Compact object coalescences and gamma-ray bursts in the gravitational-wave era

Raphaël Duque

Coalescences d'objets compacts et sursauts gamma à l'ère des ondes gravitationnelles,

thèse de doctorat présentée par

M. Raphaël Duque

préparée

à l'Institut d'astrophysique de Paris

sous la direction de

Prof. Frédéric Daigne et Dr. Robert Mochkovitch

en vue d'obtenir le grade de

Docteur ès sciences de Sorbonne Université,
spécialité Astronomie et astrophysique.

Présentée et soutenue publiquement le 10 septembre 2021, devant un jury composé de :

Marie-Christine Angonin	Professeur des universités – Sorbonne Université	Présidente du jury
Guillaume Dubus	Directeur de recherche – Centre national de la recherche scientifique	Rapporteur
Giancarlo Ghirlanda	*Primo Ricercatore – Istituto Nazionale di Astrofisica*	Rapporteur
Samaya Nissanke	*Associate Professor – University of Amsterdam*	Examinatrice
Shiho Kobayashi	*Professor – Liverpool John Moores University*	Examinateur
Frédéric Daigne	Professeur des universités – Sorbonne Université	Directeur de thèse
Robert Mochkovitch	Directeur de recherche – Centre national de la recherche scientifique	Co-directeur de thèse

Présente version compilée le 1er octobre 2021.

Compact Object Coalescences and Gamma-Ray Bursts in the Gravitational-Wave Era,

a PhD thesis presented by

Mr. Raphaël Duque

after study at the

Institut d'astrophysique de Paris

under the supervision of

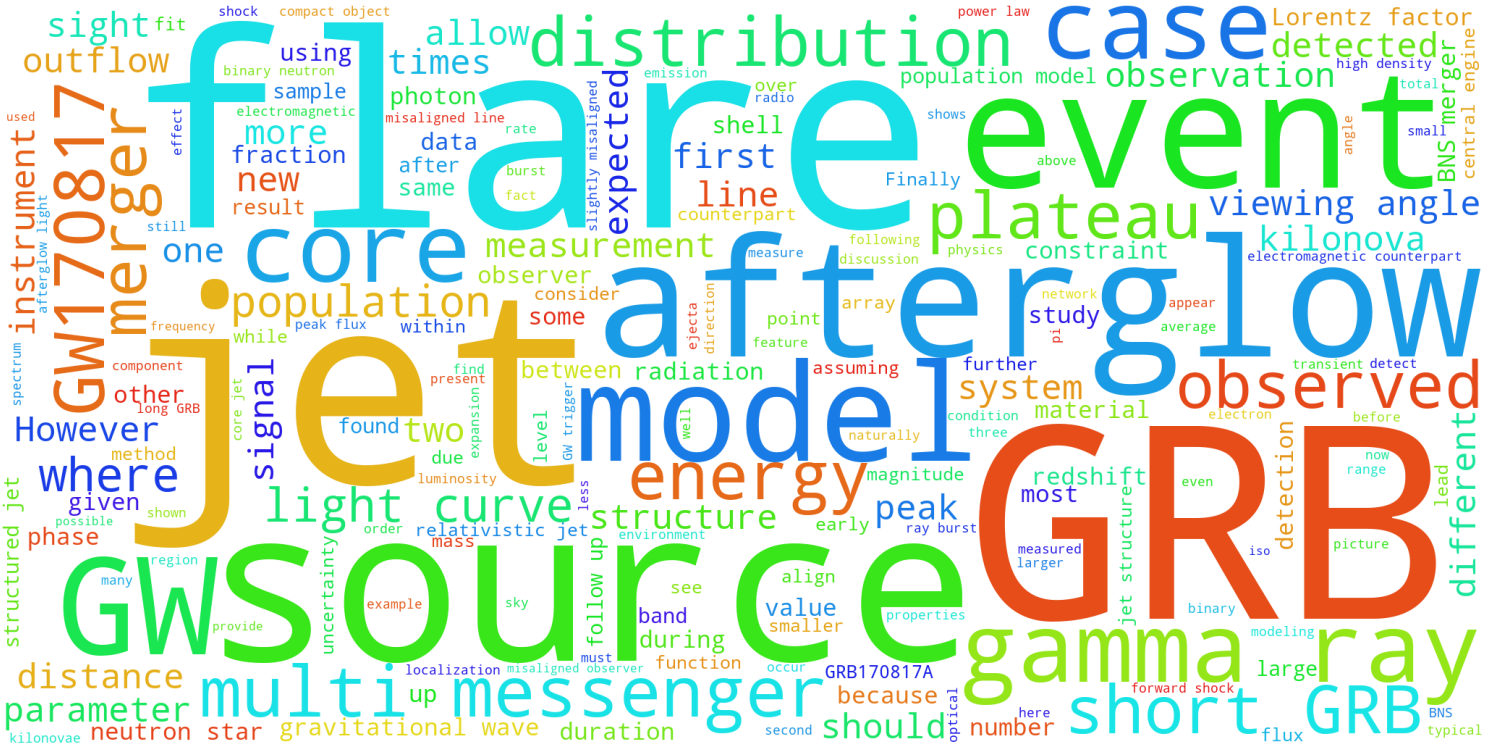
Prof. Frédéric Daigne and Dr. Robert Mochkovitch

Qu'il y ait des luminaires au firmament du ciel, pour séparer le jour et la nuit ;
qu'ils servent de signes pour marquer les fêtes, les jours et les années ;
et qu'ils soient, au firmament du ciel, des luminaires pour éclairer la terre.

Gn 1, 14-15

Let there be lights in the dome of the sky, to separate day from night;
Let them mark the fixed times, the days and the years,
and serve as luminaries in the dome of the sky, to shed light upon the earth.

Gen. 1.14-15



Nuage lexical généré avec `WordCloud for Python` (Mueller, 2020).

Remerciements

Une thèse de doctorat est comme une longue croisière. En fait, une traversée. On part dans l'inconnu, et au soir du premier jour en mer, on réalise que la réussite tient beaucoup à son voilier. Une coque pour l'abri, une quille pour la stabilité, un gréement pour toucher le vent ; voilà tout. À présent que cette traversée est finie, je souhaiterais remercier chaleureusement mes parents de m'avoir façonné avec amour, mon frère et ma sœur d'avoir grandi avec moi jusqu'à ce jour et de m'avoir accompagné au bord de la mer. On ne peut pas dessiner meilleur voilier que celui que vous m'avez donné.

En mer, on tient à son compas comme à ses vivres. Il permet de garder le cap, d'estimer la dérive et d'anticiper le prochain bord. Je remercie du fond du cœur mes directeurs de thèse Robert et Frédéric d'avoir encadré mes travaux avec bienveillance et patience ; de m'avoir enseigné la navigation et, mieux, la vie en mer. Je remercie également mon parrain de thèse Pierre Guillard pour son écoute et ses conseils.

Au delà de ses connaissances, on ne fait route sereinement qu'en ayant confiance dans les instruments, les cartes et autres almanachs qui sous-tendent la navigation. Je tiens à remercier les directeurs MM. Bernardeau et Bouchet, les secrétariats et toutes les équipes administratives et techniques qui garantissent le fonctionnement fluide de l'Institut d'astrophysique de Paris. De même, Thierry Fouchet et l'École doctorale « Astronomie et astrophysique d'Île-de-France ».

En cours de traversée, on rencontre d'autres voiliers. Pour certains, c'est aussi la première fois en mer. Souvent ils ont pris d'autres routes vers des eaux lointaines. Heureusement, la radio permet des contacts réguliers qui nourrissent notre culture générale de la mer, et fournissent un réconfort dans le gros temps. Je remercie les autres doctorants de ma génération et de la suivante pour leur amitié—en particulier Jesse, Claire, Valentin, Clément, François, Simon, Jean-Baptiste et Valentina. On en croise aussi d'autres plus expérimentés. Naviguer quelques temps bord à bord avec eux est une vraie leçon, au delà de l'exploration que l'on mène ensemble. Je pense à mes collègues et collaborateurs Paz, Éric, Simone, Irina et tous ceux avec qui j'ai partagé le bonheur de la recherche scientifique ces dernières années.

À l'approche de la destination, on regarde vers le sillage, on met les livres de bord en ordre et les carnets de voyage au propre. Sur ce dernier long bord, on réalise les mille ports visités, livres lus, et étoiles comptées les nuits sans vent. Valdo, Samuel et Jérémy, qui ont appareillé d'autres rivages ; ma belle famille, la campagne au bord de la mer ; Théo et Marcouf, ports animés ; Colette, île d'un autre temps ; Jérôme, Emmanuelle et François, mouillages paisibles.

Une fois arrivé, on est accueilli par des navires qui ont de l'histoire, parmi ceux qui ont découvert les parages que l'on a fréquentés. On leur confie journaux de bord et carnets de voyages, nos découvertes et vues nouvelles sur la mer. Alors, ils peuvent déterminer si c'est sûr pour les autres et pour soi que l'on retourne en mer. Je remercie respectueusement les deux rapporteurs de ce manuscrit pour leur relecture attentive et l'ensemble du jury pour le temps consacré à la revue critique de mes travaux.

Enfin, même après cette route je ne sais si c'est elle le vent, les vagues, la Lune, le Soleil ; peut-être tout à la fois. Cette force qui échappe aux cartes, aux compas et aux livres de bord mais dont on sait que, sans elle, tout cela n'aurait pas de sens. Sans doute la mer elle-même, mon épouse Joséphine.

Acknowledgments

A doctoral thesis is like a long cruise. In fact, a crossing. You set out into the unknown, and on the evening of the first day at sea, you realize that success is tied to the quality of your boat. A hull for shelter, a keel for stability, a rig to touch the wind; that's all. Now that this crossing is over, I would like to warmly thank my parents for having shaped me with love, my brother and sister for having grown up by my sides up to this day and accompanied me to the seaside. There is no ship greater than the one you have given me.

At sea, you hold on to your compass as strongly as to your food and water. It allows to keep the course, to measure the drift and to anticipate the next tack. I thank from the bottom of my heart my thesis advisors Robert and Frédéric for their benevolence and patience, for having taught me the art of navigation and, better still, life at sea. I would also like to thank my parrain de thèse *Pierre Guillard for his insightful advice year after year.*

Beyond your know-how, you can only sail serenely if you trust your instruments, charts and other almanacs that support your navigation. I would like to thank the directors Mr. Bernardeau and Mr. Bouchet, the secretariats and all the administrative and technical teams who allow a smooth functioning of the Institut d'astrophysique de Paris. *Likewise, Thierry Fouchet and the* École doctorale « Astronomie et astrophysique d'Île-de-France ».

During the crossing, you meet other sailors. For some it is also the first time at sea. Often they have taken other routes to distant waters. Fortunately, the radio allows for regular contact providing a general culture of the sea and comfort in heavy weather. I thank the other PhD students of my generation and the next for their friendship—in particular Jesse, Claire, Valentin, Clément, François, Simon, Jean-Baptiste and Valentina. You also meet others with more experience. Sailing beside them for a while is a real lesson, beyond the explorations we accomplish together. I would like to mention my colleagues and collaborators Paz, Éric, Simone, Irina and all those with whom I have shared my scientific reflection these last years.

As you approach the destination, you look back to the wake, you put the log books and travel notes in order. On this last long tack, you realize the thousand harbors visited, books read, and stars counted on windless nights. Valdo, Samuel and Jérémy, who sailed off other shores; my in-laws, the countryside by the sea; Théo and Marcouf, lively harbors; Colette, an island from another time; Jérôme, Emmanuelle and François, peaceful anchorages.

Once you reach the other side, you are welcomed by ships that have history, among those that discovered the waters you visited. You submit to them logbooks, discoveries and new views on the sea. They can then determine whether it is sound for the other seafarers and for you that you return to sea. I respectfully thank the two referees of this thesis manuscript for their attentive reading and the entire jury for the time devoted to the critical review of my work.

Finally, even after this sail I do not know if she is the wind, the waves, the Moon, the Sun; perhaps all at once. This force that does not fit into maps, compasses and logbooks, but without which we know that all this would make no sense. Probably the sea itself, my wife Joséphine.

Résumé court

L'ère de l'astronomie multi-messagers avec ondes gravitationnelles a débuté avec l'observation du signal gravitationnel provenant de la coalescence d'une binaire d'étoiles à neutrons, dénommée GW170817. Ce signal fut suivi d'un effort astronomique historique qui a révélé un riche ensemble de contreparties électromagnétiques : un sursaut gamma court faible, un signal transitoire thermique de kilonova, une rémanence multi-longueurs d'onde de longue durée. Cet événement et son ensemble de données multi-messagers ont permis des avancées majeures en astrophysique, et l'ère des ondes gravitationnelles est forte de promesses avec les prochaines campagnes d'observation de signaux gravitationnels et un suivi électromagnétique de plus en plus efficace et mieux organisé. Du point de vue de la science des sursauts gamma, cet événement a permis de faire la première étude fine de la structure des jets relativistes qui sont lancés par les coalescences d'objets compacts et dans lesquels certains sursauts gamma sont produits. Cette thèse de doctorat présente des travaux commencés peu après GW170817 qui explorent trois pistes de recherche motivées par le début de l'ère multi-messagers. Premièrement, nous établissons des perspectives de population pour les contreparties électromagnétiques des signaux gravitationnels de coalescence d'objets compacts à venir dans l'ère des ondes gravitationnelles. Ensuite, nous analysons les perspectives d'exploitation de ces événements pour mesurer la constante de Hubble et étudier les environnements des coalescences d'objets compacts. Enfin, nous revenons sur les leçons de GW170817 sur les jets relativistes pour explorer les conséquences possibles de la structure des jets sur les rémanences des sursauts gamma. Dans l'ensemble, le travail de doctorat présenté ici peut être vu comme une série de premiers pas dans l'étude des coalescences d'objets compacts et des sursauts gamma dans le contexte naissant de l'astronomie multi-messagers. Il ouvre de nombreuses perspectives d'études futures, que nous décrivons pour finir.

Abstract

The era of multi-messenger astronomy with gravitational waves began with the observation of the gravitational-wave signal from the merger of a binary neutron star merger, dubbed GW170817. This signal was followed by a historic astronomical effort that revealed a rich set of electromagnetic counterparts: a weak short gamma-ray burst, a thermal kilonova transient, a long-lived multi-wavelength afterglow. This event and its multi-messenger dataset afforded major breakthroughs in astrophysics, and the gravitational-wave era holds many promises with upcoming campaigns to observe further gravitational-wave signals from compact object mergers and increasingly efficient and better-organized electromagnetic follow-up. From the point of view of gamma-ray burst science, this event allowed unique insight into the structure of the relativistic jets that are launched from the mergers of compact objects and in which some gamma-ray bursts are produced. This doctoral thesis presents work started shortly after GW170817 that explores three avenues of research motivated by the inauguration of the multi-messenger era. First, we make population prospects for the electromagnetic counterparts to the gravitational-wave signals upcoming in the multi-messenger era. Second, we analyze the prospects of leveraging these events to measure the Hubble constant and study the environments of compact object mergers. Finally, we return to the lessons of GW170817 on relativistic jets to explore the possible consequences of jet structure on the observation of gamma-ray bursts. Overall, the doctoral work presented here should be understood as a series of first steps into the study of compact object mergers and gamma-ray bursts in the nascent context of multi-messenger astronomy. It opens up many prospects for future study, which are also described.

Résumé étendu

Les sursauts gamma sont de brefs mais intenses émissions de lumière gamma d'origine cosmologique (Chap. 1). Ces phénomènes transitoires sont composés d'une première phase dite « prompte » dans le domaine gamma et hautement variable, et d'une seconde phase dite « rémanente » plus durable et observée des bandes radio aux rayons X. L'émission prompte est due à la dissipation de l'énergie interne, cinétique et magnétique contenue dans le jet relativiste lancé par la source compacte à l'origine du sursaut gamma. L'émission rémanente est associée au rayonnement émis au niveau du choc fort formé lorsque le jet décélère au contact du milieu environnant la source. La classe des sursauts gamma dits « courts » – de durée courte – est issue des coalescences d'objets compacts, en particulier les binaires d'étoiles à neutrons (BNS). Au cours de leur dernières orbites avant de coalescer, ces dernières sont de puissantes sources d'ondes gravitationnelles, que l'on peut détecter depuis peu grâce aux interféromètres gravitationnels (Chap. 2). Que les coalescences d'objets compacts soient des sources à la fois d'ondes lumineuses et gravitationnelles ouvre la perspective de l'étude multi-messagers de ce phénomène. En 2017, cette perspective s'est réalisée, avec le signal d'ondes gravitationnelles dénommé GW170817 issu d'une coalescence d'étoiles à neutrons et doté de nombreuses contreparties électromagnétiques (Chap. 3). Cet événement a présenté un signal gravitationnel de phase spiralante de la binaire, un sursaut gamma court et faible, une émission thermique transitoire nommée kilonova, une rémanence multi-longueurs d'onde de très longue durée, et pourrait actuellement montrer les signes d'une rémanence de kilonova. L'exploitation combinée de ces signaux a permis de réaliser des progrès révolutionnaires en astrophysique. En particulier, une campagne d'observation de la source avec de l'interférométrie radio à très longue base (VLBI) a permis de sonder en détails la structure de l'écoulement provenant de la fusion des étoiles, révélant un jet relativiste structuré. Au vu de cet événement, les questions suivantes sont légitimes : Quelles sont les perspectives de nouvelles observations de contreparties lumineuses des signaux d'ondes gravitationnelles de coalescences d'objets compacts ? Quelles nouvelles méthodes utilisant la complémentarité de tous ces messagers pouvons-nous concevoir pour exploiter ces événements à venir ? L'événement historique GW170817 a prouvé que les coalescences de BNS – et très probablement d'autres sursauts gamma – lancent des jets relativistes structurés ; quelles sont les conséquences de cette structure sur l'interprétation des observations des sursauts gamma ?

Afin de répondre à ces questions, nous développons un modèle de population multi-messagers couvrant toutes les contreparties électromagnétiques mentionnées ci-dessus (Chap. 4, à l'exception de la rémanence de kilonova). Alors que la physique des ondes gravitationnelles et des sursauts gamma permettent une modélisation détaillée du signal de phase spiralante et de la rémanence, la nouveauté des kilonovas comme objets astronomiques ne permet qu'une modélisation assez incertaine de celles-ci ; néanmoins, notre étude est fondée sur une modélisation à jour des kilonovas étalonnée avec les observations de GW170817. De même, l'incertitude latente sur la physique des sursauts courts observés loin de l'axe du jet – comme GRB170817A – ne nous permet de considérer que les sursauts brillants dans notre modèle, c'est-à-dire ceux observés le long ou près de l'axe de leur jet. En ce qui concerne le modèle de population à proprement parler, nous utilisons des prescriptions de paramètres motivées par les contraintes déduites des observations des sursauts gamma courts pour tous les paramètres physiques, en particulier l'énergie du jet. Pour cette dernière, la distribution de probabilité est définie à partir des fonctions de luminosité des sursauts gamma courts. Le modèle de détection que nous adoptons permet d'étudier comment la population attendue de sources évolue avec la sensibilité des instruments gravitationnels et électromagnétiques et donc de replacer nos résultats dans le contexte du développement instrumental futur et des stratégies d'observation. Notre modèle de population est basé sur des critères de flux des contreparties électromagnétiques. De ce fait il porte sur la population de sources détectables par les instruments de suivi des alertes gravitationnelles. Celle-ci est distincte de la

population des sources qui seront effectivement détectées lors des futures campagnes, du fait des limitations dans la couverture du ciel par les efforts de suivi et la difficulté à déterminer la source liée à l'alerte parmi toutes les sources variables ou transitoires identifiées.

Parmi les résultats principaux de cette étude de population (Chap. 5), nous constatons que la communauté a été très chanceuse d'observer l'événement GW170817 : il était extrêmement improbable dans les conditions instrumentales de sa détection d'observer une telle pléthore de contreparties électromagnétiques. Au fur et à mesure que l'horizon gravitationnel s'éloigne grâce au développement instrumental, davantage de détections de fusion d'objets compacts dans le domaine gravitationnel sont attendues. Cependant, celles-ci devraient être suivies par des contreparties électromagnétiques dans une fraction de plus en plus faible de cas, à moins que la sensibilité du domaine électromagnétique n'augmente fortement, ce qui représente un défi pour la communauté astronomique. Naturellement, le nombre total d'événements dont les contreparties sont détectables augmente avec le temps. Les sources des détections gravitationnelles s'éloignant de plus en plus, les premiers signaux électromagnétiques dont la détection devient extrêmement improbable sont le sursaut gamma court – sans tenir compte du cas rare d'un système aligné, qui, lui, serait observable à distance cosmologique – et le déplacement angulaire du jet sur le ciel mesuré par l'imagerie VLBI. À la sensibilité nominale du réseau mondial d'interféromètres gravitationnels, seule la kilonova devrait rester accessible dans une fraction significative (environ 50 %) des 100 détections gravitationnelles annuelles de BNS. La courbe de lumière de la rémanence du jet devrait être détectable dans environ un tiers de ces événements dotés de kilonovas, mais avec de la patience : la plupart de ces rémanences devraient émerger au-dessus du seuil radio seulement quelques centaines de jours après le signal gravitationnel. Bien qu'elles soient accessibles aux instruments de suivi, la détection de ces contreparties a constitué un défi au cours de la récente campagne d'observation gravitationnelle « O3 » de la collaboration LIGO-Virgo-Kagra, pendant laquelle un autre signal de fusion de BNS et deux signaux de fusion de binaires étoile à neutron-trou noir ont été détectés. Cette difficulté provient des trop grandes incertitudes sur la position de la source dans le ciel fournie par les données gravitationnelles. Ceci est empiré par le nombre considérable de phénomènes transitoires potentiellement associés à classer et à suivre dans la zone d'incertitude. La limitation due aux grandes cartes de positionnement devrait être dépassée à la fois par la mise en service d'un plus grand nombre d'interféromètres gravitationnels – fournissant des cartes de localisation plus petites – et par l'arrivée d'instruments optiques sensibles à grand champ et à haute cadence dans les deux hémisphères, tels que le ZTF et le LSST. De plus, nous constatons que ces instruments pourraient sonder une population nombreuse de kilonovas orphelines (c'est-a-dire non détectées comme contreparties d'un signal gravitationnel) et une population d'associations de kilonovas avec des sursauts gamma brillants. Ces sources pourraient jouer un rôle important dans l'amélioration des modèles de kilonova et dans l'évaluation plus approfondie du lien entre sursauts courts et coalescences d'objets compacts.

