij}(\mathrm{c}t) \equiv \int T^{00}(\mathrm{c}t, \vec{y}) y^i y^j \mathrm{d}^3\vec{y}. \tag{1.62}$$

Here the brackets $\left[...\right]_r$ signify that the quantity inside has to be evaluated at the retarded time t_r.

The gravitational field due to Eq. 1.59 can not explain the generation of GWs, because it stays constant in time. Therefore, it has to be Eq. 1.61, the *quadrupole formula*, that is responsible for them. For a simplified source, with slowly moving particles, the gravitational field is mostly due to its mass density ρ (see Sec. 1.1.3), so that $T^{00} \approx \rho \mathrm{c}^2$. Then the quadrupole moment tensor is equal to the quadrupole moment of the matter density:

$$I^{ij}(\mathrm{c}t) = \mathrm{c}^2 \int \rho(\mathrm{c}t, \vec{x}) x^i x^j \mathrm{d}^3\vec{x}. \tag{1.63}$$

In other words, a source with only a density monopole (e.g. a stationary mass) or a dipole (e.g. a mass on a linear trajectory) do not generate any GWs.

1.3.2 Gravitational waves from black hole binaries

A simple example of a matter distribution with a non-zero quadrupole moment is that of two (massive) particles orbiting each other, which essentially describes the physics of two black holes in orbit. We can therefore use this example to study binary black holes (BBHs). The resulting GWs can be derived by modelling the two black holes as point masses m_1 and m_2, orbiting at radii R_1 and R_2 around a common centre of mass. We will assume the black holes (BHs) are far enough apart and move slowly enough, that any relativistic effects on their motion can be neglected (and the Lorentz factor $\gamma = 0$). First, we define the total mass M, the reduced mass μ, and the separation R:

$$M \equiv m_1 + m_2, \tag{1.64}$$

$$\mu \equiv \frac{m_1 m_2}{M}, \tag{1.65}$$

$$R \equiv R_1 + R_2. \tag{1.66}$$

Using that the centre of mass is the weighted sum of orbital radii, $m_1 R_1 = m_2 R_2$, we find:

$$R_A = \frac{\mu}{m_A} R, \quad A = 1, 2. \tag{1.67}$$

It is useful to choose the coordinate system such that the centre of the binary is at rest at the origin, and to choose the position of the observer at a distance r along the z-axis, so that later on the TT-gauge from section 1.2.3 can be used. A completely "face-on" BBH would be in the xy-plane, but in general, the BBH can be at any inclination angle ι with the xy-plane. The phase of the BBH can be expressed as $\omega_{\mathrm{o}} t$, where ω_{o} is the angular orbital frequency. With these conventions, the positions of the black holes are:

$$\vec{x}_A(t) = \frac{\mu}{m_A} R\,\hat{\epsilon}(t), \tag{1.68}$$

where we have defined the unit vector $\hat{\epsilon}$ as:

$$\hat{\epsilon}(t) = \big(\cos(\omega_{\mathrm{o}} t), \sin(\omega_{\mathrm{o}} t)\cos(\iota), \sin(\iota)\big). \tag{1.69}$$

Treating the black holes as point particles, the matter density distribution of the binary is given by two delta functions:

$$\rho(\mathrm{c}t, \vec{x}) = m_1 \delta(\vec{x} - \vec{x}_1(t)) + m_2 \delta(\vec{x} - \vec{x}_2(t)). \tag{1.70}$$

This can be filled in the equation (1.62) for the quadrupole moment tensor to find:

$$I^{ij}(\mathrm{c}t) = \mathrm{c}^2 \mu R^2\, \hat{\epsilon}^i(t)\, \hat{\epsilon}^j(t). \tag{1.71}$$

The GWs can be found from the radiative part of the metric perturbation, $h^{ij}(t)$ from Eq. 1.61. The other terms (with $\mu = 0$ or $\nu = 0$) are set to zero. Substituting Eq. 1.68 into Eq. 1.71, taking two time derivatives, and using some trigonometric identities results in:

$$\bar{h}^{\mu\nu}(\mathrm{c}t) = \frac{4}{r}\frac{\mathrm{G}\mu R^2 \omega_{\mathrm{o}}^2}{\mathrm{c}^4} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \cos(2\omega_{\mathrm{o}} t_{\mathrm{r}}) & \sin(2\omega_{\mathrm{o}} t_{\mathrm{r}})\cos(\iota) & \sin(2\omega_{\mathrm{o}} t_{\mathrm{r}})\sin(\iota) \\ 0 & \sin(2\omega_{\mathrm{o}} t_{\mathrm{r}})\cos(\iota) & -\cos(2\omega_{\mathrm{o}} t_{\mathrm{r}})\cos(\iota)^2 & -\cos(2\omega_{\mathrm{o}} t_{\mathrm{r}})\cos(\iota)\sin(\iota) \\ 0 & \sin(2\omega_{\mathrm{o}} t_{\mathrm{r}})\sin(\iota) & -\cos(2\omega_{\mathrm{o}} t_{\mathrm{r}})\cos(\iota)\sin(\iota) & -\cos(2\omega_{\mathrm{o}} t_{\mathrm{r}})\sin(\iota)^2 \end{pmatrix}. \tag{1.72}$$

The right-hand-side has to be evaluated at the retarded time $t_{\mathrm{r}} = t - r/\mathrm{c}$.

This expression can be simplified by adopting the TT-gauge, applying the gauge transformation given in Eq. 1.38 and Eq. 1.39. Then, the GW takes the familiar form of the TT-gauge with only transverse components. In general, the two polarisation amplitudes, h^{+} and $h^{\times}$, are a mix of the two polarisation states, $\bar{h}^{+}$ and $\bar{h}^{\times}$ as defined below. The "mixing" can be expressed using the *polarisation angle* ψ $(0 \leq \psi < \pi)$. This angle quantifies the rotation of $\hat{l}$, the normal to the orbital plane of the binary, about $\hat{k}$, the GW propagation direction (i.e. the line-of-sight), with respect to the z-axis of the chosen coordinate system; $\cos\psi = (\hat{l} \times \hat{k}) \cdot (\hat{z} \times \hat{k})$. Additionally, the GW includes an arbitrary phase offset ϕ_0:

$$h^{\mu\nu}_{\mathrm{TT}}(\mathrm{c}t) = \frac{4}{r}\frac{\mathrm{G}\mu R^2\omega_{\mathrm{o}}^2}{\mathrm{c}^4}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h^{+}(t_{\mathrm{r}}) & h^{\times}(t_{\mathrm{r}}) & 0 \\ 0 & h^{\times}(t_{\mathrm{r}}) & -h^{+}(t_{\mathrm{r}}) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{1.73}$$

$$\begin{aligned} h^{+} &= \bar{h}^{+}\cos 2\psi - \bar{h}^{\times}\sin 2\psi, \\ h^{\times} &= \bar{h}^{+}\sin 2\psi + \bar{h}^{\times}\cos 2\psi, \end{aligned} \tag{1.74}$$

$$\begin{aligned} \bar{h}^{+} &= \cos(2\omega_{\mathrm{o}}t_{\mathrm{r}} + \phi_0)\big(\frac{1}{2} + \frac{1}{2}\cos(\iota)^2\big), \\ \bar{h}^{\times} &= \sin(2\omega_{\mathrm{o}}t_{\mathrm{r}} + \phi_0)\cos(\iota). \end{aligned} \tag{1.75}$$

In these last equations, the factor 2 in front of ω_{o} means that the GW frequency f (angular frequency ω) is always double that of the orbital frequency:

$$f = \frac{\omega}{2\pi} = \frac{\omega_{\mathrm{o}}}{\pi}. \tag{1.76}$$

This is due to the quadrupolar nature of GWs; having two terms $\hat{\epsilon}^i$ and $\hat{\epsilon}^j$ in Eq. 1.71 has led to multiplying two sinusoidal terms ($\cos(\omega_{\mathrm{o}}t_{\mathrm{r}})$ and/or $\sin(\omega_{\mathrm{o}}t_{\mathrm{r}})$), which always results in a doubled angle with trigonometric identities.

In the Newtonian limit, the binary's orbit is given by Kepler's third law (for a circular orbit):

$$R^3\omega_{\mathrm{o}}^2 = \mathrm{G}M. \tag{1.77}$$

Using this to eliminate R from Eq. 1.73, the prefactor, which we can define as the GW amplitude A, can be rewritten as follows:

$$A \equiv \frac{4}{r}\frac{\mathrm{G}\mu R^2\omega_{\mathrm{o}}^2}{\mathrm{c}^4} = \frac{4}{r}\left(\frac{\mathrm{G}\mathcal{M}_{\mathrm{c}}}{\mathrm{c}^2}\right)^{5/3}\left(\frac{\omega_{\mathrm{o}}}{\mathrm{c}}\right)^{2/3} \tag{1.78}$$

where we have introduced the *chirp mass*:

$$\mathcal{M}_{\mathrm{c}} \equiv \frac{(m_1 m_2)^{3/5}}{M^{1/5}}. \tag{1.79}$$

So far, we have not taken into account that the orbit can change. Because GWs actually carry away energy, the orbital energy decreases and the binary shrinks. This in turn increases the orbital frequency (and thus the GW frequency). As long as this *inspiral* of the binary is adiabatic, i.e. the separation shrinks slowly compared to the orbital motion, the change in frequency can be computed on the basis of the classical binary motion. It is not shown here, but the result of equating the GW energy to a loss in orbital energy is the following expression for the GW frequency (Eq. 1.76):

$$f = \frac{1}{\pi}\left(\frac{\mathrm{G}\mathcal{M}_\mathrm{c}}{\mathrm{c}^3}\right)^{-5/8}\left(\frac{5}{256}\frac{1}{\tau}\right)^{3/8}. \tag{1.80}$$

Here, τ is the time until the presumed end point of the inspiral, where the two black holes collide (usually chosen to be $t = 0$). The amplitude from Eq. 1.78 in terms of the time-dependent frequency is:

$$A = \frac{4}{r}\left(\frac{\mathrm{G}\mathcal{M}_\mathrm{c}}{\mathrm{c}^2}\right)^{5/3}\left(\frac{\pi f(t_\mathrm{r})}{\mathrm{c}}\right)^{2/3} \tag{1.81}$$

The phase of the GW now has to be determined by integrating over the changing frequency, starting at some reference point t_0 with reference phase ϕ_0, so that we replace in Eq. 1.75:

$$\phi = 2\omega_\mathrm{o} t + \phi_0 \to \int_{t_0}^{t} \mathrm{d}t' f(t') + \phi_0. \tag{1.82}$$

More accurate approximations of the binary motion and GW phase can be obtained with an iterative process: taking the resulting GW to compute the orbital energy loss, then the GW frequency, which in turn results in a more accurate phase etc. However for BBHs far from coalescence, often the monochromatic wave of Eq. 1.73 is accurate enough to model their GWs.

1.3.3 The gravitational wave spectrum

For a BBH, the frequency of GWs emitted is determined by the chirp mass and separation of the two objects. Using the classical relation from Eq. 1.77, the GW frequency is:

$$f = \frac{1}{\pi}\sqrt{\frac{\mathrm{G}M}{R^3}}. \tag{1.83}$$

Because the mass range of black holes is very wide - from stellar remnants of order $10\mathrm{M}_\odot$ to super massive black holes (SMBHs) of $10^5 - 10^{10}\mathrm{M}_\odot$ - there is a vast spectrum of GWs. Additionally, BBHs are not the only source of (potentially) observable GWs. Figure 1.3 shows an overview of the GW spectrum, with the main source categories and sensitivity curves of ground-based, space-based, and Pulsar Timing Array (PTA) detectors.

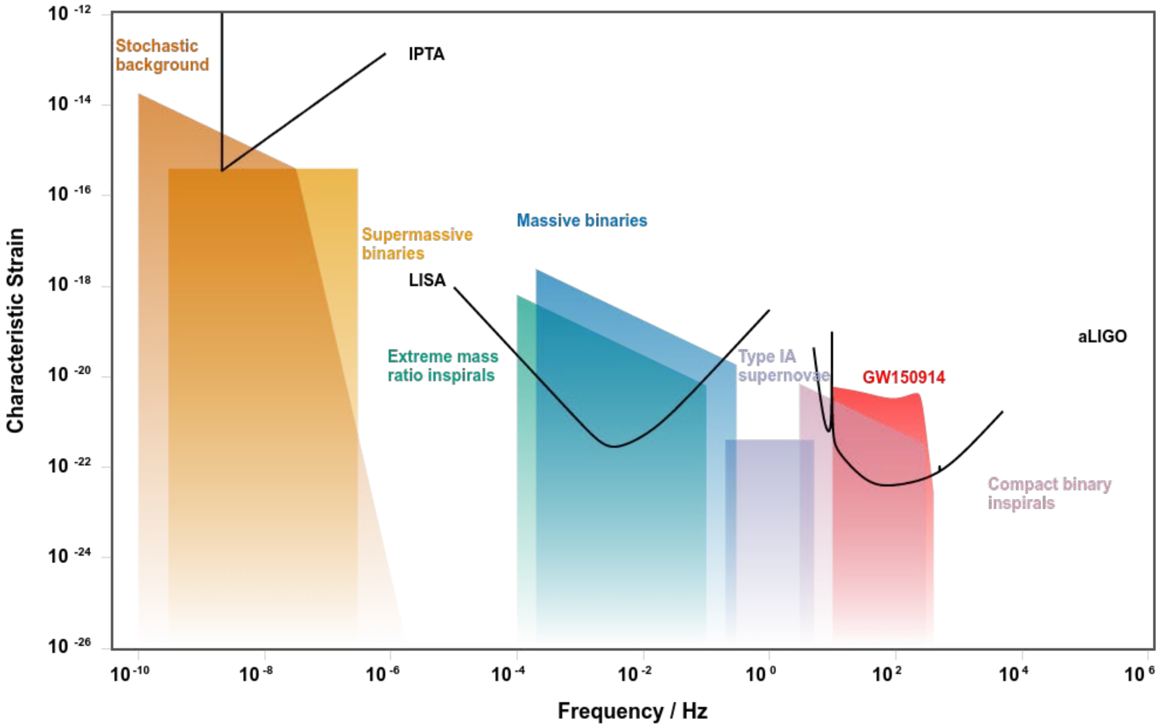

Figure 1.3: The GW spectrum of several different source types (see text). For each source, the characteristic strain (related to the GW amplitude, see Eq. 17 in [1] for an exact definition) is plotted against frequency as a coloured block. Sensitivity curves (characteristic strain of the noise) of different detectors are over-plotted as black lines. Figure made with `http://gwplotter.com/` [1].

Binary black holes across the spectrum

PTA detectors observe at the lowest frequency end of the GW spectrum, and are therefore used to search for the most massive sources. With their frequency band of approximately $3 \times 10^{-9} - 10^{-6}$Hz, PTAs have the potential to observe individual SMBH binaries with masses of at least $\sim 10^8 M_\odot$ [20], and a background signal build up of all SMBH binaries [22] ("supermassive binaries" and "stochastic background" in Fig. 1.3). PTAs and their potential for detection will be discussed in detail in the next chapter.

The LISA detector will be sensitive at much higher frequencies than PTAs, around $10^{-4} - 10^{-1}$Hz [18]. As such, it will be observing smaller BBHs with masses around $10^3 - 10^7 M_\odot$ ("massive binaries" in Fig. 1.3). A special case are the extreme mass ratio inspirals (EMRIs) (in Fig. 1.3), where a stellar mass black hole ($10 - 60 M_\odot$) inspirals around a massive black hole ($10^5 - 10^6 M_\odot$). These highly relativistic systems spend many orbits in the LISA band and as such provide a tool for precise measurements of the environment near the central black hole [18].

Some of the BBHs will continue their inspiral and become visible in the LIGO+Virgo sensitivity band [23], roughly $30-10^4$Hz [24]. Moreover, these detectors can observe (and have already detected) many "stellar mass" binaries, with frequencies too high for LISA to observe. The mass range of binaries visible to ground-based detectors can be determined using the concept of an innermost stable circular orbit (ISCO) of a black hole of mass M:

$$R_{\mathrm{ISCO}} = \frac{6\mathrm{G}M}{\mathrm{c}^2}. \tag{1.84}$$

This is the closest a test particle can get to the black hole in a stable orbit. As a rough approximation, we can use it as the smallest separation possible in a binary of mass M. Then the maximum frequency before merger is found by substituting this R into Eq. 1.83:

$$f_{\max} = \frac{1}{\pi}\frac{\mathrm{c}^3}{6^{3/2}\mathrm{G}M}. \tag{1.85}$$

This means that advanced LIGO can (in principle) detect BBH inspirals and mergers from the smallest stellar mass black holes to massive black holes up to $\sim$150$\mathrm{M}_\odot$. GW150914 (in Fig. 1.3) was the first detection of GWs with LIGO, generated by two black holes inspiralling, and subsequently merging into a new black hole of $\sim$62$\mathrm{M}_\odot$ [17].

Other gravitational wave sources

A binary neutron star (BNS) is similar to a BBH, in that it is reasonably approximated by two point masses orbiting each other. Hence, the generated GWs are comparable, and because neutron stars are smaller than (stellar mass) black holes, they fall into the sensitivity band of ground-based detectors. The first GW observation of a BNS was made by advanced LIGO and advanced Virgo [25]. Because black holes and neutron stars are collectively called compact objects, the category in Fig. 1.3 is labelled as "compact binary inspirals".

As discussed in Sec. 1.3.1, any mass quadrupole can generate GWs. The quadrupole moment of a spinning sphere or other axisymmetric object is zero, but with additional asymmetry, it becomes non-zero. It is thought that this scenario may be applicable to neutron stars, which are known to be spinning. If there is some deformity on the neutron star's surface – possibly due to its magnetic field or stresses in the dense matter – it will emit a continuous gravitational wave at twice its rotational frequency [26, and references therein]. An exploding star may also have a non-zero quadrupole moment, so long as the explosion is not spherically symmetric. It may therefore be possible to observe a burst of GWs when a star that is close enough goes supernova [e.g. 26] (Fig. 1.3 specifically includes a type IA supernova, from an accreting white dwarf).

In theory, it is possible that some GWs could be of cosmic origin. A potential source are *cosmic strings*, formed in the early universe. The collective of oscillating string loops could generate a GW background, which has so far been constrained by PTA observations [27]. GW bursts from the cusps of string loops are potentially detectable by LIGO and LISA [27]. It is also thought that a *relic background* of GWs from the early universe exists, akin to the cosmic microwave background in electromagnetic radiation. Upper limits on its strength have been placed with PTA observations, but it is unlikely to be detected in the future [28, and referenced therein].

Chapter 2

Introduction to Pulsar Timing Arrays

> *"By the end of my PhD I could swing a sledgehammer."*
>
> ---
>
> Jocelyn Bell Burnell

The term Pulsar Timing Array (PTA) was coined by Foster and Backer in 1989 [29, 30]. Their idea was to combine timing observations from multiple pulsars to search for effects common in all of them: "perturbations in the Earth's orbit", and "a cosmic background of gravitational radiation". They also proposed to establish a long-term time standard from *pulsar timing*. The hunt for GWs with PTAs is still ongoing, but as sensitivity increases with each added observation, a PTA detection of GWs is expected in the near future [31, 32].

In this chapter, we first describe in detail how PTAs operate in Sec. 2.1, including pulsar timing observations (Sec. 2.1.1), current PTA collaborations (Sec. 2.1.2) and the response of a PTA to a GW (Sec. 2.1.3). Second, in Sec. 2.2, the idea of detecting super massive black hole (SMBH) binaries is argued thoroughly (this was previously mentioned in the discussion about the GW spectrum in Sec. 1.3.3). We move on to the necessary tools from data analysis needed for the rest of this thesis in Sec. 2.3. The last Sec. 2.3.3 lays the groundwork for so called null streams, which are the basis for the methodology developed and described in the rest of this work.

2.1 Pulsars as gravitational wave detectors

2.1.1 Pulsar timing

A pulsar is a rotating neutron star that sends out a beam of radiation from its magnetic poles, pointing to the Earth at some moment during its orbit. This "lighthouse effect" means radio telescopes are able to see them as a regular pulse of light coming from the pulsar's direction (hence the name). Observing these objects over extended periods allows the construction of a timing model (TM), a prediction of the arrival times of future pulses, based on the recorded arrivals of those in the past. Some pulsars, especially ones with millisecond periods, have very stable frequencies. The most regular ones can be timed with sub-microsecond precision [33], making them very precise clocks. These clocks, placed throughout the galaxy, can be used as a reference for any effects that may influence the pulse travel time to Earth – such as GWs.

Pulse folding

The previous statement that pulsars are observed through regular radio pulses oversimplifies the task somewhat. For most pulsars, the signal is not strong enough to observe individual pulses. It is found by "folding" the data; overlapping consecutive pulses so that the signals add up. To achieve effective folding, however, one needs to know how long each folded data segment should be so that it contains one pulse. In other words, one requires the pulsar period or frequency. To this end, a first guess is made with a Fourier transform of the unfolded data. The measurement is then refined in an iterative process where subsequent stretches of data are folded using an updated pulsar frequency at each step (sometimes the original data can be re-folded as well) [34].

Another complication to pulsar timing is that the pulses are not instantaneous, but show a varying luminosity over their short duration. This is called the *pulse profile* (as an example, the pulses of PSR B1919+21 can be seen in Fig. 2.1). Individual pulse profiles vary greatly, but when averaged over a large number of pulses, they tend to be very regular. So for each epoch of observation (usually of order one hour in length), a pulse profile is taken from the folded data. Modelling pulse profiles from repeated measurements allows the construction of a "template" profile. To pinpoint a time of arrival (TOA) for each pulse, a reference point in the template profile is chosen. After fitting the observed profile to the template, the shift with respect to this reference point is the recorded pulse TOA for that epoch [34].

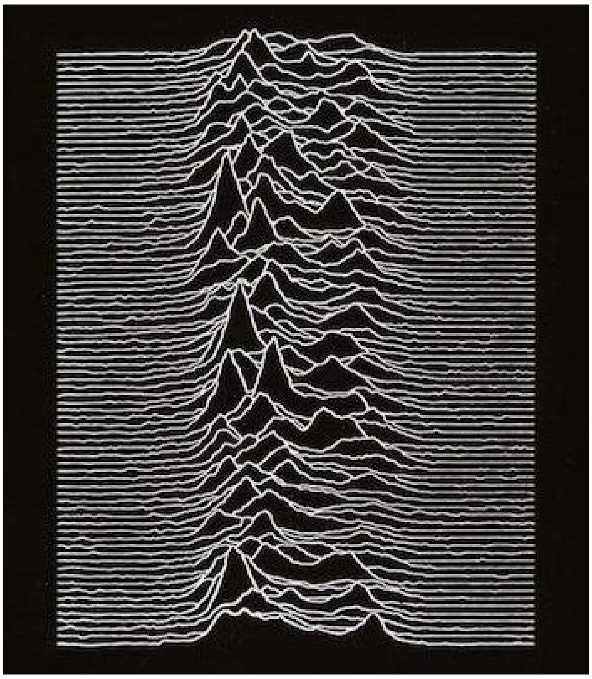

Figure 2.1: Artwork by Peter Saville showing the pulse profiles of PSR B1919+21, horizontally stacked as if "folded" (the imagine was adapted from a plot by Harold Craft, a radio astronomer working at the Arecibo Observatory). PSR B1919+21 is the first ever observed radio pulsar, discovered by Jocelyn Bell and Antony Hewish [8] in 1967. The artwork gained fame as the album cover of Unknown Pleasures by Joy Division, released in 1979.

Timing model

After obtaining a handful of TOAs, a timing model (TM) can be constructed that fits the observations. The discrepancy of the model is measured by the *residuals*, the differences between modelled and measured TOAs. A TM consisting of only the pulsar frequency is not enough to accurately model even the most regular millisecond pulsars. First of all, pulsars lose energy through radiation, and so their rotation slows down over time. After enough TOAs are accumulated, the spin-down can be included in the TM. It is actually the complete TM, not just the pulsar frequency, that is used in the iterative process of folding data, obtaining a TOAs, refitting the model, and so on.

There are plenty other contributions to the TOAs that can be included in the TM for long term observations. For example, the Earth's motion around the sun means the light travel time between the pulsar and Earth changes (as long as the pulsar is not directly above or below the ecliptic). Therefore, precise knowledge of the position of the solar system barycenter (SSB) with respect to the Earth (the ephemeris) is needed for an accurate TM. Current PTAs have actually become sensitive enough to measure corrections to the assumed ephemeris [35], as was one of the original ideas of Foster and Backer [29]. Analogous to the Earth's orbit, should the observed pulsar be part of a binary, its own orbit will affect the TOAs as well. Observing a pulsar in a binary system (or even a binary pulsar [36]) allows precision measurements of the orbital parameters (e.g. the period and companion mass). In the well-known study by Hulse & Taylor of such a system [37], the observed orbital evolution matched the predicted energy loss from GW emission precisely, making it the first indirect observation of GWs [38].

Free electrons between the pulsar and Earth, in the interstellar medium (ISM), can delay the arrival time of the photons. This effect scales with the inverse of the photon frequency squared, and can therefore be measured in multi-frequency observations. The difference in delay between high and low frequency photons is called the dispersion measure (DM), and can be used as a proxy for the free electron density along the line of sight to the pulsar. The DM is often time-dependent – either due to the Earth's movement changing the line of sight, or due to ISM particles moving in or out – and is modelled accordingly in the TM. Pulsar timing is so precise that the differences between time standards used to record TOAs have to be taken into account as well. Generally speaking, a hydrogen maser is used as a clock at the telescope, after which locally recorded times are converted to UTC (Coordinated Universal Time) with GPS and published time standard corrections. Even so, there can still be a residual clock error of a few nanoseconds that needs to be taken into account [34].

All TOAs will have some amount of added white noise from the radio telescope and back-end systems. However, often additional non-Gaussian noise is found which is not covered by the TM. This "timing noise" is not fully understood but may be (in part) intrinsic to the pulsar [39, 40]. In the end, pulsar timing is very sensitive to an incomplete or incorrect TM, as well as unexplained noise sources, so there will always be a finite accuracy to measured TOAs. Because a potential GW has such a small effect on the TOAs compared to the other ingredients described here, it is nigh impossible to measure it with a single pulsar. The key to finding GWs with pulsar timing is therefore that the same GW affects the TOAs of many pulsars. Hence PTAs are created to search for a correlated signal in multiple pulsars.

2.1.2 PTA collaborations

In principle, a PTA is just the combination of pulsar timing observations from multiple pulsars, with the aim of studying effects common in all pulsars. In practice, combining data involves multiple telescopes as well. A telescope's location determines which part of the sky it can access, and thus the set of pulsars it can observe. Several PTA collaborations exist, which, based on geographical location, carry out long term observations of different sets of pulsars (some pulsars are observed by more than one collaboration). These are the European Pulsar Timing Array (EPTA)[1], who use five radio telescopes across Europe[2] (Westerbork Synthesis Radio Telescope in the Netherlands, Effelsberg Radio Telescope in Germany, Lovell Telescope at Jodrell Bank Observatory in the UK, Nançay Radio Observatory in France, and Sardinia Radio Telescope in Italy); the North American Nanohertz Observatory for Gravitational Waves (NANOGrav)[3], who use the Green Bank telescope in the US and Arecibo in Puerto Rico; and the Parkes Pulsar Timing Array (PPTA)[4], who use the Parkes Observatory in Australia. These collaborations release their own combined data sets, the most recent (full) ones being EPTA data release (DR) 1 [42], the NANOGrav 11-year data set [43], and the PPTA DR1 [44] (extended in [45]). Furthermore, some pulsar timing data sets of one or a few pulsars have been published, usually in conjunction with a pulsar specific study (e.g. testing gravity with J1713+0747 [46], various studies on B1855+09 and B1937+21 [47], and high precision timing of four PPTA pulsars [48]).

---

[1] `www.epta.eu.org`

[2] Six, if we count LEAP (The Large European Array for Pulsars) separately. LEAP is a project in which all EPTA telescopes are used in conjunction to make one TOA, effectively creating a much larger telescope [41].

[3] `http://nanograv.org/`

[4] `www.atnf.csiro.au/research/pulsar/ppta/`

The aforementioned PTA collaborations form an umbrella consortium called the International Pulsar Timing Array (IPTA)[5]. With the aim of detecting GWs, a greater sensitivity is achieved by accumulating pulsar TOAs (for pulsars observed by more than one PTA), and by increasing the set of pulsars. Two combined data sets have been released so far as IPTA DR1 [11] and DR2 [33]. Combining pulsar timing data from different telescopes and from different collaborations is, however, not trivial. For example, the EPTA and PPTA collaborations produce one TOA per observing epoch, averaged over frequency, whereas NANOGrav produces a separate TOA for each frequency sub-band. This means that when modelling the white noise in the TOAs, a different model is needed for NANOGrav TOAs [33]. Another complication comes from combining data sets from different observing systems – either different telescopes or different back-end systems on the same telescope. Because each system has an unknown delay when processing data, possible time offsets between the data sets are fitted for in the TM of the combined data ("JUMPs") [33].

A relatively recent addition to (I)PTA potential is the Five-hundred-meter Aperture Spherical radio Telescope (FAST), located in Southwest China. This sensitive telescope is expected to play a major role in future PTA work, both by discovering new pulsars and by making high-precision observations of already known ones [49]. Furthermore, (I)PTA sensitivity will greatly increase with the future additions of the Square Kilometre Array (SKA) [50] monitoring the Southern hemisphere and the next-generation Very Large Array (ngVLA) [51] monitoring the Northern hemisphere.

2.1.3 PTA response to gravitational waves

Previously in Sec. 1.2.3, the h^+ and $h^\times$ GW polarisations were illustrated using a simplified PTA: a ring of pulsars around the Earth, with a GW travelling straight through it. We derived the induced Doppler shift in the pulsars' signals to be Eq. 1.53. The same derivation can be done for a pulsar placed anywhere in the sky, with unit vector $\hat{p}^i$ pointing to the pulsar, and a GW travelling in any direction $\hat{\Omega}$ (see e.g. [52]). The resulting Doppler shift is:

$$z(t,\hat{\Omega}) = \frac{1}{2}\frac{\hat{p}^i\hat{p}^j}{1+\hat{\Omega}\cdot\hat{p}}\Delta h_{ij}. \tag{2.1}$$

Here, $\Delta h^{ij} = h^{ij}_{\mathrm{SSB}} - h^{ij}_{\mathrm{P}}$ is still the difference between the GW strain at the SSB and at the pulsar (the time difference between the actual TOA measurement on Earth and the SSB is included in the TM). Now, h^{ij} is any combination of the two polarisation states.

---

[5] `www.ipta4gw.org/`

As a simplification, sometimes the *pulsar term* $h^{ij}_{\mathrm{P}} = h^{ij}(t_{\mathrm{P}}, \vec{x}_{\mathrm{P}})$ is neglected and only the *Earth term*, h^{ij}_{SSB} , is considered in the signal model. If the GW signal comes from a (slowly) inspiralling BBH, its frequency will have increased in the ~1000 years it takes for the pulsar's light to travel to the Earth (see Eq. 1.80). As a result, the pulsar and Earth terms have different frequencies. Whether this difference is large enough to distinguish the terms in observations depends on the pulsar distance and angle to the GW line of sight, and on the evolutionary stage of the binary. Rosado *et al.* [22] found this to be the case in about 50% of expected detections[6]. Either way, because the pulsar distances vary, only the Earth terms are the same for all pulsars. This makes them more useful for PTA analysis, and the pulsar terms are often considered noise. In this thesis, we make the simplifying assumption that the pulsar terms can be ignored, and write h^{ij} as a shorthand for $h^{ij}_{\mathrm{SSB}} = h^{ij}(t_{\mathrm{SSB}}, \vec{x}_{\mathrm{SSB}})$, which now replaces Δh^{ij} in Eq. 2.1.

It is convenient to write out the metric perturbation in terms of two polarisation tensors. First we choose a coordinate system (the same as in [52]), where the first basis vector is the GW propagation direction $\hat{\Omega}$, and the two orthogonal vectors are named $\hat{m}$ and $\hat{n}$. Given two sky angles $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$, these can be expressed as follows:

$$\begin{aligned}
\hat{\Omega} &= \big(\sin\theta\cos\phi,\ \ \sin\theta\sin\phi,\ \ \cos\theta\big) \\
\hat{m} &= \big(\sin\phi,\ \ -\cos\phi,\ \ 0\big) \\
\hat{n} &= \big(\cos\theta\cos\phi,\ \ \cos\theta\sin\phi,\ \ -\sin\theta\big).
\end{aligned} \tag{2.2}$$

Then, two polarisation tensors can be constructed from $\hat{m}$ and $\hat{n}$:

$$\begin{aligned}
e^{+}_{ij}(\hat{\Omega}) &= \hat{m}_i\hat{m}_j - \hat{n}_i\hat{n}_j \\
e^{\times}_{ij}(\hat{\Omega}) &= \hat{m}_i\hat{n}_j + \hat{n}_i\hat{m}_j.
\end{aligned} \tag{2.3}$$

These form a basis for any GWs travelling along $\hat{\Omega}$ in the transverse-traceless gauge. The metric perturbation in terms of the two polarisation amplitudes, $h^{+}(t)$ and $h^{\times}(t)$, is:

$$h_{ij}(t) = h^{+}(t)e^{+}_{ij} + h^{\times}(t)e^{\times}_{ij}. \tag{2.4}$$

Now Eq. 2.1 – neglecting the pulsar term – can also be written as:

$$z(t, \hat{\Omega}) = \frac{1}{2}\frac{\hat{p}^i\hat{p}^j}{1+\hat{\Omega}\cdot\hat{p}}(e^{+}_{ij}h^{+} + e^{\times}_{ij}h^{\times}) \equiv F^{+}(\hat{\Omega})h^{+} + F^{\times}(\hat{\Omega})h^{\times}, \tag{2.5}$$

---

[6] For the stochastic background (SB), a signal at all frequencies (see Sec. 2.2.3), there are always pulsar terms present; the pulsar term frequency of one binary will overlap with the Earth term frequency of another.

where we have defined the response functions F^+ and $F^\times$.

In pulsar timing, the Doppler shift can not be measured directly in the pulsar TOAs. Rather, a shift in TOAs with respect to the TM builds up over time in the residuals. So the actual observable due to the GW signal is the response to the time-integrated h^+ and $h^\times$. If we define $H^+(t) = \int_0^t \mathrm{d}t' h^+(t')$ and analogously $H^\times(t)$, the response in the residuals is:

$$r(t, \hat{\Omega}) = F^+(\hat{\Omega})H^+ + F^\times(\hat{\Omega})H^\times. \tag{2.6}$$

Between the different pulsars in the PTA, the response functions change, but the H^+ and $H^\times$ are held in common.

The pulsar response in Eq. 2.6 only depends on the angle between the pulsar and GW source locations (via the dot product $\hat{\Omega}\cdot\hat{p}$), and not on their absolute locations. It is minimised when $\hat{\Omega}\cdot\hat{p} = 1$, i.e. when the pulsar stands directly opposite the GW source in the sky. The smaller the angle between pulsar and source, the stronger the GW response. The expression would become problematic when $\hat{\Omega} \cdot \hat{p} = -1$, but this would require the pulsar and GW source to be at the exact same sky location, which is only a mathematical possibility and not a physical one. When $\hat{\Omega} \cdot \hat{p}$ gets close to -1, both the numerator and denominator of the response function become very small, so it is well defined in the limit.

2.2 Sensitivity to super massive black hole binaries

In our discussion of the GW spectrum in Sec. 1.3.3, it was said that PTAs are sensitive to the most massive BBHs, in the form of individual sources or a stochastic background. In this section, we justify this statement using the PTA response as derived in the previous section, and the properties of current PTAs.

2.2.1 Estimating the PTA sensitivity

PTA frequency band

For a GW to be measurable in the PTA residuals, it needs to fall within the PTA frequency band, which in Sec. 1.3.3 was said to be around $3\times10^{-9} - 10^{-6}$Hz. The minimum frequency is determined by the time span of observations T, as any signal that fluctuates on a longer time scale is not noticeable. The longest observed pulsar in the IPTA DR2 is J1713+0747 with 22.5 years, which gives a bound of $f_{\mathrm{min}} = \frac{1}{T} \approx 1.4 \times 10^{-9}$Hz. Because an increased sensitivity is obtained by using multiple pulsars, a more practical lower bound will be somewhat higher (the minimum of $\sim 3 \times 10^{-9}$Hz above is for $T \approx 10$ years).

PTA data is not regularly sampled, as each TOA is determined from observations taken a few days to weeks apart. This irregular sampling means the usual Nyquist frequency does not apply; the maximum frequency due to the time in between samples is effectively infinite (see, for example, [53]). The maximum frequency is however bounded by the integration time ΔT of the observations. Because each TOA is determined from a stretch of data of about 1 hour long, signal fluctuations on shorter time scales can not be measured. This gives us an upper bound of $f_{\mathrm{min}} = \frac{1}{\Delta T} \approx 2.8 \times 10^{-4}$Hz. As will become clear later on in this discussion, PTA sensitivity decreases with increasing frequency, so that the upper bound is effectively lower, usually taken to be $\sim 10^{-6}$Hz. Because the sensitivity is highest at the lowest frequency, the exact upper bound is not very relevant to the question at hand.

Signal from a single SMBH binary

A GW signal – assuming it falls within the PTA frequency band – can only be detected if its response in the TOA residuals is distinguishable from the noise. Therefore, the signal strength required is directly related to the precision with which the TOAs can be measured. Although pulsar timing noise is complicated, the overall noise level is summarised in the root mean square (rms) of the residuals, r_{noise}, which can be compared directly to the GW response r_{GW}. The smallest noise rms reported in the IPTA DR2 is about 110 ns, for pulsar J0437-4715 [33].

The amplitude of r_{GW} is given by Eq. 2.6. The response functions F^{+} and $F^{\times}$ are geometrical factors of order unity, which we will neglect for the following order of magnitude argument. The same holds for the polarisation state mixing (Eq. 1.74), so that $r_{\mathrm{GW}} \sim \bar{H}^{+/\times}$. For the GW model, consider the case of a slowly evolving SMBH binary, which is approximately monochromatic on the PTA observation time scale of $\mathcal{O}(10)$ years [20]. The GW polarisations are then given by Eq. 1.75, so the time-integrated $\bar{H}^{+}$ and $\bar{H}^{\times}$ are of the form:

$$\bar{H}^{+}(t) = \frac{A}{\omega}\frac{1}{2}\left(1 + \cos(\iota)^2\right)\sin(\omega t + \phi_0) \qquad \bar{H}^{\times}(t) = -\frac{A}{\omega}\cos\iota\cos(\omega t + \phi_0). \tag{2.7}$$

We consider the amplitude A in the frame of the observer:

$$A \equiv \frac{4}{D_{\mathrm{l}}}\left(\frac{\mathrm{G}\mathcal{M}_{\mathrm{c}}}{\mathrm{c}^2}\right)^{5/3}\left(\frac{\pi f}{\mathrm{c}}\right)^{2/3}. \tag{2.8}$$

Hence, f is the observed GW frequency, $\mathcal{M}_{\mathrm{c}}$ is the redshifted chirp mass (equal to $(1+z)\mathcal{M}_{\mathrm{c}}^{(\mathrm{source})}$), and the relevant distance is the luminosity distance D_{l}. However, for this estimate, we do not take into account the relatively minor effects of the redshift $z \lesssim 1$ on the other quantities. Additionally, we have absorbed the factor A into the polarisation states from Eq. 1.75 (rather than leaving it as an overall factor as in Eq. 1.73). At their maximum (for a face-on source), the residuals are $r_{\mathrm{GW}} \sim \frac{A}{\omega} \propto f^{-1/3}$, which justifies the earlier statement that PTA sensitivity is highest at the lowest frequencies.

