Detecting low frequency gravitational waves : with the international pulsar timing array telescopes

Mikel Falxa

UNIVERSITÉ PARIS CITÉ

ÉCOLE DOCTORALE 560

SCIENCES DE LA TERRE ET DE L'ENVIRONNEMENT ET PHYSIQUE DE L'UNIVERS

AstroParticules et Cosmologie (APC)

groupe Cosmologie et Gravitation

Detecting low frequency Gravitational Waves

with the International Pulsar Timing Array telescopes

Par MIKEL FALXA

Thèse de doctorat de PHYSIQUE DE L'UNIVERS

Co-dirigée par STANISLAV BABAK
et par ANTOINE PETITEAU

Présentée et soutenue publiquement le 16 décembre 2022 devant le jury composé de :

STANISLAV BABAK	Directeur de Recherche (APC)	Directeur de Thèse
ÉRIC CHASSANDE-MOTTIN	Directeur de Recherche (APC)	Président du jury
ANTOINE PETITEAU	Chercheur (APC/CEA)	Co-Directeur de Thèse
ALBERTO SESANA	Professor (Università degli studi di Milano Bicocca)	Examinateur
CATERINA TIBURZI	Researcher (Osservatorio Astronomico di Cagliari)	Examinateur
JORIS VERBIEST	Professor (Bielefeld University)	Rapporteur
SARAH VIGELAND	Assistant Professor (University of Wisconsin Milwaukee)	Rapporteur

ACKNOWLEDGEMENTS

Lehenik eskertu nahi ditut ene famili osoa eta lagunak, bereziki ene aitamak, milesker sustengaturik eta beti hor izanik.

I thank Stas and Antoine for their sympathy and always good advice throughout my PhD, as well as all the people in the PTA collaboration whom I have been working with.

I would also like to thank my examiners and reporters for accepting to evaluate my thesis and Alberto for wearing the flamingo shirt matching his hair at my defence.

Finally, I thank all my lab mates at APC and the shelf in our office.

CONTENTS

LIST OF FIGURES

RÉSUMÉ

La Relativité Générale (RG) prédisait l'existence des Ondes Gravitationnelles (OG) il y a 100 ans de cela. Dans cette théorie, l'espace-temps n'est plus une entité rigide et il peut être courbé par la présence de matière et d'énergie. Les OGs sont précisément le résultat de la propagation de ces courbures dans l'espace-temps. Elles sont de très faible amplitude et leur détection représente un véritable défi technologique. Seules des sources très massives peuvent émettre des OGs suffisamment puissantes pour être mesurées. En 2016, la collaboration LIGO/Virgo détectait pour la première fois une OG provenant de la fusion de deux trous noirs. Cette découverte a révélé un tout nouveau spectre d'observations astrophysiques qui est aujourd'hui un domaine de recherche très actif. Pour détecter une OG, il faut pouvoir mesurer les courbures de l'espace-temps induites par le passage de celle-ci. Une méthode, basée sur un réseau de pulsars chronométrés ou Pulsar Timing Array (PTA), propose d'utiliser les pulsars (des étoiles à neutron en rotation rapide qui sont perçues sur Terre comme des séries de pulses de lumière très régulièrement espacés dans le temps) comme des "horloges cosmiques" pour mesurer les perturbations dues à l'OG. Ce manuscript se concentre sur la détection d'OGs basse fréquence (nanohertz) en utilisant les données de chronométrage haute précision d'un réseau de pulsars milliseconde obtenu à l'aide de radiotéléscopes.

Le premier chapitre donne un aperçu du bagage théorique nécessaire pour comprendre l'origine des OGs. Pour comprendre la RG, il est essentiel de comprendre les bases de la géométrie différentielle et Riemanienne. Elles donnent un cadre d'étude des espaces courbes qui est le coeur de la description mathématique de la RG. L'équation centrale de la RG est l'équation de champs d'Einstein. Elle décrit la relation qui existe entre la géométrie de l'espace-temps d'une part et la présence de matière d'autre part. Pour des très petites perturbations de la métrique de l'espace-temps, il est possible de montrer que cette équation s'apparente à une équation de propagation d'une onde. Cette dérivation est détaillée dans ce chapitre car elle correspond à la description mathémathique d'une OG qui nous sera utile pour la suite. Typiquement, les sources d'OG seront des systèmes binaires d'objets très massifs (trous noirs, étoiles à neutrons, ...) comme il a été le cas pour la première OG détectée en 2016. Nous présentons ici la forme des OGs émises par des systèmes binaires en orbite circulaire l'un autour de l'autre. Il apparaît que l'amplitude de ces ondes est proportionnelle à la masse des objets du système puis inversément proportionnelle à la distance de la source. Il en vient alors que plus le système binaire est massif ou plus il est proche, plus les effets dues à l'OG émise se feront sentir. Les objets les plus massifs de l'Univers que l'on connaît à ce jour sont les trous noirs supermassifs. Leur origine reste encore un mystère car il est difficile d'établir un modèle de formation de ces astres en un temps aussi "court" que celui de l'âge de l'Univers. Notamment, l'observation de noyaux de galaxies actifs très anciens compor-

tant des trous noirs extrêmement massifs en leur centre indique que ces objets se sont formés rapidement, probablement par fusion d'autres trous noirs supermassifs. Cependant, nous ne disposons d'aucune observation de fusion de trous noirs supermassifs à ce jour. Nous argumentons qu'un réseau de pulsars chronométrés (Pulsar Timing Array ou PTA) pourrait détecter des OGs générées par des systèmes de trous noirs super massifs binaires. Comme indiqué précédemment, les pulsars peuvent êtres utilisés comme des "horloges cosmiques" donnant accès à une mesure des propriétés de l'espace-temps dans notre Galaxie. Nous dérivons une expression mathématique pour les résidus temporels induits par les OGs dans la donnée chronométrée des pulsars. Lorsqu'un tel signal est stable et monochromatique, on le qualifie d'Onde Gravitationnelle Continue (OGC). Ce manuscript est centré sur l'étude des OGCs émises par des systèmes binaires de trous noirs supermassifs.

Dans le second chapitre, nous présentons les méthodes d'analyse de donnée pour la détection d'OGs dans PTA. Pour cela, nous expliquons brièvement l'inférence Bayesienne qui est au coeur de ces méthodes. Elle nous permet d'évaluer la significativité statistique de nos modèles ainsi que les valeurs des paramètres de signaux que l'on cherche à détecter en fonction de la donnée mesurée (dans notre cas, la donnée chronométrée des pulsars). Une quantité centrale de l'inférence Bayesienne est le Facteur de Bayes. Il donne une mesure comparative des "poids" statistiques de différents modèles et est crucial pour sélectionner ceux qui reproduisent le mieux la donnée observée. Nos modèles et méthodes d'analyse doivent être capables de : (i) quantifier tous les bruits présents dans la donnée pour pouvoir isoler le signal d'OG désiré (ii) estimer un facteur de Bayes pour les OGs que l'on cherche à détecter pour affirmer leur présence dans la donnée (iii) établir une limite haute d'amplitude de l'OG pour sa détection dans le cas où on ne détecte rien. Les principaux bruits, outre les erreurs systématiques et bruits des appareils de mesures (ici les téléscopes radio), présents dans la donnée de chronométrage sont intrinsèques aux pulsars et corrélés dans le temps. Nous avons des bruits dits "rouges", dominants en basse fréquence, dus à une irrégularité dans la rotation même des pulsars et induisant de légères variations dans les temps d'arrivée des pulses chronométrés sur Terre. Nous avons aussi un bruit de mesure de dispersion, du à la propagation des photons composants les pulses dans le milieu interstellaire. Durant son voyage depuis le pulsar jusqu'à la Terre, un pulse de lumière interagit avec les électrons interstellaire, ce qui a pour effet de retarder son temps d'arrivée. Lorsque la densité d'électrons varie sur l'axe pulsar-Terre, cela induit une perturbation dans le temps d'arrivée des pulses chronométrés. La présence de ces bruits détériore notre sensibilité aux OGs et ils doivent être modélisés avec soin. Enfin, nous présentons comment ces modèles sont testés grâce à des méthodes de Monte-Carlo par chaînes de Markov (MCMC) dont le fonctionnement est expliqué.

Le troisième chapitre présente les résultats obtenus avec le second ensemble de données de la collaboration PTA internationale (IPTA DR2) qui combine les données de :

l'observatoire d'OG nanohertz d'Amérique du Nord (NANOGrav), le PTA Européen (EPTA) et le Parkes PTA (PPTA) en Australie. Cet ensemble de donnée étant assez conséquent, les ressources en calcul nécessaires pour l'analyser seront elles aussi condisérables. Pour palier ce problème, il est possible de classer les pulsars en fonction de leur sensibilité moyenne à une OGC de fréquence donnée. Pour ce faire, de fausses OGCs sont simulées dans la donnée de chronométrage pour estimer le rapport signal sur bruit de chaque pulsar. Ce rapport signal sur bruit quantifie la réponse individuelle des pulsars à l'OGC injectée et leur classement se fait suivant les valeurs obtenues. Il apparaît alors que certains pulsars dominent complètement le réseau de par leur sensibilité supérieure. En sélectionnant un sous-ensemble de pulsars qui complètent au moins 95% de la sensibilité totale du réseau, nous réduisons la donnée nécessaire aux analyses, et donc le temps de calcul, tout en conservant une bonne sensibilité aux OGCs. Après avoir classé les pulsars, nous appliquons les méthodes présentées dans le chapitre précédent pour tester la présence d'OGCs dans IPTA DR2. Nous n'avons trouvé aucune preuve de leur présence. En conséquence, nous avons établi une limite haute de détection pour l'amplitude des OGCs et avons trouvé que IPTA DR2 donne la meilleure sensibilité à ce jour pour des signaux dont la fréquence est plus haute que 11 nHz. Dernièrement, les collaborations PTA ont mis en évidence la présence d'un bruit "rouge" commun à tous les pulsars (ayant la même densité spectrale de puissance entre chaque pulsar). L'origine de ce bruit est encore à déterminer. Son inclusion dans notre modèle détériore la sensibilité aux OGCs du réseau en basse fréquence. De plus, nous montrons qu'une mauvaise modélisation des bruits intrisèques aux pulsars peut mener à une fausse détection. En général, les bruits sont décrits de la même façon pour tous les pulsars. Cependant, des études récentes ont montré qu'adapter nos modèles de bruits individuellement pour chaque pulsar donnait une meilleure représentation de la donnée. Les meilleurs modèles de bruits sont sélectionnés par calcul du facteur de Bayes (introduit dans le second chapitre). Avant l'utilisation des modèles "sur-mesure", nous détections un signal similaire à une OGC de 51nHz, supporté par le facteur de Bayes. L'introduction des nouveaux modèles a complètement absorbé cette composante. L'origine de ce signal est encore à détérminer mais il est probable qu'il résulte d'interactions complexes entre différentes éléments de notre modèle. C'est pourquoi une bonne modélisation des modèles de bruits dans PTA est cruciale à la recherche d'OGs car des erreurs dans celui-ci peuvent mener à des fausses détections.

Le dernier chapitre présente une nouvelle méthode basée sur les Estimations par Noyaux (Kernel Density Estimation ou KDE) qui a été développée durant mon doctorat pour optimiser les techniques d'analyse de donnée pour la recherche d'OGs et en particulier pour améliorer le temps effectif d'échantillonage des méthodes MCMC. Le KDE est une méthode d'estimation de densité de probabilité des paramètres d'un ensemble de donnée. Pour estimer ces densités de probabilité, il est possible d'utiliser des histogrammes.

Contrairement à ces derniers, le KDE a l'avantage d'être une fonction continue et définie dans tout l'espace. Pour qu'il décrive fidèlement la donnée, il faut l'adapter à celle-ci à l'aide de méthodes d'optimisation parfois coûteuses en temps de calcul. Nous présentons ici une nouvelle méthode d'optimisation rapide par linéarisation d'un problème de minimisation d'erreur quadratique. Ensuite, nous présentons une nouvelle méthode d'estimation des corrélations entre paires de paramètres basée sur la divergence de Jensen-Shannon à partir de laquelle un algorithme de groupement des paramètres est développé. Les KDEs produits avec ces méthodes sont utilisés en MCMC pour proposer de nouveaux points dans les régions intéressantes de l'espace des paramètres pour aider l'échantillonage. D'abord, nous avons testé les KDEs avec la donnée IPTA DR2 pour la recherche d'OGCs. Dans ce cas, le temps d'échantillonage effectif était réduit d'un facteur 2 ou 3 ce qui indiquait un bon fonctionnement de la méthode. Par la suite, nous avons testé une méthode adaptative du KDE avec de la donnée simulée de LISA (futur observatoire spatial d'OG) pour la recherche de signaux gravitationnels de binaires galactiques (systèmes binaires de naines blanches). Elle s'est avérée moins efficace avec la donnée LISA pour laquelle le temps effectif d'échantillonage n'était que légèrement réduit (de 17%).

En résumé, ce manuscript explore différents aspects de l'analyse de donnée pour la recherche de signaux gravitationnels. Il est principalement axé sur la donnée PTA (en particulier la donnée de la collaboration internationale IPTA) et la recherche d'OGC. Ces travaux ont mené à l'écriture de 3 articles scientifiques (dont 2 en tant que premier auteur) qui sont, alors que j'écris ces lignes, en cours de soumission à des revues scientifiques. Pour le moment, aucune détection d'OGC n'est à signaler dans la donnée IPTA. L'observation des pulsars et leur chronométrage est constamment en cours, donnant accès à toujours plus de donnée. De plus, la construction de nouveaux téléscopes radio haute précision (comme MeerKAT ou SKA) devrait augmenter notre sensibilité aux OGs, faisant de la première détection d'un signal gravitationnel en provenance d'un système binaire de trous noirs supermassifs un espoir bien concret pour les années à venir.

Mots clés: Onde gravitationnelle - Pulsars - Trou noir supermassif - Analyse de données

ABSTRACT : FRANÇAIS

Ce manuscript est centré sur la détection d'Ondes Gravitationnelles (OG) basse fréquence en utilisant les données de chronométrage haute précision d'un réseau de pulsars milliseconde obtenues à l'aide de radiotéléscopes. Le premier chapitre donne un apercu du bagage théorique nécessaire pour comprendre l'origine des OGs. Après avoir présenté les bases de la Relativité Générale (RG), nous argumentons qu'un réseau de pulsars chronométrés (Pulsar Timing Array ou PTA) pourrait détecter des OGs générées par des systèmes de Trous Noirs Super Massifs Binaires (TNSMB) et nous dérivons une expression mathématique pour les résidus temporels qui seraient induits dans la donnée chronométrée. Lorsqu'un tel signal est stable et monochromatique, on le qualifie d'Onde Gravitationnelle Continue (OGC). Dans le second chapitre, nous présentons les méthodes d'analyse de donnée pour la détection d'OGs dans PTA. Nous expliquons brièvement l'inférence Bayesienne et nous détaillons comment les bruits intrinsèques aux pulsars peuvent être modélisés. Nous montrons que pour estimer la significativité statistique de nos modèles de signal et de bruits, nous calculons le Facteur de Bayes (FB) en utilisant des hypermodèles et des méthodes de Monte-Carlo par chaînes de Markov (MCMC). Dans le chapite trois, nous discutons des résultats obtenus avec le second ensemble de données de la collaboration PTA internationale (IPTA DR2) qui combine les données de : l'observatoire d'OG nanohertz d'Amérique du Nord (NANOGrav), le PTA Européen (EPTA) et le Parkes PTA (PPTA) en Australie. Nous avons cherché des OGC mais n'avons trouvé aucune preuve de leur présence dans la donnée. Nous avons établi une limite haute de détection pour l'amplitude des OGCs et avons trouvé que IPTA DR2 donne la meilleure sensibilité à ce jour pour des signaux dont la fréquence est plus haute que 11 nHz. Il a été découvert il y a peu qu'un bruit en basse fréquence et commun à tous les pulsars était présent dans la donnée. L'inclusion de ce dernier dans le modèle détériore la sensibilité aux OGCs du réseau. De plus, nous montrons qu'une mauvaise modélisation des bruits intrisèques aux pulsars peut mener à une fausse détection. Le dernier chapitre présente une nouvelle méthode basée sur les Estimations par Noyaux (Kernel Density Estimation ou KDE) qui a été développée durant mon doctorat pour optimiser les techniques d'analyse de donnée pour la recherche d'OGs. Cette méthode est utilisée avec MCMC pour améliorer son temps effectif d'échantillonage. Elle a été testée avec IPTA DR2 pour la recherche d'OGCs puis avec Sangria, une donnée simulée du futur projet d'observatoire spatial d'OG (le Laser Interferometer Space Antenna ou LISA), pour la recherche de Binaires Galactiques (BG). Nous démontrons que la méthode est efficace pour IPTA DR2 mais est moins performante avec LISA.

Mots clés: Onde gravitationnelle - Pulsars - Trou noir supermassif - Analyse de données

ABSTRACT : ENGLISH

This manuscript focuses on the detection of low frequency Gravitational Waves (GW) using high precision timing observations of millisecond pulsars collected with radio telescopes. The first chapter gives an overview of the theoretical background necessary to understand the origins of GWs. After presenting the basics of General Relativity, we argue that a Pulsar Timing Array (PTA) could detect GWs generated by Super Massive Black Hole Binaries (SMBHBs) and derive a mathematical expression for the expected timing residuals induced in the pulsar timing data. When such a signal is stable and monochromatic, we call it a Continuous Gravitational Wave (CGW). In the second chapter, we present data analysis methods for the detection of GW signals in PTA. We briefly explain Bayesian inference and detail how the noises intrinsic to pulsars can be modeled. We show that to assess the significance of our noise and signal models, we compute the Bayes Factor (BF) using hypermodels and Markov Chain Monte-Carlo (MCMC). In chapter three, we discuss the results obtained with the international PTA second data release (IPTA DR2) combining data from : the North American Nanohertz Observatory for Gravitational Waves (NANOGrav), the European Pulsar Timing Array (EPTA) and the Parkes Pulsar Timing Array (PPTA). We have searched for CGW signals and found no evidence of their presence in the data. We have set an upper limit on the amplitude of CGW and found that IPTA DR2 gives the best sensitivity to date for signals with frequencies above 11 nHz. It has been recently discovered that a low frequency noise common to all pulsars was present in the data. The inclusion of the latter in the model deteriorates the sensitivity of the array to CGW. Moreover, we show that a bad modeling of the intrinsic noise properties of pulsars can lead to false detection. The last chapter presents a new method based on Kernel Density Estimation (KDE) that was developed during my PhD to optimize the data analysis pipelines for GW search. The method is used with MCMC to improve the effective sampling time. It was tested with IPTA DR2 for CGW search and the simulated Laser Interferometer Space Antenna (LISA) dataset "Sangria" for Galactic Binaries (GBs) search. We show that the method is efficient for IPTA DR2 but is less performing with LISA.

Keywords: Gravitational wave - Pulsar - Super massive black hole - Data analysis

CHAPTER 1

THEORETICAL BACKGROUND

Contents

1.1 General relativity

Albert Einstein's General relativity (GR) came out in 1915 as a new theory of gravity and a generalization of special relativity. It no longer describes gravity as an attractive force but rather as a manifestation of the curvature of space-time that is modified in the presence of masses, where free-falling objects are actually inertial and following geodesic trajectories.

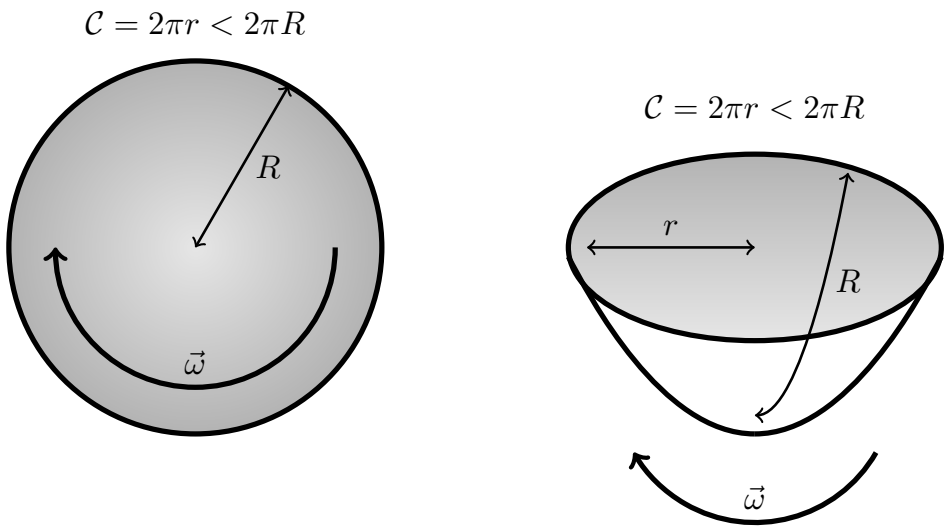

Figure 1.1: Illustration of the Ehrenfest paradox with on the left, a rotating disk with angular velocity $\vec{\omega}$ seen from above. When it rotates, it should see its circumference $\mathcal{C}$ decrease due to relativistic length contraction while its radius R is unaffected. On the right, we have the resolution of the paradox showing that a curved geometry allows smaller circumference without change in radius R.

The latter is a consequence of the equivalence principle saying that no distinction can be made between the motion of a falling object on a planet or in an accelerated frame of reference. The gravitational mass is the same as the inertial mass. Therefore, a free-falling object experiencing weightlessness is in an inertial frame of reference.

Special relativity already tells us that a particle in motion will experience time dilation and length contraction in the tangent direction to its trajectory, which was raising contradictions in the case of circular motions and orbits. One famous *Gedankenexperiment* that is behind the intuition of curvature in GR is called the Ehrenfest paradox : a rotating disk of radius R would, due to length contraction, see its circumference smaller than $2\pi R$. But its radius R, in the direction perpendicular to its trajectory, would remain unchanged. This is only possible in curved space (see figure 1.1). Planets orbiting stars should then be experiencing a curvature of space while following an inertial trajectory. It is possible if heavy masses curve space-time and orbits are geodesics in a curved space.

After years of development, the so called Einstein field equations were derived to express the equivalence between curvature of space-time and distribution of energy. To quote John Archibald Wheeler : "Spacetime tells matter how to move; matter tells spacetime how to curve". The introduction of curvature in the equations of physics led to countless predictions that were all verified experimentally like the anomalous precession of Mercury, gravitational lensing, the expansion of the universe, the existence of black holes, and more recently, gravitational waves, the main topic of this thesis.

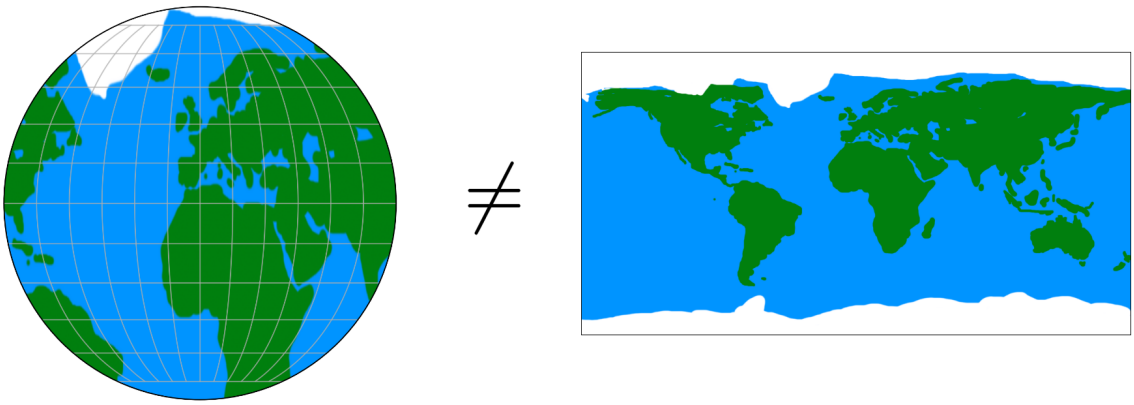

Figure 1.2: The spherical surface of the the Earth (positive curvature) is fundamentally different from a plane (null curvature). Every planisphere of the Earth that we have are just approximations where length measurements are distorted.

1.1.1 The notion of curved space

From simple 1D curve to more elaborate n-D hypersurfaces, the curvature is a measurement of how much we deviate from a straight line or a flat surface. In curved space, many of the general results of flat euclidian space geometry are no longer true. For instance, the sum of the 3 angles of a triangle may not equal 180^o or two locally parallel straight lines (geodesics) may eventually cross.

Carl Friedrich Gauss worked extensively on the matter and derived a scalar quantity, the Gaussian curvature, giving a mathematical measurement of how curved is a surface. He realized that for 2D surfaces, this quantity was invariant under isometric transformations and independent of the 3D space in which it was embedded. This observation is stated in the Theorema Egregium.

The curvature is an intrinsic property of the space. A curved space cannot be folded back into a flat space or the other way around. For example, a world map, as a 2D flat representation of the spherical surface of the Earth, does not accurately reproduce distances and areas. There exists no direct mapping of a curved surface on a flat one that comes without scale distortion issues. The sphericity of the Earth implies positive curvature that is not shared by a flat sheet of paper of null curvature. To that extent, length and angle measurements are fundamentally different in the two.

A pupil of Gauss, Bernhard Riemann, continuing the work previously undertaken by his teacher, generalized the mathematical description of the curvature for N-dimensional spaces. In GR, we make use of the Riemannian geometry to characterize the curvature of space-time.

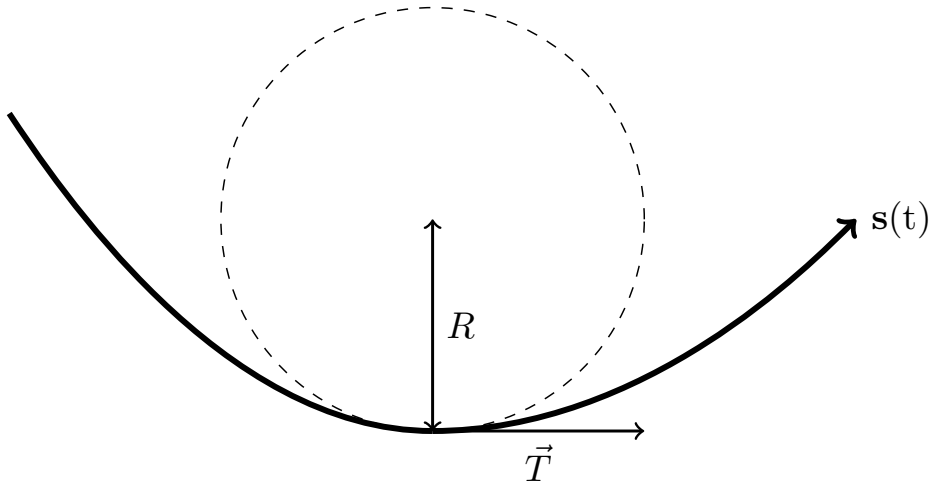

Figure 1.3: We can associate a tangent circle of radius R to each point of **s(t)**. The inverse of the radius defines the curvature. This is the simplest way to think about curvature in 1D.

1.1.2 Mathematical description of a curved space

Historically, different approaches were used to mathematically describe the curvature. The first and maybe simplest is to define a tangent osculating circle at any point of a given parametrized curve $s(t)$ (see figure). The curvature κ is defined as the inverse of the radius of the circle. A straight line has an osculating circle of infinite radius, hence no curvature. This concept can easily be generalized in differential geometry for oriented surfaces of higher dimensions. However it requires the d-dimensional surface to be embedded in a $d+1$-dimensional space.

Riemannian geometry allows the description of the curvature without immersing the surface in a $d+1$ space, following the Theorema Egregium defining it as an intrisic property of the surface. It is done using the concepts of geodesic deviation (see figure 1.5).

Here we will present the standard "toolbox" that is used in GR for calculation in curved space. One can refer to references for more details (see Thorne et al. 2000; Schutz 2022; Hausner 2016).

Basis vectors Basis vectors are tangent vectors at each point of the parameterized surface $f(x^\mu)$, we have :

$$e_\mu = \partial_\mu f. \tag{1.1}$$

They define the basis of a local tangent (and flat) space at point x^μ. We use the notation :

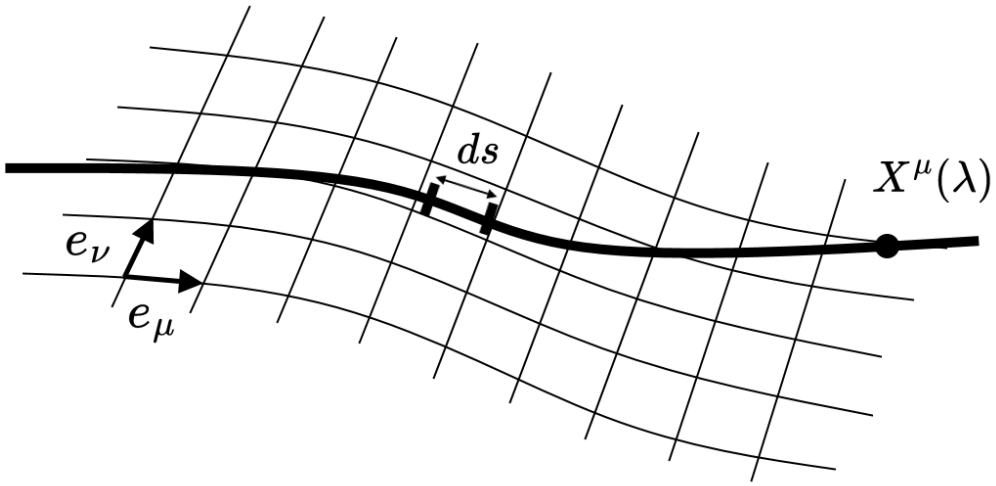

Figure 1.4: In curved space, basis vectors e_μ are function of coordinates x^μ and straight lines are parametrized geodesics $X^\mu(\lambda)$.

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu}, \tag{1.2}$$

to express the derivative with respect to coordinate x^μ.

Metric tensor The metric tensor is the mathematical object that redefines scalar product hence distances/angle measurement. In Euclidian flat space, it is just the identity matrix, but more generally we have :

$$g_{\mu\nu} = e_\mu \cdot e_\nu, \tag{1.3}$$

that is not necessarily diagonal or the identity. It quantifies length and angle distortions at every point in space.

Length measurement Length measurement depends on the metric tensor. We define the element on length of a parametrized curve with components X^μ as :

$$ds^2 = g_{\mu\nu} dX^\mu dX^\nu, \tag{1.4}$$

which defines the scalar product of infinitesimal vector $d\vec{X}$ with itself.

Christoffel symbols The Christoffel symbols $\Gamma^\beta_{\mu\alpha}$ are related to the derivative in the μ direction of basis vector $\mathbf{e}_\alpha$ as :

$$\partial_\mu \mathbf{e}_\alpha = \Gamma^\beta_{\mu\alpha} \mathbf{e}_\beta, \tag{1.5}$$

They are the coefficients that quantify how basis $\mathbf{e}_\alpha$ varies with respect to coordinate μ. They can be expressed in terms of the derivatives of the metrix tensor $g_{\mu\nu}$.

$$\Gamma^\beta_{\mu\alpha} = \frac{1}{2} g^{\beta\lambda} \Big(\partial_\alpha g_{\lambda\mu} + \partial_\mu g_{\lambda\alpha} - \partial_\lambda g_{\mu\alpha} \Big) \tag{1.6}$$

Covariant derivative The covariant derivative is the derivative of a vector $\vec{v}$ taking into account the variations of the basis vectors $\mathbf{e}_\alpha$:

$$\nabla_\mu \vec{v} = \nabla_\mu v^\alpha \mathbf{e}_\alpha = \Big(\partial_\mu v^\beta + v^\alpha \Gamma^\beta_{\mu\alpha} \Big) \mathbf{e}_\beta. \tag{1.7}$$

Geodesic equation The geodesic equation is the equation of a straight line[1] in curved space. Physically, it corresponds to the equation of motion of a non-accelerated object in a curved space.

[1] A straight line is the shortest path between two points. The geodesic equation extends this deifinition to curved space.

$$\frac{d^2X^\mu}{d\lambda^2} + \Gamma^\mu_{\alpha\beta}\frac{dX^\alpha}{d\lambda}\frac{dX^\beta}{d\lambda} = 0, \tag{1.8}$$

with λ the parameter of the curve (corresponding to the proper time τ for massive objects).

Riemann tensor The Riemann tensor is a measurement of geodesic deviation. Indeed, in flat space, two neighbouring geodesics are parallel straight lines but this is no longer true in curved space. We can quantify this effect by measuring the difference in variation of the basis vector $\mathbf{e}_\alpha$ on a closed loop in two directions μ and ν (figure 1.5). Using equation 1.5 we get :

$$R^\beta_{\mu\nu\alpha}\mathbf{e}_\beta = \left(\partial_\mu\partial_\nu - \partial_\nu\partial_\mu\right)\mathbf{e}_\alpha = \left(\partial_\mu\Gamma^\beta_{\nu\alpha} - \partial_\nu\Gamma^\beta_{\mu\alpha} + \Gamma^\lambda_{\nu\alpha}\Gamma^\beta_{\mu\lambda} - \Gamma^\lambda_{\mu\alpha}\Gamma^\beta_{\nu\lambda}\right)\mathbf{e}_\beta. \tag{1.9}$$

This definition can be generalized to any vector using parallel transport and covariant derivatives (Hausner 2016). The Riemann tensor appears when calculating the evolution of the separation vector between two objects free-falling side by side. Physically, it corresponds to the tidal forces.

Ricci tensor The Ricci tensor is obtained from the Riemann tensor $R^\beta_{\mu\nu\alpha}$ by contraction of first index β and third index ν [2]. It is analog to taking the trace of the matrix containing all possible deviation vectors $\vec{R}$ associated with $\vec{e}_\alpha$ being first transported in the μ direction (see figure 1.5), quantifying the strain felt in the directions perpendicular to μ.

$$R_{\mu\alpha} = R^\beta_{\mu\beta\alpha} \tag{1.10}$$

It contains information about volume deformation along geodesics. Component $R_{\mu\alpha}$ is a measurement of volume change when transporting basis vector $\vec{e}_\alpha$ along direction μ. A null Ricci tensor does not necessarily mean that there is no curvature. It just corresponds to volume conservation along geodesics (i.e. experienced contraction and dilation in directions perpendicular to μ compensate). This tensor is symmetric and is more often noted $R_{\mu\nu}$.

Einstein field equations

The Einstein field equations are the central equations of GR. They tell how space-time geometry is affected by the distribution of energy within it. Finding a solution to this equation gives the metric tensor $g_{\mu\nu}$ that we will use to study the dynamics of objects using the geodesic equation.

[2] Contraction with second index gives the same result with opposite sign and contraction with last index is null. This is due to the antisymmetric properties of the Riemann tensor.

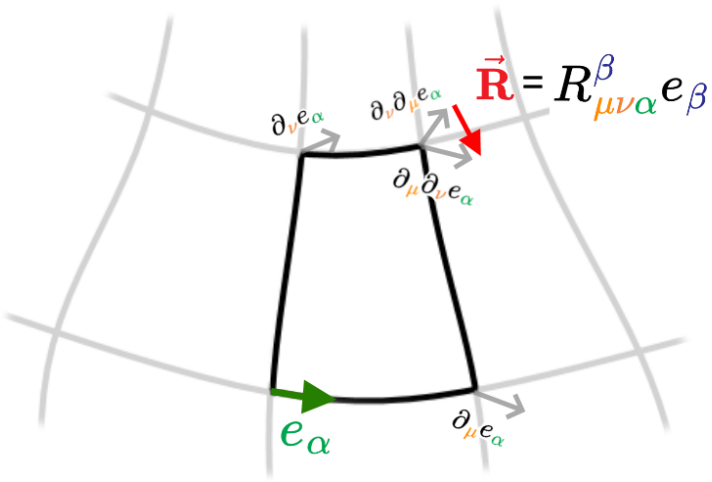

Figure 1.5: The Riemann tensor contains the components of the deviation vectors $\vec{R}$ for all possible transportation of basis vector e_α on a closed loop in directions μ and ν.

The distribution of energy and matter is described by the stress energy tensor $T_{\mu\nu}$. It follows the continuity relation :

$$\nabla^{\mu} T_{\mu\nu} = 0, \tag{1.11}$$

expressing the local conservation of energy and momentum.

At first sight, it could be tempting to write $R_{\mu\nu} = \kappa T_{\mu\nu}$ to map the curvature of space-time to the presence of energy. But this cannot work because $R_{\mu\nu}$ does not have a null covariant derivative. Nonetheless, it possible to build a curvature tensor from the Ricci tensor $R_{\mu\nu}$ that also has null covariant derivative. We call it the Einstein tensor $G_{\mu\nu}$:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}, \tag{1.12}$$

$$\nabla^{\mu} G_{\mu\nu} = 0, \tag{1.13}$$

We can now empirically postulate an equation that relates the stress-energy tensor $T_{\mu\nu}$ and the Einstein tensor $G_{\mu\nu}$ as :

$$G_{\mu\nu} = \kappa T_{\mu\nu}. \tag{1.14}$$

The value of the constant κ is set in such a way that equation 1.14 reduces to Poisson's equation for the gravitational potential in the Newtonian limit. Including the cosmological constant Λ, we have the final expression for the Einstein field equations :

$$\boxed{G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}}. \tag{1.15}$$

For a given $T_{\mu\nu}$, this equation gives a system of non linear second order differential equations for the metric tensor $g_{\mu\nu}$. It is possible to linearize this equation by introducing small perturbations around the flat-space metric.

1.2 Gravitational waves

The following section is a mathematical introduction to gravitational waves inspired from the present references : Maggiore 2007; Babak 2020; Carroll 1997.

1.2.1 Linearized Einstein field equations

We define a metric tensor that is :

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{1.16}$$

with inverse :

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}, \tag{1.17}$$

where $h_{\mu\nu} << \eta^{\mu\nu}$, is a small perturbation we add to the flat-space Minkowski metric $\eta_{\mu\nu}$.

To express the Einstein field equations using this approximation, we first need to write the Ricci tensor $R_{\mu\nu}$. Keeping only first order terms in $h_{\mu\nu}$ and getting rid of higher order terms. We have for the Christoffel symbols :

$$\Gamma^{\beta}_{\mu\alpha} = \frac{1}{2}\eta^{\beta\lambda}\Big(\partial_\alpha h_{\lambda\mu} + \partial_\mu h_{\lambda\alpha} - \partial_\lambda h_{\mu\alpha}\Big) + \mathcal{O}(h^2), \tag{1.18}$$

which gives for the Ricci tensor :

$$\begin{aligned} R_{\mu\nu} &= \partial_\lambda \Gamma^{\lambda}_{\mu\nu} - \partial_\mu \Gamma^{\lambda}_{\lambda\nu} + \mathcal{O}(\Gamma^2) \\ &= \frac{1}{2}\Big(\partial^\beta \partial_\nu h_{\beta\mu} + \partial^\beta \partial_\mu h_{\beta\nu} - \partial^\beta \partial_\beta h_{\mu\nu} - \partial_\mu \partial_\nu h^{\lambda}_{\lambda}\Big), \end{aligned} \tag{1.19}$$

and the Ricci scalar :

$$R = \partial^\beta \partial^\lambda h_{\beta\lambda} - \partial^\beta \partial_\beta h^{\lambda}_{\lambda} \tag{1.20}$$

We can now write the Einstein tensor :

$$G_{\mu\nu} = \frac{1}{2}\Big(\partial^\beta \partial_\nu h_{\beta\mu} + \partial^\beta \partial_\mu h_{\beta\nu} - \Box h_{\mu\nu} - \partial_\mu \partial_\nu h - \eta_{\mu\nu}\big(\partial^\beta \partial^\lambda h_{\beta\lambda} - \Box h\big)\Big). \tag{1.21}$$

Where we have used to simplify notation convention :

$$h^{\lambda}_{\lambda} = h \quad \text{and} \quad \partial^\beta \partial_\beta = \Box \tag{1.22}$$

The physics of GR do not depend on the choice of coordinate system. That is to say, for any local coordinate transformation :

$$x^\mu \to x^\mu + \xi^\mu(x), \tag{1.23}$$

the Einstein field equations will remain unchanged. For such a coordinate transformation, the metric perturbation $h_{\mu\nu}$ rewrites as :

$$h_{\mu\nu} \to h_{\mu\nu} + \partial_\nu \xi_\mu + \partial_\mu \xi_\nu. \tag{1.24}$$

One can verify that the additional terms actually cancel out in equation 1.21. That means we have a gauge (coordinate) freedom which we can use to simplify the expression. We introduce the trace reversed metric[3] that is

[3] At that stage, the trace reversed metric is introduced to render the "wave equation"-like properties of the linearized Einstein field equations.

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h, \tag{1.25}$$

for which the gauge condition becomes :

$$\bar{h}_{\mu\nu} \rightarrow \bar{h}_{\mu\nu} + \partial_\nu \xi_\mu + \partial_\mu \xi_\nu - \eta_{\mu\nu}\partial^\beta \xi_\beta, \tag{1.26}$$

We can rewrite the linearised Einstein tensor in term of this trace reversed metric, yielding

$$G_{\mu\nu} = \frac{1}{2}\Big(-\Box \bar{h}_{\mu\nu} + \partial^\beta \partial_\nu \bar{h}_{\mu\beta} + \partial^\beta \partial_\mu \bar{h}_{\nu\beta} - \frac{1}{2}\eta_{\mu\nu}\partial^\beta \partial^\lambda \bar{h}_{\beta\lambda}\Big). \tag{1.27}$$

Using the coordinate freedom expressed in (1.26) we can now derive the divergence of $\bar{h}_{\mu\nu}$ as

$$\partial^\mu \bar{h}_{\mu\nu} \rightarrow \partial^\mu \bar{h}_{\mu\nu} + \Box \xi_\nu, \tag{1.28}$$

then choose the ξ_μ in such a way so that they obey $\Box \xi_\nu = -\partial^\mu \bar{h}_{\mu\nu}$ to give the Lorenz gauge condition :

$$\partial^\mu \bar{h}_{\mu\nu} = 0. \tag{1.29}$$

Using that gauge, the expression of the Einstein tensor enormously simplifies to become :

$$G_{\mu\nu} = -\frac{1}{2}\Box \bar{h}_{\mu\nu} \tag{1.30}$$

giving the final expression of the linearized Einstein field equations for the trace reversed metric :

$$\boxed{\Box \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu}} \tag{1.31}$$

We can already see that this equation corresponds to the equation of propagation of a wave with a source term. Finding solutions to (1.31) give periodic first order perturbations of the metric tensor that we identify as gravitational waves.

1.2.2 Polarization tensor

When considering gravitational waves propagating in the vacuum, we can set the stress-energy tensor to be $T_{\mu\nu} = 0$. In that case, equation (1.31) takes the simple form :

$$\Box \bar{h}_{\mu\nu} = 0. \tag{1.32}$$

Solutions to the latter can be expressed as a linear combination of plane waves with polarization tensor $\epsilon_{\mu\nu}$:

$$\bar{h}_{\mu\nu} = \epsilon_{\mu\nu} e^{ik^\sigma x_\sigma}, \tag{1.33}$$

with k^μ the wave vector.