Bien que rares, les rémanences de ces événements devraient être extrêmement intéressantes, car elles devraient être le plus souvent observées avec un fort angle de vue par rapport à l'axe du jet, comme GW170817. Cette configuration géométrique permet une étude détaillée de la structure du jet et des paramètres du système et suggère donc des applications intéressantes pour mesurer la constante de Hubble (Chap. 6) et pour étudier les environnements des fusions de BNS (Chap. 7).

Les signaux gravitationnels des coalescences d'objets compacts sont fondamentalement différents des signaux électromagnétiques car la distance de la source peut être mesurée directement à partir de la forme d'onde. Dans de telles mesures, la principale dégénérescence dans les paramètres du système est entre la distance et l'angle d'inclinaison orbitale de la binaire. Lorsque l'on tente de mesurer la constante de Hubble sur un système donné en utilisant le décalage vers le rouge de la source acquise de sa galaxie hôte par une contrepartie de kilonova, cette dégénérescence intrinsèque aux données gravitationnelles limite fortement la précision de la mesure. Heureusement, l'angle d'inclinaison du système est une information que l'on peut également obtenir grâce à l'émission rémanente du jet. Ceci suggère que les rémanences de jets sont un moyen de lever la dégénérescence distance-inclinaison et d'améliorer les mesures de la constante de Hubble ; c'était le cas de GW170817, où l'apport des données de la rémanence a permis de tripler la précision de la mesure de la constante de Hubble. Toutefois, les rémanences étant rares, il est peu probable qu'elles soient détectées dans la plupart des fusions déclenchant une détection gravitationnelle. Dans le Chap. 6, nous montrons que, bien que chaque rémanence fournisse une mesure très améliorée de la constante de Hubble, la rareté de ces contreparties compense largement, de sorte que, dans l'ensemble, les rémanences ne permettront pas de contraindre la constante plus rapidement qu'avec les seules données gravitationnelles et les décalages vers

le rouge fournis par les kilonovas. Bien que cela disqualifie les rémanences pour les efforts de cosmologie multi-messagers, nous montrons que les kilonovas, plus fréquentes, pourraient jouer un rôle, à condition que les modèles s'améliorent au point qu'une seule courbe de lumière de kilonova permette de mesurer l'angle d'inclinaison avec une précision de dix degrés. Ce stade devrait être atteint grâce à l'accumulation d'un large échantillon de kilonovas avec différents angles de vue. Ensuite, il faut faire attention au traitement des effets de sélection lors de l'utilisation des kilonovas pour mesurer la constante de Hubble. En effet, nous montrons que ces effets ne sont pas négligeables dans le régime des décalages vers le rouge sondé par les configurations nominales des interféromètres gravitationnels.

En raison des longs délais entre la formation et la fusion des BNS, on s'attend généralement à ce que leurs coalescences – et les sursauts courts – se produisent dans des milieux peu denses, raréfiés. C'était le cas de GW170817, avec la densité du milieu autour de la coalescence mesurée par divers méthodes à $\lesssim 10^{-2}$ cm^{-3}. Au fil des années, un grand nombre de sondes indirectes de l'environnement des fusions de BNS se sont accumulées : l'analyse de l'échantillon encore assez réduit de rémanences des sursauts courts ; les grandes distances angulaires entre les sursauts courts et leurs galaxies hôtes ; les abondances des éléments chimiques issus du processus rapide de capture de neutrons dans divers environnements galactiques ; l'observation des systèmes binaires de notre Galaxie. Prises dans leur ensemble, ces sondes ne permettent pas de conclure à l'existence d'une population de fusions à haute densité, survenant dans des régions dont la densité est supérieure à 1 cm^{-3}. Bien que certaines études sur l'évolution binaire suggèrent que certains systèmes pourraient fusionner en des temps bien inférieurs à ceux généralement attendus (et donc peut-être dans des milieux de plus haute densité), l'existence d'une telle population de coalescences de BNS reste incertaine. Dans le Chap. 7, nous concevons une méthode multi-messagers basée sur les données gravitationnelles et l'observation de courbes de lumière de rémanences qui permet de contraindre la densité du milieu extérieur d'événements donnés. Alors que la modélisation de la rémanence des sursauts classiques produit une forte dégénérescence entre les paramètres du système – dont la densité –, la combinaison des données gravitationnelles avec le flux au pic de la rémanence et les informations sur son temps de pic permettent de contraindre suffisamment la densité pour classer les événements comme étant de faible ou de forte densité. Cette méthode permettrait d'étudier une population hypothétique d'événements à haute densité dans l'Univers local à l'aide des interféromètres actuels, et autour du pic de formation des étoiles avec des instruments de troisième génération. De plus, le flux de la rémanence dépend fortement de la densité, de sorte que, s'il existe effectivement une population de coalescences à haute densité, celles-ci devraient être sur-représentées dans les échantillons détectés, au point que seuls quelques événements devraient permettre de les révéler – ou, si aucun n'est observé, de les contraindre fortement. Dans cette méthode, c'est l'exigence de détecter le signal gravitationnel antérieur à la rémanence qui conduit à la fois à la mesure de la densité et à l'effet statistique selon lequel cette population, si elle existe, devrait être révélée tôt dans l'ère multi-messagers.

L'événement GW170817 a suscité un regain d'intérêt pour la structure des écoulements en jeu dans l'émission des sursauts gamma ; il a montré comment des jets structurés relativistes – un cœur de jet très énergétique et relativiste entouré de matière moins extrême – peuvent façonner la courbe de lumière de la rémanence observée à partir de lignes de visée considérablement éloignées par rapport à l'axe du jet. De telles rémanences portent la signature d'effets géométriques liés aux délais de propagation de la lumière entre le cœur et l'observateur et à l'ouverture retardée du cône de focalisation relativiste du cœur. Au Chap. 8, nous motivons l'étude des jets structurés et des jets relativistes en général dans divers contextes astrophysiques. Nous exposons leurs principales caractéristiques et les processus physiques qui façonnent ces écoulements. Dans le contexte des sursauts gamma, la structure de l'écoulement est principalement déterminée par l'interaction du jet lancé par le moteur central avec la couche de matière dense entourant celui-ci : l'enveloppe de l'étoile en effondrement pour les sursauts longs, les matériaux éjectés lors de la fusion pour les sursauts courts. Nous résumons les effets connus de la nature relativiste de ces jets sur la physique des sursauts. En particulier, nous présentons un argumentaire prouvant que les angles d'observation jusqu'auxquels les jets des sursauts gamma cosmologiques sont observés ne peuvent pas être plus grands qu'environ deux fois l'angle d'ouverture du cœur du jet : contrairement au jet de GW170817 – un événement très proche –, les jets classiques des sursauts gamma sont observés sur des lignes de visée tout au plus légèrement désaxées. Néanmoins, sur ces lignes de visée proches du cœur du jet, les effets relativistes et géométriques jouent toujours un rôle prédominant dans le modelage de l'émission prompte et de la rémanence des sursauts. Dans le cadre géométrique d'observateurs légèrement désaxés par rapport à des jets structurés, nous développons deux modèles pour expliquer chacun des phénomènes récurrents des rémanences de sursauts

gamma : les plateaux (de longues phases de flux quasi-constant observées dans les rémanences de certains sursauts, Chap. 9) et les éruptions (des regains soudains et transitoires de flux dans certaines rémanences, pouvant être de plusieurs ordres de grandeurs, Chap. 10).

Le premier modèle est en fait une amélioration d'une proposition faite peu après la découverte des plateaux en rayons X dans les rémanences des sursauts, motivée par l'observation historique de la structure du jet dans le cas de GW170817 et les leçons sur les jets structurés que l'on en a tirées. Ce modèle est basé sur le mécanisme physique suivant: pendant les premières phases de la rémanence, le jet central est trop éloigné pour que son rayonnement atteigne un observateur désaxé, du fait de la focalisation relativiste ; initialement, l'observateur ne voit que le rayonnement du matériau proche de sa ligne de visée et moins énergétique. Lorsque la structure décélère progressivement, le cône de focalisation du matériau de plus en plus proche du cœur s'ouvre pour inclure la ligne de visée de l'observateur. En raison de la structure du jet, la matière qui est ainsi progressivement révélée est plus énergétique intrinsèquement, mais moins amplifiée par effet Doppler. Ce mécanisme peut produire des phases longues de flux presque plat pendant la rémanence, c'est-à-dire un comportement de plateau. Une fois le cœur révélé à l'observateur, plus aucun matériau n'a encore à rejoindre la ligne de visée de l'observateur, et la courbe de lumière passe alors à une phase de décroissance ordinaire, comme pour un observateur aligné avec le jet. Au Chap. 9, nous explorons en détail ce modèle en montrant à la fois des courbes de lumière synthétiques présentant un comportement en plateau et des développements analytiques permettant de prédire la durée du plateau et son niveau de flux en fonction de la structure du jet et de la position de l'observateur. De plus, ces développements analytiques montrent que certaines corrélations observées entre les propriétés du plateau et de l'émission prompte dans les sursauts avec plateau sont des conséquences naturelles de notre interprétation géométrique. Notamment, les plateaux plus brillants ont une durée de vie plus courte et sont associés à une émission prompte plus brillante.

Motivés par le succès du modèle de plateau, et notant les tendances apparentes entre la présence de plateaux et d'éruptions dans les rémanences de sursauts gamma, nous proposons une nouvelle interprétation pour ces éruptions dans la même configuration géométrique : un observateur légèrement désaxé par rapport à un jet structuré. Dans le Chap. 10, nous suggérons que les éruptions de rémanence des rayons X sont en fait les signatures, pour des observateurs légèrement désaxés, d'une dissipation prompte se produisant dans le cœur du jet, alors que pour ces mêmes observateurs, l'émission prompte provient de la matière en aval de la ligne de visée, dans la structure latérale. Alors qu'un observateur aligné détecte les impulsions promptes du cœur du jet dans la bande des rayons gamma, les observateurs légèrement désaxés détectent cette émission dans les rayons X en raison d'un effet Doppler moins favorable ; ils les détectent naturellement plus tard en raison du temps de parcours de la lumière depuis le cœur. L'hypothèse d'une origine commune pour les éruptions et les impulsions promptes fournit naturellement les similitudes observées entre ces deux phénomènes tels que leurs profils temporels et le retard spectral. Les calculs d'ordre de grandeur montrent qu'une éruption typique peut effectivement être reproduite dans ce mécanisme en supposant que les énergies totales dissipées dans de tels épisodes sont cohérentes avec ce que nous savons de la dissipation prompte dans les sursauts gamma. Notre vision s'applique surtout à la classe des éruptions précoces se produisant moins de 1000 secondes après l'émission prompte, typiquement pendant la phase de décroissance rapide de la rémanence ou le plateau précoce. Ces éruptions précoces sont statistiquement dominantes dans la population et ont probablement une origine distincte des éruptions tardives. Une analyse fine montre que notre modèle favorise l'émergence d'éruptions au profils temporels piqués et suggère que les durées des éruptions devraient être étroitement distribuées, deux propriétés remarquables et effectivement observées dans les échantillons d'éruptions des rémanences X de sursauts gamma. Enfin, des ajustements de courbes de lumière aux données réelles montrent que le modèle est capable de reproduire les diverses morphologies des éruptions avec des valeurs de paramètres raisonnables et en maintenant une ligne de visée légèrement désaxée. Ces ajustements révèlent également la sensibilité du modèle à l'intervalle de temps pendant lequel on permet l'éjection des coquilles responsables des éruptions, c'est-à-dire la durée de l'activité du moteur central. De meilleurs ajustements sont trouvés en permettant une activité d'éjection dans le cœur du jet plus longue que la durée du sursaut parent, bien que beaucoup plus courte que le temps d'occurrence de l'éruption. La durée de l'activité du moteur central est liée à la physique à plus petite échelle autour de ce moteur, et la durée du sursaut n'est qu'une sous-estimation de cette durée. De plus, comme l'émission prompte de l'observateur mal aligné provient de la matière sur sa ligne de visée, la durée du sursaut n'est qu'une estimation de la durée de l'éjection dans cette direction, et non dans le cœur. Néanmoins, l'exigence apparente d'une activité d'éjection qui dure plus longtemps que le sursaut lui-même pose la question de la variation de l'activité d'éjection du

moteur d'une direction à une autre. Nous notons que nos deux modèles pour les plateaux et les éruptions sont établis dans la même configuration géométrique, et s'appliquent en principe à la fois aux sursauts longs et courts.

Le travail de doctorat présenté dans ce manuscrit explore de nombreuses pistes de recherche en astrophysique des hautes énergies ouvertes avec l'ère multi-messagers des ondes gravitationnelles : prédire les populations à venir d'événements multi-messagers, chercher comment tirer parti de ces événements au profit de la compréhension des environnements des coalescences de BNS ou de la cosmologie de précision, et explorer les conséquences des premières leçons de cette nouvelle ère sur la physique des sursauts gamma. Au Chap. 11, nous présentons plusieurs développements immédiats possibles pour poursuivre ces recherches. Nous proposons : l'extension de notre modèle de population à d'autres observables de rémanence, telles que la polarisation du signal radio ; la révision des mécanismes standard de sursaut gamma dans le contexte de jets relativistes structurés avec des lignes de visée désaxées ; la préparation de l'exploitation des rémanences légèrement relativistes pour étudier les environnements des coalescences en élucidant la structure de l'écoulement de kilonova ; la consolidation de notre suggestion d'utiliser des contreparties de rémanence pour étudier la distribution des délais entre la formation et la fusion des BNS ; le développement de notre interprétation géométrique des plateaux et des éruptions des sursauts gamma en étudiant l'activité simultanée plateau-éruption et en confrontant le modèle aux statistiques du phénomène. Ces propositions sont multidisciplinaires : physique classique de la rémanence des sursauts gamma, simulations hydrodynamiques, structures galactiques, inférence statistique, etc. Pour chacun des projets, nous cherchons à fournir les premières étapes pour entrer dans l'étude, les principaux points sensibles à considérer et les résultats auxquels on peut s'attendre. Naturellement, au delà de tous ces développements théoriques enthousiasmants, les résultats présentés dans ce manuscrit doivent être éclairés avant tout par de nouvelles observations multi-messagers, et nous prévoyons que le contexte instrumental des prochaines campagnes d'observation gravitationnelle le permettront.

Contents

Preface

The present doctoral thesis is an account of the work accomplished during the three years of my preparation to the PhD diploma—formally, from October 1st 2018 to September 30th 2021. I did this work under the supervision of Frédéric Daigne and Robert Mochkovitch within the "high-energy astrophysics and primordial universe" group of the *Institut d'astrophysique de Paris*. This three-year endeavor started nearly one year after the observation of GW170817, the binary neutron star merger that inaugurated the gravitational-wave multi-messenger era with its gravitational-wave signal and many electromagnetic counterparts. The era of multi-messenger astronomy with gravitational waves was well anticipated by the high-energy community globally, and this thesis should be seen as a record of the first steps our group took when this era finally came.

The three parts that compose the main text are all self-contained, as the potentially required information from other parts is summarized in the introductive chapters of each part. However, I recommend a linear reading of the chapters within a given part. Readers in a hurry can turn directly to Chap. 11, featuring a summary of the results of my doctoral work as well as proposals for future study.

The work presented here is based on a number of physics and astrophysics derivations, developed by our group during the doctoral studies or drawn from a number of courses, books or articles. The technical appendices are meant as a standalone compilation of these derivations such that the seasoned reader can know exactly which calculation or prescription we used to derive the results we present, and so that the newcomer to our group—or field in general—can find the essential physics required to understand our work tidied in a single place. I wish that these appendices avoid others the hassle of finding and connecting the many pieces of physics that were necessary to this work and to pursue it.

Margin notes beside section headers indicate text drawing from material published by myself in collaboration with others. In these notes, the reader will find the references to these articles I published placed after the following symbols: A '+' symbol means that the text draws from my published work with noteworthy additions; A '‖' symbol denotes text that is essentially the same as in the referenced source I published, up to language or reformulation; A '×' symbol means a synthesis was done with material from different articles I published, with possible additions. Similarly, when figures were reproduced from publications by myself with other collaborators, the relevant citation is made in square brackets at the end of the figure caption.

Part I

From gamma-ray bursts and gravitational waves to the multi-messenger era

Chapter 1

Gamma-ray bursts

Abstract

We describe gamma-ray bursts and provide the most important features of their prompt light curves and spectra. We briefly describe various gamma-ray detector technologies used in this field. Following a chronological approach, we progressively build the general picture of gamma-ray burst science, passing by the main theoretical milestones: cosmological origin, relativistic motion, jetted outflow, prompt and afterglow radiation phases. We describe the dichotomy between long and short gamma-ray bursts in terms of progenitors, astrophysical settings and observations. We provide a concise overview of gamma-ray burst phenomenology. We finish by opening up to the perspectives of the multi-messenger era for gamma-ray burst science through the description of the upcoming *SVOM* mission.

1.1 From gamma rays to gamma-ray bursts

Gamma-ray bursts (GRBs) are a class of non-repeating transient and spatially localized signals in the hard X-ray to gamma-ray band (Klebesadel et al., 1973; Piran, 2004). Because the Earth's atmosphere is opaque to the high-energy light composing GRBs, these transient signals are detected with space-borne instruments. They were therefore only discovered during the space age—in 1967—, and acknowledged as astrophysical phenomena in 1973. Because they can only be observed from space and due to their transient nature, it is a technical challenge to localize the source of a GRB in the sky. So much so that the distance scale and thus the energy scale of these objects—required to develop any astrophysical interpretation—were only firmly established in 1997. In this sense, the science of GRBs is a *young* branch of astrophysics.

Instrumental development progressively led to discovering GRB components in the high-energy (1 GeV, Abdo et al. 2009; Piron 2016) and then very-high-energy range (1 TeV, MAGIC Collaboration et al. 2019). Current instruments trigger on GRBs every one to two days, at positions uniformly distributed in the sky.

A specificity of gamma-ray astronomy is that instruments detect and count the arrival of individual high-energy photons. They can also measure their energies and possibly their incoming direction. The onset of a GRB manifests as a sudden increase in the detection rate of these photons, which can be identified by electronics aboard the satellite in a process known as the *trigger*. The collection of thus detected photons can be used to build the light curves and spectra of GRBs. Before delving onto the salient features of GRBs, we will briefly describe the high-energy photon detection technologies.

Depending on the energy range and localizing requirements, high-energy astronomical instruments can rely on various designs:

- A *gamma-ray detection module* is composed of a scintillator crystal made of heavy elements. Photons with energies from a few keV to tens of MeV—depending on the chemical composition and thickness—interact with the scintillator, producing lower energy photons picked up by photomultipliers which can measure their energy and, in turn, that of the incoming gamma ray. The response of such modules is not isotropic, such that comparing the photon counts in various modules aboard a same platform can provide a rough localization of the source ($\sim$ 5deg). These gamma-ray modules equip most GRB missions. Timing the GRB photons in modules across different satellites allows to triangulate the source and localize it much more precisely, depending on the available baselines between satellites. In Fig. 1.1 (top left), we show the placement of three gamma-ray detection modules (GRD A, B and C) aboard the platform of the *Space Variable Objects Monitor* mission (*SVOM*, launch scheduled in 2022). Together, these modules constitute the Gamma-Ray Monitor (GRM) of *SVOM*, its main GRB

detection instrument. For more information on *SVOM*, report to Sec. 1.5.

- For lower energy photons (less than a few hundred keV), the crystal can be organized as a detection plane composed of pixels that directly absorb incoming photons, measuring their arrival time and energy. Placing a *coded aperture mask* (i.e., a perforated opaque plane) in front of the detector plane allows to reconstruct the sky position of the source by analyzing the shadow cast by the mask onto the detector. Coded masks offer imaging capabilities over a large field of view with point-spread functions of tens of arcmin, at the expense of losing some source flux. In Fig. 1.1 (top right), we show the coded mask and detector array of the Burst Alert Telescope (BAT) aboard the *Neil Gehrels Swift Observatory* (*Swift*, launched in 2004). *SVOM* will also carry a coded-mask imaging instrument called ECLAIRs, figured in Fig. 1.1 (top left).

- *Compton scattering instruments* are effective for gamma-rays of energies of hundreds of keV to a few MeV. These instruments are composed of two layers of crystal serving as targets. Incoming photons Compton scatter on an electron in the first layer; The locus of the scattering and the electron's energy are measured by a series of photomultipliers monitoring the first layer. The scattered, softer photon is then absorbed by the second layer; Its energy is measured upon absorption. Combining the scattered photon and electron energies with the two loci of interaction allows to reconstruct the incoming photon's energy as well as its origin on a circle in the sky. Using many photons from a single source allows to pinpoint the source by triangulation. In Fig. 1.1 (bottom left), we depict the principle of Compton scattering instruments. A famous example was the Imaging Compton Telescope aboard the *Compton Gamma Ray Observatory* (COMPTEL aboard *CGRO*, on orbit 1991–2000). An advantage of this method is to allow to measure the polarization of the incoming photon.

- *Pair-production instruments* are closer to particle physics detectors; They are effective for photons from tens of MeV to hundreds of GeV. They are composed of many conversion foils interleaved with particle tracking layers. The incoming photon interacts with heavy nuclei in the conversion foils, producing an electron-positron pair. The silica-based trackers track the two particles in the layers down to a calorimeter at the bottom of the instrument. With the pair's trajectory and energy measured by the calorimeter, the initial photon's energy and incoming direction can be inferred. The Large Area Telescope (LAT) aboard the *Fermi Gamma-ray Space Telescope* (*Fermi*, launched 2008) is a pair-production instrument (Fig. 1.1, bottom right). It completes the Gamma-Ray Monitor (GBM, composed of twelve gamma-ray detection modules), which is the primary GRB-detecting instrument aboard *Fermi*.