The GW amplitude depends on the source parameters: its distance r, chirp mass $\mathcal{M}_{\mathrm{c}}$, and GW frequency $f = \omega/2\pi$. To maximise sensitivity, we can consider f to be the lower bound of the PTA band, $f \approx 3 \times 10^{-9}$Hz. If we require a certain amplitude A, there is a degeneracy between the distance and mass of the source, meaning that a binary could be detected as long as it is close enough, or equally, as long as it is massive enough. Because a small distance severely constrains the observable volume (which scales as r^3), for now we set out to find sources up to $z \approx 1$, which corresponds roughly to a luminosity distance of 6.7Gpc (assuming a standard, flat ΛCDM universe [54]). One can then ask what the corresponding minimum chirp mass is needed for large enough r_{GW}. For convenience, we express the chirp mass in multiples of $10^9\mathrm{M}_\odot$, so that the residuals scale as:

$$\begin{aligned} r_{\mathrm{GW}} &\approx \frac{A}{\omega} = \frac{4}{r}\left(\frac{\mathrm{G}\mathcal{M}_{\mathrm{c}}}{\mathrm{c}^2}\right)^{5/3}\left(\frac{\pi}{\mathrm{c}}\right)^{2/3} f^{-1/3}\frac{1}{2\pi} \\ &\approx 2.0\mathrm{ns}\left(\frac{r}{6.7\mathrm{Gpc}}\right)^{-1}\left(\frac{\mathcal{M}_{\mathrm{c}}}{10^9\mathrm{M}_\odot}\right)^{5/3}\left(\frac{f}{3\times 10^{-9}\mathrm{Hz}}\right)^{-1/3}. \end{aligned} \tag{2.9}$$

Therefore, to get $r_{\mathrm{GW}} \geq r_{\mathrm{noise}}$, we would need $\mathcal{M}_{\mathrm{c}} \geq 1.2 \times 10^{10}\mathrm{M}_\odot$.

We have not yet taken into account that the GW can be measured with multiple pulsars. The signal-to-noise ratio (S/N, ρ in Eqs.) of a single pulsar can be approximated (using the overall noise level) as $\rho_i = \frac{r_{\mathrm{GW}}}{r_{\mathrm{noise}}}$, such that the S/N of all pulsars is given by $\rho^2 = \sum_i \rho_i^2$. As an estimate, assume that each pulsar contributes the same S/N, such that the total is $\rho = \frac{r_{\mathrm{GW}}}{r_{\mathrm{noise}}} \times \sqrt{\mathrm{P}}$, where P is the number of pulsars. For a sinusoidal signal, the S/N further scales with the number of cycles in the data, which is given by $N_{\mathrm{cycles}} = T \times f$. As we have chosen our frequency to be the miminum, $f_{\mathrm{min}} = \frac{1}{T}$, there would be exactly one cycle in the data. It shows, however, that PTA sensitivity increases over time, even without improving TOA precision or adding more pulsars to the array[7].

---

[7] It also means the previously found scaling of $r_{\mathrm{GW}} \propto f^{-1/3}$ does not hold for the S/N. However, the most massive sources spend little time at higher frequencies (see T_{PTA} in the next Sec. 2.2.1), so PTA are more likely to detect an individual SMBH binary at a low frequency (see e.g. [22]).

Using P = 65 from IPTA DR2, we find:

$$\begin{aligned}\rho &= \frac{r_{\rm GW}}{r_{\rm noise}}\sqrt{\rm P}\\ &\approx 0.14\left(\frac{r_{\rm noise}}{110{\rm ns}}\right)^{-1}\left(\frac{\rm P}{65}\right)^{1/2}\left(\frac{r}{6.7{\rm Gpc}}\right)^{-1}\left(\frac{\mathcal{M}_{\rm c}}{10^9{\rm M}_\odot}\right)^{5/3}\left(\frac{f}{3\times10^{-9}{\rm Hz}}\right)^{-1/3}.\end{aligned} \tag{2.10}$$

It is not a trivial matter to determined the proper S/N threshold that would constitute a detection. For this order of magnitude estimate, we choose to consider $\mathrm{S/N} \gtrsim 1$, keeping in mind that this is a very low threshold. Using the parameters above, to reach $\mathrm{S/N} \gtrsim 1$, we require $\mathcal{M}_{\rm c} \gtrsim 3\times10^9{\rm M}_\odot$. Because we used the lowest reported noise level from the IPTA DR2 for all pulsars, this estimate is a bit too optimistic. On the other hand, looking to the future of IPTA with the additions of FAST, SKA and ngVLA (see Sec. 2.1.2), it is not unrealistic to expect noise levels to go down as well as the number of pulsars to go up.

SMBH binary population

Eq. 2.10 has established that a PTA has the potential to detect a binary with $\mathcal{M}_{\rm c} \gtrsim 3\times10^9{\rm M}_\odot$ within $z \approx 1$. The question remains whether these sources are astrophysically feasible, and if so, how many there are at any given time. It is well established that the most massive black holes reside in the centers of galaxies, and that the comic history is full of galaxies merging. The expectation is that the two black holes, after the merger of their galactic hosts, both fall to the center of the gravitational potential of the newly formed galaxy and form a binary. If the binary is close enough for GW emission to dominate their orbital dynamics, they will inspiral and emit the GW signal from Eq. 1.73 (e.g. [55]). This last assumption (for which the binary needs to get well below a separation of around 1 pc), is the most uncertain part of the scenario. It is termed "the last parsec problem" and, potentially, PTA observations could resolve it. It will be discussed in more detail in Sec. 2.2.2, but for now we will assume the binaries do indeed get close enough to enter the GW-dominated inspiral regime.

There is known to be a connection between the mass of a SMBH and its host galaxy, from observational evidence. More massive black holes reside in more massive galaxies, although the exact relations reported in the literature vary significantly [e.g. 56, 57, 58]. Most often, these relations take the form:

$$\log_{10}\left(\frac{M}{{\rm M}_\odot}\right) = \alpha + \beta\log_{10}\left(\frac{M_{\rm b}}{10^{11}{\rm M}_\odot}\right), \tag{2.11}$$

where $M_{\rm b}$ is the mass of the bulge of the host galaxy, and M is the mass of the black hole. Typical values for the parameters are $\alpha \sim 8.1$ and $\beta \sim 1.2$, with a scatter on the relation of about 0.3 dex (see, for example, [10], who have made a compilation of 14 M–M_b relations from the literature, which is also shown in Fig. 4.2). For a given total binary mass, the chirp mass from Eq. 1.79 is maximised for an equal mass binary ($m_1 = m_2$). Assuming this scenario, we can ignore the conversion factor ~1 between M and $\mathcal{M}_\mathrm{c}$, as well as the factor 2 between single black hole mass and total binary mass. Then, substituting the previous constraint ($M \sim \mathcal{M}_\mathrm{c} \gtrsim 3 \times 10^9 \mathrm{M}_\odot$) into Eq. 2.11 using the typical values for α and β gives $M_\mathrm{b} \gtrsim 6 \times 10^{11} \mathrm{M}_\odot$. The most massive galaxies are almost all ellipticals [e.g. 59], for which the total (stellar) mass is of the same order of magnitude as the bulge mass, so we will not distinguish between the two for this argument.

From observational data, the number density of massive elliptical galaxies with $M \sim 6 \times 10^{11} \mathrm{M}_\odot$ is about $\rho_\mathrm{ME} \sim 10^{-4} \mathrm{Mpc}^{-3}$ (and falls off steeply at higher masses)[e.g. Fig. 6 in 60]. In the comoving volume of the Universe within $z = 1$, there are therefore about

$$N = \rho_\mathrm{ME}\, V \approx 10^{-4} \mathrm{Mpc}^{-3}\, 1.6 \times 10^{11} \mathrm{Mpc}^3 = 1.6 \times 10^7 \tag{2.12}$$

galaxies that could potentially host a massive enough black hole. However, to host a binary, the galaxy would have to have undergone a recent enough merger. Because we are assuming roughly equal mass binaries, the merger needs to be between roughly equal mass galaxies as well, i.e. a major merger. Using data from the CANDLES survey, [61] find the major merger rate of massive ($M > 2 \times 10^{10} \mathrm{M}_\odot$) galaxies evolves from ~$0.15 \mathrm{Gyr}^{-1}$ (per galaxy) at $z = 1$ to ~$0.03 \mathrm{Gyr}^{-1}$ at $z = 0$ (with a significant spread in results between different observational studies and theoretical predictions). For this estimate, the average merger rate can be taken to be $R \sim 0.1 \mathrm{Gyr}^{-1}$ per galaxy (meaning a massive galaxy experiences, on average, ~1 major merger since $z = 1$, 7.7Gyr ago).

For the binary to be observable *now*, the merger needs to have happened recently enough such that the binary is still inspiralling when observations are taken. This window, T_PTA, is the time a binary spends in the PTA band[8]. It can be computed with Eq. 1.80, which relates the time to coalescence τ to the GW frequency f. Rewriting this and differentiating to get $\frac{\mathrm{d}t}{\mathrm{d}f} = -\frac{\mathrm{d}\tau}{\mathrm{d}f}$, allows to compute the time in band as an integral over $\mathrm{d}f$, from $f_\mathrm{min} = 3 \times 10^{-9}$Hz to f_max~10^{-6}Hz (the PTA band as previously determined). Because the binary speeds up during inspiral, the time spend near the maximum frequency is negligible, and the upper bound of the integral can be replaced with

---

[8] For the fraction of mergers observable now, it does not actually matter how long after merger it takes for the binary to *enter* the PTA band, as long as it is significantly below a Hubble time.

infinity. The result, filling in the previously found minimum chirp mass, is:

$$T_{\mathrm{PTA}} = 1.8 \times 10^5 \mathrm{yr} \left(\frac{\mathcal{M}_{\mathrm{c}}}{3 \times 10^9 \mathrm{M}_\odot} \right)^{-5/3} \left(\frac{f_{\min}}{3 \times 10^{-9} \mathrm{Hz}} \right)^{-8/3} . \tag{2.13}$$

Expected number of sources

Taking all of this together, the potential number of PTA sources within $z = 1$ is

$$N \times R \times T_{\mathrm{PTA}} \approx 1.6 \times 10^7 \cdot 1 \times 10^{-10} \mathrm{yr}^{-1} \cdot 1.8 \times 10^5 \mathrm{yr} \approx 300. \tag{2.14}$$

This is quite a large number, but not all of these sources will be resolvable as we used $\mathrm{S/N} > 1$ as a threshold. Instead, the GWs from these sources add up, and form the stochastic background (SB) (more in Sec. 2.2.3). If one of these binaries is particularly close, say within 1Gpc, it may be individually detectable (this effectively increases the S/N threshold to $\sim$7, given the scaling in Eq. 2.10). Scaling the expected number with the fractional volume, we have:

$$N \times \frac{V_{<1\mathrm{Gpc}}}{V_{<6.7\mathrm{Gpc}}} \approx 300 \left(\frac{1}{6.7} \right)^3 \approx 1, \tag{2.15}$$

indicating that there is a very real possibility of observing a resolvable binary with PTA, especially as pulsar timing precision goes up (keeping in mind that we used the lowest noise rms in the IPTA DR2 as the standard.)

Studies on the topic, such as [62] and [22], take into account many models for galaxy evolution and $M - M_{\mathrm{b}}$ relations, and use full distributions rather than ballpark numbers in their computations. Overall, they find it likely the SB will be detected first, but there is a significant probability for detection of one to a few individually resolvable binaries as well. From the predictions in Fig. 6 of [22], if a single binary is detected with current IPTA, it is most likely to be around $z \approx 0.3$, with $\mathcal{M}_{\mathrm{c}} \approx 5 \times 10^9 \mathrm{M}_\odot$, and a frequency near the lower bound of the IPTA band (which depends on the total observation time); so similar to the parameters we used in this estimate. With the additional sensitivity of SKA, a detected binary is more likely to have a somewhat lower mass, which makes sense as those sources are more abundant. Given that the most massive sources lie in the tail of the SMBH mass distribution, whether (or how soon) a detection will be made depends strongly on how distant the few heaviest binaries around us happen to be [9].

---

[9] It also depends on the sky position of these sources relative to the PTA pulsars, which we would find if we were taking the geometrical response functions from Eq. 2.5 into account.

2.2.2 The last parsec problem

It is still an open question whether SMBHs, after their host galaxies merge, form a close enough binary to be observed by PTA. Initially, when two massive galaxies meet, their cores merge under dynamical friction. The two SMBHs, contained in the galaxy cores, are also affected by dynamical friction of the dense stellar environment of the newly merged core, and gravitate towards the centre. Once they get close enough, they become gravitationally bound and form a binary [63] (at a separation of many parsecs [64]). Even in a wide orbit, the binary will emit GWs, but from Eq. 1.83 it is clear that the frequency would be much too low to observe with PTA (and so would the amplitude). In order for the binary to shrink, it needs to lose energy and angular momentum. GW radiation does carry these away, but only does so efficiently at much smaller separations, below 1 parsec (e.g. [28]). One mechanism that can shrink the binary is the interaction with stars orbiting near the center of the galaxy. They take on kinetic energy from the binary and are shot away (a gravitational slingshot) [65]. As these stars are depleted from the central region of the galaxy however, this process may be insufficient to bring the binary into the GW dominated regime. This issue is known as the "last parsec problem" [28] or "final parsec problem" [65].

A possible solution to the last parsec problem is to relax the assumption that the central regions of the merging galaxies are spherical. This triaxiality means there continue to be stellar orbits that can interact with the SMBH binary and shrink it down to the GW-dominated stage [e.g. 66]. Another solution could be to introduce a separate mechanism that can further shrink the binary. For example, gas in the galactic centre could work similar to stars and carry away energy and angular momentum [63]. However, [67] recently found that gas accreting onto the binary actually expands it, rather than shrinking it. Even if gas around the black hole could shrink the binary, it remains an issue that there would need to be just the right amount of gas to shrink the binary, but not so much that GWs have no time to take over [28].

It is possible that all – or at least a significant fraction of – SMBH binaries "stall", i.e. do not merge within a Hubble time. What fraction this is directly affects the strength of the SB and the probability of having a massive and close enough binary to observe by itself. However, even in the most pessimistic scenario where all binaries stall, [68] found that there will still be a SB signal due to unequal mass binaries inspiralling from larger separations. This SB will have a significantly reduced amplitude, but will eventually be observable by PTA as observation time increases and additions such as SKA boost sensitivity. Moreover, according to simulations by [69] and [70], subsequent galaxy mergers will turn some of the stalling binaries into triples, whose dynamical interaction causes them to merge. The resulting SB would be a factor 2-3 dimmer than one based on the assumption that all binaries merge by themselves, which is still observable with (future) PTA [69].

2.2.3 Stochastic background from unresolved binaries

As previously mentioned, the many signals from SMBH binaries that are not individually resolvable add to form a stochastic, sky-wide signal, commonly known as the stochastic background (SB) (or Gravitational Wave Background, GWB). Assuming all binaries are circular (which is fairly reasonable, given that the galactic environment tends to circularise the binary e.g. [65]) and only evolve due to GW emission, the spectrum of the SB is $h \propto f^{-2/3}$ [71]. Here, h is the "characteristic strain", or the amplitude of the GW averaged over all sources. Relaxing these assumptions changes the spectrum slightly (see [28] and references therein). It is usually assumed there are enough sources contributing to the SB so that it is isotropic. However, there may be anisotropies at the high frequency end of the PTA band, as those sources are fainter and thus there are effectively fewer that contribute to the overall signal [72].

Hellings & Downs correlations

The SB is a stochastic process with a red spectrum, as are several types of pulsar timing noise, such as clock errors or effects related to the SSB position (see Sec. 2.1.1). Therefore, to identify the SB, it is crucial to employ the specific PTA response (Eq. 2.6) due to the quadrupolar nature of GWs, which is not shared with other stochastic processes in the residuals. A resolvable GW source can be found by modelling the response directly (or by using the response to construct null streams, but this is a topic for later on in Sec. 2.3.3). Contrarily for the SB, the exact h^+ and $h^\times$ are unknown, but it can be modelled through the correlation of the signal between pulsars. This correlation, C_{ij}, is the expectation value of $r_i \times r_j$ (the product of the residuals from pulsar i and pulsar j). It was calculated by Hellings & Downs [9] to be:

$$C_{ij} = \frac{1-\cos\gamma_{ij}}{2} \ln\left(\frac{1-\cos\gamma_{ij}}{2}\right) - \frac{1}{6}\,\frac{1-\cos\gamma_{ij}}{2} + \frac{1}{3}. \tag{2.16}$$

Because the response functions from Eq. 2.6 only depend on the angle between the pulsar and GW propagation direction, this correlation only depends on the angle γ_{ij} between the pulsar sky positions.

Eq. 2.16 is visualised in Fig. 2.2, commonly known as the *Hellings & Downs curve*. The correlation is maximised when the pulsars are closest to each other. Noticeable is that there are two zero points in the correlation; for two pulsars ~49 deg or ~121 deg apart. This curve is thought to be the "smoking gun" for PTA GW detection, because it displays the unique quadrupolar correlation for GWs, distinguishing it from correlated noise (such as noise due to clock or ephemeris errors which are monopolar and dipolar in nature, respectively [73, 74]). The shape of the curve is better measured with many PTA pulsar pairs, therefore a PTA with pulsars separated at a range of angles works best (see e.g. [28]).

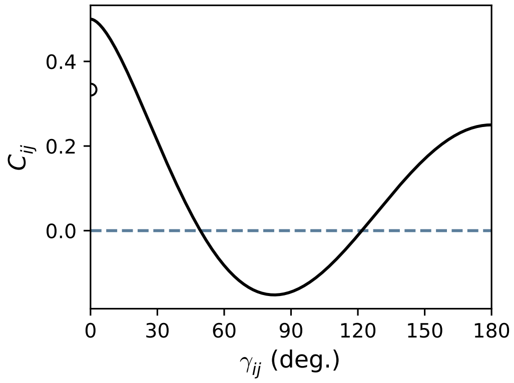

Figure 2.2: The Hellings and Downs [9] curve. The correlation between the residuals of pulsar i and pulsar j due to a GW, as a function of the angle γ_{ij} between the pulsars' sky locations.

2.2.4 Current upper limits on GWs from SMBH binaries

So far, PTAs have not detected GWs yet. Nevertheless, because both resolvable SMBH binaries and the SB are continuous sources, absence of a detection and knowledge of the detectors' sensitivity can be used to place upper limits on the signals' strength. Because PTA pulsars are not evenly distributed across the sky, and because the PTA response depends on the pulsars' locations, the sensitivity to resolvable sources varies with sky location. Upper limits are usually given as a sky average amplitude of the GW, at the most sensitive frequency. Some recent ones are given in Table 2.1 (including one based on data from a single pulsar). To facilitate a comparison with the expected amplitude in the residuals from Eq. 2.9, the last column shows $A/\omega = A/(2\pi f)$.

These limits only exclude the most optimistic scenarios in which there happens to exist a very massive and nearby SMBH binary (for example, [75] excludes any binaries with $f \lesssim 70\,\mathrm{nHz}$ and $\mathcal{M}_\mathrm{c} > 10^{10}\mathrm{M}_\odot$ within 1 Gpc), but are are still compatible with astrophysical models [75], [77], [78].

PTA (+ DR)	reference + (year)	$A(f)$	f (nHz)	$\frac{A}{2\pi f}$ (ns)
EPTA DR1	[75] (2016)	9×10^{-15}	6	∼240
PPTA	[76] (2014)	1.7×10^{-14}	10	∼270
NANOGrav (11 yr)	[77] (2019)	7.3×10^{-15}	8	∼145
single pulsar J1713+0747	[78] (2018)	1.4×10^{-14}	20	∼110

Table 2.1: Recent upper limits on resolvable SMBH binaries in PTA data.

Because these limits are getting near the predictions from astrophysical models and PTA sensitivity increases over time, future PTA observations can inform these models either through detections or through more constraining limits.

For the SB, upper limits are usually expressed as the characteristic strain (average GW amplitude) at a reference frequency of 1/yr $\approx$ 32nHz. This is assuming the GW follows the standard power spectrum of $h \propto f^{-2/3}$ [71]. (The actual sensitivity at 1/yr is actually very bad, because this is the frequency at which the Earth's motion affects the residuals. Any signal at exactly this frequency would be absorbed in the TM fit.) Recently reported upper limits are given in Table 2.2. As an equivalent to the computed residuals amplitude for the resolvable sources, here we give $h/(2\pi f)$ at $f = 10$nHz, where $h \propto f^{-2/3}$ is used to scale h. Note that this number is less applicable here, because the SB is not a monochromatic signal, so if a signal were present, its power would be spread across the spectrum.

PTA (+ DR)	reference + (year)	$A(f = \mathrm{yr}^{-1})$	$\frac{A(f=10\mathrm{nHz})}{2\pi \times 10\mathrm{nHz}}$ (ns)
EPTA DR1	[79] (2015)	3×10^{-15}	∼100
PPTA	[48] (2015)	1×10^{-15}	∼35
NANOGrav (11 yr)	[80] (2018)	1.45×10^{-15}	∼50
IPTA (DR1)	[11] (2016)	1.7×10^{-15}	∼60

Table 2.2: Recent upper limits on the stochastic background from SMBH binaries in PTA data.

The upper limit from PPTA [48] is noticeably smaller than subsequent limits from NANOGrav [80] and IPTA [11]. It was discovered in this NANOGrav analysis that PTA data is sensitive to small errors in the ephemeris (the position of the SSB). The inclusion of some degrees of freedom of the ephemeris in the fit for a SB signal (a method called BayesEphem [35]) affects the upper limit somewhat, which can at least partially explain the difference with the PPTA upper limit. The upper limit from IPTA presented here is from DR1, which includes NANOGrav data up to their 5-year DR [81]. It is therefore not surprising that the IPTA upper limit is higher than the more recent one from NANOGrav alone. Because it takes time to combine data, the IPTA DR2 [33] includes the NANOGrav 9-year DR [82] (as well as EPTA DR1 [42], the extended PPTA DR1 [45] and additional PPTA data from four pulsars reported in [48]). Results from searching for the SB in IPTA DR2 are being worked on but, at the time of writing, have not been published yet [83]. As with the resolvable source upper limits, current limits are compatible with all astrophysical models, but are beginning to constrain the most optimistic scenarios and will increasingly do so, should no detection occur as PTA sensitivity increases [10].

2.3 PTA data analysis

In the search for resolvable binaries or the SB in PTA data, both Frequentist and Bayesian methods are employed. Either approach has its advantages and, in general, PTA science benefits from having multiple analysis methods available (especially to improve confidence in a future detection). For the work presented in this thesis, we have opted for a Bayesian framework. The following Sec. 2.3.1 provides an overview of Bayesian statistics. The Gaussian likelihood, which is often employed in both Bayesian and Frequentist methods, is discussed in Sec. 2.3.2. Lastly, we close the introduction by beginning the discussion on null streams, providing an outlook on the works presented in the next three chapters.

2.3.1 Bayesian statistics

The cornerstone of Bayesian statistics is the following formula (Bayes' theorem; the underlying ideas were developed by Thomas Bayes in the 18th century, the formula was developed later by Pierre-Simon Laplace):

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B)} \tag{2.17}$$

In this equation, A and B are both events that occur with probabilities $\mathrm{P}(A)$ and $\mathrm{P}(B)$, respectively. The other probabilities are conditional, $\mathrm{P}(A|B)$ is the probability that event A occurs given that event B does.

The formula can be interpreted as a method to update one's prior belief about an event or statement (A), after learning new information (B). In this context, $\mathrm{P}(A)$ is the *prior probability* of A, in other words, the belief we have that A is true before we learn anything. B is the new information we obtain, for example the data from an experiment. It has an associated *evidence* $\mathrm{P}(B)$, which functions as a normalisation for the formula. $\mathrm{P}(A|B)$ is what we want to obtain, our updated belief about A after learning that B is true, or the *posterior probability* of A. The crux of Bayes' theorem is that we can obtain this by instead computing $\mathrm{P}(B|A)$, the *likelihood*.

In data analysis, usually we assume some model $\mathcal{M}$ which has one or more parameters $\vec{\theta}$ that we are trying to measure. Given the data d, Bayes' theorem can be adapted using probability densities as follows:

$$\mathrm{p}(\vec{\theta}|d,\mathcal{M}) = \frac{\mathrm{p}(d|\vec{\theta},\mathcal{M})\mathrm{p}(\vec{\theta}|\mathcal{M})}{\mathrm{P}(d|\mathcal{M})}. \tag{2.18}$$

The $\mathcal{M}$ after the conditional bar in each probability signifies that model $\mathcal{M}$ is assumed throughout. If the values of the parameters $\vec{\theta}$ are completely unknown in advance, its priors, $\mathrm{p}(\vec{\theta}|\mathcal{M})$, should be chosen such that they influence the result as little as possible. This usually means they are uniform ("flat") probability distributions within the possible range of the parameters. For some parameters, it makes sense to choose a non-uniform distribution; for example, when measuring the sky location of a GW source given by a polar angle $0 < \theta < \pi$ and azimuthal angle $0 < \phi < 2\pi$, one needs to use a distribution uniform in $\cos(\theta)$ (so not uniform in θ), to make the prior flat across the sky. The evidence, or marginalised likelihood, can be computed by integrating the likelihood over the parameters:

$$\mathrm{P}(d|\mathcal{M}) = \int \mathrm{d}\vec{\theta}\,\mathrm{p}(d|\vec{\theta},\mathcal{M})\,\mathrm{p}(\vec{\theta}|\mathcal{M}). \tag{2.19}$$

If one is only interested in measuring the parameters $\vec{\theta}$ (parameter estimation), the evidence is simply a normalisation for the rhs of Eq. 2.18. In that case, one can limit the computation to the numerator of Eq. 2.18 and numerically normalise the resulting posterior distributions on $\vec{\theta}$. However, the evidence also be used to compare different models, say $\mathcal{M}_1$ and $\mathcal{M}_2$, in the following way:

$$\frac{\mathrm{P}(\mathcal{M}_1|d)}{\mathrm{P}(\mathcal{M}_2|d)} = \frac{\mathrm{P}(d|\mathcal{M}_1)\mathrm{P}(\mathcal{M}_1)}{\mathrm{P}(d|\mathcal{M}_2)\mathrm{P}(\mathcal{M}_2)}, \tag{2.20}$$

where we have applied Eq. 2.17 to both the numerator and denominator in the *odds ratio* above. It reflects the posterior believe in model $\mathcal{M}_1$ relative to model $\mathcal{M}_2$. This ratio includes the priors on both models, which are often left out of the computation because there is no reason to favour one over the other. In that case, what is computed is the *Bayes factor*:

$$\frac{\mathrm{P}(d|\mathcal{M}_1)}{\mathrm{P}(d|\mathcal{M}_2)}. \tag{2.21}$$

2.3.2 Gaussian likelihood

Because the priors are chosen and the evidence can be computed from the likelihood or as a numerical normalisation, the likelihood is the focus of a Bayesian computation. If we had a perfect model $\mathcal{M}$ that, given the "correct" parameters would fit the data exactly, and for any other parameters would not, the likelihood function would be a delta function: $\mathcal{L}(\vec{\theta}) = \delta(d - \mathcal{M}(\vec{\theta}))$. Unfortunately, any realistic experiment has to deal with noise. We can assume the data is the sum of a signal s and some noise n. Usually the data will be a series of discrete points over time (such as pulsar timing observations), but they could also be points in the frequency domain or a set of one or more measurements without temporal relation. Without loss of generality, we write data, signal and noise as vectors:

$$\vec{d} = \vec{s} + \vec{n}. \tag{2.22}$$

Or equivalently, each vector as a set of k points $\vec{d} = \{d_i\}$ with $i \in [1, k]$. The signal is modelled by $\mathcal{M}$, such that $s_i = \mathcal{M}(\vec{\theta_0})_i$ for the "correct" parameters $\vec{\theta_0}$. To get a likelihood for this scenario, some assumptions need to be made about the noise. The usual one is that the noise is Gaussian, i.e. the points n_i are drawn from a multivariate Gaussian distribution:

$$n_i \sim \mathcal{N}(\vec{0}, \Sigma). \tag{2.23}$$

where Σ is its $(k \times k)$ covariance matrix:

$$\Sigma_{ij} = \langle n_i n_j \rangle. \tag{2.24}$$

It is a positive definite matrix with an inverse Σ^{-1}. The noise distribution has zero mean (if it did not, we could set it to zero by adding the mean to the signal model instead). With these assumption, the likelihood is given by a Gaussian probability distribution:

$$\mathcal{L}(\vec{\theta}) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp\left[-\frac{1}{2}\left(d_i - \mathcal{M}(\vec{\theta})_i\right)^* \left(\Sigma^{-1}\right)^{ij} \left(d_j - \mathcal{M}(\vec{\theta})_j\right)\right]. \tag{2.25}$$

Here, $|\Sigma|$ is the determinant of the covariance matrix. The first vector term in the exponential is a complex conjugate (the *), which is only needed if the data has complex values. The covariance matrix is always real. The term in the exponential can be interpreted as the distance between data points d_i and model points $\mathcal{M}(\vec{\theta})_i$, weighted by the covariance matrix Σ; the likelihood is maximised when this "distance" is minimised.

Noise modelling

In the simplest case, the noise in each measurement is independent. Then, the covariance matrix is diagonal:

$$\Sigma_{ij} = \langle n_i n_j \rangle = \sigma_i^2 \delta_{ij}, \tag{2.26}$$

where σ_i^2 is the variance for point i (which may or may not be all the same). For example, in a very simple model of a PTA in which only white noise is considered, i may label the pulsars and σ_i^2 would be the noise rms associated with that pulsars.

In reality, pulsar timing noise is an amalgamation of different deterministic and stochastic processes. Any deterministic contributions to the data that are considered noise (i.e. not part of the signal one is interested in) should be included in the model. After all, from a mathematical perspective it is not relevant whether a signal is considered interesting or not. Stochastic processes can be modelled as well, but are part of the noise n and affect the covariance matrix Σ. If these noise models depend on unknown parameters, they can be treated in the same way as the signal model parameters. In a typical Bayesian analysis of PTA residuals, (a subset of) noise parameters are fit for at the same time as TM and GW model parameters (see, for example, the NANOGrav search for resolvable binaries in their 11 year DR [77]).

2.3.3 Null streams in PTA data analysis

Any data analysis method needs to distinguish between signal and noise, for both detection (for example using Eq. 2.20) and parameter estimation (for example using Eq. 2.18). The methods that form the basis of the works presented in this thesis aim to do so by constructing *null streams*. In a collection of data streams from detectors that measure a common signal, null streams are combinations of data streams in which the signal is cancelled ("nulled"). It is possible to do so as long as the number of detectors is greater than the number of degrees of freedom in the signal, so that it is overconstrained. For a GW signal, assuming GR is correct, the degrees of freedom are the two polarisations + and ×. In a PTA with P pulsars, each sensitive to the same GW, it is therefore possible to construct $\mathrm{P}-2$ independent null streams. This construction (see Sec. 3.1.1 for details) is based on the detector response functions, given by Eq. 2.6 for PTAs. Therefore, any form of noise which has a different response would not be cancelled, providing the distinguishing feature we are looking for.

Uses of null streams in the literature

The null stream formalism is not bound to PTAs, as it can potentially be used whenever a signal is measured by enough independent detectors. It has previously been applied across the GW spectrum. For networks of ground-based detectors, the method has been proposed to discriminate between signal and unmodelled noise [3, 84, 85, 86]. In the context of LISA, the Sagnac configuration of the detector provides an example of a null stream; the interferometer channels are combined to cancel out GW signals, thus serving as detector calibration by assessing the instrumental noise level [87]. More recently, [88, 89] have adapted those techniques to PTA, and investigated the benefits of using null streams to reconstruct GW signal properties and quantify detection confidence in a frequentist framework. [90] constructed an alternative statistic by combining null streams, specifically for the purpose of localising a resolvable source.

Null streams in this work

We again apply the null stream formalism to resolvable PTA sources, but use a Bayesian framework as described in Sec. 2.3.1. Because the null streams are constructed from response functions that depend on the GW propagation direction, only the GW signal from a given direction is cancelled. This means the null streams can be used to localise the source of a signal, if it is present in the data [88, 90]. In Chapter 3, we take a first step to doing so by writing a standard PTA Gaussian likelihood in terms of the null streams. Although this likelihood is mathematically equivalent to the standard form, its specific sky parameter dependence allows a somewhat different approach to recovering the sky location of a simulated signal. We use this method to assess the localisation capabilities of different simulated PTAs, potentially useful for future observation or analysis strategies.

In Chapter 4, we use the localisation pipeline to assess how many potential host galaxies there would be for a likely future PTA detection of a SMBH binary. To enable multi-messenger astronomy with GWs, one needs to identify in which galaxy the binary resides. This can be done using electromagnetic observations, but a short list of candidates significantly reduces the effort and increases the chances of finding the host. We combine (in a Bayesian way) the information obtained from the localisation pipeline with galaxy properties – location, mass, and distance – from a mock catalogue (developing a method for future use with a real catalogue).

We expand on the work from Chapter 3 in Chapter 5 in two ways. First, we address an issue that arises from a need to work in the frequency domain (FD) when constructing null streams from PTA data. Because the data are not evenly sampled, this can not be done using a Fast Fourier Transform (FFT) and we instead apply an adapted Fourier transform (Sec. 5.1.1). Second, starting from the likelihood from Chapter 3 (Eq. 3.9), we construct a null stream likelihood that is now fully independent of the GW model, and investigate its effectiveness. The results of each work are presented and discussed at the end of each chapter, and an overall conclusion is given at the end of this thesis.

Chapter 3

Paper I: Sky localisation of resolvable PTA sources using null streams

If you can throw away the need to be certain, and accept (...) that you're just interested in confidence and believes instead, that's a really powerful thing.

Hannah Fry on Bayesian statistics

We aim to create a Bayesian framework in which null streams can be used for parameter estimation (and perhaps detection) of resolvable sources in PTA data. As a first step, we use a null-stream-based likelihood to localise SMBH binaries. The work presented in this chapter has previously been published in [2], and most of the text and figures have been adapted from there (see Declaration for details.)

Our method is detailed in Sec. 3.1, including the null-stream construction in 3.1.1, and localisation in 3.1.2. For this work, we use idealised, simulated data, which is converted to the FD as explained in Sec. 3.1.3. The results detailing the performance of this method are presented and discussed in Sec. 3.2, specifically looking at the scaling of the localisation precision with the number of pulsars in the (simulated) PTA and the signal-to-noise ratio (S/N, ρ in Eqs.). With simulations based on the recent PTA data releases (DRs), we investigate the sky-localisation capabilities of current PTAs. These results, and implications for future PTA strategies, are discussed in Sec. 3.3. The main findings are summarised in Sec. 3.4.

3.1 Method

As previously discussed qualitatively in Sec. 2.3.3, PTA data can be combined into null streams because the number of pulsars P exceeds the number of GW polarisations (which is 2). The construction employs the known response of a pulsar, given in Eq. 2.6. It does not include the pulsar term, which would be worthwhile to add in future work to deal with (almost) monochromatic sources [91] and to improve localisation in general [89]. The response can be written in the form of a matrix equation as follows:

$$\vec{r}(t,\hat{\Omega}) = \begin{pmatrix} F_1^+ & F_1^\times \\ F_2^+ & F_2^\times \\ \vdots & \vdots \\ F_\mathrm{P}^+ & F_\mathrm{P}^\times \end{pmatrix} \begin{pmatrix} H^+(t) \\ H^\times(t) \end{pmatrix} \equiv \mathcal{F}(\hat{\Omega}) \begin{pmatrix} H^+(t) \\ H^\times(t) \end{pmatrix}, \tag{3.1}$$

where we have defined the response matrix $\mathcal{F}(\hat{\Omega})$ (the response functions in $\mathcal{F}$ all depend on the GW propagation direction $\hat{\Omega}$, but that dependence will be dropped to shorten notation in the following). It will later be useful to define the columns of $\mathcal{F}$ as $\vec{F}^+$ and $\vec{F}^\times$. $\mathcal{F}$ depends on the location of the source $-\hat{\Omega}$, but not on the parameters of the specific functional form of the polarisations, which makes the following null-stream construction applicable to any single-source GW.

3.1.1 Null-stream construction

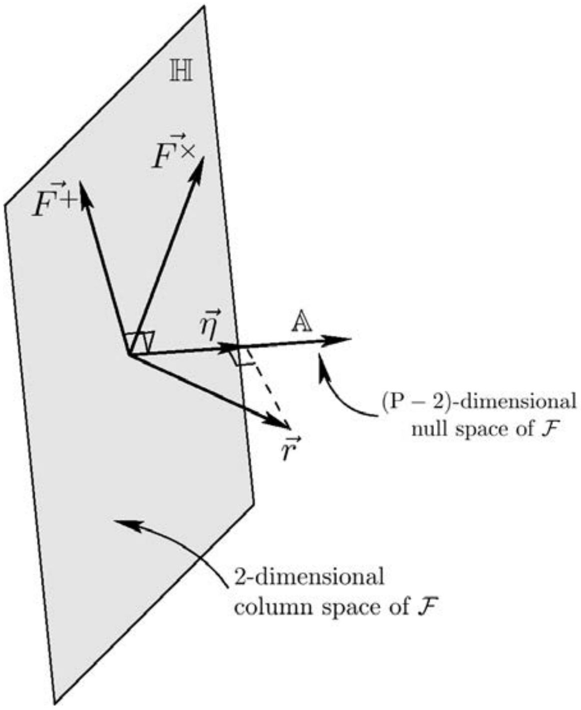

Figure 3.1: Illustration of spaces involved in the mapping $\mathcal{F}$ from GW polarisations to residuals $\vec{r}$. $\mathcal{F}$ is built from the two column vectors $\vec{F}^{+}$ and $\vec{F}^{\times}$, which span the column space of $\mathcal{F}$ or equivalently, the GW-polarisation space $\mathbb{H}$. Orthogonally to $\mathbb{H}$ lies the null space $\mathbb{A}$, which contains the null streams $\vec{\eta}$. This figure has been adapted from Fig. 2 of [3].

The matrix $\mathcal{F}$ can be interpreted as a mapping of the GW polarisations $\vec{H} \equiv (H^{+}, H^{\times})$ in a two-dimensional space $\mathbb{H}$, to the residuals $\vec{r}$ in a P-dimensional space[1]. This suggests that at each point in the domain of $\mathcal{F}$, there exists an additional (P − 2)-dimensional space orthogonal to $\mathbb{H}$. We call this the null space $\mathbb{A}$, spanned by a set of independent null streams (these spaces are illustrated in Fig. 3.1). There are different possible sets of null streams and different methods of obtaining them (e.g. our method differs from the one employed in [88, 89]). It is possible to impose the additional restriction that the null streams are orthogonal to each other (rather than just independent). The method to do so described here has been adapted from [3] and [86]. In short, we find a projection to the null space $\mathbb{A}$ that maps the residuals to null streams.

By inverting the mapping $\mathcal{F}$, the residuals can be reverted to the signal $\vec{H}$. The matrix $\mathcal{F}$

---

[1] We have dropped the time dependence in the GW polarisations and residuals, as the null-stream construction applies to any given time stamp of the quantities (or frequency bin when working in the FD).

is not square, and so clearly not invertible, but it is possible to use the Moore-Penrose inverse or pseudo-inverse $\mathcal{F}_{\mathrm{MP}}^{-1} \equiv (\mathcal{F}^{\top}\mathcal{F})^{-1}\mathcal{F}^{\top}$ instead. For data which is a combination of the residuals and noise given by $\vec{d} = \vec{r} + \vec{n} = \mathcal{F}\vec{H} + \vec{n}$, the maximum likelihood solution for $\vec{H}$ is given by this inversion $\hat{\vec{H}} = \mathcal{F}_{\mathrm{MP}}^{-1}\vec{d}$ [86]. $\mathcal{F}_{\mathrm{MP}}^{-1}$ can be seen as a projection of the data (or residuals) onto $\mathbb{H}$, which is the column space of $\mathcal{F}$. The null streams $\vec{\eta}$ are found by projecting onto the orthogonal space $\mathbb{A}$ instead, which ensures any component of $\vec{H}$ is nulled (see Fig. 3.1).