This expression of the linearized Einstein field equations in the vacuum offer additional coordinate freedom. Indeed, plugging equation (1.26) in (1.32) we have :

$$\Box \bar{h}_{\mu\nu} + \Box\Big(\partial_\nu \xi_\mu + \partial_\mu \xi_\nu - \eta_{\mu\nu}\partial^\beta \xi_\beta\Big) = 0, \tag{1.34}$$

that is satisfied if the ξ_μ follow the wave equation $\Box\xi_\mu = 0$. Solving $\xi_\mu = C_\mu e^{ik^\sigma x_\sigma}$ and using the plane wave solution of $\bar{h}_{\mu\nu}$, the gauge condition (1.26) becomes :

$$\bar{h}_{\mu\nu} = \epsilon_{\mu\nu} e^{ik^\sigma x_\sigma} \rightarrow \Big\{\epsilon_{\mu\nu} + i\Big(k_\nu C_\mu + k_\mu C_\nu - \eta_{\mu\nu} k^\beta C_\beta\Big)\Big\} e^{ik^\sigma x_\sigma}, \tag{1.35}$$

and the trace of the transformed polarization tensor gives :

$$\epsilon^\mu_\mu \rightarrow \epsilon^\mu_\mu - 2ik^\mu C_\mu \tag{1.36}$$

There are infinite number of ways to pick the C_μ without affecting the Lorenz gauge[4]. We can fix the C_μ (hence the ξ_μ) so they cancel the trace of $\bar{h}_{\mu\nu}$. This choice of coordinate system is called the Transverse Traceless (TT) gauge. If the trace is null, one immediately sees that we have :

$$\bar{h}_{\mu\nu} = h_{\mu\nu} \tag{1.37}$$

So plane wave solution of $\bar{h}_{\mu\nu}$ are the same for the initial metric perturbation $h_{\mu\nu}$. Given equation (1.32), the harmonic gauge and the TT gauge, we have :

$$\begin{aligned} k_\mu k^\mu &= 0 \\ k^\mu \epsilon_{\mu\nu} &= 0 \\ \epsilon^\mu_\mu &= 0 \end{aligned} \tag{1.38}$$

The first condition shows that k^μ is light-like so the wave is propagating at the speed of light c, the second condition expresses the transversality of the wave and the last condition gives the null trace of the polarization tensor. Therefore, for a wave propagating along k^3 with wave vector $k^\mu = (k^0, 0, 0, k^3)$, we have :

$$\begin{aligned} (k^0)^2 - (k^3)^2 &= 0 \\ \epsilon_{0\nu} = \epsilon_{3\nu} &= 0 \\ \epsilon_{11} + \epsilon_{22} &= 0. \end{aligned} \tag{1.39}$$

[4]Or De Donder gauge, or harmonic gauge... it has got many names.

Because $h_{\mu\nu}$ is symmetric, we also have for the two last unconstrained components of $\epsilon_{\mu\nu}$:

$$\epsilon_{12} = \epsilon_{21}. \tag{1.40}$$

Finally, we define $\epsilon_{11} = h_+$ and $\epsilon_{12} = h_\times$ to get the polarization tensor of a GW :

$$\epsilon_{\mu\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{1.41}$$

This tensor can be decomposed in two independent components as :

$$\epsilon_{\mu\nu} = \epsilon^+_{\mu\nu} + \epsilon^\times_{\mu\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & 0 & 0 \\ 0 & 0 & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & h_\times & 0 \\ 0 & h_\times & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{1.42}$$

that we call the "plus" and "cross" polarizations.

1.2.3 Quadrupole formula

In the previous section we have presented a solution to the linearized Einstein field equations in the vacuum but we did not explain how GWs are generated. Just like the electromagnetic waves originate from electric charges in motion, we expect GWs to be produced by moving masses. However, there is a fundamental difference with GWs. Take a spherically symmetric gravitational potential $\Phi(r) \propto M/r$ with $M = \sum_i m_i$ the total mass of the source object and write its multipole expansion :

$$\Phi(r - R) \propto \frac{1}{R}\sum_i m_i - \frac{1}{R^3}\sum_i m_i \vec{r}_i \cdot \vec{R} + ..., \tag{1.43}$$

In virtue of the momentum $\vec{p}$ conservation principle, the massive dipole $\sum_i m_i \vec{r}_i$ is a constant. Indeed, we have in the center of mass frame :

$$\sum_i \vec{p}_i = \sum_i \frac{d}{dt}\left(m_i \vec{r}_i\right) = \vec{0}, \tag{1.44}$$

Therefore, it is impossible to generate time-dependent GWs with a massive dipole. The source must have asymmetric (at least quadrupolar) features to be able to radiate GWs. In Einstein field equations, this condition is satisfied by $\nabla^\mu T_{\mu\nu} = 0$.

To solve equation (1.32) we make use of Green's function formalism. The solution for the wave equation has already been studied and we can find in the literature the expression of the Green function for the wave operator $\Box$:

$$G(t,\mathbf{x},t',\mathbf{x'}) = -\frac{\delta(t'-[t-[\mathbf{x}-\mathbf{x'}]])}{4\pi|\mathbf{x}-\mathbf{x'}|}. \tag{1.45}$$

Using this result it comes naturally that the $h_{\mu\nu}$ will be expressed in terms of the Green function as :

$$\bar{h}_{\mu\nu} = \frac{4G}{c^4}\int \mathbf{dx'}\frac{T_{\mu\nu}(t'-[t-[\mathbf{x}-\mathbf{x'}]])}{|\mathbf{x}-\mathbf{x'}|}. \tag{1.46}$$

The distance to the source is usually very large. To that extent, $|\mathbf{x}-\mathbf{x'}|$ becomes the cosmological luminosity distance d_L, taking into account the expansion of the universe for high redshift objects. Moreover, in the TT gauge, we only have 2 independent spatial components of h_{ij}. Using $\partial^\mu T_{\mu\nu} = 0$ and $T_{00} >> T_{ij}$ in the non-relativistic limit, it is possible to show after some calculation that (see Babak 2020) :

$$\partial_0^2\int \mathbf{dx}(T^{00}x_i x_j) = 2\int \mathbf{dx}T_{ij} \tag{1.47}$$

which is only valid for isolated sources and leads to :

$$\bar{h}_{ij} = \frac{2G}{d_L c^4}\partial_0^2\int \mathbf{dx'}T^{00}x'_i x'_j, \tag{1.48}$$

where $T^{00} = \rho c^2$ is the rest energy density. Defining the quadrupole moment tensor as :

$$I_{ij} = \int \mathbf{dx'}\rho x'_i x'_j, \tag{1.49}$$

and using its expression in the TT gauge :

$$\mathcal{I}_{ij} = I_{ij} - \frac{1}{3}\delta_{ij}\text{trace}(I) \tag{1.50}$$

we have the final expression of our quadrupole formula for the generation of GWs :

$$\boxed{h_{ij} = \frac{2G}{d_L c^4}\ddot{\mathcal{I}}_{ij}}. \tag{1.51}$$

The prefactor being extremely small, GWs have very low amplitude. The only way to produce loud and detectable GWs is to have a very massive source. Furthermore, as expected according to the initial comment in this section, the lowest possible order of moment is the quadrupole because of the conservation of energy and momentum. Gravitational dipoles are forbidden by the equations of GR. It means that for GWs to be produced, a system must present asymmetries in the distribution of its mass. We can find such asymmetries in orbiting binary systems.

1.2.4 Circular binary system

Consider a binary system of reduced mass μ and total mass $M = m_a + m_b$ in circular orbit in the x-y plane at orbital frequency ω with separation r. We have for the positions of the two objects $\vec{x}_a$ and $\vec{x}_b$ in the center of mass frame :

$$\begin{aligned} \vec{x}_a(t) &= \frac{r\mu}{m_a}\big(\cos(\omega t), \sin(\omega t), 0\big) \\ \vec{x}_b(t) &= -\frac{m_a}{m_b}\vec{x}_a(t). \end{aligned} \tag{1.52}$$

For two point objects, the mass distribution $\rho(\vec{x})$ is :

$$\rho(\vec{x}) = m_a\delta(\vec{x} - \vec{x}_a) + m_b\delta(\vec{x} - \vec{x}_b). \tag{1.53}$$

We can now calculate the components of quadrupole moment $\mathcal{I}_{ij}$ using equations (1.49) and (1.50) which yield :

$$\mathcal{I}_{ij} = \frac{1}{2}\mu r^2 \begin{bmatrix} \Big(1/3 + \cos(2\omega t)\Big) & \sin(2\omega t) & 0 \\ \sin(2\omega t) & \Big(1/3 - \cos(2\omega t)\Big) & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{1.54}$$

and then take the second time derivative of $\mathcal{I}_{ij}$:

$$\ddot{\mathcal{I}}_{ij} = -2\mu r^2\omega^2 \begin{bmatrix} \cos(2\omega t) & \sin(2\omega t) & 0 \\ \sin(2\omega t) & -\cos(2\omega t) & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{1.55}$$

We use Kepler's third law to constrain the value of r :

$$\begin{aligned} r^3\omega^2 &= GM \\ \Leftrightarrow r &= \Big(\frac{GM}{\omega^2}\Big)^{\frac{1}{3}}, \end{aligned} \tag{1.56}$$

now using (1.51) we have for the two independent components h_+ and $h_\times$:

$$\begin{aligned} h_+ &= -\frac{4\mathcal{M}_c^{5/3}G^{5/3}\omega^{2/3}}{d_L c^4}\cos(2\omega t) \\ h_\times &= -\frac{4\mathcal{M}_c^{5/3}G^{5/3}\omega^{2/3}}{d_L c^4}\sin(2\omega t), \end{aligned} \tag{1.57}$$

where we have introduced the chirp mass $\mathcal{M}_c = \mu^{3/5}M^{2/5}$. We can see that the GW signal oscillates at twice the orbital frequency. To that extent, we have :

$$\omega = \pi f_{gw} \tag{1.58}$$

with f_{gw} the GW signal frequency. Like the luminosity distance d_L, the frequency of the signal can be subject to redshift when the binary system is far away. Therefore, for redhsift z, it transforms as :

$$f_{gw} \rightarrow \frac{f_{gw}}{1+z} \tag{1.59}$$

A gravitational binary system will radiate energy in the form of GWs. As a consequence, the orbital frequency ω will increase with time and inversely, the separation r will decrease until the two objects eventually merge. We can estimate this frequency and separation evolution if we know the amount of energy radiated by the system per unit time dE/dt given by (Carroll 1997; Schutz 2022) :

$$\frac{dE}{dt} = \frac{G}{5c^5}\langle \dddot{\mathcal{I}}_{ij} \dddot{\mathcal{I}}^{ij} \rangle, \tag{1.60}$$

where $\dddot{\mathcal{I}}_{ij}$ is the third time derivative of the quadrupole and $\langle . \rangle$ denote the averaging over some periods of the wave. After some straightforward calculation, we have :

$$\frac{dE}{dt} = -\frac{32}{5c^5} G^{7/3} M^{4/3} \mu^2 \omega^{10/3} = -\frac{32 G^4 \mu^2 M^3}{5c^5 r^5}, \tag{1.61}$$

writing ω in terms of r using Kepler's third law. Making use of the Virial theorem for a gravitational system, we also have :

$$E = -\frac{1}{2}\frac{GM\mu}{r} \Leftrightarrow \frac{dE}{dt} = \frac{1}{2}\frac{GM\mu}{r^2}\frac{dr}{dt}. \tag{1.62}$$

Equating the two last expressions, we get an equation for r that we can solve. Once solved, it gives :

$$r(t) = \left(r_0^4 - \frac{256}{5}\frac{G^3 M^2 \mu}{c^5} t\right)^{1/4} = r_0\left(1 - \frac{t}{\tau_c}\right)^{1/4}, \tag{1.63}$$

where we have defined r_0 the initial separation and τ_c the coalescence time. At $t = \tau_c$, separation $r(\tau_c) = 0$, thus the two orbiting objects come into contact and merge[5]. Again, using Kepler's third law and the chirp mass $\mathcal{M}_c$, we can write the evolution of the orbital frequency $\omega(t)$:

$$\omega(t) = \omega_0\left(1 - \frac{256}{5}\frac{G^{5/3}\mathcal{M}_c^{5/3}}{c^5}\omega_0^{8/3} t\right)^{-3/8} = \omega_0\left(1 - \frac{t}{\tau_c}\right)^{-3/8}, \tag{1.64}$$

where ω_0 is the initial orbital frequency, showing that indeed, $\omega(t)$ goes up with time, hence, so does the GW frequency $f_{gw}(t)$. It means that the phase $\Phi(t)$ of the GW will evolve with time like :

[5]This is true considering the point-mass approximation. In reality, for two objects of radius L, we can trivially show using $r(t_c) = L$ that merger happens at $t_c = \tau_c\left(1 - (L/r_0)^4\right)$ which is less than τ_c. We deduce from this that bigger objects merge sooner.

$$\Phi(t) = \Phi_0 + \int_{t_0}^{t} dt\omega(t) = \Phi_0 + \frac{c^5}{32G^{5/3}\mathcal{M}^{5/3}}\left(\omega_0^{-5/3} - \omega(t)^{-5/3}\right) \tag{1.65}$$

In equation (1.57) we give the h_+ and $h_\times$ components of $h_{\mu\nu}$ for a binary system orbiting in the x-y plane with angular momentum $\vec{L}$, emitting GW in the z direction. For a more general description, we need to rotate around the x-axis the transverse quadrupole moment by an inclination angle ι giving :

$$\begin{aligned} h_+ &= \frac{4\mathcal{M}_c^{5/3}G^{5/3}\omega^{2/3}}{d_L c^4}(1+\cos^2\iota)\cos(2\Phi(t)) \\ h_\times &= \frac{4\mathcal{M}_c^{5/3}G^{5/3}\omega^{2/3}}{d_L c^4}\cos\iota\sin(2\Phi(t)). \end{aligned} \tag{1.66}$$

Finally, we can perform a rotation in the polarization plane by an angle 2ψ, introducing the polarization basis following $\hat{e}_1 \cdot \hat{e}_2 = 0$ and $\hat{e}_1 \cdot \hat{k} = \hat{e}_2 \cdot \hat{k} = 0$. We have for e_{ij}^A the + and × GW polarization tensor :

$$\begin{aligned} \epsilon_{ij}^+ &= \hat{e}_1 \otimes \hat{e}_1 - \hat{e}_2 \otimes \hat{e}_2 \\ \epsilon_{ij}^\times &= \hat{e}_1 \otimes \hat{e}_2 + \hat{e}_1 \otimes \hat{e}_2 \end{aligned}, \tag{1.67}$$

$$\begin{bmatrix} h_+ \\ h_\times \end{bmatrix} \to \begin{bmatrix} \cos 2\psi & -\sin 2\psi \\ \sin 2\psi & \cos 2\psi \end{bmatrix} \begin{bmatrix} h_+ \\ h_\times \end{bmatrix}, \tag{1.68}$$

and we have $h_{\mu\nu}$ in its final form :

$$\begin{aligned} h_{\mu\nu} &= \sum_{A=+,\times} \epsilon_{\mu\nu}^A h_A = \epsilon_{\mu\nu}^+ h_+ + \epsilon_{\mu\nu}^\times h_\times \\ h_+ &= \frac{2\mathcal{M}_c^{5/3}G^{5/3}\omega^{2/3}}{d_L c^4}\Big((1+\cos^2\iota)\cos 2\psi\cos(2\Phi(t)) - 2\cos\iota\sin 2\psi\sin(2\Phi(t))\Big) \\ h_\times &= \frac{2\mathcal{M}_c^{5/3}G^{5/3}\omega^{2/3}}{d_L c^4}\Big((1+\cos^2\iota)\sin 2\psi\cos(2\Phi(t)) + 2\cos\iota\cos 2\psi\sin(2\Phi(t))\Big) \\ \Phi(t) &= \Phi_0 + \frac{c^5}{32G^{5/3}\mathcal{M}^{5/3}}\left(\omega_0^{-5/3} - \omega(t)^{-5/3}\right) \\ \omega(t) &= \pi f_{gw}(t) = \omega_0\Big(1 - \frac{256}{5}\frac{G^{5/3}\mathcal{M}_c^{5/3}}{c^5}\omega_0^{8/3}t\Big)^{-3/8} \end{aligned} \tag{1.69}$$

In summary, we have 5 parameters that characterize our GW :

$\mathcal{M}_c$	Chirp mass (kg)
d_L	Luminosity distance (m)
ι	Inclination angle (rad)
ψ	Polarization angle (rad)
Φ_0	Initial phase (Rad)
f_{gw}	GW signal frequency (Hz)

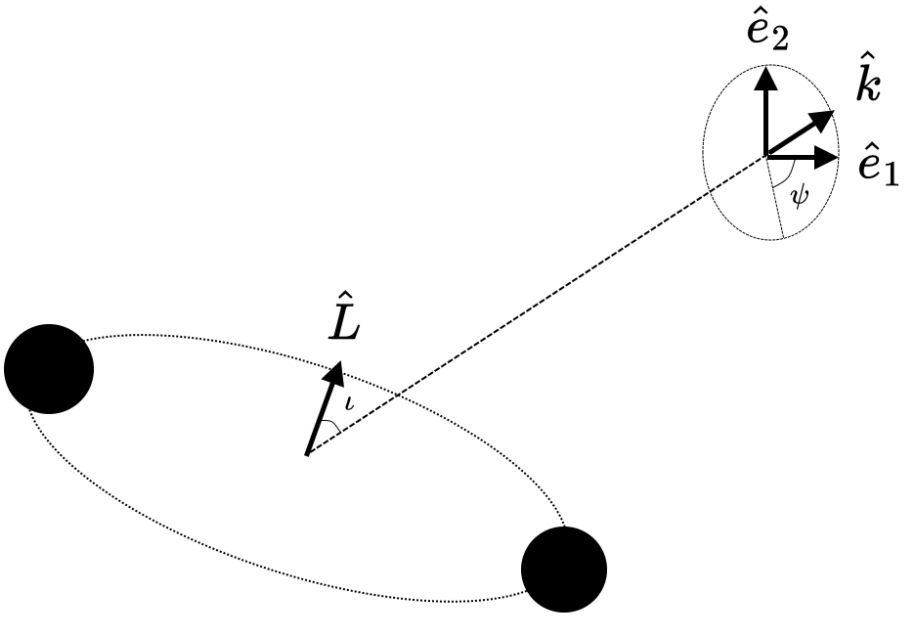

Figure 1.6: Illustration of the geometry of the SMBHB. Here, $\hat{L}$ represents the rotation axis of the binary, $\hat{k}$ the direction of propagation of the GW with polarization basis $\{\hat{e}_1, \hat{e}_2\}$, the polarization angle ψ and finally ι the inclination (i.e. the angle between $\hat{L}$ and $\hat{k}$).

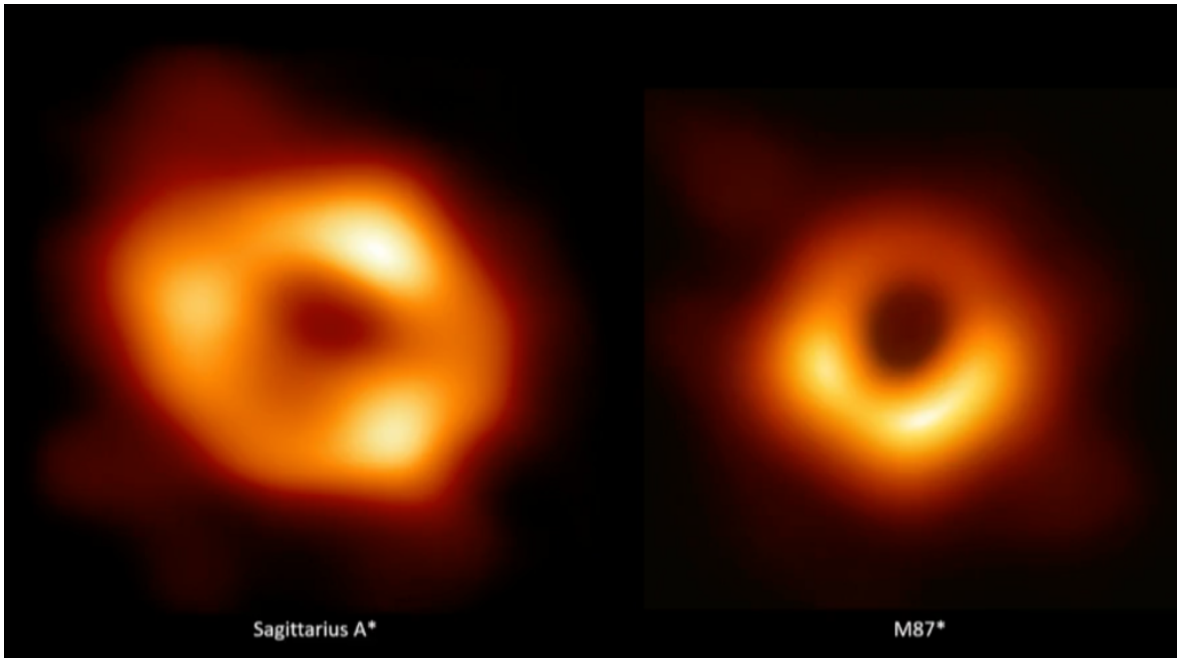

Figure 1.7: First images of SMBH captured by the Event Horizon Telescope.

1.3 Supermassive black holes

In 1916, Karl Schwarzschild came up with a metric tensor $g_{\mu\nu}$ solution to the Einstein field equations in the vaccum (see Schwarzschild 1999, a translation to English of the original 1916 paper), describing the geometry of spacetime in spherical coordinates when a non-rotating point mass M is present in the center of the frame. This solution is consistent with classical gravity in the Newtonian limit but it predicts peculiar behaviour when approaching the center. Using that metric the length element ds^2 writes as :

$$ds^2 = \big(1 - \frac{2GM}{rc^2}\big)c^2dt^2 - \big(1 - \frac{2GM}{rc^2}\big)^{-1}dr^2 - r^2d\Omega^2, \tag{1.70}$$

where r is the distance to the center for an external (far away) observer and $d\Omega = \sin^2\theta d\theta d\phi$ is the element of solid angle. One immediately sees that for $r = 2GM/c^2$, the coordinate system diverges. This radius of singularity [6] is called the Schwarzschild radius and it defines the boundaries of what we call the event horizon. Beyond this limit, nothing can escape and this region of spacetime appears completely black to external observers : we have a Black Hole (BH).

These astrophysical objects are completely characterized by their mass that can range from a few solar masses $M_\odot$ to billions of solar masses. If it is greater than $10^6 M_\odot$, we call it a supermassive black hole (SMBH). They are the largest category of black holes and

[6] Note that this is not a true singularity but rather a coordinate singularity. It is possible to make it disappear by chosing an apprioriate coordinate system, like for instance, Lemaître coordinates.

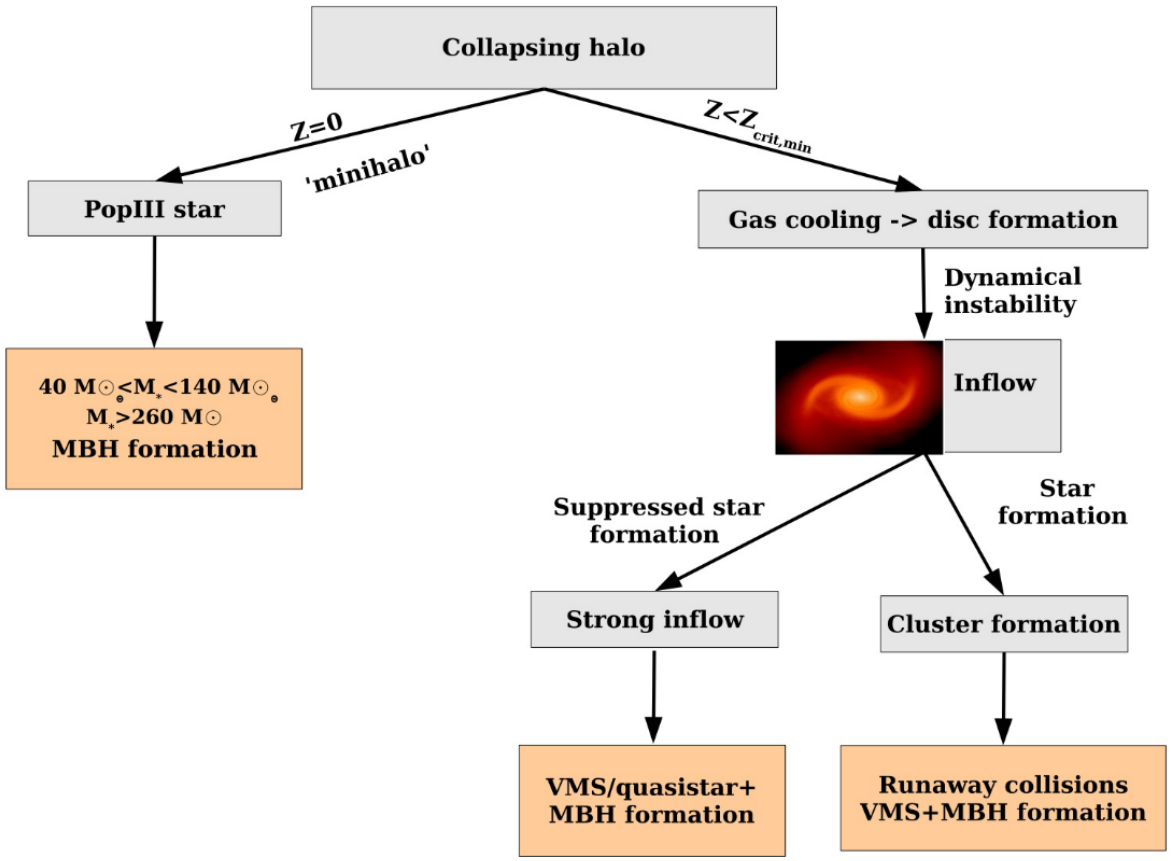

Figure 1.8: Illustration of the possible scenarios for MBH formation. Z is the metallicity of the halo, i.e. the presence of elements heavier than helium (figure taken from Volonteri 2010).

we expect to find them at the center of galaxies. How they form is still an open question in astrophysics. The observation of Active Galactic Nuclei (AGN) at very high redshift suggests that they were already present in the early Universe. The proposed scenarios for their formation rely on the production of black hole "seeds" that already have masses around $10^{3-6} M_{\odot}$. Such seeds could be produced by the collapse of first generation stars (Pop III) in large protogalactic gas clouds (Volonteri 2012). On figure 1.8, we see that the clouds can experience direct collapse or fragment into individual clumps giving birth to a cluster of smaller stars. In both cases, these early stars are often massive enough to end up exploding as supernovae or collide in the cluster to produce BHs. Later, the seeds can grow through accretion of the surrounding gas and BH merger, leading to black holes with huge masses, (Super)-Massive Black Holes (MBH). Inspiralling Supermassive Black Hole Binary (SMBHB) systems could produce loud and stable low frequency GWs that would still be detectable at large distances in spacetime.

The most accepted scenario to explain the birth of such SMBH pairs is by means of

galaxy merger. When their nuclei get closer, the stars and gas making up the galaxies steal kinetic energy from the SMBHs through gravitational interaction. The stars get ejected by slingshot effect and the gas generate dynamical friction. This loss of energy will drive the SMBHs closer and closer until they form a stable binary, separated by a distance that is only in the order of some parsecs. At that stage, there might not be enough matter present between the two SMBHs to significantly extract energy from the system. The only way it can give away energy is through GW emission that are extremely weak. It is estimated that such an inspiralling phase should be as long as the age of the Universe before entering the merging phase where the GW emissions become efficient. The observations seem to contradict this prediction because we see very massive AGNs at high redshift whose formation we can explain only with SMBHB merger. This is the "final parsec" problem. The mechanisms allowing the SMBHs to merge within a reasonable amount of time are still unclear but some solutions have been proposed. In Milosavljevic 2003 they describe two of the suggested explanations : (i) they show that taking into account the shape of the galaxies and the inhomogeneous distribution of matter within them could make the dynamical friction more efficient, allowing the SMBHs to enter merging phase, (ii) they explain that a N-SMBH (galaxy) merger could solve to the issue, where the inspiralling phase is described as a N-body problem which could allow the SMBHs to get closer under certain conditions.

In practice, it is possible to detect such GWs thanks to the huge mass of the SMBHs (i.e. huge chirp mass $\mathcal{M}_c$ of the binary system) that make up for loud GWs. A large population of SMBHB would produce a superposition of many GWs filling-up the entire Universe[7]. The resulting signal can de divided into two main categories : the stochastic Gravitational Wave Background (GWB) and Continuous Waves (CW).

1.3.1 Gravitational wave background

The vast population of observable galaxies indicates that we probably have a large population of SMBH in the Universe. The superposition of many individual SMBHB GW signals would result in a low frequency noise spread on a broad band. For a population of circular SMBHB, given the formula for GW emissions, we can show that such a noise signal has a frequency spectrum following a powerlaw. We define the characteristic strain of the GWB as (Taylor 2021):

$$h_c^2(f) \equiv f S_h(f) \propto f^{-2} \frac{d\rho}{d\ln f}, \tag{1.71}$$

with $S_h(f)$ the power spectral density for the considered population of sources (here circular SMBHBs) and $d\rho/d\ln f$ the energy density of sources in a logarithmic interval of

[7] We can only see the local Universe.

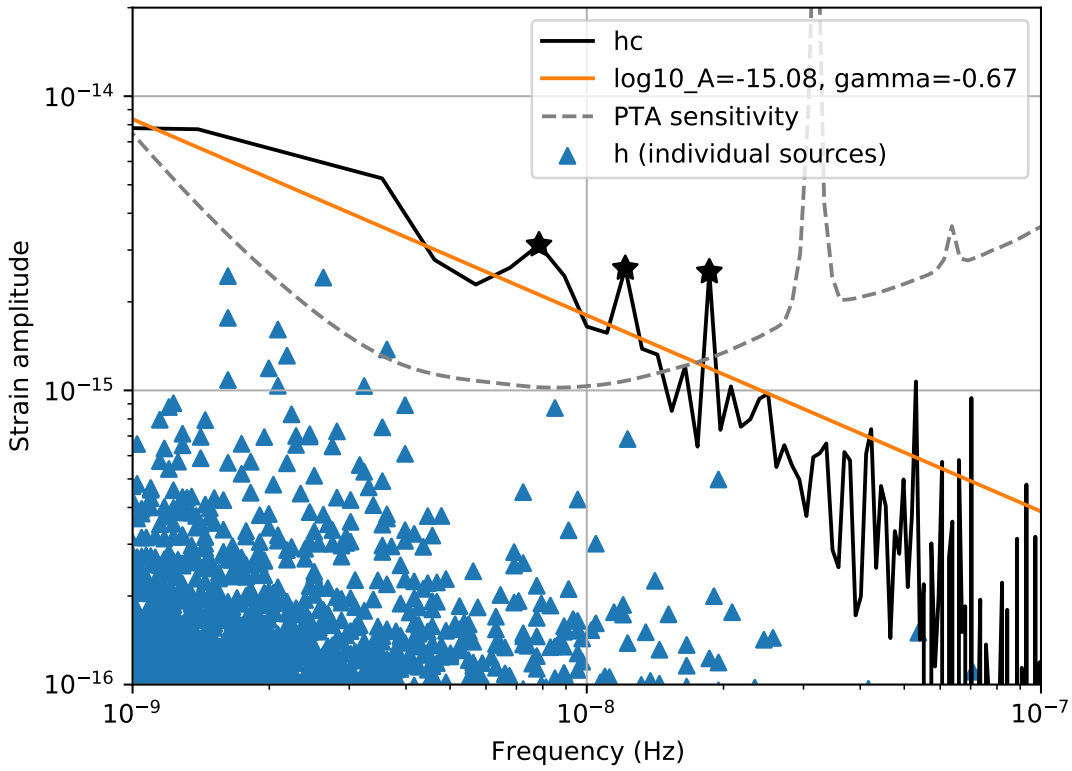

Figure 1.9: A simulation of SMBHB population in the universe where y-axis is GW amplitude and x-axis is GW frequency. Each blue triangle represents a single SMBHB source. The total sum of these signals gives the characteristic strain of the GWB (black line) which can be approximated by a powerlaw (orange line) of index -2/3. The black stars are individual binaries resolved on top of the background (CGW). The dotted line displays the sensitivity of IPTA, showing that both GWB and CGW should be detectable at low frequencies. (Acknowledgement : we have used simulated population of binaries from A.Sesana to produce this plot.)

frequency.

Considering that we have N_0 emitting binaries and ignoring the effects due to redshift[8], we can approximate the density $d\rho/d\ln f$ for circular binaries using equations (1.61) and (1.76) :

$$\frac{d\rho}{d\ln f} \sim N_0 \frac{dE}{d\ln f} \propto f \frac{dE}{dt}\frac{dt}{df} \propto f \times f^{10/3} \times f^{-11/3} \propto f^{2/3}, \tag{1.72}$$

yielding a frequency dependence following a powerlaw with spectral index 2/3.

Given equation 1.71, we have a h_c that follows a powerlaw as :

$$h_c(f) \propto f^{-2/3}. \tag{1.73}$$

On figure 1.9, we see a simulated population of individual SMBHBs with their corresponding GW strain amplitude h and frequency f. The higher frequency range appears less crowded because the frequency evolution rate of the circular SMBHB increases as $f^{11/3}$ (equation 1.76) so the binaries spend less time there. As a consequence, the characteristic strain h_c decreases with frequency, like the GW energy density, and the signal looses its stochasticity.

When we search for such stochastic signal, we model the characteristic strain as a powerlaw with amplitude A_{GWB} and spectral index γ_{GWB} within our prior assumptions on their values :

$$h_c = A_{GWB}\left(\frac{f}{1\text{yr}^{-1}}\right)^{\gamma_{GWB}}. \tag{1.74}$$

The detection of the GWB and the measurement of (A_{GWB}, γ_{GWB}) would give crucial information about the population and distribution of SMBHBs across the Universe. It is the main target of PTA collaborations.

1.3.2 Continuous waves

The amplitude of a GW decreases with distance as $1/d_L$. Some of the SMBHB responsible for the stochastic GWB may be closer to us and appear brighter in their GW emission, resulting in GW signal resolved at single frequencies on top of the GWB (see figure 1.9). These GWs are called Continuous Gravitational Waves (CGW) and are described as deterministic signals. According to equation (1.64) showing the evolution of GW frequency with time, CGWs should see their frequency increase. However, for supermassive systems, this evolution can be neglected over some orbital periods (for frequencies of $\leq 10^{-7}$

[8]This is a big approximation, for the sake of quickly demonstrating the expected frequency dependence of the characteristic strain. For a realistic derivation we would have to take into account redshift, mass distributions, mass ratios and even eccentricity of the binaries (see Taylor 2021).

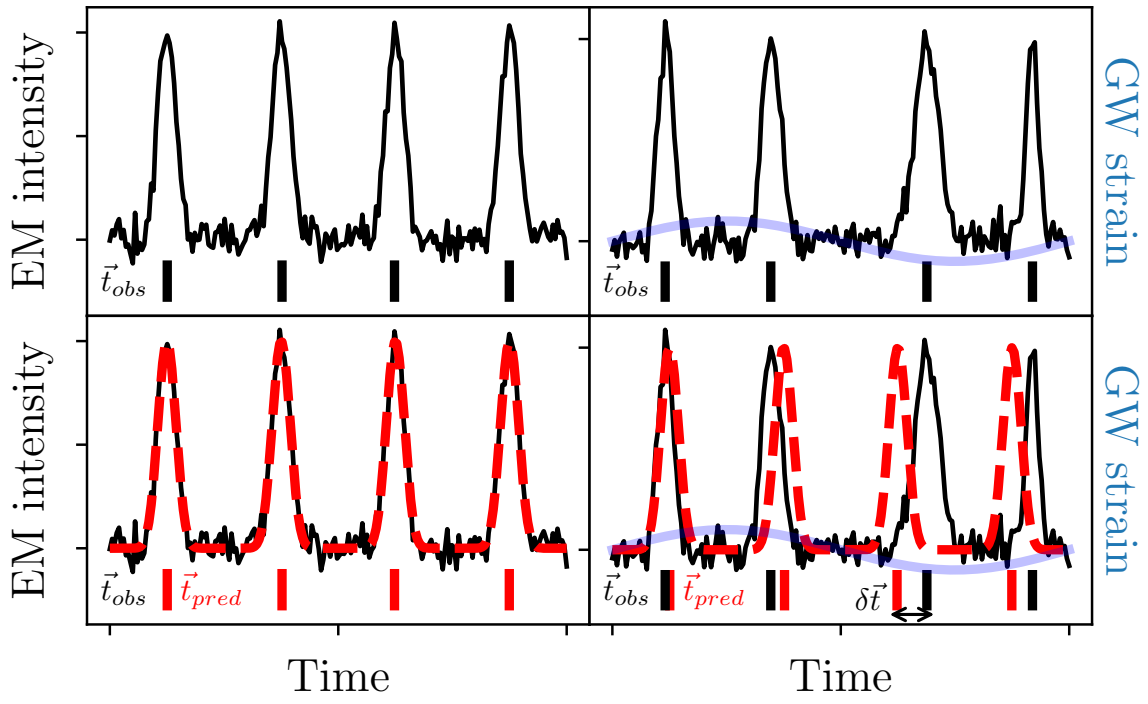

Figure 1.10: On the left panels, a simple illustration (fake data) of what is seen by the radiotelescope. We evaluate the true TOAs $\vec{t}_{obs}$ (black) and produce a predicted value $\vec{t}_{pred}$ (red) by fitting a timing model. On the right panels, the same thing but in the presence of GW signal. The timing model does not account for the delay induced by the GW and timing residuals $\delta\vec{t}$ are measured.

Hz only, see section 1.4.1 for more details). This manuscript focuses on this type of signal and we will present in the next section how it is possible to detect them using pulsars.

1.4 Pulsar Timing Array

A Pulsar Timing Array (PTA) is an array of Milliecond Pulsars (MSPs) observed from the Earth using radiotelescopes to monitor the time of arrivals (TOAs) $\vec{t}$ of their pulses. Among these pulsars, some have been observed for long periods of time T>20 years giving us robust datasets to work with. This timing data allows to use pulsars as "cosmic clocks". We can predict the TOAs by characterizing the properties of the pulsar and modeling perturbations of physical origin. The difference between the predicted TOAs $\vec{t}_{pred}$ and the measured TOAs $\vec{t}_{obs}$ is called the timing residuals $\delta\vec{t}$. If our PTA is sensitive enough, the effect of GWs should be present and detectable in the timing residuals $\delta\vec{t}$.

MSPs usually have a characteristic and well defined pulse profile. However, the observation of a single pulse can be challenging because of its low amplitude and the ambient noise levels. For that reason, we measure series of consecutive pulses that we stack to-

gether to obtain an averaged pulse profile with a significantly better signal to noise ratio. We then cross-correlate a previously determined template of the pulse profile[9] to extract a value of the TOA. One new TOA is obtained with a cadence of about 1 week. We can estimate the uncertainty of a TOA measurement as (Lorimer and Kramer 2012) :

$$\sigma_{TOA} \simeq \frac{W}{S/N}, \tag{1.75}$$

with W the width of the pulse and S/N the signal to noise ratio of the measured (averaged) pulse profile.

Typically, we have $\sigma_{TOA} \simeq 10^{-6}$s so we are accurate to the micro-second, which is remarkable for astrophysical observations. This timing precision led physicists to believe that the effect of GWs could be noticeable in the TOAs. This idea was first proposed in Sazhin 1978 and Detweiler 1979 where it is argued that only GW sources with masses around $10^{8-10} M_\odot$ could be detectable within the sensitivity σ_{TOA} of our MSP timing measurements. The natural candidates for such GW sources are SMBHBs.

1.4.1 Sensitivity of the array

The long period of observation T of the PTA makes it sensitive to GWs down to the nanohertz band ($1/T \sim 10^{-9}$ Hz). The cadence of observation is about 1 week[10] so signals with frequencies above $1/7\text{days} \sim 10^{-6}$ Hz are under sampled. Moreover, many observation gaps are present in the dataset, sometimes of several months due to maintenance of the radio-telescopes. As a consequence, the TOAs are not evenly sampled and it is not possible to directly visualize the Fourier spectrum of the pulsars. Instead, we use Gaussian processes to model the time-correlated features intrinsic to pulsars and interpolate between gaps (see section 2.2.2).

In that window of frequency, the GWB should give rise to measurable effects thanks to the high precision timing measurements of MSPs (see figure 1.9). For CGWs, we can argue whether the frequency evolution of the SMBHBs is going to introduce a bias in the measurement of the signal (Sesana and Vecchio 2010). We derive from equation (1.64) :

$$\frac{d\omega}{dt} = \frac{3}{8}\frac{1}{\tau_c}\omega^{11/3} = \frac{96}{5}\frac{G^{5/3}\mathcal{M}_c^{5/3}}{c^5}\omega^{11/3}, \tag{1.76}$$

the expression giving the rate of change in GW frequency (recall that $f_{gw} = \pi\omega$) for a binary system with orbital period ω. In the context of PTA, we are sensitive to frequencies around $f_{gw} \sim 10^{-8}$ Hz and binaries with a chirp mass $\mathcal{M}_c \sim 10^9 M_\odot$. For such systems, the coalescence time τ_c is in the order of 10^{4-5} years so the change in frequency that occurs

[9] For curiosity, many pulse profile templates can be found at : https://www.jb.man.ac.uk/research/pulsar/Resources/epn/

[10] Note true for legacy data because of gaps.

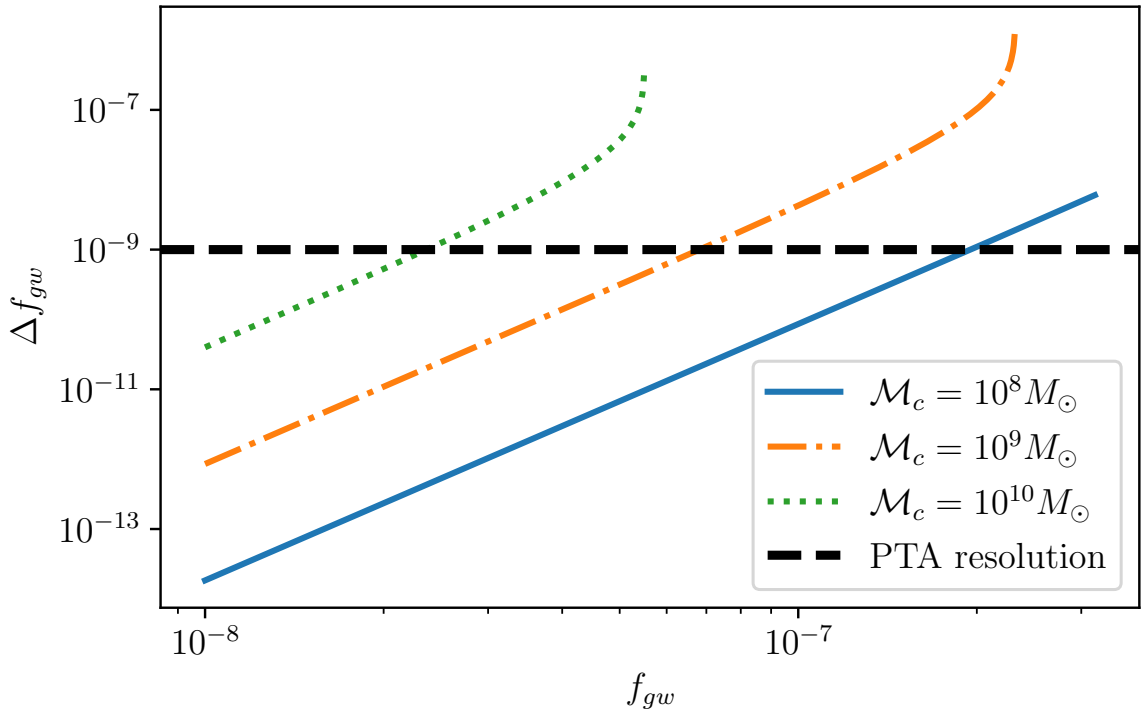

Figure 1.11: The evolution of frequency Δf_{gw} for 10 years of observation and three chirp masses compared to PTA resolution.

within the period of observation T (in the order of 10 years) is expected to be small. Using the previous equation, we can evaluate this change of frequency Δf_{gw} to first order :

$$\Delta f_{gw} \simeq \frac{df_{gw}}{dt} \times T \simeq \frac{96}{5} \frac{G^{5/3} \mathcal{M}_c^{5/3}}{c^5} f_{gw}^{11/3} \times T. \tag{1.77}$$

For the considered values of the parameters $\mathcal{M}_c$, f_{gw} and T we have a Δf_{gw} that is less than 10^{-9} Hz, hence below PTA frequency resolution. To that extent, the CGW sources can be considered stable and monochromatic (i.e. non evolving).

This approximation is no longer true for signals with higher frequencies. The very strong dependence of Δf_{gw} on f_{gw} can produce differences that are resolvable by the array. On figure 1.11, we plot Δf_{gw} (not approximated as in equation 1.77) as a function of the initial signal frequency f_{gw}, for 10 years of observation and three different values of chirp mass $\mathcal{M}_c$, to compare it with a typical PTA resolution of 1 nHz. It is clear that discrepancies between true and measured signal frequencies will appear around 100 nHz (or less for large $\mathcal{M}_c$) where Δf_{gw} exceeds the PTA resolution. We even start to see very strong frequency evolution for the largest values of chirp mass. It means that the non-evolving source approximation no longer stands. To search for signal at frequencies higher than ~100 nHz, we must take into account the evolution of the SMBHB. If we do not, 100 nHz should be a reasonable limit up to which the Δf_{gw} can be negligible for $\mathcal{M}_c \leq 10^9 M_\odot$.