- Finally, for gamma-rays with energies larger than hundreds of GeV, ground-based *imaging atmospheric Cherenkov telescopes* can be used. The atmosphere serves as a target for pair production by the incoming photon; The produced pair is itself very energetic, leading to a cascade of high-energy particles. This air shower is superluminal, such that Cherenkov light is emitted. This light is focused by telescopes onto photomultipliers, allowing to record the shower's pattern and, ultimately, to reconstruct the incoming photon's energy and direction. This inference is improved through stereoscopy by using an array of telescopes.

The capability to locate GRBs in the sky played a fundamental role in the topic's development, through the subsequent discovery of the *afterglow phase* of GRBs, which we will comment shortly. However, before the era of afterglows, GRB science only relied on the gamma-ray data proper, i.e., the set of gamma-ray photon arrival times and energies. We refer to this gamma-ray phase of GRBs as the *prompt phase*.

The *gamma-ray light curve* of a GRB is the time sequence of the rate of photon in a given energy band. In Fig. 1.2 we provide a number of gamma-ray transients detected by the Anti-Coincidence Shield aboard the *International Gamma-Ray Astrophysics Laboratory* (ACS, *INTEGRAL*, launched 2002, Rau et al. 2005). The ACS is meant to protect another instrument aboard *INTEGRAL* from charged particles and gamma-rays in some directions, but can be used as a nearly omni-directional gamma-ray detector for photons with energies $\gtrsim 75$ keV; It cannot determine the energies of the photons. All the GRBs listed in this figure were also detected by other missions. The photon events are binned in time intervals of 50 ms. This figure illustrates the diversity in GRB light curves as well as the very short variability timescales present in these signals. Indeed, we observe signals with a single smooth *pulse* (e.g., GRB041212[1]), a few such pulses (e.g., GRB040302), or highly structured events (e.g., GRB041211). Pulses are the primary structures of GRB light curves; They can be difficult to define in highly structured bursts. GRB041212 presents the prototypical *fast-rise-exponential-decay* (FRED) profile, that captures most GRB pulses (Norris et al., 2005); Other pulses can be much sharper (e.g., GRB030814). We also notice that GRBs are highly variable sources, with variability timescales often smaller than the binning time, and sometimes smaller than 10 ms (MacLachlan et al., 2013). Finally, we note the wide range in GRB durations, from sub-second events to tens of seconds.

Because of the difficulty to locate the source—even for GRBs detected by multiple instruments—, known repeating sources of gamma-rays sometimes cannot be excluded as the origin. For example, triangulation cannot exclude the soft gamma-ray repeater SGR 0525-66 as the origin of GRB030711. This poses the question of the definition of the sample of GRBs, to which we return in Sec. 1.4. The very bright and short gamma-

[1]GRBs are named according to their detection date in the format YYMMDD. In case of multiple detections, a letter is appended in the order of brightness of the GRBs. The first GRB detected was GRB670702, by the *Vela* satellites.

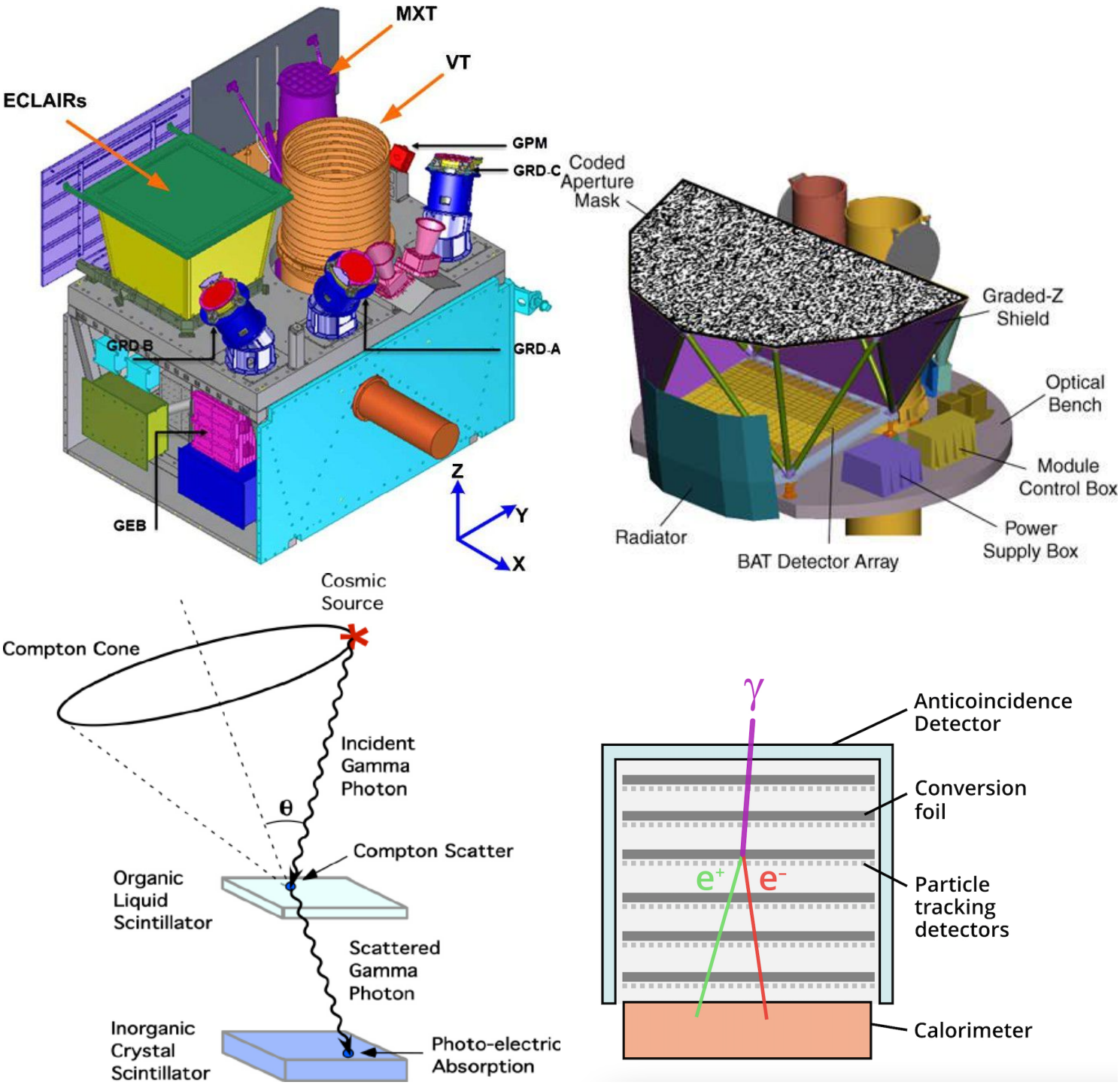


Figure 1.1: **Various gamma-ray detection technologies.** Top left: The *SVOM* platform featuring multiple instruments, among which the Gamma-Ray Monitor, composed of three gamma-ray detection modules (GRD A, B and C) and the coded-mask imager ECLAIRs (He et al., 2020). For more information on *SVOM*, see Sec. 1.5. Top right: The coded aperture mask and detection array of the Burst Alert Telescope aboard *Swift*. The shadow cast by the source on the detector plane allows imaging capabilities on a large field of view (credits: NASA Goddard Space Flight Center). Bottom left: Principle of Compton-scattering instruments. The information of the loci of Compton scattering and photo-absorption together with the scattered electron and photon energies allow to infer the incoming photon's energy and to place its incident direction on a circle (Compton cone, Bloser & Ryan 2008). Bottom right: Principle of a pair-creation instrument. The gamma-ray photon interacts in the conversion layers to create an electron-positron pair which is tracked and then detected and measured by the calorimeter, allowing to infer the incident photon's energy and direction (credits: NASA Goddard Space Flight Center).

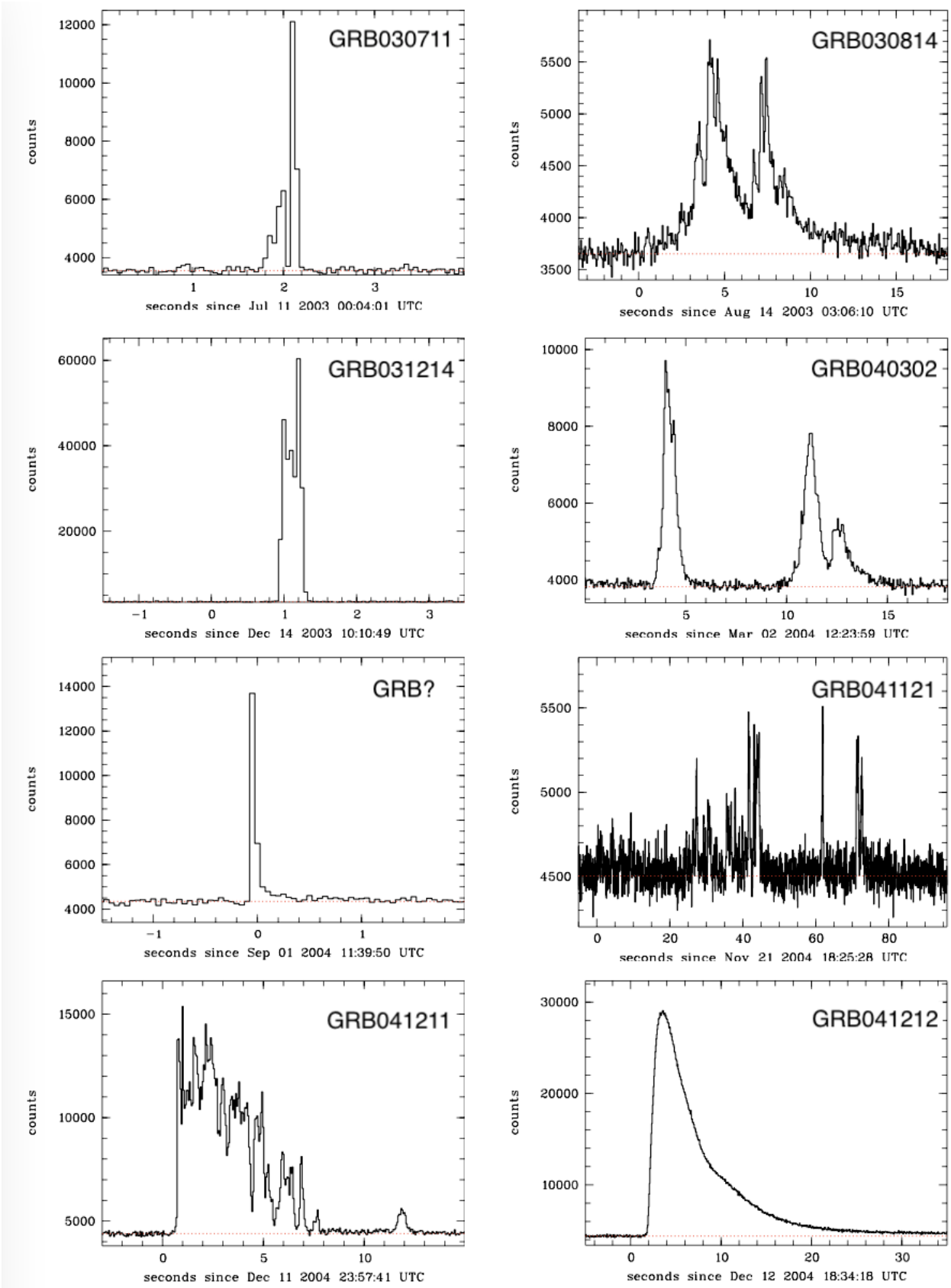


Figure 1.2: **Sample of GRB light curves,** detected by the *INTEGRAL*/SPI-ACS. All are GRBs confirmed by other GRB detectors, except the transient marked "GRB?". Counts are binned in time intervals of 50 ms, and the instrument is sensitive to photons with energies $\gtrsim 75$ keV. This sample reveals the main characteristics of GRBs: short variability timescales, a diversity in morphology and number of pulses, and a wide range of durations. See text for details (Rau et al., 2005).

ray transient marked "GRB?" in Fig. 1.2 was not detected by any other instrument; Its GRB nature is not confirmed.

The gamma-ray light curve allows to study the duration of GRBs, which is captured—though underestimated—by the T_{90} statistic. By time-ordering all the photons with their measured energies, one defines the T_{90} as the interval between the times when 5% and 95% of the cumulative detected energy were reached. This procedure is done after subtracting the instrumental and environmental background to the signal. GRBs present a very broad range of durations, with T_{90}'s from tens of milliseconds to hundreds of seconds. This distribution presents a bimodality separating *short GRBs* from *long GRBs* with an approximate divide at $T_{90} \sim 2$ s (Kouveliotou et al., 1993); We will come back to this classification in detail in Sec. 1.3.

The *gamma-ray spectrum* of a GRB is built from its empirical definition: the distribution in energy of the photons. In Fig. 1.3, we provide the gamma-ray spectrum of GRB990123, a particularly bright GRB with good photon statistics. The spectrum was compiled from observations of the four instruments aboard *CGRO* (Briggs et al., 1999). The top panel shows the *photon spectrum* in photon/cm^2/s/MeV (i.e., the time-averaged number of photons per unit detector surface per unit energy) and the bottom panel features the E^2N_E *spectrum* in erg/cm^2/s, where every photon was weighed by its energy squared. The E^2N_E spectrum is also known as the νF_ν spectrum, in reference to the spectral flux density F_ν in erg/cm^2/s/Hz which is more familiar in a modeling context.

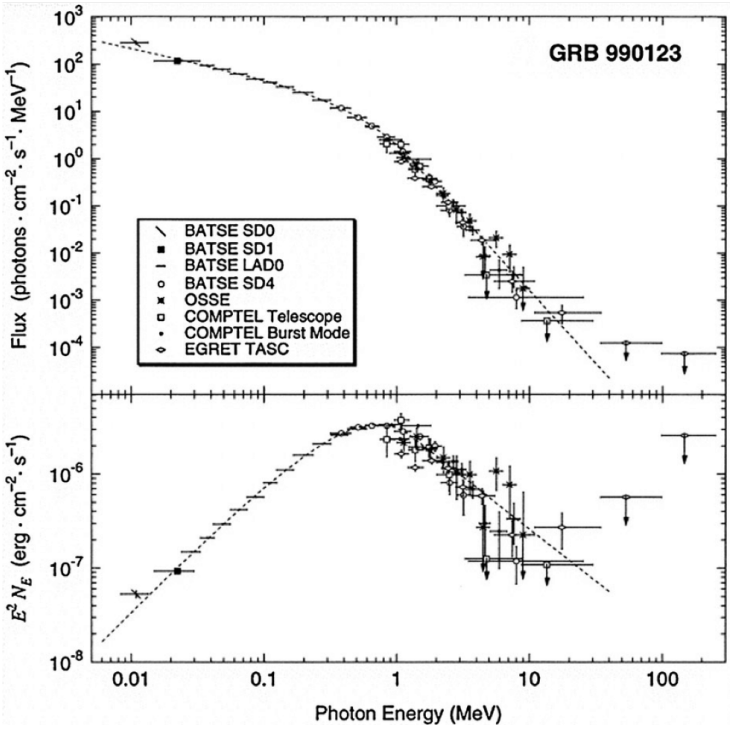


Figure 1.3: **Gamma-ray spectrum of GRB990123.** Top: Photon spectrum. Bottom: E^2N_E spectrum, also known as the νF_ν spectrum. This spectrum is prototypical with a broken power-law shape, peaking at $E_p \sim 1$ MeV (Briggs et al., 1999).

There is less diversity in GRB spectra than in their light curves. Their notable feature is the prominence of power-law segments in the observed spectrum: two smoothly connected segments, an exponentially attenuated power-law (also known as *Comptonized spectrum*) or the so-called *Band function* (an attenuated power-law smoothly connecting to a power-law, Band et al. 1993). For bursts in which the high-energy segments are not well constrained, a single power-law over the entire energy band can be fit. The spectra are therefore non-thermal. In the low-energy range before the spectral break, the photon spectra N_E show spectral indices $\alpha_{\rm ph} \geq -2$, where $N_E \propto E^{\alpha_{\rm ph}}$. They therefore usually present a peak in the E^2N_E-spectrum, as illustrated in the bottom panel of Fig. 1.3. Indeed, if $\alpha_{\rm ph} \geq -2$, the E^2N_E spectrum will exhibit a low-energy slope of $\alpha_{\rm ph} + 2 > 0$ and thus a peak. This *peak energy* is denoted by E_p, it is the photon energy at which most of the GRB's energy was emitted. In GRB spectral catalogs, each of the spectral models is fit to determine the best-fitting one and thus measure the peak energy and the power-law slopes in the photon spectrum, denoted $\alpha_{\rm ph}$ and $\beta_{\rm ph}$ before and after the peak. The most common best-fit model is the Comptonized spectrum; In the ten-year catalog of GRBs detected by *Fermi*/GBM, the observed photon spectrum low-energy slopes are distributed as $\alpha_{\rm ph} = -1.1 \pm 0.45$ (median and 68% confidence interval, Poolakkil et al. 2021). For those bursts with two well-defined segments, the peak energies are distributed as $E_p = 180^{+307}_{-88}$ keV and the high-energy slopes as $\beta_{\rm ph} = -2.2 \pm 0.3$. The very large energy band (8 keV–40 MeV) of *Fermi*/GBM makes it specially fit to measure these parameters.

The gamma-ray photons also allow to define the *fluence* of a GRB. This is the total energy carried by the detected gamma-ray photons per unit detection surface, denoted by $\mathcal{F}$ and usually given in units of erg/cm^2. The fluence can be understood as the time-integrated flux of the gamma-ray signal, and is the first step to determining the energetics of the source of the burst. Among GRBs, the fluence is also broadly distributed from 10^{-4} to 10^{-7} erg/cm^2; The lower limit is of course due to instrumental limitations and not the actual physics of the sources.

Gamma-ray data can only provide the light curve, spectrum and derived quantities such as the duration and fluence. However, any physical interpretation of GRBs must consider the dissipated energy scale of the sources, requiring the distance which the gamma-ray data alone cannot provide. The question of whether GRBs were Galactic or cosmological dominated the first decades of the science of GRBs. Assuming the source of GRBs radiate isotropically, the total dissipated gamma-ray energy in an event can be estimated as:

$$E_{\gamma,\rm iso} = 4\pi D_L^2 \mathcal{F} \tag{1.1}$$

where D_L is the luminosity distance to the source and we neglected effects of redshift.

If the sources of GRB are galactic (at $D_L \sim 10$ kpc), the above mentioned fluences correspond to $E_{\gamma,\rm iso} \sim 10^{40-43}$ erg, 7 to 10 orders of magnitude less than the total output of

a core-collapse supernova. If the sources are cosmological (at $D_L \sim 1$ Gpc), the total gamma-ray energies are rather 10^{50-53} erg, making them the most energetic electromagnetic phenomena in the Universe.

1.2 A brief history of GRB science: milestones and revolutions

1.2.1 The cosmological origin of GRBs

The interrogation on the distance scale to the sources of GRBs remained until the launch of the *CGRO* in 1991, which brought a revolution in GRB science. The *CGRO* carried the Burst And Transient Source Experiment (BATSE), with an improved capability to locate GRB sources thanks to its eight gamma-ray detection modules. In the meantime, a number of theoretical models developed in both the Galactic- and cosmological-origin frameworks; For a compilation of models at the onset of the BATSE era—25 years after the discovery of GRBs—, see Nemiroff (1994). BATSE detected approximately 300 GRBs per year, and the locations of the first progressively established that GRBs were distributed uniformly in the sky (Paciesas et al., 1999; Goldstein et al., 2013). In Fig. 1.4, we represent the sky location of the 2356 GRBs detected during the first ten years of operation of the *Fermi*/GBM. This distribution shows no anisotropy (von Kienlin et al., 2020), confirming and updating the result of BATSE at the time. A uniform distribution of GRBs in the sky tipped the scale towards a cosmological origin, a hypothesis that was disfavored at first because of its drastic consequences on the energetics of the sources.

1.2.2 Relativistic motion in the source of GRBs

A cosmological origin for GRBs leads to the first milestone in the theoretical study of these sources, known as the *compactness puzzle*. It stems from the following properties of GRBs: source variability on timescales $\delta t \lesssim 100$ ms, a non-thermal spectrum with photons of energies $\epsilon \sim 1$ MeV, and an extreme isotropic-equivalent energy output of $E_{\gamma,\rm iso} \sim 10^{52}$ erg.

The variability timescale informs us on the size of the emitting region: To observe coherent emission, the emission region must have uniform physical conditions on the scale of $R \sim c\delta t$. For simplicity, assume the source is spherical, with a width of ΔR, as in Fig. 1.5 (left). Observing photons from the source requires that the source be optically thin to scattering and absorption for this photon. In particular, high-energy photons are prone to gamma-gamma annihilation into a lepton pair. For a photon of energy ϵ_1, annihilation is most likely with photons of energies ϵ_2 such that:

$$\epsilon_1 \epsilon_2 \sim 8(m_e c^2)^2 \tag{1.2}$$

with m_e the electron's mass.

The emitting region's total energy density is given by:

$$u_\gamma = \frac{E_{\gamma,\rm iso}}{\frac{4}{3}\pi R^2 \Delta R} \tag{1.3}$$

and the number density of photons of energy ϵ_1 in the source is:

$$n_{\rm ph}(\epsilon_1) = u_\gamma \xi(\epsilon_1) \tag{1.4}$$

where ξ is the photon-number normalized spectrum (in erg^{-1}), such that $\int {\rm d}\epsilon \xi(\epsilon) = 1$. Neglecting an order-unity constant, we have:

$$\xi(\epsilon) \sim \frac{1}{E_p}\left(\frac{\epsilon}{E_p}\right)^{\beta_{\rm ph}} \tag{1.5}$$

assuming the photon spectrum has a single power-law segment.