Say a basis for this null space is $\{\hat{e}_i\}$, with $i \in \{1, 2, .., \mathrm{P} - 2\}$. Each basis vector is orthogonal to the column vectors of $\mathcal{F}$, i.e. $\hat{e}_i \cdot \vec{F^{+}} = \hat{e}_i \cdot \vec{F^{\times}} = 0$. Then a suitable null-space projection matrix is given by $\mathcal{A}$ with rows $\hat{e}_i^{\top}$, because $\mathcal{A}\mathcal{F} = 0$ as per the relation above. Given $\mathcal{A}$, the null streams can be computed as follows:

$$\mathcal{A}\vec{d} = \mathcal{A}\left(\mathcal{F}\vec{H} + \vec{n}\right) = \vec{\eta} + \mathcal{A}\vec{n}. \tag{3.2}$$

Here the first term $\vec{\eta}$ is a vector of $\mathrm{P} - 2$ zeroes; the null streams. We choose to keep the symbol $\vec{\eta}$ in the notation (rather than just using zeros), to specify the null streams constructed in this way, and to label them as η_i, with $i \in [1, \mathrm{P} - 2]$. This construction is also possible for data that does not contain the residuals response as in Eq. 3.1 – for example, a GW signal from a different direction – in which case $\eta_i \neq 0$.

To find the basis $\{\hat{e}_i\}$, consider the projection operators $\mathcal{P} = \mathcal{F}\mathcal{F}_{\mathrm{MP}}^{-1}$, and $\mathcal{S} = \mathcal{I} - \mathcal{P}$, where $\mathcal{I}$ is the $(\mathrm{P} \times \mathrm{P})$ identity matrix [3, 86]. $\mathcal{P}$ projects onto the column space:

$$\mathcal{P}\mathcal{F} = \mathcal{F}\mathcal{F}_{\mathrm{MP}}^{-1}\mathcal{F} = \mathcal{F}, \tag{3.3}$$

and $\mathcal{S}$ onto the null space:

$$\mathcal{S}\mathcal{F} = (\mathcal{I} - \mathcal{P})\mathcal{F} = \mathcal{F} - \mathcal{F} = 0. \tag{3.4}$$

However, $\mathcal{S}$ is a $(\mathrm{P} \times \mathrm{P})$ matrix whereas the null space only has $\mathrm{P} - 2$ dimensions. $\mathcal{S}$ can be reduced to $((\mathrm{P} - 2) \times \mathrm{P})$ with the QR-decomposition. Computing it[2] allows to factorise any real, square matrix into $\mathcal{S} = \mathcal{Q}\mathcal{R}$, where $\mathcal{Q}$ is an orthogonal matrix and $\mathcal{R}$ an upper triangular matrix. It is then given that, if $\mathcal{S}$ has rank r, the first r columns of $\mathcal{Q}$ form an orthonormal basis for the column vectors of $\mathcal{S}$. Therefore, the first $\mathrm{P} - 2$ columns of $\mathcal{Q}$ from this decomposition form the basis $\{\hat{e}_i\}$ that we needed to construct $\mathcal{A}$.

---

[2] The QR-decomposition is computed in PYTHON using the LINALG package `www.numpy.org/doc/stable/reference/generated/numpy.linalg.qr`

We can combine the projections $\mathcal{F}_{\mathrm{MP}}^{-1}$ and $\mathcal{A}$ into one matrix $\mathcal{N}$, by stacking the two rows of $\mathcal{F}_{\mathrm{MP}}^{-1}$ and the $\mathrm{P}-2$ rows of $\mathcal{A}$ on top of each other. The result is a square $\mathrm{P}\times\mathrm{P}$ matrix. It is not proved here but we can numerically verify that this matrix is invertible (provided the pulsar locations are all distinct, which we assume to be the case). It projects the data onto the two reconstructed GW polarisations as well as onto the $\mathrm{P}-2$ null streams:

$$\mathcal{N}\vec{d} = \begin{pmatrix} (\mathcal{F}_{\mathrm{MP}}^{-1})_1 \\ (\mathcal{F}_{\mathrm{MP}}^{-1})_1 \\ \hat{e}_1^\top \\ \vdots \\ \hat{e}_{\mathrm{P}-2}^\top \end{pmatrix} \left(\mathcal{F}\vec{H} + \vec{n}\right) = \begin{pmatrix} H^+ \\ H^\times \\ \eta_1 \\ \vdots \\ \eta_{\mathrm{P}-2} \end{pmatrix} + \mathcal{N}\vec{n} \tag{3.5}$$

$$\equiv \vec{\zeta} + \mathcal{N}\vec{n}. \tag{3.6}$$

In the last step, $\vec{\zeta}$ is defined as the combined vector of (reconstructed) polarisations and null streams. The inverse null-stream matrix $\mathcal{N}^{-1}$ turns the null-stream vector back into residuals: $\mathcal{N}^{-1}\vec{\zeta} = \vec{r}$.

3.1.2 GW-source localisation

The null-stream matrix $\mathcal{N}$ only nulls a GW from a given direction $\hat{\Omega}$. Under the assumption that a single-source signal is present in the data, denoted $\mathcal{H}_{\mathrm{sig}}$, the null streams can be used to find the signal's source location: we vary $\hat{\Omega}$ until the null streams are minimised and the reconstructed GW polarisations match our model for $\vec{H}$ as closely as possible. This can be quantified using Bayes' theorem (Eq. 2.18) for the posterior probability distribution on the sky location (assuming $\mathcal{H}_{\mathrm{sig}}$):

$$p(\hat{\Omega}|\vec{d}, \mathcal{H}_{\mathrm{sig}}) = \frac{p(\vec{d}|\hat{\Omega}, \mathcal{H}_{\mathrm{sig}})\, p(\hat{\Omega}|\mathcal{H}_{\mathrm{sig}})}{p(\vec{d})}. \tag{3.7}$$

The prior on the sky location, $p(\hat{\Omega}|\mathcal{H}_{\mathrm{sig}})$, is naturally chosen to be a uniform distribution, because there is no reason to *a priori* prefer particular source locations over others (we are not taking into account the location of the Virgo Cluster, for example). The likelihood can be based on Eq. 2.25, assuming data given by $\vec{d} = \mathcal{N}^{-1}(\hat{\Omega})\vec{H} + \vec{n}$, with Gaussian noise $\vec{n}$ characterised by a covariance matrix Σ. We require a model for the GW polarisations, say $H^+(\vec{\lambda})$ and $H^\times(\vec{\lambda})$ with parameters $\vec{\lambda}$ (it is intentionally left open for now whether data and model are continuous or discrete, time domain (TD) or FD, a series or single value; so far, the logic applies to any). The log likelihood is given by:

$$\log(\mathcal{L})(\hat{\Omega}, \vec{\lambda}) = -\frac{1}{2}\left((\vec{d} - \mathcal{N}^{-1}\vec{H})^\dagger\, \Sigma^{-1}\, (\vec{d} - \mathcal{N}^{-1}\vec{H})\right) + \text{norm}. \tag{3.8}$$

The normalisation is not explicitly written, as we will end up computing it numerically when needed later on. Here the null-stream matrix $\mathcal{N}(\hat{\Omega})$ depends on the source sky location and $\vec{H}(\vec{\lambda})$ is now the GW model which depends on its parameters $\vec{\lambda}$ (these dependencies are suppressed above to simplify the notation).

Eq. 3.8 can be rewritten by strategically inserting some identity matrices of the form $\mathcal{I} = \mathcal{N}^{-1}\mathcal{N}$, as follows:

$$\begin{aligned}\log(\mathcal{L})(\hat{\Omega}, \vec{\lambda}) &= -\frac{1}{2}\left((\vec{d} - \mathcal{N}^{-1}\vec{H})^{\dagger}\,(\mathcal{N}^{-1}\mathcal{N})^{\dagger}\,\Sigma^{-1}\,(\mathcal{N}^{-1}\mathcal{N})\,(\vec{d} - \mathcal{N}^{-1}\vec{H})\right) + \text{norm.}\\ &= -\frac{1}{2}\left((\mathcal{N}\vec{d} - \vec{H})^{\dagger}(\mathcal{N}^{-\dagger}\Sigma^{-1}\mathcal{N}^{-1})(\mathcal{N}\vec{d} - \vec{H})\right) + \text{norm..} \end{aligned} \tag{3.9}$$

This looks like a Gaussian likelihood for data $\mathcal{N}\vec{d}$, which now depends on $\hat{\Omega}$, a model $\vec{H}(\vec{\lambda})$ and a transformed covariance matrix:

$$Z \equiv (\mathcal{N}\Sigma\mathcal{N}^{\dagger})^{-1}. \tag{3.10}$$

It is mathematically equivalent to Eq. 3.8, but will be computed differently in practice. To obtain the posteriors on the source sky location, as we set out to do in Eq. 3.7, the likelihood has to be marginalised over the GW parameters $\vec{\lambda}$:

$$p(d|\hat{\Omega}, \mathcal{H}_{\text{sig}}) = \mathcal{L}(\hat{\Omega}) = \int \mathrm{d}\vec{\lambda}\,\mathcal{L}(\hat{\Omega}, \vec{\lambda})\,\mathrm{p}(\vec{\lambda}). \tag{3.11}$$

To exploit the separation of $\hat{\Omega}$ and $\vec{\lambda}$ dependencies in the null stream-based likelihood, the marginalisation is done while keeping $\hat{\Omega}$ constant, such that the terms involving $\mathcal{N}$ only need to be computed once for each $\hat{\Omega}$.

3.1.3 Application to simulated data

SMBH binary signal

The null-stream construction and likelihoods from the previous section are generally applicable to any single-source GW, if one has a model for H^{+} and $H^{\times}$. Here, we specifically consider resolvable SMBH binaries in PTA. Assuming the binaries are circularised and effectively monochromatic [20, 92], this model is given by Eq. 2.7 and Eq. 2.8, including the "mixing" of polarisation states as in Eq. 1.74. Instead of considering all the binary parameters in the GW model, we consider the amplitude A as an overall scaling factor (there is always a degeneracy between the distance D_{l} and chirp mass $\mathcal{M}_{\text{c}}$ in this model anyway). Then the vector of parameters is $\vec{\lambda} = (A, \iota, \psi, \omega, \phi_0)$, the amplitude, inclination angle, polarisation angle, angular GW frequency and phase offset, respectively.

Frequency domain

One drawback of the null-stream construction is that it requires taking linear combinations of data from different detectors at *the same* point in time or frequency. In practice, PTA TOAs are not obtained at the same time, so using TD data would require interpolation at best, and extrapolation at worst (some pulsars have been observed since much earlier dates than others). A Fourier transform (FT) may provide a solution, although some form of interpolation is still required for typical unevenly sampled PTA data. This was addressed, for example, by [88], who used linear interpolation between data points. Alternatively, Fourier coefficients for an arbitrary basis of frequencies can be directly estimated via a likelihood calculation for any type of data, as demonstrated in [93]. In Sec. 5.1.1, we discuss applying a straightforward FT to unevenly sampled data. In the following, we make the simplifying assumption that FD data is available. Simulated data is first created in the TD by sampling with constant cadence Δt, allowing to compute the FD data with the FFT algorithm [13].

Because this study addresses parameter estimation and not detection, we assume a SMBH binary has already been detected at known frequency f_0. The likelihood computation can be sped up by restricting the data to only the Fourier component at this frequency. The easiest way to achieve this is by adjusting the cadence to be a multiple of $(\mathrm{nf}_0)^{-1}$, where n is the number of sampled points. Then, the frequency bin with index $\mathrm{f}_0/\Delta\mathrm{f} = \mathrm{f}_0\mathrm{n}\Delta\mathrm{t}$ fully contains the signal power. If this convenience of simulated data is not available, the signal power is spread over multiple frequency bins but can still be recovered.

The GW model from Eq. 2.7 needs to be adapted to the FD. The FT of the sine and cosine are delta functions, which, integrated over frequency, have a nonzero contribution at $f = -\mathrm{f}_0$ and $f = \mathrm{f}_0$. The TD data is real and so the FD data at negative frequencies are the complex conjugate of the same values at positive frequencies, allowing us to only use the positive ones. The discrete FT has an additional factor of $1/\Delta\mathrm{f} = \mathrm{T}$ (the signal power is spread over one frequency bin), such that the model becomes:

$$\begin{aligned} \tilde{\tilde{H}}^{+}(f = \mathrm{f}_0) &= \mathrm{T}\frac{A}{4\omega_0}(1 + \cos(\iota)^2)\exp(i(3\pi/2 + \phi_0)) \\ \tilde{\tilde{H}}^{\times}(f = \mathrm{f}_0) &= \mathrm{T}\frac{A}{2\omega_0}\cos(\iota)\exp(i(\pi + \phi_0)). \end{aligned} \tag{3.12}$$

The mixing with the polarisation angle ψ from Eq. 1.74 and the amplitude A from Eq. 2.8 are unaffected.

Analytical and numerical marginalisation

The null-stream likelihood constructed in Eq. 3.9 for a generic set of data points $\vec{d}$ can be easily applied to our choice for simplified FD data. Because the only non-zero contribution is at $f = \mathrm{f}_0$, the data vector (before applying $\mathcal{N}$) can be replaced by a single point $\tilde{d}(\mathrm{f}_0) \equiv \tilde{d}_0$. The complete model for $\mathcal{N}\tilde{d}_0$ is $\tilde{\zeta}_0 = (\tilde{H}_0^{+}, \tilde{H}_0^{\times}, 0, \ldots, 0)$, a vector of length P (the subscript 0 indicates the quantity is evaluated at $f = \mathrm{f}_0$). The model's null-stream entries are zero, as is the expectation for the null streams constructed using the "correct" sky location. Then the marginalised (see Eq. 3.11) likelihood is:

$$\mathcal{L}(\hat{\Omega}) = \text{norm.} \int \mathrm{d}\vec{\lambda} \, \exp\left[-\frac{1}{2}\left((\mathcal{N}\tilde{d}_0 - \tilde{\zeta}_0)^{\dagger} Z^{-1} (\mathcal{N}\tilde{d}_0 - \tilde{\zeta}_0)\right)\right]. \tag{3.13}$$

There is a trick that can be applied to analytically marginalise over the phase ϕ (assuming a uniform prior $p(\phi_0) \in [0, 2\pi]$) [e.g. 94]. It applies whenever the model can be written as $A_0 \exp(i\phi)$ for some A_0, which is true for Eq. 3.12 (and trivially true for all zeros in $\tilde{\zeta}_0$). The marginalised likelihood (without normalisation), is:

$$\begin{aligned}\mathcal{L}(\hat{\Omega}, A, \iota, \psi) &= \int_0^{2\pi} \mathrm{d}\phi_0 \, \mathcal{L}(\hat{\Omega}, A, \iota, \psi, \phi_0)\, p(\phi_0) \\ &\propto \exp\left[-\frac{1}{2}\left((\mathcal{N}\tilde{d}_0)^{\dagger} Z^{-1} (\mathcal{N}\tilde{d}_0) + A_0^{\dagger} Z^{-1} A_0\right)\right] \times I_0\left[(\mathcal{N}\tilde{d}_0)^{\dagger} Z^{-1} A_0\right]\Bigg|_{\phi_0 = 0},\end{aligned} \tag{3.14}$$

where I_0 is the modified Bessel function of the first kind. The other GW parameters – A, ι and ψ – cannot be marginalised analytically. We therefore integrate the above likelihood numerically with flat priors for ψ in the range $0 - \pi$ and A in the range $0 - 10^{-12}$. The prior for the inclination is uniform in $\cos\iota$, from -1 to 1.

3.2 Results: sky-localisation performance

To investigate the performance of our localisation method, we ran a set of simulations in which a signal according to the SMBH binary model (Eq. 3.12) is injected into a noise realisation. The method described in the previous section can, in principle, be used in conjunction with any noise model by specifying the covariance matrix Σ (in the FD, or by transforming a TD covariance matrix). Here we do not aim to model realistic noise, and use only white noise simulated in the TD (see also Eq. 3.15), which remains white in the FD as well.

The marginalised likelihood (as in Eq. 3.13) is then computed at each point in a grid of sky locations. This grid consists of 12,288 equal area pixels made using the HEALPIX algorithm [12] via HEALPY[3]. Because the likelihood was not normalised to begin with, this is now done numerically (dividing each pixel value by their total sum). As a measure of localisation precision, we define Ω_{90} as the minimum sky area that contains 90% of the posterior on $\hat{\Omega}$ (or 90% of the likelihood, given a flat prior on the sky position). This quantity can be expressed as a fraction (or percentage), or in square degrees (since the whole sky is $4\pi\,\mathrm{sr.} \approx 4.1 \times 10^4\,\mathrm{deg}^2$). From the sky grid of likelihood values, Ω_{90} is easily computed by ordering the pixels by decreasing likelihood and computing the cumulative sum; the fraction of pixels needed to reach 90% is Ω_{90}.

Simulations were carried out with a varying number of pulsars P in the simulated PTA, and over a range of S/N. We chose the values $\mathrm{P} = 3, 5, 10, 20, 30, 50$ and 100. 3 is the very minimum number of pulsars needed to construct at least one null stream, and 50 is about the number of pulsars in the IPTA DR1 [11]. However, IPTA pulsars are not all equally good timers and most of the information is carried by ~10 pulsars with the smallest noise rms. In this respect, $\mathrm{P} = 50$ is more comparable to what might be achieved in the future with the addition of SKA [50] and ngVLA [51].

The set of S/N values used is $1 - 10, 15, 20$ and 30. This is the cumulative S/N in the PTA, i.e. summed over the pulsars:

$$\rho^2 = \sum_{p=1}^{\mathrm{P}} \sum_{i=1}^{n} \frac{r_{i,p}^2}{\sigma_p^2}. \tag{3.15}$$

Here, $\{r_{i,p}\}$ with $i \in \{1, \ldots, \mathrm{n}\}$, is the time series of n residuals from pulsar p. The noise model consists of white noise in the residuals with rms σ_p for each pulsar (which corresponds to $\sqrt{\mathrm{n}}\Delta\mathrm{t}\,\sigma_p$ in the FD). Here[4], all σ_p are set to 100 ns. To adjust the S/N to a desired value, the overall amplitude A of the injected signal is varied.

For each pair of P and S/N values, 10 simulations were performed injecting a source at $\theta = \pi/2$, $\phi = 0$, with a frequency of 20 nHz, and pulsars at randomised locations (with a uniform prior over the sky). These random choices are seeded such that for a given P, for each S/N the same 10 PTA configurations are used. The injected signals are for optimally oriented sources with $\iota = 0$ and $\psi = \pi/8$. $n = 300$ data points were simulated with a cadence of 10^6 s, such that the data contain 6 full cycles of the signal.

---

[3] `healpy.readthedocs.io`

[4] A more sophisticated noise model could be implemented by using a covariance matrix Σ and taking the product $r_i \Sigma_{ij}^{-1} r_j$.

3.2.1 Scaling with signal-to-noise ratio at fixed P

We investigate the sky localisation as a function of the two main parameters identifying the detection, namely the S/N and the number of pulsars in the array, P. First, P is fixed while varying S/N, and in the following Sec. 3.2.2, S/N is fixed while varying P. In Fig. 3.2, the results for P $= 3, 10$ and 100 are shown with points indicating the mean Ω_{90} of 10 simulations for each combination of P and S/N, with error bars showing the span of results (from minimum to maximum). At low S/N, Ω_{90} is limited to 90% of the sky, as there is little or no information gained from the data. For $5 < \rho < 10$, data become informative and the sky localisation rapidly improves, eventually converging to a $\Omega_{90} \propto \rho^{-2}$ relation at high S/N. This behaviour is expected; at high S/N the likelihood surface can be approximated by a multivariate Gaussian around the true value of the source parameters [95]. Parameter determination then follows the theoretical scaling $\Delta\lambda \propto \rho^{-1}$. Sky localisation is given by a combination of the two angle parameters θ and ϕ (or equivalently right ascension and declination), such that the scaling $\Omega_{90} \propto \rho^{-2}$ is recovered.

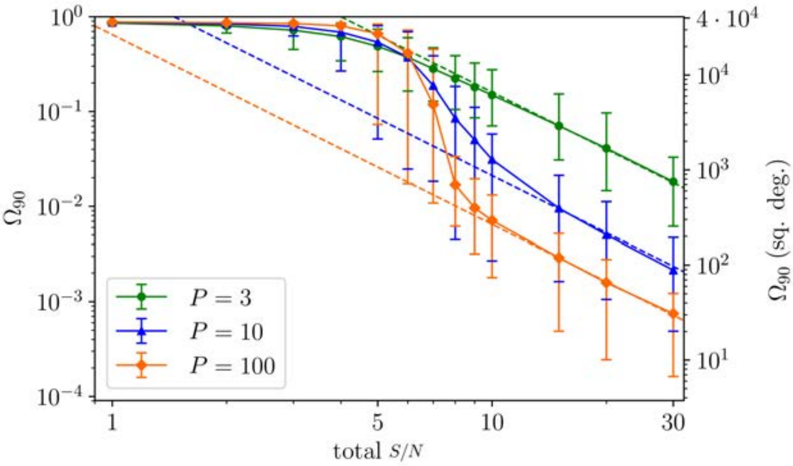

Figure 3.2: Fraction of the sky area containing 90% of the likelihood (Ω_{90}) versus PTA total S/N. Points show the mean value of a set of 10 simulations with randomly placed pulsars, and error bars show the total span of values (minimum to maximum). A power law of $\Omega_{90} \propto \rho^{-2}$ is fitted to the last three points in each curve and plotted as a dashed line. For visibility, only P $= 3, 10$ and 100 pulsars (green circles, blue triangles and orange diamonds, respectively) are included here; Fig. 3.3 shows all curves without error bars.

.

In the region around S/N from ∼5 to ∼10, a transition occurs between the two regimes (from non-informative to informative data). In Fig. 3.3, the means of the 10 runs for all values of P are

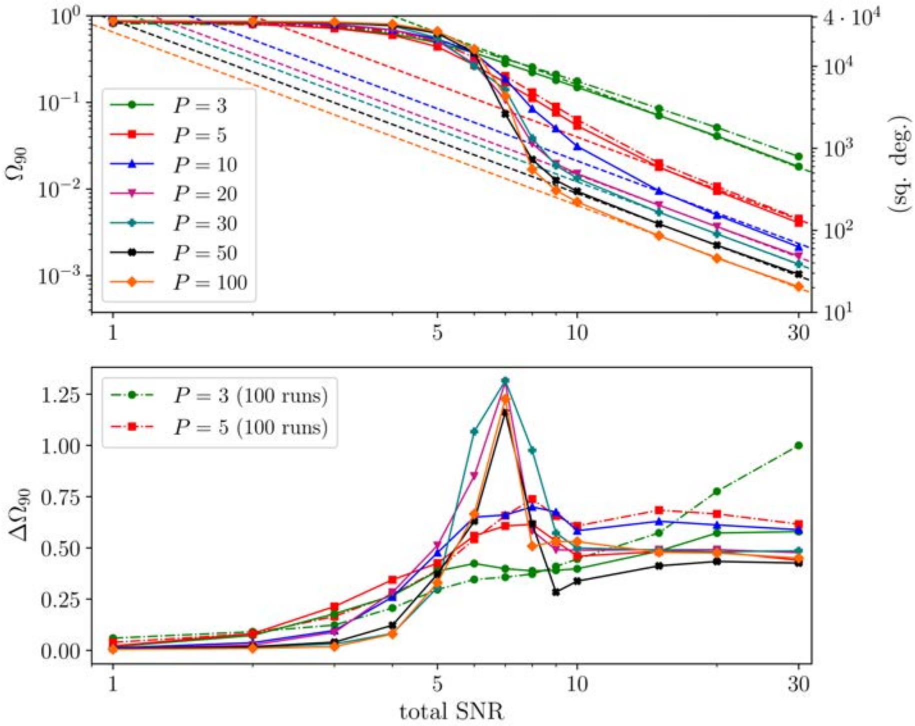

Figure 3.3: Top panel: Fraction of the sky area containing 90% of the likelihood (Ω_{90}) vs PTA total S/N, as in Fig. 3.2. Bottom panel: The spread in results of 10 simulations, computed as the standard deviation normalised by the mean. Dashed lines show values obtained with a re-run of 100 realisations per combination of P and S/N (as verification).

plotted, along with the normalised standard deviation, $\Delta\Omega_{90}$, in the bottom panel (normalised meaning divided by the mean of the corresponding set). During the transition, the mean Ω_{90} behaves similarly for all P, but $\Delta\Omega_{90}$ stands out as being much larger for larger P. An explanation is that for low P the sky localisation is still quite poor during the transition; regardless of the pulsar configuration, Ω_{90} contains a significant fraction of the sky. This is illustrated by the example sky maps in Fig. 3.6, showing Ω_{90} for all 10 random PTAs with P = 3 and S/N= 7. Conversely, for large P, the information carried by the data in the transition region strongly depends on the specific pulsar locations. This is exemplified by the sky maps with P = 30, S/N= 7 in Fig. 3.7. When some pulsars are placed close to the source (e.g. the first panel), the source sky location is determined to relatively high precision despite the relatively low S/N= 7. On the other hand, when there are no pulsars located close to the source (e.g. the third panel), sky localisation is poor and Ω_{90} can span a significant portion of the sky.

3.2.2 Scaling with P at fixed signal-to-noise ratio

The mean Ω_{90} (from 10 runs) as a function of P is shown in Fig. 3.4 for all S/N. For S/N $\lesssim 6$, data are not informative and there is little dependence of the sky localisation on P. As data become informative for S/N $\gtrsim 7$, sky localisations appears to benefit from increasing P. For P $\gtrsim 9$, the improvement in sky localisation precision is well approximated by $\Omega_{90} \propto \mathrm{P}^{-1/2}$, especially for the few highest S/N values.

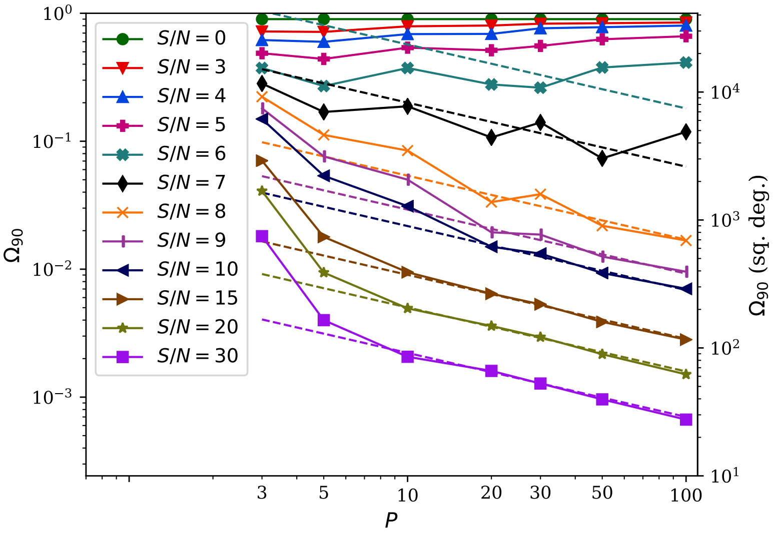

Figure 3.4: Fraction of the sky area containing 90% of the likelihood (Ω_{90}) vs number of pulsars P in the simulated PTA. Each point is the mean value of 10 random realisations of a PTA with total S/N as indicated in the inset label. A power law $\Omega_{90} \propto \mathrm{P}^{-1/2}$ is fitted to each curve, ignoring the first two points ($N = 3, 5$).

A possible explanation for this scaling behaviour can be given by the average (over random PTA realisations) angular distance of the closest pulsars to the source. The expected angular distance between the line of sight to the source and the closest pulsar scales as $\mathrm{P}^{-1/2}$ (for uniform randomly distributed pulsars). These closest pulsars contribute most to the sky localisation (the antenna patters are modulated on the smallest scales close to the pulsar). This conclusion is however not trivial and would need to be tested with $\mathrm{P} > 100$. First, sky localisation depends on the complex interplay of the antenna beam patterns of all the pulsars contributing to the array. Second, if the total S/N is kept constant, not only does the distance to the closest pulsar scale with $\mathrm{P}^{-1/2}$, but the S/N contributed by each individual pulsar decreases, so that the $\Omega_{90} \propto \mathrm{P}^{-1/2}$ scaling is not obvious. In any case, our systematic study indicates that for foreseeable future detections (involving a realistic number of pulsars up to 100 and S/N in the range 6-to-30), $\Omega_{90} \propto \mathrm{P}^{-1/2}$ provides a good empirical fit to the sky-localisation scaling.

3.2.3 Dependence on source orientation

So far we have considered optimally oriented sources, i.e. face on systems for which the two wave polarisations equally contribute to the signal ($\psi = \pi/8$), resulting in a circularly polarised wave. In this case the polarisation angle only affects the initial phase offset of the injected signal ($\tilde{\tilde{H}}^{+}$ and $\tilde{\tilde{H}}^{\times}$ from Eq. 3.12 only differ by a constant phase). Although the sky-localisation scaling obtained in the previous sections is expected to hold for any source inclination and polarisation, the normalisation of Ω_{90} may depend on those quantities. To assess this dependence, we proceed as follows. We fix a PTA of 10 pulsars and a source location in the sky. We then perform 100 simulations picking the source parameters from a 10×10 uniform grid in inclination ι and polarisation ψ. The former is chosen from $\cos(\iota) \in [0, 1]$ ranging from an edge-on to a face-on source, not including the equivalent orientations facing the other way ($\cos\iota < 0$). Similarly, the polarisation is chosen from $\psi \in [0, \pi/2)$, to avoid degeneracy with ϕ_0 at $\psi > \pi/2$ (the closed range means the final bin falls at $9\pi/20$). For this particular experiment, we used noiseless data.

The bottom right panel of Fig. 3.5 shows the resulting Ω_{90} on each point of the aforementioned grid. The bottom left and top right panels show Ω_{90} averaged over the inclination and polarisation angle, respectively. Firstly, there is essentially no dependence of Ω_{90} on ι and ψ so long as the former is smaller than $\sim\pi/3$. This includes about 50% of all binaries, assuming a uniform-on-the-sphere distribution of ι. Secondly, the average sky localisation degrades for $\iota > \pi/3$. However, compared to the reference value of $\Omega_{90} = 0.028$ for the face-on case, the worst $\iota - \psi$ combination results in $\Omega_{90} = 0.046$, which is a factor 1.6 worse. The average sky localisation of all the orientations with $\iota > \pi/3$, is only a factor 1.2 worse than the face-on case. We therefore conclude that the sky-localisation figures presented in Sec. 3.2.1 and 3.2.2 are a fair representation of PTA capabilities for general SMBH binaries.

3.2.4 Comparison with previous results

We compare our results to two previous studies that systematically investigate sky-localisation precision as a function of the number of pulsars and/or S/N [20, 92]. In [20], Sesana & Vecchio (S&V)

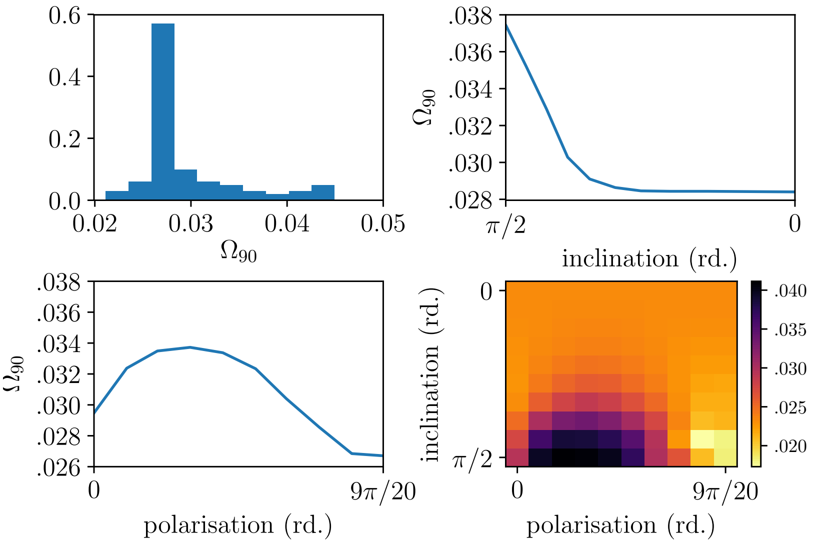

Figure 3.5: Distribution of sky localisations (Ω_{90}) obtained with varying inclination ι and polarisation ψ angle of the source. Top left: Normalised histogram of Ω_{90} of all 100 runs. Top right: Distribution of Ω_{90} for varying inclination (averaged over ψ). Bottom left: Distribution of Ω_{90} for varying polarisation angle (averaged over ι). Bottom right: All Ω_{90} shown as a colour plot for the grid of 10×10 polarisation and inclination values used.

investigated sky localisation of individual sources with PTA using the Fisher Matrix formalism. The main result is shown in their Fig. 7, where the sky-localisation precision, $\Delta\Omega$, is plotted against S/N and number of pulsars. Although results are overall compatible, there are several differences that are worth highlighting.

First, since they employ the Fisher Matrix formalism, S&V find a perfect ρ^{-2} scaling down to S/N$=5$. Our more realistic approach shows that this scaling kicks in only for S/N$\gtrsim 10$, whereas for lower values, sky-localisation performances are much poorer. For example, at S/N$=5$ our simulated PTAs have essentially no localisation power and even at S/N$=7$, typical performances are a factor of $\sim$3 worse than the ρ^{-2} extrapolation predicts. This point is of particular relevance because a PTA signal builds up slowly over time, which means that the first confident single-source detection will necessarily have low S/N. PTAs will therefore have limited capabilities to determine the source parameters in the early stages of detection.

Second, S&V found that the $\mathrm{P}^{-1/2}$ scaling does not hold in general. Their Fig. 7 shows that the sky-localisation improvement flattens out for $\mathrm{P} > 100$, even though $\mathrm{P}^{-1/2}$ might provide a reasonable fit in the $10 \leq N \leq 100$ range investigated in this work. It is likely that a saturation point is reached when the average contribution to the S/N of the closest pulsars is of order 1. At that point, the signal added in each pulsar (if we keep the total S/N fixed) will be below the typical noise level, and no information about the source sky localisation can be gained.

Third, the overall normalisation of the sky-localisation performance is different. For P = 100 and S/N= 10, S&V find a median $\Delta\Omega \approx 40\mathrm{deg}^2$, to be compared to our value of about $200\mathrm{deg}^2$. This is partly due to the different definition of $\Delta\Omega$, which in their study is the region of the sky with probability $e^{-1} \approx 0.63$ of hosting the source. For a multivariate Gaussian likelihood surface, this area is a factor 2.3 smaller than that enclosing the 90% probability that we use. The equivalent 90% probability region of S&V is therefore $\approx 100\mathrm{deg}^2$, which is only a factor of two smaller than what we find. Fisher Matrix calculations provide a limit to the obtainable sky-localisation precision. Even for P = 100 and S/N= 10, we find that the likelihood function is highly non-Gaussian, resulting in a somewhat worse localisation performance compared to the theoretical limit.

In [92], Taylor *et al.* (THGW) constructed a Bayesian pipeline for detection and parameter estimation of *eccentric* binaries and carried out a systematic investigation of parameter errors as a function of S/N. Although the addition of eccentricity increases the complexity of the problem, we do not expect this parameter to couple with the sky localisation, and their results should be comparable with ours. The relevant result for comparison is reported in THGW Fig. 9, which shows $\Delta\Omega$ as a function of S/N for a PTA of 18 pulsars. This PTA is based on the properties of the data used for the NANOGrav 9-year upper limit on the SB [96]. The trend of $\Delta\Omega$ with S/N is very similar to what we found, showing an initial 'transition phase' up to about S/N$\approx$ 8, then settling into the ρ^{-2} behaviour predicted in the strong signal limit. The overall normalisation of the curve is also comparable. At S/N= 20, THGW find a 95% probability region (Ω_{95}) of $\approx 500\mathrm{deg}^2$, which is a factor of a few bigger than Ω_{90} as shown in our Fig. 3.2 for 10 and 20 pulsars, but comparable to the 5 pulsar case. This is likely due to the fact that the 18 pulsars THGW use are not randomly distributed in the sky and have different noise rms, therefore only the few best contribute significantly to the sky localisation. Overall, we deem our results to be in agreement with those in [92].

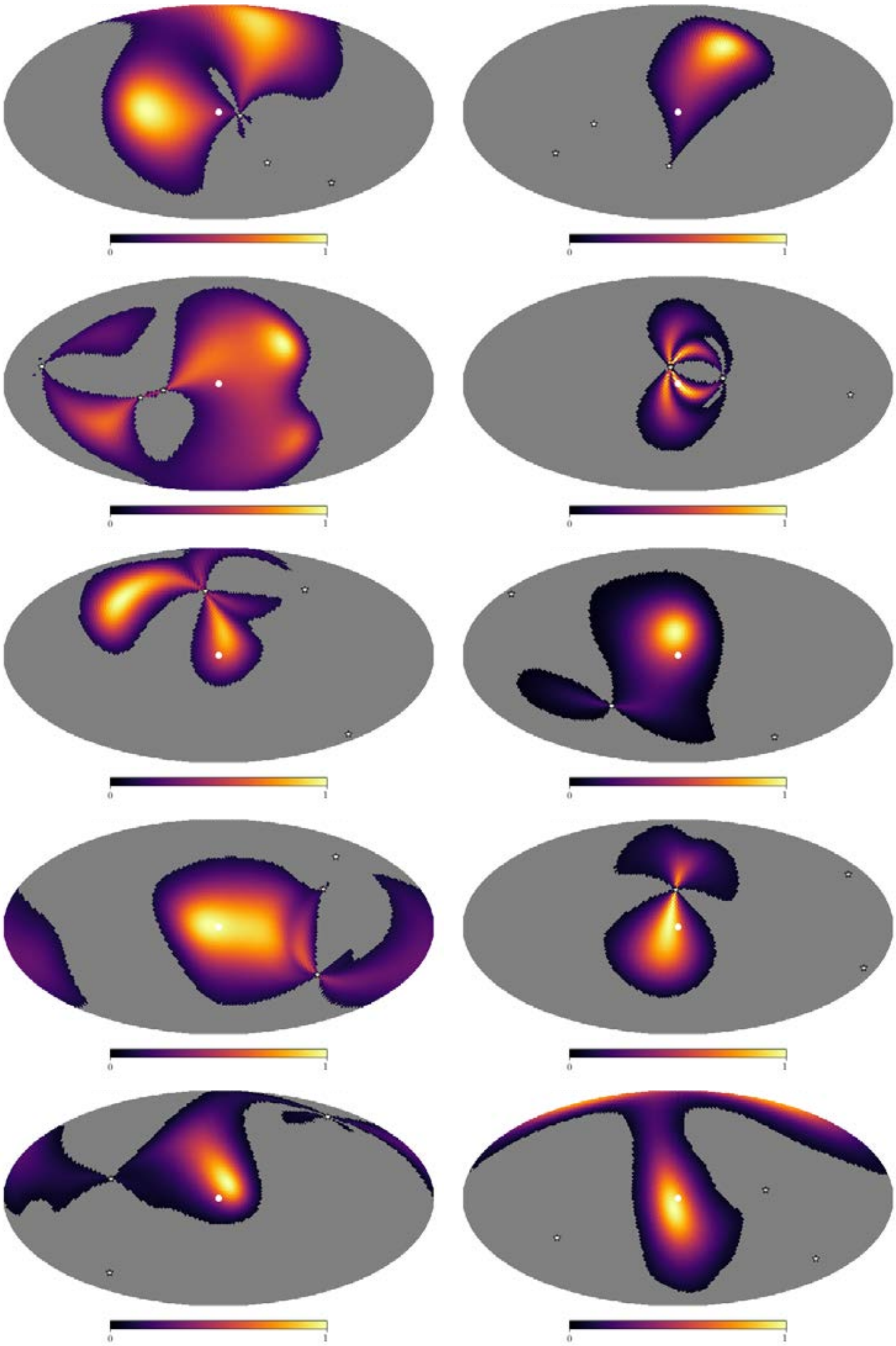

Figure 3.6: Sky maps of 10 different PTA configurations with 3 pulsars, at a total S/N of 7. The injected source is always located in the middle of the map and indicated with a circle marker. The positions of the pulsars are marked with stars. Pixels not contributing to Ω_{90} are masked in grey. Ω_{90} ranges from 0.128 to 0.469 ($\Delta\Omega_{90} = 0.563$ dex).