1.4.2 Millisecond pulsars

Pulsars appear as the last state of evolution of stars with masses around $9M_{\odot}$. When nuclear fusion cycles are over within its core, the produced iron core collapses under the gravitational pull producing a gigantic explosion known as a supernova of type II. During this event, the whole star shrinks down into a small volume, leaving behind an extremely compact and highly magnetized neutron star with a diameter of about 10km. The angular momentum of the star before the supernova is conserved and transferred to the tiny residual neutron star, giving it a high rotational speed. The combination of both high magnetization and high spin velocity produces beams of electromagnetic emissions that we perceive as pulses, hence the name, Pulsar (Pulsating staR). They were originally detected in the radio frequency band. Some of them are extremely stable in their rotation, which is why they are interesting objects to study, opening a whole spectrum of very precise astrophysical measurements.

Due to their emissions, pulsars loose energy with time. As a consequence, they loose angular momentum and their rotational speed decreases. This "spin-down" can cause the pulsar to fade away and not be detectable by radiotelescopes. If the pulsar is in a binary system, it can get re-accelerated by stealing angular momentum from its companion through gas accretion. This process is called the "recycling" of pulsars (Alpar et al. 1982). Once equilibrium is reached, such pulsars can have periods of rotation in the order of the millisecond (i.e. millisecond pulsars). Moreover, due to their history and age, they are much more stable than young pulsars that just went through a supernova (more prone to glitches/starquakes).

MSPs in our galaxy are monitored by several collaborations throughout the world using their radio observatories to collect timing data. Lately, efforts were made to put in common all these datasets which gave birth to the international collaboration.

1.4.3 The International Pulsar Timing Array

The IPTA DR2 consists of 65 stable MSP with the duration of observations up to 30 years (Perera 2019; Antoniadis 2022). It combines the pulsar timing data acquired by three PTA collaborations, namely :

- the **European Pulsar Timing Array** EPTA first data release (Desvignes 2016) combining data from the Nancay radio-telescope (France), the Effelsberg radio-telescope (Deutschland), the Lovell Telescope (UK), the Westerbork Synthesis radio-telescope (Nederland) and the Sardinia radio-telescope (Italia),

- the **North American Nanohertz Observatory for Gravitational Waves** NANOGRAV 9 year data release (Arzoumanian 2016) with data from the Green Bank Telescope (USA) and the Arecibo Radio telescope (Puerto Rico) who recently passed away,

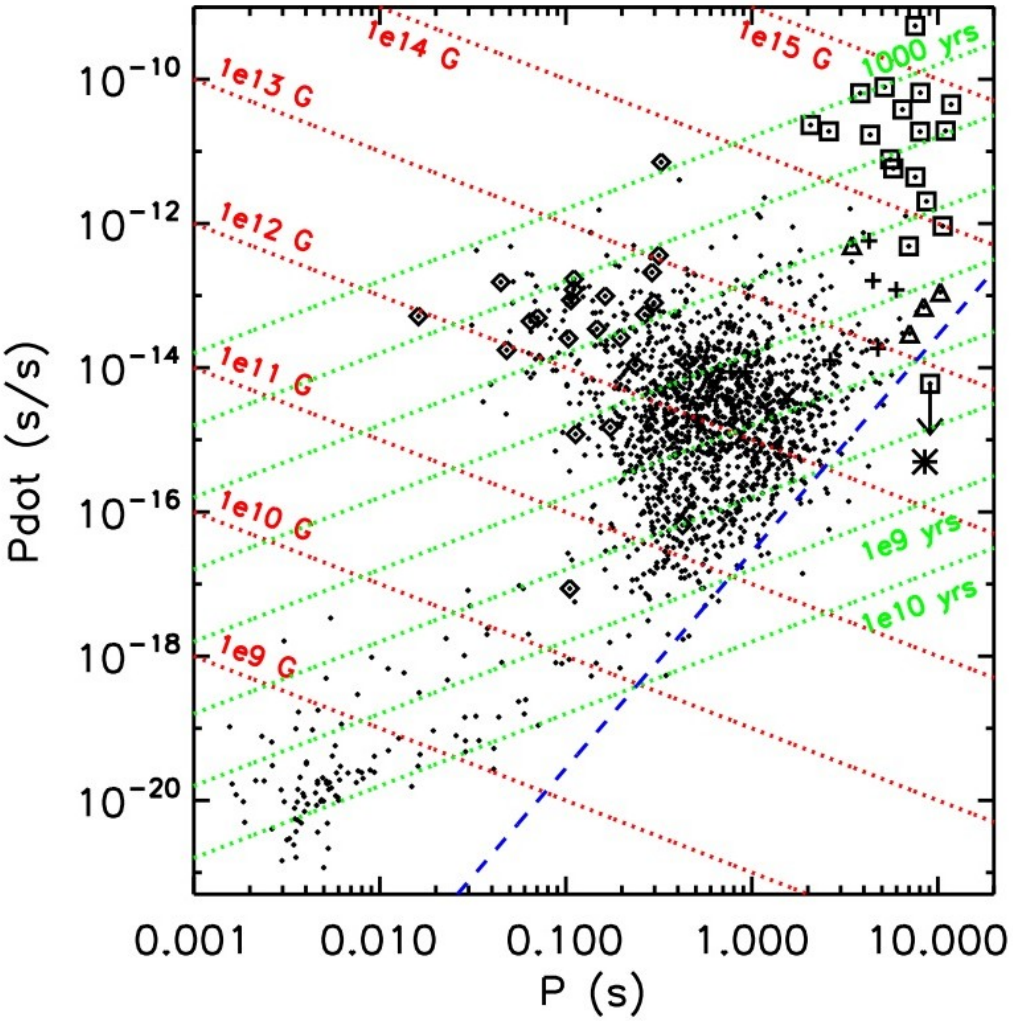

Figure 1.12: The population of known pulsars given their spin period (x-axis) and spin period first time derivative (y-axis) (plot taken from Tiengo et al. 2011). MSPs lie on the bottom left corner after getting recycled. The dashed blue line is the limit of the pulsar graveyard below which no pulsar (or very few) is expected to radiate because of electron-positron pair production (Chen and Ruderman 1993).

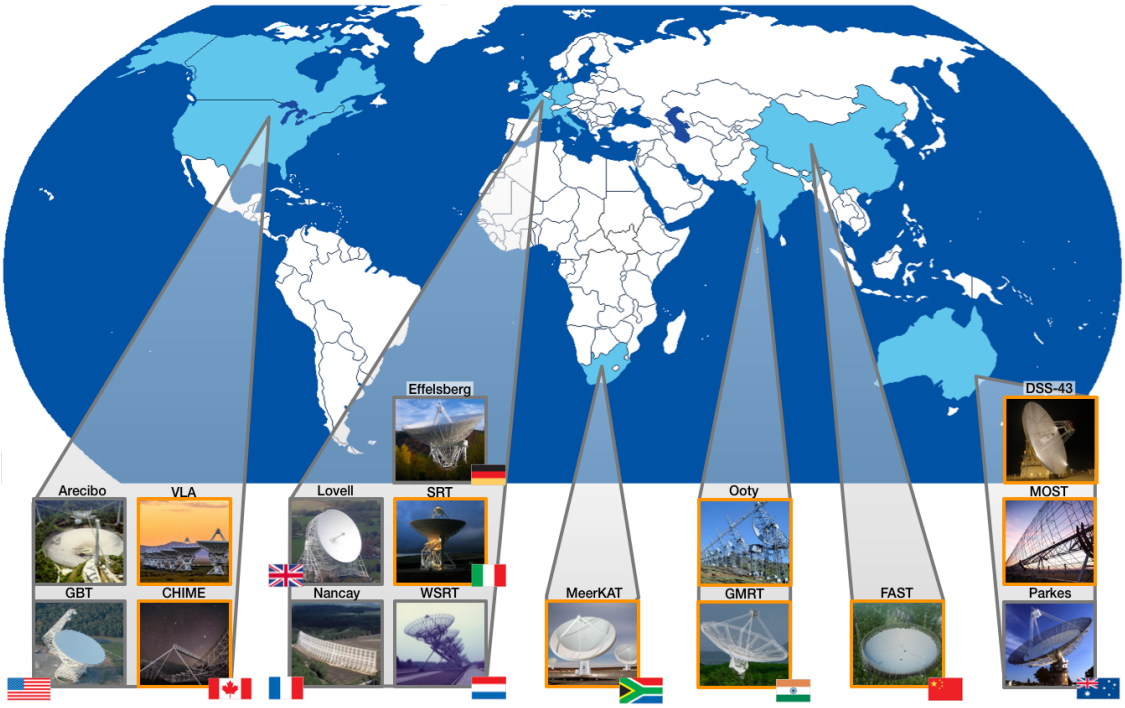

Figure 1.13: The list of IPTA telescopes and their locations in the various countries of the collaboration (picture taken from the NANOGRAV website).

- The **Parkes Pulsar Timing Array** PPTA first data release (Manchester 2013a) using the Parkes Radio telescope (Australia).

The combination of all data is superior to the datasets of each collaboration. We have already observed the improvement in the detection of the common red noise process in Antoniadis 2022 by using IPTA DR2. This upgrade in sensitivity can be explained by :

- a better sky coverage providing better localization of GW signals,
- a better decoupling and identification of noise components due to the increased number of backends giving better constraints on noise parameters,
- the reduction of the number of gaps in the data due to absence of observations.

Today, the IPTA also counts the Indian PTA (InPTA) collaboration (Tarafdar et al. 2022) and the South-African MeerKAT radio-telescope (Bailes et al. 2020) within its collaborators. For now, the only available combination is the DR2 of which we make extensive use in this manuscript. The IPTA collaboration is already working on the third data release that will combine latest datasets to give the best possible sensitivity to date.

1.4.4 PTA response to GW signal

We have presented in section 1.2 a solution to the linearized Einstein field equations in the TT gauge that gave GWs. The first order perturbation to the flat-space metric for a GW propagating along the $\hat{z}$ axis was :

$$h_{\mu\nu}^{TT} = \sum_{A=+,\times} \epsilon_{\mu\nu}^{A} h_A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{1.78}$$

where we see the two independent polarization states h_+ and $h_\times$.

For a pulsar in the sky, let $\hat{n}$ be the unit vector pointing from the Earth to the pulsar and σ^μ the four-momentum of a light-like signal of frequency ν travelling from the pulsar to the Earth. We can write the geodesic equation for σ^μ as :

$$\frac{d\sigma^\mu}{d\lambda} + \Gamma^\mu_{\alpha\beta}\sigma^\alpha\sigma^\beta = 0, \tag{1.79}$$

with

$$\sigma^\mu = (\nu, -\nu\hat{n}) = \frac{dX^\mu}{d\lambda} = \frac{d}{d\lambda}(t, \vec{x}). \tag{1.80}$$

In theory, the GW should influence the position of the pulsar $\hat{n}$ but this effect is considered small and neglected here. Expressing the Christoffel symbols for $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ and keeping only the first order terms, we get :

$$\begin{aligned} \Gamma^\mu_{\alpha\beta} &= \frac{1}{2} g^{\mu\delta}\Big(\partial_\alpha g_{\beta\delta} + \partial_\beta g_{\delta\alpha} - \partial_\delta g_{\alpha\beta}\Big) \\ &\simeq \frac{1}{2}\eta^{\mu\delta}\Big(\partial_\alpha h_{\beta\delta} + \partial_\beta h_{\delta\alpha} - \partial_\delta h_{\alpha\beta}\Big) + \mathcal{O}(h^2) \end{aligned} \tag{1.81}$$

We plug this expression of $\Gamma^\mu_{\alpha\beta}$ into the geodesic equation and focus on the time component of σ^μ :

$$\begin{aligned} &\frac{d\sigma^0}{d\lambda} + \Gamma^0_{\alpha\beta}\sigma^\alpha\sigma^\beta = 0 \\ \Leftrightarrow &\frac{d\nu}{d\lambda} + \frac{1}{2}\eta^{00}\big(-\partial_0 h_{\alpha\beta}\big)\sigma^\alpha\sigma^\beta = 0 \\ \Leftrightarrow &\frac{d\nu}{d\lambda} + \frac{1}{2}\frac{\partial h_{ij}}{\partial t}\nu^2\hat{n}^i\hat{n}^j = 0. \end{aligned} \tag{1.82}$$

We only have spatial components in $h_{\mu\nu}$ so we denote the metric perturbation by h_{ij}, omitting the $h_{0\nu}$ and $h_{\mu 0}$ terms that are null. To account for the fact that GWs propagate at the speed of light, we have to define the delayed wave solution :

$$h_{ij} = h_{ij}(t - \hat{k}\cdot\vec{x}), \tag{1.83}$$

with $\hat{k}$ the direction of propagation of the GW. This gives for the total derivative of h_{ij} with respect to λ:

$$\begin{aligned}\frac{dh_{ij}}{d\lambda} &= \frac{\partial h_{ij}}{\partial t}\frac{dt}{d\lambda} + \frac{\partial h_{ij}}{\partial(\hat{k}\cdot\vec{x})}\frac{d(\hat{k}\cdot\vec{x})}{d\lambda} \\ &= \frac{\partial h_{ij}}{\partial t}\nu - \frac{\partial h_{ij}}{\partial t}\hat{k}\cdot\frac{d\vec{x}}{d\lambda} \\ &= \frac{\partial h_{ij}}{\partial t}\nu - \frac{\partial h_{ij}}{\partial t}\hat{k}\cdot(-\nu\hat{n}) \\ &= \frac{\partial h_{ij}}{\partial t}\nu\big(1 + \hat{k}\cdot\hat{n}\big),\end{aligned} \tag{1.84}$$

and for the time derivative :

$$\frac{\partial h_{ij}}{\partial t} = \frac{dh_{ij}}{d\lambda}\Big(\frac{1}{1+\hat{k}\cdot\hat{n}}\Big)\frac{1}{\nu}. \tag{1.85}$$

Now, combining (1.82) and (1.85), we get :

$$\frac{1}{\nu}\frac{d\nu}{d\lambda} = -\frac{1}{2}\Big(\frac{\hat{n}^i\hat{n}^j}{1+\hat{k}\cdot\hat{n}}\Big)\frac{dh_{ij}}{d\lambda}. \tag{1.86}$$

This equation shows how the frequency of a signal is affected by the presence of a GW. In fact, if we consider a pulsar that is emitting pulses of light at rate ν_0, we can integrate this last expression between the times of emission of the pulse at pulsar t_p and reception t on Earth. This yields the redshift (or blueshift) induced in the TOAs of a pulsar :

$$\ln\frac{\nu(t)}{\nu_0} = -\frac{1}{2}\sum_{A=+,\times}\Big(\frac{\hat{n}^i\hat{n}^j}{1+\hat{k}\cdot\hat{n}}\Big)\epsilon^A_{ij}(h_A(t) - h_A(t_p)) \tag{1.87}$$

with $\nu(t)$ the frequency at reception that we can define as $\nu(t) = \nu_0 + \delta\nu$ for small variation of ν giving $\ln\big(\nu_0 + \delta\nu\big)/\nu_0 \simeq \delta\nu/\nu_0$.

We can relate the time at the pulsar t_p with the time on Earth t using the relation :

$$t_p = t - \frac{L}{c}(1 + \hat{k}\cdot\hat{n}), \tag{1.88}$$

where L is the distance Earth-pulsar. The right hand side term represents the delay due to the propagation of the pulse, taking into account the geometry of the system (see figure 1.14). We see that for $\hat{k}\cdot\hat{n} = -1$ we have $t_p = t$. In that particular case, the pulses and GW propagate on the same axis at same velocity c (the emitting SMBHB is right behind the pulsar), canceling the delay between the measured time at Earth and pulsar, hence the frequency shift induced in the TOAs. This is called the "surfing" effect.

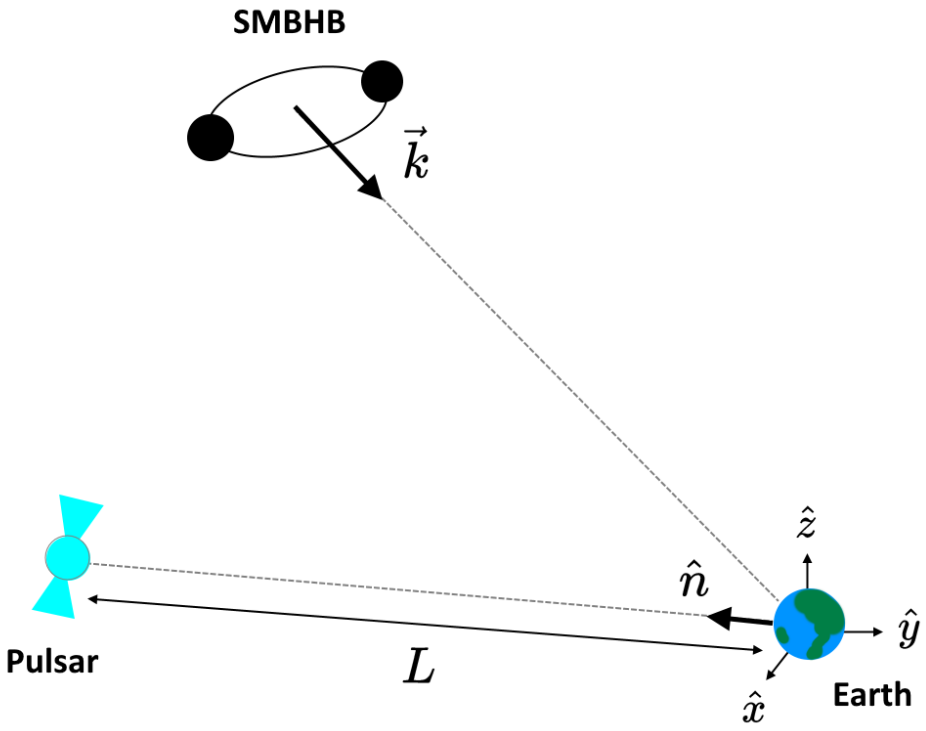

Figure 1.14: The geometry of the PTA in the radio observatory (Earth) frame. The SMBHB emitting a GW signal in the direction $\vec{k}$ is going to influence the TOAs of the pulsar observed on Earth. The angle between $\vec{k}$ and the unit vector $\hat{n}$ pointing to the sky position of the pulsar seen from Earth determines how the GW is felt by the pulsar according to the antenna pattern fuction $F^{+,\times}$. The distance Earth-pulsar L affects the delay in the pulsar term.

We can also calculate the scalar product between the geometrical prefactor function of $\hat{n}$ and the ϵ^A_{ij} using the definition of the polarization basis ($\hat{e}_1$, $\hat{e}_2$) and the expression of vectors in spherical coordinates :

$$\begin{aligned}
\hat{k} &= -\sin\theta\cos\phi\hat{x} - \sin\theta\sin\phi\hat{y} - \cos\theta\hat{z} \\
\hat{e}_1 &= -\sin\phi\hat{x} + \cos\phi\hat{y} \\
\hat{e}_2 &= -\cos\theta\cos\phi\hat{x} - \cos\theta\sin\phi\hat{y} + \sin\theta\hat{z}
\end{aligned} \tag{1.89}$$

This gives the antenna pattern functions F^A for both polarizations, expressing the response of the PTA as a function of the sky position $\hat{n}$ of the pulsar and the direction of propagation of the GW $\hat{k}$. We can write them as :

$$F^+(\hat{k}) = \frac{1}{2}\frac{(\hat{e}_1\cdot\hat{n})^2 - (\hat{e}_2\cdot\hat{n})^2}{1+\hat{k}\cdot\hat{n}}, \tag{1.90}$$

$$F^\times(\hat{k}) = \frac{(\hat{e}_1\cdot\hat{n})(\hat{e}_2\cdot\hat{n})}{1+\hat{k}\cdot\hat{n}}, \tag{1.91}$$

allowing us to rewrite equation (1.87) :

$$\frac{\delta\nu}{\nu_0} = -\sum_{A=+,\times} F^A(\hat{k})\big(h_A(t) - h_A(t_p)\big) \tag{1.92}$$

To get the timing residuals $\delta\vec{t}_{GW}$ induced in the TOAs by the passing GW signal, one needs to integrate again $\delta\nu/\nu_0$ to calculate the phase shift produced by this frequency modulation of the pulsar signal :

$$\int_0^t dt\frac{\delta\nu}{\nu_0} = -\sum_{A=+,\times} F^A(\hat{k})\int_0^t dt\big(h_A(t) - h_A(t_p)\big). \tag{1.93}$$

This requires to integrate the h_{ij} terms, which cannot be done analytically. However, it is possible to approximate this integral by neglecting the evolution of the orbital frequency $\omega(t)$ of the SMBHB (see section 1.4.1) producing the GWs over the course of PTA observation (of about 20 years). Given equations in (1.69), for non polarized signal ($\psi = 0$) we have :

$$\begin{aligned}
h_+ &\propto \omega(t)^{2/3}\cos(2\Phi(t)) \\
h_\times &\propto \omega(t)^{2/3}\sin(2\Phi(t)),
\end{aligned} \tag{1.94}$$

and for the integrals, neglecting the evolution of $\omega(t)$ and using the chain rule to relate Φ and t :

$$\begin{aligned}
\int dt h_+ &\propto \omega^{2/3}\int\left(\frac{d\Phi}{dt}\right)^{-1} d\Phi\cos(2\Phi(t)) \\
\int dt h_\times &\propto \omega^{2/3}\int\left(\frac{d\Phi}{dt}\right)^{-1} d\Phi\sin(2\Phi(t)),
\end{aligned} \tag{1.95}$$

with $d\Phi/dt = \omega$, we finally get :

$$\begin{aligned} \int dt h_+ &\propto \frac{1}{2}\omega^{-1/3}\sin(2\Phi(t)) \\ \int dt h_\times &\propto -\frac{1}{2}\omega^{-1/3}\cos(2\Phi(t)). \end{aligned} \tag{1.96}$$

After re-rotating the polarization basis by an angle 2ψ, we can write the timing residuals $\delta\vec{t}_{GW}$ induced in the TOAs of a pulsar by an individual SMBHB emitting CWs :

$$\boxed{\begin{aligned} \delta\vec{t}_{GW} &= \sum_{A=+,\times} F^A(\hat{k})\big(s_A(t) - s_A(t_p)\big) \\ s_+(t) &= \frac{\mathcal{M}_c^{5/3}G^{5/3}}{d_L c^4}\omega(t)^{-1/3}\Big((1+\cos^2\iota)\cos 2\psi\sin(2\Phi(t)) + 2\cos\iota\sin 2\psi\cos(2\Phi(t))\Big) \\ s_\times(t) &= \frac{\mathcal{M}_c^{5/3}G^{5/3}}{d_L c^4}\omega(t)^{-1/3}\Big((1+\cos^2\iota)\sin 2\psi\sin(2\Phi(t)) - 2\cos\iota\cos 2\psi\cos(2\Phi(t))\Big) \\ t_p &= t - \frac{L}{c}(1+\hat{k}\cdot\hat{n}). \end{aligned}} \tag{1.97}$$

Note that we have two terms acting in $\delta\vec{t}_{GW}$, the Earth term $s_A(t)$ and the pulsar term $s_A(t_p)$ (for both polarizations s_+ and $s_\times$ expressed above). We see that the amplitude of the timing residuals depends on $\omega(t)^{-1/3}$. It means that the sensitivity of the array decreases for signals with higher frequencies where the array experiences mainly white noise. We can expect an improved sensitivity for the low frequency signals but it is also mitigated by the presence of time correlated noises (red noise, dipsersion measure noise) as well as our limited time of observation (see section 1.4.1). The negative effect of the $\omega(t)^{-1/3}$ factor is reflected in the behaviour of the PTA sensitivity curve on figure 1.9 where we see its degradation as we go higher in frequency.

Even though we have neglected the evolution of $\omega(t)$ to approximate the integral in (1.95), it is not possible to neglect its evolution over the course of the whole travel of a pulse from pulsar to Earth. In section 1.4.1, we have explained why it is a reasonable approximation to consider the CGW sources to be monochromatic and non evolving during the time of observation (~10 years) for low frequency signals. However, as demonstrated here, PTA is sensitive to the difference in GW amplitude h felt by the pulsar and our radio-observatories on Earth. Pulsars are on average at a distance of about 1kpc, meaning that it takes ~ 1000 years for their radio emissions to reach us. For such timescale, it is no longer possible to ignore the frequency evolution of the SMBHB system. Therefore, the pulsar term in $\delta\vec{t}_{GW}$ can have different frequency and phase than the Earth term. For pulsar α, the frequency $\omega(t_{p,\alpha})$ and phase $\Phi(t_{p,\alpha})$ of the pulsar term are obtained using equations (1.64) and (1.65) at pulsar time t_p (equation 1.88). We can think of two particular cases (Babak 2016; Ellis et al. 2012) :

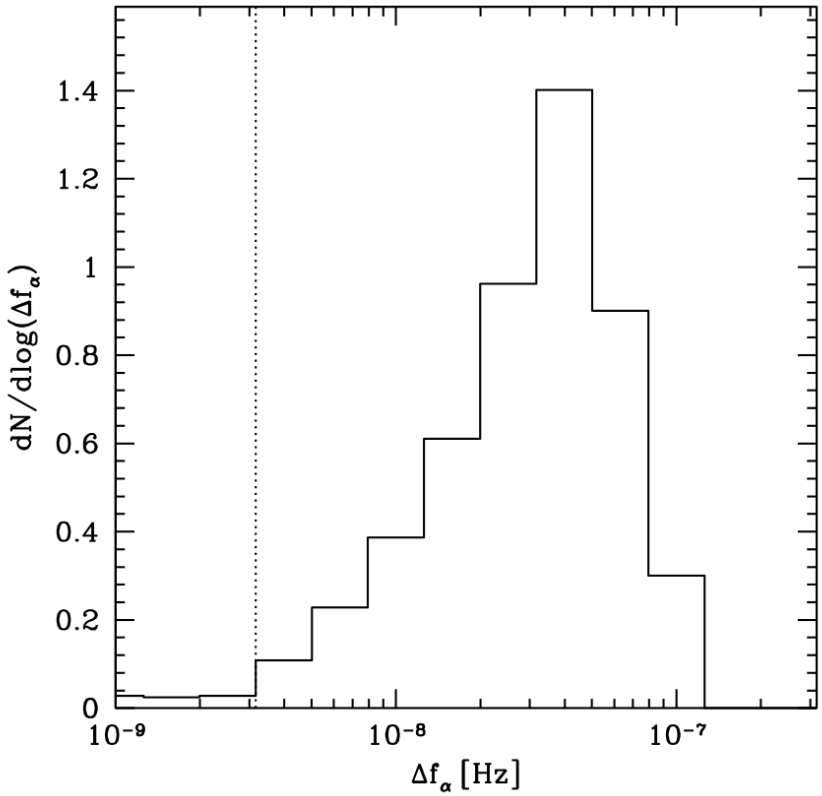

Figure 1.15: Distribution of Δf_α as shown in Sesana and Vecchio 2010. The vertical dotted line at ~3 nHz represents the frequency resolution of the array.

- $\Delta f_\alpha = \left[\omega(t_{p,\alpha}) - \omega(t)\right]/2\pi < 1/T$ where the pulsar term and the Earth term have nearly the same frequency and their difference is below the resolution $1/T$ of the array. However, the phase $\Phi(t_{p,\alpha})$ will be different for each pulsar because of their disparate distances and sky locations, destroying the phase coherency of the pulsar terms, while the Earth phase $\Phi(t)$ will be the same for all pulsar.

- $\Delta f_\alpha = \left[\omega(t_{p,\alpha}) - \omega(t)\right]/2\pi > 1/T$ where the pulsar term and the Earth term have different frequencies that the array can resolve. In that case, they can be treated as separate components. We can then include the pulsar term in our analysis or consider it to be an extra source of noise because of its random phase and frequency (different for all pulsars), focusing on the Earth terms that are actually in phase.

In Sesana and Vecchio 2010, they estimate the distribution of Δf_α for 100 isotropically distributed pulsars across the sky at 1kpc distance using a realistic simulation of SMBHB population (figure 1.15). They found that the frequency difference between pulsar and Earth term is mainly above the resolution of the array. To that extent, they can be treated separately.

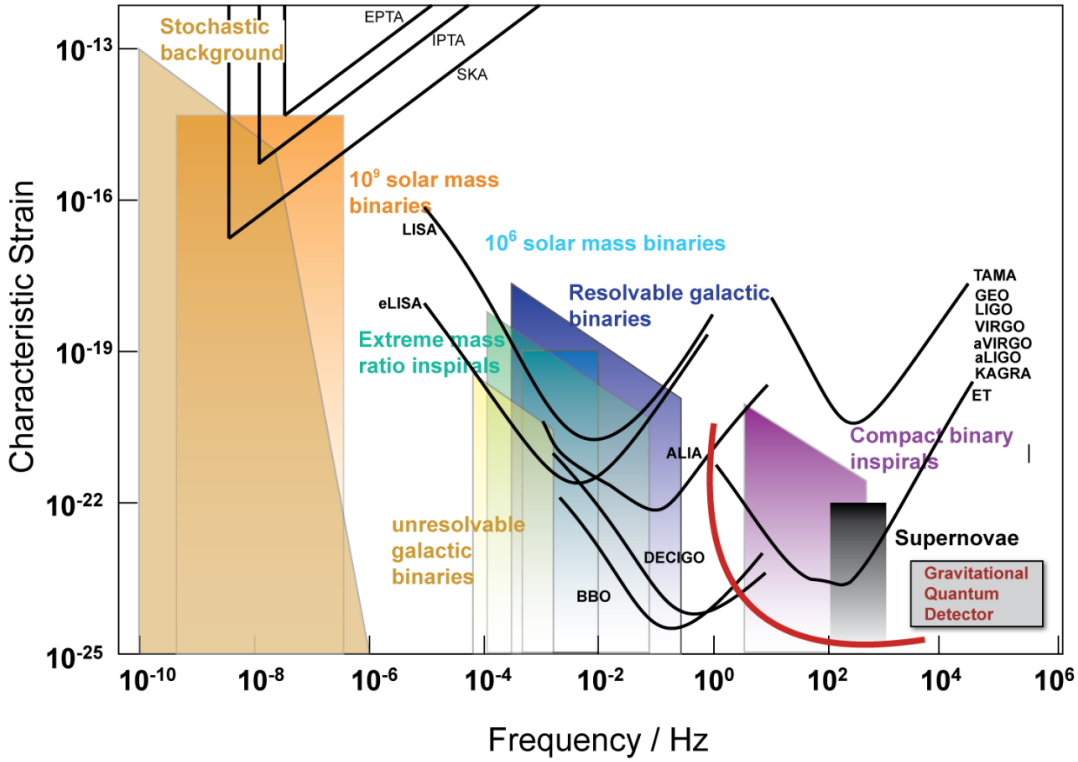

Figure 1.16: GW detectors sensitivities from Thermal noise in BEC-phononic gravitational wave detectors https://www.researchgate.net/publication/275974267_Thermal_noise_in_BEC-phononic_gravitational_wave_detectors

1.5 Multi-band gravitational wave astronomy

As we have seen in the previous section, PTA aims to detect very low frequency GW signal in the nanohertz band. But the efforts of the scientific community gave birth to other detectors allowing to probe the universe for the presence of GW at higher frequencies (between 1-10^3Hz). Some of these detectors are already functioning, some others are still being designed. We will give a brief overview of the main projects that are currently being developed (see figure 1.16).

1.5.1 Ground based detectors : LIGO/Virgo

The GW detectors of the LIGO (Laser Interferometer Gravitational-Wave Observatory) and Virgo collaborations follow the same concept. They are ground-based Michelson interferometers (LIGO in the USA and Virgo in Italy) with arms that are 3 to 4 km long. A GW passing through the detector modulates the length of the arms resulting in an interference pattern that can be measured[11] (Abbott et al. 2020). This technology is suited for the

[11] To be more precise, the GW affects the proper distance between the beam splitter and the mirrors.

detection of GWs emitted by merging stellar-mass black hole (or neutron star) binary systems because they can reach frequencies of ~ 10Hz (see figure 1.16), where the detector is most sensitive and the GW amplitude is loudest. SMBHBs are invisible to LIGO/Virgo because they merge before even getting to such high frequencies due their massive size.

The first direct detection of a GW merger signal was announced by the collaborations on 11 February 2016. The development of a new generation of ground-based detectors (like Einstein telescope, Maggiore et al. 2020) is currently an active field of research.

1.5.2 Space based detector : LISA

The Laser Interferometer Space Antenna (LISA) is an international project of space-based GW detector in heliocentric orbit. Like LIGO and Virgo, it uses laser interferometry to detect GWs, but with a different "triangle-shaped" configuration with arms that are ~2.5 million kilometers long. This massive size brings technical difficulties that are worked around using Time-Delay Interferometry to measure the frequency shift $\delta\nu/\nu$ of the lasers where the signature of GWs will be imprinted (for details, see Amaro-Seoane et al. 2017). It is most sensitive for GW signals with frequencies of ~ 10^{-3} Hz (see figure 1.16).

The Galactic Binaries (GBs) present in the Milky way (white dwarf binaries) will be detectable by LISA. They share similarities with the PTA CGWs in the sense that they will be seen as deterministic components resolved on top of a stochastic signal generated by the large population of GBs (the stochastic "foreground"). The equations governing their evolution are the same as the CGWs, but here for higher frequencies (~ 10^{-3} Hz) and lighter systems ($\mathcal{M}_c \sim 1M_\odot$). In chapter 4, we develop and test a data analysis method using simulated LISA data for the detection of GBs (see section 4.4).

CHAPTER 2

DATA ANALYSIS METHODS FOR PTA

Contents

2.1 Bayesian analysis

The topic of Bayesian analysis is very broad. Here, we will focus on the concepts that are interesting for our work. For more details, see Sivia and Skilling 2006.

2.1.1 Likelihood function

In statistics, the likelihood function $\mathcal{L}(\vec{\delta}t|\vec{\theta}_A, \mathcal{M}_A)$ assesses how likely it is to observe data $\delta\vec{t}$ given a set of parameters $\vec{\theta}_A$ of model $\mathcal{M}_A$. For example, under the assumption that

the samples in $\delta\vec{t}$ are independent and Gaussian distributed with a standard deviation σ_i that is determined by parameters $\vec{\theta}_A$[1], the likelihood writes as :

$$\mathcal{L}(\vec{\delta}t|\vec{\theta}_A,\mathcal{M}_A) = \prod_i \left\{ \frac{1}{\sqrt{2\pi}\sigma_i(\vec{\theta}_A)} \exp\left(-\frac{1}{2}\frac{\delta t_i^2}{\sigma_i^2(\vec{\theta}_A)} \right) \right\} = \frac{1}{\sqrt{|2\pi\mathbf{N}|}} \exp\left(-\frac{1}{2}\vec{\delta}t\mathbf{N}^{-1}\vec{\delta}t \right), \quad (2.1)$$

with $|.|$ denoting the determinant and $\mathbf{N} = \text{diag}\{\sigma_i(\vec{\theta}_A)\}$ the diagonal matrix containing the $\sigma_i(\vec{\theta}_A)$.

Intuitively, we want the values of parameters $\vec{\theta}_A$ that maximize the likelihood to match the data $\delta\vec{t}$ that was actually observed. Finding the maximum likelihood thus gives estimators of the parameters $\vec{\theta}_A$. However, working with fixed values does not provide information about the probability distribution of the $\vec{\theta}_A$.

2.1.2 Likelihood ratio test

We can compare two models A and B by taking the ratio of their respective likelihood functions. We often work with the logarithm of this ratio that we note Λ :

$$\ln\Lambda = \ln\frac{\mathcal{L}(\vec{\delta}t|\vec{\theta}_A,\mathcal{M}_A)}{\mathcal{L}(\vec{\delta}t|\vec{\theta}_B,\mathcal{M}_B)}. \quad (2.2)$$

This quantity can be used to measure the statistical significance of a model with respect to another. In the case of signal detection, we compare a model with noise + the anticipated signal to the model with only noise. We refer to the latter as the "null hypothesis" often noted $\mathcal{H}_0$. In the particular case where the two models are nested (i.e. when $\vec{\theta}_A = \vec{\theta}_B \bigcup \vec{\theta}_s$ with $\vec{\theta}_s$ the parameters associated with the additional signal present in A) and for large number of samples, this ratio follows a chi squared distribution (see Willk's theorem in Wilks 1938).

2.1.3 Posterior distribution

In the Bayesian framework, we set up a model and define a prior probability distribution for each of the parameters. This prior probability is a key point to the approach because model parameters are now treated as random variables with probability distributions.

We want to update our prior knowledge $\pi(\vec{\theta}_A|\mathcal{M}_A)$ of parameters $\vec{\theta}_A$ for a model $\mathcal{M}_A$, by expressing the probability of observing them given some data $\delta\vec{t}$. Using the rules of conditional probability we get the posterior probability :

[1] In section 2.2.2, we show that in PTA noise models the σ_i are function of two parameters EFAC and EQUAD characterizing the white noise.

$$p(\vec{\theta}_A|\vec{\delta t},\mathcal{M}_A) = \frac{\Pr(\vec{\delta t}|\vec{\theta}_A,\mathcal{M}_A)\Pr(\vec{\theta}_A|\mathcal{M}_A)}{\Pr(\vec{\delta t}|\mathcal{M}_A)} = \frac{\mathcal{L}(\vec{\delta t}|\vec{\theta}_A,\mathcal{M}_A)\pi(\vec{\theta}_A|\mathcal{M}_A)}{\mathcal{Z}_A}. \tag{2.3}$$

with $\vec{\theta}_A$ the vector of model parameters, $\mathcal{L}(\vec{\delta t}|\vec{\theta}_A,\mathcal{M}_A)$ the likelihood, $\pi(\vec{\theta}_A|\mathcal{M}_A)$ the prior probability and $\mathcal{Z}_A = p(\vec{\delta t}|\mathcal{M}_A)$ the evidence, expressing the probability of measuring data $\vec{\delta t}$ given model $\mathcal{M}_A$, acting as a normalizing constant.

In order to assess the statistical significance of a model $\mathcal{M}_A$ with respect to another model $\mathcal{M}_B$, we need to compare the odds $\Pr(\mathcal{M}_A|\vec{\delta t})$ and $\Pr(\mathcal{M}_B|\vec{\delta t})$ of the considered models given data $\vec{\delta t}$. Using conditional probabilities (Bayes theorem) we get the odds ratio :

$$\frac{\Pr(\mathcal{M}_A|\vec{\delta t})}{\Pr(\mathcal{M}_B|\vec{\delta t})} = \frac{\frac{\Pr(\vec{\delta t}|\mathcal{M}_A)\Pr(\mathcal{M}_A)}{\Pr(\vec{\delta t})}}{\frac{\Pr(\vec{\delta t}|\mathcal{M}_B)\Pr(\mathcal{M}_B)}{\Pr(\vec{\delta t})}} = \frac{\mathcal{Z}_A \times p(\mathcal{M}_A)}{\mathcal{Z}_B \times p(\mathcal{M}_B)}. \tag{2.4}$$

The evidence $\mathcal{Z}$ corresponds to the integral of the posterior probability distribution over whole parameter space (fully marginalized posterior) :

$$\mathcal{Z}_A = \int d\vec{\theta}_A p(\vec{\theta}_A|\vec{\delta t},\mathcal{M}_A) = \int d\vec{\theta}_A \mathcal{L}(\vec{\delta t}|\vec{\theta}_A,\mathcal{M}_A)\pi(\vec{\theta}_A|\mathcal{M}_A). \tag{2.5}$$

Assuming we are agnostic about models $\mathcal{M}_A$ and $\mathcal{M}_B$, we can take the two probabilities $p(\mathcal{M}_A)$ and $p(\mathcal{M}_B)$ to be equal. In that particular case, the odds ratio reduces to the Bayes factor :

$$\mathcal{B}_{AB} = \frac{\mathcal{Z}_A}{\mathcal{Z}_B}, \tag{2.6}$$

where $\mathcal{B}_{AB}$ is the Bayes factor, the quantity that we use for model selection. The bigger the $\mathcal{B}_{AB}$, the preferred model A is over model B. In GW data analysis, we use it to search for the presence of GWs in the data by comparing a model with a signal against another model without that signal. For large values of $\mathcal{B}_{AB}$, we can say that we have detected that signal. To decide which model is best, we can refer to Jeffrey's scale (Jeffrey 1998):

$\log_{10}\mathcal{B}_{AB}$	Strength of evidence
0 to 1/2	Not decisive
1/2 to 1	Substantial
1 to 2	Strong
>2	Decisive

It is important to note that other scales exist (Kass and Raftery 1995), so deciding whether the Bayes factor is significant may seem rather arbitrary. To that extent, it is possible to claim detection only when its value is high enough (>10^{2-3}). We must remain cau-

tious when dealing with intermediate values of $\mathcal{B}_{AB}$ that are sensitive to our own "bias" factors.

This quantity is often difficult to compute since we can only approximate the integral in (2.5) that yields the evidence $\mathcal{Z}$, because of our poor knowledge of the posterior distribution. To estimate the latter, we resort to sampling algorithms allowing us to explore the parameter space.

2.2 Modelling PTA data

2.2.1 The timing model

In order to get the timing residuals $\delta\vec{t}$ of a pulsar, we need to fit a timing model that gives a predicted value $\vec{t}_{pred}$ of the TOAs $\vec{t}_{obs}$. This predicted value needs to account for all physical effects responsible for the delays induced in the TOAs. Here, we only give a brief overview of the mechanisms at work. For a complete description, please see Edwards et al. 2006. A least square fit is performed to estimate parameter values for :

- The **spin frequency** and **spin-down** rate measuring the deceleration of the pulsar due to its emissions,
- The **sky position** and **proper motion** of the pulsar in the sky,
- The **dispersion measure** due to the propagation of the pulses through the interstellar medium with which they can interact during their journey to Earth,
- The **binary orbital effects** when the pulsar orbits a massive companion that can induce GR effects like Shapiro delay or Einstein delay (or simply Doppler shift due to the orbital motion).

The geomoetry of the problem is expressed in the Solar System Barycenter (SSB) that we estimate using Solar System Epheremides (SSE) providing the position of the planets at a given time. This translation calls for an additional delay, the Roemer delay given by :

$$\Delta_R = -\frac{\vec{r}\cdot\vec{R}_{BB}}{c}, \tag{2.7}$$

with $\vec{r}$ the position of the radio observatory in the SSB frame and $\vec{R}_{BB}$ the unit vector pointing to the sky position of the pulsar in the SSB frame. If this effect is not accurately quantified (i.e. errors in the SSE estimations hence in the position of the SSB), it can leave some artifacts in the timing residuals that will appear as a signal correlated between pulsars (Tiburzi et al. 2016).

For our analyses, we used a pre-fitted dataset including the timing residuals $\delta\vec{t}$ with the associated values of timing model parameters for all pulsars so we did not have to perform the timing model fit ourselves. However, there are still unmodelled noises in the $\delta\vec{t}$ that we will present in the next section.

2.2.2 Noises in PTA

To use PTA as a detector we need to identify all the sources of noise that will interfere with the desired signal (Verbiest and Shaifullah 2018). Each noise will be represented by a set of parameters in the model. One of the drawbacks of the PTA being a pulsar based detector is that it is not possible to characterize noises beforehand. They should be set as free parameters while running detection scripts, often leading to high dimensional parameter space and heavy computational cost.

We can group the different noises in three categories : white noise, red noise and dispersion noise (Chalumeau 2021, Goncharov et al. 2020). We will first explain what each noise category describes, then we will show the final expression of the PTA marginalised likelihood (equations 2.21 and 2.31).

White noise The white noise describes the observational errors dominating the high frequency of PTA sensitivity. The radio-telescopes used to produce the PTA datasets are often subject to random interferences, hence the evaluation of TOA uncertainties σ for one pulsar may be biased (Taylor 1992; Shannon et al. 2014). The white noise also deals with the statistics of pulse to pulse variations, hence the uncertainty on the determination of the averaged pulse profile (see section 1.4.1). These effets are taken into account with a rescaling of the uncertainties (see Chalumeau 2021) by a global multiplicative factor EFAC and a constant EQUAD added in quadrature as :

$$\hat{\sigma}_k^2 = \mathrm{E}_{f,k}^2 \times \sigma_{TOA,k}^2 + \mathrm{E}_{q,k}^2 \tag{2.8}$$

Each $E_{f,k}$ (EFAC), $E_{q,k}$ (EQUAD) pairs are different for each radiometer backend k and are applied as global rescaling parameters to every TOA uncertainty $\sigma_{TOA,k}$ coming from backend k.