Therefore, the optical depth to gamma-gamma annihilation of a photon of energy ϵ_2 is:

$$\tau_{\gamma\gamma}(\epsilon_2) \sim n_{\rm ph}(\epsilon_1)\Delta R \sigma_T \tag{1.6}$$

$$= \frac{E_{\gamma,\rm iso}}{\frac{4}{3}\pi R^2}\frac{1}{E_p}\left(\frac{\epsilon_1}{E_p}\right)^{\beta_{\rm ph}}\sigma_T \tag{1.7}$$

$$= \frac{E_{\gamma,\rm iso}}{\frac{4}{3}\pi (c\delta t)^2} E_p^{-1-\beta_{\rm ph}}\left(\frac{8 m_e^2 c^4}{\epsilon_2}\right)^{\beta_{\rm ph}}\sigma_T \tag{1.8}$$

where we used the Thompson cross section as an approximation for the gamma-gamma annihilation process.

From Eq. 1.8, we understand that a harder spectrum, a smaller variability timescale or a larger total energy output will increase the optical depth. Numerically, we find:

$$\tau_{\gamma\gamma} = 3\times 10^{12}\left(\frac{E_{\gamma,\rm iso}}{10^{52}\ {\rm erg}}\right)\left(\frac{\delta t}{100\ {\rm ms}}\right)^{-2}\left(\frac{E_p}{200\ {\rm keV}}\right)^{1.2}\left(\frac{\epsilon_2}{1\ {\rm MeV}}\right)^{2.2} \tag{1.9}$$

where we adopted $\beta_{\rm ph} = -2.2$.

We thus find that the optical depth to observed photons to gamma-gamma annihilation within the source is $\tau_{\gamma\gamma} \gg 1$ for typical bursts, which is the compactness puzzle.

The solution to this puzzle comes by introducing relativistic motion in the source, as originally proposed by Rees (1966) for radio sources. The corresponding physical setup is shown in Fig. 1.5 (right): The shell has relativistic motion with Lorentz factor Γ. The first consequence is the constraint on the source size stemming from the variability timescale: Because of relativistic beaming, the observer only detects radiation from a portion of the shell an angular distance at most $1/\Gamma$ from their line of sight. This portion now has a transverse size of $2R(1-\cos 1/\Gamma) \sim R/\Gamma^2$ such that the source size can now be much larger than estimated from the variability: $R \sim c\delta t \Gamma^2$. Second, the condition for photon annihilation must be written in the shell's rest frame, such that it is now $\frac{\epsilon_1}{\Gamma}\frac{\epsilon_2}{\Gamma} \sim 8(m_e c^2)^2$. Therefore, changing ϵ_1 for ϵ_1/Γ^2 and R for $c\delta t\Gamma^2$ in Eq. 1.8, we find that relativistic motion changes the optical depth by:

$$\tau_{\gamma\gamma,\rm Rel.} = \Gamma^{2\beta_{\rm ph}-2}\tau_{\gamma\gamma} \tag{1.10}$$

With $\beta_{\rm ph} \sim -2.2$, we find that the optical depth can be brought down to $\lesssim 1$ with $\Gamma \gtrsim 100$. The observation of such high-energy photons from such variable sources thus proves the ultra-relativistic nature of GRB sources.

This calculation affords intuition on this puzzle, but it is greatly simplified (e.g., it assumes a single power-law spectrum) and only considers gamma-gamma annihilation as the process

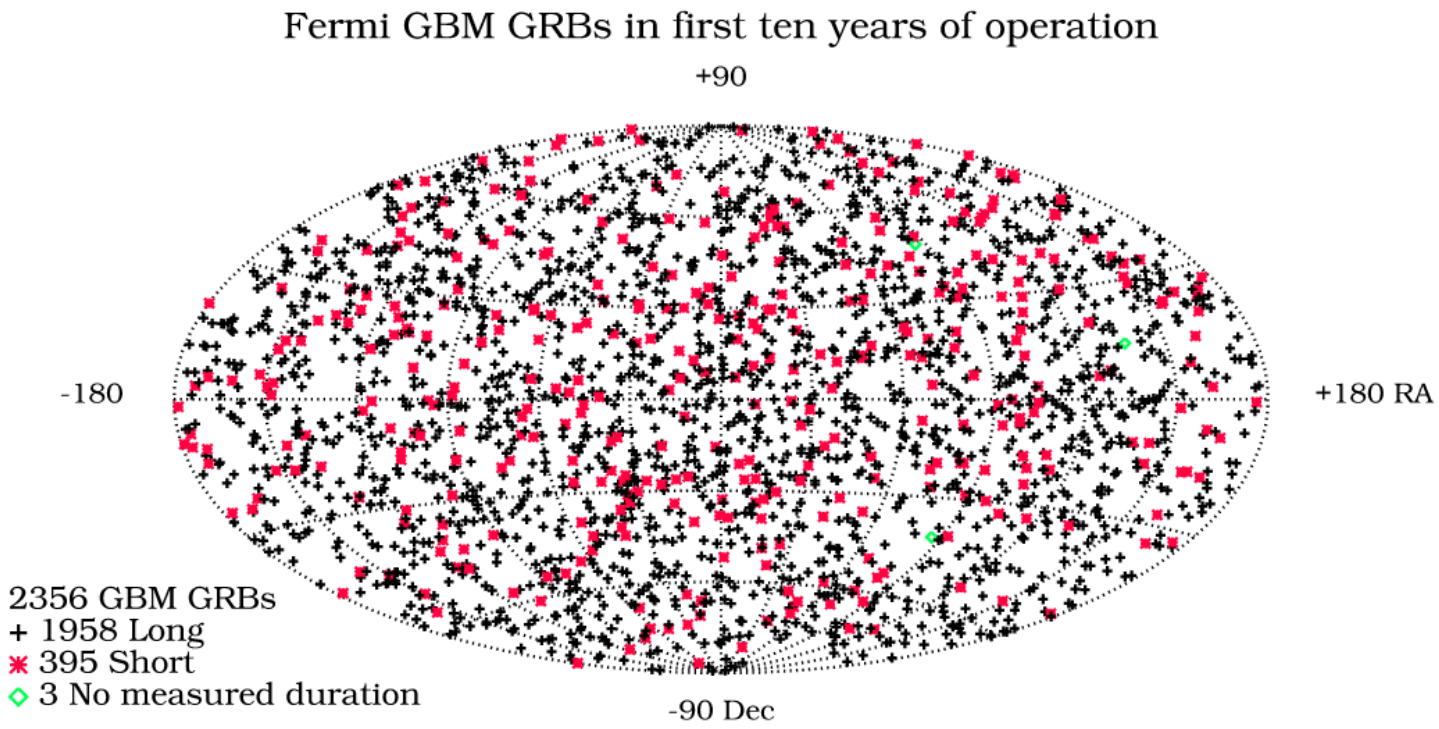


Figure 1.4: **Sky locations of the first 2356 GRBs detected by *Fermi*/GBM.** It shows no anisotropy, confirming the result of *CGRO*/BATSE which hinted towards a cosmological origin for GRBs as of the early 1990s (von Kienlin et al., 2020).

Figure 1.5: **Physical setup for the compactness puzzle calculation.** Left: The GRB-emitting region source is a spherical shell of thickness ΔR at a radius R from the central engine. Right: If the region is relativistic, only the blue portion of angular size $1/\Gamma$ is visible to the observer; The red portion is relativistically beamed away.

to which gamma-ray can be lost. Other processes, such as scattering on the gamma-gamma-created pairs and on the ambient electrons in the source must also be considered. Calculations show that pair creation is indeed often the dominant process, and considering actual GRBs with different spectral shapes and also in the higher-energy bands shows that generally, the compactness constraint is $\Gamma \gtrsim 100$ (Lithwick & Sari, 2001; Hascoët et al., 2012). Note that the modified annihilation condition on the photons' energies can also be interpreted by the fact that in the observer frame, the source photons are all nearly parallel because of relativistic beaming, and therefore their interaction is suppressed.

We also note that a variability on the scale of tens of milliseconds indicates emission regions on the scale of tens of kilometers. Such large energies on such small scales point towards an origin of GRB in the cataclysmic formation or interaction of compact objects. In Sec. 1.3, we will detail the now-established origin of GRBs: in the collapse of massive stars and the merging of compact objects.

The solution to the compactness puzzle leads to the *baryon loading puzzle:* With estimated outflow energies on the order of the dissipated energy $E_{\gamma,\mathrm{iso}} \sim 10^{52}$ erg and Lorentz factors of $\Gamma \sim 100$, the baryonic content of the outflow should be on the order of:

$$M_{\mathrm{bar.}} \sim 5 \times 10^{-5} \mathrm{M}_{\odot} \left(\frac{E_{\gamma,\mathrm{iso}}}{10^{52}\ \mathrm{erg}} \right) \left(\frac{\Gamma}{100} \right)^{-1} \tag{1.11}$$

Such a small ejected mass from high-energy phenomena such as GRBs is puzzling, especially knowing the large ejected masses $\gtrsim 1\mathrm{M}_{\odot}$ from other phenomena such as supernovae. We will present some suggested solutions to the baryon-loading puzzle in Sec. 1.2.6.

1.2.3 The afterglow phase of GRBs

BATSE data strongly suggested a cosmological origin for GRBs. The evidence was, however, quite indirect and did not allow to pinpoint the distance to individual sources. In 1996, the *Satellite per Astronomia X* (*BeppoSAX*) was launched with its two coded-mask X-ray imagers, the Wild Field Cameras (WFC). The platform was able to downlink localization data on detected transients, such that the sources location could be calculated on the ground within hours and circulated to the community through the Gamma-ray Coordinates Network (GCN). The GCN Circulars and Notices are still used today as the main information relay for high-energy transient astronomy; It has now also been invested with a crucial role for multi-messenger astronomy (Sec. 2.4).

The WFCs detected GRB970228 and, subsequently, a fading X-ray signal hours after the trigger. Optical facilities on the ground detected a fading counterpart at a location consistent with the X-ray data (van Paradijs et al., 1997). Deep imaging at this location later identified a distant galaxy, the first candidate *host galaxy* of a GRB.

Confirmation came when the WFCs detected GRB970508 which was localized within four hours after trigger. Such an early localization allowed for greater and sooner follow-up searches; Radio facilities joined the effort and detected a counterpart. A comparison of optical images of the X-ray error box during the first days after trigger identified a fading object within a distant galaxy. The redshift of this host galaxy was measured at $z = 0.835$, making it the first accurate determination of the distance to a GRB and a spectacular confirmation of their cosmological origin (Metzger et al., 1997). Subsequent bursts with measured redshifts—which were still a minority of all detected bursts at the time—were all found to be cosmological.

The fading radiation detected in the radio, optical and X-ray bands is known as the *afterglow emission*; Its detection constitutes a revolution in GRB science. The afterglow is usually detected first in the X-ray band by instruments on the GRB mission platform, and then in the optical and eventually the radio bands after searches in the X-ray error box on the ground. Multi-wavelength missions such as *Swift* carry optical telescopes to observe the afterglow from space. Beyond providing the source's distance through the identification of a host galaxy, the afterglow radiation allowed for major advances in understanding the physics of GRBs.

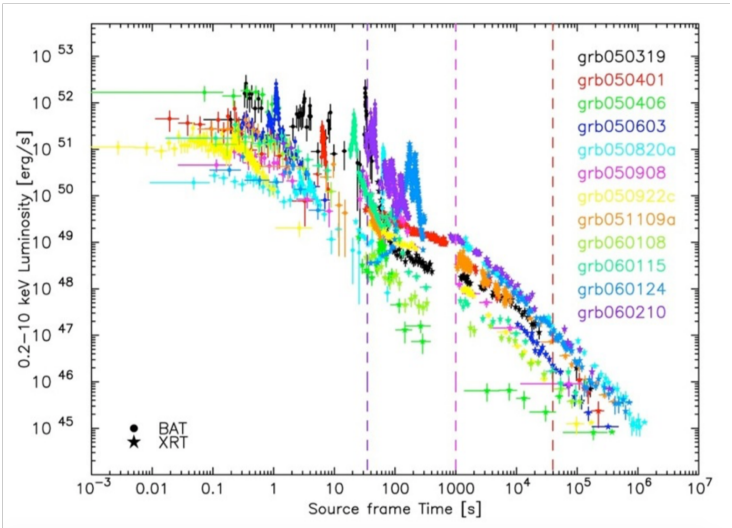


Figure 1.6: **Sample of X-ray afterglow light curves** observed by *Swift*/XRT and BAT (Mangano et al., 2006).

BeppoSAX's localization capabilities allowed to observe the afterglow as of ~ 1000 s after the trigger; a phase now known as the *late afterglow*. Fig. 1.6 allows to study the late afterglow, through X-ray afterglow light curves after the 1000 s mark as detected by the X-Ray Telescope (XRT) aboard *Swift*. These light curves present the canonical behavior of this phase: X-ray afterglows have a power-law behavior in both time and frequency coordinates with a spectral flux in $F_\nu(t) \propto t^x \nu^y$ with $x \sim -1$ and $y \sim -1$; It is thus also non-thermal radiation. In some cases, the afterglow suddenly steepens to $x \sim -2$. Extrapolating the X-ray light curves towards earlier times shows a continuity between the end of the prompt phase and the afterglow. The late afterglow shows almost no variability, it is thus associated with physics on much longer timescales than the prompt phase. It

is now understood as radiation from the shock formed at the front of the continuously decelerating outflow from the GRB source as it penetrates the circum-burst medium; The reverse shock running up the ejecta can contribute to the earlier afterglow phases (Kobayashi, 2000). In this strong forward shock, material is heated and cools through the synchrotron process, producing a spectrum of power-law segments. The modeling of synchrotron radiation from a decelerating outflow is at the heart of the doctoral work presented in this manuscript, and will be vastly developed throughout Parts II and III.

The X-ray flux measured in the 0.1–10 keV band of afterglows at 11 hours after trigger is remarkably tightly distributed as a log-normal with an average value of 5×10^{-13} erg/cm^2/s and a standard deviation of 0.43 (De Pasquale et al., 2003). The total radiated energy in the X-ray afterglow is approximately a few percent of the gamma-ray energy dissipated in the prompt phase. The X-ray afterglow can start while the prompt phase still has some significant activity in the gamma-ray band.

1.2.4 Jets in GRB outflows

The late optical and radio afterglows are detected in about half of well-localized X-ray afterglows. Some of these afterglows show that the steepening to t^{-2} is in fact achromatic: It occurs simultaneously in all bands. An example of such an achromatic break is shown in the afterglow of GRB090426 in six optical bands in Fig. 1.7 (left). Such breaks are interpreted as evidence for a jetted outflow in GRB sources, allowing to decrease the total energy budget for these events. Indeed, the measurement of redshifts for more and more events led to enormous estimates of total dissipated gamma-ray energies in GRBs, according to Eq. 1.1. Assuming the sources are isotropic, we find energies up to 10^{54-55} erg, three orders of magnitude more than the kinetic energy of a core-collapse supernova. The record is currently held by GRB080919C, with $E_{\gamma,\rm iso} \sim 9 \times 10^{54}$ erg (Briggs et al., 2009). Such energies pose a serious challenge to stellar progenitors to GRBs. If the outflow is jetted and covers only a solid angle Ω around the central engine, the actual energy budget of GRBs is much less: Supposing an outflow in the form of a (double) conical jet with half-opening angle θ_j, the real dissipated energy is:

$$E_{\gamma,\rm real} = E_{\gamma,\rm iso}\frac{\Omega}{4\pi} \tag{1.12}$$

$$= (1-\cos\theta_j)E_{\gamma,\rm iso} \tag{1.13}$$

$$\sim \frac{\theta_j^2}{2}E_{\gamma,\rm iso} \tag{1.14}$$

for small viewing angles. For a fiducial $\theta_j \sim 6$ deg, the total energy budget is ~ 200 smaller than the isotropic-equivalent energy, on the order of core-collapse kinetic energies.

Observationally, the jetted nature of the outflow can only manifest once the outflow has started to decelerate. Assuming the observer is looking down the jet, initially they can only detect radiation from a cone of angular size $1/\Gamma$, generally smaller than the conical jet's geometrical size. At this stage, any material outside the cone would be debeamed from their line of sight, such that it is as if the outflow were isotropic. When the jet decelerates, its beaming cone opens up and the observer receives radiation from material at further and further latitudes from their line of sight. When the beaming cone reaches the jet's geometric edge, the solid angle of contributing material suddenly saturates, and the still-opening beaming cone produces a break in the light curve: The observer realizes there is material missing around the jet, it is the *jet break*, illustrated in Fig. 1.7 (right). This break should be achromatic, as in the multi-wavelength afterglow shown in Fig. 1.7 (left). Leveraging such observations, it is possible to estimate the jet angles of GRBs. Fong et al. (2012) find an interval of $2-20$ deg for a sample of long GRBs; Such measurements are however quite uncertain as they depend on the deceleration dynamics of the outflow and therefore on the circum-burst environment (Sec. 4.4.5, Appendix D). Note that such measurements can only be done if a multi-wavelength light curve is available and the achromaticity of the break is effectively noted. In practice, true breaks attributable to the jet break are rare; Other late breaks in afterglows are indeed chromatic, requiring another—likely spectral—interpretation (Panaitescu et al., 2006). The presence of a relativistic jet in GRB outflows is further supported by the observation of radio scintillation in afterglows (e.g. Frail et al., 1997), as well as super-luminal source expansion observed on very close-by events (Pihlström et al., 2008). Finally, we note that the relativistic jet nature of GRB outflows has an incidence on the observed rate of GRBs: The requirement of approximate alignment with the jet due to relativistic beaming leads to correct the estimation of the actual rate of GRB explosions from their detection rate by the same $\Omega/4\pi$ factor in Eq. 1.14, known as the *beaming factor*.

The presence of relativistic jets in GRBs is thus robustly implied by prompt and late afterglow observations. These relativistic jets play a key role in shaping the phenomenology of GRB prompt and afterglow emission. Naturally, a simple conical geometry is unlikely to actually exist. Evidence shows that these jets rather possess *angular structure*, with a relativistic cone defining a preferential outflow direction surrounded by less extreme material; We will present this evidence in Chap. 8. An important part of the work accomplished during the doctoral studies presented in this manuscript consisted in further exploring the consequences of relativistic jets and their structure on GRB observations; The reader eager to discover this work can report to Chaps. 9 and 10. Today, relativistic jets and their structure are a prime ingredient of the general discussion in the high-energy astrophysics community.

1.2.5 The *Swift* era: a rich phenomenology in the early afterglow

Another era in GRB science started with the launch of the *Swift* mission in 2004. Its large-field-of-view coded-mask instrument (the BAT) and swift slewing capabilities permitted to search and detect afterglows much earlier; This meant revealing the *early afterglow* phase ($\lesssim 1000$ s after trigger) and a much more frequent measurement of GRB redshifts. From

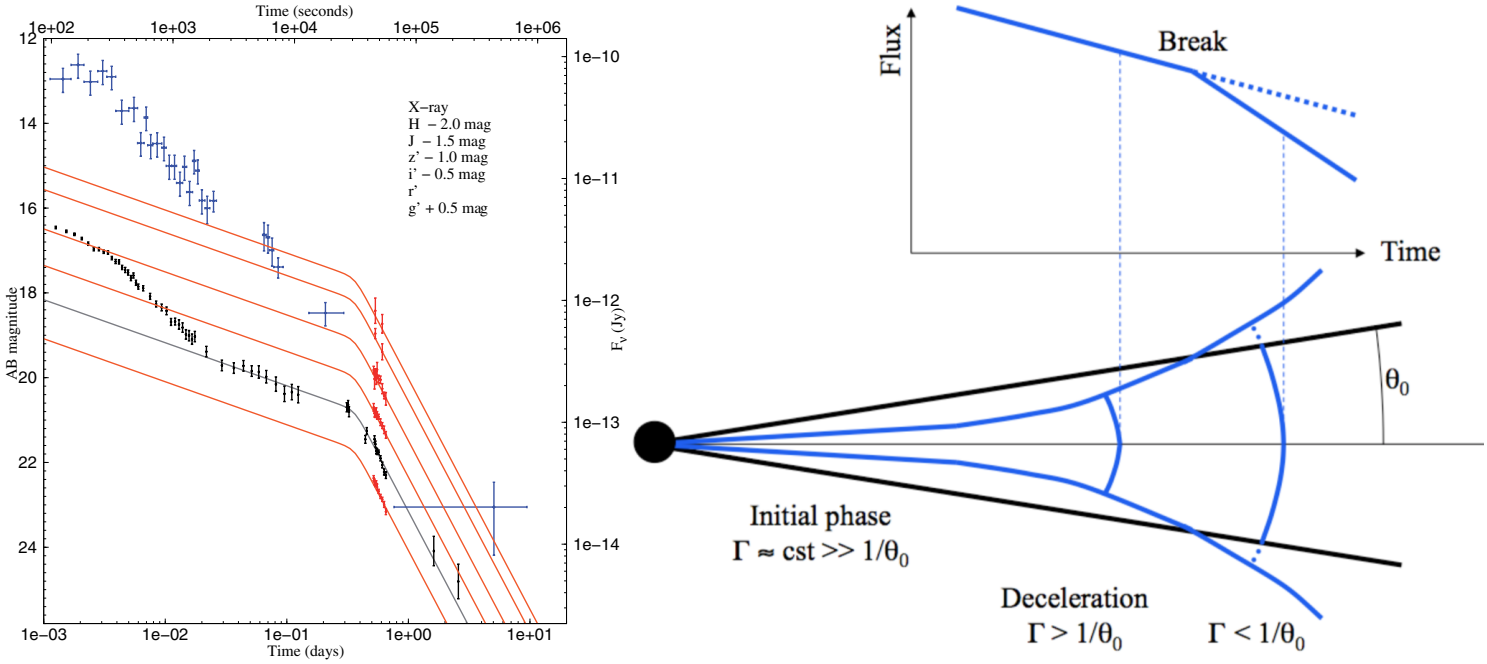


Figure 1.7: **Example and explanation of the jet break.** Left: An example of an achromatic jet break, in GRB090426 (Nicuesa Guelbenzu et al., 2011). Right, bottom: The blue lines represent the progressive opening of the relativistic beaming cone; The black lines are the physical cone of the jet. Once the beaming cone reaches the edge of the jet, the solid angle of visible material suddenly saturates, producing a break in the light curve. Right, top: Corresponding light curve feature.

only a handful of redshifts measured among the ~ 300 detected per year during the BATSE era, *Swift* detects ~ 100 GRBs per year, among which $\sim 36\%$ have measured redshifts, from $z = 0.03$ to $z = 9.38$ (Amati et al., 2002a; Lien et al., 2016). This constituted a revolution in the study of GRB energetics and a development of GRB phenomenology we briefly present in Sec. 1.4.