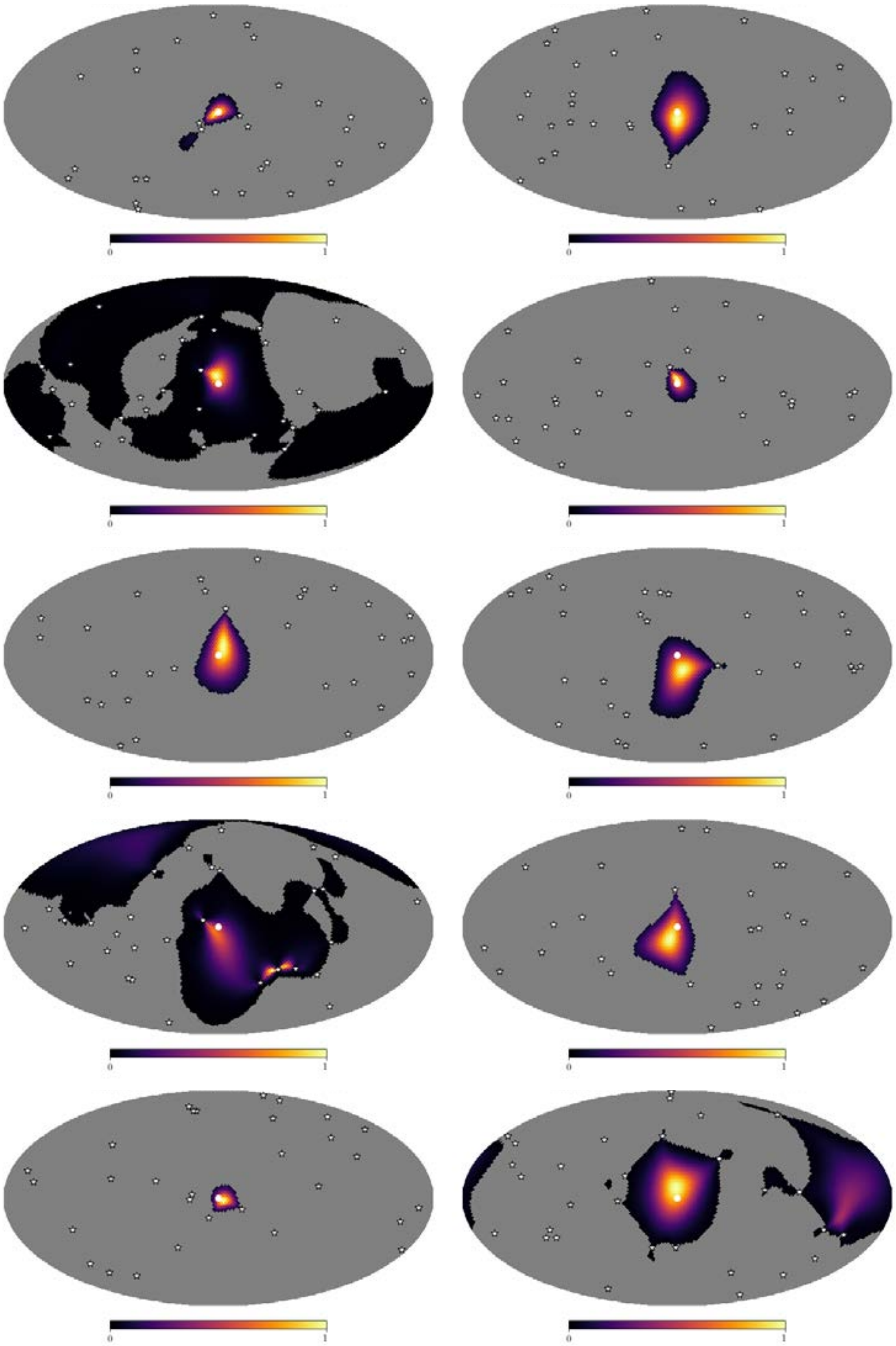

Figure 3.7: Sky maps of 10 different PTA configurations with 30 pulsars, at a total S/N of 7. The injected source is always located in the middle of the map and indicated with a circle marker. The positions of the pulsars are marked with stars. Pixels not contributing to Ω_{90} are masked in grey. Ω_{90} ranges from 0.0082 to 0.575 ($\Delta\Omega_{90} = 1.84$ dex).

3.3 Implications for existing Pulsar Timing Arrays

The null-stream formalism developed in this work can be used to assess sky-localisation capabilities of current PTAs. In the previous section, we demonstrated the beneficial effect on sky localisation of higher S/N and larger number of pulsars. The obvious way to increase S/N and number of pulsars is to combine individual PTA data sets under the umbrella of the IPTA. In this section we therefore focus on the potential gain in individual source localisation by creating IPTA data sets.

With the aforementioned goal in mind, we need to compare the capabilities of an IPTA data set to those of the individual PTA data sets that went into its production. The only official IPTA DR to date is IPTA DR1 [11][5]. The relevant DRs are therefore:

- EPTA DR1 [42], consisting of 42 millisecond pulsars (MSPs) monitored with radio telescopes at Effelsberg, Jodrell Bank, Nancay and Westerbork;
- the extended PPTA DR1 [45], consisting of 20 MSPs monitored with the Parkes radio telescope;
- NANOGrav 5-year DR [81], consisting of 17 MSPs, monitored with the Arecibo and Green bank radio telescopes;
- IPTA DR1 [11], consisting of the combination of the three aforementioned data sets, for a total of 49 MSPs.

Several MSPs are monitored by multiple regional PTAs, and so the number of MSPs in IPTA does not correspond to the sum of those in the regional ones. By combining pulsar data sets, however, IPTA features more high quality pulsars than the regional PTAs. We also stress that the chosen DRs are the ones that were used to build IPTA DR1, and not necessarily the most recent regional PTA DRs, which is the meaningful thing to do given our aim to assess the benefit of combining PTA data specifically.

The current implementation of our technique permits the use of a different noise rms and sky location for each individual pulsar, but is only applicable to evenly sampled data spanning the same observation time. The PTA data set properties therefore need to be modified, keeping as close as possible to the originals. For each PTA we compute an average data set length (observation time) $\bar{T}$ as

$$\bar{T} = \frac{1}{\mathrm{P}} \sum_{p=1}^{\mathrm{P}} T_p, \tag{3.16}$$

where the index p runs over all pulsars in the array and T_p is the data set length of the p-th pulsar.

---

[5] At the time of writing of this study. IPTA DR2 was released in late 2019 [33].

PTA	P	$\bar{n}$	$\bar{T}$[yr]	S/N	Ω_{90}[deg^2]	$\mathcal{R}$	$\tilde{\mathcal{R}}$
EPTA	42	592	12.7	19.4	4492	22.0	3.3
PPTA	20	186	6.3	21.8	949	5.0	2.2
NANOGrav	17	50	4.8	8.0	14172	102.2	18.8
IPTA	49	1401	11.1	28.7	167	–	–

Table 3.1: For each PTA, we list the number of pulsars P, the average number of TOAs per pulsar $\bar{n}$ and the average data set length $\bar{T}$. Also listed are the performances of each PTA for a face on source with $A = 10^{-13.5}$ averaged over position in the sky: S/N, sky localisation Ω_{90}, and improvement factors $\mathcal{R}$ and $\tilde{\mathcal{R}}$ of IPTA compared to regional PTAs (see text for details).

Likewise, we compute an average number of TOAs per pulsar $\bar{n}$ as

$$\bar{n} = \frac{1}{N}\sum_{p=1}^{N} n_p, \tag{3.17}$$

where n_p is the number of TOAs of the p-th pulsar in the array. $\bar{n}$ is rounded to the next integer. The values of $\bar{T}$ and $\bar{n}$ for each PTA are given in Table 3.1. Individual pulsar locations and noise rms values are used as reported in [42] (their Table 1 under rms) for EPTA, in [44] (their Table 7 under Rms res.) for PPTA, in [81] (their Table 2 under rms) for NANOGrav, and in [11] (their Table 4 under Residual rms) for IPTA.

Now that we have specified the properties of the PTAs, we conduct our experiment by considering a face-on circular SMBH binary producing a monochromatic signal with frequency $f = 20$nHz and amplitude $A = 10^{-13.5}$, well within the reach of all PTAs. We place the source in turn at 48 different points in a grid over the sky and use the null-stream technique described in Sec. 3.1 to compute the Ω_{90} sky localisation. The resulting map of Ω_{90} values is an impression of how well the PTA would perform at localising a source placed across the sky (without taking into account the varying probability of detecting a source at different sky locations).

Results are presented in Fig. 3.8, where colour maps have been interpolated from the Ω_{90} values at 48 grid points. Firstly, the uneven pulsar distribution in the sky results in a very source-position-dependent sky-localisation precision. This is particularly true for EPTA and NANOGrav that have localising power mostly in the left side of the map, where all the best pulsars are concentrated, but it still holds for PPTA and IPTA to a lesser extent.

Secondly, the sky-localisation performance differs between PTAs. Due to the limited number of good pulsars and of the short data span, the NANOGrav 5-year DR performs poorly. EPTA and PPTA on the other hand have comparable capabilities, although the latter performs better in the right

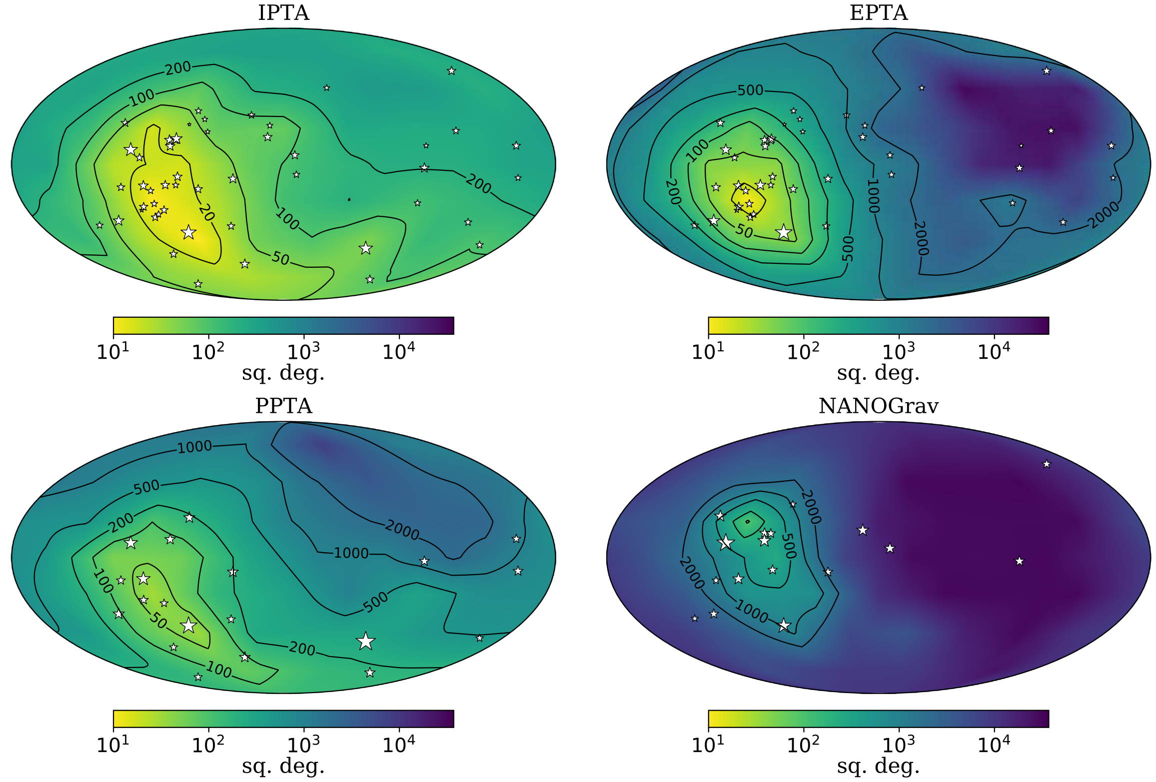

Figure 3.8: Ω_{90} as a function of source sky location for IPTA, EPTA, PPTA and NANOGrav (f.l.t.r. and f.t.t.b.). The simulated PTAs are approximates of the IPTA DR1 and its constituent data sets: EPTA DR1, the extended PPTA DR1 and the NANOGrav 5-year DR (see text for details). The maps are interpolated from 48 pixels for which a value of Ω_{90} was obtained by placing a source with a fixed amplitude ($1.0 \times 10^{-13.5}$) in the middle of that pixel. Contours are plotted at Ω_{90} = 20, 50, 100, 200, 500 and 1000 deg^2.

half of the map. The IPTA gives the best localisation overall. The injected source can be localised to better than 20deg^2 over a region of the sky of about 3500deg^2, whereas a comparable precision is achieved only by EPTA, on a smaller region of < 1000deg^2. Overall, IPTA can locate the source to better than 500deg^2 regardless of its sky location and to better than 200deg^2 over two thirds of the sky. For comparison, PPTA can locate the source to better than 500deg^2 in about half of the sky, and in some regions localisation is worse than 2000deg^2. On average, IPTA can localise the source within 167deg^2 whereas the averages for EPTA, PPTA and NANOGrav are 4492deg^2 949deg^2 and 14172deg^2, respectively (see also Table 3.1).

We can define a relative improvement factor of IPTA sky localisation with respect to a regional

PTA as a function of the source location $\hat{\Omega}$ as

$$\mathcal{R}_X(\hat{\Omega}) = \frac{\Omega_{90,X}(\hat{\Omega})}{\Omega_{90,\mathrm{IPTA}}(\hat{\Omega})}, \tag{3.18}$$

where X stands for EPTA, PPTA or NANOGrav. This relative improvement is shown in Fig. 3.9. Compared to the best individual PTA data set (PPTA), sky localisation improves by more than a factor of two virtually everywhere in the sky, and up to a factor of ten in some regions, confirming the superior performance of IPTA.

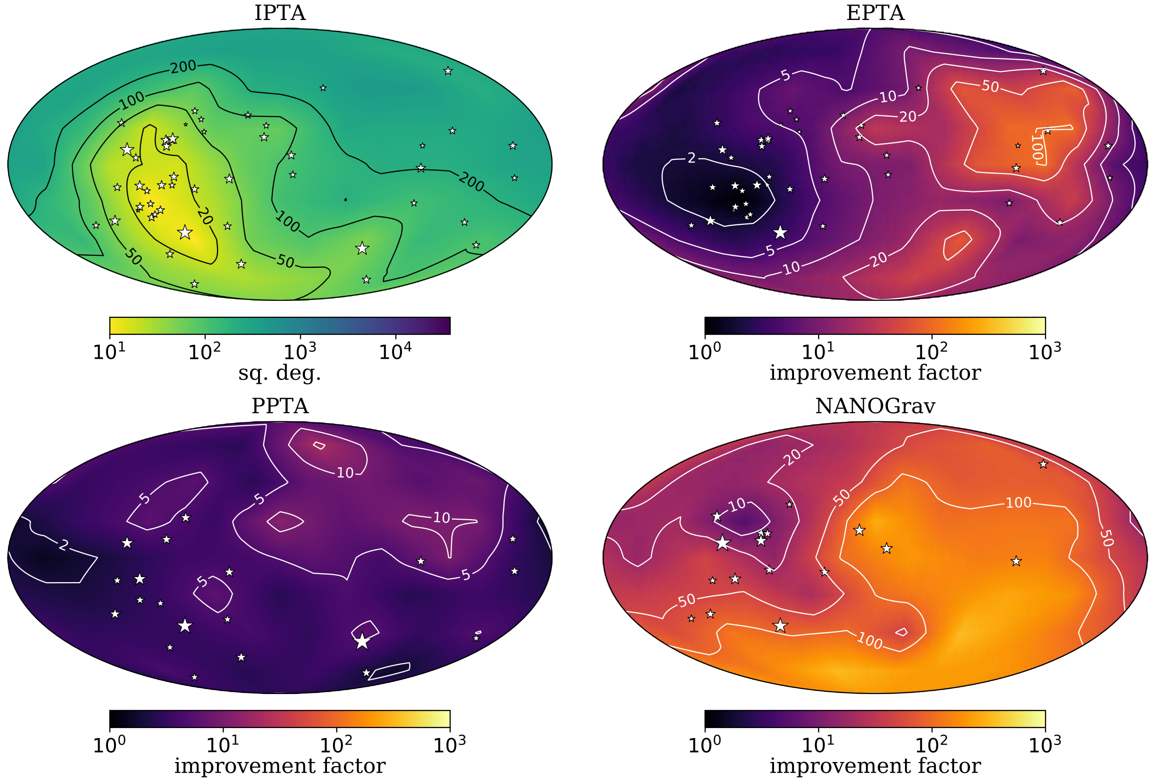

Figure 3.9: The top left panel is the same as in Fig. 3.8. The remaining panels show the improvement factor $\mathcal{R}_X(\hat{\Omega})$ of IPTA compared to EPTA, PPTA and NANOGrav (see definition in the main text and Eq. 3.18). The maps are interpolated from 48 pixels, as in Fig. 3.8. Contours are plotted at $\mathcal{R}_X(\hat{\Omega}) = 2$, 5, 10, 20, 50 and 100.

As shown in Sec. 3.2, the sky localisation naturally improves as ρ^{-2}, but also (although to a lesser extent) as more pulsars are added to the array, even when keeping the total S/N fixed. We therefore investigate whether the improvement of IPTA compared to regional PTAs goes beyond the expected S/N scaling. We define the sky dependent 'S/N-gain-normalised' improvement factor $\tilde{\mathcal{R}}_X(\hat{\Omega})$

as

$$\tilde{\mathcal{R}}_X(\hat{\Omega}) = \mathcal{R}_X(\hat{\Omega}) \times \left(\frac{\rho_X}{\rho_{\mathrm{IPTA}}}\right)^2 . \tag{3.19}$$

By normalising $\mathcal{R}$ with the square of the S/N ratios, $\tilde{\mathcal{R}}$ quantifies the improvement brought by the better IPTA sky coverage.

Results are shown in Fig. 3.10 and highlight that IPTA benefits indeed go beyond the S/N increase. The threshold value of $\tilde{\mathcal{R}} = 1$, indicating improvement is exactly as expected from the S/N, is surpassed by all regional PTAs in most of the sky. Exceptions (which may show that the theoretical S/N scaling is not yet achieved by these examples in the moderate-S/N regime, or may be due to numerical inaccuracy) are a fourth, a sixteenth and a forty eighth of the sky for PPTA, EPTA and NANOGrav, respectively. These are the areas where sky localisation is best for the regional PTAs.

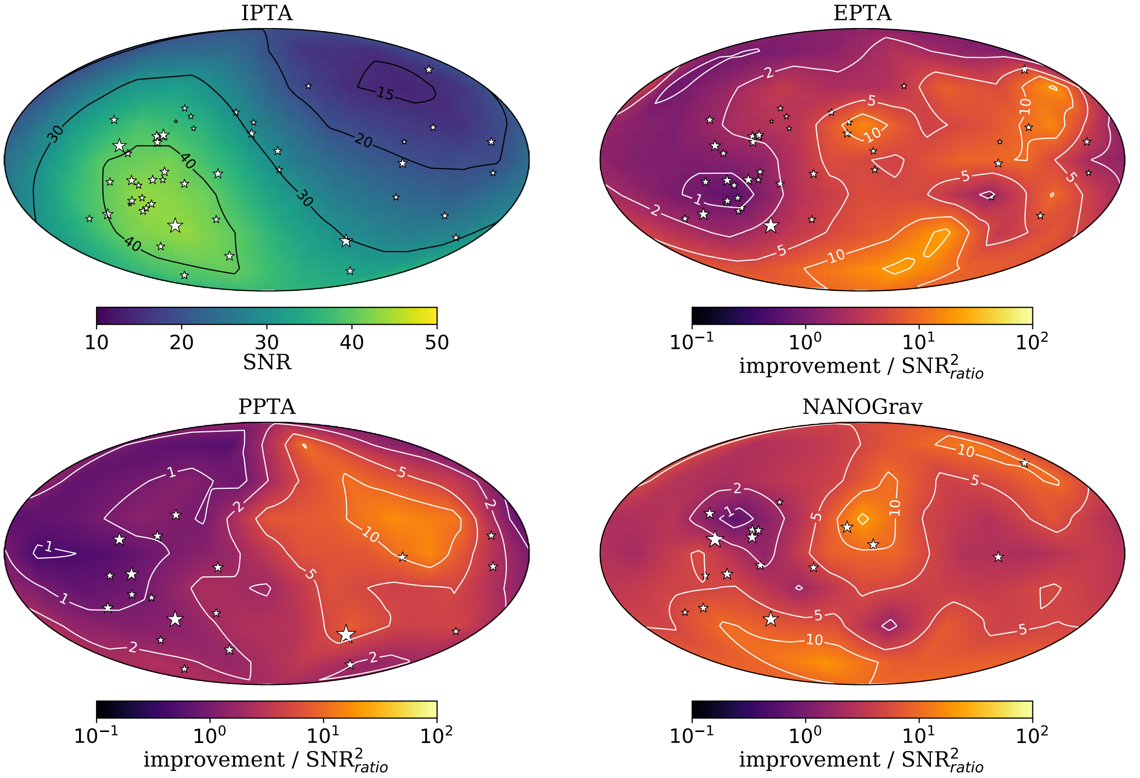

Figure 3.10: The top left panel shows the source S/N as a function of sky location in the IPTA. The remaining plot show the "S/N-gain-normalised" improvement factor $\tilde{\mathcal{R}}_X(\hat{\Omega})$ of IPTA compared to EPTA, PPTA and NANOGrav (see definition in the main text and Eq. 3.19). The maps are interpolated from 48 pixels, as in Fig. 3.8. Contours are plotted at a value of $\tilde{\mathcal{R}}_X(\hat{\Omega}) = 1$, 2, 5, 10 and 20.

In all cases, gain factors of up to 10 are found in parts of the sky, where the beneficial effect of better sky coverage of IPTA is maximised. Averaged over the sky, $\tilde{\mathcal{R}} = 3.3, 2.2, 18.8$ for EPTA, PPTA and NANOGrav respectively, certifying the benefits of IPTA data combination. We remark that the great improvements compared to NANOGrav are simply because only their 5-year data set was included in IPTA DR1. Repeating the analysis for IPTA DR2 [33], which includes the NANOGrav 9-year DR, will allow the verification of the benefits of IPTA when combining data sets of more comparable quality.

We caution that these results have been obtained by using an average timespan and sampling rate for all pulsars of each specific array. In practice, PTA data are not evenly sampled and the timespan of observations varies from pulsar to pulsar. Although the methods would certainly be more complicated, we expect that using more realistic PTA data would only have a minor impact on our conclusions about the localisation *potential* of the PTAs. We have considered typical resolvable sources at a frequency of 20nHz, which is well within the sensitivity band of PTA whether the data is evenly or unevenly sampled. Furthermore, although the cadence and timespan of individual pulsars are different, they usually lie within a factor of two of the average values that we assumed in Table 3.1, suggesting that by using the actual TOA timestamps of each pulsar we should reach similar conclusions. Nonetheless, it is important to verify these expectations by employing an algorithm that can handle the complexity of more realistic data sets, an extension that we plan to explore in future work.

3.4 Conclusions

In this chapter, we have introduced a general mathematical description for the construction of null streams from the response to a resolvable PTA signal. This method is general, works both in the TD and FD and can be applied to any deterministic waveform. We then provided a Bayesian framework to extract the source parameters by exploring the likelihood given by the comparison of the constructed null streams and theoretical model. As proof of concept, we applied our method to the special case of a monochromatic source generated by a circular SMBH binary, considering the Earth term only in the PTA response function. We used this setup to carry out a systematic investigation of PTA sky-localisation capabilities as a function of the array parameters using the sky region containing 90% of the source location likelihood distribution (Ω_{90}) as figure of merit.

We found that for S/N $\gtrsim 10$, Ω_{90} scales as ρ^{-2}, as expected from theoretical arguments in the high S/N limit. However, we find that at low S/N this scaling breaks down, and the source cannot be well-localised. A transition between the two regimes is found for $5 \lesssim$ S/N $\lesssim 10$, in which the Ω_{90} improvement is much steeper than the theoretical scaling. Ω_{90} is also found to scale as the inverse square root of the number of pulsars in the array $\mathrm{P}^{-1/2}$, at least for $10 < \mathrm{P} < 100$ and S/N $\gtrsim 8$. As a reference point, the median Ω_{90} for a source observed with S/N = 10 in an array of 100 equal MSPs randomly distributed in the sky is about 200deg^2. These results are generally consistent with previous findings based on Fisher Matrix calculation, although there are significant differences in the $5 <$ S/N < 10 transition region and in the Ω_{90} normalisation.

We then used our formalism to investigate the sky-localisation capabilities of regional PTAs compared to IPTA. We found that the benefits of combining data in the IPTA DR go beyond the mere gain in S/N due to the accumulation of a larger amount of data. When normalised by expected S/N gain, IPTA is found to perform a factor between ~2 and ~20 better than regional PTAs. This is because combining PTA data provides a better sky coverage and increases the number of high quality pulsars that contribute informative data to the detection. These findings demonstrate that combining regional data under the IPTA umbrella maximises the scientific potential of PTAs as GW detectors.

The framework we applied in this study can be improved in several ways and extended to study a number of problems relevant to PTA data analysis. In particular, our current implementation requires that data are taken at simultaneous times in all pulsars if working in the TD, or that data can be reliably transformed to the FD. One of our primary future goals is to develop an implementation that can handle arbitrary data sets, with unevenly sampled data, gaps, and different time spans, thus allowing the assembly of a pipeline that can be applied to real data[6].

We also considered only the Earth term of the signal which may or may not be appropriate for the loudest SMBH binaries, as shown in [22]. If the frequency of the pulsar and Earth term cannot be separated, then, while the Earth terms may still be cancelled by the null-stream method, there will remain a contribution from the pulsar terms. This could be treated as excess unmodelled noise, or may be modelled explicitly by introducing an additional amplitude and phase term for each pulsar. Efficient methods exist to either maximise or marginalise the calculation over these parameters, as shown for example by [89] and [97], and this is another avenue for improving our method.

---

[6] This will be addressed in Chapter 5

Of great interest is also the expansion of the formalism to treat the cases of multiple deterministic sources and stochastic GW backgrounds. Besides the determination of source parameters, the null-stream formalism provides a powerful tool to validate candidate GW signals and assess detection significance, a possibility that we want to explore in the context of Bayesian model selection.

Chapter 4

Paper II: Associating host galaxy candidates to massive black hole binaries resolved by Pulsar Timing Arrays

"Gravitational Wave Astronomers Hit Mother Lode."

Scientific American on GW170817

Having developed a method to localise resolvable PTA signals, it is interesting to look at the possibility to identify the galaxy that hosts the signal's source. In this chapter, we present a method to rank candidate host galaxies. This work has previously been published in [4], and most of the text and figures have been adapted from there (see Declaration for details.)

4.1 Multi-messenger astronomy with Pulsar Timing Arrays

Multi-messenger astronomy with GWs – specifically, observing the same object with both GWs and electromagnetic (EM) radiation – is a powerful tool for understanding the Universe. The first example of this was the coalescence of a binary neutron star (BNS), observed as GW170817 by LIGO+Virgo [25] and as a bright EM signal at all wavelengths [98]. These observations confirmed theoretical models – the connection between short gamma ray bursts and BNS mergers [99] and the synthesis of heavy elements through r-processes [100] – and opened up the door for cosmology using BNS mergers as standard sirens [101, 102, 103]. All of this has been achieved thanks to the sky localisation ($\sim$31deg^2) and distance information provided by LIGO+Virgo, an intense follow-up campaign, and the presence of a bright EM counterpart that could easily be distinguished from other candidates.

Given the anticipation of detecting individual sources with LISA and PTAs, there is also potential for multi-messenger astronomy at lower GW frequencies. The main candidate for these resolvable sources are (super) massive binary black holes, which – unlike a BNS – do not necessarily have an EM counterpart. It is however possible that a signature could be detected from the gas that accretes onto the binary during inspiral. Although it is not clear exactly what that signature would look like [see, e.g. 104], a range of possibilities have been proposed, from periodicity [e.g. 105] to peculiar features [e.g. 106] and EM chirps [e.g. 107, specifically for LISA]. If PTAs observe an individual SMBH binary in the near future, three main differences between this detection and GW170817 will make finding the EM counterpart more difficult: First, the sky localisation is expected to be relatively poor (of order $\sim$100deg.^{2} [20, and Sec. 3.2]). Second, the GW is likely to be monochromatic (on the timescale of observations) which means there is no measurable chirp to break the degeneracy between the source mass and distance. Last, the long period of PTA signals (or order years) makes it more difficult to identify the host galaxy in relatively short observations [105, 106, 108].

It is therefore crucial to find a way to identify the most promising host galaxy candidates among the millions of objects falling within the source sky location error-box. In this paper, we develop a Bayesian framework to identify the most likely hosts by matching the information contained in a hypothetical PTA detection to candidate galaxy properties. The key point around which our analysis is built is that individually resolvable sources in the PTA band necessarily have a large strain amplitude A [22, 109], which can result only from particularly massive and/or nearby binaries (see also Sec. 2.2.1). This method allows us to exclude most candidate galaxies in the sky error box, providing a shortlist for further investigation.

To demonstrate our method, we simulate PTA data with an injected GW signal based on the expected properties of the first single source to be detected by PTA, using the procedure from [22]. The null-stream analysis from the previous Chapter 3 is adapted to compute a three-dimensional likelihood of the signal amplitude A, and source location θ, ϕ. For the galaxies, we use a mock catalogue extracted from the synthetic all-sky maps made by [110] from the Millennium Simulation [111] , which is complete enough so no viable candidates are missed (for details on the mock catalogue, see Sec. 4.3.3). We then rank these galaxies using our Bayesian framework. If applying our method with a future detection using real, observational catalogues, completeness may be an issue, but this does not affect the assessment of the method in this work.

The chapter is organised as follows. In Section 4.2 we lay out the mathematical basis of our experiment, including the construction of a likelihood from null-streams and the Bayesian framework for the computation of a host galaxy probability. This framework is then applied in Section 4.3 to a number of representative simulations with results laid out in Section 4.4 and the main conclusions and outlook presented in Section 4.5.

4.2 Mathematical framework

4.2.1 Signal model and null-stream sky localisation

We consider a signal from a circular, monochromatic-during-observation SMBH binary as in the previous Chapter (Sec. 3.1.3), with the GW model given by Eqs. 2.7, 2.8 and 1.74. Because the chirp mass $\mathcal{M}_{\mathrm{c}}$ and distance D_{l} are degenerate, we still consider the amplitude A as a parameter on its own for the purposes of computing the likelihood. However, to relate it to the galaxies' properties, the definition of A given in Eq. 2.8 becomes important:

$$A \equiv \frac{4}{D_{\mathrm{l}}} \left(\frac{\mathrm{G}\mathcal{M}_{\mathrm{c}}}{\mathrm{c}^2} \right)^{5/3} \left(\frac{\pi f}{\mathrm{c}} \right)^{2/3} . \tag{4.1}$$

As in Eq. 2.8, the amplitude is expressed in the frame of the observer. This means $\mathcal{M}_{\mathrm{c}}$ is the redshifted chirp mass, which is related to the intrinsic source component masses, $M_1 \geq M_2$, as:

$$\mathcal{M}_{\mathrm{c}} = (1+z)\mathcal{M}_{\mathrm{c}}^{(\mathrm{src})} = (1+z)\frac{(M_1 M_2)^{3/5}}{(M_1+M_2)^{1/5}}, \tag{4.2}$$

and D_{l} is the luminosity distance:

$$D_{\mathrm{l}} = (1+z)D_{\mathrm{H}}\int_0^z \frac{\mathrm{d}z'}{E(z')}. \tag{4.3}$$

In the above, z is the source redshift, $D_{\mathrm{H}} = \mathrm{c}/\mathrm{H}_0$ (the Hubble distance) and $E(z) = \sqrt{\Omega_{\mathrm{M}}(1+z)^3 + \Omega_\Lambda}$, with Ω_{M} and Ω_Λ being the fractional mass and cosmological constant energy content and H_0 the Hubble constant. (We assume a standard flat ΛCDM Universe with $\Omega_{\mathrm{M}} = 0.308$ and $\mathrm{H}_0 = 67.8\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$ from the Planck 2015 results [54]).

To extract information from our simulated data, we use the null-stream-based likelihood constructed in the previous chapter, Eq. 3.9. Now marginalising over one fewer parameter yields the likelihood $\mathcal{L}(A, \theta, \phi)$ (see Sec. 4.3.2 for details on parameter sampling). An example output is given in Fig. 4.1. The following method for host galaxy selection takes a generic input $\mathcal{L}(A, \theta, \phi)$, and so could be used in conjunction with any method for sky localisation and obtaining the likelihood for A.

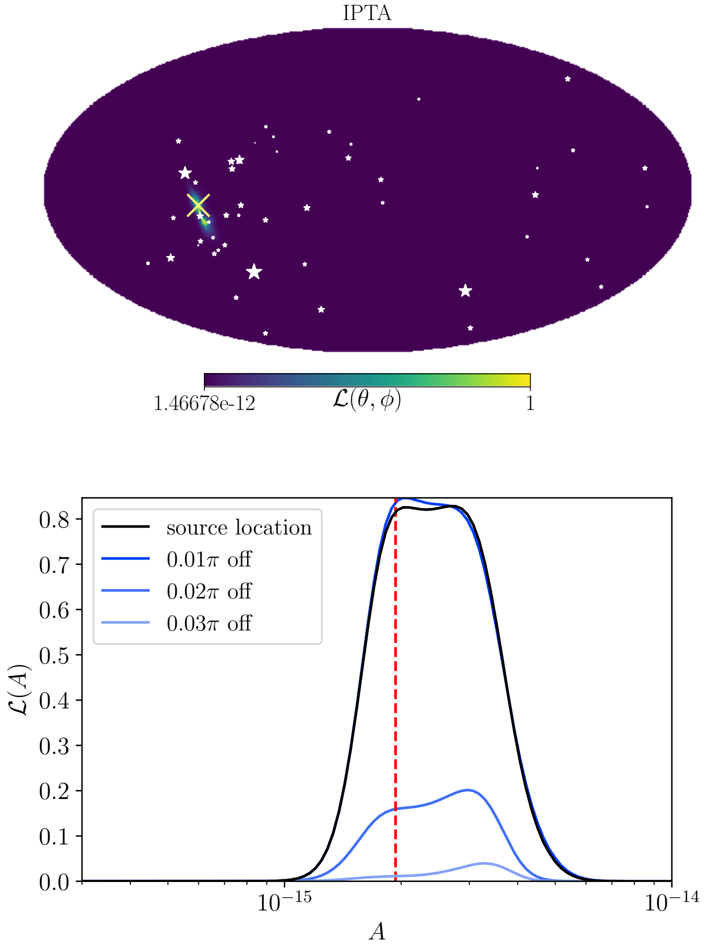

Figure 4.1: Example of $\mathcal{L}(A, \theta, \phi)$ as output by the null-stream pipeline from Chapter 3. The injected signal is for our chosen source A, with S/N = 12 (see Section 4.3).
Top: Likelihood marginalised over amplitude A (i.e. $\mathcal{L}(\theta, \phi)$, with an arbitrary normalisation). The IPTA pulsars are marked with stars, where the size of the star corresponds to the noise level of the pulsar (with bigger stars for lower noise). The yellow cross indicates the position of the injected source.
Bottom: $\mathcal{L}(A|\theta, \phi)$ at the source position (θ_s, ϕ_s) (in black) and at some offset positions $(\theta_s, \phi_s + \Delta)$ (in blue). The likelihoods are normalised only with respect to each other. The red dashed line is placed at the injected amplitude value.

4.2.2 Bayesian inference for galaxy host

Our goal is to combine the likelihood information $\mathcal{L}(A, \theta, \phi)$ with individual galaxy properties to assess the probability of each given galaxy to be the host of the detected GW source. To formalise the question: given the detection of a signal with 3-D likelihood described by $\mathcal{L}(A, \theta, \phi)$, what is the probability that a galaxy G_i described by a set of observed parameters $\vec{\lambda}$ – known with prior probability $p(\vec{\lambda}|G_i)$ – is the host of the signal source? Answering this requires a theoretical model that connects the strength and location of a putative GW signal to observable galaxy parameters.

Since SMBH binaries reside in the centres of galaxies, the sky coordinates of each specific galaxy (θ_G, ϕ_G) coincide with the sky coordinates of the putative GW source. We therefore have $\theta_G = \theta$ and $\phi_G = \phi$. Furthermore, we see from Eqs. 4.1 and 4.2 that the GW amplitude A depends on the source chirp mass $\mathcal{M}_{\rm c}^{\rm (src)}$ and luminosity distance $D_{\rm l}$. The latter can be measured from the galaxy spectroscopic redshift – or with a larger error, from the photometric redshift – by assuming a fiducial cosmology. $\mathcal{M}_{\rm c}^{\rm (src)}$ can be written in terms of the total binary mass M, and mass ratio $q = M_2/M_1$ (with $M_1 \geq M_2$) as: $\mathcal{M}_{\rm c}^{\rm (src)} = Mq^{3/5}/(1+q)^{6/5}$. We can assume the total mass to be related to the bulge mass via an $M - M_{\rm b}$-relation of the form [10]

$$\log_{10}\left(\frac{M}{\rm M_\odot}\right) = \alpha + \beta \log_{10}\left(\frac{M_{\rm b}}{10^{11}{\rm M_\odot}}\right) , \tag{4.4}$$

which connects the total binary mass to the observable galaxy bulge stellar mass $M_{\rm b}$. If we group the $M - M_{\rm b}$ constants α and β with the galaxy parameters, the vector of seven parameters

$$\vec{\lambda} = (M_{\rm b}, D_{\rm l}, \theta, \phi, q, \alpha, \beta), \tag{4.5}$$

is sufficient to connect a specific galaxy to the GW strain. All of them but q, α and β can be directly extracted from observations.

Formally the full calculation can be cast in term of Bayes' theorem. Let $P(G_i|d)$ be the probability of galaxy G_i being the host galaxy, given some data d, then:

$$P(G_i|d) = \frac{P(G_i)}{P(d)} P(d|G_i) = \frac{P(G_i)}{P(d)} \int p(d|\vec{\lambda}) p(\vec{\lambda}|G_i)\, {\rm d}\vec{\lambda}, \tag{4.6}$$

where $P(d) = \sum_i P(d|G_i)$ is the likelihood of the data marginalised over all galaxies (or evidence). $P(G_i)$ is the prior probability of G_i being the host, which we take to be a constant, having no reason *a priori* to prefer any particular galaxy. To compare galaxies, only the shape of the distribution $P(G_i|d)$

is needed. The constant prefactor, $P(G_i)/P(d)$, can therefore be disregarded and we are left with only the likelihood $P(d|G_i)$. The likelihood of a specific galaxy G_i to be the host of the GW source is given by the integral in Eq. 4.6 and is composed of the probability of the data given the source parameters $p(d|\vec{\lambda})$, times the prior distribution on these parameters $p(\vec{\lambda}|G_i)$, integrated over all the relevant variables given in Eq. 4.5.