We assume that the noises are gaussian and the residuals should behave as a normal distribution if the model is matching well with the data. Including these white noise parameters allows us to rescale the residuals so they fit this assumption.

Another white noise component is ECORR. It takes into account the jitter-like effect (pulse to pulse variation) that is correlated between TOAs for a given observation epoch. Note that this noise is only modelled in the NANOGRAV data and is not to be considered in other datasets[2] (Lentati et al. 2016).

[2]For example, in the EPTA dataset, the jitter-like effect is considered to be included in the EQUAD term.

In general, including white noise leads to a lot of additional parameters because of the many different backend-receiver systems in radio-telescopes. For that reason, it is common practice to evaluate the maximum likelihood values of EFAC, EQUAD and ECORR for the TOAs of each pulsar separately, and use these values as fixed for GW analysis that employs the whole PTA array. This allows to considerably reduce the size of the parameter space and the computational cost (Aggarwal 2019).

Red noise The red noise is a time-correlated noise intrinsic to each pulsar dominating at low frequencies (Shannon and Cordes 2010, Haasteren and Levin 2012). It takes into account possible stochasticity in pulsar rotation. We describe it as a Gaussian process decomposed on a Fourier basis with finite number of components. Its power spectral density is expressed as :

$$S_{RN}(f) = \frac{A_{RN}^2}{12\pi^2}\mathrm{yr}^3\left(\frac{f}{\mathrm{yr}^{-1}}\right)^{-\gamma_{RN}} \tag{2.9}$$

with A_{RN} the amplitude and γ_{RN} the spectral index.

In the case of an individual red noise, different values of A_{RN} and γ_{RN} are associated with each pulsar. To describe common red noise processes, we set the same values of A_{CRN} and γ_{CRN} for all pulsars, meaning that they share a red noise with same PSD. Lately, such common processes have been discovered by all PTA collaborations S Chen et al. 2021; Goncharov et al. 2021; Antoniadis 2022; Arzoumanian et al. 2020.

Dispersion measure variation The photons emitted from the pulsars get scattered by the interstellar medium on their way to Earth. The variability of the electron density on the axis Earth-pulsar introduces additional time-correlated noise in the data which resembles a low frequency red noise. However, photons with different frequencies will interact differently with the interstellar medium. Thus, observing incoming photons at different frequency bands allows us to distinguish this dispersion noise from a regular red noise. Its power spectral density writes as follows :

$$S_{DM}(f) = \frac{A_{DM}^2}{12\pi^2}\mathrm{yr}^3\left(\frac{f}{\mathrm{yr}^{-1}}\right)^{-\gamma_{DM}}\left(\frac{1400\mathrm{MHz}}{f_\nu}\right)^2 \tag{2.10}$$

with A_{DM} the amplitude, γ_{DM} the spectral index and f_ν the photon frequency[3].

The dispersion measure (DM) noise is individual to each pulsar because it depends on the density of the interstellar medium along the axis Earth-pulsar.

[3]Not to be confused with the spin velocity of the pulsar sending photon packets (pulses) with a rate ν. The frequency f_ν represents the "colour" of the photons measured on that frequency band.

Bayesephem The physics of PTA is expressed in the solar system barycenter (SSB) frame. To translate from Earth's reference frame to SSB, we need a precise knowledge of Earth's position with respect to the SSB (Hobbs et al. 2006, Vallisneri 2020). We get this information via solar system ephemeris (SSE). If the translation is not performed correctly, one can observe timing residuals in the TOAs due to errors in the determination of planetary motion (thus, an error in the determination of the SSB). Small deviations can lead to noticeable delay and interfere with GW detection. Bayesephem is a method that proposes to mitigate the systematic errors in the SSE by including deterministic corrections in the timing residuals (Roemer delay). Gaussian weights are given to small deviations in the orbital parameters of planets, allowing to absorb uncertainties on planetary position and correct errors on the determination of the SSB.

2.2.3 Single pulsar likelihood

To get the timing residuals $\vec{\delta t}_\alpha$ of one pulsar, we need to fit a timing model to the observed TOAs, $\vec{t}_{obs,\alpha}$. This timing model depends on many parameters (pulsar position, pulsar rotation frequency, spin down, parallax, ...) and gives a predicted value of the TOAs, $\vec{t}_{pred,\alpha}$ (Hobbs et al. 2006). The difference between the observed and predicted values of the TOAs gives the timing residuals. Small imperfections are present in the timing model and assumed to be linear so first order corrections are included for better precision (Lentati et al. 2013), with $\mathbf{M}$ the design matrix (containing the first derivatives of $\delta\vec{t}_{pred,\alpha}$ with respect to the timing model parameter errors) and ϵ the vector of small offset in timing parameters.

$$\vec{\delta t'}_\alpha = \vec{t}_{obs,\alpha} - \vec{t}_{pred,\alpha} = \mathbf{M}\epsilon + n. \tag{2.11}$$

The noises n are assumed to be Gaussian and stationary. It follows that the timing residuals $\delta\vec{t}_\alpha$ behave as a normal distribution as developped in Haasteren and Vallisneri 2014. Given the noises that we presented in the previous section, we can write for n :

$$n = n_{WN} + n_{RN} + n_{DM}, \tag{2.12}$$

where we have the correlated RN and DM noises[4] and the uncorrelated WN. We model the noises n as Gaussian processes (Haasteren and Vallisneri 2014) and we can express them in time domain as a Fourier decomposition to the N_fth harmonic :

$$n(\vec{t}_{obs,\alpha}) = \sum_l^{N_f} X_l \cos(2\pi f_l \vec{t}_{obs,\alpha}) + Y_l \sin(2\pi f_l \vec{t}_{obs,\alpha}) = \mathbf{F}a, \tag{2.13}$$

with

[4] In reality, many other noises can be added here, but only these two will be used in this manuscript

$$a = \begin{bmatrix} X_1 \\ Y_1 \\ X_2 \\ \vdots \\ X_{N_f} \\ Y_{N_f} \end{bmatrix} \quad \text{and} \quad \mathbf{F} = \begin{bmatrix} \cos(2\pi t_0 f_1) & \sin(2\pi t_0 f_1) & \dots & \cos(2\pi t_0 f_{N_f}) & \sin(2\pi t_0 f_{N_f}) \\ \cos(2\pi t_1 f_1) & \sin(2\pi t_1 f_1) & \dots & \cos(2\pi t_1 f_{N_f}) & \sin(2\pi t_1 f_{N_f}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \cos(2\pi t_{N_t-1} f_1) & \sin(2\pi t_{N_t-1} f_1) & \dots & \cos(2\pi t_{N_t-1} f_{N_f}) & \sin(2\pi t_{N_t-1} f_{N_f}) \end{bmatrix} \tag{2.14}$$

where N_t is the number of samples in $\vec{t}_{obs,\alpha}$, X_l and Y_l are the Gaussian distributed weights, the $\cos(2\pi t f_l)$ and $\sin(2\pi t f_l)$ are the basis functions and the f_l are the N_f linearly spaced frequencies starting from lowest frequency $f_0 = 1/T$ with T the total timespan of observation of the pulsar (Haasteren and Vallisneri 2014) $T = \max\{\vec{t}_{obs,\alpha}\} - \min\{\vec{t}_{obs,\alpha}\}$.

In theory, a Fourier basis is orthogonal when the $\vec{t}_{obs,\alpha}$ are evenly spaced. This is not the case for a real PTA dataset because of stochasticity in the sampled values of TOAs. We still consider that the basis is approximately orthogonal so we can write the PSD Σ as a function of the Fourier coefficients a :

$$\Sigma = < a^\top a > = \text{diag}\{S_n(f_L, \vec{\theta}_{HP})\}/T, \tag{2.15}$$

with $< . >$ denoting the averaging over the noise realizations, $S_n(f_L, \vec{\theta}_{HP})$ the PSD of the modeled noise evaluated at Fourier frequency f_L for hyper-parameters $\vec{\theta}_{HP}$ corresponding to, in the case of RN and DM, the amplitudes A_{RN}, A_{DM} and spectral indices γ_{RN}, γ_{DM}. Then, in the "weight-space" view of Gaussian processes (Haasteren and Vallisneri 2014), the Fourier coefficients a follow a Gaussian prior probability of :

$$\pi(a, \Sigma) = \frac{\exp\left(-\frac{1}{2} a^\top (\Sigma)^{-1} a\right)}{\sqrt{|2\pi\Sigma|}}. \tag{2.16}$$

We can also write the covariance matrix for noise n as :

$$< n^\top n > = \mathbf{F}^\top < a^\top a > \mathbf{F} = \mathbf{F}^\top \Sigma \mathbf{F}. \tag{2.17}$$

This result will be useful to express the total covariance matrix of the likelihood. The final expression for the residuals δt is :

$$\vec{\delta} t_\alpha = \vec{\delta t'}_\alpha - \mathbf{M}\epsilon - \mathbf{F}_{RN} a_{RN} - \mathbf{F}_{DM} a_{DM} \tag{2.18}$$

Assuming Gaussian likelihood for the $\vec{\delta t}_\alpha$ with uncorrelated white noise (therefore assuming a likelihood of the form 2.1), we have for the posterior distribution :

$$p(\delta\vec{t}_\alpha|\vec{\theta}_\alpha) = \frac{\exp\left(-\frac{1}{2}\vec{\delta t}_\alpha \mathbf{N}^{-1}\vec{\delta t}_\alpha\right)}{\sqrt{|2\pi\mathbf{N}|}} \frac{\exp\left(-\frac{1}{2}a^\top(\Sigma)^{-1}a\right)}{\sqrt{|2\pi\Sigma|}}\pi(\epsilon)\pi(\vec{\theta}_{HP}). \tag{2.19}$$

with $\mathbf{N}$ the diagonal matrix of the $\hat{\sigma}_i$ presented in equation (2.8), the Gaussian prior on a, $\pi(\epsilon)$ the prior on the timing model parameter offset and $\pi(\vec{\theta}_{HP})$ the prior on the hyper-parameters.

The posterior distribution can be analytically marginalized with respect to the a thanks to their Gaussian behaviour. However, $\pi(\epsilon)$ is a uniform distribution on the whole real axis and requires a different approach for marginalization. It is possible to mimick a uniform $\pi(\epsilon)$ using a Gaussian prior in the limit :

$$\lim_{\lambda\to\infty} \frac{\exp\left(-\frac{1}{2}\epsilon^\top(\lambda\mathbf{I})^{-1}\epsilon\right)}{\sqrt{|2\pi\lambda\mathbf{I}|}} = \pi(\epsilon). \tag{2.20}$$

with $\mathbf{I}$ the identity matrix. In that approximation, the ϵ are Gaussian weights and the design matrix M plays the role of basis functions. This method was first shown in Haasteren and Vallisneri 2014, allowing marginalisation on all parameters at once and giving the marginalised posterior distribution where the noise hyper-parameters $\vec{\theta}_{HP}$ (the amplitude and spectral indices of the Gaussian processes) appear in the covariance matrix $\mathbf{C}$ of the new marginalised likelihood $\mathcal{L}(\vec{\delta t}_\alpha|\vec{\theta},\mathbf{C}_\alpha)$ as

$$\begin{aligned} p(\delta\vec{t}_\alpha|\vec{\theta}_\alpha) &= \mathcal{L}(\vec{\delta t}|\vec{\theta}_\alpha,\mathbf{C}_\alpha)\pi(\vec{\theta}_{HP}) \\ &= \frac{1}{\sqrt{|2\pi\mathbf{C}_\alpha|}}\exp\left(-\frac{1}{2}\vec{\delta t'}_\alpha\mathbf{C}_\alpha^{-1}\vec{\delta t'}_\alpha\right)\pi(\vec{\theta}_{HP}), \end{aligned} \tag{2.21}$$

with[5] :

$$\begin{aligned} \mathbf{C}_\alpha &= \mathbf{N} + \mathbf{C}_{TM,\alpha} + \mathbf{C}_{RN,\alpha} + \mathbf{C}_{DM,\alpha} \\ &= \mathbf{N} + \mathbf{M}^\top\lambda\mathbf{I}\mathbf{M} + \mathbf{F}_{RN}^\top\Sigma_{RN}\mathbf{F}_{RN} + \mathbf{F}_{DM}^\top\Sigma_{DM}\mathbf{F}_{DM}. \end{aligned} \tag{2.22}$$

Bare in mind that this timing model marginalisation "trick" is an approximation as λ cannot be numerically set to ∞ and we use instead very large values (10^{40}). It is more convenient to work with this expression of the posterior for Bayesian analyses because the covariance matrix can be written as :

$$\mathbf{C}_\alpha = \mathbf{N} + \mathrm{T}^\top\Phi\mathrm{T}. \tag{2.23}$$

[5]Note that we omit the index α for matrices $\mathbf{N}$, $\mathbf{M}$, $\mathbf{F}$ and Σ to avoid overcrowded equations but it should actually be there because they describe properties of pulsar α.

$$\mathrm{T} = \begin{bmatrix} M \\ F_{RN} \\ F_{DM} \end{bmatrix} \Phi = \begin{bmatrix} \lambda\mathbf{I} & 0 & 0 \\ 0 & \Sigma_{RN} & 0 \\ 0 & 0 & \Sigma_{DM} \end{bmatrix}. \tag{2.24}$$

In this form, the covariance matrix can be inverted using the Woodbury lemma :

$$\begin{aligned} \mathbf{C}_\alpha^{-1} &= (\mathbf{N} + \mathrm{T}^\top \Phi \mathrm{T})^{-1} \\ &= \mathbf{N}^{-1} - \mathbf{N}^{-1}\mathrm{T}(\Phi^{-1} + \mathrm{T}^\top \mathbf{N}^{-1}\mathrm{T})^{-1}\mathrm{T}^\top \mathbf{N}^{-1}, \end{aligned} \tag{2.25}$$

which is used to optimize numerical computation. In ENTERPRISE, the argument of the exponential $\delta\vec{t}_\alpha \mathbf{C}_\alpha^{-1} \delta\vec{t}_\alpha$ is computed by first solving the linear system for x and taking the dot product with $\delta\vec{t}_\alpha \mathbf{N}^{-1}\mathrm{T}$ from the left :

$$\begin{aligned} &(\Phi^{-1} + \mathrm{T}^\top \mathbf{N}^{-1}\mathrm{T})\vec{x} = \mathrm{T}^\top \mathbf{N}^{-1} \delta\vec{t}_\alpha \\ \Leftrightarrow &\delta\vec{t}_\alpha \mathbf{N}^{-1}\mathrm{T} \cdot \vec{x} = \delta\vec{t}_\alpha^\top \mathbf{N}^{-1}\mathrm{T}(\Phi^{-1} + \mathrm{T}^\top \mathbf{N}^{-1}\mathrm{T})^{-1}\mathrm{T}^\top \mathbf{N}^{-1}\delta\vec{t}_\alpha, \end{aligned} \tag{2.26}$$

giving,

$$\delta\vec{t}_\alpha \mathbf{C}_\alpha^{-1} \delta\vec{t}_\alpha = \delta\vec{t}_\alpha \mathbf{N}^{-1}\delta\vec{t}_\alpha - \delta\vec{t}_\alpha \mathbf{N}^{-1}\mathrm{T} \cdot \vec{x} \tag{2.27}$$

Finally, if we want to add the contribution of any deterministic signal to our model (CGW signal, BAYESEPHEM), we need to transform the residuals as follows :

$$\delta\vec{t}_\alpha \to \delta\vec{t}_\alpha - s(\vec{t}_{obs,\alpha}, \vec{\theta}) \tag{2.28}$$

where $s(\vec{t}_{obs,\alpha}, \vec{\theta})$ is a deterministic signal, function of time.

2.2.4 Multi pulsar likelihood

In the case of deterministic GGW signal, the likelihood for an array of pulsars is simply the product of the single pulsar likelihoods. For N pulsars indexed by α, we have the total timing residual vector :

$$\delta\vec{t} = \bigcup_\alpha \delta t_\alpha. \tag{2.29}$$

When we are using the whole array, the Fourier basis that is used for GP noises is defined from the total observation timespan of the array ($T = \max\{\bigcup_\alpha \delta\vec{t}_{obs,\alpha}\} - \min\{\bigcup_\alpha \delta\vec{t}_{obs,\alpha}\}$) and the covariance matrix is :

$$\mathbf{C}_{ab} = \begin{bmatrix} \mathbf{C}_0 & 0 & 0 & 0 \\ 0 & \mathbf{C}_1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & C_N \end{bmatrix}. \tag{2.30}$$

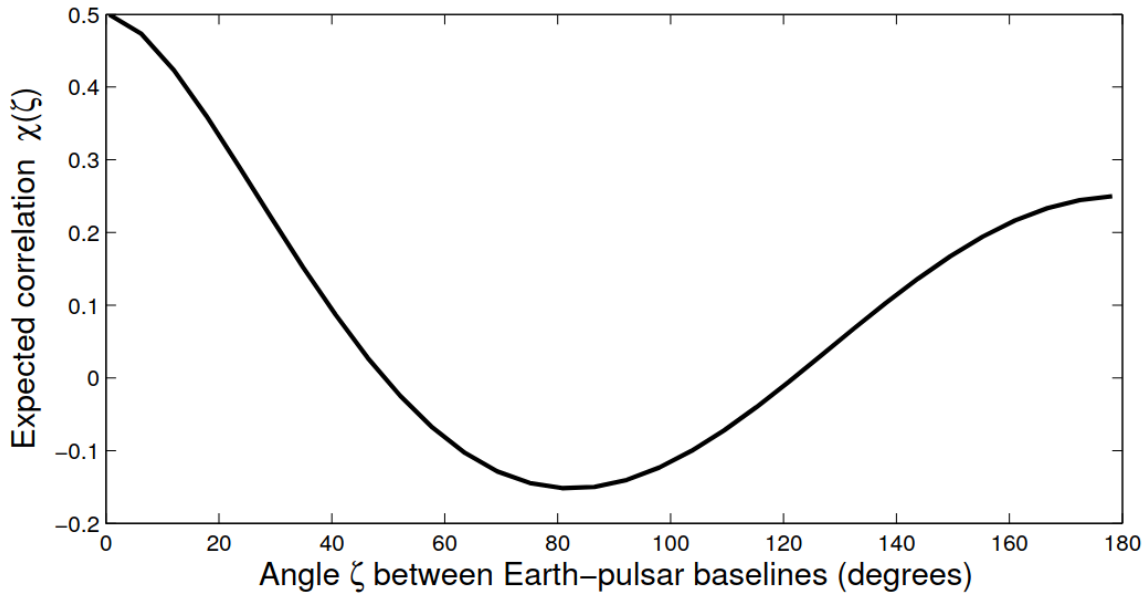

Figure 2.1: Hellings and Down correlation with χ a function of angular separation ζ between two pulsars a and b (plot taken from Jenet and Romano 2014)

In this thesis, we are interested only in CGW signal so the covariance matrix we use is always block diagonal and the total likelihood factorises as the product of single pulsar likelihoods :

$$\mathcal{L}_{PTA}(\vec{\delta t}|\vec{\theta}, \mathbf{C}_{ab}) = \prod_{\alpha} \mathcal{L}(\vec{\delta t}|\vec{\theta}_{\alpha}, \mathbf{C}_{\alpha}), \tag{2.31}$$

where the $\mathbf{C}_{\alpha}$ appear in the diagonal of the $\mathbf{C}_{ab}$.

2.2.5 Correlated signals

When including correlated signals to the model, the covariance matrix presented above is no longer block diagonal. The off diagonal terms can follow various correlation patterns χ_{ab} that are function of the angular separation ζ between pulsars.

Hellings and Down PTA collaborations mainly target quadrupolar correlation patterns because they agree with the presence of GW signal. We can recover such correlations by calculating the overlap between the antenna pattern functions of two pulsars with angular separation ζ. This yields the Hellings and Down (HD) curve that is expected to be the signature of the GWB (see Hellings and Downs 1983).

The GWB has a correlation matrix $\mathbf{C}_{GWB}$ similar to that of a common red noise but with additional correlation terms χ_{ab} that relate two pulsar a and b given their angular separation ζ_{ab} (see figure 2.1) :

$$\mathbf{C}_{GWB} = \chi_{ab}\mathbf{C}_{RN}, \tag{2.32}$$

with $\chi_{ab} = 1$ if $a = b$.

Dipolar correlation Dipolar correlations can originate from errors in the SSE (Tiburzi et al. 2016). The position of the observatories on Earth with respect to the SSB must be known to high precision, otherwise measurable (and correlated between pulsars) effects will appear in the timing residuals when the translation to the SSB is made (see section 2.2.1). In that case, the correlation function is of the form :

$$\chi_{ab} \propto \cos(\zeta_{ab}). \tag{2.33}$$

A dipolar correlation is somewhat similar to a HD curve. This resemblance can cause problems when searching for GWB because the two signals may interact in the model. A correct modelisation of the SSE is then crucial to PTA efficiency (Vallisneri 2020).

Monopolar correlation A signal with monopolar correlation is fully correlated between pulsars. There is no dependence in the angular separation ζ, hence the off-diagonal terms of the covariance matrix are equivalent. We have :

$$\chi_{ab} = 1. \tag{2.34}$$

One possible origin for monopolar features is clock correction errors. The atomic clock systems in the radio-telescopes allowing local time measurements can accumulate errors that must be corrected in order to keep them synchronized. If the correction is not done correctly, it can cause a low frequency noise in the timing residuals of the TOAs measured with the affected radio-telescope (Tiburzi et al. 2016).

2.3 Markov Chain Monte Carlo

When dealing with high dimensional models and complex datasets, we often work with difficult posterior probability distributions that are not analytically invertible or for which finding maxima is not trivial. Furthermore, it is almost impossible to visualize the probability distributions of individual parameters because of the tremendous amount of computation time that it would represent. For illustration, if we want to plot a posterior probability distribution $p(\vec{\theta})$ of a d-dimensional parameter space on a grid of N points per parameter, we would need N^d evaluations of the probability function $p(\vec{\theta})$. Assume that one computation of $p(\vec{\theta})$ is fast ($\sim 10^{-3} s$) and we want a grid of $N = 10$ points (low resolution) for a 10-dimensional parameter space. This plot would take :

$$10^{-3} \times 10^{10} = 10^7 s \simeq 3 \text{ months and } 3 \text{ weeks}, \tag{2.35}$$

which is ridiculously long for such a low resolution plot. To tackle the issue, numerical algorithms were proposed, one of the most popular being Markov Chain Monte Carlo (MCMC). It allows to sample points from a target distribution $p(\vec{\theta})$ while focusing on the interesting regions, in a limited number of evaluations of the function. Before presenting the sampling algorithm, we must explain what a Markov process is in order to justify the motivation behind it. For more details, see Brooks et al. 2011.

2.3.1 Markov process

A Markov process (or Markov chain) is a stochastic process defining a series of events in which each event only depends on the previous. Such a process is said to be "without memory" or simply "Markovian". Let $X = \{A, B, C, ...\}$ a random variable following the rules of a Markov process and $\vec{p}(t)$ a vector expressing the probabilities of finding X at states $\{A, B, C, ...\}$ at discrete time t. For a set of samples $\{X_T\}$, the previous no-memory condition is satisfied by :

$$\Pr(X_T | X_{T-\Delta t}, X_{T-2\Delta t}, X_{T-3\Delta t}, ...) = \Pr(X_T | X_{T-\Delta t}). \tag{2.36}$$

Because of this condition, we can define a transition matrix Q that is independent of time, relating two successive events at times t and $t + \Delta t$ in the chain so that the components of probability vector $\vec{p}(t + \Delta t)$ change as :

$$p_j(t + \Delta t) = \sum_i Q_{ij} p_i(t). \tag{2.37}$$

To preserve probabilities $p_j(t + \Delta t)$ to be between 0 and 1 and $\vec{p}(t + \Delta t)$ to be normalized, Q must be positive and respect the condition :

$$\sum_j Q_{ij} = 1 \tag{2.38}$$

When vector $\vec{p}(t + \Delta t)$ is different from $\vec{p}(t)$, we are in transient regime and probabilities are function of time. But when vector $\vec{p}(t + \Delta t)$ is equal to $\vec{p}(t)$, we have a stationary process for which we will denote probability vector as $\vec{\mathcal{P}}$ so that :

$$\mathcal{P}_j = \sum_i Q_{ij} \mathcal{P}_i. \tag{2.39}$$

The time independence of stationary processes is a key element that is utilized in the MCMC algorithms. Nonetheless, stationary does not necessarily mean equilibrium and it is custom to define a stronger condition to force stationarity.

2.3.2 Detailed balance

The detailed balance is a way of defining equilibrium by reversibility of a process. Indeed, if a process is reversible it is stationary, but not always the other way around. Detailed balance is expressed as :

$$Q_{ij}\mathcal{P}_i = Q_{ji}\mathcal{P}_j \tag{2.40}$$

If this equality is respected, we have :

$$\begin{aligned} & Q_{ij}\mathcal{P}_i = Q_{ji}\mathcal{P}_j \\ \Leftrightarrow & \sum_i Q_{ij}\mathcal{P}_i = \sum_i Q_{ji}\mathcal{P}_j \\ \Leftrightarrow & \sum_i Q_{ij}\mathcal{P}_i = \mathcal{P}_j, \end{aligned} \tag{2.41}$$

where we have used (2.38) to go from second to last line. In other words, this ensures that going forwards or backwards in time has no effect on the system. A reversible Markov process must fulfill this condition and the core sampling algorithm of MCMC is entirely based on it.

2.3.3 The sampling algorithm

Until now, we have been working with a finite number of components for probability vector p_i and transition matrix Q_{ij}. We can easily generalize to continuous functions by saying :

$$\begin{aligned} & p_i \rightarrow p(\vec{\theta}) \\ & Q_{ij} \rightarrow q(\vec{\theta}|\vec{\theta}^*), \end{aligned} \tag{2.42}$$

where $\vec{\theta}$ is a vector in a d-dimensional parameter space following probability distribution $p(\vec{\theta})$. In that case, the detailed balance condition is :

$$q(\vec{\theta}|\vec{\theta}^*)\,p(\vec{\theta}) = q(\vec{\theta}^*|\vec{\theta})\,p(\vec{\theta}^*) \tag{2.43}$$

The idea behind the sampling algorithm is to use the properties of a stationary Markov process to draw samples from a target distribution $p(\vec{\theta})$. In other words, starting from an initial sample $\vec{\theta}_0$, we want to build a Markov chain (i.e. a set of N samples $\{\vec{\theta}_N\}$) that performs a random walk in the parameter space following the probability of $p(\vec{\theta})$. By ensuring that the target $p(\vec{\theta})$ is the stationary distribution of the Markov process, we ensure that the $\vec{\theta}$ are sampled from $p(\vec{\theta})$. The resulting set of samples $\{\vec{\theta}_N\}$ can later be post-processed to study the probability distributions of the model parameters.

For any probability distribution $p(\vec{\theta})$, we want to find a transition matrix $q(\vec{\theta}^*|\vec{\theta})$ that always respects the detailed balance condition. We first introduce a proposal distribution $g(\vec{\theta}^*|\vec{\theta})$ from which it is possible to draw new samples $\vec{\theta}^*$ given $\vec{\theta}$, the current sample (or current state of the Markov chain) [6]. Then we define an acceptance probability $\alpha(\vec{\theta}^*,\vec{\theta})$ that decides if the new sample is accepted or rejected. We can write the transition kernel as the product of the two :

$$q(\vec{\theta}^*|\vec{\theta}) = g(\vec{\theta}^*|\vec{\theta})\alpha(\vec{\theta}^*,\vec{\theta}), \tag{2.44}$$

and plugging this expression in (2.43), we get :

$$\frac{\alpha(\vec{\theta}^*,\vec{\theta})}{\alpha(\vec{\theta},\vec{\theta}^*)} = \frac{p(\vec{\theta}^*)g(\vec{\theta}|\vec{\theta}^*)}{p(\vec{\theta})g(\vec{\theta}^*|\vec{\theta})}. \tag{2.45}$$

We can say that if a new jump going from $\vec{\theta}$ to $\vec{\theta}^*$ is not accepted, then the same jump but reversed would always be accepted, that is to say, $\alpha(\vec{\theta},\vec{\theta}^*) = 1$. Thus, we have :

$$\alpha(\vec{\theta}^*,\vec{\theta}) = \min\Big\{1, \frac{p(\vec{\theta}^*)g(\vec{\theta}|\vec{\theta}^*)}{p(\vec{\theta})g(\vec{\theta}^*|\vec{\theta})}\Big\}, \tag{2.46}$$

where the *min* ensures that $0 \leq \alpha(\vec{\theta}^*,\vec{\theta}) \leq 1$. This is called the Metropolis-Hastings choice (Hastings 1970).

The acceptance probability will entirely determine our random walk in the parameter space, which by construction, respects the detailed balance condition for target distribution $p(\vec{\theta})$. Starting from sample $\vec{\theta}$, we jump to sample $\vec{\theta}^*$ with a probability $\alpha(\vec{\theta}^*,\vec{\theta})$ or stay at sample $\vec{\theta}$ with probability $1-\alpha(\vec{\theta}^*,\vec{\theta})$. The process is repeated N times and obtained samples at each step are concatenated into a chain (Markov chain) that forms our final set of samples $\{\vec{\theta}_N\}$ following the target probability distribution $p(\vec{\theta})$.

Let us summarize the sampling algorithm in steps :

- For current (or initial) sample $\vec{\theta}_0$ compute $p(\vec{\theta}_0)$
- Draw new sample $\vec{\theta}^*$ from proposal distribution $g(\vec{\theta}^*|\vec{\theta}_0)$ and compute $p(\vec{\theta}^*)$
- Compute the acceptance probability $\alpha(\vec{\theta}^*,\vec{\theta}_0) = \min\Big\{1, \frac{p(\vec{\theta}^*)g(\vec{\theta}_0|\vec{\theta}^*)}{p(\vec{\theta}_0)g(\vec{\theta}^*|\vec{\theta}_0)}\Big\}$
- Accept new sample with probability $\alpha(\vec{\theta}^*,\vec{\theta})$ and set $\vec{\theta}_0 = \vec{\theta}^*$ **OR** reject new sample with probability $1-\alpha(\vec{\theta}^*,\vec{\theta})$ and set $\vec{\theta}_0 = \vec{\theta}$
- Append $\vec{\theta}_0$ to chain of samples and repeat

[6] The proposal distribution can be anything, from a Gaussian distribution centered on $\vec{\theta}$ to a uniform distribution, as long as the output sample is random.

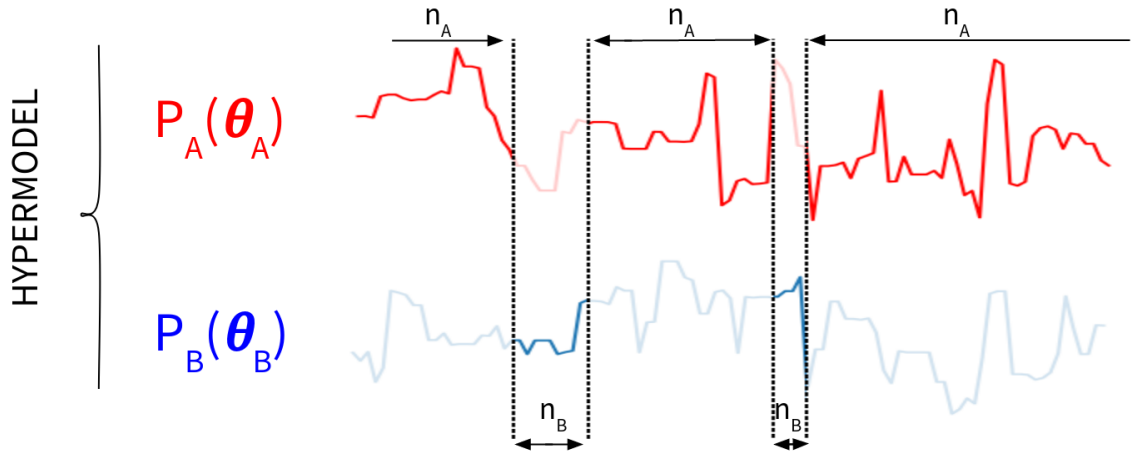

Figure 2.2: Illustration of the hypermodel method. Here, the hypermodel is the union of two different models A and B (red and blue) with posterior distributions P_A and P_B functions of parameters $\vec{\theta}_A$ and $\vec{\theta}_B$. As the sampling goes on, the sampler can jump between models by selecting which posterior is active. It will spend more time on a model if it is favoured by the Metropolis-Hasting choice (model A in this example). The time spent on models A and B corresponds respectively to the total number of samples n_A and n_B.

2.3.4 Hypermodels for model selection

To compute the Bayes factor, we can use a method presented in Hee et al. 2015 called hypermodels. A hypermodel is the union of two or several submodels. Its vector of parameters $\vec{\theta}$ is :

$$\vec{\theta} = \bigcup_{i=0}^{N-1} \vec{\theta}_i \tag{2.47}$$

with $\vec{\theta}_i$ the parameters for submodel i and N is the total number of submodels.

We introduce an additional integer parameter j with uniform probability distribution $n(j) = \{0,1,2,...,N-1\}$. Throughout sampling, we draw an integer k from $n(j)$ that determines which submodel is going to be active for the likelihood computation so that :

$$\mathcal{L}(\vec{\theta}, j = k|\delta\vec{t}) = \mathcal{L}_k(\vec{\theta}_k|\delta\vec{t}) \tag{2.48}$$

where $\mathcal{L}_k(\vec{\theta}_k|\delta\vec{t})$ the likelihood of submodel k. In that case, the posterior distribution of the hypermodel is :

$$
\begin{aligned}
p(\vec{\theta}, j|\delta\vec{t}) &= p(\vec{\theta}_0 \cup \vec{\theta}_1 \cup ... \cup \vec{\theta}_i \cup ... \cup \vec{\theta}_{N-1}|\delta\vec{t})n(j) \\
&= p\Big(\prod_{i=0}^{N-1} \cup\vec{\theta}_i|\delta\vec{t}\Big)n(j) \\
&= \mathcal{L}\Big(\prod_{i=0}^{N-1} \cup\vec{\theta}_i|\delta\vec{t}\Big)\prod_{i=0}^{N-1}\pi(\vec{\theta}_i)n(j)/\mathcal{Z},
\end{aligned}
\tag{2.49}
$$

where $\mathcal{Z}$ is the evidence of the hypermodel acting as a normalizing constant and $\pi(\vec{\theta}_i)$ the prior probabilities for parameters $\vec{\theta}_i$ of submodel i. Fixing $j = k$, we get :

$$
p(\vec{\theta}, j = k|\vec{\delta}t) = \mathcal{L}_k(\vec{\theta}_k|\vec{\delta}t)\prod_{i=0}^{N-1}\pi(\vec{\theta}_i)n(j = k)/\mathcal{Z}, \tag{2.50}
$$

which the active posterior distribution when an integer k is drawn for $n(j)$. If we marginalize this expression for all $\vec{\theta}$, we obtain :

$$
\begin{aligned}
p(j = k|\delta\vec{t}) &= \int d\vec{\theta}\, p(\vec{\theta}, j = k|\vec{\delta}t) \\
&= \int d\vec{\theta}\mathcal{L}_k(\vec{\theta}_k|\vec{\delta}t)\prod_{i=0}^{N-1}\pi(\vec{\theta}_i)n(j = k)/\mathcal{Z} \\
&= \int d\vec{\theta}_k\mathcal{L}_k(\vec{\theta}_k|\vec{\delta}t)\pi(\vec{\theta}_k)\prod_{\substack{i=0\\ i\neq k}}^{N-1}\int d\vec{\theta}_i\pi(\vec{\theta}_i)n(j = k)/\mathcal{Z}.
\end{aligned}
\tag{2.51}
$$

Using the definition of the evidence $\mathcal{Z}_k$ and the fact that prior probabily distributions $\pi(\vec{\theta}_i)$ are normalized :

$$
\begin{aligned}
\int d\vec{\theta}_k\mathcal{L}_k(\vec{\theta}_k|\vec{\delta}t)\pi(\vec{\theta}_k) &= \mathcal{Z}_k \\
\prod_{\substack{i=0\\ i\neq k}}^{N-1}\int d\vec{\theta}_i\pi(\vec{\theta}_i) &= 1,
\end{aligned}
\tag{2.52}
$$

equation (2.51) greatly simplifies to :

$$
\begin{aligned}
p(j = k|\delta\vec{t}) &= \mathcal{Z}_k n(j = k)/\mathcal{Z} \\
\Leftrightarrow \mathcal{Z}_k &= \frac{p(j = k|\delta\vec{t})}{n(j = k)}\mathcal{Z}.
\end{aligned}
\tag{2.53}
$$

The $p(j = k|\delta\vec{t})$ corresponds to the odds ratio for parameter j to be at value k. So for a chain of N samples, if n_k is the number of samples corresponding to $j = k$, we have :

$$
p(j = k|\delta\vec{t}) = \frac{n_k}{N}, \tag{2.54}
$$

and for two different models k and k', we directly see that :

$$
\frac{\mathcal{Z}_{k'}}{\mathcal{Z}_k} = \frac{n_{k'}}{n_k}, \tag{2.55}
$$

where we have used $n(j = k) = n(j = k')$ because $n(j)$ is a uniform distribution. Recall from section 2.1.3 that the ratio of evidences corresponds to the Bayes factor which is the quantity that we used to do model selection. Therefore, using hypermodels, the evaluation of the Bayes factor between models A and B is :

$$\mathcal{B}_{AB} = \frac{n_A}{n_B} \tag{2.56}$$

with n_A and n_B respectively the number of final samples for which submodel A was active and submodel B was active. They represent the time spent by the sampler on one model. Comparing them shows which model is preferred according to the MCMC rules.

The big advantage of this technique is that it requires no direct evidence estimates which are usually complicated multidimensional integrals[7]. However, it can sometimes be a bit long to run because we need the sampler to efficiently explore every submodel. This requires many samples for which we need to run our MCMC for longer.

[7]Other methods like parallel-tempering or nested sampling can estimate these difficult integrals with fairly good accuracy.

CHAPTER 3

SEARCHING FOR CONTINUOUS GRAVITATIONAL WAVE SIGNAL WITH PTA

Contents

The International Pulsar Timing Array 2nd data release is the combination of datasets from worldwide collaborations. In this study, we search for continuous waves: gravitational wave signals produced by individual supermassive black hole binaries in the local universe. We consider binaries on circular orbits and neglect the evolution of orbital frequency over the observational span. We find no evidence for such signals and set sky averaged 95% upper limits on their amplitude h_{95}. The most sensitive frequency is 10nHz with $h_{95} = 9.1 \times 10^{-15}$. We achieved the best upper limit to date at low and high frequencies of the PTA band thanks to improved effective cadence of observations. In our analysis, we have taken into account the recently discovered common red noise process, which has an impact at low frequencies. We also

find that the peculiar noise features present in some pulsars data must be taken into account to reduce the false alarm. We show that using custom noise models is essential in searching for continuous gravitational wave signals and setting the upper limit.

3.1 Introduction

The goal of the Pulsar Timing Array (PTA) collaborations is to detect gravitational wave (GW) signals in the nanohertz band, where we expect to see a gravitational wave background (GWB) produced by the superposition of GW signals from the population of supermassive black hole binaries (SMBHBs) (Maiorano et al. 2021). Some individual SMBHBs might be brighter than most and stand above the stochastic signal; those are individually resolvable sources. The binaries detectable in the PTA band are in the orbits with the period from a few months to a few years and emit almost monochromatic GWs continuously during decades; we refer to those signals as continuous GWs (CGWs) (Ellis et al. 2012; Babak 2016; Aggarwal 2019; Corbin and Cornish 2010).

The GWs affect propagation of the radio emission from millisecond pulsars leaving an imprint in the time of arrival (TOA) of pulses observed with the radio telescopes. CGWs impact TOAs from all observed millisecond pulsars in a deterministic manner characterized by parameters of the SMBHBs. In this work, we consider the data combined by the International Pulsar Timing Array (IPTA). IPTA is a consortium of NANOGRAV (NANOGrav Collaboration et al. 2015), European Pulsar Timing Array (EPTA) (Desvignes et al. 2016), Australian (PPTA) (Manchester et al. 2013b) and Indian Pulsar Timing Array (InPTA) (Tarafdar et al. 2022) collaborations. In particular, we analyze the second data released by IPTA (IPTA DR2) described in details in Perera 2019.

Recently, PTA collaborations have reported on the discovery of the common red noise signal, that is the stochastic signal with the spectral shape common to all pulsars in the array. Its high statistical significance was demonstrated independently by three collaborations (Arzoumanian et al. 2020; S Chen et al. 2021; Goncharov et al. 2021) and, with even higher statistical confidence, was assessed using the IPTA DR2 (Antoniadis 2022). We do not yet know the nature of this process, and its interpretation as GW background is inconclusive: the data is not informative enough to resolve the Helllings-Downs spatial correlations (Hellings and Downs 1983), which should be present in the case of the GW signal.

In this work, we search for continuous GWs which could be present in the data in addition to the stochastic GWB. Following the steps of previous studies (Babak 2016; Aggarwal 2019; Zhu et al. 2014; Becsy et al. 2022; Becsy and Cornish 2020), we search for a single GW signal produced by a SMBHB binary in a circular orbit. In our study, we neglect the pulsar

terms during the search and setting an upper limit on GW amplitude. However, we do an in-depth analysis of the (weak) candidate events identified as plausible GW signals. In the followup analysis on the restricted parameter space (frequency and sky position), we extend our model to include (i) pulsar term, (ii) eccentricity, (iii) extend the model beyond the assumption of a single source. For the first time, we have included in the analysis the common red component as a part of the total noise model.

The main results of the paper can be summarized as follows. We did not detect any CGW signal and set an upper limit on GW amplitude. We have found that the noise model plays a crucial role in interpreting PTA observations. The detailed analysis performed on the most promising candidate event revealed that it could be explained by a time-correlated high-frequency noise in one of the pulsars.

This chapter is organized as follows. In the next Section, we will briefly describe the IPTA DR2 dataset and the data model used in the analysis. Most of the material needed for this Section is available in the literature, and we heavily refer to it, keeping only parts which are necessary for further presentation. In Section 3.4 we describe the methodology which we have followed to get our results presented in Section 3.5. In Section 3.6 we give a detailed follow-up study of a most promising GW candidate event and demonstrate the importance of noise modelling at high frequencies. We conclude with Section 3.7. **Throughout this chapter, we adopt geometrical units G=c=1.**

3.2 The noise and signal model

3.2.1 The noise model

In all that follows, the noise model that we use is based on what is presented in section 2.2.2. That is to say, we use a PTA likelihood with marginalized timing model, including red noise and dispersion measure noise. The covariance matrix is block diagonal because we do not include common correlated processes between pulsars. However, we do include a common red noise process (except for the ranking) because it was found in the IPTA DR2 with good statistical significance (Antoniadis 2022). To summarize, the covariance matrix $\mathbf{C}_\alpha$ for the noise model of one pulsar is :

$$\mathbf{C}_\alpha = \mathbf{N}_\alpha + \mathbf{C}_{TM,\alpha} + \mathbf{C}_{DM,\alpha} + \mathbf{C}_{RN,\alpha} + \mathbf{C}_{CRN}, \tag{3.1}$$

with $\mathbf{N}_\alpha$ the uncorrelated white noise, $\mathbf{C}_{TM,\alpha}$ the term associated with the marginalized timing model, $\mathbf{C}_{DM,\alpha}$ the dispersion measure noise, $\mathbf{C}_{RN,\alpha}$ the individual red noise and $\mathbf{C}_{CRN}$ the common red noise.

3.2.2 The signal model

For the signal model, we consider monochromatic CGW signal as presented in section 1.4.4. The timing residuals expected in the PTA data for a CGW signal coming from a circular SMBHB:

$$s_a(t,\hat{\Omega}) = -\sum_A F^A(\hat{\Omega})[s_A(t) - s_A(t-\tau_a)] \tag{3.2}$$

with $\tau_\alpha = t - t_{p,\alpha}$ the delay between Earth time t and pulsar time $t_{p,\alpha}$ and :

$$\begin{aligned} s_+(t) = \frac{\mathcal{M}^{5/3}}{d_L\omega(t)^{1/3}} & \left[\sin[2\Phi(t)](1+\cos^2\iota)\cos 2\psi\right. \\ & \left. +2\cos[2\Phi(t)]\cos\iota\sin 2\psi\right], \end{aligned} \tag{3.3}$$

$$\begin{aligned} s_\times(t) = \frac{\mathcal{M}^{5/3}}{d_L\omega(t)^{1/3}} & \left[\sin[2\Phi(t)](1+\cos^2\iota)\cos 2\psi\right. \\ & \left. -2\cos[2\Phi(t)]\cos\iota\sin 2\psi\right], \end{aligned} \tag{3.4}$$

where $\mathcal{M}$ is the chirp mass, d_L the luminosity distance, $\omega(t)$ the CGW orbital angular frequency, ι is the orbital inclination to the line of sight, ψ is a polarization angle and $\Phi(t)$ is the phase of CGW.