Fig. 1.6 testifies the rich diversity of behaviors revealed by *Swift* in the early afterglow, in contrast with the late afterglow. The *canonical light curve* presents the following phases (Nousek et al., 2006):

- It starts with the *early steep decay* (ESD) which is continuously connected to the end of the prompt phase in both the soft and hard X-ray bands; It decays steeply with $x < -3$. This phase is commonly associated with high-latitude emission, that is the continuous arrival of the last prompt photons from regions further and further from the observer's line of sight; Their arrival is delayed because of the longer light travel time (e.g. Zhang et al., 2006).

- The ESD can lead to a *plateau*: a phase of very shallow decrease or slight increase of the afterglow flux, that can be imperceptible or extended for hundreds of seconds.

- Only after the ESD—or plateau, if present—does the *regular decay* mentioned earlier appear, with $x \sim -1$. Once again, this phase is associated with radiation from the shock formed by the GRB outflow's deceleration. The steepening to $x \sim -2$ of jet-break or spectral origin can then follow.

- Above these strictly power-law decreasing phases, *Swift* revealed the *flares* observed in about half of the early afterglows. Flares are sudden and transient rebrightenings of the afterglow, sometimes by orders of magnitude.

Afterglow plateaus and flares themselves feature a rich phenomenology and constitute an entire branch of GRB studies; Their origin is still an active subject of debate. In Chaps. 9 and 10, we will introduce afterglow plateaus and flares in more depth and present original models based on the structure of GRB jets that seem consistent with the phenomenology revealed by the numerous catalogs built from *Swift* data.

1.2.6 The GRB phenomenon: global picture

From the general facts above, we can draw a gross picture for the GRB phenomenon. Transient activity in the vicinity of a compact source—known as the *central engine*—leads to the launching of an ultra-relativistic jetted outflow with a poor baryon content that reaches Lorentz factors of $\Gamma \gtrsim 100$ and isotropic-equivalent energies of $E_{\rm jet,iso} \gtrsim 10^{52}$ erg of kinetic, magnetic or thermal nature. Some of the jet's energy is converted to radiation, producing the prompt emission through a non-thermal process or the transformation (e.g., by Compton scattering) of thermal energy of the jet. The outflow with the remaining energy then penetrates the circum-burst medium, producing a long-lived strong forward shock and a shorter-lived

reverse shock. Jet and circum-burst material are heated at the shock, where particles are accelerated to non-thermal distributions and magnetic fields form, such that durable synchrotron radiation is produced as the outflow decelerates, which is the long-lasting, multi-wavelength afterglow.

The relativistic ejection can occur through the accretion-ejection process around the central engine, for example by the Blandford & Znajek (1977) process in the case of a large scale magnetic field threading the accretion disk. It can also occur by depositing energy in the cloud of material around the engine through the annihilation of neutrinos and antineutrinos coming from the hot accretion disk (Mochkovitch et al., 1993); By adiabatic expansion the cloud can accelerate to ultra-relativistic speeds. In the accretion-ejection picture, the threading of the magnetic field lines across the black hole's horizon allows to remove their baryon content, solving the baryon-loading puzzle (Sec. 8.2.3 Levinson & Eichler, 1993). In the neutrino-antineutrino picture, the energy is deposited within the cloud from an external source, such that the injected energy is independent from the initial density in baryons, which can be very small.

The prompt dissipation mechanism within the jet could also be due to non-thermal radiation from a population of leptons accelerated by dissipation of magnetic energy by reconnection (*magnetic reconnection models*, e.g. Giannios & Spruit, 2005) or by internal shocks formed when shells in the outflow with different velocities collide (*internal shock models*, Rees & Meszaros, 1994; Kobayashi et al., 1997; Daigne & Mochkovitch, 1998). The other main family of models advanced is partial comptonization of the black-body radiation released when the hot outflow becomes optically thin to the observer after expansion, the so-called *dissipative photospheric models* (e.g. Beloborodov & Mészáros, 2017).

1.3 Long and short GRBs

Another fundamental result of the BATSE era that has been confirmed ever since is the bimodality of the durations of GRBs, underlined by two-dimensional clustering in the *spectral-hardness-duration plane*, leading to the definition of short and long GRBs. The spectral hardness is the ratio of photon counts between a higher-energy band and a lower-energy band of a same instrument, such that harder spectra correspond to larger spectral hardness. In Fig. 1.8, we reproduce the distribution of the ten-year catalog of GRBs detected by *Fermi*/GBM in this plane using the 10–50 keV and 50–300 keV bands (von Kienlin et al., 2020). There appears a bimodality in duration and a trend by which shorter bursts are harder than longer bursts; Because of the wide distribution of spectral hardness in long GRBs, the bimodality in this quantity is less obvious.

The simultaneous fitting of two Gaussian sub-populations in $\log T_{90}$ (top panel) suggests that the median short GRB has a duration of $\overline{T_{90,\mathrm{s}}} = 1.05$ s while the median long has $\overline{T_{90,\mathrm{l}}} = 30$ s; short GRBs represent $\sim 25\%$ of the observed population. According to this fit, the two sub-populations have equal frequency at $T_{90} \sim 4$ s, placing the short–long divide at this duration. Because of the different energy ranges and sensitivities of GRB detectors, the measured T_{90} vary from an instrument to another, as do the divide between short and long GRBs in catalogs; The bimodality remains however.

The afterglow era allowed to study the host galaxies of GRBs and in particular to compare the two classes to GRBs: The host galaxies of long GRBs were generally faint, blue and very star-forming (Le Floc'h et al., 2003; Savaglio et al., 2009), whereas short GRBs are not selective of galaxy type nor star-formation activity (Berger, 2014). Furthermore, long GRBs tend to occur close to their host galaxy centers, in UV-bright regions thus associated with massive stars (Lyman et al., 2017). Conversely, short GRBs can occur with very large offsets from their host galaxies (Fong & Berger 2013, Chap. 7).

Early on, it was suggested that long GRBs could be produced in the collapse of massive stars, it is the *collapsar model* (Woosley, 1993; Paczyński, 1998): The progenitor star collapses to a black hole and its envelope forms an accretion disk around the compact object, starting the scenario described in Sec. 1.2.6. This picture was confirmed with the space-time coincidence of GRB980425 and SN1998bw, in which the afterglow spectrum progressively led to a supernova spectrum (Galama et al., 1998). Other examples followed with similar spectral signatures (e.g., GRB030329, Hjorth et al. 2003) and emergence of the supernova optical signal during the afterglow (e.g., GRB050525A, Della Valle et al. 2006). When observed in association with long GRBs, supernovae are systematically of the Ic broad-line type (Woosley & Bloom, 2006). The question of the conditions on the progenitor (mass, metallicity, etc.) so that collapse leads to long GRB emission is still not clear; For an observational and statistical perspective, see Palmerio et al. (2019).

The progenitors of short GRBs remained elusive until very recently. Paczynski (1986); Eichler et al. (1989) suggested the merger of binary neutron stars: The various ejecta from such a merger—ejected by tidal forces, sheer pressure between the colliding objects, etc.—would form an accretion disk around the merger product, a black hole. This would once again lead to scenarios as in Sec. 1.2.6. The binary neutron star (or generally binary compact object) would only merge a long time after its formation, because its orbital decay due to gravitational radiation is generally slow; This would explain the large host galaxy offsets, and the generally lower redshifts of observed short GRBs with respect to long GRBs (Guetta & Piran, 2006). This progenitor scenario for at least some short GRBs was confirmed spectacularly with GRB170817A, a short GRB—though a very peculiar one—that was the first electromagnetic counterpart to a gravitational-wave binary neutron star merger signal. This event kicked off the era of *multi-messenger astronomy* with gravitational waves, where information carried by both electromagnetic and gravitational-wave signals are combined to study astronomical phenomena. It motivated essentially all of the doctoral work presented in this manuscript. We will introduce gravitational waves in Chap 2, and Chap. 3 is dedicated to this event.

While observed short GRBs are indeed harder than long GRBs ($\sim$ 450 keV vs. 200 keV), it is not clear that they are

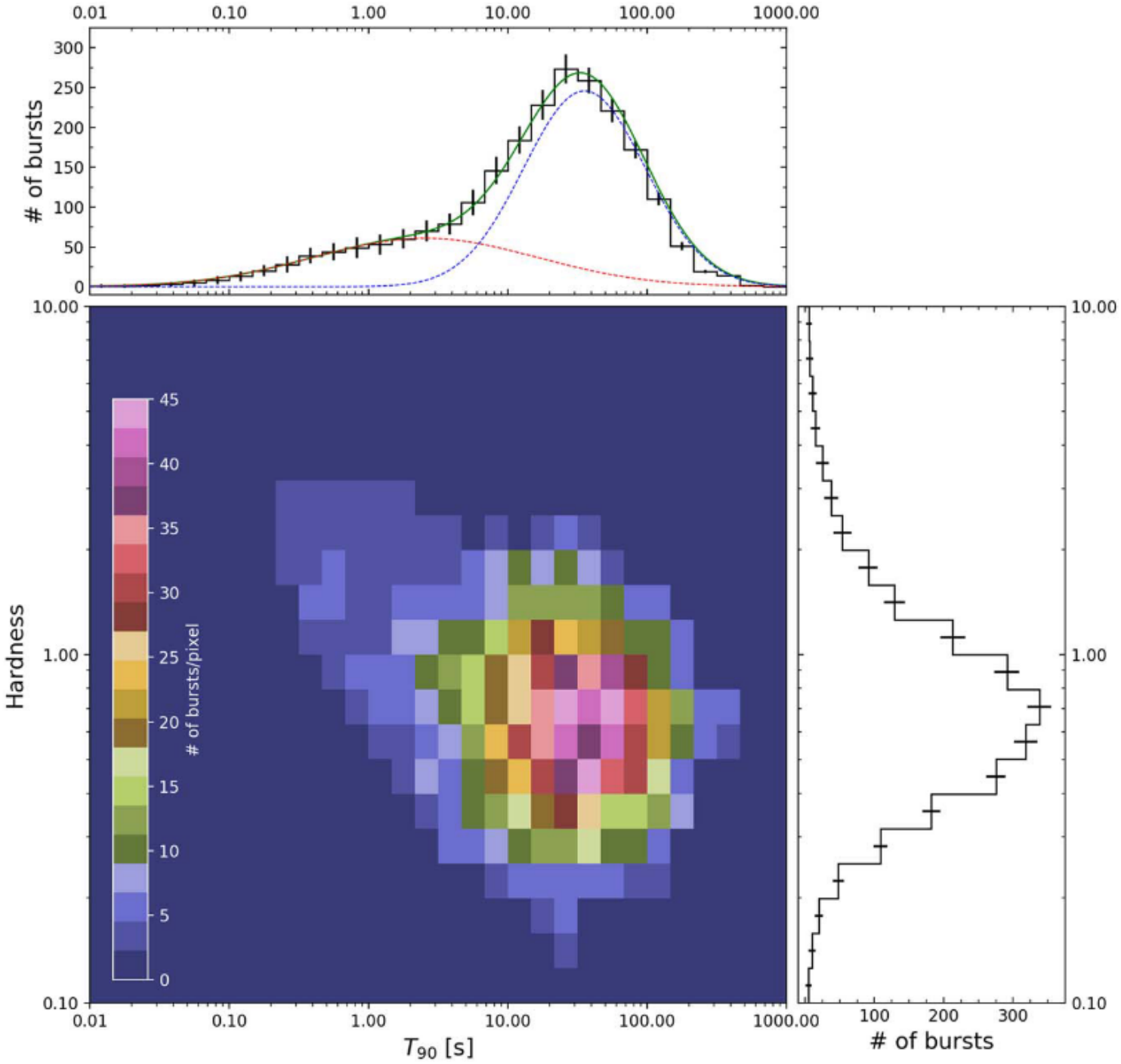


Figure 1.8: **Two-dimensional distribution of the spectral hardness vs. duration,** for the fourth *Fermi*/GBM catalog of GRBs. The top and right panels show one-dimensional projections for T_{90} and spectral hardness, respectively. The top panels shows a fit of two Gaussian sub-populations in duration, revealing the dichotomy between short and long GRBs (von Kienlin et al., 2020).

intrinsically harder, as the low statistics of short GRBs with redshift do not allow to conclude (Poolakkil et al., 2021). From a temporal point of view, pulses in long and short GRBs are very similar (Guiriec et al., 2010); Most short GRBs are single-pulsed, contrarily to long GRBs (Hakkila et al., 2018).

The physical setup is therefore quite similar for short and long GRBs: an accretion disk around a stellar-mass black hole. However, the lower mass of the disk following a compact object merger with respect to a collapsar and possibly its different rotation profile or large scale magnetic field could simply explain the shorter duration, lower gamma-ray dissipated energy, luminosity and dimmer afterglows of short GRBs with respect to long GRBs (Ghirlanda et al., 2009; Berger, 2014; Poolakkil et al., 2021).

Where short and long GRBs should differ is in the immediate environment of the burst: Massive stars are expected to launch winds and the outflows of collapsar-induced GRBs should thus penetrate this type of environment, signing the afterglow phase particularly; Conversely, binary compact objects are expected to merge in a uniform, likely rarefied medium, producing visible signatures in the afterglow. Nonetheless, studies have found that afterglows from a large number of long GRBs are better fit assuming a uniform circum-burst medium (e.g., Curran et al., 2011). Similarly, evidence has accumulated over the years that blurs the association of short GRBs with rarefied media. In Chap. 7, we will summarize this evidence and more generally pose the question of the delay between binary formation and merger, and its role in taking binaries to rarefied, very offset merging locations. We will describe a multi-messenger method that aims to provide a more direct probe of the environments of binary compact object mergers—and thus short GRBs—, relying on merger gravitational-wave signals and afterglows.

Finally, in Sec. 1.2 we mentioned the potential structure of GRB jets and its role in shaping the prompt and afterglow signals. In this respect, we also expect a difference between short and long GRBs. While there can be many origins for the structure of GRB jets, we will show in Chap. 8 that structure is prominently determined by the interaction of the jet launched by the central engine with the layer of dense material surrounding it: the collapsar's envelope or ejecta from the merger. The different velocities and masses of these two types of layers should imprint different structures. Currently, only the extraordinary imagery of the jet in GRB170817A allowed a detailed insight into a GRB jet (Sec. 3.3). Apart from this singular event, only recent numerical simulations allow some insight into these differences; We mention them in Sec. 8.3. They generally conclude that the larger mass, thickness and ejection duration in collapsar scenarios lead to more interaction and particularly more mixing between jet and layer material and thus jet loading; The ultimate structure should therefore be shallower and wider. In Chap. 9 and 10, we will study the influence of structure on GRB signals though afterglow models. They could contribute to better understanding these structures and study their differences in long and short GRBs.

1.4 A concise phenomenology of GRBs

To provide a more complete picture, we will briefly summarize the phenomenology of GRBs.

The *Swift* era bringing many more redshift measurements, a series of correlations between various GRB source-frame observables was established. For long GRBs, these are, e.g, positive correlations between: the peak energy of the GRB spectrum measured at peak flux and the isotropic-equivalent luminosity at this moment ($E_{p,\mathrm{peak}}$–$L_{\mathrm{iso,peak}}$, Yonetoku et al. 2004); spectral peak energy and isotropic-equivalent gamma-ray energy (E_p–$E_{\gamma,\mathrm{iso}}$, Amati et al. 2002b); spectral peak energy and beaming-corrected gamma-ray energy (E_p–$E_{\gamma,\mathrm{real}}$, Ghirlanda et al. 2004b). Though with smaller statistics, it seems such correlations also hold for short GRBs; most notably, the $E_{p,\mathrm{peak}}$–$L_{\mathrm{iso,peak}}$ correlation (Ghirlanda et al., 2009; Ghirlanda et al., 2015). Whether these correlations are intrinsic or shaped by selection effects is still debated, while some efforts have been done to disentangle the contributions from both possibilities (Palmerio & Daigne, 2021). The isotropic-equivalent energies of GRBs are broadly distributed as $E_{\gamma,\mathrm{iso}} \sim 10^{53\pm1}$ erg (Amati, 2006), and the beaming-corrected energies more tightly $E_{\gamma,\mathrm{real}} \sim 10^{51\pm0.5}$ erg (Ghirlanda et al., 2004b). While the more energetic GRBs are the rarest, there is evidence for an intrinsic cutoff in the distribution of $E_{\gamma,\mathrm{iso}}$, at $\sim 1{-}3{\times}10^{54}$ erg $\sim 1 M_\odot c^2$ (Atteia et al., 2017). In some serendipitous cases, an optical flash coincident with the prompt phase was observed (Akerlof et al., 1999), possibly with strong linear polarization (Troja et al., 2017a); The prompt activity in the optical band is still elusive and would be better understood with optical monitoring of gamma-ray instruments' fields of view (Sec. 1.5).

From the temporal point of view, it seems that there is no preferred variability timescale in GRB light curve: The power density spectrum reveals a simple power-law (Beloborodov et al., 2000); This is naturally best established for very long GRBs with more data. GRBs generally exhibit *spectral lag*, by which pulses peak early when viewed in higher-energy bands (e.g. Bernardini et al., 2015). Moreover, more luminous bursts display smaller spectral lag, for given energy bands (Hakkila et al., 2008; Ukwatta et al., 2010). Similarly, when viewed in higher energy bands, GRB pulses have smaller widths, with $w(E) \propto E^{-0.4}$ (Norris et al., 1996). Generally, bursts present a hard-to-soft evolution (Ford et al., 1995); For very bright bursts in which time-resolved spectral evolution can be studied in individual pulses, it was shown that the peak energy decreases approximately as $E_p \sim t^{-1}$ (Peng et al., 2009). There is evidence for some precursor episodes in $\sim$ 10% of *Swift*/BAT GRBs (Troja et al., 2010); These episodes have similar temporal and spectral properties as the main prompt episode and occur up to tens of seconds before the main pulse.

Among long GRBs, there is a class of *ultra long GRBs*, with $T_{90} \geq 1000$ s that is statistically distinct from classical long GRBs (Boër et al., 2013); They represent $\sim$ 2% of the *Swift*/BAT catalog of GRBs (Lien et al., 2016). Their extreme longevity and energetics could suggest a separate class of progenitors, such as low-metallicity blue super-giants, or generally population III stars (e.g. Gendre et al., 2013). Among short

GRBs, there is a class of bursts with *extended emission*, presenting softer, less luminous emission episodes lasting for tens of seconds after the main pulse (Norris & Bonnell, 2006); These represent $\sim 1\%$ of *Swift*/BAT GRBs (Lien et al., 2016). The similarity with afterglow plateaus has led some to suggest continuous energy injection to a forward shock from slower ejecta (Matsumoto et al., 2020), or the spindown of the newly-formed magnetar (Metzger et al., 2008). These extended episodes with durations similar to long GRBs also remind us of the finding that the variability timescales and spectra of short GRBs are similar to those during the first seconds of long GRBs (Nakar & Piran, 2002; Ghirlanda et al., 2004a).

Finally, we note the existence of low-luminosity events, dubbed low-luminosity GRBs, X-ray-rich GRBs, or X-ray flashes; Examples include GRB980425 (Waxman, 2004) or GRB031203 (Ramirez-Ruiz et al., 2005). The low-luminosity events are generally softer, smoother than other long GRBs and single-pulsed (Schulze et al., 2014, and references therein); Because of their small redshifts, these events are those with SN associations (Sec. 1.3). The volumetric rate of such low-luminosity events is at least ten times higher than the extrapolation of the inferred long GRB luminosity function down to their luminosities of $L_{\rm iso} \lesssim 10^{48}$ erg/s (Cobb et al., 2006; Soderberg et al., 2006; Liang et al., 2007). This poses the question of the nature of these low-luminosity transients and the role of the jet in these cases (Sec. 8.2.3).

1.5 *SVOM*: a mission for multi-messenger high-energy astrophysics

The latest revolution in GRB science is of course GRB170817A and the opening of the era of multi-messenger astronomy with gravitational waves. Throughout Chaps. 4 and 5, we will outline the signals expected from multi-messenger sources in this era; In Chaps. 6 and 7, we will describe how to leverage the multi-messenger datasets to the benefit of cosmology and GRB science. In this context of transient multi-messenger sources, a specific instrumental setup is advocated for, with the following capabilities:

- trigger on GRBs and rapidly slew to carry out multi-wavelength follow-up from the satellite platform;
- rapidly downlink and circulate GRB data to the ground for an extended follow-up effort;
- listen to transient alerts from other GRB or gravitational-wave facilities in order to follow-up from space;
- respond to such alerts with a dedicated ground segment with both shallow, large-field-of-view searching capacities and deep, small field of view instruments.

Such a profile is incarnated in the Space Variable Objects Monitor mission (*SVOM*, launch scheduled in 2022). The information in this section is summarized from the *SVOM* white paper Wei et al. (2016).