Now we need to specify how to compute $p(d|\vec{\lambda})$. First, the amplitude A is independent of θ, ϕ and so:

$$p(d|\vec{\lambda}) = p(d|M_{\mathrm{b}}, D_{\mathrm{l}}, \theta, \phi, q, \alpha, \beta) = p(d|A, \theta, \phi)\, p(A|M_{\mathrm{b}}, D_{\mathrm{l}}, q, \alpha, \beta). \tag{4.7}$$

Second, A is a function of the source chirp mass $\mathcal{M}_{\mathrm{c}}^{(\mathrm{src})}$ and distance only, we can therefore write

$$p(A|M_{\mathrm{b}}, D_{\mathrm{l}}, q, \alpha, \beta) = p(A|\mathcal{M}_{\mathrm{c}}^{(\mathrm{src})}, D_{\mathrm{l}})\, p(\mathcal{M}_{\mathrm{c}}^{(\mathrm{src})}|M_{\mathrm{b}}, q, \alpha, \beta). \tag{4.8}$$

Last, $\mathcal{M}_{\mathrm{c}}^{(\mathrm{src})}$ is a function of q and M, and the latter is related to M_{b} by the $M - M_{\mathrm{b}}$-relation. We therefore have

$$p(\mathcal{M}_{\mathrm{c}}^{(\mathrm{src})}|M_{\mathrm{b}}, q, \alpha, \beta) = p(\mathcal{M}_{\mathrm{c}}^{(\mathrm{src})}|M, q)\, p(M|M_{\mathrm{b}}, \alpha, \beta). \tag{4.9}$$

Putting the chain together we get:

$$\begin{aligned} p(d|G_i) = \int & p(d|A, \theta, \phi)\, p(A|\mathcal{M}_{\mathrm{c}}^{(\mathrm{src})}, D_{\mathrm{l}})\, p(\mathcal{M}_{\mathrm{c}}^{(\mathrm{src})}|M, q)\, p(M|M_{\mathrm{b}}, \alpha, \beta) p(M_{\mathrm{b}}, D_{\mathrm{l}}, \theta, \phi, q, \alpha, \beta|G_i) \\ & \mathrm{d}M_{\mathrm{b}}\, \mathrm{d}D_{\mathrm{l}}\, \mathrm{d}\theta\, \mathrm{d}\phi\, \mathrm{d}q\, \mathrm{d}\alpha\, \mathrm{d}\beta\, \mathrm{d}M. \end{aligned} \tag{4.10}$$

We can now further specify the individual elements of Eq. 4.10 and address how to compute them in practice. $p(\vec{\lambda}|G_i) = p(M_b, D_{\mathrm{l}}, \theta, \phi, q, \alpha, \beta|G_i)$ describes the prior knowledge of each galaxy property and the underlying $M - M_{\mathrm{b}}$ constants. We assume that all five galaxy parameters – so excluding α and β – are independent so that the prior can be factorised as $p(\vec{\lambda}|G_i) = \prod_{j=1}^{5} p(\lambda_j|G_i)$. In particular:

- M_{b} in real surveys is generally obtained from the galaxy luminosity via bulge-disk decomposition. M_{b} is then computed from the bulge luminosity by assuming a stellar mass function. Typical uncertainties in this procedure can be up to a factor of two [112]. Nonetheless, as a first approximation, we take M_{b} to be known exactly, reducing the prior $p(M_{\mathrm{b}})$ to a delta function (so the integral over M_{b} drops out).
- D_{l} can be computed from the spectroscopic redshift of the galaxy z, via Eq. 4.3. Uncertainties on the cosmological parameters $H_0, \Omega_{\mathrm{M}}, \Omega_{\Lambda}$ are of the order of a few percent [54] and weak

lensing is subdominant for the $z < 1$ galaxies relevant here [113]. We therefore also assume D_l to be known exactly, reducing the prior $p(D_\mathrm{l})$ to a delta function, and dropping the integration over D_l from the likelihood marginalisation. The effect of including uncertainties on M_b and D_l is discussed in Sec. 4.4.3.

- θ, ϕ are generally determined with arcsecond precision, which for any practical purposes can be treated as delta functions as well.
- q, the binary mass ratio, is essentially undetermined. We therefore use a broad prior, flat in $\log(q)$, with range $-2 \leq \log_{10}(q) \leq 0$ (so that $0.01 \leq q \leq 1$).

The impact of changing the adopted priors in the calculation are discussed in Section 4.4.3.

The term $p(d|A, \theta, \phi)$ is directly proportional to the likelihood in the amplitude-sky location-space $\mathcal{L}(A, \theta, \phi)$, which is returned as a numerical function with finite resolution by our null-stream-based parameter estimation pipeline (see also Sec. 4.3.2). Given the values of A, θ and ϕ from the priors, we select the numerical value from the sky pixel at (θ, ϕ) and the closest sampled amplitude to A. The sampling range (10^{-17}–10^{-14}) is big enough to cover the area of interest, so for values of A outside this range, the likelihood is set to zero.

The GW quadrupole formula determines the factor $p(A|\mathcal{M}_\mathrm{c}, D_\mathrm{l})$. Given the system chirp mass and distance, the amplitude is univocally determined by Eq. 4.1. We can thus write

$$p(A|\mathcal{M}_\mathrm{c}^{(\mathrm{src})}, D_\mathrm{l}) = \delta\left(A - 4\frac{(\mathrm{G}\mathcal{M}_\mathrm{c}^{(\mathrm{src})})^{5/3}(\pi f)^{2/3}}{D_\mathrm{l}}\right). \tag{4.11}$$

Similarly, $p(\mathcal{M}_\mathrm{c}^{(\mathrm{src})}|M, q)$ is obtained from the definition of $\mathcal{M}_\mathrm{c}^{(\mathrm{src})}$ in terms of M, q:

$$p(\mathcal{M}_\mathrm{c}^{(\mathrm{src})}|M, q) = \delta\left(\mathcal{M}_\mathrm{c}^{(\mathrm{src})} - \frac{Mq^{3/5}}{(1+q)^{6/5}}\right). \tag{4.12}$$

The probability distribution $p(M|M_\mathrm{b}, \alpha, \beta)$ is a core ingredient of the calculation. The possibility of ranking galaxy hosts stems from the simple fact that extremely massive black holes are hosted in extremely massive galaxies, a relation that has to be handled with care. Once a specific $M - M_\mathrm{b}$ relation of the form given by Eq. 4.4 with intrinsic dispersion ϵ is given, the BH total mass probability is described by a log-normal prior

$$p(M|M_\mathrm{b}) = \frac{1}{\sqrt{2\pi\epsilon^2}} \exp\left\{-\frac{\left[\log\frac{M}{\mathrm{M}_\odot} - \left(\alpha + \beta\log\frac{M_\mathrm{b}}{10^{11}\mathrm{M}_\odot}\right)\right]^2}{2\epsilon^2}\right\}, \tag{4.13}$$

that we integrate from -3ϵ to $+3\epsilon$ around the minimum and maximum expectation values of M (the range of M values being due to the spread in (α, β)).

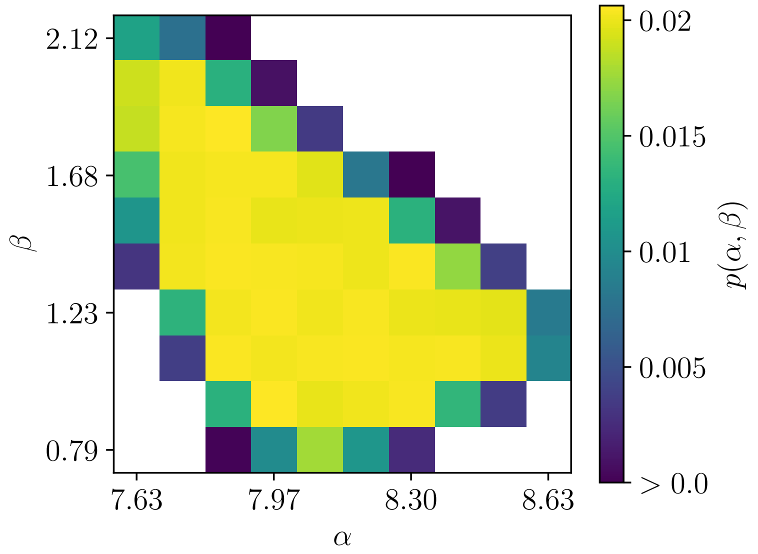

Figure 4.2: Prior on the $M - M_{\rm b}$ constants (α, β) constructed from the compilation of $M - M_{\rm b}$-relations in [10], see text in Sec. 4.2.2. The prior is binned in a 10×10 regular grid with $\alpha \in [7.63, 8.63]$ and $\beta \in [0.79, 2.14]$. The pixels are normalised such that their sum is one. Some combinations (α, β) have zero prior weight and are masked in white.

The $M - M_{\rm b}$ relation is quite uncertain, as demonstrated by its many different flavours found in the literature. Using the compilation of $M - M_{\rm b}$ relations of [10] we construct an observationally motivated prior distribution in (α, β) by the following procedure. We make many random draws of the pair (α, β) uniformly from the ranges $\alpha \in [7.63, 8.63]$ and $\beta \in [0.79, 2.14]$, and consider the pair valid if the resulting $M - M_{\rm b}$ line falls within the region enclosed by the compiled sample of relations in the range $10^6 {\rm M}_\odot < M < 10^{10} {\rm M}_\odot$. The resulting probability distribution $p(\alpha, \beta)$ is shown in Fig. 4.2. We then marginalise over the parameters (α, β) in the computation of $p(d|G_i)$ in Eq. 4.10. We assume $\epsilon = 0.3$ throughout, which is the typical relation dispersion value reported in the literature.

Using these assumption, Eq. 4.10 reduces to a four dimensional integral over M, q, α and β. In the following, we show results where $M - M_{\rm b}$ is always marginalised over (α, β) and we discuss the impact of assuming one specific scaling relation instead in Sec. 4.4.3. In practice, we transform variables to $\log_{10}(M)$ and $\log_{10}(q)$ and perform the numerical integration in log space for these parameters.

4.3 Practical implementation

4.3.1 Source selection

We test the candidate host galaxy selection method using simulated data of plausible future PTA detections. These simulations are based on work by Rosado *et al.* [22], who have studied large scale simulations of SMBH binary populations and the resulting GW signals. They constructed 2000 models (combinations of different observed SMBH mass functions, pair fractions and $M - M_{\rm b}$-relations) and drew several Monte Carlo realisations of each model, to build realistic binary populations. They then considered the sensitivity of several PTAs as a function of time and used simple detection statistics to declare detection of either individual binaries or of the SB. Although they find that it's more likely that the background is detected first, eventually, individual sources can also be confidently identified. For each of the simulations they record the properties of the first SMBH binary to be individually resolved by the PTA under consideration. Therefore, their procedure informs the likely parameters of the first resolvable binary. We use it here to get the parameters for our test injections, as described in the following.

The signal-to-noise ratio (S/N, ρ in Eqs.) of a circular SMBH binary in an array of P pulsars can be written as

$$\rho = \left[\sum_{i=1}^{\rm P} \rho_i^2 \right]^{1/2}, \tag{4.14}$$

where the S/N in the i-th pulsar is

$$\rho_i^2 = \frac{A^2}{4\pi^2 f^2 S_i} \mathcal{R}(\vec{\delta}). \tag{4.15}$$

Here, A is the GW amplitude given by Eq. 4.1 and f is the observed GW frequency. $\mathcal{R}(\vec{\delta})$ is a factor of order unity that depends of the geometry of the system – including source sky location and inclination, wave polarisation angle and pulsar sky location – and on the duration of the PTA observation T; see [22] for the full expression. S_i is the noise in the i-th pulsar which we consider to be of the form

$$S_i = 2\Delta t \sigma_i^2 + S_{h,\rm rest}, \tag{4.16}$$

where the first term on the rhs is the rms noise level of the timing residuals and the second term is the level of confusion noise given by all other sources contributing to the overall GW signal.

To select suitable individual sources, we construct a mock version of the IPTA using the 49 pulsars of IPTA DR1 [33]. The mock IPTA uses the actual sky location and rms noise of each pulsar,

source	$\mathcal{M}\,[\mathrm{M}_\odot]$	z	$f\,[\mathrm{nHz}]$	A
A	3.18×10^9	0.62	7.44	0.96×10^{-15}
B	5.36×10^9	0.57	5.94	2.05×10^{-15}
C	3.69×10^9	0.18	5.18	2.40×10^{-15}

Table 4.1: Properties of the three test sources selected for this study, modelled after predictions of future PTA detections.

but simplifies the observations to be evenly sampled with a cadence of $\Delta t = 2$ weeks, and a timespan of $T = 10$ years. Next, we generate 50 realisations of a realistic population of circular, GW driven BBHs, based on one of the models presented in [114]. The number of realisations is chosen to produce a sample of individually resolvable sources that is large enough to freely pick sources in desired regions in the sky (see below). We use a fairly optimistic model, resulting in a characteristic GW strain $A(f = \mathrm{yr}^{-1}) \approx 1.3 \times 10^{15}$, which is just at the edge of the most recent PTA limits [11, 48, 79, 80] (see also Sec. 2.2.4).

In each model realisation, we select the loudest sources one-by-one and use all remaining binaries to consistently compute $S_{h,\mathrm{rest}}$ from Eq. 4.16. All potentially resolvable GW sources had S/N < 2 in the adopted setup. This is a good sanity check for our simulation; it is expected that no observable sources result from this procedure, given that no single SMBH binary has been detected to date with the real IPTA either. To increase the S/N, we suppress the noise by multiplying each rms residual σ_i by a fudge factor $\eta < 1$. After decreasing η to 0.2, we observe ∼30 sources (in 50 GW signal realisations) at S/N $\gtrsim 5$. We select three of those sources, which we name A, B and C.

Relevant parameters of the selected sources are listed in Table 4.1 and their location in the sky, relative to the IPTA pulsars, can be seen in Fig. 4.4. We have intentionally picked three sources in areas of different IPTA pulsar density. Because the response functions depend on the angular distance between the pulsar and the GW propagation direction (Eq. 2.5), the localisation behaviour is different for sources that are close to (good) pulsars than for those in relatively empty regions of the sky (see also Sec. 4.4.1). Parameters listed in Table 4.1 are consistent with the predictions of Rosado *et al.* shown in their Fig. 6 [22]. The first resolvable sources are likely to be at low frequencies (few nHz) and can come from SMBH binaries at moderate redshifts (up to $z \approx 1$).

4.3.2 Source injection and likelihood evaluation

We simulate data from a synthetic PTA, based on the IPTA DR1 [11], as was done before to study IPTA source localisation in Sec. 3.3. Again, the sky localisation and residuals rms level are the same as for IPTA DR1, but practical limitations on the localisation method mean the cadence and observation time have to be the same for each pulsar, so we average over them. After simulating signals for each of the previously selected sources, we adjust the total observation time and/or reduce the noise in each pulsar by a constant factor to set the S/N to the desired values 7, 10, 12 and 15 (see Tab. 4.2 for details). We choose 7 as the smallest S/N value because it ensures a confident detection according to the $\mathcal{F}$ statistic adopted by [22]; assuming a typical PTA and a false alarm probability of 0.001, a source with S/N $= 7$ has a detection probability of ≈ 0.9. For each setup, a likelihood $\mathcal{L}(A, \theta, \phi)$ is obtained using three different realisations of random white noise in the null stream pipeline. Summarising, we run a total of 36 simulations featuring:

- *three* different sources: A, B, C;
- *four* values of detection S/N $=$ 7, 10, 12, and 15;
- *three* independent white noise realisations.

	source	A		B		C	
S/N	% rms IPTA	T (yr)	ΔT (s) $\times 10^5$	T (yr)	ΔT (s) $\times 10^5$	T (yr)	ΔT (s) $\times 10^5$
7	100	12.8	2.12	10.7	2.03	12.2	2.76
10	80	21.3	2.71	16.0	2.33	12.2	2.11
12	80	29.8	2.64	26.7	2.69	18.4	2.20
15	70	34.0	2.52	32.0	2.70	24.5	2.54

Table 4.2: Adjustments made to the simulated IPTA-like array in order to fix S/N of the three injected sources A, B and C. The pulsar locations are kept the same as in IPTA DR1 [11], as are the relative rms noise levels of each pulsar. For S/N ≥ 10, the noise is decreased by a constant factor in all pulsars. The cadence ΔT and observation time T are averaged over for all pulsars. Then T is adjusted to set the S/N at specific values, keeping ΔT as close to the IPTA DR1 value as the null-stream localisation method allows.

The likelihood is evaluated on a three-dimensional grid in amplitude (A) and sky location (θ, ϕ). A is evenly sampled in log space between 10^{-17}–10^{-14}. The sky location parameters θ (polar coordinate from 0–π) and ϕ (azimuthal coordinate from 0–2π) are sampled over grid of equal area pixels, constructed with the HEALPIX algorithm [12] via HEALPY[1]. HEALPIX allows the user to define a grid refinement parameter n, which results in a number of pixels $N_{\rm pix} = 12n^2$. We choose $n = 32$, giving $N_{\rm pix} = 12288$ pixels of approximately equal area of 3.36 deg^2. For the likelihood calculation we use θ and ϕ at the mid-point of each pixel. The sky error-box, Ω_{90}, is then determined in the same way as was done for the previous study (see the start of Sec. 3.2). We repeat the computation of the likelihood inside Ω_{90} on a grid with the next level of refinement ($n = 64$).

4.3.3 Mock galaxy catalogue for host selection

Having determined $\mathcal{L}(A, \theta, \phi)$, we now need to draw a set of properties of potential hosts from a realistic galaxy population. We use a mock realisation of the observed sky extracted from the Millennium Run [111]. The simulation evolves dark matter particles over a volume $(500/h\,{\rm Mpc})^3$, reconstructing the clustering of dark matter halos. Semi-analytic galaxy formation models are then used to populate halos with galaxies, tracking their star formation, accretion and merger history. Although not 'state of the art', the large volume of the Millennium Run (683.7 Mpc side [111]), compared to more recent large scale, fully hydro-dynamical, simulations such as Illustris (105.6 Mpc side [115]) and EAGLE (100 Mpc side [116]), is relevant for our work. It ensures more statistical variation in the resulting galaxies, and in particular a better sampling of the high mass tail of the distribution, which is where the best candidate galaxies reside. We use the simulated sky maps from [110] that employ the semi-analytic model of [117], which has been shown to reproduce a number of observed properties of galaxies, including luminosity function, morphology and clustering.

The sky maps are flux-limited to $i < 21.0$ [see 110, for full details]. This results in galaxy catalogues that are complete down to stellar masses of $\approx 10^{11}{\rm M}_\odot$ at $z = 0.5$ and $\approx 4 \times 10^{11}{\rm M}_\odot$ at $z = 1$. We will show in Section 4.4 that all credible hosts are above these completeness limits. We downloaded all galaxies with stellar masses of $5 \times 10^{10}{\rm M}_\odot$ and higher at $z \leq 1$, which resulted in about 50 million objects. For each galaxy we store the bulge mass $M_{\rm b}$, the coordinates in the sky (θ, ϕ) and the apparent redshift z. The latter is then converted to $D_{\rm l}$ by assuming our fiducial cosmology (flat ΛCDM with $h_0 = 0.73$, $\Omega_{\rm M} = 0.25$). This information, together with a prior on the BBH mass ratio q and the aforementioned assumptions for the $M - M_{\rm b}$-relation, is all we need to perform the calculation outlined in Section 4.2.2.

To limit data size, only galaxies that fall within Ω_{90} are considered, which contain most of the

---

[1] healpy.readthedocs.io

relevant information. The simplifying assumption is made that one of the galaxies in Ω_{90} is the true source of the PTA signal, but there is – by definition – a 10% probability it falls outside the error-box. For each galaxy, the likelihood of being the GW source host is finally computed with Eq. 4.10, where $M_{\mathrm{b}}, D_{\mathrm{l}}, \theta$ and ϕ are fixed by the chosen galaxy and M, q, α and β are integrated over as described in Sec. 4.2.2.

4.4 Results and discussion

For each of the 36 experimental setups (Sec. 4.3.2), we use the null stream pipeline to obtain $\mathcal{L}(A, \theta, \phi)$ and determine Ω_{90}, the results of which we discuss here first in Section 4.4.1. Then, we perform the $p(d|G_i)$ calculation as described in Section 4.2.2 for each galaxy within Ω_{90}. From all $p(d|G_i)$, we obtain a cumulative likelihood distribution. These results are shown in Section 4.4.2.

4.4.1 sky localisation

First consider the behaviour of Ω_{90} with increasing S/N, shown in Fig. 4.3 for the three selected sources A, B and C. The expected trend $\Omega_{90} \propto \rho^{-2}$ is roughly followed by all sources, albeit not perfectly, due to the small numbers of performed simulations for each case. An exception is source A at S/N < 10, which shows a much steeper slope. Although this is consistent with the 'transition zone' identified in the previous study (Sec. 3.2.1) – signalling the S/N at which the data start to be informative – sources B and C do not behave the same way.

We conjecture that this can be explained fully by the specific position of the sources, relative to the pulsars (see Fig. 4.4). Source A is close to the location of the best pulsars. At the marginal detection level, the PTA combined S/N (Eq. 4.14) is mostly due to the contribution of these few, good pulsars (or possibly only one good pulsar). The other pulsars have very low individual S/N (Eq. 4.15). Therefore, the source is effectively only triangulated by one or a few pulsars, making localisation poor. At higher combined S/N (around 10), more pulsars contribute to the triangulation as their individual S/N increase. As such, there is a steep improvement in Ω_{90}, marking the transition to the $\Omega_{90} \propto \rho^{-2}$ regime. Conversely, sources B and C are relatively far away from the majority of the best pulsars. Hence a combined S/N$= 7$ already requires significant contributions from several different pulsars, making triangulation more effective. Sources B and C follow the standard S/N scaling, which source A joins as well, after the transition (where the shaded areas cross in Fig. 4.3).

Apart from the trend, the overall localisation precision of the three sources vary by a factor of

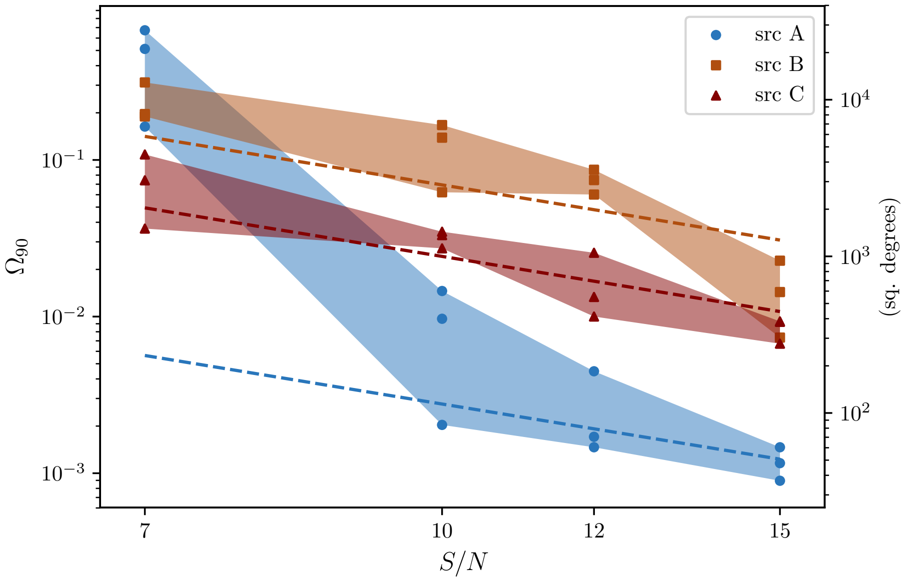

Figure 4.3: sky-localisation precision for the selected sources A, B, C at S/N $= 7, 10, 12$ and 15. Three markers at each S/N are the results of three runs with different white noise realisations. The dashed lines give the best fit of $\Omega_{90} \propto \rho^{-2}$ to the points at S/N≥ 10.

∼20 between them. This is due to both the inhomogeneous distribution of pulsars in the sky [20] as well as the different quality (noise level) of the pulsars in the arrays [75], which is expected to cause a difference in localisation. The best localisation, at high S/N, is achieved for source A, sitting in the 'sweet spot' of the array (where most of the pulsars, including the best ones, are). However, there is not simply a monotonic increase of Ω_{90} for sources further away, since the furthest source C has a better localisation than source B. This is also expected; sources that are antipodal to the 'sweet spot' are better localised than sources at a ∼90 deg angle, due to the shape of the PTA response functions [e.g. see Fig. 10 in 20].

A further investigation of this is visualised in Fig. 4.4. Here we inject a source with the same parameters as A at 192 different locations in the sky into white noise, using a synthetic IPTA-like array. The S/N is set to 12 everywhere, by scaling the amplitude of the GW signal. The map shows the resulting localisation Ω_{90} at each point. A dipolar structure of Ω_{90} is noticeable. Sources near the 'sweet spot' of clustered pulsars, and to a lesser extent, sources near the antipodal point, are localised better than those in between. This antipodal symmetry in the pulsar response functions is due to the quadrupolar nature of GWs, as was also shown by [20].

In any case, the huge scatter in Ω_{90} warns of a potential risk of an anisotropic sky coverage of PTA pulsars. Should the loudest resolvable GW sources be positioned at unfavourable locations, their detection, even at moderate S/N ≈ 12, would allow localisation accuracies of about 2000deg^2 only (an area containing ~2 million galaxies in our catalogue prior to selection). This means 'bad luck' could easily jeopardise any effort to identify a possible EM counterpart.

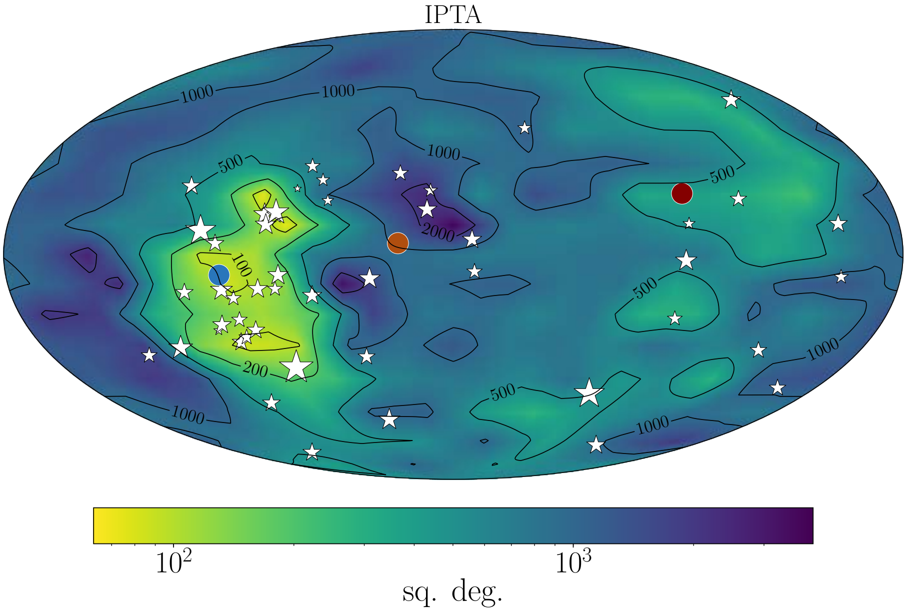

Figure 4.4: localisation capability of an IPTA-like array (based on DR1 [11]) for a source at fixed S/N = 12. This map is interpolated from 192 localisation values obtained by injecting a source at 192 locations forming a grid of equal sky area pixels (a HEALPIXgrid with $n = 4$ [12]). The IPTA pulsars are marked with stars, where the size of the star corresponds to the noise level of the pulsar (with bigger stars for lower noise). The circles indicate the positions of our selected sources A (blue, left), B (orange, middle) and C (red, right).

4.4.2 Host candidate population

Number of credible host candidates

Our main results consist of a set of $p(d|G_i)$ for the galaxies $\{G_i\}$ within Ω_{90} for each of the 36 experimental setups (Sec. 4.3). First, we compute the cumulative likelihood distribution from these $p(d|G_i)$. We then define N_x to be the minimum number of galaxies needed to sum to x% of the total likelihood $\sum_i p(d|G_i)$. In other words, by checking the first N_x best candidates, we expect to find the host galaxy among them x% of the time. Specifically, we look at N_{50} and N_{90} as proxies for the expected number of candidate host galaxies. An example cumulative distribution can be seen in

Fig. 4.5 for source A at S/N = 15 (the first random noise realisation). Within $\Omega_{90} \approx 60$ deg^2, there are ~1.2×10^5 galaxies $\{G_i\}$ in our mock catalogue, which would make detailed follow-ups for host identification impractical. After applying our technique, we have $N_{90} = 409$ and $N_{50} = 34$.

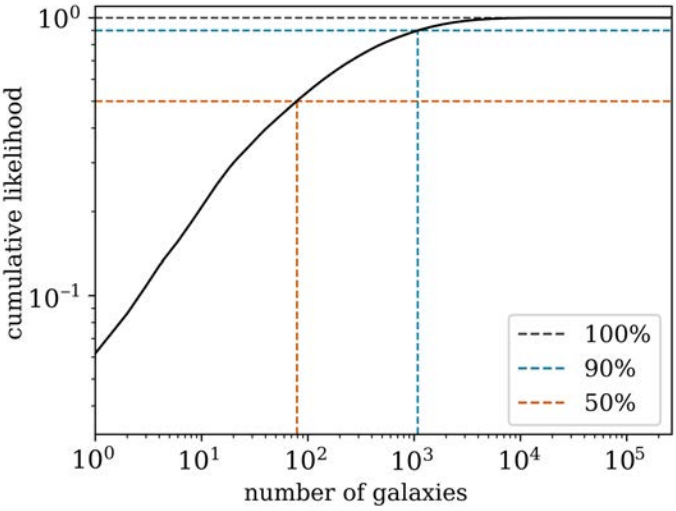

Figure 4.5: Cumulative likelihood of $p(d|G_i)$. The likelihood data d is for the IPTA setup (as described in the text) with source A at S/N = 15 (one of the random noise realisations). Vertical dashed lines identify the number of galaxies making up 50% (orange) and 90% (blue) of the total likelihood.

The collection of N_{50} and N_{90} of all experimental cases for which we obtained results can be found in Fig. 4.6. We fit two power laws $N_\mathrm{x} = c(\Omega_{90}/\Omega^*_{90})^p$, with parameters c, p (the power) and Ω^*_{90}, and x being either 50 or 90, by minimising the sum of squared differences between the predicted log values and the log of the data points. The best fitting powers are 0.64 for N_{50} and 0.65 for N_{90}. Naively one would expect a linear proportionality between Ω_{90} and the number of potential hosts, which we do not have, but there is significant scatter on the relation.

Tighter fits are obtained by treating the points for different injected sources separately, with best fit powers as in Table 4.3. These numbers show that fits to individual source data points are generally steeper and therefore closer to the expected linear dependence. The shallower global fit seems to be (partly) caused by the larger N_{50} and N_{90} for source A with respect to sources B and C at sky localisations of $\approx 300\,\mathrm{deg}^2$, which is not very surprising. Source A has S/N= 10 around this localisation area, while source B and C have higher S/N = 15. Consequently, N_{90} and N_{50} for source A includes galaxies with a lower bulge mass than for B and C, resulting in a larger N_{90} and N_{50}.

So while there is clearly a relation between the size of the sky error-box and the number of

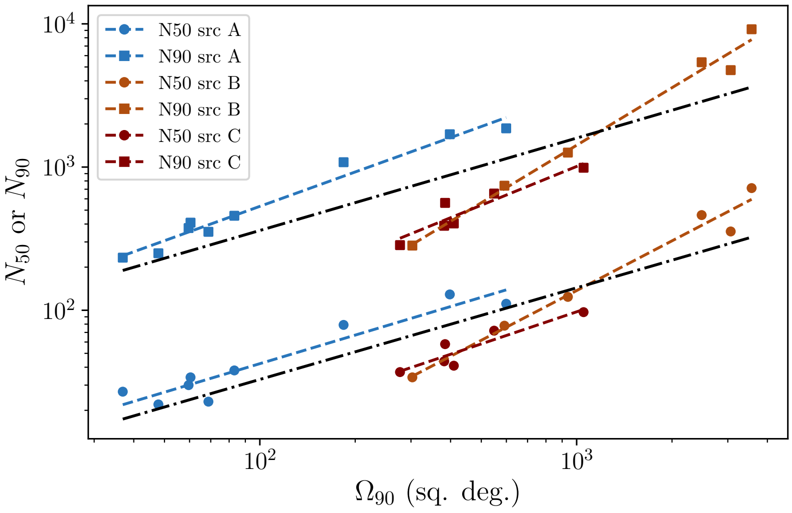

Figure 4.6: Number of candidate galaxies adding up to 50% (N_{50}, circle markers) or 90% (N_{90}, square markers) of the total likelihood to host the detected source, versus the sky-localisation area Ω_{90} for that detection. Results are shown for source A (blue) at S/N $= 10, 12$ and 15, and sources B (orange) and C (red) at S/N $=$ 12 and 15. For each S/N three noise realisations give a cluster of points at similar Ω_{90} values. Dashed lines show fitted power laws per source (see Table 4.3 for the best-fit powers). Dot-dashed lines are fits to N_{50} for all sources, with power 0.64, and to N_{90}, with power 0.65.

candidate host galaxies, significant scatter is due to the detailed source properties. Nonetheless, as a rule of thumb, we expect that for a resolvable PTA signal located in the sky with a precision of $\approx 100\,\mathrm{deg}^2$, we can identify few hundreds (few tens) galaxies that include the signal host with 90% (50%) confidence. Compared to all galaxies with stellar mass $> 5 \times 10^{11}\mathrm{M}_\odot$ at $z < 1$ falling in the error-box, these numbers restrict the pool of realistic hosts by nearly three (four) orders of magnitude, making realistic detailed follow-up campaigns feasible.

There are two points missing in Fig. 4.6, which are N_{90} and N_{50} for source A at S/N$=7$. This case was very poorly localised with $\Omega_{90} \approx 2.8 \times 10^4$deg.^{2} (67% of the sky). The expected number of

source	N_{50} power	N_{90} power
A	0.66	0.80
B	1.15	1.34
C	0.74	0.90
all	0.64	0.65

Table 4.3: Best fit powers for the power law fits to N_{50} and N_{90} as in Fig. 4.6. These are obtained by minimising the sum of squared errors on the log N_x values.

candidate hosts becomes very large, and also disobeys the trend discussed above. This is probably caused by $\mathcal{L}(\theta, \phi)$ not having a single peak in this case, which means the sky location of a potential host is much less informative and the list of candidates is barely restricted.

Host candidate sky distribution and clustering

Apart from the number of galaxies that make up a significant fraction of the likelihood $\sum_i p(d|G_i)$, we can also look at the properties of these galaxies. The parameters from the mock galaxy catalogue are $M_{\mathrm{b}}, D_{\mathrm{l}}, \theta, \phi$. First, we look at θ and ϕ by plotting the sky locations of the galaxies within N_{50} or N_{90} for the example case (source A at S/N$=15$), in Fig. 4.7. They are shown on top of the sky location likelihood $\mathcal{L}(\theta, \phi)$ of the injected source. The galaxies necessarily follow the localisation area, because we only used the galaxies that fall within Ω_{90}. Moreover, there is a relatively high concentration of N_{50} galaxies in the highest likelihood pixels. Hence, $\mathcal{L}(\theta, \phi)$ must contribute more to the selection of candidates than simply what we get from a straightforward cut using Ω_{90}.

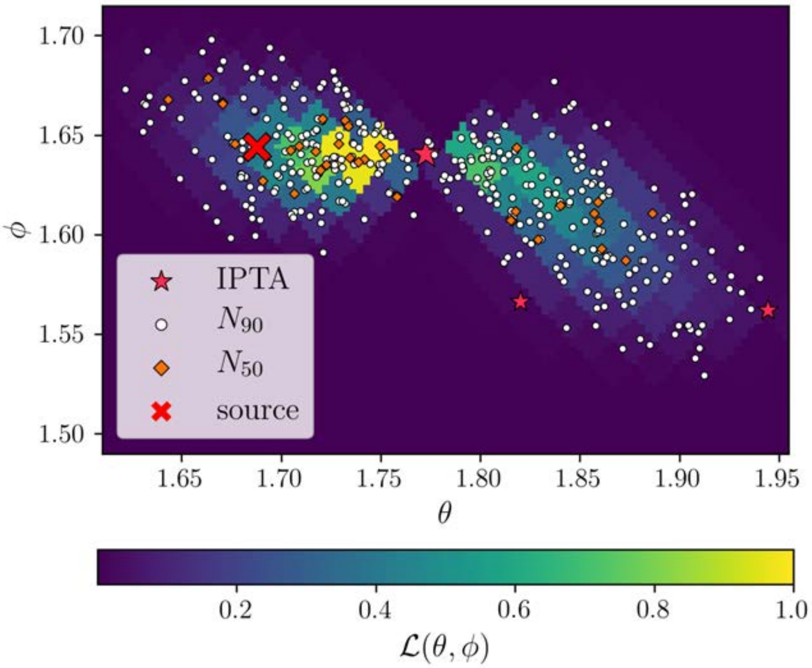

Figure 4.7: Locations of the best candidate host galaxies on top of the sky location likelihood for the injected source A (located at the red cross). The PTA has pulsar locations (pink stars) and relative noise levels of the IPTA DR1, but is adjusted such that the total S/N $= 15$ (see text). The 34 best candidates sum to 50% of the likelihood to be the host galaxy (N_{50} in orange diamonds) and an additional 375 sum to 90% (N_{90} in white circles). For this example, the $M - M_{\mathrm{b}}$-relation is marginalised over priors obtained from the literature (see text).

We further investigate this statement considering the clustering of good candidate galaxies – the N_{50} galaxies – for all of the experimental cases. Fig. 4.8 simultaneously shows a measure of the concentration of the localisation likelihood $\mathcal{L}(\theta, \phi)$, and of the concentration of N_{50} galaxies. The sky error-box Ω_{90} consists of a number of pixels $N_{\rm pix}$ that are sorted in descending $\mathcal{L}(\mathrm{pix}_i)$ order. Starting with an area $n_{\rm pix} = 1$ of one pixel (the best one), we iteratively increase $n_{\rm pix}$ by adding the next best pixel. On the x-axis of the figure, we plot the size of the included area relative to Ω_{90}, i.e. $n_{\rm pix}/N_{\rm pix}$. The concentration of $\mathcal{L}(\theta, \phi)$ in each area is also given by the fraction included in $n_{\rm pix}$, i.e. $\sum_i^n \mathcal{L}(\mathrm{pix}_i)/\sum_i^N \mathcal{L}(\mathrm{pix}_i)$. Compare this to the concentration of good candidate hosts, similarly defined as the fractional number of N_{50} galaxies in each area. The distributions are spread out, but there is no significant difference between the $\mathcal{L}(\theta, \phi)$ and N_{50} concentrations, i.e. the host probability follows the sky-localisation distribution. This confirms that it is valuable to include detailed localisation information when selecting candidate host galaxies, rather than only making a cut based on Ω_{90} (or on a different localisation criterion).

Host candidate mass and redshift

Second, we consider the other two parameters from the catalogue, the bulge mass $M_{\rm b}$ and luminosity distance $D_{\rm l}$. Fig. 4.9 shows their distribution among candidate hosts for the example case, where $D_{\rm l}$ has been converted into redshift. This figure best visualises the key idea behind our method. Since $A \propto M^{5/3}/D_{\rm l}$ – and there is a proportionality $M \propto M_{\rm b}^{\beta}$ and an almost linear proportionality between $D_{\rm l}$ and z at $z < 1$ – there is only a stripe in the mass–redshift plane defining the region of possible galaxy hosts. Moreover, since the first detection of a resolved PTA source will necessarily involve a very strong signal from a very massive binary system, this region lies at the highest masses. Due to the steep decay of the high mass end of the galaxy mass function, only few credible host candidates can be identified.

In the example shown, galaxies belonging to N_{50} or N_{90} are roughly bound by a line of slope $3/(5\beta)$ in the $\log D_{\rm l}(z) - \log M_{\rm b}$ plane (where β is the $M - M_{\rm b}$ constant marginalised over our prior), as expected by the GW amplitude scaling (Eq. 4.1). There is, however, a large mixing of galaxies with different likelihoods in this plane, due to their specific sky location. For example, there are a few very massive galaxies that nonetheless fall into the lowest 10% of the likelihood, which is due to an unfavoured sky position.