The F^A are the antenna pattern functions (Babak and Sesana 2012; Sesana and Vecchio 2010; Ellis et al. 2012; Taylor et al. 2016) given as

$$F^+(\hat{\Omega}) = \frac{1}{2}\frac{(\hat{m}\cdot\hat{p})^2 - (\hat{n}\cdot\hat{p})^2}{1+\hat{\Omega}\cdot\hat{p}}, \tag{3.5}$$

$$F^\times(\hat{\Omega}) = \frac{(\hat{m}\cdot\hat{p})(\hat{n}\cdot\hat{p})^2}{1+\hat{\Omega}\cdot\hat{p}}. \tag{3.6}$$

In this work[1], we neglect the pulsar term considering it as a part of the noise, assuming that the source has evolved sufficiently over τ_α to move the pulsar term off the earth-term frequency. Including pulsar term should improve the parameter estimation but comes with a huge price of the increase in the complexity of the likelihood surface and the dimensionality of parameter space (2 additional parameters per pulsar for phase and frequency of pulsar term, e.g. see Corbin and Cornish 2010). We foresee the possibility of following up the candidate events (identified using the earth term only) with the extended signal model (pulsar term, eccentric orbit) on the reduced parameter space. We also neglect the evolution of the GW frequency ($\omega_o = 2\pi f_{gw}$) over the observation time. The frequency evolution becomes potentially measurable for the heavy sources emitting at frequency $\geq 10^{-7}$Hz, neglecting the frequency evolution does not prevent us from detecting the sources but introduces a bias in the measured GW frequency (overestimating

[1] Based on the submitted IPTA collaboration paper where I am corresponding author.

CGW parameter	Range
$\log_{10} h$	[-18, -11]
f_{gw} (Hz)	$[10^{-9}, 10^{-7}]$
ϕ_0	$[0, 2\pi]$
$\cos\iota$	[-1, 1]
ψ	$[0, \pi]$
θ	$[0, \pi]$
ϕ	$[0, 2\pi]$

Table 3.1: List of the CGW parameters as defined in our model with their respected ranges.

it), for more details see conclusion in Petiteau et al. 2013. So the CGW phase takes a very simple form:

$$\Phi(t) = \omega_o t + \phi_0/2, \tag{3.7}$$

where ϕ_0 is initial orbital phase. Finally, the CGW amplitude h is a function of $\mathcal{M}$, d_L and f_{gw} given by

$$h = \frac{2\mathcal{M}^{5/3}(\pi f_{gw})^{-2/3}}{d_L}. \tag{3.8}$$

We consider the model containing only one CGW signal. This model still detects multiple CGW if they are present in the data at the non-overlapping Fourier frequencies (see Babak and Sesana 2012 for discussion). If we find more than one candidate with sufficient statistical significance as potential GW sources, we will conduct additional investigations extending our model to include several CGWs. We start the analysis with 1 CGW source characterized by 7 parameters summarized in table 3.1 together with their prior range (we always assume a uniform prior[2]).

3.3 Ranking the pulsars

Within the pulsars that make up the PTA, some may be more sensitive to the presence of GW signal, depending on their noise properties and time of observation. To characterize the contribution of individual pulsars we can evaluate the signal to noise ratios (SNR) separately for each pulsar after injecting fake GW signal in their TOAs. It is interesting to rank them according to their SNR to see which contribute the most to the sensitivity of the array. We eventually discard the worst pulsars to lighten the data and the computational cost without significant loss in detectability.

[2]For setting an upper limit we use uniform prior on the amplitude of GW strain

3.3.1 Ranking method

In L. Speri 2022[3] we present two methods (Chimera and SNR_B maximization, see reference and appendix 6.5 for details about the methods) to rank pulsars with respect to GWB, which is a stochastic and correlated signal. CGW are deterministic signals and their analysis has been treated separately from the stochastic GWB. CGWs are included in the model as a periodic delay applied to the timing residuals $\vec{\delta t}$ while the effect of the GWB is included in the covariance matrix $\mathbf{C}$ of the likelihood. This fundamental difference between the two signals and their mathematical description calls for a different ranking method.

Here, we want to rank pulsars according to their response to a CGW signal. One way to proceed is to inject a large number of fake CGW signals with randomized parameters except for fixed frequency and amplitude (Babak 2016). Then, for each pulsar, the CGW signal-to-noise ratio is computed for each injection and averaged numerically. In this way, we have the average response of each individual pulsar in the array at a given frequency of the CGW signal. This averaging can also be done analytically, as shown in the following paragraph. Note that we refer to the signal-to-noise ratio of CGWs using the acronym SNR. However, we use the symbol ρ to distinguish the SNR of CGWs from the previously defined SNRs.

In the likelihood of Eq.(2.21), the inclusion of a deterministic signal is performed by changing the timing residuals as $\delta\vec{t} \to \delta\vec{t} - \vec{s}(\vec{\theta})$ where $\vec{s}(\vec{\theta})$ is the signal template we aim to measure. In that case, the likelihood can be rewritten as:

$$\ln\mathcal{L} = -\frac{1}{2}[(\delta\vec{t}|\delta\vec{t}) + (\vec{s}|\vec{s}) - 2(\delta\vec{t}|\vec{s}) + \mathrm{Tr}\ln 2\pi\mathbf{C}]\,, \tag{3.9}$$

where we have introduced the noise weighted inner product $(\vec{x}|\vec{y}) = x^T\mathbf{C}^{-1}y$.

We can now calculate this expression for the hypothesis of the presence of a CGW ($\mathcal{H}_1$) versus its absence ($\mathcal{H}_2$) (see section 2.1.2). The expectation value of the log-likelihood ratio becomes:

$$\langle\ln\Lambda\rangle_{\mathcal{H}_1} = \left\langle \ln\left(\frac{p(\delta\vec{t}|\vec{s})}{p(\delta\vec{t}|\vec{0})}\right)\right\rangle_{\mathcal{H}_1} = \langle(\delta\vec{t}|\vec{s}) - \frac{1}{2}(\vec{s}|\vec{s})\rangle_{\mathcal{H}_1} = \frac{1}{2}(\vec{s}|\vec{s})\,, \tag{3.10}$$

where $\rho_{\mathrm{Opt}} = \sqrt{(\vec{s}|\vec{s})}$ is the optimal SNR for the CGW source.

Since the source parameters are not known a priori, we average ρ^2_{Opt} over gravitational wave polarization ψ, initial phase ϕ_0, inclination ι, and sky location (θ,ϕ). To do so, we analytically compute the integral over the defined bounds of the CGW parameters:

$$\rho^2 = \int_0^{\pi}\frac{d\psi}{\pi}\int_0^{2\pi}\frac{d\phi_0}{2\pi}\int_1^{-1}\frac{d\cos\iota}{2}\int_1^{-1}\frac{d\cos\theta}{2}\int_0^{2\pi}\frac{d\phi}{2\pi}(\vec{s}|\vec{s})\,. \tag{3.11}$$

[3] I am co-author of the paper for which I worked on the ranking method for CGW signals.

Using the formula for a CGW signal from a circular SMBHB, $\vec{s}(t,\Omega)$, as presented in section 1.4.4, the SNR^2 averaged over CGW parameters takes this simple form (see appendix for calculus development):

$$\rho^2(h,f) = \frac{4}{15}\Big(\frac{h}{2\pi f}\Big)^2 \Big[(\cos 2\pi f t|\cos 2\pi f t) + (\sin 2\pi f t|\sin 2\pi f t)\Big], \tag{3.12}$$

with

$$h = \frac{2\mathcal{M}^{5/3}(\pi f)^{2/3}}{d_L}, \tag{3.13}$$

where f and h are the gravitational wave frequency and amplitude, $\mathcal{M}$ is the chirp mass and d_L is the luminosity distance. For pulsar a, we evaluate ρ_a^2 at the TOAs $\vec{t}_a$. We consider an Earth-term SNR because the inclusion of the pulsar term requires assuming a chirp mass and having a good knowledge of pulsar distance to estimate frequency evolution. Furthermore, the pulsar term will only promote pulsars at large distances with typically low SNRs (large-distance pulsars typically have a worse timing and their radio signal is weaker, leading to a degradation in the distance estimation). We are interested in the relative SNR contribution of pulsars at fixed signal frequency thus including pulsar term constitutes additional complexity that is beyond our purposes. If pulsar term and earth term are at same frequency (but different phase), their averaging yields twice the same average SNR^2 therefore leaving the relative contribution unchanged. The correlated noises (e.g. intrinsic and dispersion measure noises) are taken into account in the covariance matrix $\mathbf{C}$ of the noise-weighted inner product of the cosine and sine terms.

Common (correlated) processes were not included in our noise model, so the covariance matrix is block-diagonal. In this way, the likelihood can be factorized and SNR^{2}s can be computed independently for each pulsar. In practice, we should incorporate the common processes in the noise model, but this adds another level of complexity that is irrelevant for the goal of the selection procedure[4]. The ultimate goal is identification of the best pulsars for CGW detection, and therefore, only the intrinsic properties of the pulsars were considered.

We estimate the relative contribution of one pulsar to the total SNR of the array using the normalized SNR^2:

$$\bar{\rho}_a^2(f) = \frac{\rho_a^2(h,f)}{\sum_b \rho_b^2(h,f)}, \tag{3.14}$$

Note that the amplitude h cancels out in this expression and the CGW frequency f remains the only parameter. Therefore we can fix h to any value without affecting the ranking.

We construct the cumulative sum of the normalized SNR^{2}s of the pulsars ranked from best to worst. We fix a threshold value for the SNR^2 cumulative sum above which pulsar

[4]Furthermore, detectable CGW signals must be louder than the GWB. Since the GWB is stronger at lower frequencies, CGW signals are more likely to be found at high frequencies.

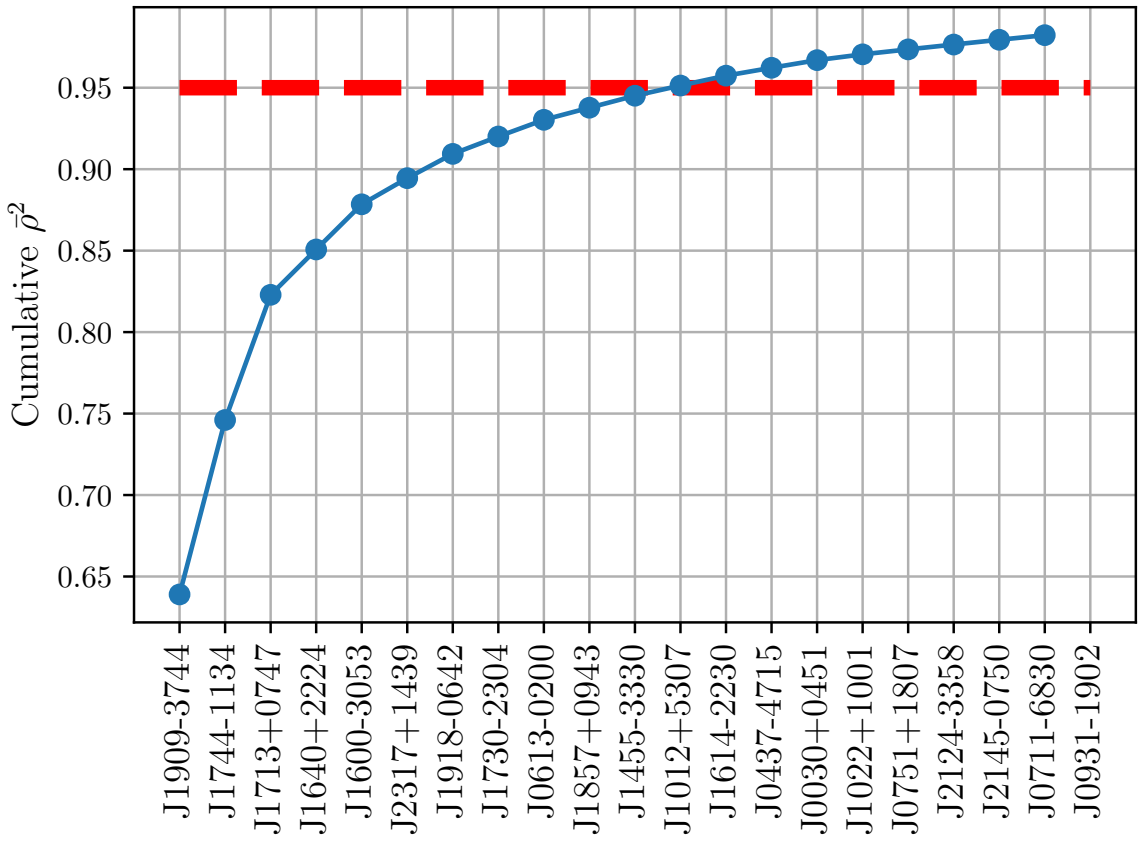

Figure 3.1: Cumulative $\bar{\rho}^2$ plot for the pulsars in the IPTA DR2 at CGW frequency of 5nHz. The pulsars above the red dashed line contribute less than 5% of the total SNR^2. This means only 12 pulsars out of 65 contribute on average to 95% of the total SNR^2 of the array at 5nHz. Note that, while only the best 22 pulsars in the array are shown in the figure, the normalized total SNR has been evaluated using all 65 pulsars in the array.

contributions to the total SNR^2 are not considered significant. This value was chosen to be 0.95. The process is illustrated in Figure 3.1 for pulsars from the IPTA second data release (Perera 2019).

Due to the strong dependence of $\bar{\rho}_a^2(f)$ on f, the resultant CGW pulsar ranking is also frequency dependent. This can be clearly seen from Figure 3.2. In our analysis, we use 100 log-spaced frequency bins between 10^{-9} and 10^{-7} Hz. Ranking lists were obtained separately for each frequency bin. In order to construct the final ranking catalog of best pulsars at a given frequency range, the lists at each frequency are merged together. This procedure ensures that we will gain at least, no matter the CGW frequency, 95% of the total SNR^2 of the array.

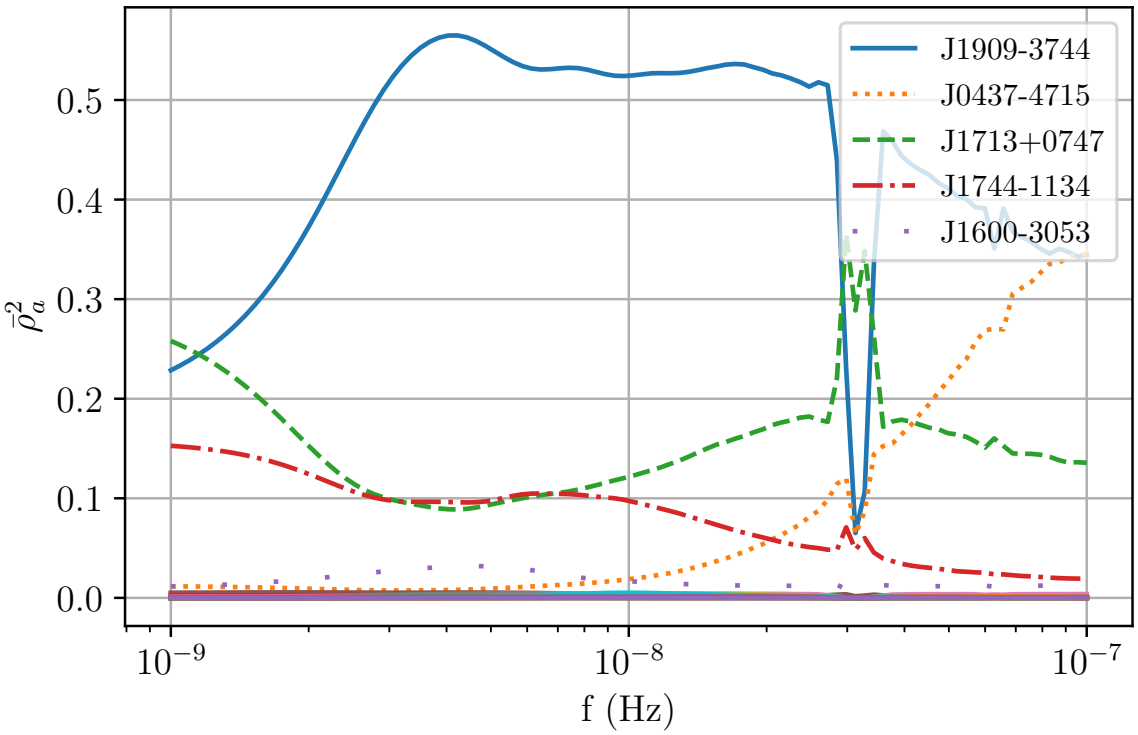

Figure 3.2: Normailzed $\bar{\rho}_a^2$ of the five best pulsars of the IPTA DR2, at different CGW frequencies. The glitches at the right of the plots are due to the one year and half-year peaks.

3.3.2 Ranking results

We now test the performance of the CGW ranking method using noise-parameter values previously extracted from individual pulsar noise analyses of the latest IPTA data release (Perera 2019) and a "realistic" simulated EPTA dataset consisting in 40 pulsars with same RMS and sky locations as real EPTA dataset (Desvignes 2016) but with white noise only, ignoring intrinsic red noise and dispersion measure noise, and injected GWB (L. Speri 2022).

Because the ranking method is based on an exact noise-averaged formula, it is unnecessary to simulate noise realizations to test its performance. However, we still want to prove that the selected pulsars recover most of the total SNR in the presence of a true (i.e. non-averaged) signal. We test this by comparing the fraction of total SNR^2 obtained using the CGW ranked pulsars to that obtained from a random pulsar selection. For an array of N pulsars, the fraction of total SNR^2, given a list of $M < N$ pulsars, is defined as:

$$\rho_M^2 = \sum_{a=1}^{M} \bar{\rho}_a^2, \qquad \text{with } 0 < \rho_M^2 < 1, \tag{3.15}$$

where $\bar{\rho}_a^2$ is the normalized SNR^2 defined in Eq (3.14).

After extracting the list of best pulsars, we test the selection procedure as follows:

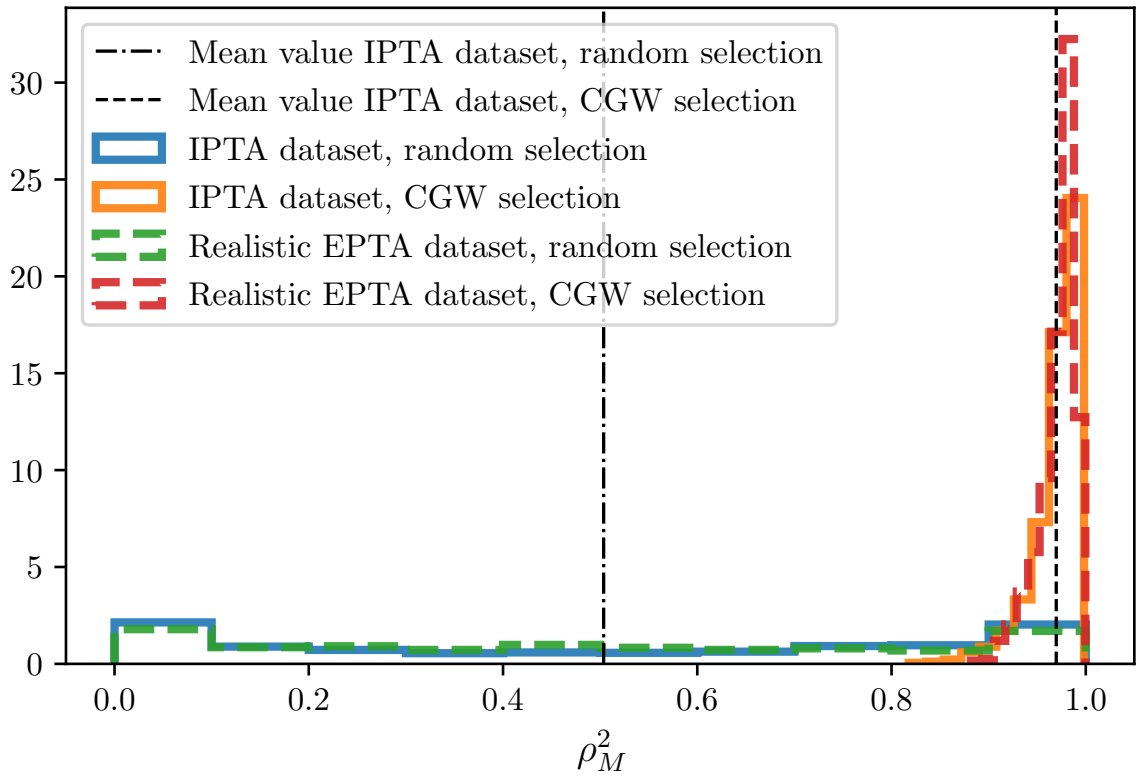

Figure 3.3: Distribution of the SNR^2 coverage for 1000 different sets of CGW parameters. The distributions are obtained with the list of best pulsars and random selection for a real IPTA dataset and the realistic EPTA dataset.

- We draw the CGW signal parameters $\vec{\theta}$ from a uniform distribution with bounds defined as in the integral of Eq. (3.11), and with frequency between 1 and 100 nHz. As pointed out in section 3.3.1 the strain amplitude has no influence on the ranking and therefore we fix it to $h = 10^{-14}$.

- We compute the non-averaged optimal SNR $\rho_{\text{Opt}} = \sqrt{(\vec{s}|\vec{s})}$ for each pulsar for a CGW signal $\vec{s}(t,\vec{\theta})$ and we use this quantity to calculate the normalized $\bar{\rho}_a^2$ defined in Eq. (3.14).

- We compute $\rho^2_{M-\text{CGW}}$ for the list of best selected pulsars and $\rho^2_{M-\text{rand}}$ for a random subset of pulsars of random size M.

- We repeat the previous steps one thousand times.

This gives us 1000 values of $\rho^2_{M-\text{CGW}}$ and $\rho^2_{M-\text{rand}}$ that we plot as histograms on Figure 3.3. For the IPTA dataset, the distribution of fractional $\rho^2_{M-\text{CGW}}$ for the selected pulsars is narrowly peaked around a mean value 0.97. The random selection $\rho^2_{M-\text{rand}}$ gives an almost uniform distribution with 0.50 mean value. The distribution is not uniform because ρ_a^2 is not uniform and a few ρ_a^2 values are much bigger while many others are very small. Similar results are obtained for the realistic EPTA dataset. We find that the number of pulsars which gives 95% of the SNR^2 is 22 for both datasets.

Now we briefly discuss the comparison between the CGW and GWB selection methods. Focusing on the realistic EPTA dataset, we find an overlap between the identified best pulsars with the CGW method and GWB method as shown in Table 3.2. This time we run the Chimera and SNR$_\text{B}$-maximization ranking for brief explanation of the methods) without fixing the six initial pulsars of the EPTA. We find that 17 pulsars are common to all three selection methods (highlighted in bold).

In summary, when true CGW signals are injected in the data, the CGW ranking method selects the pulsars which provides most of the SNR of the array, whereas a random selection is inefficient. This method extracts the few best pulsars to optimize the search for a CGW signal.

3.3.3 Sensitivity curve

CGW strain amplitude for $\rho = 1$

It is possible to plot the PTA sensitivity to CGW signal using the SNR formula in equation 3.12. Indeed, the total SNR^2 of the array ρ^2 is the quadratic sum of the individual SNRs ρ_a^2 of pulsars. We can invert the expression 3.12 to express the amplitude h of the CGW as a function of ρ^2

CGW ranking	Chimera method	SNR_B maximization
J0030+0451	**J0030+0451**	**J0030+0451**
J0613−0200	J0034−0534	**J0613−0200**
J0751+1807	**J0613−0200**	J0621+1002
J1012+5307	J0621+1002	**J0751+1807**
J1022+1001	**J0751+1807**	J1022+1001
J1024−0719	J1012+5307	**J1024−0719**
J1600−3053	**J1024−0719**	**J1600−3053**
J1640+2224	J1455−3330	**J1640+2224**
J1713+0747	**J1600−3053**	**J1713+0747**
J1730−2304	**J1640+2224**	**J1730−2304**
J1744−1134	**J1713+0747**	**J1744−1134**
J1751−2857	**J1730−2304**	**J1751−2857**
J1804−2717	**J1744−1134**	J1801−1417
J1853+1303	**J1751−2857**	**J1804−2717**
J1857+0943	J1801−1417	J1843−1113
J1909−3744	**J1804−2717**	J1853+1303
J1910+1256	J1843−1113	**J1857+0943**
J1911+1347	**J1857+0943**	**J1909−3744**
J1918−0642	**J1909−3744**	**J1910+1256**
J2010−1323	**J1910+1256**	J1911+1347
J2124−3358	J1911−1114	J1911−1114
J2145−0750	**J1918−0642**	**J1918−0642**
	J2010−1323	**J2010−1323**
	J2124−3358	**J2124−3358**
	J2322+2057	J2322+2057

Table 3.2: List of the first 22 pulsars selected with the CGW ranking method and the 25 pulsars selected with the Chimera method and SNR_B-maximization in the realistic EPTA dataset. Bold font indicate the 17 pulsars that are selected by all three methods described in L. Speri 2022.

$$\rho^2(h,f) = \sum_a \rho_a^2(h,f), \tag{3.16}$$

$$h(f) = \frac{\sqrt{15}\pi f \rho(f)}{\sqrt{(\cos 2\pi f t|\cos 2\pi f t) + (\sin 2\pi f t|\sin 2\pi f t)}}, \tag{3.17}$$

giving h as a function of the CGW frequency f and the noise parameters included in the covariance matrix of the noise weighted inner products.

By setting ρ to be equal to 1 we obtain the value of CGW amplitude corresponding to the threshold above which the signal starts to become louder than the noise[5]. This dependence on the noise becomes obvious when we plot the sensitivity curve : the threshold amplitude h as a function of frequency f.

Effect of noises on the sensitivity

Naturally, we expect the chosen noise model to have a strong influence on the sensitivity of the array. This is a major concern when dealing with PTA data because we cannot characterize the noises present in the array beforehand, we have to fit for them. A wrong modelization of the noises could lead to fake detections.

In Bayesian analysis, we treat noise parameters as random variables with probability distributions. In equation 3.17 we explicitly see the dependence of h on noise model parameters present in the noise weighted inner products (denominator of the expression), meaning that the threshold amplitude of detection should itself have a probability distribution. We propose to test two things :

- Study the influence of different values of noise parameters for the same model,
- Examine the influence of the size of the Fourier basis for the noises modelled as Gaussian processes (RN and DM, see section 2.2.2).

For the first point, we use previous MCMC noise analyses that were performed on individual pulsars to recover the posterior distributions of the noise parameters. We produce a 100 different sensitivity curves using 100 different samples randomly drawn from the posteriors of the noise parameters. We use the noise model presented in section 2.2.2 with a timing model, white noise, red noise and dispersion measure noise. Figure 3.4 shows that amplitude $h(f)$ now becomes a broad distribution. It appears less constrained at low frequencies due to the stronger presence of RN and DM noise. The sensitivity we obtain with fixed maximum likelihood value of noise parameters seems to give a better sensitivity than the average distributions of h around 10nHz. Therefore, in the case of PTA, taking into account the variability of the noise parameters has a direct influence on the detectability of a signal.

[5]Here, the SNR is for a single frequency signal.

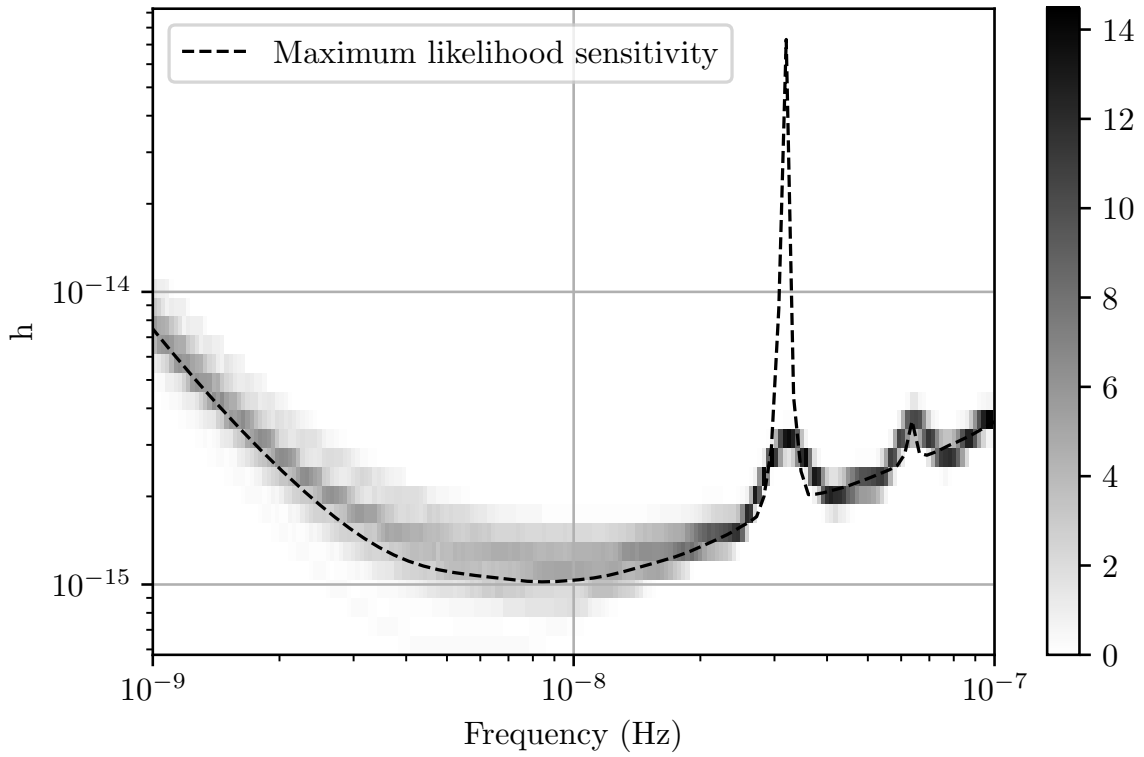

Figure 3.4: Binned sensitivity curve for 100 different realization of noise parameters. The grey colour scale represents the binning of h at a given frequency. The dashed line represents the sensitivity curve for maximum likelihood values of parameters.

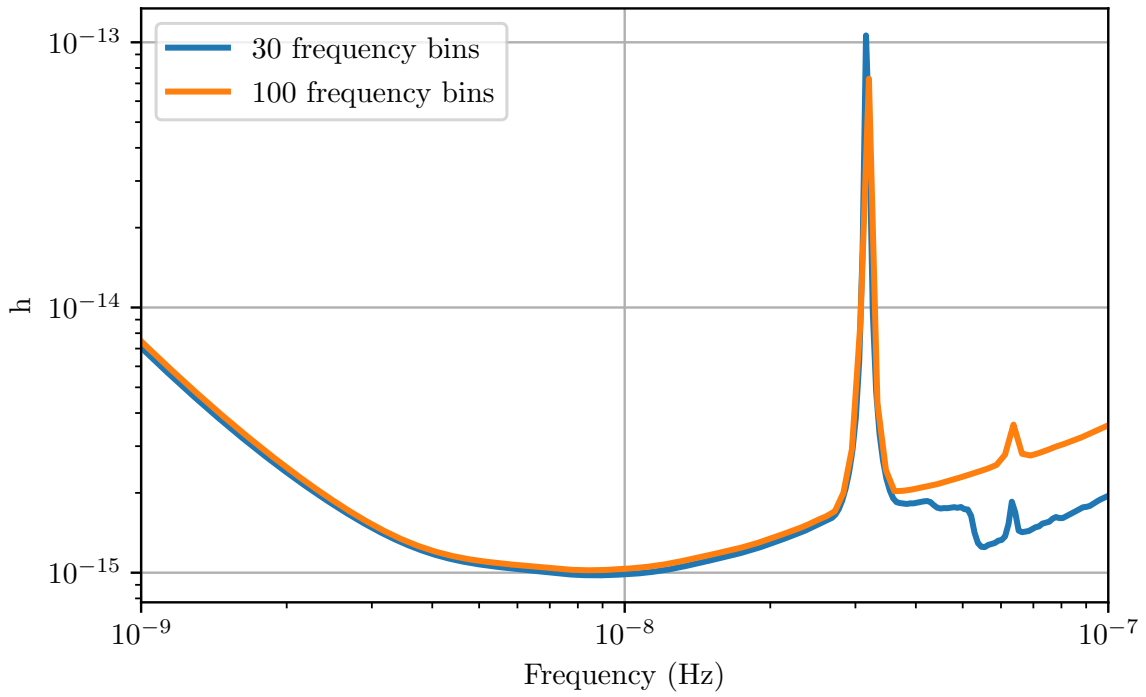

Figure 3.5: Effect of the Fourier basis size for red noise of the sensitivity curves. We compare two models using 30 and 100 frequency bins for RN and DM Fourier basis. Using different numbers of frequency bins can have an impact on the noise modelling at high frequencies because it moves the frequency after which only white noise is present in the array.

For the second point, we plot two different sensitivity curves with same noise model but different number of frequency bins (30 and 100) for RN and DM with maximum likelihood values of parameters. We clearly see on figure 3.5 that using a short basis has an impact on the predicted sensitivity at high frequencies. In section 2.2.2 we explained that for one pulsar the Fourier basis is made of the N harmonic frequencies of $1/T$, T being the total time of observation for that pulsar. Logically, choosing a N that is too small will brutally turn off the effect of the noise after the frequency N/T, producing a "step" in the sensitivity that we see on figure 3.5. This effect will be even more pronounced for red noises with low spectral index. For very steep RN, the white noise rapidly takes over and the step will not appear. All things considered, having a short Fourier basis will not have a strong impact on signal searches at low frequency but it is crucial at higher frequencies.

3.4 Method

We work within the Bayesian framework and start with running the search for the CGW signal. As mentioned above, we sample parameters of CGW together with the noise pa-

rameters for RN and DM. We keep the white noise parameters fixed and marginalize over the timing model and BAYESEPHEM parameters. We made two runs: with and without CRN, to check how much it affects our result. We always use Markov chain Monte-Carlo (MCMC) sampler (Ellis and Haasteren 2017), and we use ENTERPRISE (Ellis et al. 2020) software to construct the models and compute the likelihood and prior probability.

We extensively use the single pulsar noise explorations runs performed before the main analysis. We have converted posteriors for the RN and DM into 2D histograms ($\{A_{RN,DM}, \gamma_{RN,DM}\}$) and use them as one of a proposal for those parameters. This empirical proposal improves the efficiency of MCMC and reduces the autocorrelation length of the chain (Aggarwal 2019).

During the search, we compute the Bayes factor (BF), comparing the null model (noise only) against the model where we have a CGW signal on the logarithmically spaced frequencies. In the absence of the detection, we proceed to setting an upper limit, building a 95% sensitivity curve. During the upper limit analysis, we used a uniform prior on the amplitude of the GW signal.

IPTA data contains 53-millisecond pulsars; however, not all of them are equally sensitive to the CGW. We have selected 21 pulsars which, on average, recover 95% of the array's total signal-to-noise ratio (SNR) to CGW. The selected pulsars are depicted on the projected sky map in Figure 3.6 as red stars; we have used large green stars and annotation for the four best timers. The ranking method is briefly outlined in Babak 2016 and in greater detail in L. Speri 2022. This significantly reduces the computational cost without much affecting the final results.

Historically we have performed the search with the noise model with the uniform settings across all pulsars in the array; namely, we have used 30 frequency bins for modelling both RN and DM processes (as it was done in preceding work Arzoumanian et al. 2014; Aggarwal 2019). Using this uniform setting we have obtained quite erroneous results, and after a long investigation, we realised that this noise model does not adequately describe the observational data (see section 3.6). We have switched to another noise model where we have used the custom-made noise model for the six best EPTA pulsars (see Chalumeau 2021 for details), and for other pulsars, we changed the number of used frequency bins: *RN30DM100* – 30 bins for the RN and 100 bins for DM modelling. We will continue using this short-hand notation for the noise model, showing explicitly the number of Fourier frequencies used by the base functions in the Gaussian process describing the corresponding noise. The results presented in the next section were obtained using the "custom" made noise model; we postpone the detailed discussion on the noise model selection and influence until Section 3.6.

3.4.1 Model selection

During the search, we consider two competing models: noise only and noise with 1 CGW signal. We compute the Bayes factor (BF) to measure which model is preferred by the IPTA DR2 dataset. In particular we employ hyper-model jumps to compute BF following the methods outlined in Hee et al. 2015 and implemented in the `enterprise extensions` (Taylor et al. 2021). In this method, we introduce a hyper-parameter which indexes the models; then, we perform sampling inside each model and in this parameter. The BF then is a fraction of the samples the chain spends in each model. For example, if we consider two models $\mathcal{M}_A$ and $\mathcal{M}_B$ with the hyper-parameter n, then the BF is the ratio (see section)

$$\mathcal{B}^A_B = \frac{n_A}{n_B} \tag{3.18}$$

where n_A, n_B counts the number of samples in the chain corresponding to the models A, B. In our previous investigations, we have found that this method gives quite a reliable result, e.g. see Chalumeau 2021 where comparison is done against the evidence computation obtained with `Dynesty` nested sampling.

Later in this work we will compute BF between the models where we also vary the noise.

3.4.2 Upper limit

For obtaining the upper limit, we again use MCMC assuming a uniform prior distribution for $\log h$ within the bounds $[-18, -11]$ on the fixed set of f_{gw}. We use marginalized posterior probability distribution for the CGW amplitude $p(h)$ to set 95% upper limit h_{95} defined as :

$$0.95 = \int_0^{h_{95}} p(h)\,dh. \tag{3.19}$$

We use a grid of 100 logarithmically spaced frequencies between 10^{-9} and 10^{-7} Hz. The lowest bound determined by the IPTA observational span $1/T_{obs}$ that is $\approx 10^{-9}$ Hz while the upper bound, 10^{-7} Hz, is constrained by reduction in the sensitivity due to response ($\propto f^{-1/3}$) and by our assumption that GW frequency does not evolve. We should take into account the frequency evolution of the GW signal above 10^{-7} Hz during the analysis, as discussed in Petiteau et al. 2013. As mentioned above, we compute the upper limit for two models of noise; with and without CRN. The inclusion of CRN adds two more parameters to the model (an amplitude parameter A_{CRN} and a spectral index γ_{CRN}), which we sample together with parameters of CGW and other noise components.

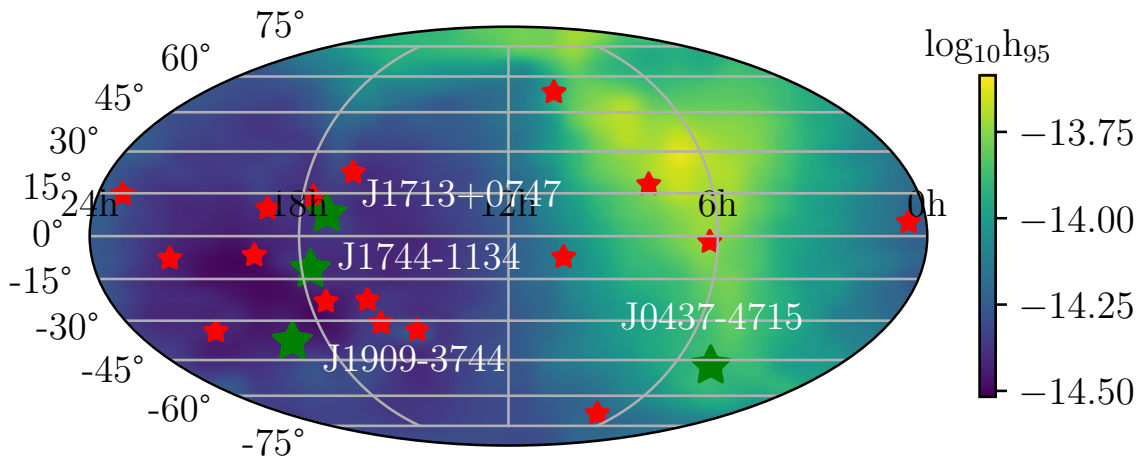

Figure 3.6: 2D sky sensitivity of IPTA for CGW signal around most sensitive frequency. The figure was obtained computing the 95% upper limit for h on 12 patches across the sky. Gaussian interpolation was used to smoothen it across boundaries.

3.5 Results

We compute BF between two models $\mathcal{B}^{custom+CGW}_{custom}$ for a noise only model $\mathcal{M}_{custom}$ and noise+CGW signal model $\mathcal{M}_{custom+CGW}$. The subscript "custom" corresponds to the noise model we have used for these main results and distinguishes it from other noise models considered in the next section. Similarly to the upper limit run, we have used a log-uniform prior on $\log h$ bounded by [-18, -11], and the BF was computed on the grid of 100 CGW frequencies (f_{gw}) between 1 and 100 nHz.

The main result of this work is summarized in Figure 3.7. In the lower panel we plot the BF, and, as one can see, the noise model is usually preferred. There are few spikes where BF reaches 1, this is definitely not a detection, however PTA efforts probably should monitor carefully those frequencies in the future extended IPTA datasets. The excess in BF was also used as identification of the CGW candidates to follow. These results show that there is nothing to follow[6]: $\mathcal{B}^{custom+CGW}_{custom} \leq 1$.

The top panel of Figure 3.7 shows the upper bound of the 95% central credible region for CGW strain computed at the same set of frequencies. Note the "spiky" features at several frequencies (8.1, 13, 16) nHz corresponding to the outlier indicating potential candidates of CGW. However the corresponding BFs are (0.95, 0.33, 0.34) indicating no statistical significance in the analyzed data.

We have computed the upper limit using uniform in GW strain prior with and without CRN on the fine frequency grid; the results are present in Figure 3.9. The upper limit slightly worsens at low frequencies when we add the CRN; this is understandable as we

[6]Note that the highest spike corresponds to frequencies close to f_{yr} and should be discarded.

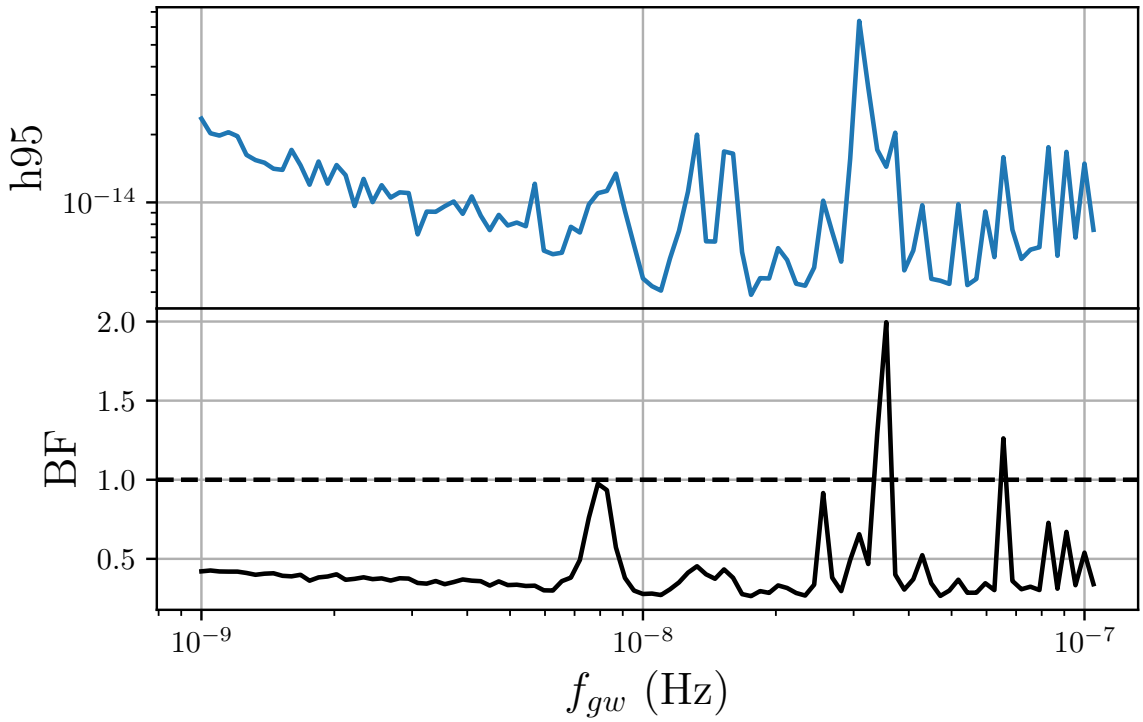

Figure 3.7: On the top panel, the upper bound of the 95% central credible region on $\log h$ obtained with a log-uniform prior (with CRN). Bottom panel is the associated BF at given CGW frequency. The black dashed line shows where BF is equal to 1. The two peaks around 35 and 70 nHz with BF>1 are the 1yr and 2yr and should not be taken into account.

need a higher amplitude of CGW to get the same SNR when raising the noise floor. The most sensitive frequency of IPTA DR2 is 10.2 nHz with $h_{95} = 9.1 \times 10^{-15}$ in both cases with and without CRN component.