The instruments aboard the space platform are described in Fig. 1.1 (top left): ECLAIRs is a large field of view (FoV, 2 sr) coded-mask instrument that innovates in its low-energy-reaching range: 4–150 keV; Its localizing capabilities are $\leq$ 12 arcmin. The Gamma-Ray Monitor (GRM) is composed of three gamma-ray detection modules (GRD A, B and C), each sensitive to the 30 keV–5 MeV energy range on a 3 sr FoV; Where their FoVs coincide—over a region of $\sim$ 2 sr centered on the ECLAIRs FoV—, they collectively have $5-10$ deg localization capabilities. The Visible Telescope (VT) and Microchannel X-ray Telescope (MXT) are both smaller FoV (VT: 26^2 arcmin^2; MXT: 1 deg^2) instruments with localizing capabilities ($\leq$ 10 arcsec point-spread functions), aligned with ECLAIRs. The VT is an optical instrument with simultaneous multi-band imaging (red, blue); The MXT is a "lobster-eye" instrument, sensitive to the 0.2–10keV band—essentially the same as *Swift*/XRT.

The space segment is completed by an extended ground segment, composed of: the Ground Wide-Angle Cameras (GWACs), two geographically separate sets of robotic cameras each with a FoV of 1.5 sr and reaching magnitudes $\sim$ 16 in 10-second exposures; and the Ground Follow-up Telescopes (GFTs), two geographically separate 1-meter class small FoV instruments ($\leq$ 20 arcmin^2) with optical–near infrared cameras.

Using GRB catalogs from previous missions, simulations show that ECLAIRs should detect 50–80 GRBs per year and be sensitive to a number of different GRB classes: classic short and longs GRBs but also softer events—low-luminosity GRBs, X-ray flashes, etc.—thanks to the low threshold of the coded-mask instrument (Antier-Farfar, 2016); The latter is naturally an interesting class because of its unclear nature. Short GRBs are of prime importance in the multi-messenger context. ECLAIRs alone has a moderate efficiency in short GRB detection because of its low high-energy threshold. However, combining ECLAIRs and GRM allows a large spectral range (up to 5 MeV) on the $\lesssim$ 2 sr where their FoV coincide, with good localizing features. The large range should allow a robust reconstruction of the short GRB spectra, even those with multiple components. While the GRM should be most sensitive to the short hard pulse, ECLAIRs should allow to investigate the extended emission of short GRBs down to unprobed low energies.

SVOM will adopt a anti-solar attitude outside of triggered or scheduled observations. The GWACs will continuously monitor ECLAIRs FoV such that $\sim$ 16% of ECLAIRs-triggered GRB will also be covered in the optical. This innovation should permit deeper insight in the still elusive optical features of the prompt emission and generally the prompt physics, through detections or upper limits.

For the brightest GRBs, *SVOM* will slew to allow follow-up by the MXT and VT. In the meantime, an extended VHF network will recover and circulate the most important trigger data such as localization to the ground segment and the astronomical community. Simulations show that the MXT can detect and localize 90% of X-ray afterglows for ECLAIRs triggers; It is expected that $\sim$ 2/3 of ECLAIRs triggers will be

followed by at least 5 min of visibility for MXT and the VT. The optical afterglow can be searched for both by the VT and by the GFT in the visible and near-infrared, affording early photometric redshift measurements. Overall, the rationale for *SVOM* is to build a sample of GRBs smaller than those of *Swift* and *Fermi*, but with better completeness in redshifts ($\sim 66\%$) and multi-wavelength afterglow coverage. This should make of *SVOM* a unique tool for the physics of GRBs, for the use of GRBs for cosmology and for host galaxy studies.

In the multi-messenger perspective, the large FoV of ECLAIRs and GRM should allow detections or constraints on high-energy counterparts to gravitational-wave triggers and high-energy neutrino events. Furthermore, the special sensitivity to low-luminosity events is advantageous as these are promising sources for neutrinos. In rapid response to triggers, the GWACs can be used to search very large gravitational-wave error boxes as has already been done in the recent O3 run. Their goal is optical counterpart detection and localization; The characterization or deeper follow-up and photometric sampling can then be taken over by the GFTs. The GWACs are however very limited in magnitude, and identification of interesting counterpart candidates likely requires spectroscopy from third-party facilities.

From space, the tiling of error boxes in search for X-ray counterparts is a natural strategy for the MXT and its 1deg^2 FoV, several times larger than that of *Swift*/XRT. The GFTs on the ground can rather adopt the galaxy-targeting strategy and of course characterize counterpart candidates from the GWACs or other instruments, and sample confirmed counterparts, as long as they are localized to within ~ 30 arcmin.

To summarize, *SVOM* is a versatile mission with an innovative instrumental profile deployed in space and on the ground. This profile is adapted to a GRB perspective, especially for the study of the elusive low-luminosity bursts; It aims to build a smaller yet more redshift- and afterglow-complete sample. It is also tailored to multi-messenger astronomy with rapid response capabilities and the possibility to both detect and localize counterpart candidates, and then characterize and sample these sources in the optical and X-ray bands.

1.6 Conclusion

GRBs are transient, non-repeating and spatially-localized gamma-ray signals of cosmological origin. They are composed of two radiative phases: the very variable and diverse prompt phase, with non-thermal spectra peaking at energies of hundreds of keV and lasting up to hundreds of seconds; and the longer-lived and smoother afterglow phase, which is observed from the radio to higher-energy bands. GRBs are associated with ultra-relativistic jets launched from compact objects; They are the most energetic electromagnetic phenomena known. The prompt emission is produced by dissipation of energy within the jet, through a mechanism which is still elusive. The afterglow is associated with the continuous deceleration of the jet in the circum-burst environment, producing a hard shock from which accelerated particles radiate mainly through the sychrotron process. The durations of GRB signals are bimodal: Short GRBs last less than ~ 2 s and are produced in the merger of compact objects; Long GRBs are produced in the collapse of massive stars, they are generally softer than short GRBs.

Though still young, GRB science has already experienced numerous revolutions driven by instrumental improvement: discovery in 1967; evidence of cosmological origin through the uniformity of sources in the sky as of 1991; confirmation by the first afterglow detections and redshifts in 1997; more frequent redshifts and revealing of the rich diversity in the early afterglow as of 2004; observation of coincident gamma-ray and gravitational-wave radiation from a source in 2017. This last revolution—the multi-messenger era with gravitational waves—holds much promise for GRB science, as we will start to show through the doctoral work presented in this manuscript.

In Chap. 2, we will introduce the other messenger of this new multi-messenger astrophysics: gravitational waves. In Chap. 3, we will detail the event that inaugurated this era: GW170817, GRB170817A and their counterparts. In Parts II and III, we will describe the doctoral work accomplished which constitutes our first steps into the study of GRBs in the multi-messenger era. In Chap. 11 we will propose some developments to this work.

Chapter 2

Gravitational waves: A new messenger from the extreme Universe

Abstract

We introduce gravitational waves and motivate the mergers of compact objects as promising sources of transient gravitational wave bursts. We derive the signal expected from such mergers and briefly describe the optics-based instrumentation deployed to detect these signals. We relate the first merger events detected in the gravitational domain and introduce multi-messenger astronomy with gravitational waves as a new era for astrophysics. We describe current and future networks of gravitational-wave detectors, how they should perform and how the astronomical community organizes to follow-up the signals detected by the network.

2.1 Gravitational waves and their sources

Gravitational waves (GW) are the propagation of disturbances of the space-time metric. Observationally, they manifest as the oscillation in time of the physical length between two points as measured by light travel time. The propagation of the gravitational interaction at finite speed was suggested by Poincaré (1905) and gravitational waves were predicted shortly after the formulation of the theory of general relativity by Einstein (1915, GR). In the context of GR, GW come about as solutions to a wave equation obtained by linearizing the Einstein equations around a flat and empty Universe solution.

Far away from concentrations of energy and matter, the space-time metric can be decomposed as $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, where η is the Minkowski metric and h is a perturbation with $h_{\mu\nu} \ll 1$. With an adapted choice of coordinates and gauge, the wave equation reads (Einstein, 1916):

$$\Box h_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu} \tag{2.1}$$

where $\Box$ is the usual flat-space-time d'Alembert operator $\eta^{\mu\nu}\partial_\mu\partial_\nu$ and T is the stress-energy tensor.

In vacuum, $T = 0$ and the general solutions of Eq. 2.1 are a superposition of transverse monochromatic plane waves with two possible polarization states denoted h_+ and $h_\times$:

$$h(\overrightarrow{x},t) = h_+(\overrightarrow{x},t)(e_x\otimes e_x - e_y\otimes e_y) + h_\times(\overrightarrow{x},t)(e_x\otimes e_y + e_y\otimes e_x) \tag{2.2}$$

where z is the direction of propagation, x and y are orthogonal directions and the fundamental modes are $h_\circ = a_\circ e^{i(\omega_\circ t - k_\circ z)}$ for both polarizations $\circ = +, \times$. GWs are non-dispersive and propagate at the speed of light: $\omega_\circ = ck_\circ$. Extensions of GR or alternative theories predict the existence of additional polarization states; However the first observed GW signals from compact object mergers have strongly constrained these hypothetical states (Abbott et al. 2017a, Sec. 2.3).

GWs are transverse waves, with effects only in the plane orthogonal to propagation. A purely +-polarized wave will successively dilate and contract lengths in a direction denoted $\overrightarrow{n_+}$, and a purely ×-polarized wave in the $\overrightarrow{n_\times}$ direction. There is an angle of $\pi/4$ between $\overrightarrow{n_+}$ and $\overrightarrow{n_\times}$, as illustrated in Fig. 2.1; hence the "+" and "×" denominations. Generally, if we measure a length L along a unitary direction $\overrightarrow{n}$, the passage of the wave will affect L by:

$$\frac{\Delta L}{L} = \frac{1}{2} h_{ij} n^i n^j \tag{2.3}$$

Coming back to Eq. 2.1, the GW radiation far away from a gravitating source can be determined by the mass distribution and motion of the source[1]. In case of a source made purely

[1]The derivations presented in this section and the next draw from the course *Objets compacts et phénomènes associés* by Prof. F. Daigne.

of matter with a small velocity, the *quadrupole formula* writes (Einstein, 1918):

$$h_{ij}(\overrightarrow{x},t) = \frac{2G}{c^4 r}\left[P_i^k P_j^l - \frac{1}{2}P_{ij}P^{kl}\right]\ddot{Q}_{kl}\left(t-\frac{r}{c}\right) \tag{2.4}$$

where $Q_{ij}(t)$ is the Newtonian quadrupolar moment of the source:

$$Q_{ij}(t) = \int_{\text{source}} \mathrm{d}^3\overrightarrow{x}\,\rho(\overrightarrow{x},t)\left(x_i x_j - \frac{1}{2}\|\overrightarrow{x}\|^2\delta_{ij}\right) \tag{2.5}$$

and P is the projection operator on the plane transverse to the direction to the source $P_{ij}(\overrightarrow{x}) = \delta_{ij} - \frac{x_i x_j}{r^2}$. For a purely material source, the Newtonian mass density is $\rho = T^{00}/c^2$. The second-derivative nature of Eq. 2.4 naturally motivates sources in acceleration—such as tight binary systems—as promising GW emitters.

GWs transport energy, linear and angular momentum. The energy flux from a source of GWs is:

$$F_{\text{GW}}(\overrightarrow{x},t) = \frac{c^3}{16\pi G}h^2(\overrightarrow{x},t) \tag{2.6}$$

where h is the amplitude of the wave, i.e., the root-mean-square average of the two polarization states: $h^2 = \langle \dot{h}_+^2 + \dot{h}_\times^2 \rangle$.

More generally, the GW luminosity of a source is:

$$L_{\text{GW}} = \frac{G}{5c^5}\langle \dddot{Q}_{ij}\dddot{Q}^{ij}\rangle \tag{2.7}$$

Eq. 2.7 allows to discuss the possible sources of GW radiation. Considering a source of mass M and radius R, the quadrupole Q can be estimated in order of magnitude by $Q \sim sMR^2$, where s is a dimensionless factor capturing the geometry of the source: For a spherically-symmetric source, the quadrupole vanishes and $s = 0$; For asymmetric sources $s \geq 0$. A time-variable quadrupole requires a source in motion; We denote by v the velocity of the source. The characteristic dynamical timescale is thus $t \sim R/v$. We then find that:

$$L_{\text{GW}} \sim \frac{G}{5c^5}\left(\frac{Q}{t^3}\right)^2 \tag{2.8}$$

$$= \frac{G}{5c^5}\frac{s^2M^2v^2}{R^2} \tag{2.9}$$

$$= L^0_{\text{GW}}s^2\Xi^2\beta^6 \tag{2.10}$$

where we introduced the fundamental GW luminosity $L^0_{\text{GW}} = c^5/5G$, the source's compactness $\Xi = GM/Rc^2$ and the dimensionless velocity $\beta = v/c$.

The fundamental GW luminosity is colossal:

$$L^0_{\text{GW}} = 7\times10^{58}\ \text{erg/s} \tag{2.11}$$

$$= 2\times10^{25}L_\odot \tag{2.12}$$

$$= 4\times10^4 M_\odot c^2/\text{s} \tag{2.13}$$

However, to reach such luminosities, Eq. 2.10 teaches us that the source must respect a number of conditions: It must be strongly asymmetrical (large s); It must be relativistic in both the special (large β) and general (large Ξ) sense. The large exponents in Eq. 2.10 make these conditions strong.

This calculation points to the interaction or formation of compact objects as efficient production sites for GW radiation. For example, the formation of a black hole in an asymmetrical core-collapse supernova (for a review, see Abdikamalov et al., 2020); Historically, this source was perceived as the most promising for transient GWs (Virgo Collaboration, 2009). The sources that are the most interesting for GRB science and its many bridges to other domains of astrophysics are the merger of binary compact objects, or *compact binary coalescences*. What's more, they are much more asymmetrical than core-collapse supernovae and, indeed, these revealed the first and most prominent sources for GW astronomy. Before moving to describe this phenomenon, we simply mention that GWs can also be continuously produced by large-scale structures or phenomena such as inflation, phase transitions in the early Universe or cosmic strings (e.g., Christensen, 2019). Such sources are naturally of great interest to cosmology.

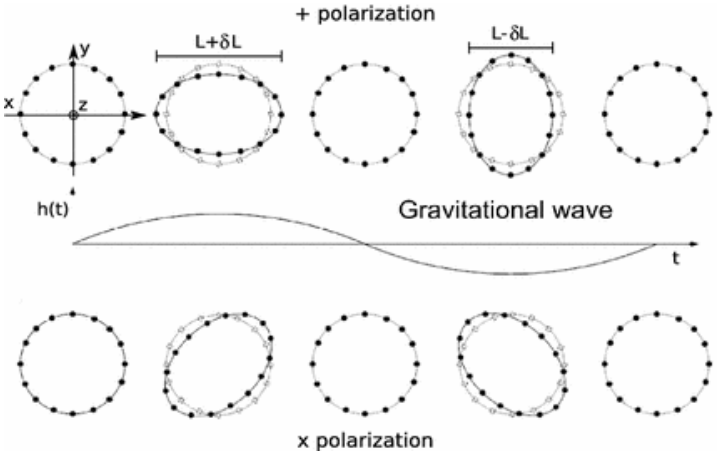


Figure 2.1: **Two polarization modes of gravitational waves.** In the transverse plane, the two polarization modes predicted by GR are named "+" and "×" because of their effects on the variation of the distances measured between particles disposed as a circle, as illustrated here (Casanueva Diaz, 2018).

2.2 Gravitational waves from compact binary coalescences

Consider a binary compact object in a circular orbit. We adopt this hypothesis to simplify the derivation; Because of angular momentum losses by GW radiation, binaries are expected to have circularized by the time they are close to merging.

Eq. 2.7 shows that this asymmetrical system will lose energy to GW radiation, and thus the orbits will become tighter, possibly leading to the merger of the two component objects. To describe the system's evolution, we will follow the lowest-order Newtonian approximation: We suppose the two objects are point-like; We suppose a quasi-static evolution such that

the orbits are Keplerian at all times. To detect merger signals in instrumental data, more sophisticated *Post-Newtonian expansions* are used; They expand the orbital dynamics and flux of energy and angular momentum at infinity in ever-increasing orders in v/c, the velocity of the orbiting objects (Blanchet & Damour, 1989).

Denote by $M = M_1 + M_2$ the total mass of the binary and a the time-varying orbital separation, i.e., the sum of the semi-major axes of the two orbits. Writing the mass density as $\rho = M_1\delta_{\vec{x_1}} + M_2\delta_{\vec{x_2}}$, Eq. 2.7 allows to determine the luminosity in GW radiation:

$$L_{\rm GW} = \frac{32G^4}{5c^5}\frac{M^3\mu^2}{a^5} \tag{2.14}$$

where $\mu = M_1M_2/M$ is the system's reduced mass.

The system's energy is the usual Keplerian energy:

$$E = -\frac{1}{2}\frac{GM\mu}{a} \tag{2.15}$$

Finally, equating the GW flux at infinity with the decrease in orbital energy $\dot{E} = -L_{\rm GW}$, we arrive at the follow ordinary equation:

$$\dot{a} = -\frac{64G^3}{5c^5}\frac{M^2\mu}{a^3} \tag{2.16}$$

which is readily solved as:

$$a(t) = a_0\left(1 - \frac{t}{\tau_c}\right)^{1/4} \tag{2.17}$$

where we have denoted by a_0 the binary's initial separation and by τ_c the binary's *inspiral time*:

$$\tau_c = \frac{5c^5}{256G^3}\frac{a_0^4}{M^2\mu} \tag{2.18}$$

$$= 14\ {\rm Gyr}\left(\frac{M}{3\ M_\odot}\right)^{-2}\left(\frac{\mu}{0.75\ M_\odot}\right)^{-1}\left(\frac{a_0}{5\ R_\odot}\right)^{4} \tag{2.19}$$

For a non-zero initial eccentricity e_0, we find:

$$\tau_c = \frac{5c^5}{256G^3}\frac{a_0^4}{M^2\mu}(1-e_0)^{7/2} \tag{2.20}$$

The inspiral time is such that $a(\tau_c) = 0$; It is the time before the binary separation formally vanishes. Of course, the physical extent of the objects implies that a binary neutron star will come into contact and produce an explosion before this time, and that binary black holes will coalesce. First, it is remarkable that τ_c is *finite*. Second, the GW luminosity $L_{\rm GW}$ has a very strong dependence on the binary separation: $L_{\rm GW} \propto a^{-5}$. Therefore, the system will emit a burst of GW right before the merger, when a is the smallest; The more compact the objects, the smaller the orbit can become before merging and the brighter the ultimate GW burst. Third, τ_c is very sensitive to both the initial separation and eccentricity. Fourth, the inspiral time is generally expected to be very long: Eq. 2.19 is normalized to near-Chandrasekhar masses $M_1 = M_2 = 1.5M_\odot$ and a separation close to that of the Hulse–Taylor binary pulsar.

The Husle Taylor system has been monitored for nearly four decades and its timing reveals an orbital decay due to GW radiation consistent with GR to within 0.2% (Hulse & Taylor, 1975; Weisberg & Huang, 2016); The first years of monitoring of this system had indirectly proved the existence of GWs. Its small, non-zero eccentricity shortens its inspiral time, it is expected to merge in ~ 300 Myr.

It is thus expected that binary compact object mergers manifest as bursts of GW; This fact was acknowledged during the "GR revival" (Dyson, 1963). Using Eq. 2.4, we can determine the time-dependent GW signal, also known as the *waveform*. For simplicity, we place the observer on a line of sight orthogonal to the orbital plane, and suppose that the detection plane is parallel to the orbital plane; To signify this particular setup, we add the superscript $^\perp$. In this case, the projection operators vanish and we have:

$$\begin{cases} h_+^\perp(t) &= -h(t)\cos\left(2\omega\left(t - \frac{D_L}{c}\right)\right) \\ h_\times^\perp(t) &= -h(t)\sin\left(2\omega\left(t - \frac{D_L}{c}\right)\right) \end{cases} \tag{2.21}$$

where D_L denotes the luminosity distance to the source and $\omega(t)$ the time-varying orbital frequency. The terms in the cos and sin are evaluations of ω at the retarded time $t_{\rm ret.} = t - D_L/c$, not multiplications. We thus understand that the GW signal has a frequency that is twice the orbital frequency of the system: $\nu(t) = 2\omega(t)$.

The GW amplitude is:

$$h(t) = \frac{4G}{c^4}\frac{\mu a^2\omega^2}{D_L} \tag{2.22}$$

$$= \frac{16^{1/3}G^{5/3}}{c^4}\frac{\mathcal{M}^{5/3}\nu(t)^{2/3}}{D_L} \tag{2.23}$$

Also, note the $1/D_L$ dependence of the GW signal, such that a increase in instrument sensitivity opens up a much larger detection volume $\propto D_L^3$ than in the electromagnetic domain ($\propto D_L^{3/2}$). As the system inspirals, ω increases and so does the GW amplitude; Using the equation for $a(t)$ (Eq. 2.17), we can predict the GW signal frequency as a function of time:

$$\frac{\nu(t)}{\nu(0)} = \left(1 - \frac{t}{\tau_c}\right)^{-3/8} \tag{2.24}$$

$$\to_{t\to\tau_c} \infty \tag{2.25}$$

The *GW inspiral signal* is thus a quasi-monochromatic signal with frequency $\nu(t) \propto (1 - t/\tau_c)^{-3/8}$ and amplitude $h(t) \propto (1 - t/\tau_c)^{-1/4}$, which both formally diverge at merger. The characteristic amplitude of the signal is given by the *chirp mass*:

$$\mathcal{M} = \frac{(M_1M_2)^{3/5}}{M^{1/5}} \tag{2.26}$$

introduced in Eq. 2.23.