The N_{50} candidates are found across the whole range of redshifts, which is no surprise as the

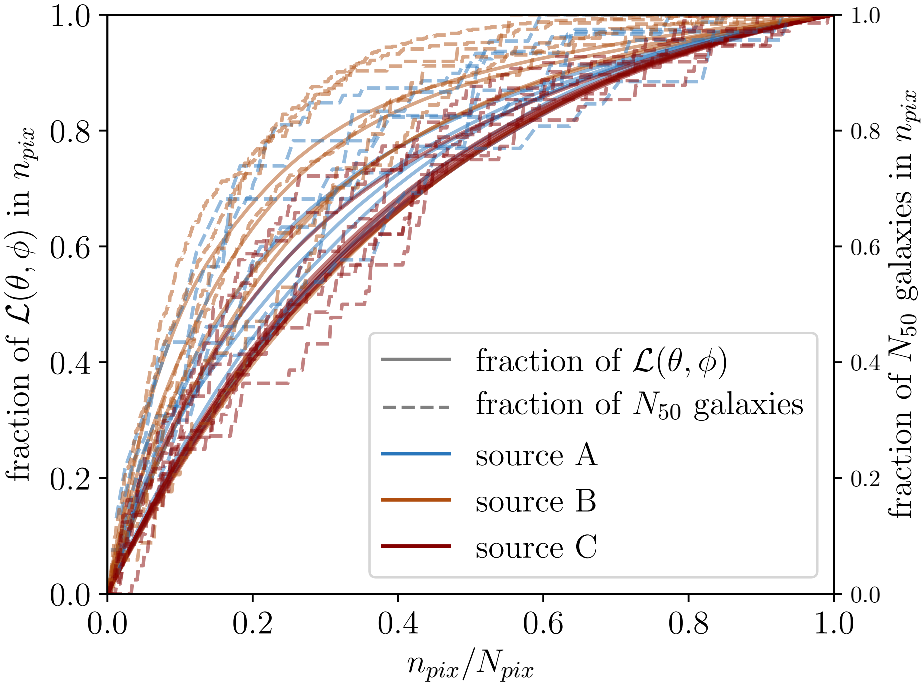

Figure 4.8: Comparison between the concentration of the sky-localisation likelihood and of the locations of good candidate host galaxies. With the fractional area of Ω_{90} on the x-axis, the fractional localisation likelihood in this area on the first y-axis (left, solid lines), and the fractional number of N_{50} galaxies on the second y-axis (right, dashed lines) (see text for details.) The quantities are normalised between 0 and 1 so that all experimental cases fit on the same scale. This plot includes all three injected sources (A in blue, B in orange and C in red), for S/N = 12 and 15.

redshift of the injected source A is 0.62 (Table 4.1). To explore further we look at the redshifts of candidate host galaxies for all injected sources and S/N values. Fig. 4.10 shows a number of histograms of z on a logarithmic scale. For each source A, B and C, the results of S/N = 12 and 15 (with three noise realisations per S/N), are combined. We make a comparison between the redshift distribution of the candidate galaxies pre-selected within the sky error-box (the filled histograms), and the N_{90} candidates selected with our method (the hatched histograms). The N_{90} candidates generally prefer lower redshifts. However, for all injected sources, there are a significant number of candidates at redshifts $z > 0.6$. Even though the injected redshift for source C ($z = 0.18$) is much lower than for sources A and B, the redshift distributions of candidate hosts differ only slightly, which reflects the fact that redshift is degenerate with mass in determining the GW amplitude.

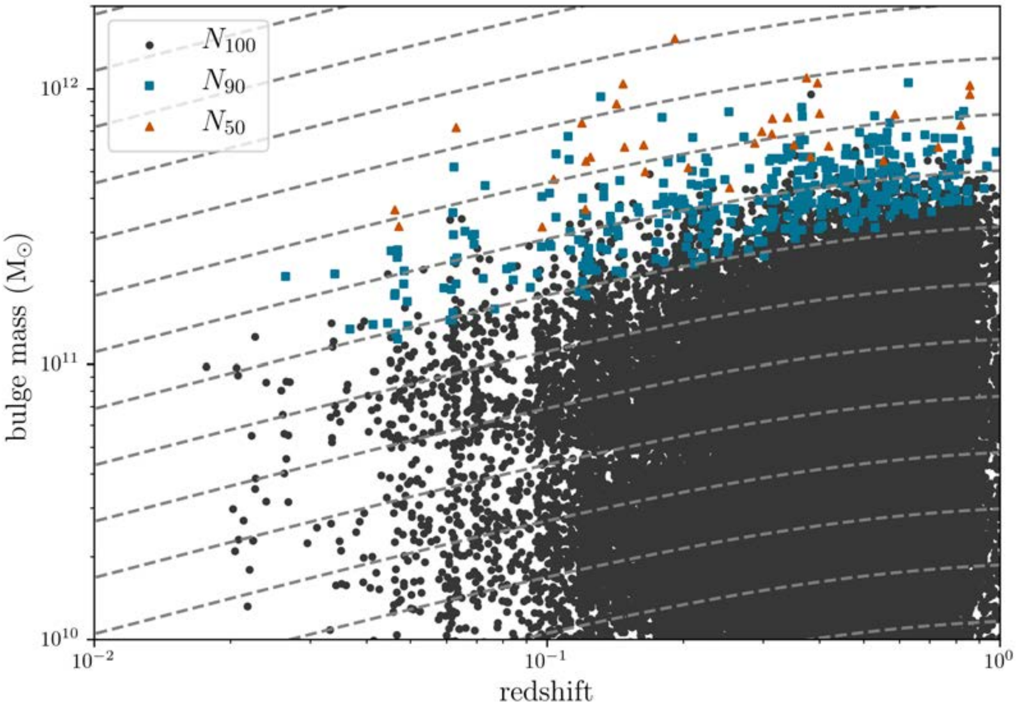

Figure 4.9: Distribution of bulge masses and redshifts of the candidate host galaxies of the example case; source A with S/N = 15. Blue squares mark galaxies that make up N_{90} and orange triangles mark the best candidates, which make up N_{50}. All other galaxies that fall within the sky-localisation error-box Ω_{90}, but form the lowest 10% of the total likelihood $\sum_i p(d|G_i)$, are marked with (dark) grey circles. Dashed grey lines are lines of constant GW amplitude (as in Eq. 4.1).

The turnover in the pre-selected number of galaxies seen in Fig. 4.10 at $z > 0.5$ is due to the $i = 21$ flux limit of the adopted galaxy catalogue, resulting in a severe incompleteness of lower mass galaxies at higher redshifts. Fig. 4.9 however, shows that typical galaxies belonging to N_{90} have $M_\mathrm{b} \gtrsim 2 \times 10^{11} \mathrm{M}_\odot$ at $z \approx 0.5$ and $M_\mathrm{b} \gtrsim 3 \times 10^{11}$ at $z \approx 1$. The catalogue is therefore complete in the relevant mass-redshift range for potential GW host galaxies. Fig. 4.10 also shows that the distribution of credible galaxy hosts is somewhat peaked around $z \lesssim 0.2$, even though two out of three of our selected signals (A and B) have $z \approx 0.6$. These sources were selected according to their sky location, so A, B and C are not an unbiased sample and are not necessarily representative of the actual redshift distribution of the first resolved PTA signals. There are several systems at $z > 0.5$ in the sample of 30 resolvable sources found in Sec. 4.3, and also [22] found that the peak of the expected first resolved PTA sources is at $z \approx 0.5$.

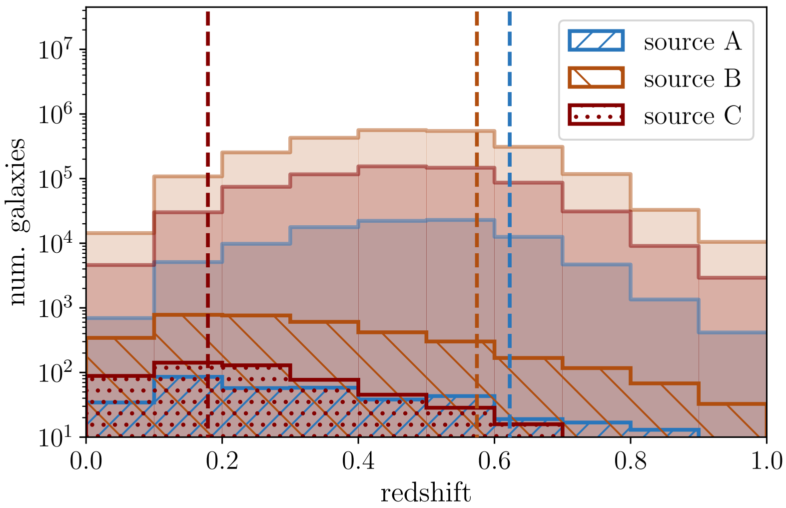

Figure 4.10: Logarithmic histogram of the redshifts of candidate host galaxies per source. The counts from the six experimental cases with S/N = 12 and 15, and three noise realisations, are averaged over. The foreground (hatched) histograms are N_{90} candidates, and the background (filled) histograms are all (i.e. N_{100}) candidates from the selected sky error-box. Injected redshift values for each source are indicated by a dashed line (see also Table 4.1).

High z sources are more common despite there being fewer potential host galaxies at such redshifts. This indicates that the likelihood of a galaxy to be a host is not only connected to its sky location and its position in the M–z plane, which are the factors considered in this work. The other key parameter is likely to be the absolute galaxy mass (regardless of redshift). There is evidence – both from observations and from cosmological simulations [see, e.g., results compiled in figure 1 of 118] – that the galaxy merger rate at low redshift is a strong function of the galaxy mass, with massive galaxies merging more often. Since the SMBH binary population simulated in Sec. 4.3 consistently takes this fact into account, the resulting population is naturally skewed towards high masses. Conversely, our host selection method only ranks galaxies based on the GW amplitude, determined by the combination of redshift and bulge mass, and therefore favours relatively more lighter galaxies at lower redshifts. These lower mass candidates are less likely to have undergone a major merger (and hence to host a GW source), compared to the few more massive candidates at higher redshifts. This suggests that combining our method with a (prior) probability of hosting a SMBH binary based on galaxy mass only [119, 120] can somewhat break the intrinsic mass–redshift degeneracy, further reducing the numbers of credible galaxy hosts.

4.4.3 Assumptions and approximations

Although simulations performed in this work are realistic in many aspects, some assumptions and choices had to be made to make their total runtime manageable.

Related to the galaxy mass

Several assumptions were made in the connection of the chirp mass of the GW source to the bulge mass of the host galaxy. First, we assumed a log-flat prior on $-2 \leq \log_{10}(q) \leq 0$, based on the broad q distribution of merging binaries found in cosmological simulations [121]. Although this is not necessarily representative of the q distribution of real SMBH binaries, we tested that different choices for the prior have only a minor impact on the results [see also 122, 123, 124].

Second, we did not consider errors in the measurements of galaxy M_b and D_l. The latter does not matter if a precisely measured spectroscopic redshift is available. In that case, estimates of D_l are only affected by galaxy peculiar velocities and uncertainties in the knowledge of the cosmological parameters, resulting in a negligible few % error. Conversely, there can be a significant uncertainty on the bulge mass, or on D_l if a photometric redshift measurement is used. This is likely to impact our results, spreading the host probability distribution and thus returning more host candidates. Some tests on a limited number of setups found that including an uncertainty of a factor of two on the galaxy bulge mass results in roughly a factor of two more candidate hosts galaxies.

Last, we marginalised over the uncertainty in the $M - M_\mathrm{b}$-relation. Assuming a specific $M - M_\mathrm{b}$-relation instead can affect our results, especially if the relation predicts relatively higher or lower black hole masses than the marginalised relation. As an example, we ran some test cases assuming the $M - M_\mathrm{b}$-relation from [56], which associates relatively higher black hole masses with the given galaxy bulge mass. The number of candidate host galaxies in these cases is increased by a factor ranging between ~ 3 and ~ 8 with respect to the marginalised $M - M_\mathrm{b}$ case. Conversely, for a 'pessimistic' $M - M_\mathrm{b}$-relation such as [57] – which predicts relatively lower black hole masses especially for high-mass galaxies – the number of candidates is a factor ~ 2 to ~ 4 lower.

Limiting the computational cost

Due to computational limitations, we ran a relatively small number of simulations. Although we checked robustness of the results against the specific noise realisation, we only picked one sky location for each source. This may make cosmic variance a factor in the determination of the number of galaxy hosts. To test this, for a selected GW source, we performed some simulations with rotations of the Millennium sky, and counted N_{50} and N_{90} for each of them. Although numbers vary, the scattering is consistent with that observed in Fig. 4.6.

An important assumption of our method is that the true host of the detected GW signal is present in the galaxy catalogue. This is guaranteed only for complete catalogues. Real catalogues based on observations never are, and the simulated catalogue from [110] reflects this by selecting galaxies based on observational criteria. This results in a number of missing galaxies – more towards higher redshifts. However, for the most part these are the small galaxies (which are more difficult to observe), and those are not relevant host candidates. Since at redshifts $z \lesssim 1$ only the most massive galaxies are selected in N_{90} (see Fig. 4.9), this is unlikely to affect the results for N_{90} and N_{50}, but it is still a possible source of error. As there are good candidates up to $z = 1$, it is also possible there are a small number of potential N_{90} galaxies at $z > 1$ that were not included.

Finally, it should be kept in mind that we selected the 90% sky location credible region. By selecting N_{50} and N_{90} in this region, the actual probability to find the true host in these sets is $0.9 \times 0.5 = 0.45$ and $0.9 \times 0.9 = 0.81$, respectively.

4.5 Conclusion

In this chapter, we proposed a novel methodology to select host galaxy candidates of the first individual gravitational wave sources observed by pulsar timing arrays. Since PTA source localisation is expected to be of several tens of square degrees at best, up to a few million galaxies might end up in the sky error-box. Classifying the most promising host candidates is therefore paramount to increase the chances of true host identification via dedicated follow-ups. Our method exploits the GW strength dependence on chirp mass and distance, together with empirical SMBH mass–host galaxy correlation, to rank galaxies in the mass–redshift plane. This selection is combined, in a Bayesian way, with the sky-localisation information which we obtain with the null-stream method from Chapter 3 (but could work with any method of obtaining the likelihood on the sky position and GW amplitude). Each galaxy is assigned a probability to be the host of the detected GW signal.

To test our method, we performed realistic simulations by drawing GW sources from detailed SMBH binary population models based on observed merging galaxies, by employing the actual IPTA pulsar sky locations and rms values to build the array, and by selecting host candidates based on formation and evolution models. We considered different GW source sky positions and detection S/N and investigated the ensemble of credible host galaxy candidates. In particular, we defined N_{50} and N_{90} to be the smallest numbers of galaxies having a collective 50% and 90% chance of being the true host of the GW source respectively, assuming the true host is among the prior selection of candidates. Our key results can be summarised as follows:

- N_{50} and N_{90} are respectively nearly four and three orders of magnitude smaller than the number of galaxies obtained with a straightforward cut: stellar mass $M_* > 5 \times 10^{10}\mathrm{M}_\odot$, redshift $z < 1$, and found in the 90% confidence sky location region Ω_{90};
- N_{50} and N_{90} should roughly be proportional to Ω_{90}. We find a sub-linear proportionality, although with large scatter;
- despite the large scatter, a useful rule of thumb is that for $\Omega_{90} = 100\mathrm{deg}^2$, $N_{50} \lesssim 50$ and $N_{90} \lesssim 500$;
- although the distribution of potential hosts peaks around $z < 0.2$, it has a long tail that extends up to $z \lesssim 1$.

Our methodology can therefore effectively select the most likely host galaxy candidates, which might have a major impact on future multi-messenger observations of SMBH binaries. For typical PTA sky-localisation precision of hundreds of deg^2, instead of following up millions of galaxies, we can choose to accept the risk of missing the true host with 55% (19%) probability and monitor only the ≈ 100 (1000) most promising ones. There is significant uncertainty on these numbers however, mainly due to the uncertainty in the $M - M_\mathrm{b}$-relation (see Section 4.4.3).

The applicability of our method obviously relies on the availability of photometric and spectroscopic data from all-sky surveys necessary to identify potential galaxy hosts and to estimate their stellar (and bulge) masses. Since the most credible galaxy candidates are necessarily very massive (and/or particularly nearby), relatively shallow surveys are sufficient for this scope. Catalogues from SUSS [125, covering $\approx 1/4$ of the sky], Pan-STARRS [126, $\approx 3/4$ of the sky], LSST [127, $\approx 1/2$ of the sky] and Gaia [128, all sky] will provide enough imaging, photometric and (possibly) spectroscopic information for reliable mass estimates via, for example, spectral energy distribution fitting [see e.g. 112, 129].

A positive host identification chance increase of less than a factor of two comes at the expense of following up a factor of ten more galaxies. The follow-up strategy can therefore be optimised based on the future number of resolved PTA sources and on available observing facilities. Reducing the number of credible hosts is critical mostly because our knowledge of SMBH binary signatures is poor [see, e.g. 104]. One therefore has to collect all possible hints to build up confidence that the true host has been found. This might require multiple studies out of, for example, photometric and spectroscopic optical and IR follow-up of the candidates to unveil any observational hint of an accreting binary; deep-field imaging to look for features such as stellar tails and shells that indicate a recent merger [e.g. 130]; integral field spectroscopy to identify the presence of a 'dry' SMBH binary via kinematic signatures in the stellar distribution [131]; deep X-ray observations to unveil the presence of an obscured AGN and its possible high energy signatures [132]; and many more.

The upcoming ELT [133] and JWST [134] will be particularly suited for the optical and near infrared follow-ups mentioned above, whereas the X-ray satellite Athena [135] can potentially survey the 100 most probable hosts within less than 1 day of observation time. Clearly, the fewer the candidates, the more extensive the follow-up campaign can be, thus enhancing the chances of a positive detection. Archival data can also be used to identify hints of, e.g, periodic variability matching the frequency of the GW source. This can be done in the optical and, possibly, in X-ray with LSST and eROSITA [136] archival data, respectively.

Finally, the mismatch between the credible host redshift distribution identified with our method and the expected distribution of the first PTA sources predicted by [22] indicates that a more efficient galaxy host selection can be performed when the mass-dependent galaxy merger probability is folded into the calculation [see also 120]. By doing so, the mass–redshift degeneracy intrinsic in our method might be alleviated, further decreasing the number of credible hosts. Our method is suitable for such an extension, possibly by including the galaxy merger probability as a prior in the Bayesian computation.

Chapter 5

Null streams for realistic PTA data and a model-independent localisation method

> *"Given current theoretical models, it is likely that ultra-low frequency GWs will be detected by pulsar timing experiments within 5-10 yr. "*
>
> IPTA scientists in 2010 [137]

The null-stream localisation method from Chapter 3 needs to be adapted before it can be used on real PTA data (also discussed in Sec. 3.4). The power of the null-stream transformation to separate the dependency on the GW model and on the source sky location is also not fully utilised in the likelihood Eq. 3.9, which is still equivalent to a "standard" likelihood such as Eq. 3.8 (albeit written in a different basis). In this chapter, we make a start on these outstanding tasks. First, we address the main problem with creating null streams from real PTA data, which is that it requires coincident samples from all pulsars. A possible solution is to work in the frequency domain (FD) instead of the time domain (TD), which is discussed in Sec. 5.1.1. Second, we introduce the null streams in Sec. 5.2, which requires combining the residuals from all pulsars (Sec. 5.2.1). Finally, in Sec. 5.2.3, we we construct a GW-model-independent likelihood by marginalising over the null-stream elements that contain the GW polarisations.

This work is still in progress but some preliminary results show that there is potential for these methods, although it remains uncertain whether they can function well together. A selection of interesting results is presented and discussed in Sec. 5.4. Outstanding issues to applying the null-stream method to real PTA data, as well ideas to improve or expand upon it, are discussed in Sec. 5.5.

5.1 Handling realistic PTA data

To apply the null-stream transformation $\mathcal{N}$ from Eq. 3.6, one requires an input data point from all pulsars observed at the same time, or sampled at the same frequency. Previously in Chapter 3, we simulated idealised PTA data that was sampled with a constant cadence and at the same times for all pulsars. This allowed us to easily work in the FD and even use only one data point from each pulsar (at the frequency of the injected signal). In reality, PTA data is very different. First, observations are not taken simultaneously for all pulsars (this is not even possible with telescopes that point to a specific region of the sky at a time). Second, sampling is not regular; although there might be some periods in which a particular pulsar is observed very regularly, there are still days to weeks between observations, so the cadence is never perfectly constant. Third, some pulsars have been observed since much earlier dates than others, so the first observation date and total span of observations varies greatly between them. Last, gaps in the data are not uncommon. Observations are sometimes paused, for example when a telescope is being upgraded or repaired.

To describe a realistic PTA data set, we label the set of residuals from pulsar p as $r^{(p)}(t_j) \equiv r_j^{(p)}$, with $j \in \{1, 2, \ldots, \mathrm{N}^{(p)}\}$[1]. As described above, the times t_j can have varying cadence and the total number of samples, $\mathrm{N}^{(p)}$, can vary between pulsars (an example of simulated sampling times is shown in Fig. 5.1). We assume that the residuals contain a GW signal s and Gaussian noise n:

$$r_j^{(p)} = s_j^{(p)} + n_j^{(p)}, \tag{5.1}$$

where the signal is given by the PTA response from Eq. 2.6:

$$s_j^{(p)} = F_{(p)}^{+}(\hat{\Omega}) H^{+}(t_j; \vec{\lambda}) + F_{(p)}^{\times}(\hat{\Omega}) H^{\times}(t_j; \vec{\lambda}). \tag{5.2}$$

Here, $\hat{\Omega}$ is the GW propagation direction. The time-integrated GW polarisations H^{+} and $H^{\times}$ are assumed to be given by a model with parameters $\vec{\lambda}$. The noise is given by a Gaussian distribution:

$$\vec{n} \sim \mathcal{N}(0, \Sigma^{(p)}), \tag{5.3}$$

with covariance matrix $\Sigma^{(p)}$. Realistic PTA noise modelling does not fall within the scope of this project; here we simply include only Gaussian noise that is independent in each TD residual, so that

$$\Sigma_{jk}^{(p)} = \langle n_j^{(p)} n_k^{(p)} \rangle = \sigma_j^{(p)} \delta_{jk}. \tag{5.4}$$

The noise level $\sigma_j^{(p)}$ at time t_j is given by the rms of the residuals for pulsar (p). The rms is usually constant for a short stretch of data, but it can change during the course of PTA observations. For example, after a telescope upgrade the overall noise level may go down. In our simulations later on, we use constant rms but the following logic applies in either case. We also assume the noise is not correlated between pulsars, i.e.:

$$\langle n_j^{(p)} n_k^{(q)} \rangle = \Sigma_{jk}^{(p)} \delta_{pq}. \tag{5.5}$$

Standard likelihood

Before we move onto rewriting the likelihood in the FD and later in terms of null streams in Sec. 5.2, it is useful to write our "starting point" likelihood in terms of the chosen notation. For a data set $\vec{d}$ which is the collection of all pulsars' residuals $r_j^{(p)}$ as described above, the "standard" log of the likelihood $p(\vec{d}|\hat{\Omega}, \vec{\lambda}, \mathcal{H}_{\mathrm{sig}})$ is:

$$\log(\mathcal{L})(\hat{\Omega}, \vec{\lambda}) = \sum_{p=1}^{\mathrm{P}} \left(-\frac{1}{2} \left(x_j^{(p)} (\Sigma_{(p)}^{-1})_{jk} x_k^{(p)} \right) - \frac{1}{2} \mathrm{N}^{(p)} \log(2\pi) - \frac{1}{2} \log\left(|\Sigma^{(p)}|\right) \right). \tag{5.6}$$

---

[1] The label (p) is not a vector index, but a superscript (or subscript, if more convenient in the notation) indicating a property of $\tilde{r}$, namely, which pulsar $\tilde{r}$ belongs to. To distinguish it from indices such as j, we write (p) in brackets.

This equation is the equivalent of Eq. 3.8 in Chapter 3, for our extended notation. The quantity $\vec{x}$ is defined as $x_j^{(p)} \equiv r_j^{(p)} - s_j^{(p)}(\hat{\Omega}, \vec{\lambda})$. $|\Sigma^{(p)}|$ is the determinant of the covariance matrix for pulsar p. Although the GW model that goes into $s_j^{(p)}$ is shared across pulsars, the likelihood computation treats each of them separately, summing (in log) over the contribution from each pulsar.

5.1.1 Fourier transform

We wish to introduce the null-stream transformation $\mathcal{N}$ into the likelihood, analogous to Eq. 3.9 in Chapter 3. To do so, first we need to "line up" the pulsars' data points, as $\mathcal{N}$ is always applied to a vector of samples at equal time or equal frequency from all pulsars. In principle, this could be done by interpolating the data in the TD. However, this does not deal with unequal start times of gaps in the pulsar timing data. It is therefore more practical to transform data into the FD. If the same frequencies are used for each pulsar, the FD data are naturally lined up. Additionally, one can reduce the number of data points by choosing a set of frequencies that are fewer than the typical number of residuals (a few hundred).

For unevenly sampled data, the often-used Fast Fourier Transform (FFT) algorithm [13] does not work. Instead, we can straightforwardly extend the definition of the discrete Fourier transform (DFT) to work with variable cadence, and compute it directly. Given frequencies labelled by α, the FD residuals $\tilde{r}_\alpha^{(p)} \equiv \tilde{r}^{(p)}(f_\alpha)$ are given by:

$$\tilde{r}_\alpha = \sum_j \mathcal{F}_{\alpha j} r(t_j). \tag{5.7}$$

. The matrix $\mathcal{F}_{\alpha j}$ is defined as[2]:

$$\mathcal{F}_{\alpha j} = w_j \exp\left(-2\pi \mathrm{i} f_\alpha t_j\right). \tag{5.8}$$

The weights w_j take on the role of the cadence Δt, but are slightly adjusted to account for the difference in the time step before and after a sample at time t_j:

$$w_j = \begin{cases} t_2 - t_1 & \text{if } j = 1 \\ t_\mathrm{N} - t_{\mathrm{N}-1} & \text{if } j = \mathrm{N} \\ \frac{1}{2}(t_{j+1} - t_j) + \frac{1}{2}(t_j - t_{j-1}) & \\ \quad = \frac{1}{2}(t_{j+1} - t_{j-1}) & \text{otherwise.} \end{cases} \tag{5.9}$$

---

[2] This is the same symbol as used previously for the matrix of pulsar response functions used to construct the null-stream matrix $\mathcal{N}$. Here, $\mathcal{F}$ always refers to the Fourier transform matrix as the former is not needed any more.

If the data are evenly sampled, all weights are equal to Δt, and this general *extended Fourier transform* (EFT) reduces to the standard DFT.

Choice of frequencies

The set of frequencies f_α used in the EFT can be chosen freely. However, not every choice results in a good representation of the information in the TD data. The Nyquist frequency of irregularly sampled data is not as straightforwardly defined as for evenly sampled data. It is determined by the longest time step $\mathrm{d}t$, such that all observed times can be written as $t_0 + k\,\mathrm{d}t$, where k is an integer. Then $f_{\mathrm{Ny}} = 1/(2\mathrm{d}t)$ [53, and references therein]. This f_{Ny} easily becomes very large if the observed times are sufficiently random (which is the case for PTA data).

There is another restriction on the highest possibly frequency: One can only measure frequencies up to $1/2t_{\mathrm{int}}$, where t_{int} is the integration time of one observation [53]. Determining one pulsar TOA takes about 30 minutes (or around that time scale, at least), which would result in $f_{\max} \approx 5 \times 10^{-4}$Hz. This is much higher then the frequency of any expected resolvable PTA signals [22, especially Fig. 6], therefore, we simply choose an astrophysically motivated $f_{\max} = 10^{-7}$Hz. This is also how we can justify using fewer data points in the FD than were available in the TD; we are excluding samples outside of the frequency range of interest.

The minimum measurable frequency is determined by the total time of the observations. Let T be the time between the earliest and latest TOA in any of the pulsars, then $f_{\min} = \frac{1}{T}$. We choose the frequencies to be evenly sampled with a step equal to $f_{\min}$, up to the highest frequency lower than the cut-off $f_{\max}$. (The precise value of the highest frequency should not matter, since we choose $f_{\max}$ to be higher than any frequencies we expect a signal to be at. If signals get close to $f_{\max}$, it should be increased). The same set of $\mathrm{N_f}$ frequencies are used for all the pulsars. This is essential to achieve the coincident data points that are necessary for constructing null streams.

Fourier domain likelihood

Having determined the FD residuals $\tilde{r}_\alpha^{(p)}$ above, the likelihood from Eq. 5.6 can be expressed in the FD as follows (using the Einstein summation convention for repeated indices α and β):

$$\log(\mathcal{L})(\hat{\Omega}, \vec{\lambda}) = \sum_{p=1}^{\mathrm{P}} \left(- \left((y^{(p)})^*_\alpha \, (\tilde{\Sigma}^{-1}_{(p)})_{\alpha\beta} \, y^{(p)}_\beta \right) - \mathrm{N_f} \log{(2\pi)} - \log{(|\tilde{\Sigma}^{(p)}|)} \right). \tag{5.10}$$

Here, $\vec{y}$ is the EFT of $\vec{x}$, i.e. $y_\alpha^{(p)} = \tilde{r}_\alpha^{(p)} - \tilde{s}_\alpha^{(p)}$, and $(\vec{y})^*$ is its complex conjugate. The FD covariance matrix $\tilde{\Sigma}^{(p)}$ is obtained by transforming the TD covariance $\Sigma^{(p)}$:[3]

$$\langle y_\alpha^{(p)} y_\beta^{(p)} \rangle = \mathcal{F}_{\alpha j} \Sigma_{jk}^{(p)} \mathcal{F}_{\beta k} \equiv \tilde{\Sigma}_{\alpha\beta}^{(p)}. \tag{5.11}$$

There is an implicit sum over the frequencies in the term $(y^{(p)})_\alpha^* \, (\tilde{\Sigma}_{(p)}^{-1})_{\alpha\beta} \, y_\beta^{(p)}$. Because we have chosen to only use positive frequencies, the equivalent – because the residuals are real – contribution from the negative frequencies is not included. This is compensated by a factor of two w.r.t. Eq. 5.6 (which cancels the factor $\frac{1}{2}$). The normalisation terms are also doubled, as they should be included once for the real and once for the imaginary part of the complex variable $\vec{y}$.

5.2 Null stream likelihood

5.2.1 Concatenating the residuals from all pulsars

Index rule

At this point, it is possible to apply the null-stream transformation to a vector containing one residual from each pulsar, at a given frequency, as given in Eq. 3.6. Because we previously considered each pulsars' residuals separately, it is useful to first combine them into a single vector. This is done by concatenating them as follows:

$$\tilde{r}_A \equiv \mathrm{cat}(\tilde{r}_\alpha^{(p)}) = (\tilde{r}_1^{(1)}, \tilde{r}_2^{(1)}, \ldots, \tilde{r}_{\mathrm{N_f}}^{(1)}, \tilde{r}_1^{(2)}, \ldots, \tilde{r}_{\mathrm{N_f}}^{(2)}, \ldots, \tilde{r}_{\mathrm{N_f}}^{(\mathrm{P})}). \tag{5.12}$$

A concatenated vector is indicated with a capital index, $A = 1, 2, \ldots, \mathrm{PN_f}$. It has the following conversion rule associated:

$$A = (p-1)\mathrm{N_f} + \alpha \quad \Leftrightarrow \quad p = \lfloor A/\mathrm{N_f} \rfloor + 1, \quad \alpha = A \bmod \mathrm{N_f}. \tag{5.13}$$

Concatenation following this rule means all residuals from a single pulsar are grouped together, in order of increasing frequency. For more compact notation later on, we define the following shorthand for the indices derived from a concatenated vector index A:

$$A' \equiv \lfloor A/\mathrm{N_f} \rfloor + 1 = p(A) \tag{5.14}$$

$$\bar{A} \equiv A \bmod \mathrm{N_f} \;\; = \alpha(A). \tag{5.15}$$

---

[3] If the pulsar noise levels σ_j^p from Eq. 5.4 are time-dependent, the noise becomes correlated in the FD, i.e. $\tilde{\Sigma}$ will not be diagonal.

FD likelihood in terms of concatenated vectors

Other vectors are concatenated according to the same rule from Eq. 5.13. This also means that the previously defined $y_\alpha^{(p)} = \tilde{r}_\alpha^{(p)} - \tilde{s}_\alpha^{(p)}$ translates simply to $y_A = \tilde{r}_A - \tilde{s}_A$. Explicitly writing α and (p) as functions of A allows to find the covariance matrix for the concatenated quantity[4]:

$$\begin{aligned} y_A &= y_{\bar{A}}^{(A')} \\ &= \sum_j \mathcal{F}_{\bar{A}j}\, x_j^{(A')}, \end{aligned} \tag{5.16}$$

therefore

$$\begin{aligned} \langle y_A y_B \rangle &= \sum_{j,k} \mathcal{F}_{\bar{A}j}\, \langle x_j^{(A')} x_k^{(B')} \rangle\, \mathcal{F}_{\bar{B}k} \\ &= \tilde{\Sigma}_{\bar{A}\bar{B}}^{(A')}\, \delta_{A'B'} \equiv \Psi_{AB}. \end{aligned} \tag{5.17}$$

Although the expression for Ψ_{AB} may seem messy, it is actually quite simply a block-diagonal matrix built from the individual pulsar $\tilde{\Sigma}^{(p)}$. Its inverse is therefore also block-diagonal, constructed from $\tilde{\Sigma}_{(p)}^{-1}$:

$$\Psi_{AB}^{-1} = \left(\tilde{\Sigma}_{(A')}^{-1}\right)_{\bar{A}\bar{B}} \delta_{A'B'}. \tag{5.18}$$

We can now write the FD likelihood Eq. 5.10 as follows (with the Einstein summation written explicitly):

$$\log(\mathcal{L})(\hat{\Omega}, \vec{\lambda}) = \sum_{A,B=1}^{\mathrm{PN_f}} -\left(y_A^\dagger \Psi_{AB}^{-1} y_B\right) - \mathrm{PN_f} \log(2\pi) - \log(|\Psi|). \tag{5.19}$$

To take the normalisation out of the sum, we have used that $|\Psi| = \prod_p |\tilde{\Sigma}^{(p)}|$.

5.2.2 Introducing the null streams

Null stream from concatenated residuals

The null-stream transformation is given in Eq. 3.6 for a single vector of points from all pulsars. In our current notation, this means that for a single frequency α, we have:

$$\sum_p \mathcal{N}_{kp} \tilde{r}_\alpha^{(p)} = \sum_p \mathcal{N}_{kp} (\tilde{s}_\alpha^{(p)} + \tilde{n}_\alpha^{(p)}) = \zeta_\alpha^{(k)} + \mathcal{N}_{kp} \tilde{n}_\alpha^{(p)}. \tag{5.20}$$

---

[4] Because the concatenation-rule expressions are to be interpreted as a single index, they are written in brackets. To keep consistent with writing pulsar labels in brackets as well, some of these expressions will have double brackets.

The dependencies on the GW model and propagation direction $\hat{\Omega}$ are not written explicitly here, but the signal $\tilde{s}_\alpha^{(p)}$ depends on both (see Eq. 5.2). The null-stream matrix $\mathcal{N}$ is constructed assuming some $\hat{\Omega}'$. If and only if these are the same $\hat{\Omega} = \hat{\Omega}'$, the vector $\zeta_\alpha^{(k)}$ has the familiar null-stream properties (the label (k) now indicates the null-stream number; as with the pulsar label, there are P in total):

$$\zeta_\alpha^{(k)} = \begin{cases} \tilde{H}_\alpha^{+}(\vec{\lambda}) & \text{if } k = 1 \\ \tilde{H}_\alpha^{\times}(\vec{\lambda}) & \text{if } k = 2 \\ 0 & \text{otherwise.} \end{cases} \quad . \tag{5.21}$$

It will also be useful to define the null-stream-transformed residuals:

$$\xi_\alpha^{(k)} \equiv \sum_p \mathcal{N}_{kp} \tilde{r}_\alpha^{(p)}, \tag{5.22}$$

and the null-stream equivalent of the "data minus model" quantity y:

$$\chi_\alpha^{(k)} \equiv \xi_\alpha^{(k)} - \zeta_\alpha^{(k)} = \sum_p \mathcal{N}_{kp} y_\alpha^{(p)}. \tag{5.23}$$

In order to find the correct null-stream transformation for the concatenated vectors, we can use the same trick as in Eq. 5.16 (using the Einstein summation convention here, and in Eq. 5.25):

$$\begin{aligned} \chi_A &= \chi_{\bar{A}}^{(A')} \\ &= \mathcal{N}_{A'B'}\, y_{\bar{B}}^{(B')} \delta_{\bar{A}\bar{B}} \\ &= \mathcal{N}_{A'B'} \delta_{\bar{A}\bar{B}}\, y_B \\ &\equiv M_{AB}\, y_B. \end{aligned} \tag{5.24}$$

Although the "big" null-stream matrix M is similarly defined to the "big" covariance matrix Ψ, this one is not block-diagonal. It is actually built from P "small" $\mathcal{N}$ matrices in such a way that, had we chosen the concatenation rule differently (with residuals from one frequency of all pulsars grouped together instead of the other way around), M would have been diagonal and Ψ would not. It is, however, relatively straightforward to show that the inverse M^{-1} can be constructed by using $\mathcal{N}^{-1}$ instead of $\mathcal{N}$, similar to Eq. 5.18:

$$M_{AB}^{-1} M_{BC} = (\mathcal{N}^{-1})_{A'B'} \delta_{\bar{A}\bar{B}} \mathcal{N}_{B'C'} \delta_{\bar{B}\bar{C}} = \delta_{AC}. \tag{5.25}$$

Null stream likelihood

We wish to write the standard FD likelihood in terms of null streams, analogously to Eq. 3.9 in Chapter 3. Starting from Eq. 5.19, this can be done by substituting the following expression which makes use of Eq. 5.25:

$$y_A = \delta_{AB} y_B = M^{-1}_{AC} M_{CB}\, y_B = M^{-1}_{AC}\, \chi_C. \tag{5.26}$$

The log likelihood, now without the constant normalisation terms, then becomes:

$$\begin{aligned}\log(\mathcal{L})(\hat{\Omega}, \vec{\lambda}) &= \sum_{A,B} -\left(\sum_C M^{-1}_{AC}\, \chi_C\right)^{*} \Psi^{-1}_{AB} \left(\sum_D M^{-1}_{BD\,\chi_D}\right) \\ &= \sum_{A,B,C,D} -\chi^{*}_C M^{-1}_{AC} \Psi^{-1}_{AB} M^{-1}_{BD} \chi_D. \end{aligned} \tag{5.27}$$

We can define

$$Z_{CD} \equiv \sum_{A,B} M_{AC} \Psi_{AB} M_{BD}, \tag{5.28}$$

to find

$$\log(\mathcal{L})(\hat{\Omega}, \vec{\lambda}) = \sum_{C,D} -\chi^{*}_C Z^{-1}_{CD} \chi_D. \tag{5.29}$$

As a reminder of the dependencies on the GW model $\vec{H}(\vec{\lambda})$ and propagation direction $\hat{\Omega}$, the likelihood can also be written more explicitly as:

$$\log(\mathcal{L})(\hat{\Omega}, \vec{\lambda}) = \sum_{A,B} -\left(\xi_A(\hat{\Omega}) - \zeta_A(\vec{\lambda})\right)^{*} Z^{-1}_{AB}(\hat{\Omega}) \left(\xi_B(\hat{\Omega}) - \zeta_B(\vec{\lambda})\right). \tag{5.30}$$

Both ξ_A (the null-stream-transformed residuals) and Z_{AB} are constructed using $\mathcal{N}(\hat{\Omega})$, which in turn is constructed using the response functions that depend on $\hat{\Omega}$. The null-stream-transformed model, ζ_A, is assumed to be of the form of Eq. 5.21, and is therefore fully determined by $\vec{H}(\vec{\lambda})$. This means the dependencies on the source location $\hat{\Omega}$ and the GW parameters $\vec{\lambda}$ are completely separated. In the following section, we exploit this separation of dependencies to construct a likelihood that does not depend on the GW model at all.