In the model that includes CRN, we are allowing slope and amplitude to vary during the sampling. The recovered posterior for CRN with CGW at 1nHz (blue) is compared to the posterior obtained in Antoniadis 2022 (orange) in figure 3.8. We observe that the amplitude of CRN in the model which does not include CGW (Antoniadis 2022) is slightly higher while the slope is almost the same. This could be explained by a partial absorption of the CRN into CGW. At the same time adding the CRN to the model increases the overall noise level (therefore decreasing the SNR of the signal). The interplay between CRN and CGW appears to mitigate the effect of the CRN on the upper limits in figure 3.8.

We have also overplotted the best CGW upper limit available to date based on the analysis of the NANOGRAV 11 years dataset (Aggarwal 2019) as a dashed (green) line. Note that only the nine-year NANOGRAV data set was included in IPTA DR2. As expected, our current results are better at very low frequencies thanks to the longer observation time. Extended sky coverage, improved effective cadence of observation thanks to overlapping timing data (gaps coverage) and the addition of pulsars like J0437-4715 (only present in PPTA data Manchester 2013a), which is an excellent timer, is reflected in a much improved upper limit at 100 nHz, where we might expect the first detection of CGWs.

We want to point at the double-peak feature just above 10nHz (see Figure 3.7). This peak remains the same under the prior change (from the uniform in the amplitude to the uniform in the log-amplitude), which often corresponds to a signal present in the data. The Bayes factors for those peaks are low (0.33, 0.34). Nonetheless, we should keep an eye on those frequencies in the next data releases.

3.6 Effect of noise modeling on the CGW search

In this subsection, we consider several noise models and compute BF between those models of noise with and without CGW. The main results presented in the previous section were obtained with the custom-made noise model for the best EPTA pulsars (Chalumeau 2021) and with RN30DM100 choice of Fourier frequencies for other pulsars. The custom noise model modifies the number of Fourier frequencies as given in table 3.3 and includes additional noise components (like system noise). However, the Fourier basis for CRN is always fixed at 30 frequency bins.

Here we present results with what was considered "standard" noise settings before this work, namely RN30DM30 model. We have started analysis using this model, and the search quickly converged to a particular sky position at 51 nHz. The first peculiarity of these results is that 51 nHz is very close to Venus orbital frequency, and the second is

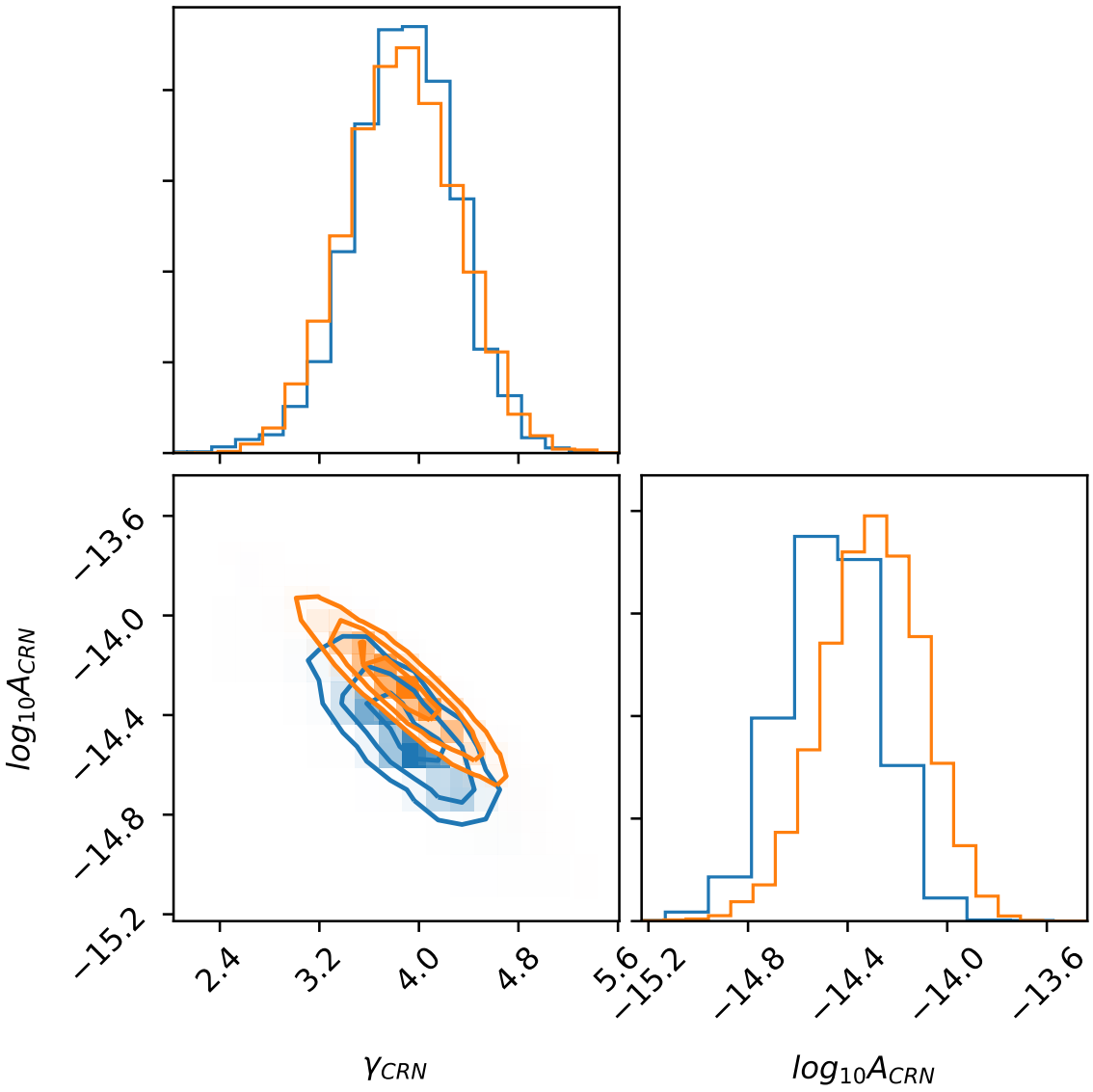

Figure 3.8: CRN parameters for IPTA DR2 without CGW signal obtained in (orange) and with CGW signal at 1 nHz (blue).

Pulsar name	RN bins	DM bins
J0613-0200	10	30
J1012+5307	150	30
J1713+0747	15	150
J1744-1134	10	100
J1909-3744	10	100
J0437-4715	30	100
Other pulsars	30	100

Table 3.3: Number of frequency bins used for individual RN and DM noise

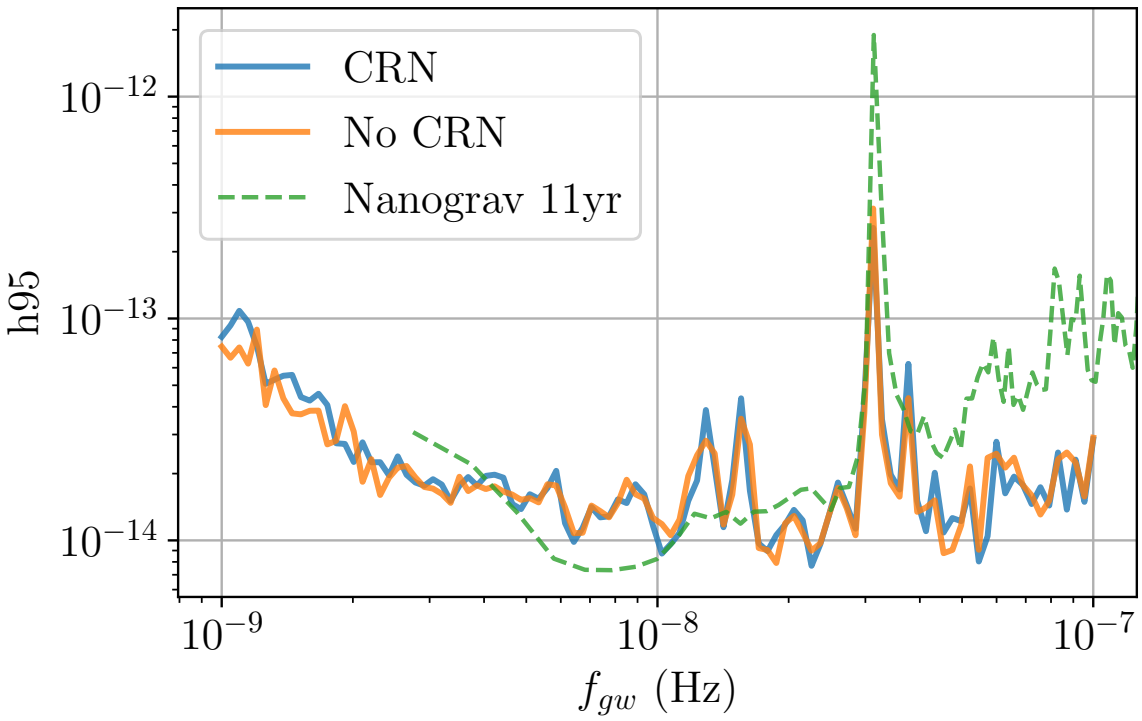

Figure 3.9: 95% sky averaged upper limit h_{95} on CGW amplitude for models with and without CRN

that the sky position had a bi-modal structure and was located very close to J1012+5307 see Figure 3.10. The Bayes factor for this event with the RN30DM30 noise model was $\mathcal{B}^{RN30\_DM30+CGW}_{RN30\_DM30} = 18$, which is not very high; however, it seriously triggered our attention by being relatively well constrained in the parameter space.

We have launched a set of investigations trying to understand this event. Using samples taken from the candidate's posterior, we have checked the contribution to the SNR from each pulsar. It came out that the main part comes from J1012+5307, but a few other pulsars (J1713+0747 and J0437-4715) also contributed not negligibly. We have checked that the zero contribution from a very good timer J1909-3744 is expected given the presumed sky position of the event.

We also conducted a set of injections of CGW signals with parameters taken from the candidate's posteriors using a simulated IPTA data (same TOAs as the real) with white noise only (RN and DM are supposed to be sub-dominant at the candidate frequency). We could not reproduce the observed results with injections even when we increased the amplitude of the simulated signal.

BAYESEPHEM does not include a contribution from the inner planets to the phenomenological model, so we consulted a group from Observatoire de la Côte d'Azur (INPOP group), inquiring if there could be an error in the Venus orbit picked up in the CGW analysis. We were reassured that the Venus parameters are known with very high preci-

sion: this is a simple planet without any moons. However, looking into the future, we probably should extend BAYESEPHEM (or alternative mitigation models) by including the perturbation of orbital elements for the inner planets. We have performed the analysis with a narrow prior around this event with the model including the pulsar term. The Bayes factor has slightly increased $BF = 21$; however, the parameter estimation did not change appreciably. We have tried several runs with an extended model that includes the orbital eccentricity (again with a narrow prior). The eccentric runs show a very poor convergence, but all of them suggest a relatively high eccentricity. Results of both models (eccentric and with pulsar term) imply that the power is not localized at one particular frequency but spread over some finite-size frequency band.

We have turned to several noise models. We have started with varying number of Fourier frequencies used in the Gaussian process for RN and DM and tried $\mathcal{M}_{RN30DM100}$ and $\mathcal{M}_{RN100DM100}$. The BF for CGW with those noise models has increased tremendously. Finally, we have tried the custom model for the six best EPTA pulsars. Most notable is the peculiar noise model for J1012+5307 (see table 3.3) which, in combination with the sky position of the candidate event (being next to it), suggests that the explanation might be in the time-correlated high-frequency noise present in that pulsar (see Appendix of Chalumeau 2021). This would also be consistent with the results of eccentric runs suggesting that this could be extended in the frequency feature. We do not know the origin of that noise; it was found empirically. Including the custom noise model reduced the BF for this event to $\mathcal{B}^{custom+CGW}_{custom} = 0.95$ and the posterior samples are not anymore constrained (see Figure 3.11 and, please, note that we have used different range for the parameters). We have also computed the Bayes factor for the noise-only models (assuming that GW signal is weak) and found that $\mathcal{B}^{RN30DM30}_{custom} =< 10^{-2}$ and similar result for RN30DM100 model suggesting that the data prefers by far the custom model.

Custom made noise models for PPTA pulsars were studied in Goncharov et al. 2020. However, the peculiar behaviour that we have found in IPTA data was mainly originating from J1012+5307 which is not observed by PPTA. For that reason, we have chosen to focus on noise models from Chalumeau 2021. On the other hand, the noise model for J0437-4715 was based on Goncharov et al. 2020 where the spectral index for RN is of 3 and the optimal number of frequency bins for DM is (at least) 91. For this pulsar, we chose 30 frequency bins for RN as it was the recommended value for spectral index > 1.5 (see Goncharov et al. 2021).

This was a useful exercise that triggered a set of investigations we would have to do in case of any CGW candidate. In addition, this section shows the importance of custom modelling noise for the best pulsars in the array, especially the noise at high frequencies, which is often partially neglected, assuming that it is dominated by the white (measurements) noise and it does not affect the search for the stochastic GW signal (which is most pronounced at low frequencies). The considered event shows how unmodelled

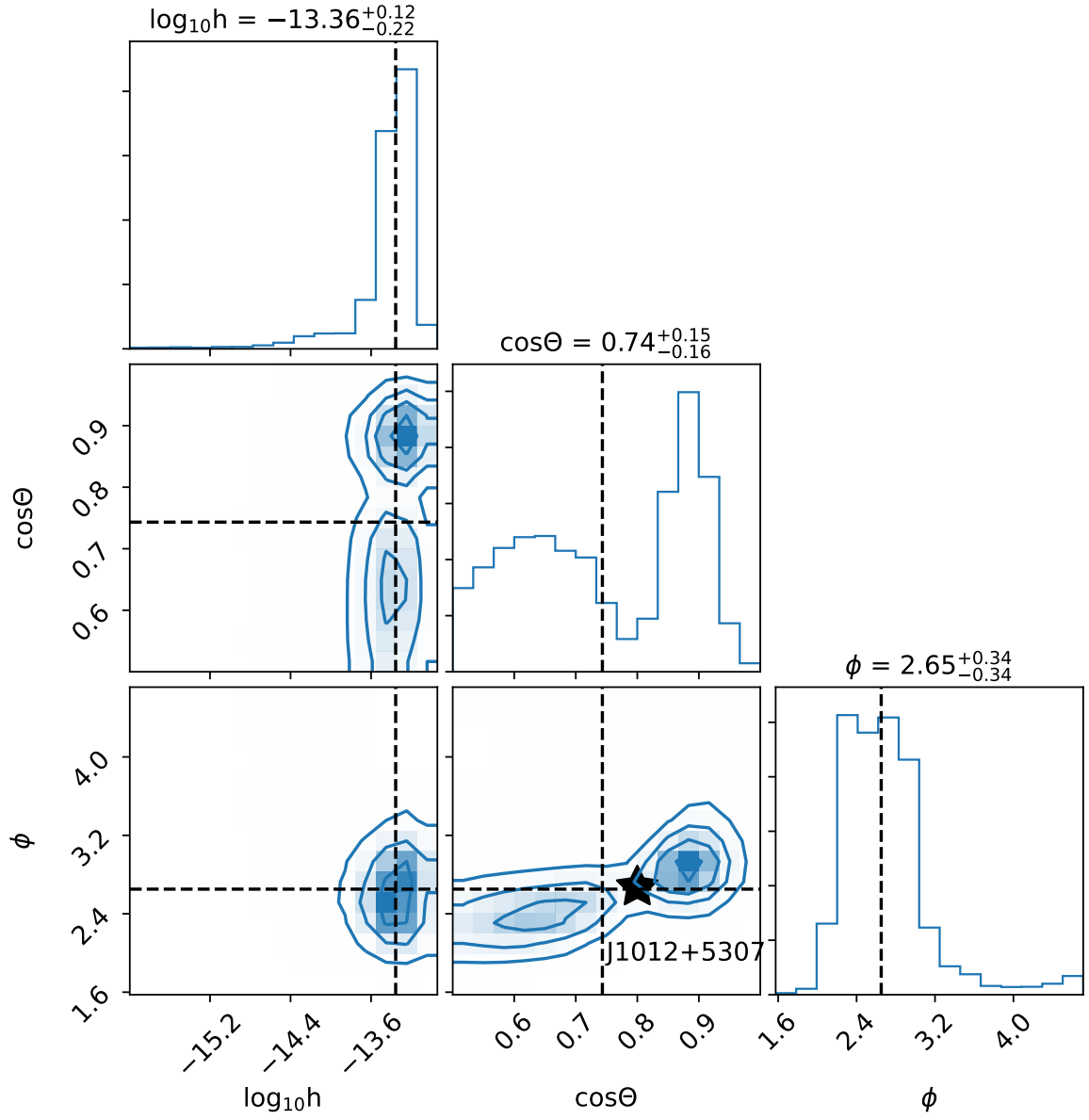

Figure 3.10: Posterior distribution for $\log h$, $\cos\theta$ and ϕ using noise model $\mathcal{M}_{RN30_DM30}$. The black star indicates the sky position of J1012+5307. The dashed lines represent the median values of the parameters.

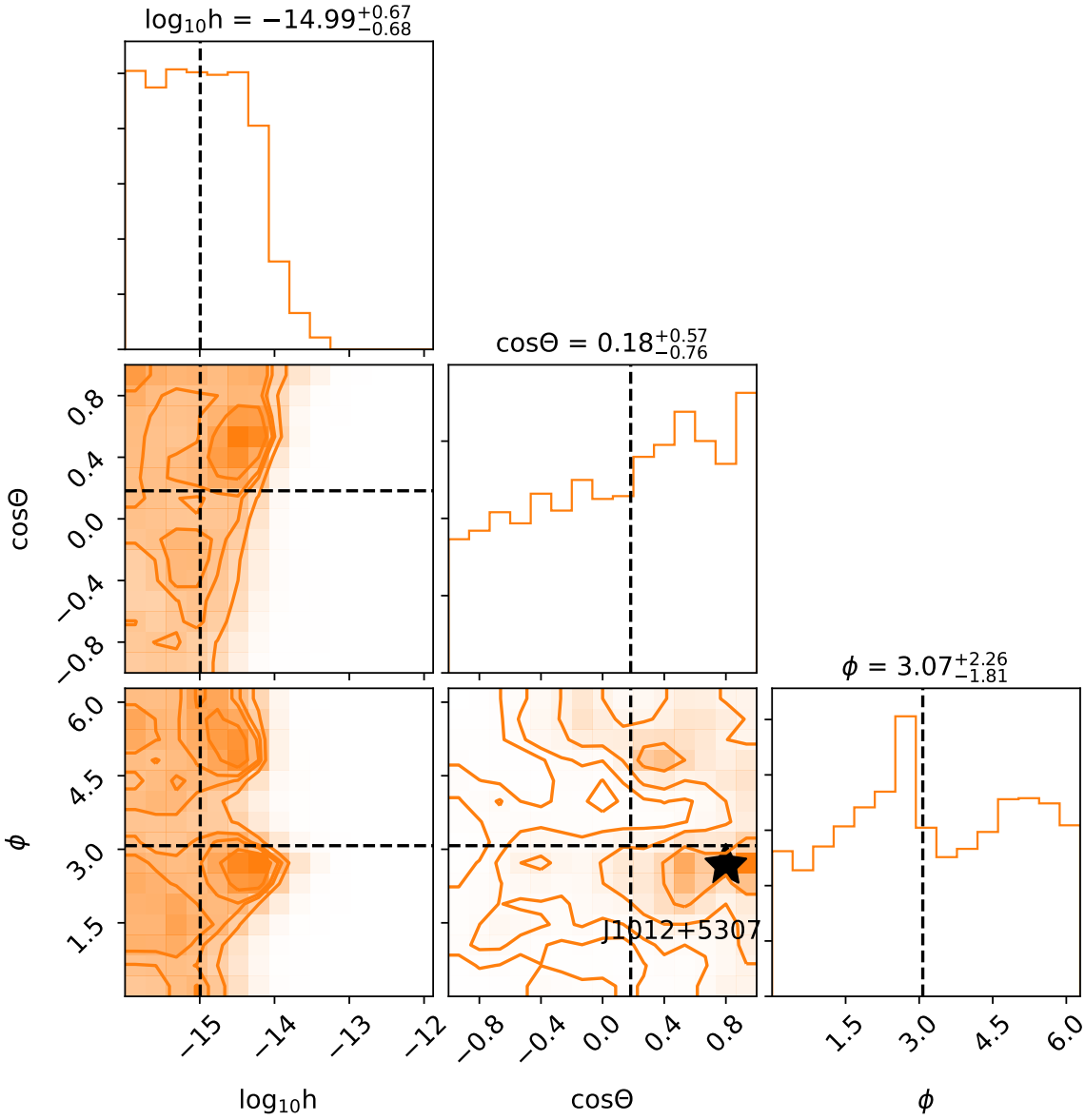

Figure 3.11: Posterior distribution for $\log h$, $\cos\theta$ and ϕ using noise model $\mathcal{M}_{custom}$. The black star indicates the sky position of J1012+5307. The dashed lines represent the median values of the parameters.

high-frequency noise could conspire for CGW signal.

3.7 Conclusion

We have searched for a continuous GW signal in the IPTA DR2 dataset. We have used the Bayesian approach and based detection criteria on the Bayes factor. We have shown that using a custom noise model for the six best EPTA pulsars is essential for the correct interpretation of the data. This is especially true for J1012+5307, which exhibits time-correlated noise at high frequencies. We found no CGW candidates using this noise model and proceeded to set the upper limit on GW strain. The addition of CRN in the noise model slightly affects the upper limit by lowering the sensitivity of the array at low frequencies. The most sensitive frequency appears to be around 10 nHz with a 95% sky averaged upper limit for CGW amplitude $h_{95} = 9.1 \times 10^{-15}$. The IPTA DR2 shows a much better upper limit than previously set at higher frequencies, making it a promising dataset to detect CGW.

During the analysis, we demonstrated the CGW candidate follow-up investigations program, which was an important exercise that should be used in subsequent PTA CGW analysis. The expected CGW signal has low SNR, and its SNR will be slowly accumulated as we get more pulsars and a longer observational span. Modelling noise in pulsar data is essential, especially at high frequencies.

This analysis was limited to circular SMBHBs using only the Earth term. The use of eccentric CGW signal and including the pulsar term might potentially improve the search; however, it brings signal complexity which might make harder the interpretation of the results and increases the parameter space. We are entering the era of very high quality and high cadence radio observations with new instruments like FAST (Hobbs et al. 2019) or SKA (Stappers et al. 2018) with sophisticated data analysis techniques. Additional investigations of the best approach to detecting CGW have to be re-investigated, probably using simulated data and/or an extended CGW signal injection campaign.

CHAPTER 4

KERNEL DENSITY ESTIMATION METHODS IN GRAVITATIONAL WAVE DATA ANALYSIS

Contents

Markov Chain Monte Carlo approach is frequently used within Bayesian framework to sample the target posterior distribution. Its efficiency strongly depends on the proposal used to build the chain. The best jump proposal is the one that closely resembles the unknown target distribution, therefore we suggest an adaptive proposal based on Kernel Density Estimation (KDE). We group parameters of the model according to their correlation and build KDE based on the already accepted points for each group. We adapt the KDE-based proposal until it stabilizes. We argue that such a proposal could be helpful in applications where the data volume is increasing and in the hypermodel sampling. We tested it on several astrophysical datasets (IPTA and LISA) and have shown that in some cases KDE-based proposal also helps to reduce the autocorrelation length of the chains. The efficiency of this proposal is reduces in case of the strong correlations between a large group of parameters.

4.1 Introduction

We live in the era of large physics and astrophysics projects and often have to deal with large and complex datasets. The data analysis usually requires large computing facilities and a single computation could sometimes last for weeks. Optimizing the analysis techniques and pipelines is then a key challenge of the data science associated with all large (astro)physical experiments.

Nowadays, it is quite common to use Bayesian framework for analysing the data. It is especially convenient if we have a parameterized data model (or several competing models) describing the data. In this approach we treat all parameters as random variables with some prior based either on some physical principles or informed from the previous independent experiments. We use the observations at hands to refine our prior knowledge and infer a posterior probability distribution function for parameters of a model, or even perform a selection among several models. Often we have to deal with multidimensional parameter space with a non-trivial likelihood function which can be evaluated only numerically. One of the most used tool to perform the numerical sampling from a target distribution is Markov Chain Monte Carlo (MCMC). Building a Markov chain that represents the desired posterior requires two key ingredients: (i) proposal suggesting how to choose point $\vec{X}_{i+1}$ given the last point in the chain $\vec{X}_i$; (ii) the detailed balance which ensures the reversibility of the chain. One of the most successful and frequently use proposal approach is to use parallel tempering: running several chains with logarithmic distributed temperature ladder (see for example Earl and Deem 2005). The hot chains play role of a proposal there and its efficiency depends on the interplay of number (and distribution) of hot chains (more chains is better) and computational demands (increase with the number of chains). Understanding the properties of the signal and the likelihood surface often leads to a custom proposal suitable for a particular problem.

Here we suggest another generic proposal based on the Kernel Density Estimation (KDE). The idea of using KDE has already been explored (Ashton and Talbot 2021; Farr et al. 2020). The main result of this paper is in a particular implementation of KDE itself and its embedding into a sampler. Even though the proposed method is very generic we will mainly discuss its implementations in the gravitational waves (GWs) data analysis.

Let us summarize the key points of the KDE-based proposal.

- KDE is used together with the set of other proposals in building a Markov chain. We assume an adaptive approach where we use the data accumulated in a chain to regularly rebuild the KDE. We repeat adjustments until the convergence criteria based on the Kulback-Leibler divergence is satisfied.

- In order to build KDE we split all parameters into several groups, where parameters in each group show evidence of mutual correlation. The KDE-based proposal

is most efficient if the full parameter space could be split into many small uncorrelated groups. The performance drops significantly if dimensionality of a group is larger than 5.

- We have build KDE with the self-optimizing bandwidth based on the distribution of a sample points provided at the input.

Note that the adaptation breaks "Markovian" properties of the chain. Either one should dismiss the parts of the chain during the adaptation or, in case of uninterrupted adaptation, assume that the chain is only asymptotically Markov. We give detailed description of implementation in the next two sections (4.2, 4.3).

We have implemented the KDE-based proposal in a particular sampler `https://gitlab.in2p3.fr/lisa-apc/mc3`. We give a detailed description of this sampler in Appendix 6.4. The main feature of this sampler is that it runs several chains either completely independently or as parallel tempering. Multi-chain run is used to compute Gelman-Rubin ratio (Gelman and Rubin 1992) to monitor the convergence.

We assess the performance of the suggested proposal in two applications to GW data analysis. In first one we analyse the data combined by International Pulsar Timing Array (IPTA) collaboration searching for a continous GW signal in the nano-Hz band. As the second dataset we use simulated LISA data publically available through LISA Data Challenge (2a) portal. We use the KDE-based proposal to infer parameters of 6 Galactic white dwarf binaries. We present the performance of our proposal for those two data analysis problems in Section 4.4, in particular we show that the KDE-based proposal allows to reduce the autocorrelation length while keeping high acceptance rate.

We conclude the paper with discussion on the limitation and possible extension of our method in Section 4.5.

4.2 Kernel Density Estimation

In this rather short section we describe our particular way of building KDE. We start with a short introduction to KDE and then give details of the bandwidth optimization that we use.

KDE is a non-parametric method used to estimate a probability density function (pdf) based on a finite set of sample points (Parzen 1962; Rosenblatt 1956). It is a smooth alternative to a histogram. The advantage of KDE is that it uses no binning and gives a continuous function interpolating (and extrapolating) across the whole parameter space. For a d-dimensional dataset $\{\vec{X}\}$ of size N and kernel $K(\mathbf{x},\vec{h})$, we have our KDE $\hat{f}(\mathbf{x},\vec{h})$:

$$\hat{f}(\mathbf{x},\vec{h}) = \frac{1}{N}\sum_{a=0}^{N-1} K(\mathbf{x}-\vec{X}_a,\vec{h}), \tag{4.1}$$

with parameter $\vec{h}$ specifying the bandwidth of the kernel. We use latin subscripts from the first half of the alphabet to enumerate the samples in the set. The main idea is to sum smooth kernel functions of $\mathbf{x}$ centered on each sample (input) data point $\vec{X}_a$. The overlaps between neighbouring kernels will add-up, shaping the PDF for the set of samples $\{\vec{X}\}$. The choice of the kernel is arbitrary and we choose to work with a gaussian kernel of the form:

$$K_g(\mathbf{x}-\vec{X}_a,\vec{h}) = \prod_{i=1}^{d} \frac{\exp\left\{-\frac{1}{2}\frac{|\mathbf{x}-\vec{X}_a|_i^2}{h_i^2}\right\}}{\sqrt{2\pi}h_i}, \tag{4.2}$$

where the h_i is the local bandwidth corresponding to i-th parameter $|\mathbf{x}-\vec{X}_a|_i$, and d is the dimensionality of the parameter space. We use the latin letter from the second half of the alphabet to enumerate particular parameters, and the vector notation corresponds to a vector in the parameter space.

4.2.1 Optimal bandwidth

A KDE has one free parameter which we want to tune, the bandwidth $\vec{h}$. Its value should be adapted to the dataset we are working with. There is no direct way of estimating it and we use an optimisation method (Jin et al. 2021) based on the minimisation of the mean squared error (MSE) ϵ^2 with respect to $\vec{h}$:

$$\epsilon^2 = \int d\mathbf{x}(\hat{f}(\mathbf{x},\vec{h}) - f(\mathbf{x}))^2, \tag{4.3}$$

$$\frac{\partial \epsilon^2}{\partial \vec{h}} = 0, \tag{4.4}$$

where $f(\mathbf{x})$ is the true pdf that we want to approximate with the KDE. The numerical way of evaluating this integral is given in the appendix 6.3.

Instead of using a global bandwidth $\vec{h}$, we can define a local bandwidth $\vec{h}_a$ for each kernel $K(\mathbf{x}-\vec{X}_a,\vec{h}_a)$ (Terrell and Scott 1992). In that case, our KDE $\hat{f}(\mathbf{x},\vec{h})$ is :

$$\hat{f}(\mathbf{x},\vec{h}) = \frac{1}{N}\sum_{a=0}^{N-1} K(\mathbf{x}-\vec{X}_a,\vec{h}_a). \tag{4.5}$$

Intuitively we expect the local bandwidth $\vec{h}_a$ to be scaled according to the local density of points. Indeed, the bandwidth is chosen so that it is narrow in the regions of parameter space where the samples are most dense and it is broad where we have fewer samples. This can ensure good interpolation and overlap between kernels, in particular, in high dimensional problem where the sample points are very sparse.

In practice we use solution of equation 4.4 to find optimal local bandwidth. The global bandwidth (if needed) could be defined as an average over all local values (by setting the input parameter `global bw = True`).

In Appendix 6.3, we have shown that each local bandwidth $\vec{h}_a = [h_{a,1}\ h_{a,2}\ ...]$ of the kernel centered on point $\vec{X}_a$ can be approximated by solving a linear system for the k nearest neighbours $\vec{X}_b$ of the form :

$$\mathbf{A}(\vec{X}_b)\begin{bmatrix}\frac{1}{h_{a,1}^2}\\ \frac{1}{h_{a,2}^2}\\ \vdots\end{bmatrix} = \vec{B}(k) \tag{4.6}$$

Matrix $\mathbf{A}$ and vector $\vec{B}$ are given in Appendix 6.3 eqn. 6.33, where we provide a detailed description of the method. They depend on the position of nearest neighbours which are the k_{near} points contained in a hypercube centered on the point $\vec{X}_a$ in the parameter space. The edge ΔX_i of the hypercube for each parameter X_i is defined as :

$$\Delta X_i = (\max_{\{\vec{X}\}} X_i - \min_{\{\vec{X}\}} X_i)/s \tag{4.7}$$

where maximisation and minimisation are performed over the set of input samples and s is a scaling parameter we call "`adapt scale`".

It might happen that a hypercube contains no point (besides the central). In that case, we cannot compute the local bandwidth and the point is discarded. Its bandwidth is later set to the global bandwidth as defined above. In case of high dimensionality and if the points in the dataset are very sparse, we change parameter s iteratively decreasing by a factor 2 until we find non-empty hypercubes. However, if this happens the evaluation of the bandwidth will probably be flawed and there is not much we can do about it, except use bigger datasets with more sample points. Often the amount of additional points that is needed to cover all “holes” could be very large incurring unmanageable computational cost. That is why good parameter grouping is essential : reducing dimensionality without loss of correlated features in the data. This will be the main subject of the next section.

4.3 Method

The main idea is to build a KDE for a given d-dimensional set of sample points $\{\vec{X}\}$. However, for a high dimensional KDE we are strongly affected by the "curse of dimensionality" (Scott and Sain 2005) because the sample sets are often limited in size leaving undercovered regions of the parameter space. In addition the efficiency of KDE is degrading if there are too many points since we place the kernel on top of each sample. For that

reason, we will split a d-dimensional parameter space into several low dimensional subspaces grouping the most correlated parameters together. We assume that sub-groups are not correlated and build the KDE for each of them $\hat{f}_\alpha(\mathbf{x}_\alpha, \vec{h}_\alpha)$, so the total KDE is the product of the low-dimensional KDEs:

$$\hat{F}(\mathbf{x}, \vec{h}) = \prod_\alpha \hat{f}_\alpha(\mathbf{x}_\alpha, \vec{h}_\alpha), \tag{4.8}$$

where the greek indices enumerate the subgroups of parameters. Forming these subgroups relies on the assessment of the correlation between parameters based on the provided set of samples $\{\vec{X}\}$ and this is the main subject of the next subsection.

4.3.1 Parameter grouping

For a d-dimensional dataset $\{\vec{X}\}$ we want to split the parameters in several sub-groups. Each sub-group will contain correlated parameters while parameters from different subgroups will be uncorrelated. We could use a covariance matrix to identify correlations, however it implicitly assumes Gaussian distribution and cannot account for any complex 2D structures between pairs of parameters. Instead we use a method based on the Jensen-Shannon divergence (JSD) (Nielsen 2019). JSD, similarly to the Kullback–Leibler (KL) divergence, measures the similarity between two distributions, but with the advantage of being symmetric and bound $0 \leq JSD \leq \ln 2$, moreover it does not make any assumptions about the distributions. For each pair of parameter (X_i, X_j) with the joined probability distribution $p(X_i, X_j)$ we compute

$$0 \leq \mathrm{JSD}\big(p(X_i, X_j) || p(X_i)p(X_j)\big) \leq \ln 2, \tag{4.9}$$

where the second distribution $p(X_i)p(X_j)$ is a product of one-dimensional (marginalized) distributions obtained by shuffling the parameters $\{X_i, X_j\}$ that supposed to destroy any correlations between parameters X_i, X_j (see figure 4.1). If the JSD is low, it means that the shuffling did not affect the dataset and the parameters did not exhibit correlation. On the other hand, if the JSD is large, the shuffling did change something and the parameters should be grouped together.

We define a JSD threshold (we usually use 0.1) which we consider that two parameters are correlated. Starting from one parameter, we iterate the process to extract all correlated (chained) pairs. Once no additional correlated parameter is found, we take the union of all correlated pairs of parameters to form a sub-group. This process is illustrated in figure 4.2.

In case of multimodality of the probability distribution function that we try to reproduce with KDE, we implemented an additional (optional) feature: clustering samples before building KDE. This feature is especially welcome when the modes are separated by

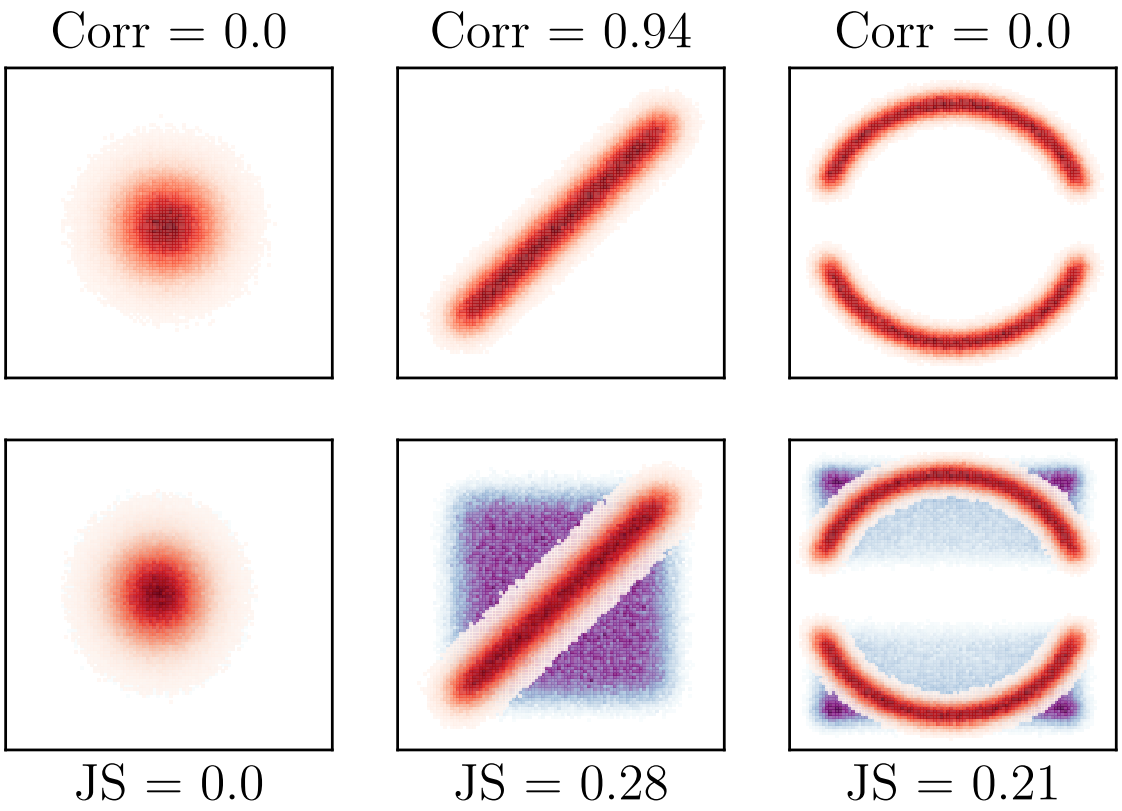

Figure 4.1: We plot three examples of datasets where we have on the left no correlation, in the middle linear correlation and on the right more elaborate features. On the top panels we have the corresponding values of the correlation coefficient based on simple evaluation of the covariance matrix. We see that it excels at finding the linear correlation but completely fails with the right panel features. On the bottom panels we have the same three datasets in red with their corresponding shuffled version that destroys correlations in blue. While the left panel remains unchanged, the others are affected and it is captured by the JS divergence.

very low probability valleys. We cluster the samples (using k-means method MacQueen 1967) and apply KDE building approach described above to each mode. This does not change the fundamental structure of the KDE but it helps the bandwidth adaptation.

4.3.2 Turning KDE into proposal

It is desirable in several application to use posterior points inferred for some parameters into a prior for another investigation. Let us give several examples:

- The data inferred from electromagnetic observations in form of samples is used as a prior for the GW experiment. In this case we can either build a joined likelihood

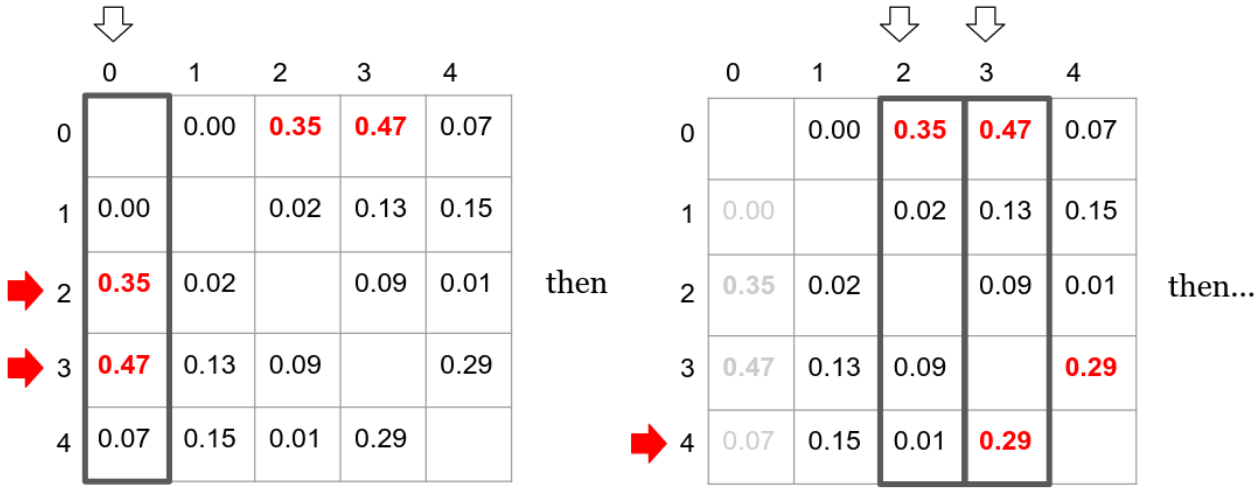

Figure 4.2: Illustration of the parameter grouping process using a JSD matrix. For this example, the JSD threshold is set to 0.25, hence the parameter sub-groups will be [0, 2, 3, 4] and [1]. Starting from parameter 0, we find that JSD for parameters 2 and 3 are above the threshold so they are both correlated with parameter 0. Then we check for parameters 2 and 3 and find that 4 is correlated with 3 while 2 sees no additional correlation. The last step would have been to check for 4 and find that there are no additional correlations. Therefore, 0, 2, 3 and 4 will form a sub-group of correlated parameters. The parameter "1" is the last parameter that does not correlate with others (according to the adopted threshold) and it will be a sub-group on its own.

or, alternatively, build KDE on the external posteriors and use it as a prior while analyzing the GW data.

- In Pulsar Timing Array (PTA) data analysis we often first investigate data acquired for each pulsar and trying to build a noise model. Later this pulsar and associated noise model is plugged into "Array" of pulsars for searching for a GW signal. It is proven to boost significantly efficiency of the GW search if we use posteriors for the noise model inferred in the first step as a proposal in the global fit later on.

- Often the data is taken continuously and we want to analyse it "on the fly"; that is true for GW data analysis. In this case you want to increment the data with a certain cadence while using information about the sources acquired from the analysis of the past data. One possibility is to turn again posterior built from the analysis of, say, first half a year of data into a prior in the analysis of the whole year of data.

- The method in Hee et al. 2016 gives a practical suggestion on how to compute the Bayes factor comparing several models without computing the evidence for each model. In this approach we introduce hyper-parameter indexing the models and

jump in this parameter (say, within MCMC) which corresponds to jumping between the models. For this method to be robust the exploration within each model must be very efficient otherwise it will lead to very long poorly converging runs or to spurious results. If we have posteriors for each (or some) model available, we can turn them into proposal and use in the hyper-model exploration.