This signal lasts until the *merger time*, when the objects' surfaces formally touch. At merger time, we have $a(t) \sim 2R$ for same-radii systems; Using Kepler's law, we find that the GW

frequency at this time is:

$$\nu_{\rm merger} = 2\omega_{\rm merger} \tag{2.27}$$

$$= 4 \text{ kHz} \left(\frac{\Xi}{0.1}\right)^{3/2} \left(\frac{M}{3\ M_\odot}\right)^{-1} \tag{2.28}$$

which is in the acoustic range. Such a signal converted to audio indeed sounds like a bird's chirp. We slightly anticipate in inviting the reader to contemplate the inspiral signal of GW170817, a binary neutron star merger detected at the dawn of the multi-messenger era: Fig. 3.3 (bottom panel) shows the signal's *time-frequency map*, i.e. the time-resolved Fourier transform, also known as the *spectrogram*. This signal was observed for the last 3000 orbits of the system, constituting the loudest chirp detected to date; The signal is therefore very clear in the time-frequency map.

We can now estimate the GW amplitude in the relevant frequency range, shortly before merger:

$$h = 3 \times 10^{-21} \left(\frac{\mathcal{M}}{1.3\ M_\odot}\right) \left(\frac{\nu}{1 \text{ kHz}}\right)^{2/3} \left(\frac{D_L}{10 \text{ Mpc}}\right)^{-1} \tag{2.29}$$

The bold simplicity of Eqs. 2.23 and 2.25 is beautiful; What else should we have expected from GR? They also contain a remarkable feature: By an algebraic manipulation, one finds that:

$$\mathcal{M} = \frac{c^3}{G} \left(\frac{5}{96\pi^{8/3}}\right)^{3/5} \dot{\nu}(t)^{3/5} \nu(t)^{-11/5} \tag{2.30}$$

such that plotting the time-frequency map allows to easily extract $\nu(t)$ and to directly measure the chirp mass. In actual observations, this is indeed the best-constrained parameter of the binary (Sec. 2.3), especially for low-masses systems such as binary neutron stars, for which the inspiral signal lasts longer; It does not even require to measure the signal amplitude precisely. Once the chirp mass is measured, the distance to the source follows from:

$$D_L = 4\pi^{2/3} \frac{G^{5/3}}{c^4} \frac{\mathcal{M}^{5/3} \nu(t)^{2/3}}{h(t)} \tag{2.31}$$

such that the distance can also be measured from the signal; It is sometimes said that compact object merger are "self-calibrating sources" because of this. In practice, this method has some serious limitations. First, instruments do not have access to $h(t)$ directly, but to the instrumental response, which depends on the technology employed and is degenerate with other binary parameters that we will discuss in Sec. 2.3—prominently, the binary's inclination with respect to the observer, which we ignored by placing the observer orthogonal to the orbital plane up to now. Second, it requires measuring $h(t)$ precisely; Current instruments are limited in data calibration to $\sim 1\%$ (Karki et al., 2016), which propagates to D_L with larger uncertainty because of this degeneracy. Third, in an astronomical context, signals are redshifted: The frequencies observed are $\nu_{\rm obs} = \nu/(1+z)$ and $\dot{\nu_{\rm obs}} = \dot{\nu}/(1+z)^2$. Thus, using Eq. 2.30 leads to measuring a chirp mass of $\mathcal{M}_{\rm obs} = (1+z)\mathcal{M}$, which is sometime called the *redshifted chirp mass* by a language abuse. Therefore, using Eq. 2.31 to measure D_L, we actually obtain:

$$(1+z) D_L(z) = 4\pi^{2/3} \frac{G^{5/3}}{c^4} \frac{{\mathcal{M}_{\rm obs}}^{5/3} \nu_{\rm obs}(t)^{2/3}}{h(t)} \tag{2.32}$$

such that there is now degeneracy between z and any of $\mathcal{M}$ or D_L.

The redshift can be cosmological or appear as such when the signal was magnified by lensing by a massive object. The latter hypothesis was evoked when the first detected signals revealed the existence of very massive binary black holes (Sec. 2.3); However, the likeliness of lensing is very small and it cannot have affected all the detected signals (Pang et al., 2020). Theoretically, the inspiral signal can contain a signature of the object's redshift introduced by tidal interactions between the component neutron stars, which have well-defined frequencies in the system's rest frame (e.g., Messenger & Read 2012, see also Chap. 6).

Naturally, the above calculation will lose predictive power in the last orbits of the binary, when the orbital velocities become relativistic and finite-size effects appear. Upon contact, the physics at play change drastically and many uncertain quantities must be taken into account, such as the equation of state of nuclear matter in the case of a binary containing a neutron star. For this *merger phase*, numerical relativity simulations are required to evolve the system and predict the GW signal (e.g. Baiotti & Rezzolla, 2017). In the case of binary black holes, there is less uncertainty in the merger phase. Finally, note that in the case of a very asymmetrical system or some cases of neutron star-black hole binaries, the lighter object can be tidally disrupted or enter the horizon before the binary becomes relativistic; Then, no actual merger phase occurs.

The merger leads to the formation of a transient compact that eventually collapses to a black hole. In the case of binary neutron stars, the nature, lifetime and behavior of the transient object is still very uncertain (Sec. 3.2). In the case of a binary black hole, the merger phase directly results in a perturbed black hole which relaxes to the stationary Kerr solution during the *ring-down phase*. The GW signal from this phase is more readily predicted than that from the merger phase, because perturbations of the Kerr metric have a robust theoretical background in the theory of quasi-normal modes (Leaver, 1985). For binary neutron stars, a ring-down signal is also predicted, however the large uncertainties on the merger phase disallow to predict the state of the perturbed black hole eventually formed at the end of the merger phase. In Fig. 2.2, we schematically describe the inspiral, merger, and ring-down phases for a binary black hole system.

The most updated merger rates for compact binary coalescences in the local Universe are $23.9^{+14.9}_{-8.6}$ Gpc^{-3}yr^{-1} for binary black holes and 320^{+490}_{-240} Gpc^{-3}yr^{-1} for binary neutron stars (90% confidence intervals, Abbott et al., 2021b). The merger rate of binaries composed of a neutron star and a black hole are currently constrained to $22-242$ Gpc^{-3}yr^{-1}, depending on the hypothesis made for the distribution of component masses in these systems (Abbott et al., 2021c).

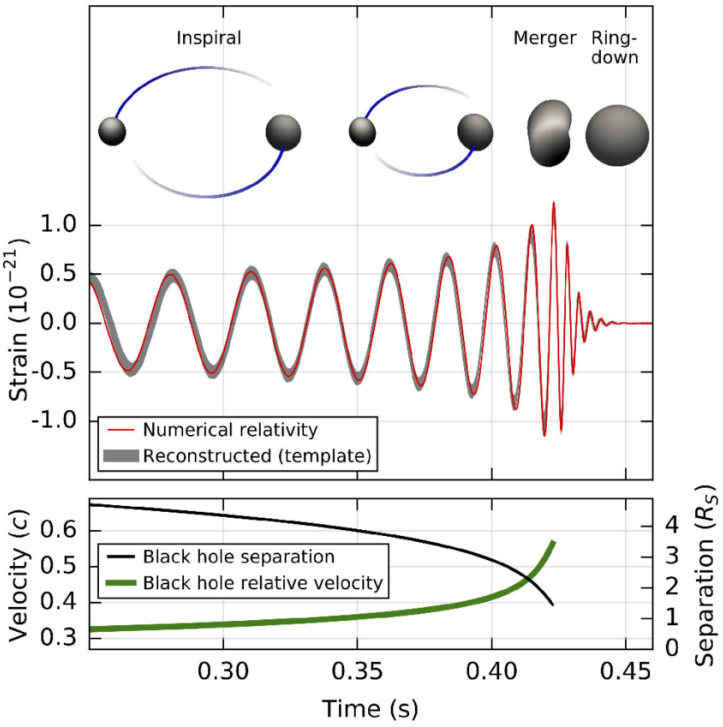


Figure 2.2: **Inspiral, merger and ring-down phases of a binary black hole merger.** The inspiral phase leads to the horizons merging; The subsequent merger phase requires numerical relativity calculations; It leads to the formation of a perturbed Kerr black hole, which relaxes to a stationary state by emitting the last GWs: the ring-down signal (Abbott et al., 2016a).

2.3 The observation of gravitational waves

2.3.1 Gravitational interferometers

Current detectors of GW bursts leverage the above-mentioned effect of GWs on distances measured by light travel time. They are optical setups where laser beams are set to follow different optical paths before being recombined and interfere together. During the passage of a GW, the different paths will be affected differently, such that the interference pattern will change. This method is called *gravitational interferometry* (Maggiore, 2007). By adapting the optical setup and with sufficiently large optical lengths, this method can be made effective for a broad range of GW frequency bands.

This method requires to minimize any variations in the optical lengths traveled by the laser beams not due to the passage of a GW; In addition, larger contrasts in the interference pattern will aid the detection of transient changes. Gravitational interferometers (IFOs) thus rely on powerful and stable lasers, and the attenuation of any parasite motion of the optical setup. The corresponding technology—high-power, high-stability lasers and super-attenuating suspensions—was developed in the 1980s, allowing to project kilometer-sized IFOs for the detection of astrophysical GW bursts.

In Fig. 2.4, an aerial view of the Virgo interferometer (Accadia et al., 2012) lodged at the European Gravitational-wave Observatory can be found (left), as well as a schematic description of its optical setup (right). It is basically a Michelson interferometer with 3-km-long perpendicular arms; The laser is a Nd–YAG 1064 nm laser with an injected power of 200 W. The laser beam is split at the beam splitter (BS) to enter the cavity formed by the *test masses* (WI, WE, NI and NE), which are silica mirrors. The beams are then recombined to reach the photodiode (B1) that measures the interference state. The cavities are Fabry-Pérot cavities allowing to virtually increase the optical length traveled by the beams, increasing the interferometric precision. The power-recycling mirror (PRM) allows to reflect power back into the instrument, further increasing the contrast of the instrument. The signal-recycling mirror (SRM) allows to tune the instrument to a given frequency band. In total, the six optical benches include thousands of optical components. The steady-state power in the Fabry-Pérot cavities is $\sim$ 100 kW. The optical setup affords about two orders of magnitude in sensitivity with respect to a simple Michelson setup.

The change in the interferometric pattern due to the passage of a transient GW can only be detected if the photodiode's measured signal varies by more than its typical fluctuations in the instrument's *static state*, in the absence of a GW. The static fluctuations have different amplitudes at different timescales; This is captured by the noise power spectrum density (PSD) defined by:

$$S_N(\nu) = \lim_{T\to\infty} \frac{1}{T} \mathbb{E}\left[\left|\int_0^T \mathrm{d}t e^{-i\nu t} P_{B_1}(t)\right|^2\right] \tag{2.33}$$

where $P_{B_1}(t)$ is the photodiode's static signal. $S_N(\nu)$ can be understood as the variance of the Fourier transform at ν of the static signal, and $S_N^{1/2}(\nu) = \sqrt{S_N(\nu)}$ as the typical amplitude of static fluctuations on timescale $1/\nu$. In practice, S_N is estimated by averaging various static runs of hundreds of seconds, so as to eliminate possible transient signals.

S_N allows to define the signal-to-noise ratio (SNR) of an astrophysical signal through:

$$\rho^2 = \int \mathrm{d}\nu \frac{\tilde{s}(\nu)\overline{\tilde{s}(\nu)}}{S_N(\nu)} \tag{2.34}$$

where $\tilde{s}$ is the Fourier transform of the instrumental response expected from the astrophysical signal.

In Fig. 2.3, we represent the *sensitivity curve*, i.e. $S_N^{1/2}$ for the design configuration of the Virgo interferometer; It features the main sources of noise and the total noise. Virgo is most sensitive to the 50–1000 Hz band, in accordance with our estimations for the chirp of binary neutron star mergers (Sec. 2.2).

The photodiode is limited by its efficiency in detecting photons; This produces a noise dubbed *shot noise*. In the absence of a GW and even with a perfectly stable laser, the power detected by the photodiode fluctuates in time through a Poisson-

type process. This produces an incompressible frequency-independent noise, represented in green in the sensitivity curve. Furthermore, the laser's frequency fluctuates at a high frequency, such that high-frequency GW-induced variations in the interference cannot be distinguished; This results in the shot noise increasing with frequency, such that the instrument is ultimately limited by shot noise at high frequency ($\nu \geq 5$ kHz).

At very low frequency ($\nu \leq 2$ Hz), the super-attenuators are less efficient and seismic activity induces motion on the optical setup. In this range, the instrument is thus limited by seismic noise (red). In the central range, the instrument is limited by the thermal motion of the super-attenuators themselves and of the test masses. Current avenues for IFO improvement include the squeezing of light at input to reduce the uncertainty on photon phases, specific coating of test masses to reduce thermal noise and light diffusion therein, and the never-ending hunt for *stray light*—scattered laser beams—in the optical setup.

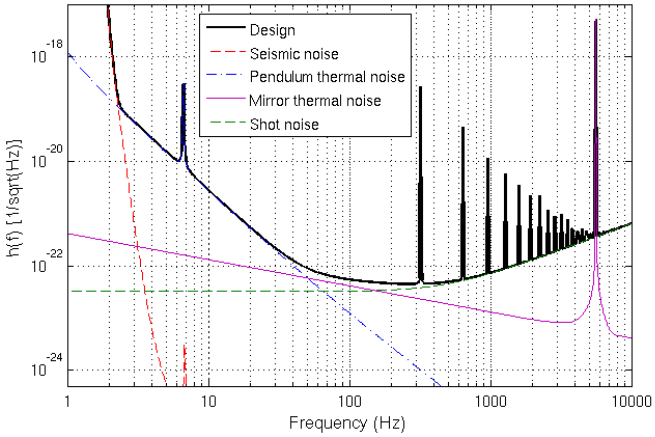


Figure 2.3: **Projected sensitivity curve for the Virgo design.** The instrument is most sensitive to the 50–1000 Hz band, in accord with theoretical predictions of the GW bursts of binary compact mergers. From the lowest to highest frequencies, Virgo's noise is mostly of seismic, thermal motion and finally photon counting origin. See text for more details (Accadia et al., 2012).

2.3.2 Response to a inspiral signal and matched-filtering

In Sec. 2.2, we determined the inspiral GW waveform measured on a line of sight aligned with the system's orbital axis, as given by $h_+^\perp$ and $h_\times^\perp$ (Eq. 2.21). In fact, the GW signal is not isotropic. Denote ι the *inclination angle*, that is, the angle between the system's orbital momentum and the line from the observer to the system. Then, the propagation direction of the GW is still radial and in a plane orthogonal to this direction, and the two polarization states are (e.g. Finn & Chernoff, 1993):

$$\begin{cases} h_+(t) &= \frac{1+\cos^2\iota}{2} h_+^\perp(t) \\ h_\times(t) &= \cos\iota\, h_\times^\perp(t) \end{cases} \tag{2.35}$$

The flux is $F_{\rm GW} \propto (1 + 6\cos^2\iota + \cos^4\iota)$ and energy is preferentially radiated in the direction of the angular momentum.

The IFOs are sensitive to the optical length of the lasers in the two arms. Denote by $\overrightarrow{n}_{\rm N}$ and $\overrightarrow{n}_{\rm W}$ unit vectors in the directions of the two arms (e.g., North and West) and L_0 their static length. The arms (~ 1 km) are much smaller than the wavelength of the GW (~ 1000 km for $\nu \sim 100$ Hz). In other words, the light travel time through the instrument is much less than the GW frequency—it is the *stationary phase approximation*. Therefore, the time-varying difference in optical paths in the two arms—the *strain*—is given by (Eq. 2.3):

$$m(t) = L_0 \left(\frac{1}{2} h_{ij}(t) n_{\rm N}^i n_{\rm N}^j - \frac{1}{2} h_{ij}(t) n_{\rm W}^i n_{\rm W}^j \right) \tag{2.36}$$

The explicit expression for m is usually given in the *wave coordinate system* described in Fig. 2.5 (left). The source's sky location is given by spherical angles θ and ϕ; The source's orientation is given by the two angles ι and ψ. The inclination angel is as defined above; The *polarization angle* ψ is measured between the projections of the y axis and of the source's angular momentum on the tangent plane (Abadie et al., 2010).

In these coordinates, we have (Finn & Chernoff, 1993):

$$\frac{m(t)}{L_0} = F_+(\theta,\phi,\psi) h_+(\iota,t) + F_\times(\theta,\phi,\psi) h_\times(\iota,t) \tag{2.37}$$

with the following *antenna patterns*:

$$\begin{cases} F_+ &= \frac{1}{2}\cos 2\psi \left(1+\cos^2\theta\right)\cos 2\phi - \sin 2\psi \cos\theta \sin\phi \\ F_\times &= \frac{1}{2}\sin 2\psi \left(1+\cos^2\theta\right)\cos 2\phi + \cos 2\psi \cos\theta \sin\phi \end{cases} \tag{2.38}$$

We thus understand that individual IFOs are directional detectors. In Fig. 2.5 (right), we represent the F_+ antenna pattern of one of the two Laser Interferometer Gravitational-wave Observatories (LIGOs, Abbott et al. 2004). This represents the sensitivity to compact binary mergers in different sky locations, on average on ι and ψ. For a purely +-polarized wave—such as the inspiral signal of an edge-on binary (Eq. 2.35)—the IFOs are most sensitive at the zenith and nadir, and have four blind spots in the detector plane, on the bisectors and their perpendiculars. The $F_\times$ antenna pattern is similar, rotated by 45 deg around the instrument. Overall, for a generic inspiral signal, a single IFO is most sensitive at zenith and nadir (both these are equivalent), and the blind spots remain, though less extended.

The signal expected from a typical merger is very weak (Eq. 2.29). Only in exceptional cases such as GW170817 (Sec. 3.2.3) or GW150914 (see below) does the inspiral waveform clearly appear in raw data. Therefore, inspiral signals are searched for using *matched filtering* techniques (Abbott et al., 2020a): A bank of numerical inspiral signals named *templates* are prepared for a set of system properties (inclination, chirp mass, etc.) and sky locations. In real time, the stream of noisy IFO data is cross-correlated against each of these templates. When a certain template produces a SNR ratio above a *triggering threshold*, the relevant data is conserved for finer analysis in search for the signal and source parameter inference.

On September 12, 2015 the two LIGO IFOs—LIGO-Hanford and LIGO-Livingston—started the first GW observing run, O1. Observing runs are long-term configurations of

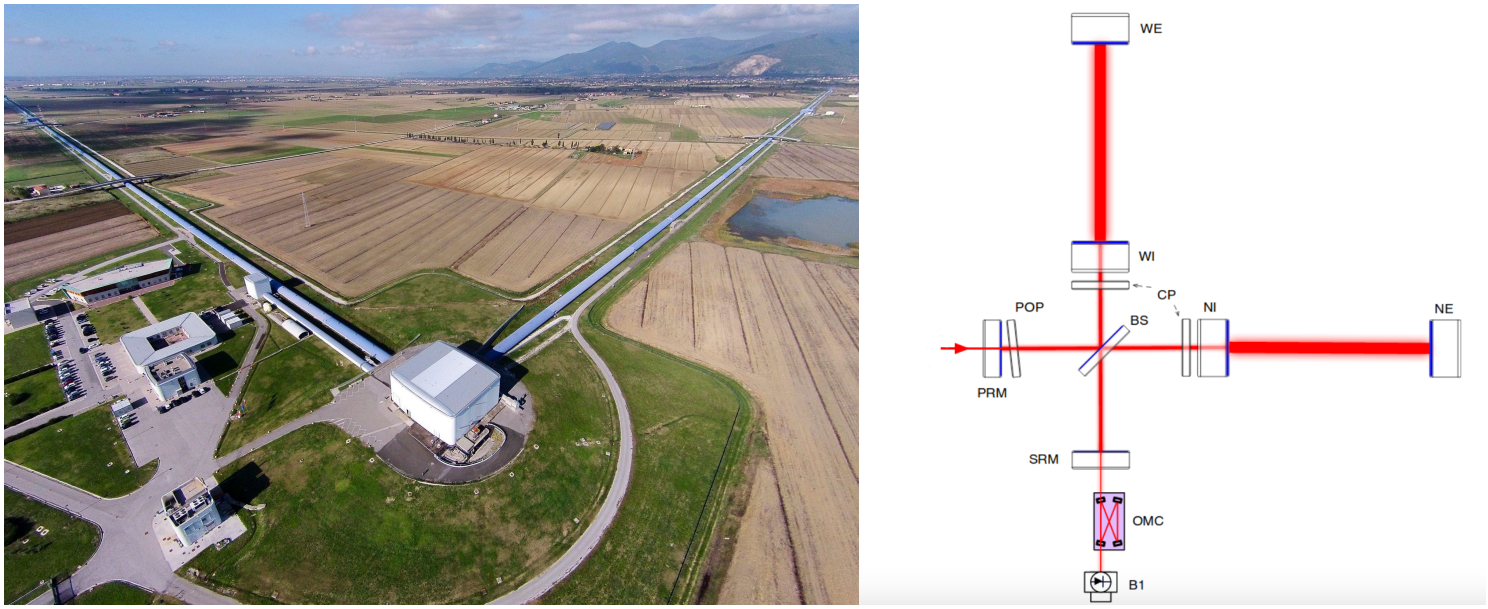


Figure 2.4: **Virgo gravitational interferometer.** Left: An aerial view of the site near Pisa, Italy. The perpendicular arms are 3 km long. The instrument is composed of six optical benches with thousands of optical components (credit: European Gravitational-wave Observatory). Right: The optical setup of Virgo. It is essentially a Michelson interferometer with Fabry-Pérot cavities in the place of arms and with power and signal recycling. See text for details.