5.2.3 GW-model-independent likelihood

To construct the GW-independent likelihood, we can marginalise over the signal components that depend on $\vec{H}$. Previously, we assumed there was some model $H^{+}(\vec{\lambda}), H^{\times}(\vec{\lambda})$ (or equivalently, $\tilde{H}^{+}(\vec{\lambda}), \tilde{H}^{\times}(\vec{\lambda})$ in the FD). Here, we treat each frequency component $\tilde{H}^{+}_{\alpha}$, $\tilde{H}^{\times}_{\alpha}$ in Eq. 5.21 as $2\mathrm{N_f}$ independent variables (actually, $4\mathrm{N_f}$ variables because each entry is a complex number). Then, the marginalised likelihood can be computed as:

$$\mathcal{L}(\hat{\Omega}) = \int \mathrm{d}^{\mathrm{N_f}} \vec{\tilde{H}}^{+} \int \mathrm{d}^{\mathrm{N_f}} \vec{\tilde{H}}^{\times} \mathcal{L}(\hat{\Omega}, \vec{H})\, \mathrm{p}(\vec{\tilde{H}}^{+}, \vec{\tilde{H}}^{\times}). \tag{5.31}$$

If we assume flat priors $\mathrm{p}(\vec{\tilde{H}}^{+}, \vec{\tilde{H}}^{\times})$, the result of this integral can be obtained from the standard integral over a multivariate Gaussian with complex variables, which is:

$$\int \mathrm{d}\vec{x} \exp\left(-\frac{1}{2}(\vec{x}^{\dagger}, \vec{y}^{\dagger}) \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^{\top} & \Sigma_{yy} \end{pmatrix}^{-1} \begin{pmatrix} \vec{x} \\ \vec{y} \end{pmatrix}\right) = \exp\left(-\frac{1}{2}\vec{y}^{\dagger}\Sigma_{yy}^{-1}\vec{y}\right) \frac{(2\pi)^{\dim(\vec{x})}|\Sigma|}{|\Sigma_{yy}|}. \tag{5.32}$$

We can designate the first $2\mathrm{N_f}$ elements of χ_A as $\vec{x}$ in the equation above (corresponding to the elements of ζ_A that contain $\tilde{H}^{+}$ and $\tilde{H}^{\times}$), and the rest as $\vec{y}$. Then, we define the "cut" variables $\bar{\chi}$ as χ without the first $2\mathrm{N_f}$ entries and $\bar{Z}$ as Z without the first $2\mathrm{N_f}$ rows and columns (the latter fills the role of Σ_{yy} in the equation above). As we are not marginalising over anything that depends on $\hat{\Omega}$, terms that depend on $\hat{\Omega}$ in Eq. 5.29 can be treated as constants during integration. The marginalised likelihood, apart from any constant factors, is:

$$\mathcal{L}(\hat{\Omega}) = \exp\left(\sum_{A,B} -\bar{\chi}_A^{*}\bar{Z}_{AB}^{-1}\bar{\chi}_B\right)\frac{|Z|}{|\bar{Z}|}. \tag{5.33}$$

Note that $\bar{Z}^{-1} \neq (\overline{Z^{-1}})$, i.e. we have to compute the inverse of $\bar{Z}$, not remove the rows and columns from Z^{-1}.

5.3 Practical implementation with PTACAKE

To simulate, manage, and analyse realistic PTA data, we created a PYTHON package called PTACAKE (Pulsar Timing Arrays: Correlations Are KEy). Apart from localisation using null streams, PTACAKE is also used to adapt the spherical harmonic analysis from [138] to realistic data. We aim to make the package publicly available in the future, so these correlation-based techniques can be developed further and used for analysis of future PTA data sets. Here we focus on the practical implementation of simulating realistic data, applying the EFT of Sec. 5.1.1 and computing the GW-model-independent likelihood from Eq. 5.33 in PTACAKE.

5.3.1 Simulating realistic PTA data

The pulsars in existing PTAs are not uniformly distributed across the sky, as they are more abundant near the galactic disk (see for example Fig. 3.8). Additionally, radio telescopes cannot observe all parts of the sky (equally well), due to their fixed location on earth. To emulate this effect in our simulated PTA data, we have constructed a weight map based on the locations of millisecond pulsars in the ATNF pulsar catalogue as of May 2018 [139] (using the complete set, minus globular cluster duplicates). Pulsar locations are then randomly generated across the sky, biased by this weight map. For our initial test runs, we vary the total number of pulsars by selecting the first P from one preconstructed list. Each pulsar is assigned a noise rms $\sigma^{(p)}$ from a Gaussian distribution with mean 10^{-7}s and standard deviation 4×10^{-8}s. Each $\sigma_p < 10^{-8}$s is set to this minimum instead. This is an optimistic scenario for PTA; the mean rms of 100ns is almost reached by the best pulsar in the IPTA DR2 [33].

We obtain results from both idealised data, and data with realistic sampling times. In the idealised case, the data is evenly sampled with a constant cadence of $\Delta t = 2 \times 10^6$s (about 23 days). All pulsars start observing from $t = 0$ and end at $t = 20$yr. The realistic data sampling has three components: First, to introduce uneven sampling, cadences Δt_i are chosen randomly from a Gaussian distribution with a mean of 2×10^6s and a standard deviation of 10^5s (about one day). To ensure no two points are too close, each $\Delta t_i < 10^5$s is set to this minimum instead. Second, a random start time is assigned to each pulsar, chosen uniformly in the range $0 - 5$yr. Finally, another level of complexity is added by introducing gaps with a chosen "expected gap spacing" $\Delta_{\text{gap}} = 5$yr and "expected gap length" $T_{\text{gap}} = 5 \times 10^6$s. Then for each pulsar, the number of gaps is sampled from a Poisson distribution with expected value $(t_{\text{end}} - t_{\text{start}})/\Delta_{\text{gap}}$. Each gap is placed at a random time during observation, with a length sampled from an exponential distribution with rate parameter T_{gap} (overlapping gaps are allowed). All random choices are seeded, such that we are able to vary other parameters between tests while keeping the actual set of sampling times the same. An example of the sampling times in the first 5 years of simulated data for a PTA with 8 pulsars can be seen in Fig. 5.1

As in the previous chapters, we simulate a signal from a monochromatic, circularised SMBH binary, given by Eq. 2.7 and Eq. 1.75 (the mixing of polarisation states). The parameters of our test signal are amplitude $A = 10^{-14}$, phase offset $\phi_0 = 0.456$, polarisation $\psi = \pi/7$, inclination $\iota = 0.4$, and GW frequency $f = 2 \times 10^{-8}$Hz (all angles are in radians). The pulsar response further depends on the source location, which we choose to be at $(\theta = 0.456\pi, \phi = 0.321)$, where $\theta \in [0, \pi]$ is the polar sky angle and $\phi \in [0, 2\pi]$ is the azimuthal sky angle. This choice is mostly random, but it is intentionally not very close to any of the (first 10) pulsars in our preconstructed list, since that would strongly affect sky localisation (as we have previously seen in Chapter 3).

We run tests both with and without noise in the simulation, but in both cases the same noise model (Eqs. 5.3, 5.4, and 5.5) is assumed. When including noise, we add this to the TD signal, as a

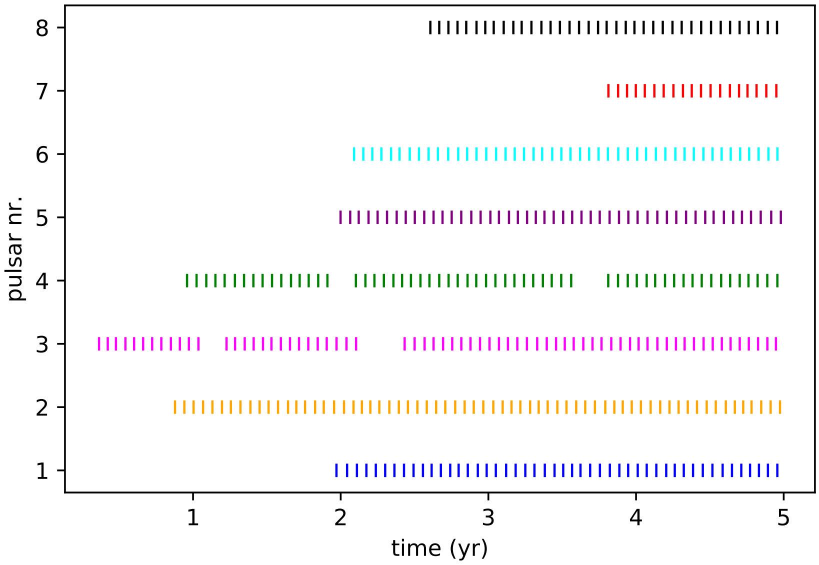

Figure 5.1: Example of the sampling times for a PTA with 8 pulsars (indicated with different colours). Only the first five years of simulated data are shown, so that it is easier to see the uneven cadences, unequal pulsar start times and data gaps.

randomly sampled value from a Gaussian distribution with zero mean, and variance equal to $\sigma^2_{(p)}$. For pulsar p, $\sigma_{(p)}$ is the pulsars' noise rms which was previously assigned. As an example, the residuals of three of the pulsars from the PTA of Fig. 5.1 are shown in Fig. 5.2. Although it is possible to perform all computations with a time-dependent $\sigma^{(p)}$, they are kept constant here. This means the TD covariance matrix $\Sigma^{(p)}$ for each pulsar is simply a $\mathrm{N}^{(p)} \times \mathrm{N}^{(p)}$ diagonal matrix with all diagonal elements equal to $\sigma^2_{(p)}$ (it is white noise).

Noise realisations are created using the same random seed, for better comparison between our test runs. Ideally, tests would be run with many different noise realisations, choices of signal parameters and PTA configuration. However, with limited available (computational) time, these preliminary results are all obtained with controlled variables as describes above. For each run (whether the data contain noise or not), we compute the (total) S/N in the TD as follows:

$$\rho^2 = \sum_{p=1}^{\mathrm{P}} \left(r_j^{(p)} \Sigma_{jk}^{(p)} r_k^{(p)} \right) . \tag{5.34}$$

Once the TD data are simulated, we determine the set of Fourier frequencies $\{f_{\mathrm{min}}, 2f_{\mathrm{min}}, \ldots, \mathrm{N_f} f_{\mathrm{min}}\}$ to use in our EFT. As discussed in Sec. 5.1.1, $f_{\mathrm{min}} = 1/T$ where T is the longest time between any two residuals in the data. This will be nearly $T = 20\mathrm{yr}$ such that $f_{\mathrm{min}} \approx 1.6 \times 10^{-9}\mathrm{Hz}$. The maxi-

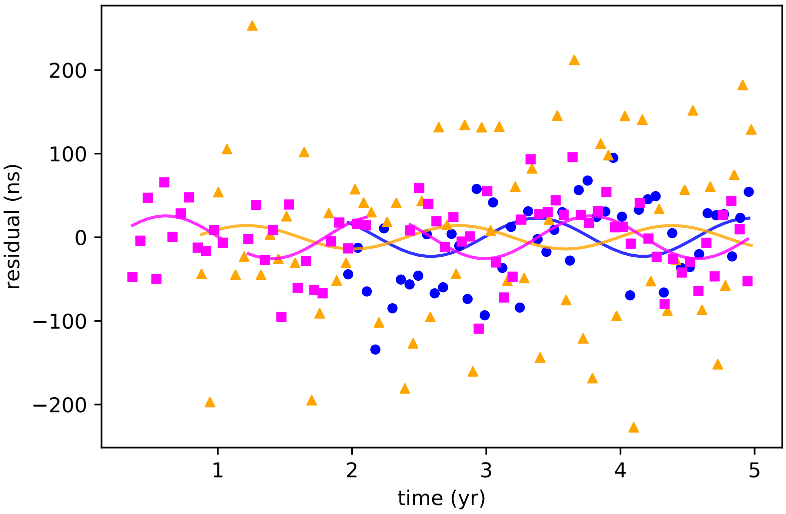

Figure 5.2: Example of the residuals from the first three pulsars of the PTA from Fig. 5.1, with corresponding colours. The residuals include the response to an injected sinusoidal GW signal (indicated with the solid lines) and Gaussian noise, according to the pulsars' noise levels. These are $\sigma_{(1)} \approx 53$ns (blue), $\sigma_{(2)} \approx 124$ns (orange), and $\sigma_{(3)} \approx 42$ns (magenta). The rms values are constant which means the noise is white. For visibility, only the first five years of simulated data are plotted.

mum frequency is cut off at $f_{\max} = 10^{-7}$Hz, so the total number of frequencies is $\mathrm{N_f} = 63$. We then compute the FD data by applying Eqs. 5.7-5.9. An example of the residuals from unevenly sampled data (with unequal start times and gaps) in the TD and FD is shown in Fig. 5.3. We have verified that for evenly sampled data, the result is the same as obtained using an FFT (as long as we choose the set of frequencies to match the FFT's).

5.3.2 Obtaining sky location posteriors and creating sky maps

The likelihood computation as described in Sec. 5.2 is done in PYTHON, making use of linear algebra tools in the NUMPY package (mainly from NUMPY.LINALG) [140, 141]. We also compute the "standard" FD likelihood from Eq. 5.19, henceforth $\mathcal{L}_{\mathrm{FD}}$, to compare to the null-stream-based, GW-model-independent likelihood from Eq. 5.33, henceforth $\mathcal{L}_{\mathrm{NS}}$. (Actually, we compute the log likelihood in both cases, which is less prone to numerical errors.) To sample each likelihood and obtain posteriors on the source sky position (θ, ϕ), we employ CPNEST [142], an implementation of a nested sampling algorithm [143]. To ensure the prior distribution is uniform on the sky, we sample over $\cos\theta \in [-1, 1]$ and $\phi \in [0, 2\pi]$. The likelihood $\mathcal{L}_{\mathrm{FD}}$ also depends on the GW model (we use the same model that was used to simulate the signal) and its parameters. The posterior samples output by CPNEST will

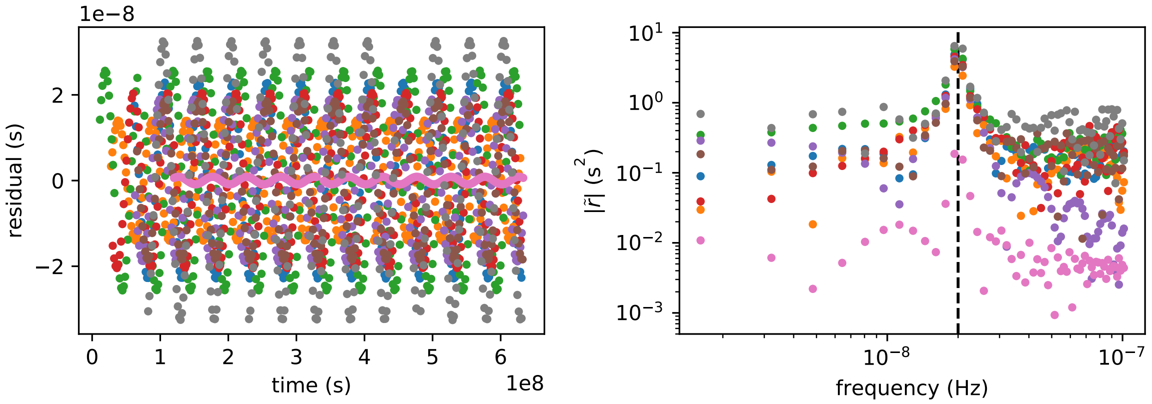

Figure 5.3: Example of simulated residuals for a PTA with 8 pulsars (indicated with different colours), with unevenly sampled data, unequal pulsar start times and data gaps. No noise is added to the injected sinusoidal signal, which has S/N $\approx$ 11 (computed with the assumed pulsar noise levels). Data are simulated in the TD (left), and converted to the FD (right) using the EFT from Sec. 5.1.1. The dashed line indicates the injected signal's frequency (2×10^{-8}Hz).

include these parameters as well as θ and ϕ, but they are not used in the comparison with $\mathcal{L}_{\rm NS}$. We use the following uniform priors (avoiding any degeneracy between the polarisation angle ψ and the phase off-set ϕ_0): $\phi_0 \in [0, 2\pi]$, $\log(A) \in [-16, -12]$, $\psi \in [0, \pi/4]$, $\cos(\iota) \in [-1, 1]$, and $f \in [0, 10^{-7}]$.

The output of CPNEST is a list of (θ, ϕ) samples that follow the posterior distribution. It is useful to estimate the (normalised) probability density at each point in the sky from these samples, which can then be used to visualise the posterior and compute a metric for the localisation precision. Because the sky is not flat, this is not a trivial task. We therefore employ the LIGO.SKYMAP software[5] [144], which uses a KDE (kernel density estimation) method to create pixelated maps of the posterior (with a maximum HEALPIX refinement parameter of $n = 64$). From these maps, we are able to compute Ω_{90}, the fraction of the sky that contains 90% of the posterior (we also used this as a measure of localisation precision in Chapter 3). Additionally, Ω_{50} and Ω_{95} are computed for 50% and 95% of the posterior, respectively. Our hypothesis is that the localisation for $\mathcal{L}_{\rm NS}$ is generally poorer than for $\mathcal{L}_{\rm FD}$, given that some of the data (the first two null streams), have not been used. The fraction of unused data is 2/P, so it is expected that for an increasing number of pulsars P, the posterior from $\mathcal{L}_{\rm NS}$ resembles the one from $\mathcal{L}_{\rm FD}$ more closely.

---

[5] `git.ligo.org/lscsoft/ligo.skymap`

5.4 Preliminary results

5.4.1 Idealised data without noise

Consider first a test run with evenly sampled data, without any noise. The posterior obtained with the "standard" likelihood $\mathcal{L}_{\mathrm{FD}}$, for a PTA with P = 10, is shown in Fig. 5.4. The signal is localised around the injected source location, within an error-box of $\Omega_{90} = 109\mathrm{deg}^2$. This is consistent with earlier results from Chapter 3; the most comparable runs have P = 10, S/N = 15, and range between $70\mathrm{deg}^2 \lesssim \Omega_{90} \lesssim 900\mathrm{deg}^2$ as shown in Fig. 3.3. (There is such a wide range in Ω_{90} values because the localisation is strongly affected by the pulsar placement). The sky maps for different size PTAs are not shown here, but the posteriors for $\mathrm{P} = 3, 5, 8, 20$ are all found around the injected source sky location as well (their localisation areas can be seen in Fig. 5.7).

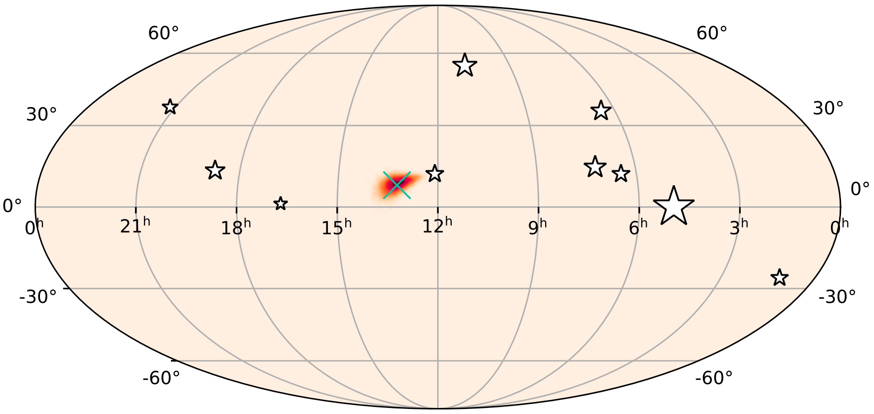

Figure 5.4: Posterior on the source sky position from $\mathcal{L}_{\mathrm{FD}}$ (the "standard" FD likelihood from Eq. 5.10). Simulated data is for a PTA with 10 pulsars (their positions are marked with stars, with bigger markers for pulsars with lower noise levels), is evenly sampled, and has no added noise. The injected source sky location is indicated with a green cross.

Using the same data, the posterior obtained with $\mathcal{L}_{\mathrm{NS}}$ does not localise the signal at the correct location, as shown in Fig. 5.5. Upon further investigation, it turns out the exponential term in Eq. 5.33 is peaked at the injected source location, but the term $|Z|/|\bar{Z}|$ is peaked elsewhere. Because the second term is bigger, this explains the misplaced posterior. At this time, it is unclear what causes this behaviour. It could be an error in our derivation of the null-stream likelihood, or in the computation of Z or $\bar{Z}$ (or something else entirely). There is no reason to believe that there is something specific about the PTA configuration used here that could cause the issue.

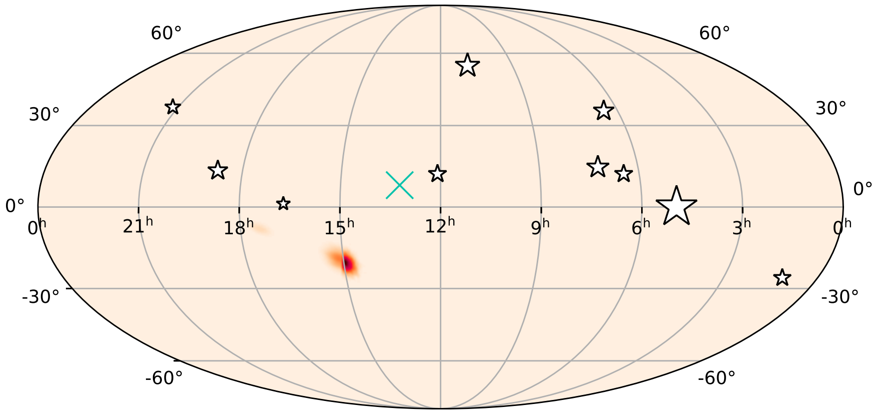

Figure 5.5: Posterior on the source sky position from $\mathcal{L}_{\rm NS}$ (the null-stream likelihood from Eq. 5.33). Simulated data and markers are the same as in Fig. 5.4.

To make use of $\mathcal{L}_{\rm NS}$, we define an adjusted likelihood in which we leave out the problematic $|Z|/|\bar{Z}|$-term:

$$\mathcal{L}_{\rm NS2} = \exp\left(\bar{\chi}_A \bar{Z}^{-1}_{AB} \bar{\chi}_B\right). \tag{5.35}$$

The posterior obtained with this likelihood (still for evenly sampled data without noise), can be seen in Fig. 5.6. This time, the posterior is well-localised, and very similar to the $\mathcal{L}_{\rm FD}$ one in Fig. 5.4. The localisation area is $\Omega_{90} = 124\text{deg}^2$, which is somewhat larger than in the $\mathcal{L}_{\rm FD}$ case. The hypothesis that the posterior from the null-stream likelihood is always wider than the "standard" one should apply to $\mathcal{L}_{\rm NS2}$ as well as $\mathcal{L}_{\rm NS}$, as the same fraction of data is unused in both. This result therefore seems to follow our expectations. However, it is possible that leaving out the $|Z|/|\bar{Z}|$-term affects the posterior width, thereby contributing to the difference in Ω_{90} between $\mathcal{L}_{\rm FD}$ and $\mathcal{L}_{\rm NS2}$.

We can further investigate the difference in localisation precision of $\mathcal{L}_{\rm FD}$ and $\mathcal{L}_{\rm NS2}$ by comparing all measures, Ω_{50}, Ω_{90}, and Ω_{95}, for a varying number of pulsars P. These values are all plotted in the top panel of Fig. 5.7. The bottom panel shows the ratio of each pair of $\Omega_{\rm x}$ (x being 50, 90 or 95) between $\mathcal{L}_{\rm NS2}$ and $\mathcal{L}_{\rm FD}$. The posteriors are all well-localised around the injected source sky location, although the P = 3 posteriors are both very wide and have multiple peaks, but that is not surprising.

The overall decrease of Ω_{50}, Ω_{90}, and Ω_{95} with increasing P follows the same trend we found in

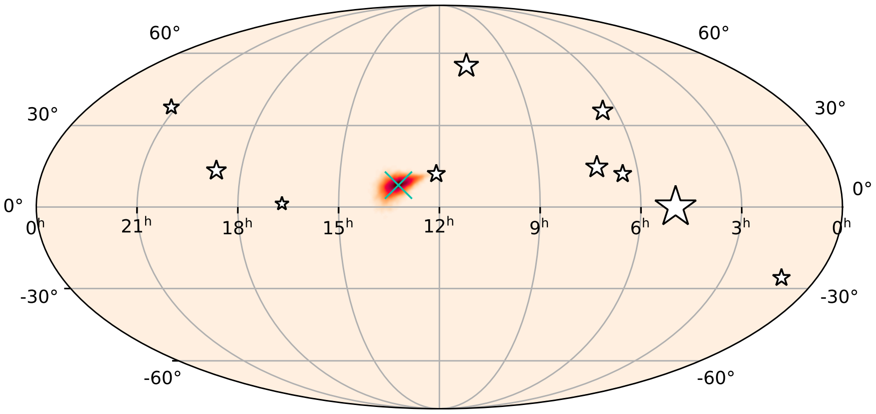

Figure 5.6: Posterior on the source sky position from $\mathcal{L}_{\rm NS2}$ (the adjusted null-stream likelihood from Eq. 5.35). Simulated data and markers are the same as in Fig. 5.4.

Chapter 3, with a steep increase in the "transition area" – for this PTA configuration between P = 8 to 10. In most cases, the $\mathcal{L}_{\rm NS2}$ posterior is wider than the $\mathcal{L}_{\rm FD}$ one, as expected. However, for P = 5, Ω_{90} and Ω_{95} are smaller for $\mathcal{L}_{\rm NS2}$ (Ω_{50} is not). This possibly means that the error of the null-stream likelihood is underestimated, as it should not be possible to have a more precise measurement with fewer data.

The expectation that the posteriors for $\mathcal{L}_{\rm FD}$ and $\mathcal{L}_{\rm NS2}$ resemble each other more for larger P seems to be true for Ω_{50}, but less so for Ω_{90} and Ω_{95}, mainly due to the unexpected values at P = 5. Looking at the bottom panel of Fig. 5.7, perhaps one could conclude that for $5 \lesssim \mathrm{P} \lesssim 8$, the peak of the posterior (Ω_{50}) is narrower for $\mathcal{L}_{\rm FD}$ than for $\mathcal{L}_{\rm NS2}$, but that the opposite is true for the larger contour areas, and that for $\mathrm{P} \gtrsim 10$ the posteriors behave as expected. However, keeping in mind that the sky localisation also depends on the specific PTA configuration, we can definitely not make this statement in general. Additionally, the finite number of samples used to estimate the posterior density means that the larger contours, Ω_{90} and Ω_{95}, are measured less accurately than Ω_{50}, because there are fewer samples near the edges of the distribution. This may especially be important for the wider posteriors of P = 3 and P = 5.

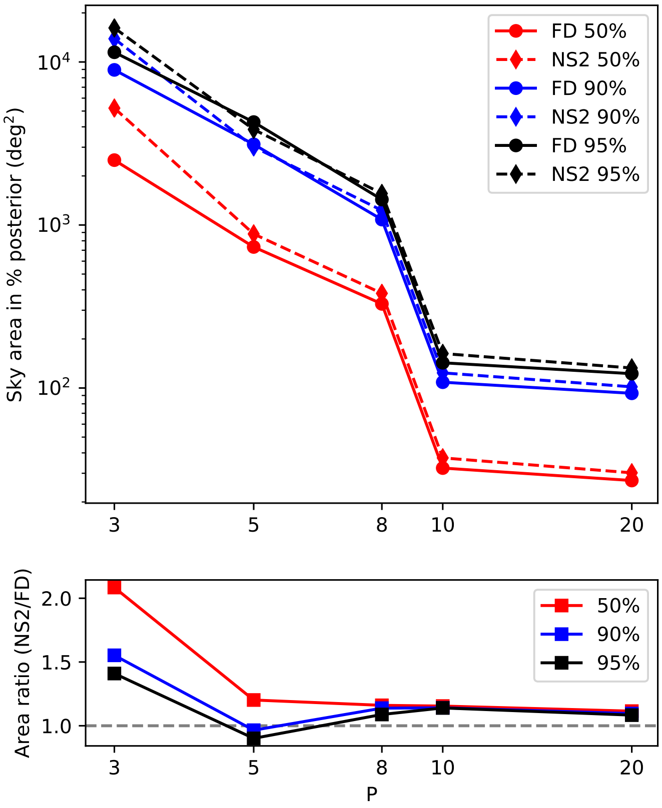

Figure 5.7: Top: sky-localisation precision comparing the use of $\mathcal{L}_{\mathrm{FD}}$ (the standard FD likelihood from Eq. 5.19, solid lines) and $\mathcal{L}_{\mathrm{NS2}}$ (the adjusted null-stream likelihood from Eq. 5.35, dashed lines). The localisation precision is measured using Ω_{50} (red), Ω_{90} (blue), and Ω_{95} (black), which are the smallest areas containing 50%, 90% and 95% of the sky posterior, respectively. On the x-axis, the number of pulsars P in the PTA is varied. The simulated data is evenly sampled and has no noise.
Bottom: Ratio between $\mathcal{L}_{\mathrm{NS2}}$ and $\mathcal{L}_{\mathrm{FD}}$ of each pair of areas in the top panel. The grey dashed line marks the expected lower limit, which is to say that above the line, the posterior from $\mathcal{L}_{\mathrm{NS2}}$ is wider than the one from $\mathcal{L}_{\mathrm{FD}}$.

5.4.2 Realistic sampling and no noise

The main reason for our computations to be in the FD is so our method can be applied to unevenly sampled data. Fig. 5.8 shows the skymap made with $\mathcal{L}_{\rm FD}$, using the full suite of data sampling complexity as described in Sec. 5.3.1: uneven cadences, unequal observation start times and data gaps. The source is still well-localised, with $\Omega_{90} = 125{\rm deg}^2$, a somewhat wider posterior than the evenly sampled case with $\Omega_{90} = 109{\rm deg}^2$. There is a difference in S/N as well; for the idealised data, S/N = 16.5, and for the realistic data – which has fewer TOAs due to unequal observation start times and data gaps – S/N = 15.2. The theoretical scaling of $\Omega_{90} \propto \rho^{-2}$ almost entirely explains the difference in Ω_{90}, but it is possible there is an additional contribution from information being lost when applying the EFT to the unevenly sampled data.

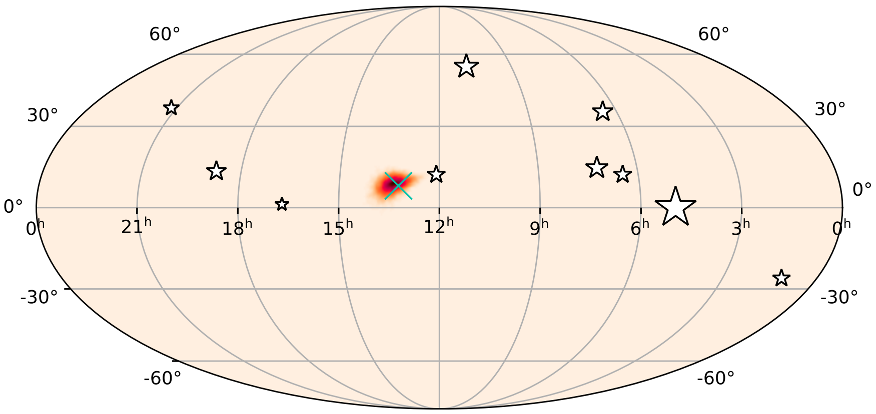

Figure 5.8: Posterior on the source sky position from $\mathcal{L}_{\rm FD}$ (the "standard" FD likelihood from Eq. 5.10). The simulated PTA has 10 pulsars (their positions are marked with stars, with bigger stars for pulsars with lower noise levels). Data is unevenly sampled, with unequal start dates for the pulsars and includes gaps, but has no added noise. The injected source sky location is indicated with a green cross.

We can compare this to the result using $\mathcal{L}_{\rm NS2}$, with the exact same simulated PTA data. The posterior, as shown in Fig. 5.9, is very similar to the $\mathcal{L}_{\rm FD}$ result. The localisation area is $\Omega_{90} \approx 140{\rm deg}^2$, which is somewhat bigger than the $\mathcal{L}_{\rm FD}$ one and therefore follows our expectations. As with the evenly sampled runs, we make a full comparison of localisation precision with varying P, using Ω_{50}, Ω_{90} and Ω_{95}. The results are plotted in Fig. 5.10. All posteriors are found around the location of the injected source (with the P = 3 posteriors being very wide with multiple peaks).

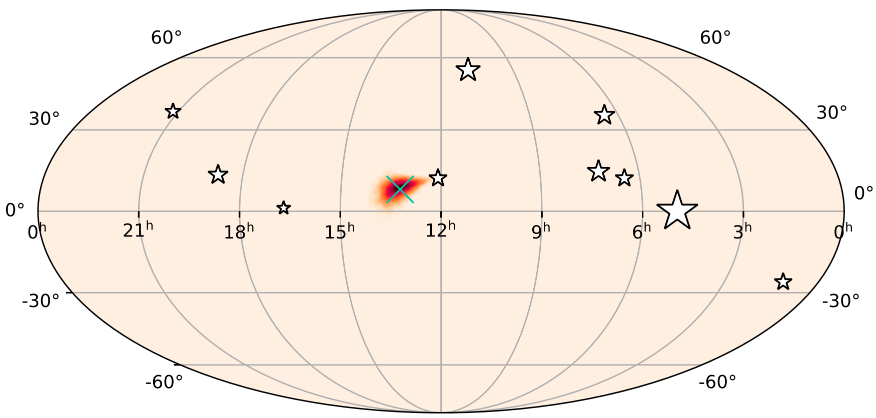

Figure 5.9: Posterior on the source sky position from $\mathcal{L}_{\mathrm{NS2}}$ (the adjusted null-stream likelihood from Eq. 5.35). Simulated data and markers are the same as in Fig. 5.8.

The same steep transition in localisation precision between P = 8 and 10 is found as in the evenly sampled case. The ratios of Ω_{50}, Ω_{90} and Ω_{95} between $\mathcal{L}_{\mathrm{NS2}}$ and $\mathcal{L}_{\mathrm{FD}}$, found in the bottom panel of Fig. 5.10, are very similar to the results in Fig. 5.7. This time, the Ω_{90} ratio for P = 5 is just above 1, and the Ω_{95} ratio is slightly higher as well. As before, the posteriors seem to resemble each other most for the highest number of pulsars.

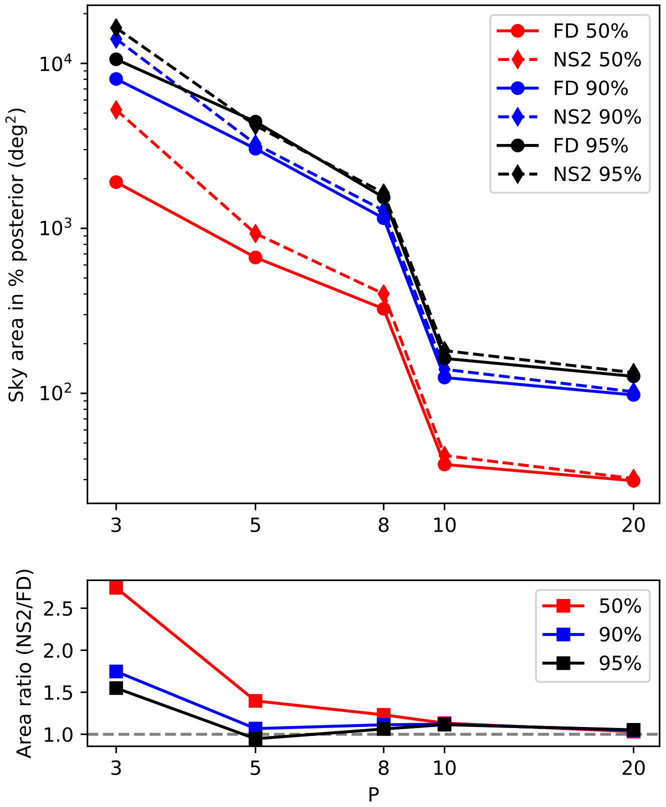

Figure 5.10: Top: sky-localisation precision comparing the use of $\mathcal{L}_{\rm FD}$ (the standard FD likelihood from Eq. 5.19, solid lines) and $\mathcal{L}_{\rm NS2}$ (the adjusted null-stream likelihood from Eq. 5.35, dashed lines). The localisation precision is measured using Ω_{50} (red), Ω_{90} (blue), and Ω_{95} (black), which are the smallest areas containing 50%, 90% and 95% of the sky posterior, respectively. On the x-axis, the number of pulsars P in the PTA is varied. The simulated data has varying cadences, unequal observation start dates for each pulsar, and includes gaps, but there is no noise added.
Bottom: Ratio of the areas plotted in the top panel. The grey dashed line marks the expected lower limit of the ratio, meaning that the posterior from $\mathcal{L}_{\rm NS2}$ is wider than the posterior from $\mathcal{L}_{\rm FD}$.

5.4.3 Runs with noise

If unlimited (computational) time was available, it would be ideal to run tests with many different noise realisations, and randomly chosen PTA configurations and source parameters. This not being the case, it is still valuable to run few tests with a specific noise realisation, to see how much it affects the results. Adding Gaussian noise to the simulated data for our benchmark run ($\mathcal{L}_{\mathrm{FD}}$ with P = 10) results in a very similar posterior as the no-noise run, as shown in Fig. 5.11. The peak of the posterior is slightly shifted compared to the no-noise run, but still very close to the injected source's sky location. The localisation areas are also comparable with $\Omega_{90} \approx 137\mathrm{deg}^2$ with noise versus $\Omega_{90} \approx 125\mathrm{deg}^2$ without.

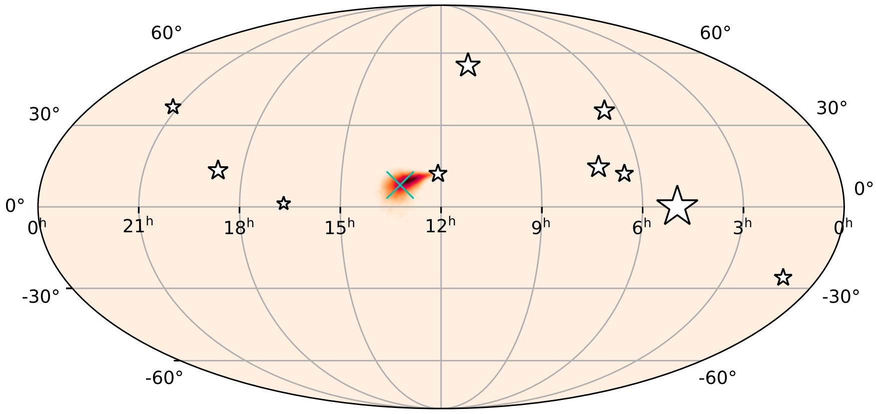

Figure 5.11: Posterior on the source sky position from $\mathcal{L}_{\mathrm{FD}}$. The simulated PTA has 10 pulsars (their positions are marked with stars, with bigger stars for pulsars with lower noise levels). Data is unevenly sampled, with unequal start dates for the pulsars, includes gaps and has white noise added according to the pulsars' noise levels. The injected source sky location is indicated with a green cross.

The posterior obtained using the null-stream likelihood $\mathcal{L}_{\mathrm{NS2}}$ however, is somewhat further shifted from the injected source sky location, as shown in Fig. 5.12. Unfortunately, if we look at the corresponding run with P = 8, shown in Fig. 5.13, it becomes clear $\mathcal{L}_{\mathrm{NS2}}$ is not functioning well at all, rather than this being a small error (this is also evident from other PTA sizes not shown here). It is unclear at the moment what causes this issue, but it appears to manifest only when complex data sampling and noise are combined. In a test with the same PTA configuration and noise level as Fig. 5.13, but with evenly sampled data, the source is well-localised again (see the skymap in Fig. 5.14). More research is required to find out when exactly this problem manifests (for example, is there a difference between data sampled with varying cadences and data with gaps?) and whether it is avoidable or not. Perhaps it is related to the issue that causes the $\mathcal{L}_{\mathrm{NS}}$ (the likelihood that includes the $|Z|/|\bar{Z}|$-term) posteriors to be misplaced.

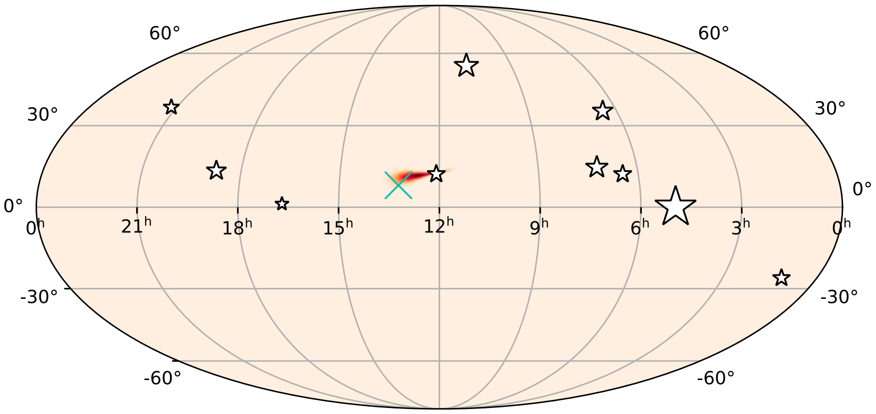

Figure 5.12: Posterior on the source sky position from $\mathcal{L}_{\mathrm{NS2}}$. Simulated data and markers are the same as in Fig. 5.11.