The main idea of grouping parameters in building KDE implies that we probably also want to make jumps within each subgroup (or in some subgroups) while keeping other parameters fixed (Gibbs-like sampling). The subgroups for a current jump are chosen randomly assuming equal probability attached to each subgroup. The randomness implies reversibility and equal probability is a warrant that we jump in all parameters evenly (on average) while performing low dimensional jumps. Let us denote the number of sub-KDEs that is used for each jump n_{kde}, then the proposal probability is

$$\hat{F}_{n_{kde}}(\mathbf{x}, \vec{h}) = \prod_{\alpha}^{n_{kde}} \hat{f}_{\alpha}(\mathbf{x}_{\alpha}, \vec{h}_{\alpha}), \tag{4.10}$$

where the subscript in $\mathbf{x}_{\alpha}$ implies that we vary only parameters that belong to that (α) subgroup. This probability[1] is used to balance the chain in the Metropolis-Hastings step of the MCMC algorithm (Hastings 1970). Choosing a point from a given sub-KDE $\hat{f}_{\alpha}(\mathbf{x}_{\alpha}, \vec{h}_{\alpha})$ is done by drawing a point from the randomly chosen kernel $K(\mathbf{x} - \vec{X}_a, h_a)$ of $\hat{f}_{\alpha}(\mathbf{x}_{\alpha}, \vec{h}_{\alpha})$ centered on $\vec{X}_a$:

$$\vec{X}_b^* \to \vec{X}_a + \mathcal{N}(\vec{0}, h_a), \tag{4.11}$$

where $\mathcal{N}(\vec{0}, h_a)$ is a normally distributed random variable with $\vec{0}$ mean and covariance matrix diag(h_a) that is the bandwidth of kernel $K(\mathbf{x} - \vec{X}_a, h_a)$. For a random set of n_{kde} sub-KDEs, the newly proposed point $\vec{X}^*$ is the union of parameters from each subgroup $\vec{X}_b^*$:

$$\vec{X}^* = \bigcup_b \vec{X}_b^*. \tag{4.12}$$

4.3.3 Adaptive proposal

In case we do not have samples from the previous investigations, we still can build KDE-based proposal using the points accepted by a running MCMC. There are several caveats which need to be considered: (i) during the burn-in and even some time after the distribution of the accepted points is quite unstable that will reflect on the KDE (ii) we are breaking the rules of MCMC, the chain is only asymptotically Markov, so that at some point we should fix the KDE-based proposal and dismiss all samples accumulated before.

[1] The probability $\hat{F}_{n_{kde}}(\mathbf{x}, \vec{h})$ corresponds to the proposal probability for the jump $q(\vec{\theta}|\vec{\theta}^*)$ presented in section 2.3.3

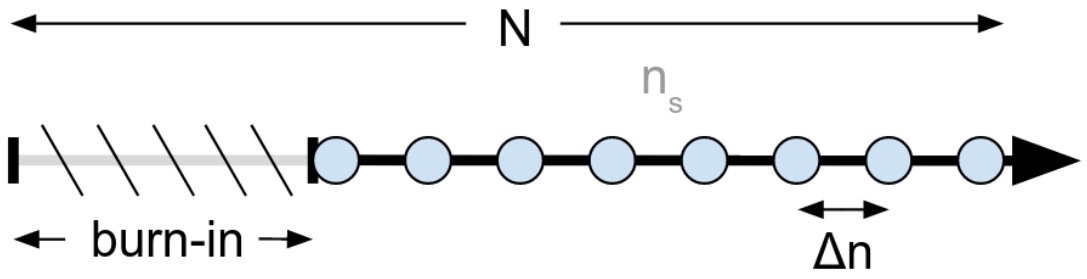

Figure 4.3: For a chain of total length N, we get rid of the burn-in, then we extract n_s linearly space samples from the remaining fraction of the chain. These n_s points are used to build the KDE.

The rest of this subsection gives the details of the practical implementation of the adaptation.

To build a KDE on the currently sampled chain, we select only a subset of the total N samples. In particular, we take n_s uniformly spaced points, from burn-in to the last point, as illustrated in Figure 4.3. The burn-in is taken to be a fraction, q, of the total chain length N and those points are dismisssed (hence, $n_s = (1-q)N$). As chain evolves, the burn-in could also grow until we get to the stationary distribution. The post-burn-in length is also increasing, and, since we keep n_s fixed, the space Δn between the selected samples grows. We expect that the quality of the KDE improves with increase in Δn because of the reduced correlation in the samples taken for building KDE, and that is true until Δn reaches typical autocorrelation length of the chain (Hogg and Foreman-Mackey 2018).

We re-build a new KDE after each N_{adapt} iterations (jumps) of the chain. We want to track the evolution of KDE and stop adapting when it has reached stability (that could be another indicator of the burn-in phase). We compare the new (re-built) KDE $\hat{F}_1(\mathbf{x}, \vec{h}_1)$ with the old $\hat{F}_0(\mathbf{x}, \vec{h}_0)$ by computing the KL divergence (Kullback and Leibler 1951):

$$\mathrm{KL}(\hat{F}_0||\hat{F}_1) = \int d\mathbf{x}\hat{F}_0(\mathbf{x})(\log\hat{F}_0(\mathbf{x}) - \log\hat{F}_1(\mathbf{x})) \tag{4.13}$$

This integral could be approximated as

$$\mathrm{KL}(\hat{F}_0||\hat{F}_1) \simeq \frac{1}{n_s}\sum_a (\log\hat{F}_0(\vec{X}_{0a}, \vec{h}_0) - \log\hat{F}_1(\vec{X}_{0a}, \vec{h}_1)), \tag{4.14}$$

where $\{\vec{X}_0\}$ is the set of n_s samples used to build $\hat{F}_0(\mathbf{x}, \vec{h}_0)$. This gives a measure of change in KDE between successive updates: ΔKL. We stop updating KDE in order to preserve the

ergodicity of the process (Atchadé and Rosenthal 2005) as soon as convergence criterion

$$\frac{|<\Delta \mathrm{KL}>|}{\sqrt{<\mathrm{KL}^2>}} < 5\% \tag{4.15}$$

is satisfied. The angular brackets denote the averaging over the last 5 updates and we demand that the average change in KL is small compared to the average KL values. ΔKL can be negative or positive, depending on the evolution of KL. If KL does not converge to a specific value, this condition ensures that it is at least oscillating around a mean value.

4.4 Results

We consider two datasets and perform search/parameter estimation using MCMC with KDE-based proposal.

In first application we consider a dataset from International PTA collaboration and perform the noise analysis for each pulsar in the array (Perera 2019). The likelihood is expected to be quite broad and unimodal but the dimensionality of the parameter space is large as well as its overall volume.

In second application we work with the simulated LISA data and search/characterize small bandwidth with several Galactic white-dwarf binaries. The likelihood in this case has more complex structure with quite strong correlation between parameters.

We quantify performance of the KDE-based proposal using the following criteria:

- Closeness between built KDE and the true distribution using **KL divergence**[2].
- The **acceptance rate** when the KDE is used as a proposal with MCMC sampler
- The **autocorrelation length** (Hogg and Foreman-Mackey 2018) of the MCMC chain when the KDE is used as proposal, to evaluate improvement in the mixing of the chain.

4.4.1 IPTA dataset

We build the KDE using $n_s = 10000$ samples from the chains generated by previous MCMC runs. We adopt the following choice of parameters for generating KDE:

- `js threshold = 0.1`
- `adapt scale = 10`
- `use kmeans = False`

[2]Because we do not know the true distribution, we have to use histograms to evaluate KL. The formula for binned KL is given in Appendix 6.2.

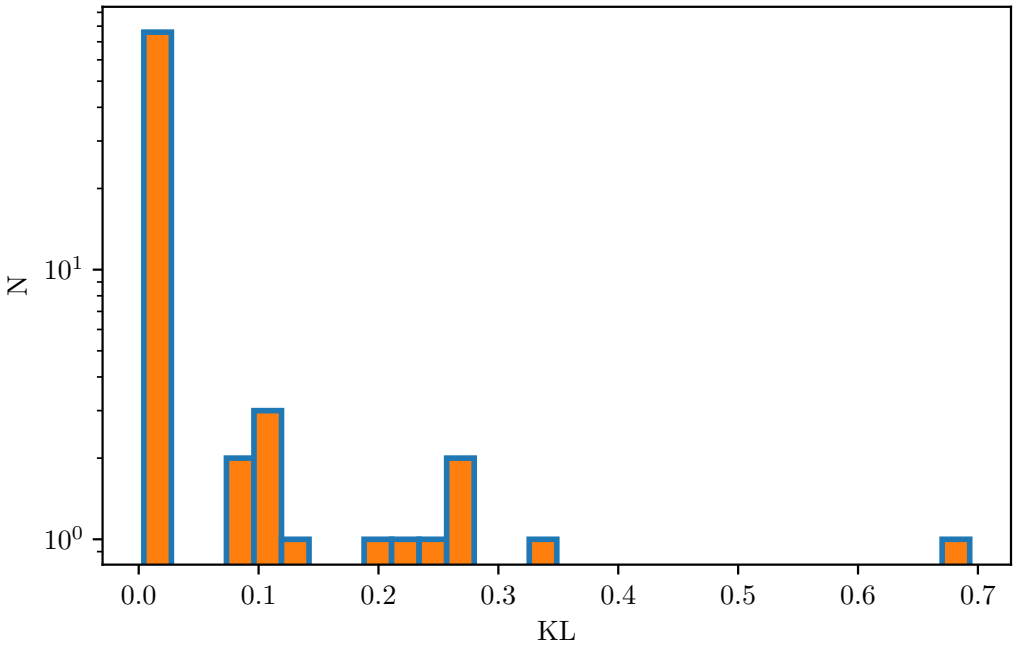

Figure 4.4: Distribution of the KL divergence for all sub-KDEs. 1-d sub-KDEs are of best quality.

- `global bw = True`
- `n kde = 1`

Within total 104 parameters, the grouping algorithm finds : **76** 1-dimensional sub-groups, **12** 2-dimensional sub-groups and **1** 4-dimensional sub-group. For each of these sub-groups we calculate the KL divergence $\mathrm{KL}(p_\alpha(\mathbf{x}_\alpha)||\hat{f}_\alpha(\mathbf{x}_\alpha,\vec{h}_\alpha))$ of the true pdf of the sub-group $p_\alpha(\mathbf{x}_\alpha)$ against the corresponding sub-KDE $\hat{f}_\alpha(\mathbf{x}_\alpha,\vec{h}_\alpha)$. A good sub-KDE should give a $\mathrm{KL}(p_\alpha(\mathbf{x}_\alpha)||\hat{f}_\alpha(\mathbf{x}_\alpha,\vec{h}_\alpha))$ that is close to 0. Because the number of parameter is large, we show the histogram plot in figure 4.4 depicting KL for each subgroup.

The impact of the dimensionality of the sub-KDE on KL can be clearly seen in Figure 4.4 . Close to 0 we have 76 1-dimensional sub-KDEs, between 0.1 and 0.4 we have 12 2-dimensional sub-KDEs, and the 4-dimensional sub-KDE has KL ≈ 0.7. As discussed in the previous section, increasing dimensionality implies sparse data samples, so we expect KL values to rise because the KDE might fail to interpolate the PDF correctly between neighbouring points, producing holes in the distribution. Computing KL allows us to quantify this effect and assess the quality of the KDE that may not be obvious by just eyeballing (see figure 4.5).

Next we will analyse the IPTA data using MCMC sampler (Ellis and Haasteren 2017) and ENTERPRISE (Ellis et al. 2020) package for computing likelihood function. We chose

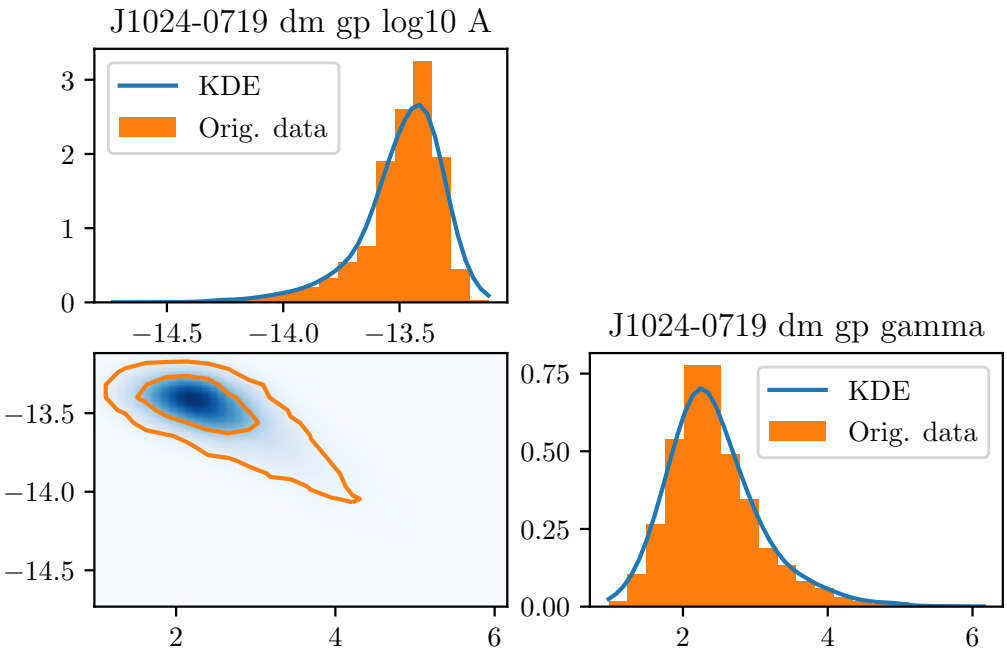

Figure 4.5: 2-dimensional corner plot of binned original dataset against smooth KDE plot. KDE is blue, original is orange. The contour levels for the original dataset on the bottom left panel are [0.1, 0.95]. For this sub-KDE, KL ≃ 0.11.

to use two jump proposals Single-Component Adaptive Metropolis (SCAM) (Haario et al. 2001) and Differential Evolution (DE) (Ter Braak 2006; Jasper A. Vrugt 2008) to compare to KDE. We search for a continuous gravitational wave signal while fitting for pulsar noise parameters (Aggarwal 2019). The KDE for the noise parameters has been built using posterior samples obtained from the preceding single pulsar analysis. We will compare three different runs:

- using a KDE-based + default jump proposals SCAM and DE (labeled as “KDE”);
- using binned empirical distributions + default proposals SCAM and DE (labeled as “Binned”);
- using only default proposals SCAM and DE (labeled as “None”).

The binned empirical distributions are essentially 2-dimensional histograms based on the same posterior samples as used in building KDE (Taylor et al. 2021). Those carry a similar spirit to KDE, being pair-wise approximation to marginalized posterior, but at the same time are fundamentally different from the KDE in the sense that the KDE is a continuous function in space that interpolates between the sample points using smoothening kernel. In addition KDE-based proposal makes grouping based on the parameter correlation that could lead to more than two dimensional group (see figure 4.4).

	KDE	SCAM	DE
Acceptance rate	0.57	0.39	0.43

Table 4.1: Acceptance for various proposals (KDE run)

	Binned	SCAM	DE
Acceptance rate	0.47	0.41	0.43

Table 4.2: Acceptance for various proposals (binned proposal run)

Tables 4.1, 4.2, compare acceptance rate of different proposals in two independent runs. One can see that KDE has the highest acceptance rate due to smart parameter grouping and interpolation between the samples incorporated in the KDE-based proposal. For the jumps we could use simultaneous jump in one or several (n_{kde}) sub-groups. Increasing n_{kde} leads to the higher dimensionality of the jumps and has a strong impact on the acceptance rate as shown in Figure 4.6. The best results are achieved if we perform jumps in one subgroup at the time, the acceptance rate decreases exponentially with n_{kde}.

A high acceptance rate is not necessarily a sign of a good proposal as it has to be paired with the low autocorrelation length. For each run, we compute the autocorrelation lengths of all parameters and compare the maximum, minimum and mean autocorrelation lengths. The low autocorrelation length implies that the samples drawn/accepted are independent. Results are presented in Table 4.4. KDE performs very well for IPTA data. Minimum autocorrelation length does not seem to be affected much by the choice of proposal but the maximum is reduced by a factor 2 when using KDE. Reducing the maximum is the most important because thinning the chain by this factor ensures that all samples are independent for all parameters. The mean autocorrelation length is just an indicator of the average performance of the proposal.

Like for the acceptance rate, we check the influence of n_{kde} on the autocorrelation length. Results are given in Table 4.5. For high values of n_{kde}, even though we have decreasing acceptance rate, the autocorrelation drops too and mixing improves. However, this result should be take with a caution, using high n_{kde} could lead to a very low acceptance point as shown in Figure 4.6. In Figure 4.7 we show the autocorrelation length as a function of n_{kde} and it indicates that the optimal number is around $n_{kde} = 5$ with ac-

n_{kde}	KDE	SCAM	DE
1	0.47	0.41	0.43
5	0.22	0.36	0.41
10	0.08	0.35	0.40

Table 4.3: Acceptance for different n_{kde}

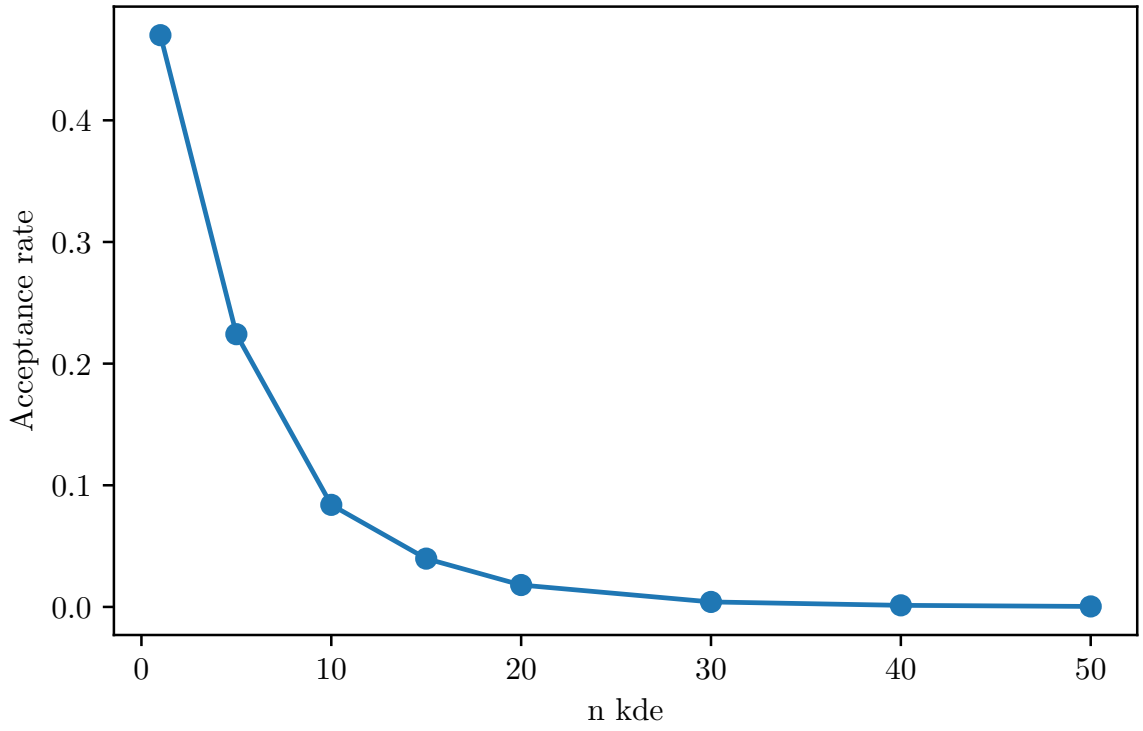

Figure 4.6: Acceptance rate for several values of n_{kde}.

	max	**min**	**mean**
KDE	578	27	113
Binned	1386	29	160
None	1032	25	208

Table 4.4: The maximum, minimum and mean autocorrelation lengths for three runs.

n_{kde}	**max**	**min**	**mean**
1	578	27	113
5	386	25	79
10	395	28	95

Table 4.5: The maximum, minimum and mean autocorrelation lengths for different n_{kde}.

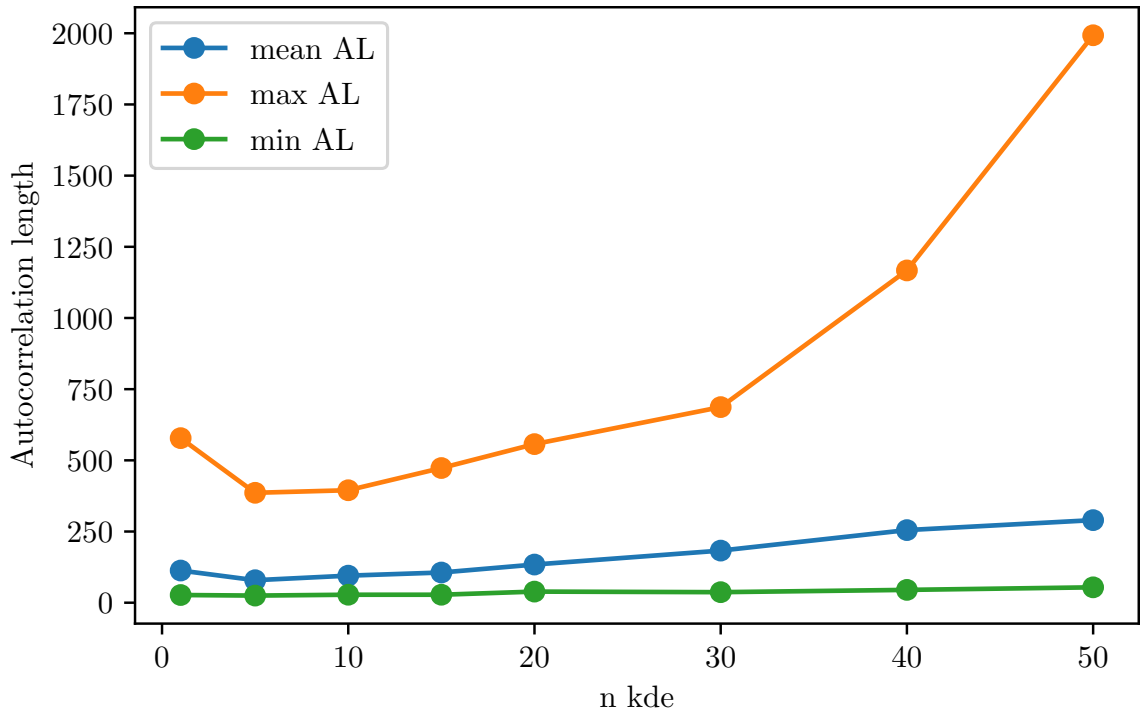

Figure 4.7: Mean, maximum and minimum autocorrelation lengths for several values of n_{kde}.

ceptance rate of 0.22 that is very close to the expected optimal acceptance rate of 0.234 (Gelman et al. 1997). Based on these results, we recommended to work with values of n_{kde} between 1 and 10, especially when the dimensionality of the parameter space is high like in this IPTA example.

4.4.2 LISA dataset

Now we turn our attention to the simulated LISA data, in particular, we use "Sangria" dataset which is 1 year long, it contains about a dozen of merging massive black hole binaries and about 30 millions of Galactic binaries https://lisa-ldc.lal.in2p3.fr/. Here we are interested in Galactic binaries in the very narrow frequency range around 4 mHz and we have removed all merging black holes. We have detected 6 sources in that frequency interval and we use KDE-based proposal together SCAM and DE[3]. This time we use home-made sampler *M3C2* (https://gitlab.in2p3.fr/lisa-apc/mc3), specifically parallel tempering version of it. We describe this sample in details in the Appendix 6.4. This sampler is using Metropolis-Hastings acceptance-rejection step as well as *slice* sampling (Neal 2003).

[3]The DE introduced in Ter Braak 2006 is using population MCMC, here we rather use it on a single chain in the spirit described in the snooker proposal in Jasper A. Vrugt 2008

Each Galactic binary is characterized by 8 parameters, so we have in total 48 parameters (Littenberg 2011). We expect some parameters (like amplitude and the orbital inclination angle) to correlate for each source, and, in addition, some parameters could correlate between the sources. Here we try to build a proposal on-the-fly. The likelihood surface for this problem is rather complex having many well separated maxima (reason for using parallel tempering). We will build KDE as we accumulate samples: adapting KDE proposal with the rate every 5000 samples and using $n_s = 5000$ samples for each chain. Besides KDE we also use SCAM proposal and slice sampling. Note that the combination of SCAM jump with Metropolis-Hastings sampling and slice (being very independent) already significantly reduce the autocorrelation of the accepted points.

First, we consider the convergence of KDE adaptation. During the burn-in stage the KDE is changing quite violently. The correlation between parameters is quite unstable which leads to fluctuation in how parameters are grouped and in the number of subgroups. As burn-in proceeds we keep track of the grouping and fix the splitting in subgroups as soon as it stabilizes (when the same grouping appears at least 5 times). Once we have fixed sub-groups we check the KL divergence between the subsequent updates of KDE (as described above). The results of KDE adaptation are presented in Figure 4.8. The consistent grouping of parameters was reached after 26 updates as indicated by a dashed red line. Then we compute KL after each update and stop adaptation once the condition 4.15 is met (see the right panel of Figure 4.8). From then on we keep KDE fixed and perform the actual sampling.

As a next step we want to check the acceptance rate of the KDE-based proposal and compare it to SCAM and slice. Note that slice is not based on the Metropolis-Hastings acceptance/rejection algorithm, however we still can introduce an effective acceptance rate as a ratio of total number of slice calls to the total number of likelihood evaluations used by slice.

We consider two cases for the subgroup jumps $n_{kde} = 1$ and $n_{kde} = 5$. Figure 4.9 compares acceptance rate for $n_{kde} = 1$. One can see that it stabilizes around 0.31 very fast, and, despite that it is lower than what we had for the PTA application, it is a very decent acceptance rate. SCAM has a similar acceptance (but usually longer autocorrelation length), while slice is worse by a factor 10 (though it usually has low autocorrelation length).

Figure 4.10 compares acceptance rate for $n_{kde} = 5$. Increase in the dimensionality of the jumps has a drastic effect on the acceptance rate, its value drops to about 0.015. It is also interesting to compare behaviour of SCAM and slice for two runs. Slice shows very stable/consistent results, while SCAM has significant fluctuations though preserving the trend. The two-dimensional KDE jumps seem to be the best option in this application.

Next we check the autocorrelation length when we run with and without KDE-based proposal. We restrict ourselves with the case $n_{kde} = 1$ since $n_{kde} = 5$ has a very poor acceptance rate. Figure 4.11 compares maximum and mean autocorrelation of two runs.

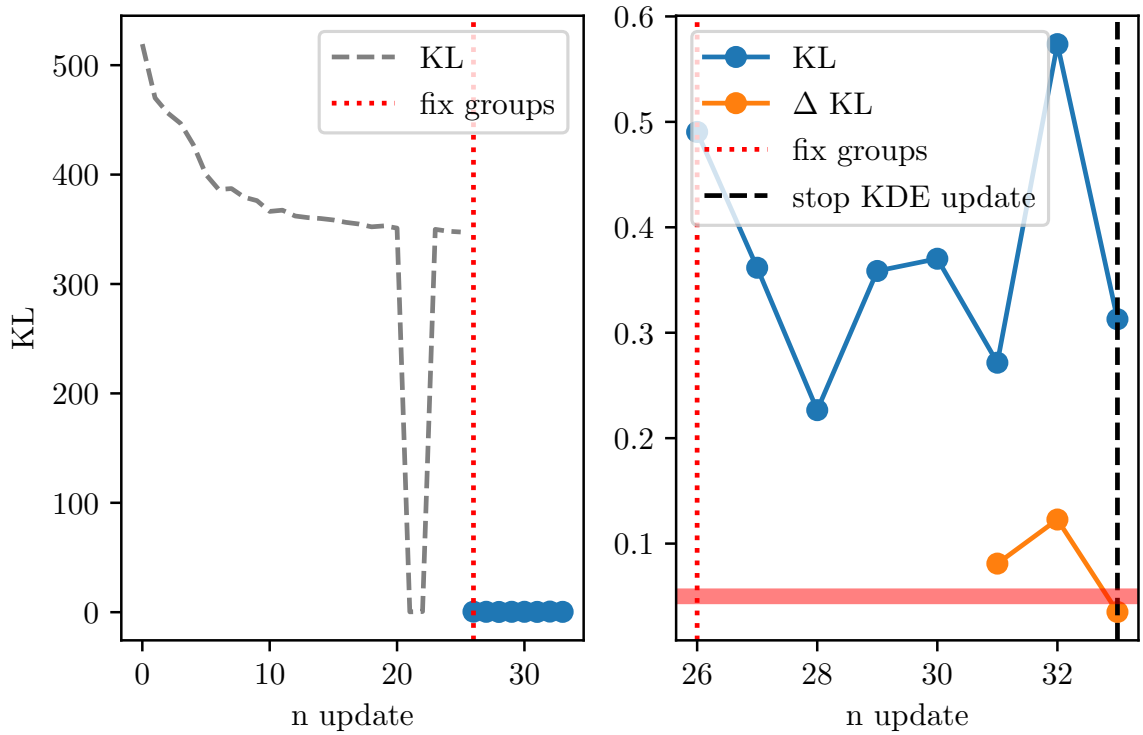

Figure 4.8: Evolution of KL and ΔKL. Left panel, evolution of KL before fixing parameter groups. Right panel, evolution of KL after fixing parameter groups, we start computing ΔKL 5 updates after the grouping was fixed. The thick red line indicates the ΔKL threshold level of 5% below which we reach convergence.

We observe that the mean value is slightly (about 17%) lower when we include KDE-based proposal and the maximum length remains the same. As we have already mentioned, mixing SCAM with Metropolis-Hastings and slice steps does reduce the autocorrelation already (compared to PTA example where we did not use slice sampling). In addition we use parallel tempering algorithm, where the hot chains could be seen as yet another jump proposal. All in all, KDE does not add much to already reduced autocorrelation run in the current analysis.

4.5 Conclusion

Bayesian formalism is a usual approach in nowadays gravitational waves data analysis. The inference of the parameters posterior distribution is often done using MCMC and the efficiency strongly depends on the proposal it uses. In this article we have presented KDE-based proposal which can be either built on-the-fly during the extended burn-in stage or constructed using posterior points from another run.

The suggested KDE-based proposal has several extended features: (i) the adaptive

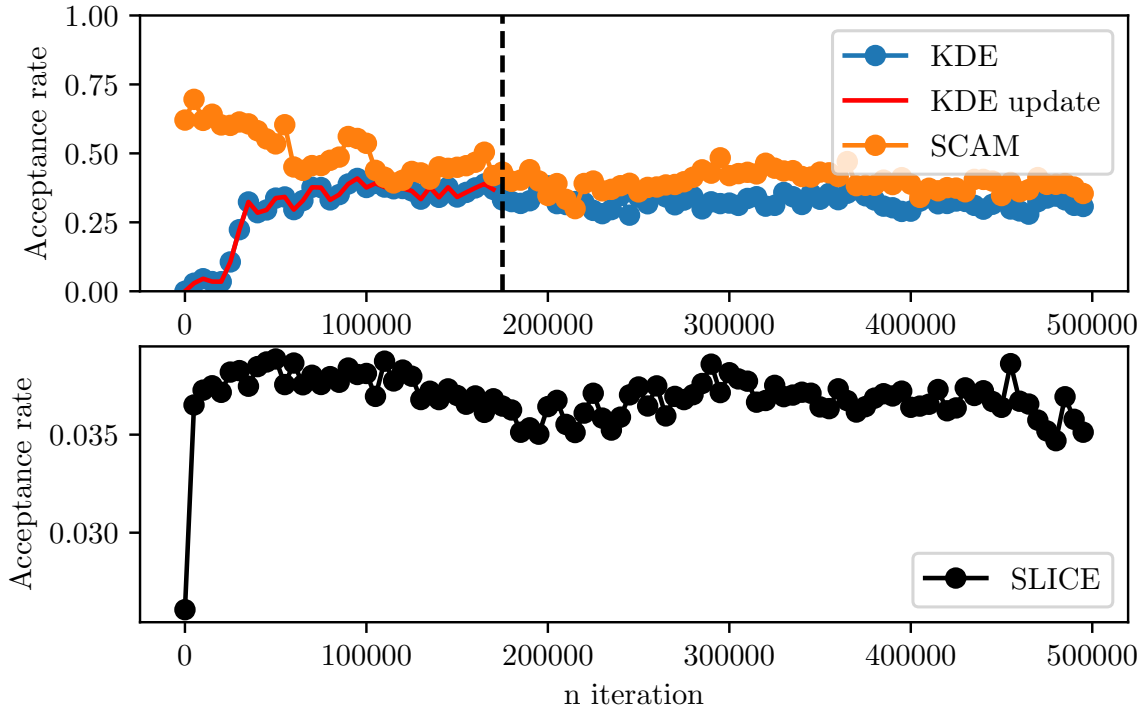

Figure 4.9: Acceptance rate of all proposals for $n_{kde} = 1$. KDE and SCAM are on top panel, SLICE on bottom panel. Red plot and black dashed line shows where KDE finally converged and stopped updating.

bandwith based on the local density of points (small bandwidth in the densely sampled regions of the parameter space); (ii) splitting parameter space into sub-groups of the correlated parameters and applying KDE on each subgroup, we identify correlations using JSD; (iii) possibility of building KDE adaptively.

We tested this proposal by running MCMC on IPTA data, using a KDE that we have built from previous MCMC runs (i.e. non adaptive case). The advantage of KDE-based proposal was clearly seen in high acceptance rate with low autocorrelation length. We have found that using rather low value $n_{kde} = 1 - 5$ (number of subgroups used in the jump simultaneously) seems to be optimal.

Another application of the KDE-based proposal was in running PTMCMC on the simulated LISA data searching for Galactic white dwarf binaries in a narrow frequency band. In this case we have built KDE adaptively during an extended burn-in stage. The addition of KDE-proposal to the sampling had only a moderate impact: it shows a decent acceptance rate (about 31%) with only small improvement in the autocorrelation length. Moreover we have shown that low-dimensional jumps are strongly preferred.

Few things could be improved, most notably in the adaptation. The threshold for grouping parameters was chosen somewhat ad hoc, and correlation of some parameters could be close to the threshold. We did observe the fluctuation in choosing the sub-

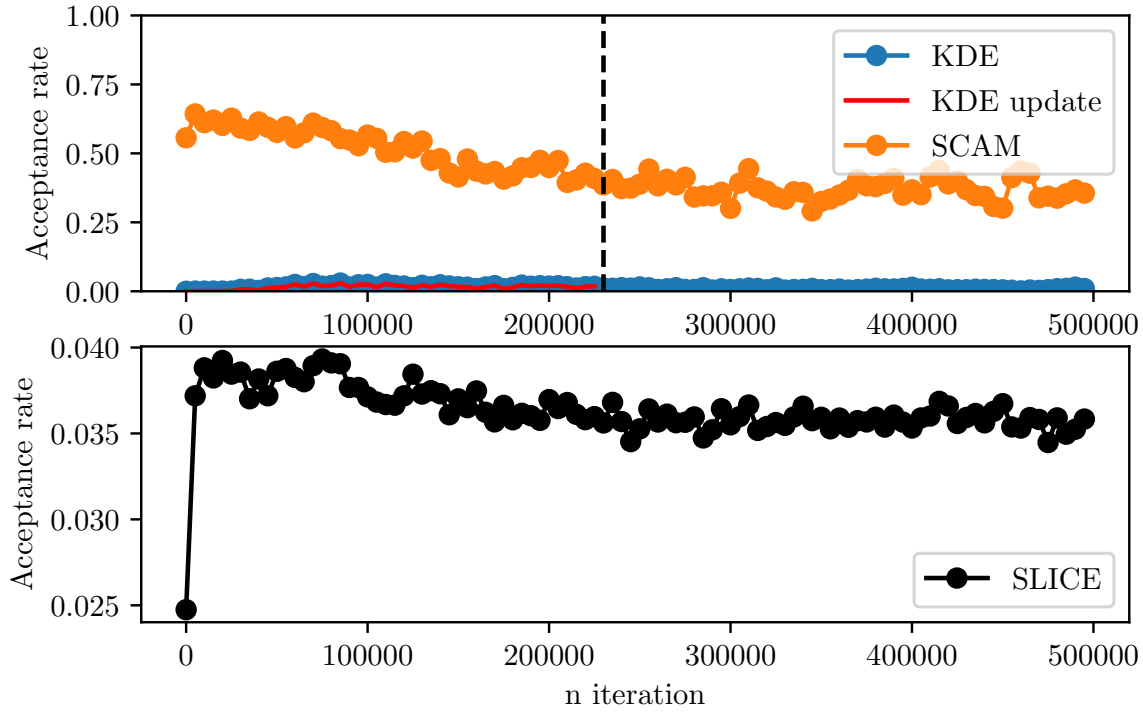

Figure 4.10: Acceptance rate of all proposals for $n_{kde} = 5$. KDE and SCAM are on top panel, SLICE on bottom panel. Red plot and black dashed line shows where KDE finally converged and stopped updating.

groups during the adaptation. One possibility could be to choose not one but two plausible grouping, build KDE for each and use two KDE proposals in a probabilistic manner. The criteria for stopping adaptation was also chosen somewhat arbitrary, and might benefit from the further tuning. Finally, we should implement an adaptive tuning for optimal n_{kde} based on the acceptance rate.

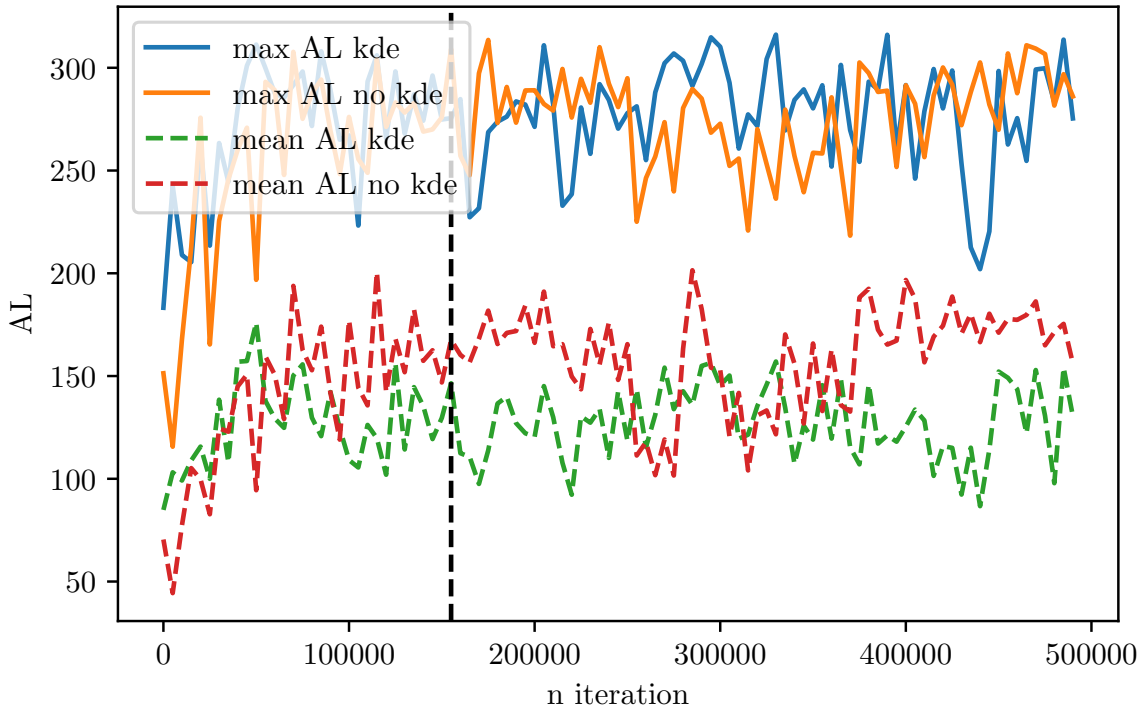

Figure 4.11: Maximum and mean autocorrelation lengths for $n_{kde} = 1$, comparing a run with KDE against a run without KDE. Black dashed line shows where KDE finally converged and stopped updating.

CHAPTER 5

SUMMARY AND PERSPECTIVES

In this PhD thesis, we have explored the low frequency realm of gravitational waves using the stability of millisecond pulsars with Pulsar Timing Arrays. More specifically, we have used the second data release of the International Pulsar Timing Array that is the combination of all the pulsar timing data collected for the past decades. It gathers high precision measurements from the European, North-American and Australian collaborations. We developed robust data analysis pipelines to search for the presence of continuous gravitational wave signals, i.e. monochromatic and deterministic gravitational wave signals induced by supermassive black hole binary systems. The manuscript was separated into four main sections : (I) the presentation of the theoretical background necessary to understand the physics of the pulsar timing array, (II) the data analysis methods that are used to search for gravitational wave signal in the data, (III) the results of this search obtained with the IPTA DR2 and finally, (IV) the presentation of a new method using kernel density estimation to optimize our data analysis pipelines.

The search for deterministic gravitational wave signal in the IPTA DR2 was conclusive. We found no evidence for such signals so we have set an upper limit on amplitude h. It appears that the IPTA DR2 is most sensitive for signals with a frequency around 10nHz where its upper limit is $h_{95} = 9.1 \times 10^{-15}$ and provides the best sensitivity to date of all PTAs at high frequencies thanks to the improved cadence of observation. Following the recent discoveries of low frequency processes common to all pulsars in the array, we need to include it into the model. It came out that the latter correlates with the common red noise, slightly deteriorating the upper limit at low frequencies. Even though the search was not fruitful, we have stumbled upon gravitational wave like signals at high frequencies with relatively high Bayes factor that triggered our attention. These features were absorbed after changing the noise model. This result is crucial to understand how the noise model directly affects the observation and must be chosen with great care. This study led to a publication for which I am corresponding author. It was shared with the international collaboration and approved by the steering committee. It was submitted to MNRAS under the name : Searching for continuous Gravitational Waves in the second data release of the International Pulsar Timing Array.

We proposed a new method using kernel density estimation in gravitational wave data analysis. It came from a desire to simplify and yet optimize the data analysis pipelines that we use. In this chapter, we present a new fast semi-analytical method to optimize

the bandwidths of the kernels. This optimal kernel density estimation is then used with MCMC as a proposal distribution. We quantify the quality of our proposal by estimating the autocorrelation length (i.e. the effective sampling time) of the produced Markov chains for different proposals. It turned out that our proposal was giving the best performance with IPTA data. We wanted to push it a step further by suggesting an adaptive proposal that would build itself as the MCMC is sampling. We tested it on simulated LISA data to search for galactic binary signals. The proposal was still reducing the effective sampling time but its performance was less probing. This work led to the writing of an article that was submitted of Physical Review D and pre-published on Arxiv (Falxa et al. 2022).

The search for gravitational wave signal is still an ongoing project. The recent detection of the common red noise process by all collaborations really has spawned renewed interest for the pulsar timing world. Hopefully, in the upcoming days, we will finally detect the signals that have been hunted for many years. The latest data releases together with the newly available radio-observatories provide the best sensitivity to date, raising hopes of gravitational wave detection in the next few years.

CHAPTER 6

APPENDIX

Contents

6.1 Averaged signal to noise ratio

In this section, we propose to develop the analytical averaging of the signal to noise ratio formula presented in section 3.3.1. This formula is developed fromg the expression of the timing residuals induced in PTA by a CGW that was demonstrated in section 1.4.4. This calculation is rather trivial. We use for the timing residuals (only Earth term)

$$s = F^{+} s_{+} + F^{\times} s_{\times}, \tag{6.1}$$

with

$$s_{+}(t) = \frac{h}{\omega}\left[-\sin[2\Phi(t)]\mathcal{I}\cos 2\psi \; -\cos[2\Phi(t)]\mathcal{J}\sin 2\psi\right] \tag{6.2}$$

$$s_{\times}(t) = \frac{h}{\omega}\left[-\sin[2\Phi(t)]\mathcal{I}\cos 2\psi \; +\cos[2\Phi(t)]\mathcal{J}\sin 2\psi\right], \tag{6.3}$$

where we have defined

$$h = \frac{2\mathcal{M}^{5/3}(\pi f)^{2/3}}{d_L} \tag{6.4}$$

$$\mathcal{I} = 1 + \cos^2\iota \;\; \mathcal{J} = 2\cos\iota. \tag{6.5}$$

The signal to noise ratio is, using the definition of the noise weighted inner product :

$$\rho^2 = (s|s) = s^T \Sigma^{-1} s \tag{6.6}$$

We will denote the averaging of a quantity A with respect to some parameter p with brackets and index $< A|A >_p$. For the signal to noise ratio ρ^2, function of the GW parameters, we have

$$\int \frac{d\psi}{\pi} \frac{d\phi_0}{2\pi} \frac{d\cos\iota}{2} \frac{d\cos\theta}{2} \frac{d\phi}{2\pi} \rho^2 =< s|s >_{\psi,\phi_0,\iota,\theta,\phi}, \tag{6.7}$$

$$(s|s) = (F^+)^2 (s_+|s_+) + (F^\times)^2 (s_\times|s_\times) + 2F^+F^\times (s^+|s^\times), \tag{6.8}$$

In the following, we omit the $\left(\frac{h}{\omega}\right)^2$ prefactor. We have for each term in 6.8

$$\begin{aligned}(s_+|s_+) \propto \mathcal{I}^2 \cos^2 2\psi (\sin 2\Phi(t)|\sin 2\Phi(t)) + \mathcal{J}^2 \sin^2 2\psi (\cos 2\Phi(t)|\cos 2\Phi(t)) \\ + 2\mathcal{I}\mathcal{J} \cos 2\psi \sin 2\psi (\sin 2\Phi(t)|2\cos\Phi(t)),\end{aligned} \tag{6.9}$$

$$\begin{aligned}(s_\times|s_\times) \propto \mathcal{I}^2 \sin^2 2\psi (\sin 2\Phi(t)|\sin 2\Phi(t)) + \mathcal{J}^2 \cos^2 2\psi (\cos 2\Phi(t)|\cos 2\Phi(t)) \\ - 2\mathcal{I}\mathcal{J} \cos 2\psi \sin 2\psi (\sin 2\Phi(t)|\cos 2\Phi(t)),\end{aligned} \tag{6.10}$$

$$\begin{aligned}(s_+|s_\times) \propto \mathcal{I}^2 \cos 2\psi \sin 2\psi (\sin 2\Phi(t)|\sin 2\Phi(t)) - \mathcal{J}^2 \cos 2\psi \sin 2\psi (\cos 2\Phi(t)|\cos 2\Phi(t)) \\ - 2\mathcal{I}\mathcal{J} (\sin^2 2\psi - \cos^2 2\psi)(\sin 2\Phi(t)|\cos 2\Phi(t)).\end{aligned} \tag{6.11}$$

We average the expression with respect to the polarization angle ψ to get

$$< s|s >_\psi = \frac{1}{2}\left(\frac{h}{\omega}\right)^2 \left[(F^+)^2 + (F^\times)^2\right] \left[\mathcal{I}^2 (\sin 2\Phi(t)|\sin 2\Phi(t)) + \mathcal{J}^2 (\cos 2\Phi(t)|\cos 2\Phi(t))\right]. \tag{6.12}$$

In the case of a monochromatic, non evolving CGW signal, we can approximate $2\Phi(t)$ by :

$$2\Phi(t) \to \omega t + \phi_0, \tag{6.13}$$

so the $\sin 2\Phi(t)$ and $\cos 2\Phi(t)$ terms become

$$\sin 2\Phi(t) = \sin\omega t \cos\phi_0 + \cos\omega t \sin\phi_0 \tag{6.14}$$

$$\cos 2\Phi(t) = \cos\omega t \cos\phi_0 - \sin\omega t \sin\phi_0, \tag{6.15}$$

yielding

$$(\sin 2\Phi(t)|\sin 2\Phi(t)) = \cos^2\phi_0(\sin\omega t|\sin\omega t) + \sin^2\phi_0(\cos\omega t|\cos\omega t) + 2\cos\phi_0\sin\phi_0(\sin\omega t|\cos\omega t), \tag{6.16}$$

$$(\cos 2\Phi(t)|\cos 2\Phi(t)) = \cos^2\phi_0(\cos\omega t|\cos\omega t) + \sin^2\phi_0(\sin\omega t|\sin\omega t) - 2\cos\phi_0\sin\phi_0(\sin\omega t|\cos\omega t). \tag{6.17}$$

After averaging with respect to the initial phase we get

$$< \sin 2\Phi(t)|2\sin\Phi(t) >_{\phi_0} = < \cos\Phi(t)|\cos\Phi(t) >_{\phi_0} = \frac{1}{2}(\sin\omega t|\sin\omega t) + \frac{1}{2}(\cos\omega t|\cos\omega t), \tag{6.18}$$

So ρ^2 becomes

$$< s|s >_{\psi,\phi_0} = \frac{1}{4}\Big(\frac{h}{\omega}\Big)^2\big[(F^+)^2 + (F^\times)^2\big]\big[\mathcal{I}^2 + \mathcal{J}^2\big]\Big[(\sin\omega t|\sin\omega t) + (\cos\omega t|\cos\omega t)\Big]. \tag{6.19}$$

We now compute the average of the $\mathcal{I}^2 + \mathcal{J}^2$ term with respect to the inclination ι

$$< \mathcal{I}^2 + \mathcal{J}^2 >_\iota = \frac{16}{5}, \tag{6.20}$$

and we have

$$< s|s >_{\psi,\phi_0,\iota} = \frac{4}{5}\Big(\frac{h}{\omega}\Big)^2\big[(F^+)^2 + (F^\times)^2\big]\Big[(\sin\omega t|\sin\omega t) + (\cos\omega t|\cos\omega t)\Big]. \tag{6.21}$$

This partially averaged, sky location dependent SNR formula can be used to rank the pulsars according to frequency of CGW and sky location. However, we want the fully averaged ρ^2 so we need to express the antenna pattern functions in terms of the CGW sky location parameters

$$F^+(\hat{\Omega}) = \frac{1}{2}\frac{(\hat{m}\cdot\hat{p})^2 - (\hat{n}\cdot\hat{p})^2}{1+\hat{\Omega}\cdot\hat{p}} \tag{6.22}$$

$$F^\times(\hat{\Omega}) = \frac{(\hat{m}\cdot\hat{p})(\hat{n}\cdot\hat{p})}{1+\hat{\Omega}\cdot\hat{p}}, \tag{6.23}$$

$$\begin{aligned}\hat{\omega} &= -(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)\\ \hat{m} &= (-\sin\phi, \cos\phi, 0)\\ \hat{n} &= -\cos\theta\cos\phi, -\cos\theta\sin\phi, \sin\theta),\end{aligned} \tag{6.24}$$

To average over sky location parameters, we place ourselves in the frame where the earth-pulsar axis unit vector $\hat{p} = (0,0,1)$, which makes the calculation easy and gives :

$$< (F^+)^2 + (F^\times)^2 >_{\theta,\phi} = \frac{1}{3}. \tag{6.25}$$

Finally, we have for the averaged signal to noise ratio to CGW signal :

$$\boxed{< s|s >_{\psi,\phi_0,\iota,\theta,\phi} = \frac{4}{15}\left(\frac{h}{\omega}\right)^2 \Big[(\cos\omega t|\cos\omega t) + (\sin\omega t|\sin\omega t)\Big]} \tag{6.26}$$

6.2 Kullback-Leibler divergence

We test the quality of our KDE by computing the KL divergence for each subgroup $\hat{f}_\alpha(\mathbf{x}_\alpha, \vec{h}_\alpha)$. Because we cannot know the true pdf for a set of samples $\{\vec{X}\}$, we have to resort to binning. From sub-KDE $\hat{f}_\alpha(\mathbf{x}_\alpha, \vec{h}_\alpha)$ built on subgroup of parameter samples $\{\vec{X}_\alpha\}$, we draw a new set of samples $\{\vec{X}^*_\alpha\}$. We estimate the corresponding normalized histogram distributions P_α and P^*_α using same grid of N bins to compute the KL divergence as Kullback and Leibler 1951 :

$$\mathrm{KL}(\mathrm{P}_\alpha||\mathrm{P}^*_\alpha) = \sum_{i=0}^{N} \mathrm{P}_{\alpha,i}\Big(\ln \mathrm{P}_{\alpha,i} - \ln \mathrm{P}^*_{\alpha,i}\Big). \tag{6.27}$$

To avoid divergence of the logarithms, we set every $\mathrm{P}_{\alpha,i}$ and $\mathrm{P}^*_{\alpha,i}$ that are equal to 0 to the minimum found value in $\mathrm{P}_\alpha \bigcup \mathrm{P}^*_\alpha$ that is not 0.

6.3 Optimal bandwidth of KDE

.

We start this appendix with defining few useful expressions that will be used in our derivations later.

- The overlap between two neighbouring kernels of same bandwidth is given by

$$\boxed{\int d\mathbf{x} K(\mathbf{x} - \vec{X}_a, \vec{h}) K(\mathbf{x} - \vec{X}_b, \vec{h}) = \prod_{i=1}^{d} \frac{\exp -\frac{1}{4}\frac{|\vec{X}_a - \vec{X}_b|_i^2}{h_i^2}}{\sqrt{\pi} 2 h_i}}. \tag{6.28}$$

  Let us remind you that d is the dimensionality.

- If a set of samples $\{\vec{X}\}$ of size N drawn from the probability density function $f(x)$, then we can approximate the averaging integral as :

$$\boxed{\int dx f(x) g(x) \simeq \frac{1}{N} \sum_a g(X_a)}, \tag{6.29}$$

  where the function g is evaluated at the sample points $\vec{X}_a$.

The main objective of this Appendix is to derive the optimal local bandwidth which is defined through the minimization of the mean square error:

$$\begin{aligned}\epsilon^2 &= \int dx(\hat{f}(x,h) - f(x))^2 \\ &= \int dx\hat{f}^2(x,h) - 2\int dx\hat{f}(x,h)f(x) + \int dx f^2(x) \\ &= \int dx\hat{f}^2(x,h) - 2\int dx\hat{f}(x,h)f(x) + const,\end{aligned} \tag{6.30}$$

where $f(x)$ is the true PDF, ($\hat{f}(x,h)$ is its KDE approximation and $const$ is the term independent of the bandwidth h. We introduce local bandwidth h_a attached to each sample point X_a. Next we assume that all points in the vicinity of each point have similar bandwidth, in other words, $h_b \approx h_a$ for k_{near} local points X_b. Using these assumptions we can approximate the first term:

$$N^2 \int dx\hat{f}(x,h_a)\hat{f}x,h_b = \sum_a \prod_{i=1}^{d} \frac{1}{2\sqrt{\pi}h_{a,i}} + \sum_a \sum_{b\neq a} \prod_{i=1}^{d} \frac{e^{-\frac{1}{4}\frac{\Delta X^2_{ab,i}}{h^2_{a,i}}}}{2\sqrt{\pi}h_{a,i}}, \tag{6.31}$$

where $h_{a,i}$ is i-th component of bandwidth attached to Gaussian kernel at $\vec{X}_a$ and $\Delta X_{ab,i} = (\vec{X}_a - \vec{X}_b)_i$ is i-th component of the a vector connecting two samples in the parameter space.

Using now second bullet equation and excluding the actual sample from the sum (for improving stability and removing the possible bias, see "leave one out estimator" Jin et al. 2021) we obtain for the second term

$$2\int dx\hat{f}(x,h_a)\hat{f}x \approx \frac{2}{N^2}\sum_a \sum_{b\neq a} \prod_{i=1}^{d} \frac{e^{-\frac{1}{2}\frac{\Delta X^2_{ab,i}}{h^2_{a,i}}}}{\sqrt{2\pi}h_{a,i}}, \tag{6.32}$$

where we have assumed $N \gg 1$, $N(N-1) \approx N^2$. Combining these terms together gives us

$$N^2(\epsilon^2 - const) = \sum_a \left\{ \prod_{i=1}^{d} \frac{1}{2\sqrt{\pi}h_{a,i}} + \sum_{b\neq a} \prod_{i=1}^{d} \frac{e^{-\frac{1}{4}\frac{\Delta X^2_{ab,i}}{h^2_{a,i}}}}{2\sqrt{\pi}h_{a,i}} - 2\sum_{b\neq a} \prod_{i=1}^{d} \frac{e^{-\frac{1}{2}\frac{\Delta X^2_{ab,i}}{h^2_{a,i}}}}{\sqrt{2\pi}h_{a,i}} \right\}.$$

Find the minimum of this expression by differentiating with respect to $h_{c,j}$ and equating it to zero:

$$0 = -\frac{1}{h_{c,j}} \prod_{i=1}^{d} \frac{1}{2\sqrt{\pi} h_{c,i}} \left\{ 1 + \sum_{b \neq c} \left[1 - \frac{1}{2} \frac{\Delta X^2_{bc,j}}{h^2_{c,j}} \right] \prod_{i=1}^{d} e^{-\frac{1}{4} \frac{\Delta X^2_{bc,i}}{h^2_{c,i}}} \right.$$
$$\left. -2(\sqrt{2})^d \left[1 - \frac{\Delta X^2_{bc,j}}{h^2_{c,j}} \right] \prod_{i=1}^{d} e^{-\frac{1}{2} \frac{\Delta X^2_{bc,i}}{h^2_{c,i}}} \right\}$$

Next we assume quite conservative approximation: $\frac{\Delta X^2_{bc,j}}{h^2_{c,j}} \ll 1$ for all points $\vec{X}_b$ in vicinity of $\vec{X}_c$ and all components j. This assumption overestimates the bandwidth and therefore conservative: this is what is used in this paper. Expanding in this small parameters and retain only the terms quadratic in this small ratio we obtain the system of linear equations for $1/h^2_{c,j}$:

$$\frac{k_{near}(2^{d/2+1}-1)-1}{(2^{d/2}-1)} = \sum_{b \neq c} \left[3 \frac{\Delta X^2_{bc,j}}{h^2_{c,j}} + \sum_{i \neq j} \frac{\Delta X^2_{bc,i}}{h^2_{c,i}} \right]. \tag{6.33}$$

Solving this system at each point $\vec{X}_c$ for each direction in the parameter space (j) gives us the desired local bandwidth $\vec{h}_c$.

As an alternative approach we can assume that the bandwidth is comparable to the distance to the neighnbours and define $h^2_{c,j} = \overline{\Delta X^2}_{c,j}(1+\varepsilon_{c,j})$, where $\overline{\Delta X^2}_{c,j} = 1/k_{near} \sum_b \Delta X^2_{bc,j}$ is the average square distance (i-th component) to the points in vicinity of $\vec{X}_c$ and assume that $\Delta X^2_{bc,j}/\overline{\Delta X^2}_{c,j} \sim 1$ and $\varepsilon_{c,j} \ll 1$ for all b, c, j. This yields

$$\prod_{i=1}^{d} e^{-\frac{1}{4} \frac{\Delta X^2_{bc,i}}{h^2_{c,i}}} \approx \left(1 + \frac{1}{4} \sum_{i}^{d} \frac{\Delta X^2_{bc,i}}{\overline{\Delta X^2}_{c,i}} \right) \prod_{i=1}^{d} e^{-\frac{1}{4} \frac{\Delta X^2_{bc,i}}{\overline{\Delta X^2}_{c,i}}}. \tag{6.34}$$

Using this approximation we arrive at the system of linear equations for $1/h^2_{c,j}$:

$$-1 + \sum_{b \neq c} 2(\sqrt{2})^d \prod_{i=1}^{d} e^{-\frac{1}{2} \frac{\Delta X^2_{bc,i}}{\overline{\Delta X^2}_{c,i}}} - P_{bc} \left(1 - \frac{1}{2} \sum_{i}^{d} \frac{\Delta X^2_{bc,i}}{\overline{\Delta X^2}_{c,i}} \right) =$$
$$\sum_{b \neq c} P_{bc} \left(3 \frac{\Delta X^2_{bc,j}}{\overline{\Delta X^2}_{c,j}} \varepsilon_{c,j} + \sum_{i \neq j} \frac{\Delta X^2_{bc,i}}{\overline{\Delta X^2}_{c,i}} \varepsilon_{c,i} \right),$$

where

$$P_{bc} \equiv \frac{1}{4} \left[\prod_{i=1}^{d} e^{-\frac{1}{4} \frac{\Delta X^2_{bc,i}}{\overline{\Delta X^2}_{c,i}}} - 4(\sqrt{2})^d \prod_{i=1}^{d} e^{-\frac{1}{2} \frac{\Delta X^2_{bc,i}}{\overline{\Delta X^2}_{c,i}}} \right].$$

Solving this system at each point $\vec{X}_c$ for each direction in the parameter space (j) gives us the desired local bandwidth $\vec{h}_c$.

6.4 The M3C2 sampler

The `M3C2` [1] (Multiple parallel Markov Chain Monte Carlo) is a python implementation of MCMC sampler. The aim of this tool is to improve the sampling robustness of complex posterior distribution, by running multiple chains in parallel. The cross check of the chain performance informs us about convergence (using Gelman-Rubin ratio Gelman and Rubin 1992). We have implemented two mechanisms of building the chain (1) using slice sampling (`slice`) and (2) Metropolis-Hastings algorithm (MH) which could be used separately or together improving the mixture of the chains and reducing the auto-correlation length. Even though the sampler is very generic, we primarily use it within the context of GW data analysis. For the Metropolis-Hastings method we have implemented a set of proposal jumps:

- `SCAM` (Single Component Adaptive Metropolis), jumps along one randomly chosen direction given by the eigen vectors of the covariance matrix Roberts and Rosenthal 2009; Haario et al. 2001
- `DE` (Differential Evolution), jumps along the direction given by difference of two randomly chosen samples of the chains, or (as in the classic implementation Ter Braak 2006; Jasper A. Vrugt 2008) by using state of different chains running in parallel.
- `ReMHA` (Regional Metropolis Hastings Algorithm), the proposal represented by a mixture of several Gaussians distributions Roberts and Rosenthal 2009; Craiu et al. 2009.

For `SCAM` we build the covariance matrix adaptively based on the accumulated samples. The use of the accumulated samples breaks the Markov property of the chain, making it asymptotically Markovian. The stability of the covariance matrix is yet another sign of the converged chain. One can stop adaptation after burn-in run. This proposal is suggested in Ellis 2013.

`ReMHA` is similarly used adaptively. We use accumulated samples during the burn-in to estimate the number of clusters using Variational Bayesian Gaussian Mixture (`skikit-learn` package) and use this Gaussian mixture probablity as a proposal. This proposal is somewhat similar to the one suggested in Craiu et al. 2009.

`DE` could be used as a proposal (snooker) described in Jasper A. Vrugt 2008 using multiple chains running in parallel or using the accumulated samples for each chain to propose the jump. There is no much difference between those two ways in case of well converged

[1] https://gitlab.in2p3.fr/lisa-apc/mc3

chains. However, the behavior and efficiency of those two implementations is very different during the burn-in stage.

In `slice` sampling, we use slicing of the parameter space either randomly or along the eigen directions of the covariance matrix, the choice is made with a probability set by user. In case of mixture of `slice` and MH, the frequency of each method is defined by a user specified weight. In addition to preset proposals available in `M3C2`, user can add custom jump-proposals using a common interface. The weights and proposals can be set individually for each running chain.

Besides running parallel independent chains, `M3C2` sampler has also parallel tempering implementation with an adaptive temperature ladder following Vousden et al. 2016. The adaptation is aiming at increasing the acceptance rate between the chains.

The multi-chain scheme of `M3C2` can be easily deployed on the multicore CPU infrastructure. Data exchange between chains, in case of parallel tempering, is restricted to its minimum level (pairwise communication between the chains), to ensure a good scalability.

6.5 Pulsar ranking for correlated signals

In L. Speri 2022, several methods are presented for the ranking of pulsars according to their contribution to the total SNR of the array, for a given correlated signal. In this manuscript, we mention SNR-B and Chimera when we cross-check the ranking obtained with the CGW method to support its validity. Here we will briefly explain what the methods consist in.

6.5.1 SNR-B

The SNR-B method computes the optimal statistics that maximize the SNR of the array in the presence of correlated signal (Rosado et al. 2015). It is given by :

$$\mathrm{SNR}_B^2 = 2\sum_{a<b}\int \frac{\Gamma_{ab}^2 S^2(f)\, T_{ab}}{\big[P_a(f)+S(f)\big]\big[P_b(f)+S(f)\big]+S^2(f)\Gamma_{ab}^2}, \tag{6.35}$$

where Γ_{ab} is the correlation coefficient between pulsars a and b, T_{ab} is the overlapping time of observation between pulsars a and b, $S(f)$ is the PSD of the expected signal and $P_a(f)$, $P_b(f)$ are the PSD of the intrinsic noises of pulsars (red noise, dispersions measure, etc).

6.5.2 Coupling matrix

The coupling matrix formalism seeks to quantify the overlap between different spatial correlation patterns between pulsars using spherical harmonics Y_{lm}. Indeed, $l = 0, 1, 2$

describe respectively a monopole, a dipole and a quadrupole, corresponding to the correlations expected for the three main types of correlated signals in PTA (see section 2.2.5). We decompose the sky distribution of pulsars on a spherical harmonic basis (Efstathiou 2004) and evaluate the overlap between the coefficients as :

$$K_{(lm)(lm)'} = \int d\Omega\, Y_{lm}(\Omega)\, W(\Omega)\, Y_{(lm)'}(\Omega), \tag{6.36}$$

with

$$W(\Omega) = \sum_a w_a \delta(\Omega - \hat{p}_a), \tag{6.37}$$

the window function, giving the positions $\hat{p}_a$ of pulsar with an associated weight w_a. The weight w_a is chosen according to the RMS value of pulsar a residuals (Roebber 2019).

For a set of pulsars, we evaluate the leakage between components with the ratio $\delta_\lambda = \lambda_{min}/\lambda_{max}$, λ being the eigenvalues of the matrix $M_{ll'}$ obtained by averaging of $K_{(lm)(lm)'}$ over the components m. Selecting the largest ratio δ_λ ensures that we have the best possible decoupling between different l.

6.5.3 Chimera

The Chimera method is a combination of the two previous methods. We define an effective score $\mathrm{SC}_{Chimera}$ of SNR-B and δ_λ (coupling matrix) according to which we rank the pulsars :

$$\mathrm{SC}_{Chimera} = \mathrm{SNR}_B \delta_\lambda. \tag{6.38}$$

In L. Speri 2022, the 25 best pulsars acquired with this method are saved as best pulsars of the array.

BIBLIOGRAPHY

[1] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, T. Akutsu, et al.
"Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA"
Living Reviews in Relativity 23.1 (2020).

[2] K. Aggarwal et al.
"The NANOGrav 11 yr Data Set: Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries"
The Astrophysical Journal 880.2 (2019), p. 116.

[3] M. A. Alpar, A. F. Cheng, M. A. Ruderman, and J. Shaham
"A new class of radio pulsars"
Nature 300.5894 (Dec. 1982), pp. 728–730.

[4] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi, et al.
"Laser interferometer space antenna"
arXiv preprint arXiv:1702.00786 (2017).

[5] J Antoniadis et al.
"The International Pulsar Timing Array second data release: Search for an isotropic gravitational wave background"
Monthly Notices of the Royal Astronomical Society 510.4 (2022), pp. 4873–4887.

[6] Arzoumanian et al.
"The NANOGrav Nine-year Data Set: Limits on the Isotropic Stochastic Gravitational Wave Background"
The Astrophysical Journal Letters 821.1, 13 (2016), p. 13.

[7] Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, S. Chatterjee, J. M. Cordes, P. B. Demorest, X. Deng, T. Dolch, J. A. Ellis, R. D. Ferdman, N. Garver-Daniels, F. Jenet, G. Jones, V. M. Kaspi, M. Koop, M. T. Lam, et al.
"Gravitational Waves from Individual Supermassive Black Hole Binaries in Circular Orbits: Limits from the North American Nanohertz Observatory for Gravitational Waves"
ApJ 794.2, 141 (Oct. 2014), p. 141.

[8] Z. Arzoumanian, P. T. Baker, H. Blumer, B. Bécsy, A. Brazier, P. R. Brook, S. Burke-Spolaor, S. Chatterjee, S. Chen, J. M. Cordes, N. J. Cornish, F. Crawford, H. T. Cromartie, M. E. Decesar, P. B. Demorest, T. Dolch, J. A. Ellis, et al.
"The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background"
The Astrophysical Journal Letters 905.2, L34 (Dec. 2020), p. L34.

[9] G. Ashton and C. Talbot
"BILBY-MCMC: an MCMC sampler for gravitational-wave inference"
Mon. Not. Roy. Astron. Soc. 507.2 (Oct. 2021), pp. 2037–2051.

[10] Y. F. Atchadé and J. S. Rosenthal
"On adaptive Markov chain Monte Carlo algorithms"
Bernoulli 11.5 (2005), pp. 815 –828.

[11] Babak 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 (Jan. 2016), pp. 1665–1679.

[12] S. Babak
Gravitational Waves from Coalescing Binaries
Morgan and Claypool, 2020.

[13] S. Babak and A. Sesana
"Resolving multiple supermassive black hole binaries with pulsar timing arrays"
Phys. Rev. D 85 (2012), p. 044034.

[14] M. Bailes, A. Jameson, F. Abbate, E. D. Barr, N. D. R. Bhat, L. Bondonneau, M. Burgay, S. J. Buchner, F. Camilo, D. J. Champion, I. Cognard, P. B. Demorest, P. C. C. Freire, T. Gautam, M. Geyer, J.-M. Griessmeier, L. Guillemot, et al.
"The MeerKAT telescope as a pulsar facility: System verification and early science results from MeerTime"
Publications of the Astronomical Society of Australia 37 (2020).

[15] B. Becsy and N. J. Cornish
"Joint search for isolated sources and an unresolved confusion background in pulsar timing array data"
Classical and Quantum Gravity 37.13 (2020), p. 135011.

[16] B. Becsy, N. J. Cornish, and M. C. Digman
"Fast Bayesian analysis of individual binaries in pulsar timing array data"
Physical Review D 105.12 (2022).

[17] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng
Handbook of markov chain monte carlo
CRC press, 2011.

[18] S. M. Carroll
Lecture Notes on General Relativity
1997.

[19] A Chalumeau et al.
"Noise analysis in the European Pulsar Timing Array data release 2 and its implications on the gravitational-wave background search"
Monthly Notices of the Royal Astronomical Society 509.4 (2021), pp. 5538–5558.

[20] K. Chen and M. Ruderman
"Pulsar Death Lines and Death Valley"
ApJ 402 (Jan. 1993), p. 264.

[21] V. Corbin and N. J. Cornish
Pulsar Timing Array Observations of Massive Black Hole Binaries
2010.

[22] R. Craiu, J. Rosenthal, and C. Yang
"Learn From Thy Neighbor: Parallel-Chain and Regional Adaptive MCMC"
Journal of the American Statistical Association 104 (Dec. 2009), pp. 1454–1466.

[23] Desvignes et al.
"High-precision timing of 42 millisecond pulsars with the European Pulsar Timing Array"
Monthly Notices of the Royal Astronomical Society 458.3 (2016), pp. 3341–3380.

[24] G. Desvignes, R. N. Caballero, L. Lentati, J. P. W. Verbiest, D. J. Champion, B. W. Stappers, G. H. Janssen, P. Lazarus, S. Osłowski, S. Babak, C. G. Bassa, P. Brem, M. Burgay, I. Cognard, J. R. Gair, E. Graikou, L. Guillemot, et al.
"High-precision timing of 42 millisecond pulsars with the European Pulsar Timing Array"
Monthly Notices of the Royal Astronomical Society 458.3 (May 2016), pp. 3341–3380.

[25] S. Detweiler
"Pulsar timing measurements and the search for gravitational waves"
ApJ 234 (Dec. 1979), pp. 1100–1104.

[26] D. J. Earl and M. W. Deem
"Parallel tempering: Theory, applications, and new perspectives"
Physical Chemistry Chemical Physics 7.23 (2005), p. 3910.

[27] R. T. Edwards, G. B. Hobbs, and R. N. Manchester
"tempo2, a new pulsar timing package – II. The timing model and precision estimates"
Monthly Notices of the Royal Astronomical Society 372.4 (Oct. 2006), pp. 1549–1574.

[28] G. Efstathiou
"Myths and truths concerning estimation of power spectra: the case for a hybrid estimator"
MNRAS 349.2 (Apr. 2004), pp. 603–626.

[29] J. A. Ellis
"A Bayesian analysis pipeline for continuous GW sources in the PTA band"
Classical and Quantum Gravity 30.22, 224004 (Nov. 2013), p. 224004.

[30] J. A. Ellis, X. Siemens, and J. D. E. Creighton
"Optimal strategies for continuous gravitational wave detection in pulsar timing arrays"
The Astrophysical Journal 756.2 (2012), p. 175.

[31] J. Ellis and R. van Haasteren
jellis18/PTMCMCSampler: Official Release
2017.

[32] J. A. Ellis, M. Vallisneri, S. R. Taylor, and P. T. Baker
ENTERPRISE: Enhanced Numerical Toolbox Enabling a Robust PulsaR Inference SuitE
Zenodo. 2020.

[33] M. Falxa, S. Babak, and M. L. Jeune
Adaptive Kernel Density Estimation proposal in gravitational wave data analysis
2022.

[34] B. Farr, W. Farr, D. Rudd, A. Price-Whelan, and D. Macleod
kombine: Kernel-density-based parallel ensemble sampler
Astrophysics Source Code Library, record ascl:2004.010. Apr. 2020.

[35] A. Gelman, W. R. Gilks, and G. O. Roberts
"Weak convergence and optimal scaling of random walk Metropolis algorithms"
The Annals of Applied Probability 7.1 (1997), pp. 110 –120.

[36] A. Gelman and D. B. Rubin
"Inference from Iterative Simulation Using Multiple Sequences"
Statistical Science 7.4 (1992), pp. 457 –472.

[37] B. Goncharov, D. J. Reardon, R. M. Shannon, X.-J. Zhu, E. Thrane, M Bailes, N. D. R. Bhat, S Dai, G Hobbs, M Kerr, R. N. Manchester, S Osłowski, A Parthasarathy, C. J. Russell, R Spiewak, N Thyagarajan, and J. B. Wang
"Identifying and mitigating noise sources in precision pulsar timing data sets"
Monthly Notices of the Royal Astronomical Society 502.1 (2020), pp. 478–493.

[38] B. Goncharov, R. M. Shannon, D. J. Reardon, G. Hobbs, A. Zic, M. Bailes, M. Curyło, S. Dai, M. Kerr, M. E. Lower, R. N. Manchester, R. Mandow, H. Middleton, M. T. Miles, A. Parthasarathy, E. Thrane, N. Thyagarajan, et al.
"On the Evidence for a Common-spectrum Process in the Search for the Nanohertz Gravitational-wave Background with the Parkes Pulsar Timing Array"
The Astrophysical Journal Letters 917.2 (2021), p. L19.

[39] H. Haario, E. Saksman, and J. Tamminen
"An Adaptive Metropolis Algorithm"
Bernoulli 7 (Apr. 2001).

[40] R. van Haasteren and Y. Levin
"Understanding and analysing time-correlated stochastic signals in pulsar timing"
Monthly Notices of the Royal Astronomical Society 428.2 (2012), pp. 1147–1159.

[41] R. van Haasteren and M. Vallisneri
"New advances in the Gaussian-process approach to pulsar-timing data analysis"
Physical Review D 90.10 (2014).

[42] W. K. Hastings
"Monte Carlo Sampling Methods Using Markov Chains and Their Applications"
Biometrika 57.1 (1970), pp. 97–109.

[43] H. Hausner
"General Relativity and Differential Geometry"
(2016).

[44] S. Hee, W. J. Handley, M. P. Hobson, and A. N. Lasenby
"Bayesian model selection without evidences: application to the dark energy equation-of-state"
Monthly Notices of the Royal Astronomical Society 455.3 (Nov. 2015), pp. 2461–2473.

[45] S. Hee, W. J. Handley, M. P. Hobson, and A. N. Lasenby
"Bayesian model selection without evidences: application to the dark energy equation-of-state"
Mon. Not. Roy. Astron. Soc. 455.3 (Jan. 2016), pp. 2461–2473.

[46] R. W. Hellings and G. S. Downs
"Upper limits on the isotropic gravitational radiation background from pulsar timing analysis."
ApJL 265 (Feb. 1983), pp. L39–L42.

[47] G. B. Hobbs, R. T. Edwards, and R. N. Manchester
"tempo2, a new pulsar-timing package - I. An overview"
Monthly Notices of the Royal Astronomical Society 369.2 (2006), pp. 655–672.

[48] G. Hobbs, S. Dai, R. N. Manchester, R. M. Shannon, M. Kerr, K.-J. Lee, and R.-X. Xu
"The role of FAST in pulsar timing arrays"
Research in Astronomy and Astrophysics 19.2 (2019), p. 020.

[49] D. W. Hogg and D. Foreman-Mackey
"Data Analysis Recipes: Using Markov Chain Monte Carlo"
The Astrophysical Journal Supplement Series 236.1 (2018), p. 11.

[50] C. J.T. B. Jasper A. Vrugt
"Differential Evolution Markov Chain with snooker updater and fewer chains"
Statistics and Computing 18.4 (Oct. 2008), pp. 435–446.

[51] Jeffrey
"The Theory of Probability"
(1998).

[52] F. A. Jenet and J. D. Romano
Understanding the gravitational-wave Hellings and Downs curve for pulsar timing arrays in terms of sound and electromagnetic waves
2014.

[53] Y. Jin, Y. He, and D. Huang
"An Improved Variable Kernel Density Estimator Based on L2 Regularization"
Mathematics 9.16 (2021).

[54] R. E. Kass and A. E. Raftery
"Bayes Factors"
Journal of the American Statistical Association 90.430 (1995), pp. 773–795.

[55] S. Kullback and R. A. Leibler
"On Information and Sufficiency"
The Annals of Mathematical Statistics 22.1 (1951), pp. 79 –86.

[56] L. Speri et al.
"Quality over Quantity: Optimizing pulsar timing array analysis for stochastic and deterministic gravitational wave signals"
Monthly Notices of the Royal Astronomical Society (2022).

[57] L. Lentati, R. M. Shannon, W. A. Coles, J. P. W. Verbiest, R. van Haasteren, J. A. Ellis, R. N. Caballero, R. N. Manchester, Z. Arzoumanian, S. Babak, C. G. Bassa, N. D. R. Bhat, P. Brem, M. Burgay, S. Burke-Spolaor, D. Champion, S. Chatterjee, et al.
"From spin noise to systematics: stochastic processes in the first International Pulsar Timing Array data release"
Monthly Notices of the Royal Astronomical Society 458.2 (2016), pp. 2161–2187.

[58] L. Lentati, P. Alexander, M. P. Hobson, F. Feroz, R. van Haasteren, K. J. Lee, and R. M. Shannon
"temponest: a Bayesian approach to pulsar timing analysis"
Monthly Notices of the Royal Astronomical Society 437.3 (Dec. 2013), pp. 3004–3023.

[59] T. B. Littenberg
"A detection pipeline for galactic binaries in LISA data"
Phys. Rev. D 84 (2011), p. 063009.

[60] D. R. Lorimer and M. Kramer
"Handbook of pulsar astronomy"
Handbook of Pulsar Astronomy (2012).

[61] J. B. MacQueen
"Some Methods for Classification and Analysis of MultiVariate Observations"
In: *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*. Ed. by L. M. L. Cam and J. Neyman. Vol. 1. University of California Press, 1967, pp. 281–297.

[62] M. Maggiore
Gravitational waves: Volume 1: Theory and experiments
OUP Oxford, 2007.

[63] M. Maggiore, C. V. D. Broeck, N. Bartolo, E. Belgacem, D. Bertacca, M. A. Bizouard, M. Branchesi, S. Clesse, S. Foffa, J. Garcí a-Bellido, S. Grimm, J. Harms, T. Hinderer, S. Matarrese, C. Palomba, M. Peloso, A. Ricciardone, et al.
"Science case for the Einstein telescope"
Journal of Cosmology and Astroparticle Physics 2020.03 (2020), pp. 050–050.

[64] M. Maiorano, F. D. Paolis, and A. A. Nucita
"Principles of Gravitational-Wave Detection with Pulsar Timing Arrays"
Symmetry 13.12 (2021), p. 2418.

[65] Manchester et al.
"The Parkes Pulsar Timing Array Project"
Publications of the Astronomical Society of Australia 30, e017 (2013), e017.

[66] R. N. Manchester, G. Hobbs, M. Bailes, W. A. Coles, W. van Straten, M. J. Keith, R. M. Shannon, N. D. R. Bhat, A. Brown, S. G. Burke-Spolaor, D. J. Champion, A. Chaudhary, R. T. Edwards, G. Hampson, A. W. Hotan, A. Jameson, F. A. Jenet, et al.
"The Parkes Pulsar Timing Array Project"
Publications of the Astronomical Society of Australia 30, e017 (Jan. 2013), e017.

[67] M. Milosavljevic
"The Final Parsec Problem"
In: *AIP Conference Proceedings*. AIP, 2003.

[68] NANOGrav Collaboration, Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. Chamberlin, S. Chatterjee, B. Christy, J. M. Cordes, N. Cornish, K. Crowter, P. B. Demorest, T. Dolch, J. A. Ellis, R. D. Ferdman, E. Fonseca, N. Garver-Daniels, M. E. Gonzalez, et al.
"The NANOGrav Nine-year Data Set: Observations, Arrival Time Measurements, and Analysis of 37 Millisecond Pulsars"
The Astrophysical Journal 813.1, 65 (Nov. 2015), p. 65.

[69] R. M. Neal
"Slice sampling"
The Annals of Statistics 31.3 (2003), pp. 705 –767.

[70] F. Nielsen
"On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means"
Entropy 21.5 (2019).

[71] E. Parzen
"On Estimation of a Probability Density Function and Mode"
The Annals of Mathematical Statistics 33.3 (1962), pp. 1065 –1076.

[72] B. B. P. Perera et al.
"The International Pulsar Timing Array: second data release"
Monthly Notices of the Royal Astronomical Society 490.4 (2019), pp. 4666–4687.

[73] A. Petiteau, S. Babak, A. Sesana, and M. de Araújo
"Resolving multiple supermassive black hole binaries with pulsar timing arrays. II. Genetic algorithm implementation"
Phys. Rev. D 87 (6 2013), p. 064036.

[74] G. Roberts and J. Rosenthal
"Examples of adaptive MCMC"
Journal of Computational and Graphical Statistics - J COMPUT GRAPH STAT 18 (June 2009).

[75] E. Roebber
"Ephemeris Errors and the Gravitational-wave Signal: Harmonic Mode Coupling in Pulsar Timing Array Searches"
The Astrophysical Journal 876.1 (2019), p. 55.

[76] 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 451.3 (2015), pp. 2417–2433.

[77] M. Rosenblatt
"Remarks on Some Nonparametric Estimates of a Density Function"
The Annals of Mathematical Statistics 27.3 (1956), pp. 832 –837.

[78] S Chen, R. N. Caballero, Y. J. Guo, A Chalumeau, K Liu, G Shaifullah, K. J. Lee, S Babak, G Desvignes, A Parthasarathy, H Hu, E van der Wateren, J Antoniadis, A.-S. B. Nielsen, C. G. Bassa, A Berthereau, M Burgay, et al.
"Common-red-signal analysis with 24-yr high-precision timing of the European Pulsar Timing Array: inferences in the stochastic gravitational-wave background search"
Monthly Notices of the Royal Astronomical Society 508.4 (2021), pp. 4970–4993.

[79] M. Sazhin
"Opportunities for detecting ultralong gravitational waves"
Soviet Astronomy 22 (1978), pp. 36–38.

[80] B. Schutz
A first course in general relativity
Cambridge university press, 2022.

[81] K. Schwarzschild
"On the gravitational field of a mass point according to Einstein's theory"
(1999).

[82] D. Scott and S. Sain
"Multi-dimensional density estimation"
Data Mining and Data Visualization 24 (Jan. 2005).

[83] A. Sesana and A. Vecchio
"Measuring the parameters of massive black hole binary systems with pulsar timing array observations of gravitational waves"
Physical Review D 81.10 (2010).

[84] R. M. Shannon, S. Osłowski, S. Dai, M. Bailes, G. Hobbs, R. N. Manchester, W. van Straten, C. A. Raithel, V. Ravi, L. Toomey, N. D. R. Bhat, S. Burke-Spolaor, W. A. Coles, M. J. Keith, M. Kerr, Y. Levin, J. M. Sarkissian, et al.
"Limitations in timing precision due to single-pulse shape variability in millisecond pulsars"
Monthly Notices of the Royal Astronomical Society 443.2 (July 2014), pp. 1463–1481.

[85] R. M. Shannon and J. M. Cordes
"ASSESSING THE ROLE OF SPIN NOISE IN THE PRECISION TIMING OF MILLISECOND PULSARS"
The Astrophysical Journal 725.2 (2010), pp. 1607–1619.

[86] D. Sivia and J. Skilling
Data analysis: a Bayesian tutorial
OUP Oxford, 2006.

[87] B. Stappers, E. Keane, M. Kramer, A. Possenti, and I. Stairs
"The prospects of pulsar timing with new-generation radio telescopes and the Square Kilometre Array"
Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences 376 (May 2018), p. 20170293.

[88] P. Tarafdar, N. K., P. Rana, J. Singha, M. A. Krishnakumar, B. C. Joshi, A. K. Paladi, N. Kolhe, N. D. Batra, N. Agarwal, A. Bathula, S. Dandapat, S. Desai, L. Dey, S. Hisano, P. Ingale, R. Kato, et al.
The Indian Pulsar Timing Array: First data release
2022.

[89] J. H. Taylor
"Pulsar Timing and Relativistic Gravity"
Philosophical Transactions: Physical Sciences and Engineering 341.1660 (1992), pp. 117–134.

[90] S. R. Taylor, M. Vallisneri, J. A. Ellis, C. M. F. Mingarelli, T. J. W. Lazio, and R. van Haasteren
"Are we there yet ? Time to detection of nanohertz gravitational wave based on pulsar timing array limits"
The Astrophysical Journal 819.1 (2016), p. L6.

[91] S. R. Taylor
The Nanohertz Gravitational Wave Astronomer
2021.

[92] S. R. Taylor, P. T. Baker, J. S. Hazboun, J. Simon, and S. J. Vigeland
enterprise extensions
v2.3.3. 2021.

[93] C. J. F. Ter Braak
"A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces"
Statistics and Computing 16.3 (Sept. 2006), pp. 239–249.

[94] G. R. Terrell and D. W. Scott
"Variable Kernel Density Estimation"
The Annals of Statistics 20.3 (1992), pp. 1236 –1265.

[95] K. S. Thorne, C. W. Misner, and J. A. Wheeler
Gravitation
Freeman San Francisco, CA, 2000.

[96] C. Tiburzi, G. Hobbs, M. Kerr, W. A. Coles, S. Dai, R. N. Manchester, A. Possenti, R. M. Shannon, and X. P. You
"A study of spatial correlations in pulsar timing array data"
MNRAS 455.4 (Feb. 2016), pp. 4339–4350.

[97] A. Tiengo, R. Mignani, A. D. Luca, P. Esposito, S. Mereghetti, A. Pellizzoni, M. Burgay, N. D'Amico, P. Esposito, A. Pellizzoni, and A. Possenti
"The first deep X-ray and optical observations of the closest isolated radio pulsar"
In: *AIP Conference Proceedings*. AIP, 2011.

[98] M. Vallisneri et al.
"Modeling the Uncertainties of Solar System Ephemerides for Robust Gravitational-wave Searches with Pulsar-timing Arrays"
The Astrophysical Journal 893.2 (2020), p. 112.

[99] J. P. W. Verbiest and G. M. Shaifullah
"Measurement uncertainty in pulsar timing array experiments"
Classical and Quantum Gravity 35.13, 133001 (July 2018), p. 133001.

[100] M. Volonteri
"The Formation and Evolution of Massive Black Holes"
Science 337.6094 (2012), pp. 544–547.

[101] M. Volonteri
"Formation of supermassive black holes"
The Astronomy and Astrophysics Review 18.3 (2010), pp. 279–315.

[102] W. D. Vousden, W. M. Farr, and I. Mandel
"Dynamic temperature selection for parallel tempering in Markov chain Monte Carlo simulations"
mnras 455.2 (Jan. 2016), pp. 1919–1937.

[103] S. S. Wilks
"The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses"
The Annals of Mathematical Statistics 9.1 (1938), pp. 60 –62.

[104] X. J. Zhu, G. Hobbs, L. Wen, W. A. Coles, J. B. Wang, R. M. Shannon, R. N. Manchester, M. Bailes, N. D. R. Bhat, S. Burke-Spolaor, S. Dai, M. J. Keith, M. Kerr, Y. Levin, D. R. Madison, S. Osłowski, V. Ravi, et al.
"An all-sky search for continuous gravitational waves in the Parkes Pulsar Timing Array data set"
MNRAS 444.4 (Nov. 2014), pp. 3709–3720.