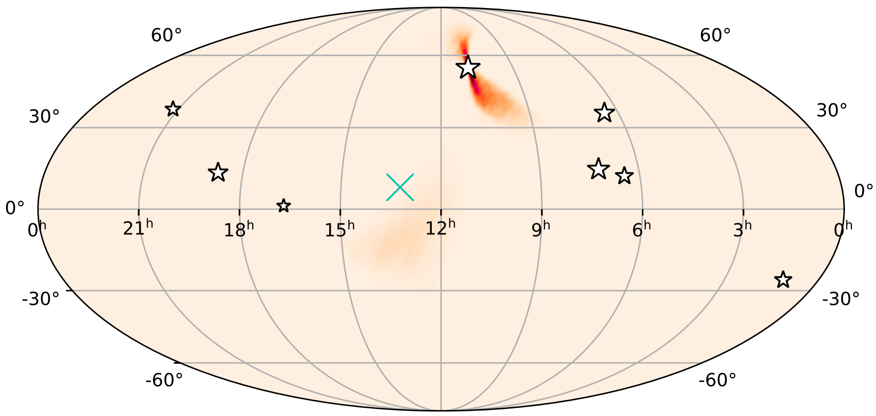

Figure 5.13: Posterior on the source sky position from $\mathcal{L}_{\rm NS2}$. The simulated PTA has 8 pulsars (their positions are marked with stars, with bigger stars for pulsars with lower noise levels). Data is unevenly sampled, with unequal start dates for the pulsars, includes gaps and has white noise added according to the pulsars' noise levels. The injected source sky location is indicated with a green cross.

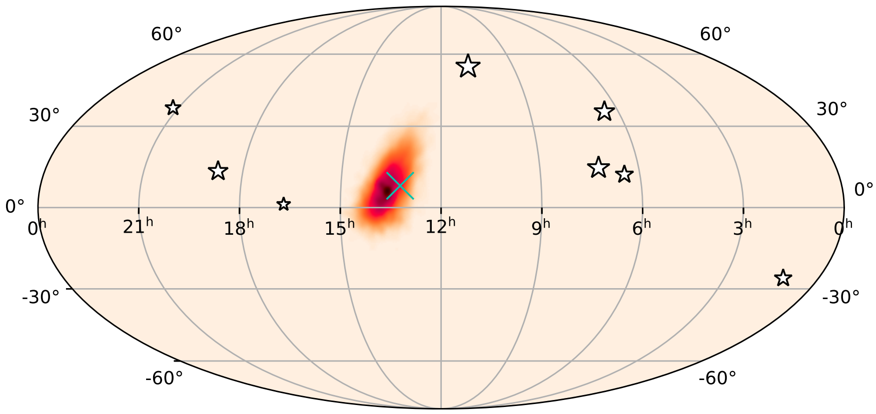

Figure 5.14: Posterior on the source sky position from $\mathcal{L}_{\rm NS2}$. The PTA configuration and plot markers are the same as in Fig. 5.13. Simulated data is evenly sampled and has added white noise.

5.5 Discussion

In this chapter, we have constructed a null-stream likelihood that can be applied to realistic PTA data with arbitrary sampling times, and is independent of the GW model. We use this likelihood (Eq. 5.33) in a nested sampling algorithm to recover the sky location of a SMBH binary signal injected into simulated data. The posterior is found to be misplaced, however, this issue seems to disappear if we adjust the likelihood to Eq. 5.35. Source sky location posteriors obtained with this method are compared to those from a "standard" FD likelihood, Eq. 5.19. The expectation is that the posteriors from the null-stream likelihood are wider, because part of the data is unused. Results are largely consistent with this hypothesis using evenly sampled data, as well as using realistic data with irregular cadences, unequal observation start times, and gaps. For some cases however, the null-stream likelihood results in an overall narrower posterior while the standard likelihood is more peaked.

When adding white noise to the simulated data, another issue manifests: the posteriors from the null-stream likelihood do not recover the injected source sky location any more. However, for evenly sampled data, adding noise is not a problem. More tests using different combinations of noise and data complexity could help uncover under what circumstances the issue occurs. It is currently unclear what the cause is, and whether it is related to the problem with the originally derived Eq. 5.35. More research is needed to investigate both issues. A specific point of interest is the choice of prior in the marginalisation of $\tilde{H}^{+}$ and $\tilde{H}^{\times}$ in Eq. 5.31. Flat priors allowed to analytically compute the marginalised likelihood with the standard integral from Eq. 5.32, but this may be possible with a different choice as well. Given that the components of $\tilde{H}^{+}$ and $\tilde{H}^{\times}$ are complex variables, perhaps a log flat prior on their amplitude would be appropriate.

If resolving these issues results in a likelihood that consistently localises the source, error estimation can be further investigated by running many test with randomised source parameters, PTA configurations and noise realisations. As we found that for some of our PTA configurations, the posteriors from the null-stream likelihood were narrower than from the "standard" likelihood, it is possible the error is underestimated. Creating a "pp-plot" from a large collection of runs would be a robust test of the error estimation. Repeating these runs with the standard likelihood would also allow a better assessment of the null-stream likelihood's performance compared to the standard case. It would be useful to assess what number of pulsars are needed to make the null-stream method viable.

5.5.1 Future work

If teething problems with the null-stream likelihood can be resolved, it has great potential for use in PTA data analysis. A source can be located on the basis of the relation between the responses of different pulsars alone, without the need to model the GW signal. Complexity in the signal (for example, a binary's eccentricity or the effect of gas on the orbital evolution) can be modelled, but it would be useful to also have an unmodelled search available in which no assumptions about the GW signal have to be made. Furthermore, with the null-stream likelihood, the source location can be measured without needing to measure any signal parameters at the same time, meaning the dimensionality of the sampling problem is greatly reduced.

Several simplifications in our method have not been addressed yet. First, we have not included the "pulsar term" in the PTA response (see also Sec. 2.1.3). Because the pulsar terms depend on the pulsar distances (modulo the GW wavelength), which are largely unknown, including it in the signal would effectively add another 2P independent variables. This means null streams cannot be constructed in this case. A more plausible approach could be to treat them as a source of noise. Second, we have greatly simplified the noise in the simulated PTA data for our null-stream analysis. Noise modelling for PTA is a complex task on its own, but it can potentially be combined with the null-stream approach without much further effort. The construction of the null-stream likelihood prescribes how to transform a given covariance matrix Σ of the TD residuals (first to the FD, then with the null-stream transformation), and this Σ could be as complex as needed to begin with. Any parameters of the noise model could be searched over at the same time as the source sky location. This would incur an additional computational cost; the transformed Σ (Eq. 5.28, and its inverse after discarding the first $2\mathrm{N_f}$ rows and columns) would have to be computed at each point in the sampling space.

Another prospect for the null-stream likelihood method is to extend it to search for a resolvable signal, rather than just localise an already detected one. Detection significance in the Bayesian framework can be addressed by computing the Bayes factor (Eq. 2.21) between a model that assumes a signal is included in the data and a model that assumes no signal is present. This would require computing the evidence from the null-stream likelihood, which is only useful if the likelihood is properly normalised. We have not included normalisation in our computation but it can in principle be done. One could search for unmodelled resolvable PTA signals, such as burst from cosmic string cusps [27], or entirely unknown quadrupolar sources in the Universe.

Conclusion

Pulsar Timing Arrays (PTAs) are expected to detect gravitational waves (GWs) from super massive black hole (SMBH) binaries in the not-too-distant future. Likely, a stochastic background built up from unresolved binaries throughout the Universe will be detected first, but few signals from the most massive and nearby binaries could be resolved individually [22]. Localising these signals to small regions of the sky is crucial to increase the chances of being able to identify the galaxy that hosts the source, thereby opening up the possibility to observe the source in the electromagnetic spectrum as well as with GWs.

We have employed the construction of *null streams* – combinations of data that cancel the GW signal from a given direction – for source localisation in a Bayesian framework. In Chapter 3, we have constructed and implemented a null-stream-based likelihood in which the dependency on the source sky location and on the GW model is separated in different terms. We have then employed this likelihood to investigate the localisation capacity of different PTA configurations, using idealised, simulated data. We found that the area to which a signal can be localised scales with the signal-to-noise ratio (S/N, ρ in Eqs.) of the injections as ρ^{-2}, as is expected from theory and also found in [20], but only for high S/N. Between $5 \lesssim \rho \lesssim 10$, there is a steep transition between very poor localisation and more precise localisation following the high-S/N scaling. Localisation is also found to be strongly affected by the particular placement of the pulsars in the PTA.

We further used this localisation method to investigate the capabilities of existing PTA experiments and the combined International Pulsar Timing Array (IPTA). There is convincing evidence that combining data sets improves source localisation, not only through an increased S/N, but through an improved sky coverage with more pulsars as well.

It is expected that the first detected SMBH binaries with PTAs will have relatively poor sky location precision (a few 100deg^2). This makes identifying the host galaxy challenging. In Chapter 4, we have constructed a Bayesian method of combining the limited information on the source sky location, and the posterior on the signal's amplitude, with the properties of galaxies from a catalogue. This allows us to assign each galaxy a probability of hosting the detected source, and rank them accordingly. We have demonstrated the method with simulated data and a mock galaxy catalogue, and found that it greatly decreases the number of candidate galaxies. By accepting a 19% (55%) probability of missing the true host, our method reduces the number of candidates by nearly three (four) orders of magnitude compared to a crude cut using a lower limit on the stellar mass of the galaxy, $M_* > 5 \times 10^{10} \mathrm{M}_\odot$, and a redshift limit of $z < 1$. We also found that, although the redshift distribution of the best candidates peaks around $z \lesssim 0.2$, it has a long tail that extends up to $z \lesssim 1$, consistent with predictions from [22].

In Chapter 5, we have further developed the null-stream-based likelihood so it can be applied to more realistic PTA data with arbitrary sampling times, by transforming the data into the frequency domain (FD). We have then used the separation of the dependence on the source sky location and on the GW model to marginalise over the null-stream components that depend on the GW, making the likelihood independent from it. This means a fraction 2/P of the data – where P is the number of pulsars in the array – is unused. As expected, this generally results in wider posteriors using the null-stream likelihood compared to posteriors obtained with a "standard" likelihood in the FD.

There are still some issues with this extended null-stream method that need to be addressed in future work. First, we have had to make an empirical adaption to the derived GW-model-independent likelihood, which could be due to an error in the derivation or in the implementation. Second, although the method seems to perform well for realistically sampled data with zero noise, or for evenly sampled data with added white noise, it does not work when those factors are combined.

Before applying the null-stream likelihood to real PTA data in future work, it would be good to include the pulsar term in the PTA response. It is also important to address the complexity of PTA noise. In principle, an arbitrarily complex noise model can be used in conjunction with the null-stream method, but this has not been tested. Lastly, it would be valuable to adapt our method to detection, rather than just localisation of already detected sources. This could be done in the Bayesian framework by computing the Bayes factor, for which one would need to determine the normalisation of the null-stream likelihood (so far, we have relied on post hoc numerical normalisation of the posteriors). Succeeding in this effort would mean that null streams could be used as an alternative search for resolvable SMBH binaries, as well as for difficult to model, or even unmodelled, signals in PTA data.

Bibliography

[1] C J Moore, R H Cole, and C P L Berry. Gravitational-wave sensitivity curves. *Classical and Quantum Gravity*, 32(1):015014, Dec 2014.

[2] Janna M. Goldstein, John Veitch, Alberto Sesana, and Alberto Vecchio. Null-stream analysis of Pulsar Timing Array data: localization of resolvable gravitational wave sources. *Monthly Notices of the Royal Astronomical Society*, 477(4):5447–5459, July 2018.

[3] S. Chatterji, A. Lazzarini, L. Stein, et al. Coherent network analysis technique for discriminating gravitational-wave bursts from instrumental noise. *Physical Review D*, 74(8):082005, October 2006.

[4] Janna M. Goldstein, Alberto Sesana, A. Miguel Holgado, and John Veitch. Associating host galaxy candidates to massive black hole binaries resolved by pulsar timing arrays. *Monthly Notices of the Royal Astronomical Society*, 485(1):248–259, May 2019.

[5] Jeroen Meidam, Ka Wa Tsang, Janna Goldstein, et al. Parametrized tests of the strong-field dynamics of general relativity using gravitational wave signals from coalescing binary black holes: Fast likelihood calculations and sensitivity of the method. *Physical Review D*, 97(4):044033, February 2018.

[6] Riccardo Buscicchio, Elinore Roebber, Janna M. Goldstein, and Christopher J. Moore. Label switching problem in Bayesian analysis for gravitational wave astronomy. *Physical Review D*, 100(8):084041, October 2019.

[7] Elinore Roebber, Riccardo Buscicchio, Alberto Vecchio, et al. Milky Way Satellites Shining Bright in Gravitational Waves. *arXiv e-prints*, page ArXiv:2002.10465, February 2020.

[8] A. Hewish, S. J. Bell, J. D. H. Pilkington, P. F. Scott, and R. A. Collins. Observation of a Rapidly Pulsating Radio Source. *Nature*, 217(5130):709–713, February 1968.

[9] R. W. Hellings and G. S. Downs. Upper limits on the isotropic gravitational radiation background from pulsar timing analysis. *The Astrophysical Journal*, 265:L39, feb 1983.

[10] Hannah Middleton, Siyuan Chen, Walter Del Pozzo, Alberto Sesana, and Alberto Vecchio. No tension between assembly models of super massive black hole binaries and pulsar observations. *Nature Communications*, 9:573, February 2018.

[11] J. P. W. Verbiest, L. Lentati, G. Hobbs, et al. The International Pulsar Timing Array: First data release. *Monthly Notices of the Royal Astronomical Society*, 458(2):1267–1288, May 2016.

[12] K. M. Górski, E. Hivon, A. J. Banday, et al. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *Astrophysical Journal*, 622(2):759–771, April 2005.

[13] James W. Cooley and John W. Tukey. An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, 19:297–301, 1965.

[14] Henri Poincaré. Sur la dynamique de l'électron. *Rendiconti del Circolo matematico di Palermo*, 21:129–176, Jan 1906.

[15] Albert Einstein. Näherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin)*, pages 688–696, Jan 1916.

[16] Albert Einstein. Über Gravitationswellen. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin)*, pages 154–167, Jan 1918.

[17] B. P Abbott, R Abbott, T. D. Abbott, et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters*, 116(6):61102, feb 2016.

[18] Pau Amaro-Seoane, Heather Audley, Stanislav Babak, et al. Laser Interferometer Space Antenna. *arXiv e-prints*, page ArXiv:1702.00786, February 2017.

[19] S. Detweiler. Pulsar timing measurements and the search for gravitational waves. *Astrophysical Journal*, 234:1100–1104, Dec 1979.

[20] Alberto Sesana and Alberto Vecchio. Measuring the parameters of massive black hole binary systems with pulsar timing array observations of gravitational waves. *Physical Review D*, 81(10):104008, May 2010.

[21] M. P. Hobson, G. P. Efstathiou, and A. N. Lasenby. *General Relativity.* 2006.

[22] P. A. Rosado, A. Sesana, and J. Gair. Expected properties of the first gravitational wave signal detected with pulsar timing arrays. *Monthly Notices of the Royal Astronomical Society*, 451:2417–2433, August 2015.

[23] Alberto Sesana. Prospects for Multiband Gravitational-Wave Astronomy after GW150914. *Physical Review Letters*, 116(23):231102, Jun 2016.

[24] J Aasi et al. Advanced ligo. *Class. Quantum Grav.*, 32(7):074001, 2015.

[25] B. P. Abbott et al. GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.*, 119:161101, Oct 2017.

[26] Paul D. Lasky. Gravitational Waves from Neutron Stars: A Review. *Publications of the Astronomical Society of Australia*, 32:e034, September 2015.

[27] Thibault Damour and Alexander Vilenkin. Gravitational radiation from cosmic (super)strings: Bursts, stochastic background, and observational windows. *Phys. Rev. D*, 71:063510, Mar 2005.

[28] Andrea N Lommen. Pulsar timing arrays: the promise of gravitational wave detection. *Reports on Progress in Physics*, 78(12):124901, 2015.

[29] R. S. Foster and D. C. Backer. Constructing a Pulsar Timing Array. *Astrophysical Journal*, 361:300, September 1990.

[30] R. S. Foster and D. C. Backer. Results from the Berkeley-NRAO Pulsar Timing Array Experiment. In *Bulletin of the American Astronomical Society*, volume 21, page 1205, September 1989.

[31] S. R. Taylor, M. Vallisneri, J. A. Ellis, et al. Are we there yet? time to detection of nanohertz gravitational waves based on pulsar-timing array limits. *Astrophysical Journal Letters*, 819(1):L6, March 2016.

[32] Xavier Siemens, Justin Ellis, Fredrick Jenet, and Joseph D. Romano. The stochastic background: scaling laws and time to detection for pulsar timing arrays. *Classical and Quantum Gravity*, 30(22):224015, November 2013.

[33] B. B. P. Perera, M. E. DeCesar, P. B. Demorest, et al. The International Pulsar Timing Array: second data release. *Monthly Notices of the Royal Astronomical Society*, 490(4):4666–4687, December 2019.

[34] Andrea N. Lommen and Paul Demorest. Pulsar timing techniques. *Classical and Quantum Gravity*, 30(22):224001, November 2013.

[35] M. Vallisneri, S. R. Taylor, J. Simon, et al. Modeling the uncertainties of solar-system ephemerides for robust gravitational-wave searches with pulsar timing arrays. *arXiv e-prints*, page ArXiv:2001.00595, January 2020.

[36] J. H. Taylor and J. M. Weisberg. A new test of general relativity - Gravitational radiation and the binary pulsar PSR 1913+16. *Astrophysical Journal*, 253:908–920, February 1982.

[37] R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. *Astrophysical Journal Letters*, 195:L51–L53, January 1975.

[38] J. M. Weisberg, J. H. Taylor, and L. A. Fowler. Gravitational waves from an orbiting pulsar. *Scientific American*, 245:74–82, October 1981.

[39] G Hobbs, A Lyne, and M Kramer. Pulsar timing noise. *Chinese Journal of Astronomy and Astrophysics*, 6(S2):169–175, oct 2006.

[40] Ryan M. Shannon and James M. Cordes. Assessing the role of spin noise in the precision timing of millisecond pulsars. *The Astrophysical Journal*, 725(2):1607–1619, nov 2010.

[41] C. G. Bassa, G. H. Janssen, R. Karuppusamy, et al. LEAP: the Large European Array for Pulsars. *Monthly Notices of the Royal Astronomical Society*, 456(2):2196–2209, February 2016.

[42] G. Desvignes, R. N. Caballero, L. Lentati, et al. High-precision timing of 42 millisecond pulsars with the European Pulsar Timing Array. *Monthly Notices of the Royal Astronomical Society*, 458(3):3341–3380, May 2016.

[43] Zaven Arzoumanian, Adam Brazier, Sarah Burke-Spolaor, et al. The NANOGrav 11-year Data Set: High-precision Timing of 45 Millisecond Pulsars. *Astrophysical Journal Supplement Series*, 235(2):37, April 2018.

[44] R. N. Manchester, G. Hobbs, M. Bailes, et al. The Parkes Pulsar Timing Array Project. *Publications of the Astronomical Society of Australia*, 30:e017, January 2013.

[45] D. J. Reardon, G. Hobbs, W. Coles, et al. Timing analysis for 20 millisecond pulsars in the Parkes Pulsar Timing Array. *Monthly Notices of the Royal Astronomical Society*, 455(2):1751–1769, January 2016.

[46] W. W. Zhu, I. H. Stairs, P. B. Demorest, et al. Testing Theories of Gravitation Using 21-Year Timing of Pulsar Binary J1713+0747. *Astrophysical Journal*, 809(1):41, August 2015.

[47] V. M. Kaspi, J. H. Taylor, and M. F. Ryba. High-Precision Timing of Millisecond Pulsars. III. Long-Term Monitoring of PSRs B1855+09 and B1937+21. *Astrophysical Journal*, 428:713, June 1994.

[48] R. M. Shannon, V. Ravi, L. T. Lentati, et al. Gravitational waves from binary supermassive black holes missing in pulsar observations. *Science*, 349(6255):1522–1525, September 2015.

[49] George Hobbs, Shi Dai, Richard N. Manchester, et al. The role of FAST in pulsar timing arrays. *Research in Astronomy and Astrophysics*, 19(2):020, February 2019.

[50] M. Kramer and B. Stappers. Pulsar Science with the SKA. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 36, April 2015.

[51] NANOGrav Collaboration. Science with the next-generation vla and pulsar timing arrays. *arXiv e-prints*, page arXiv:1810.06594, 2018.

[52] Melissa Anholm, Stefan Ballmer, Jolien D. E. Creighton, Larry R. Price, and Xavier Siemens. Optimal strategies for gravitational wave stochastic background searches in pulsar timing data. *Phys. Rev. D*, 79(8):id. 084030, sep 2008.

[53] Jacob T. VanderPlas. Understanding the Lomb-Scargle Periodogram. *Astrophysical Journal Supplement Series*, 236(1):16, May 2018.

[54] Planck Collaboration, P. A. R. Ade, N. Aghanim, et al. Planck 2015 results. XIII. Cosmological parameters. *Astronomy and Astrophysics*, 594:A13, September 2016.

[55] Mohan Rajagopal and Roger W Romani. Ultra-low frequency gravitational radiation from massive black hole binaries. 1995.

[56] John Kormendy and Luis C. Ho. Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies. *Annual Review of Astronomy and Astrophysics*, 51(1):511–653, August 2013.

[57] Francesco Shankar, Mariangela Bernardi, Ravi K. Sheth, et al. Selection bias in dynamically measured supermassive black hole samples: its consequences and the quest for the most fundamental relation. *Monthly Notices of the Royal Astronomical Society*, 460(3):3119–3142, August 2016.

[58] Kayhan Gültekin, Douglas O. Richstone, Karl Gebhardt, et al. The M-σ and M-L Relations in Galactic Bulges, and Determinations of Their Intrinsic Scatter. *Astrophysical Journal*, 698(1):198–221, June 2009.

[59] A. Sesana, E. Barausse, M. Dotti, and E. M. Rossi. Linking the Spin Evolution of Massive Black Holes to Galaxy Kinematics. *Astrophysical Journal*, 794(2):104, October 2014.

[60] Adam Muzzin, Danilo Marchesini, Mauro Stefanon, et al. The Evolution of the Stellar Mass Functions of Star-forming and Quiescent Galaxies to z = 4 from the COSMOS/UltraVISTA Survey. *Astrophysical Journal*, 777(1):18, November 2013.

[61] Kameswara Bharadwaj Mantha, Daniel H. McIntosh, Ryan Brennan, et al. Major merging history in CANDELS. I. Evolution of the incidence of massive galaxy-galaxy pairs from z = 3 to z $\sim$ 0. *Monthly Notices of the Royal Astronomical Society*, 475(2):1549–1573, April 2018.

[62] A. Sesana, A. Vecchio, and M. Volonteri. Gravitational waves from resolvable massive black hole binary systems and observations with Pulsar Timing Arrays. sep 2008.

[63] M. C. Begelman, R. D. Blandford, and M. J. Rees. Massive black hole binaries in active galactic nuclei. *Nature*, 287(5780):307–309, September 1980.

[64] N. Roos. Galaxy mergers and active galactic nuclei. *Astronomy and Astrophysics*, 104:218–228, December 1981.

[65] Miloš Milosavljević and David Merritt. The Final Parsec Problem. In Joan M. Centrella, editor, *The Astrophysics of Gravitational Wave Sources*, volume 686 of *American Institute of Physics Conference Series*, pages 201–210, October 2003.

[66] Miguel Preto, Ingo Berentzen, Peter Berczik, and Rainer Spurzem. Fast Coalescence of Massive Black Hole Binaries from Mergers of Galactic Nuclei: Implications for Low-frequency Gravitational-wave Astrophysics. *Astrophysical Journal Letters*, 732(2):L26, May 2011.

[67] Diego J. Muñoz, Ryan Miranda, and Dong Lai. Hydrodynamics of Circumbinary Accretion: Angular Momentum Transfer and Binary Orbital Evolution. *Astrophysical Journal*, 871(1):84, January 2019.

[68] Irina Dvorkin and Enrico Barausse. The nightmare scenario: measuring the stochastic gravitational wave background from stalling massive black hole binaries with pulsar timing arrays. *Monthly Notices of the Royal Astronomical Society*, 470(4):4547–4556, October 2017.

[69] Matteo Bonetti, Alberto Sesana, Enrico Barausse, and Francesco Haardt. Post-Newtonian evolution of massive black hole triplets in galactic nuclei - III. A robust lower limit to the nHz stochastic background of gravitational waves. *Monthly Notices of the Royal Astronomical Society*, 477(2):2599–2612, June 2018.

[70] Taeho Ryu, Rosalba Perna, Zoltán Haiman, Jeremiah P. Ostriker, and Nicholas C. Stone. Interactions between multiple supermassive black holes in galactic nuclei: a solution to the final parsec problem. *Monthly Notices of the Royal Astronomical Society*, 473(3):3410–3433, January 2018.

[71] E. S. Phinney. A Practical Theorem on Gravitational Wave Backgrounds. *arXiv e-prints*, pages astro–ph/0108028, August 2001.

[72] C. M. F. Mingarelli, T. Sidery, I. Mandel, and A. Vecchio. Characterizing gravitational wave stochastic background anisotropy with pulsar timing arrays. *Phys. Rev. D*, 88:062005, Sep 2013.

[73] C. Tiburzi, G. Hobbs, M. Kerr, et al. A study of spatial correlations in pulsar timing array data. *Monthly Notices of the Royal Astronomical Society*, 455(4):4339–4350, February 2016.

[74] S. R. Taylor, L. Lentati, S. Babak, et al. All correlations must die: Assessing the significance of a stochastic gravitational-wave background in pulsar timing arrays. *Physical Review D*, 95(4):042002, February 2017.

[75] S. Babak, A. Petiteau, A. Sesana, et al. European Pulsar Timing Array limits on continuous gravitational waves from individual supermassive black hole binaries. *Monthly Notices of the Royal Astronomical Society*, 455(2):1665–1679, January 2016.

[76] X. J. Zhu, G. Hobbs, L. Wen, et al. An all-sky search for continuous gravitational waves in the Parkes Pulsar Timing Array data set. *Monthly Notices of the Royal Astronomical Society*, 444(4):3709–3720, November 2014.

[77] K. Aggarwal, Z. Arzoumanian, P. T. Baker, et al. The NANOGrav 11 yr Data Set: Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries. *Astrophysical Journal*, 880(2):116, August 2019.

[78] B. B. P. Perera, B. W. Stappers, S. Babak, et al. Improving timing sensitivity in the microhertz frequency regime: limits from PSR J1713+0747 on gravitational waves produced by supermassive black hole binaries. *Monthly Notices of the Royal Astronomical Society*, 478(1):218–227, July 2018.

[79] L. Lentati, S. R. Taylor, C. M. F. Mingarelli, et al. European Pulsar Timing Array limits on an isotropic stochastic gravitational-wave background. *Monthly Notices of the Royal Astronomical Society*, 453(3):2576–2598, November 2015.

[80] Z. Arzoumanian, P. T. Baker, A. Brazier, et al. The NANOGrav 11 Year Data Set: Pulsar-timing Constraints on the Stochastic Gravitational-wave Background. *Astrophysical Journal*, 859(1):47, May 2018.

[81] P. B. Demorest, R. D. Ferdman, M. E. Gonzalez, et al. Limits on the Stochastic Gravitational Wave Background from the North American Nanohertz Observatory for Gravitational Waves. *Astrophysical Journal*, 762(2):94, January 2013.

[82] NANOGrav Collaboration, Zaven Arzoumanian, Adam Brazier, et al. The NANOGrav Nine-year Data Set: Observations, Arrival Time Measurements, and Analysis of 37 Millisecond Pulsars. *Astrophysical Journal*, 813(1):65, November 2015.

[83] C. Mingarelli, P. Baker, S. Chen, et al. New Limits on the Gravitational Wave Background from the International Pulsar Timing Array. In *American Astronomical Society Meeting Abstracts*, American Astronomical Society Meeting Abstracts, page 433.04, January 2020.

[84] L. Wen and B. F. Schutz. Coherent network detection of gravitational waves: the redundancy veto. *Classical and Quantum Gravity*, 22:S1321–S1335, September 2005.

[85] D. R. B. Yardley, G. B. Hobbs, F. A. Jenet, et al. The sensitivity of the Parkes Pulsar Timing Array to individual sources of gravitational waves. *Monthly Notices of the Royal Astronomical Society*, 407(1):669–680, September 2010.

[86] M. Rakhmanov. Rank deficiency and Tikhonov regularization in the inverse problem for gravitational-wave bursts. *Classical and Quantum Gravity*, 23:S673–S685, October 2006.

[87] D. A. Shaddock. Operating LISA as a Sagnac interferometer. *Physical Review D*, 69(2):022001, January 2004.

[88] X.-J. Zhu, L. Wen, G. Hobbs, et al. Detection and localization of single-source gravitational waves with pulsar timing arrays. *Monthly Notices of the Royal Astronomical Society*, 449:1650–1663, May 2015.

[89] X.-J. Zhu, L. Wen, J. Xiong, et al. Detection and localization of continuous gravitational waves with pulsar timing arrays: the role of pulsar terms. *Monthly Notices of the Royal Astronomical Society*, 461:1317–1327, September 2016.

[90] J. S. Hazboun and S. L. Larson. Null-stream pointing with pulsar timing arrays. *ArXiv e-prints*, page ArXiv:1607.03459, July 2016.

[91] K. J. Lee, N. Wex, M. Kramer, et al. Gravitational wave astronomy of single sources with a pulsar timing array. *Monthly Notices of the Royal Astronomical Society*, 414:3251–3264, July 2011.

[92] S. R. Taylor, E. A. Huerta, J. R. Gair, and S. T. McWilliams. Detecting Eccentric Supermassive Black Hole Binaries with Pulsar Timing Arrays: Resolvable Source Strategies. *Astrophysical Journal*, 817(1):70, January 2016.

[93] L. Lentati, P. Alexander, M. P. Hobson, et al. Hyper-efficient model-independent Bayesian method for the analysis of pulsar timing data. *Physical Review D*, 87(10):104021, May 2013.

[94] Piotr Jaranowski and Andrzej Królak. Searching for gravitational waves from known pulsars using the $\mathcal{F}$ and $\mathcal{G}$ statistics. *Classical and Quantum Gravity*, 27(19):194015, October 2010.

[95] M. Vallisneri. Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects. *Physical Review D*, 77(4):042001, February 2008.

[96] Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, et al. The NANOGrav Nine-year Data Set: Limits on the Isotropic Stochastic Gravitational Wave Background. *Astrophysical Journal*, 821(1):13, April 2016.

[97] Stephen Taylor, Justin Ellis, and Jonathan Gair. Accelerated Bayesian model-selection and parameter-estimation in continuous gravitational-wave searches with pulsar-timing arrays. *Physical Review D*, 90(10):104028, November 2014.

[98] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Multi-messenger Observations of a Binary Neutron Star Merger. *Astrophysical Journal Letters*, 848:L12, October 2017.

[99] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A. *Astrophysical Journal Letters*, 848:L13, October 2017.

[100] R. Chornock, E. Berger, D. Kasen, et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. IV. Detection of Near-infrared Signatures of r-process Nucleosynthesis with Gemini-South. *Astrophysical Journal Letters*, 848:L19, October 2017.

[101] B. F. Schutz. Determining the Hubble constant from gravitational wave observations. *Nature*, 323(6086):310–311, September 1986.

[102] B. P. Abbott, R. Abbott, T. D. Abbott, et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551:85–88, November 2017.

[103] M. Fishbach, R. Gray, I. Magaña Hernandez, et al. A Standard Siren Measurement of the Hubble Constant from GW170817 without the Electromagnetic Counterpart. *Astrophysical Journal Letters*, 871(1):L13, January 2019.

[104] M. Dotti, A. Sesana, and R. Decarli. Massive Black Hole Binaries: Dynamical Evolution and Observational Signatures. *Advances in Astronomy*, 2012:940568, 2012.

[105] A. Sesana, C. Roedig, M. T. Reynolds, and M. Dotti. Multimessenger astronomy with pulsar timing and X-ray observations of massive black hole binaries. *Monthly Notices of the Royal Astronomical Society*, 420:860–877, February 2012.

[106] Takamitsu Tanaka, Kristen Menou, and Zoltán. Haiman. Electromagnetic counterparts of supermassive black hole binaries resolved by pulsar timing arrays. *Monthly Notices of the Royal Astronomical Society*, 420(1):705–719, February 2012.

[107] Zoltán Haiman. Electromagnetic chirp of a compact binary black hole: A phase template for the gravitational wave inspiral. *Physical Review D*, 96(2):023004, July 2017.

[108] Sarah Burke-Spolaor. Multi-messenger approaches to binary supermassive black holes in the ‘continuous-wave’ regime. *Classical and Quantum Gravity*, 30(22):224013, November 2013.

[109] Luke Zoltan Kelley, Laura Blecha, Lars Hernquist, Alberto Sesana, and Stephen R. Taylor. Single sources in the low-frequency gravitational wave sky: properties and time to detection by pulsar timing arrays. *Monthly Notices of the Royal Astronomical Society*, 477:964–976, June 2018.

[110] B. M. B. Henriques, S. D. M. White, G. Lemson, et al. Confronting theoretical models with the observed evolution of the galaxy population out to z= 4. *Monthly Notices of the Royal Astronomical Society*, 421:2904–2916, April 2012.

[111] Volker Springel, Simon D. M. White, Adrian Jenkins, et al. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435(7042):629–636, June 2005.

[112] M. Longhetti and P. Saracco. Stellar mass estimates in early-type galaxies: procedures, uncertainties and models dependence. *Monthly Notices of the Royal Astronomical Society*, 394:774–794, April 2009.

[113] C. Shapiro, D. J. Bacon, M. Hendry, and B. Hoyle. Delensing gravitational wave standard sirens with shear and flexion maps. *Monthly Notices of the Royal Astronomical Society*, 404:858–866, May 2010.

[114] A. Sesana. Systematic investigation of the expected gravitational wave signal from supermassive black hole binaries in the pulsar timing band. *Monthly Notices of the Royal Astronomical Society*, 433:L1–L5, June 2013.

[115] M. Vogelsberger, S. Genel, V. Springel, et al. Properties of galaxies reproduced by a hydrodynamic simulation. *Nature*, 509(7499):177–182, May 2014.

[116] S. McAlpine, J. C. Helly, M. Schaller, et al. The EAGLE simulations of galaxy formation: Public release of halo and galaxy catalogues. *Astronomy and Computing*, 15:72–89, April 2016.

[117] Q. Guo, S. White, M. Boylan-Kolchin, et al. From dwarf spheroidals to cD galaxies: simulating the galaxy population in a ΛCDM cosmology. *Monthly Notices of the Royal Astronomical Society*, 413:101–131, May 2011.

[118] A. Sesana, F. Shankar, M. Bernardi, and R. K. Sheth. Selection bias in dynamically measured supermassive black hole samples: consequences for pulsar timing arrays. *Monthly Notices of the Royal Astronomical Society*, 463:L6–L11, November 2016.

[119] P. A. Rosado and A. Sesana. Targeting supermassive black hole binaries and gravitational wave sources for the pulsar timing array. *Monthly Notices of the Royal Astronomical Society*, 439:3986–4010, April 2014.

[120] Chiara M. F. Mingarelli, T. Joseph W. Lazio, Alberto Sesana, et al. The local nanohertz gravitational-wave landscape from supermassive black hole binaries. *Nature Astronomy*, 1:886–892, November 2017.

[121] L. Z. Kelley, L. Blecha, and L. Hernquist. Massive black hole binary mergers in dynamical galactic environments. *Monthly Notices of the Royal Astronomical Society*, 464:3131–3157, January 2017.

[122] A. Sesana, Z. Haiman, B. Kocsis, and L. Z. Kelley. Testing the Binary Hypothesis: Pulsar Timing Constraints on Supermassive Black Hole Binary Candidates. *Astrophysical Journal*, 856:42, March 2018.

[123] K. Inayoshi, K. Ichikawa, and Z. Haiman. Gravitational Waves from Supermassive Black Hole Binaries in Ultraluminous Infrared Galaxies. *Astrophysical Journal Letters*, 863:L36, August 2018.

[124] A. Miguel Holgado, Alberto Sesana, Angela Sandrinelli, et al. Pulsar timing constraints on the Fermi massive black hole binary blazar population. *MNRAS Letters*, 481(1):L74–L78, November 2018.

[125] S. Alam, F. D. Albareti, C. Allende Prieto, et al. The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III. *Astrophysical Journal Supplement Series*, 219:12, July 2015.

[126] N. Kaiser, H. Aussel, B. E. Burke, et al. Pan-STARRS: A Large Synoptic Survey Telescope Array. In J. A. Tyson and S. Wolff, editors, *Survey and Other Telescope Technologies and Discoveries*, volume 4836 of *Proceedings of the SPIE*, pages 154–164, December 2002.

[127] LSST Science Collaboration, P. A. Abell, J. Allison, et al. LSST Science Book, Version 2.0. *arXiv e-prints*, page ArXiv:0912.0201, December 2009.

[128] Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, et al. The Gaia mission. *Astronomy and Astrophysics*, 595:A1, November 2016.

[129] K. Duncan, C. J. Conselice, A. Mortlock, et al. The mass evolution of the first galaxies: stellar mass functions and star formation rates at $4 < z < 7$ in the CANDELS GOODS-South field. *Monthly Notices of the Royal Astronomical Society*, 444:2960–2984, November 2014.

[130] J. M. Lotz, P. Jonsson, T. J. Cox, and J. R. Primack. Galaxy merger morphologies and time-scales from simulations of equal-mass gas-rich disc mergers. *Monthly Notices of the Royal Astronomical Society*, 391:1137–1162, December 2008.

[131] Y. Meiron and A. Laor. The kinematic signature of the inspiral phase of massive binary black holes. *Monthly Notices of the Royal Astronomical Society*, 433:2502–2510, August 2013.

[132] M. J. Koss, L. Blecha, P. Bernhard, et al. A population of luminous accreting black holes with hidden mergers. *Nature*, 563:214–216, November 2018.

[133] R. Gilmozzi and J. Spyromilio. The European Extremely Large Telescope (E-ELT). *The Messenger*, 127, March 2007.

[134] Jonathan P. Gardner, John C. Mather, Mark Clampin, et al. The James Webb Space Telescope. *Astrophysics and Space Science Proceedings*, 10:1, January 2009.

[135] K. Nandra, D. Barret, X. Barcons, et al. The Hot and Energetic Universe: A White Paper presenting the science theme motivating the Athena+ mission. *arXiv e-prints*, page ArVix:1306.2307, June 2013.

[136] A. Merloni, P. Predehl, W. Becker, et al. eROSITA Science Book: Mapping the Structure of the Energetic Universe. *arXiv e-prints*, page ArXiv:1209.3114, September 2012.

[137] G. Hobbs, A. Archibald, Z. Arzoumanian, et al. The International Pulsar Timing Array project: using pulsars as a gravitational wave detector. *Classical and Quantum Gravity*, 27(8):084013, April 2010.

[138] Elinore Roebber. Ephemeris Errors and the Gravitational-wave Signal: Harmonic Mode Coupling in Pulsar Timing Array Searches. *Astrophysical Journal*, 876(1):55, May 2019.

[139] R. N. Manchester, G. B. Hobbs, A. Teoh, and M. Hobbs. The Australia Telescope National Facility Pulsar Catalogue. *Astronomical Journal*, 129(4):1993–2006, April 2005.

[140] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[141] S. van der Walt, S. C. Colbert, and G. Varoquaux. The numpy array: A structure for efficient numerical computation. *Computing in Science Engineering*, 13(2):22–30, 2011.

[142] Walter Del Pozzo and John Veitch. Cpnest, 2019. `https://doi.org/10.5281/zenodo.835874`.

[143] John Skilling. Nested Sampling. In Rainer Fischer, Roland Preuss, and Udo Von Toussaint, editors, *American Institute of Physics Conference Series*, volume 735 of *American Institute of Physics Conference Series*, pages 395–405, November 2004.

[144] Leo P. Singer and Larry R. Price. Rapid bayesian position reconstruction for gravitational-wave transients. *Phys. Rev. D*, 93:024013, Jan 2016.