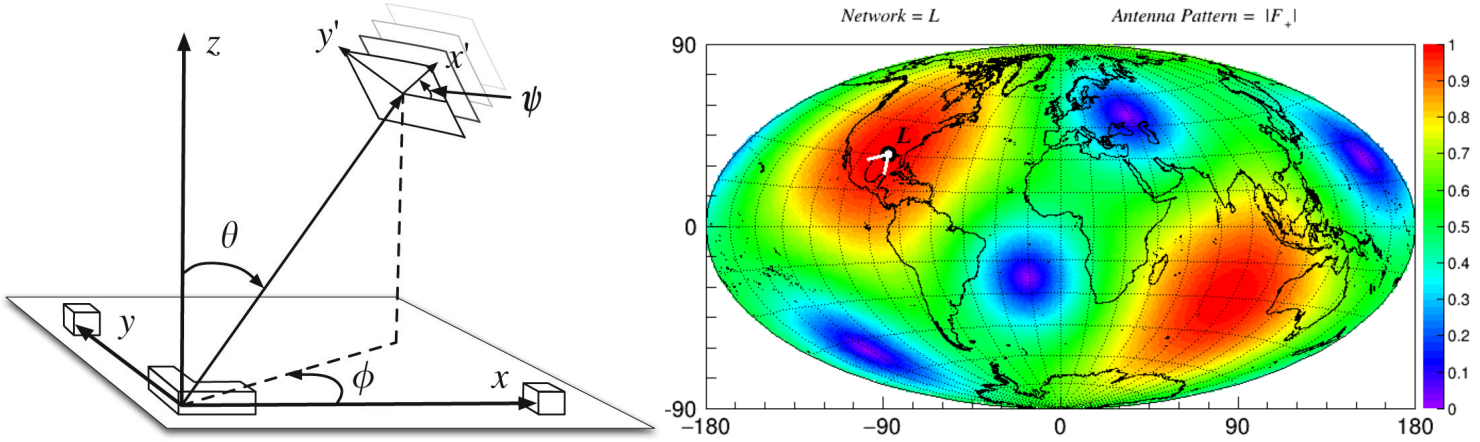


Figure 2.5: **Antenna pattern of a single gravitational-wave interferometer.** Left: The wave coordinate system allows to specify the sky position (θ, ϕ) and orientation (ι, ψ) with respect to a rectangular instrument (Abadie et al., 2010). Right: The F_+ antenna pattern for a single LIGO instrument. The pattern is symmetrical with respect to the detector's plane, and has four blind spots. The instrument is most sensitive at zenith. (Prodi, 2018).

IFOs during which no changes are done to the instrumental setup so as to keep the same sensitivity curves. The matched-filtering algorithms are run simultaneously on the data stream of all IFOs in the network and triggers are made on signals with matching source parameters in at least two of the participating instruments.

Success came on September 14, 2015: The network triggered on a candidate transient signal, later confirmed as a binary black hole (BBH) merger and dubbed GW150914[2] (Abbott et al., 2016a). After trigger, a refined Bayesian analysis on the IFOs' data allows to jointly constrain all the event's parameters—source parameters, system orientation and sky location—, including the inspiral, merger and ringdown phases when detected. In Tab. 2.1 we collect the most important inferred parameters of GW150914 (Abbott et al., 2016b). The large masses surprised the astrophysical community, but further detections saw similar mass ranges (Abbott et al., 2019b); These triggered discussion on the possible formation channels of such systems, another vibrant field awoken by the gravitational-wave era.

Table 2.1: **GW-inferred properties of GW150914, the first GW transient signal detected.** Masses are given in the source frame, and all ranges are 90% confidence intervals (Abbott et al., 2016b).

Property	GW constraint
Progenitor primary mass M_1	$36^{+5}_{-4}\ M_\odot$
Progenitor secondary mass M_2	$29^{+4}_{-4}\ M_\odot$
Chirp mass $\mathcal{M}$	$28^{+2}_{-2}\ M_\odot$
Remnant BH mass	$62^{+2}_{-2}\ M_\odot$
Luminosity distance D_L	410^{+160}_{-180} Mpc
Inclination angle ι	$20 - 170$ deg
Projected area of sky-location constraint	610 deg^2

Tab. 2.1 illustrates two weaknesses of IFOs: measurement of ι and source localization. The first comes from the dependence of h on ι through a polynomial in $\cos\iota$, which is not very steep (Eq. 2.35); Thus a given signal does not provide a good inclination constraint. This uncertainty in turn impacts the measurement of D_L as discussed previously. Adding more IFOs to the network only modestly improves the ι constraint (Veitch et al., 2012; Ghosh et al., 2016a). The sky-location constraint partially comes from the antenna patterns F_+ and $F_\times$, but is however improved by triangulation of the transient arrival times between IFOs. Even if the baselines are small, triangulation is in fact the dominant contribution to precision in localization. At the time of GW150914, only two IFOs were online, explaining the poor constraint. In time this will improve when more instruments come online, allowing for very good localization as of run O4 (Sec. 2.4).

The large uncertainty on ι and subsequently on D_L hinders some astrophysical applications of GW transient signals. Early on, it was understood that being able to measure D_L from the waveform itself (Eq. 2.31) made such signals particularly fit for cosmology and, e.g., the measurement of the Hubble constant (e.g., Schutz, 1986; Nissanke et al., 2013). This uncertainty and degeneracy thus calls for methods to improve the ι measurement for individual events, possibly by complementing the GW constraint with electromagnetic information. In Chap. 7, we will explore how multi-messenger observations of binary neutron star mergers can provide such a method, through leveraging the electromagnetic counterparts to the GW transient.

The peak GW luminosity of GW150914 was $\sim 200\ M_\odot c^2/\mathrm{s}$; The total emitted energy was $\sim 3\ M_\odot c^2$ (Abbott et al., 2016b). BBH mergers are the most powerful phenomena known in the Universe.

2.4 Multi-messenger astronomy in the GW era

2.4.1 The GW detector network

Successive observing runs saw more IFOs join the network: The O1 run (September 12, 2015 – January 19, 2016) comprised the two LIGOs; The O2 run (November 13, 2016 – August 25, 2017) was joined in August by Virgo; The O3 run (April 1, 2019 – March 27, 2020) was joined in February by the Kamioka Gravitational-Wave Detector (KAGRA, Kagra Collaboration et al. 2019), forming the global network at the time of writing, the *LIGO-Virgo-KAGRA Collaboration* (LVKC). The KAGRA IFO innovated by its cryogenized test masses, sapphire mirrors and underground location. In the near future, the network should comprise a third LIGO-like instrument, LIGO-India (Iyer, 2015); The network's F_+ antenna pattern in represented in Fig. 2.6 (right), where the positions and orientations of member instruments are also figured. In this configuration, the collective sky-location response is much more homogeneous than for a single instrument: There are no more blind spots. We define the *network SNR* of a signal as $\rho^2_{\mathrm{net.}} = \Sigma\rho_i^2$, where the ρ_i's are the SNRs in the individual IFOs. The triangulation of transients by the many IFOs opens the way to locating sources in the sky and conducting electromagnetic follow-up in the context of *GW astronomy*.

The performance of an individual interferometer and a network is captured by a number of *figures of merit*:

- An IFO's *horizon distance* H is the distance to which a standard source can produce a given trigger threshold under the best-possible observing conditions: The source is seen face-on ($\iota = 0$) and is placed at the detector's zenith. The standard sources are a double-$1.4M_\odot$ binary neutron star merger (BNS), a double-$30M_\odot$ BBH merger and a 1.4-$10M_\odot$ neutron star-black hole binary merger (NSBH).

- An IFO's *range* is the average distance to which a standard source can be detected. The average is taken over all possible source orientations and sky positions. The BNS range is the most common figure of merit for an individual IFO. In Fig. 2.6 (left), we report the achieved

[2] GW transient signals follow the same naming convention as GRBs.

(O1, O2 and O3) and projected (O4 and O5) BNS ranges for the instruments for various observing runs (Abbott et al., 2020a). In Appendix A, we provide some relations between the horizon and range, and derive a simple criterion for the detection of an inspiral signal based only on these figures of merit; The criterion is fundamental to our multi-messenger population model of Part II.

- The network's *search volume* for a given run, denoted by VT, is the time-integrated volume in which sources are detectable in GWs, on average on source orientation and polarization angle. The natural subsequent figure of merit is the number of standard events detected: During O1 and O2, the network confidently detected 10 BBHs and 1 BNS—GW170817, see Chap. 3. During O3, there were projected 1^{+12}_{-1} BNS, 0^{+19}_{-0} NSBH and 17^{+22}_{-11} BBH detections; The confident detections for the first half of O3 are already published (36 BBHs and 3 systems with at least one component mass $\leq 3\ M_\odot$ with uncertain classification, Abbott et al. 2021a); The LVK collaboration has just announced the two first confident NSBH inspiral detections during the second half of O3 (Abbott et al., 2021c). Projections for O4 are 10^{+52}_{-10} BNSs, 1^{+91}_{-1} NSBHs, 79^{+89}_{-44} BBHs (Abbott et al., 2020a). The improvement of IFO sensitivity and the addition of IFOs to the network allow the expected number of detections to make great jumps in detection rate between observing runs, especially thanks to the $1/D_L$ flux dependence of GW signals.

- In the context of multi-messenger astronomy, a pivotal point is the capability to constrain the source's sky position to allow for electromagnetic follow-up in search for counterparts. A figure of merit is the *median sky-projected size of the 90%-confidence sky location of the source*, or GW skymap. In this respect, it is mostly the number of IFOs having detected the signal that allows for small GW skymaps, through source triangulation. For GW170817, the addition of Virgo data decreased the GW skymap from a size of 190 to 16 deg^2 (Sec. 3.2.1). From O3 to O4, the addition of KAGRA will decrease the median GW skymap from 270 to 33 deg^2 for BNS mergers, even though the sources will be detected on average at larger distances (Abbott et al., 2020a).

2.4.2 GW astronomy: from trigger to follow-up

The automatic noticing of GW triggers to the astronomical community was started with O3 in 2019, through the GCN network (introduced for GRB localization circulation, Sec. 1.2.3). Within seconds after the matched-filtering algorithm identifies a candidate transient GW, a low-latency inference of the candidate's distance and localization is started with the BAYESTAR algorithm (Singer & Price, 2016). A *preliminary* GCN Notice shares the main results of the matched-filtering and of this parameter inference to the astronomical community (Abbott et al., 2020a):

- The source localization given as a projected skymap and the direction-dependent posterior distribution of the source's luminosity distance, drawing a three-dimensional error box.

- The probabilities that the candidate is in the five following empirical categories: BNS merger (both component masses are $\leq 3\ M_\odot$), BBH merger (both component masses are $\geq 5\ M\odot$), NSBH merger (one mass $\leq 3M_\odot$, the other $\geq 5M\odot$), "mass gap" merger (at least one mass between 3 and 5 $M_\odot$) and "terrestrial" (the signal is of non-astrophysical origin). This last probability is estimated by comparing the signal with the matched-filtering output run on many injections of artificial signals.

- The probability that at least one of the components is a neutron star (empirically, $\leq 3M_\odot$) and the probability that a non-zero amount of baryonic matter was ejected from the system. This last probability is deduced from theoretical models with the inferred component masses and spins (Abbott et al. 2020a and references therein); It is used to encourage adapted electromagnetic follow-up searches of the event.

Note that at this stage—seconds after GW trigger–, no human intervention occurs on the algorithms and the information circulated. Afterwards, human vetting starts to verify detector and pipeline status and data quality. Meanwhile, mid-latency parameter inference starts with full Bayesian statistics, through the LALInference algorithm (LIGO Scientific Collaboration, 2018). If the candidate passes the human vetting, then the *initial* GCN Notice and GCN Circular are sent with the refined skymaps and probabilities mentioned above, within hours after trigger. In case of a negative result from human vetting, the *retraction* GCN Notice is sent. Thereafter, further refined parameter inference is conducted with waveforms including, e.g., full treatment of component spins or neutron star tidal interaction. In case of significant improvement of parameter inference by this high-latency activity, *update* GCN Notices are sent; This can occur up to weeks after trigger.

2.4.3 GW astronomy: follow-up strategies

Once the GW trigger information is notified to the astronomical community, searches for electromagnetic counterparts can begin. Electromagnetic counterparts can come from autonomously triggering instruments, such as GRB missions. In the case of compact object mergers, the GRB occurs very early after the GW trigger and is rather used to further inform the electromagnetic follow-up on, e.g., source location. Here, we will focus on active follow-up that uses the GW-inferred information. Such follow-up usually follows three phases: candidate counterpart identification, candidate characterization and confirmation, and long-term source sampling:

- Candidate identification consists in finding variable sources in the GW skymap which could be electromagnetic counterparts to the event. Depending on instrumental capabilities, this phase can be led by three main

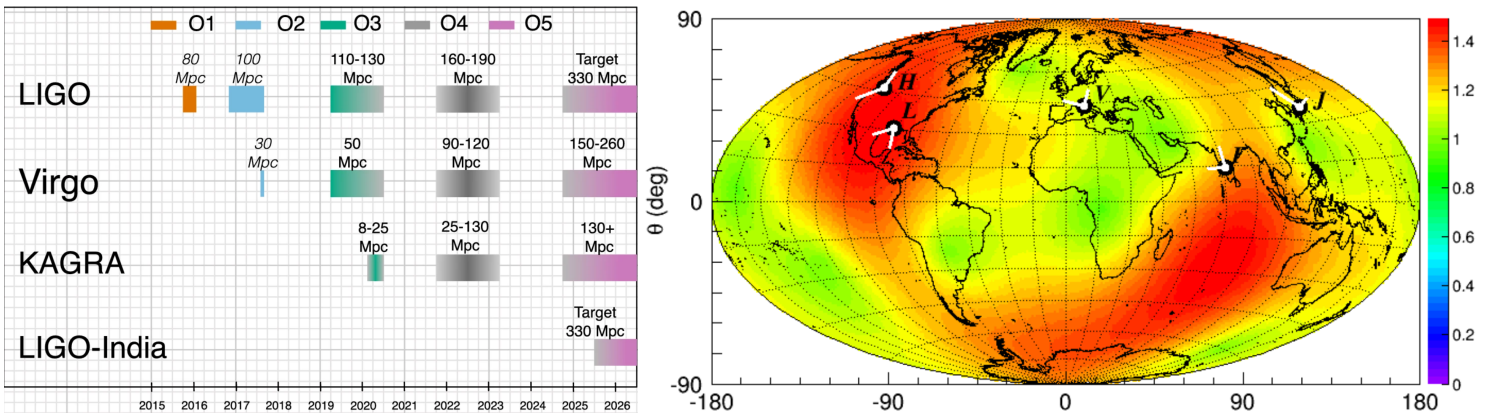


Figure 2.6: **Performance of the planned network of gravitational interferometers.** Left: Projected sensitivities of the main existing and upcoming instruments, given as individual ranges for a standard BNS merger (Abbott et al., 2020a). Right: The F_+ antenna pattern for the projected network. The pattern no longer has any blind spots, and is almost uniform on the sky (Prodi, 2018).

strategies. Survey instruments with high-cadence image the entire sky periodically; Crossing the automatically-identified variable sources from two passages with the GW skymaps can provide a list of candidate counterparts. *Tiling* consists in progressively imaging the entire GW skymap starting from the areas with more enclosed probability (Ghosh et al., 2016b); Comparison with archival images then allows to identify candidate counterparts. Finally, *galaxy targeting* starts by crossing galaxy catalogs with the GW skymap and imaging the individual galaxies in search for variable objects (Gehrels et al., 2016); Beyond their presence in areas with more or less GW probability, galaxy targeting can prioritize targets according to quantities correlated with hosting GW sources, such as galaxy type, mass or star-forming activity (Ducoin et al., 2020). Galaxy targeting is most adapted to small FoV instruments and naturally limited by galaxy catalog completeness, which is only partial as of distances of 200 Mpc (Kulkarni et al., 2018), well in reach of O4 configurations. The tiling strategy is limited by the size of the GW skymaps. During O3, only instruments such as the Zwicky Transient Facility and PanSTARRS—in fact survey instruments—were able to fully tile the northern GW skymap for the likely-BNS merger GW190425 (Coughlin et al. 2020c, Sec. 5.4). As of O4, the drastic improvement of typical skymap sizes should render tiling an effective strategy for candidate identification. Beyond the GW skymap, both galaxy targeting and tiling strategies can use the estimated source distance to adapt their search depths. While these three strategies are currently mainly realized in the optical bands, tiling in the X-ray is also effective (e.g., with *Swift*/XRT, Klingler et al. 2019) and planned (e.g., with *SVOM*/MXT, Sec. 1.5), and radio surveys could open up a new perspective for candidate identification (Dobie et al., 2021).

- Candidate counterparts are circulated within the astronomical community, leading to the candidate characterization phase. The candidate identification phase likely led to a large number of variable sources, including objects not related to the GW signal. These candidates must be characterized spectrally and photometrically in order to identify the counterpart to the GW trigger. Expected counterparts to BNS mergers, such as the thermal transients named *kilonovae* (Metzger 2019, Sec. 3.2.2) have particular spectra and color evolution that can guide the search. Candidate characterization can be limited by instrument availability and the number of variable sources to characterize. For GW190425—localized to a area of 8300 deg^2 and a distance of 150 ± 60 Mpc (Abbott et al., 2020b)—, more than fifty counterpart candidates were reported, all eventually dismissed. The identification of the counterpart is the pivotal point of electromagnetic searches, and currently the bottle-neck of multi-messenger astronomy because of the sheer number of candidates. This should be improved by the smaller skymaps expected as of O4. To help the identification, it has been advocated to release more detailed GW-inferred properties of the source, such as the chirp mass and component tidal deformability (Barbieri et al., 2020; Nicholl et al., 2021). The authors argue that using theoretical models with these parameters released in low-latency could help to predict the expected kilonova light curves and tailor the searches to the specific events.

- Once the counterpart is identified, a number of breakthroughs occur: First, pinpointing the source's location in the sky allows to lift some of the related degeneracy in the GW signal and therefore obtain even better source parameter estimations such as inclination, masses or spins from the GW data. Second, determining the host galaxy allows to measure the source redshift which further decreases uncertainty in the GW parameters by informing the source distance. Conversely, the redshift opens the way to using the source for cosmology, for example to make an independent measurement of the Hubble constant (Chap. 6).

Third, knowing the source's location allows to carry sub-threshold searches for GRB or high-energy neutrino signals, possibly revealing more counterparts (Albert et al., 2017). Finally, pinpointing the system in the sky leads to concentrate follow-up efforts on a single source and to monitor this source in all electromagnetic bands for any duration necessary, leading to long-term sampling of the electromagnetic counterparts. GW170817, the event that kicked off the multi-messenger era with gravitational waves (Chap. 3) revealed a rich set of electromagnetic counterparts and is still monitored at the time of writing, almost four years after trigger.

For a more detailed description of the limitations of these phases, please see the discussion around our multi-messenger population study, in Sec. 5.7.

In the future, the GW network will comprise IFOs with longer arms (Cosmic Explorer, 40 km, Reitze et al. 2019) or other geometries (Einstein Telescope, Punturo et al. 2014). These will have much greater reaches than the existing and near-future instruments, out to cosmological redshifts. This will affect the perspectives for electromagnetic follow-up of GW signals; We discuss this topic in more depth in the discussion of multi-messenger cosmology, in Sec. 6.8.

2.5 Conclusion

GWs are oscillations of the space-time metric. Rapidly moving, massive, asymmetrical systems such as tightly orbiting binary compact objects are powerful sources of GWs. Because of this radiation, these binaries grow closer and closer until they merge in a last burst of GWs. These merger signals are the most luminous events in the known Universe. The transient GWs emitted by the mergers of BNSs and BBHs peak in the acoustic band, and instruments based on optical interferometry allow to detect these signals. Analysis of the GW signal affords insight on the sources that is complementary to the information carried by electromagnetic radiation, opening the view for multi-messenger studies of binary compact objects.

The first such signals were detected in 2015; The very massive black holes at hand surprised the astronomical community, in a foretaste of what the GW era could bring to astrophysics. As more interferometers came online, the network's capability to locate the sources in the sky improved.

This culminated in August 2017, when the source of GW170817 was pinpointed after telescope searches in the GW-inferred skymap found a transient optical signal, opening a new multi-messenger era. In Chap. 3, we will describe this event, the rich set of electromagnetic counterparts and wealth of astrophysical breakthroughs it brought, and how it motivated the doctoral work presented in this thesis.

Chapter 3

GW170817 & GRB170817A: Welcome to a new multi-messenger era

Abstract

A new era of multi-messenger astronomy opened during the O2 gravitational-wave observing run, when the first binary neutron star merger was detected both in the gravitational domain by a three-interferometer network and in the electromagnetic domain by a historic multi-wavelength follow-up effort. This event and its rich set of counterparts—a weak short GRB, a kilonova, a long-lived afterglow—resulted in major breakthroughs in physics and astrophysics. While its first afterglow data points were not conclusive on the structure of the outflow from the merger, an extraordinary imagery campaign revealed that a relativistic jet had indeed been launched during the merger. Fine analysis of the afterglow data thus afforded unique insight into the structure of the jets that produce some GRBs. This event was the prime motivation for the work presented in this manuscript. We describe it in detail from the gravitational-wave signal to the latest observations and finally motivate the doctoral work.

3.1 Introduction: a historic event

On August 17$^{\text{th}}$ 2017, the first binary neutron star (BNS) merger was observed through gravitational-wave (GW) and electromagnetic signals. This event, GW170817, was rich with electromagnetic counterparts to the GW trigger and inaugurated a new multi-messenger era. In this chapter, we will describe the signals from GW170817 on various timescales and frequency bands, from the merger up to the latest observations. We will develop the major breakthroughs in astrophysics that this event brought, and finally motivate the questions posed by this event that the doctoral work presented in this manuscript has addressed.

The Virgo interferometer joined the two LIGO instruments in the O2 observing run on August 1$^{\text{st}}$ (Chap. 2.4.1). On August 17$^{\text{th}}$, a GW observation by the three-instrument configuration occurred: an inspiral phase of the merger of a BNS. This joint observation allowed the first astronomically significant triangulation of the source in the sky. With a slight delay with respect to the GW-determined time of merger, a weak short GRB triggered both the *Fermi*/GBM and *INTEGRAL*/SPI-ACS gamma-ray detectors. This dual detection triangulated the GRB to a sky location consistent with the GW localization box. The GW and GRB information were circulated to the astronomical community. An extraordinary ground- and space-based instrumental effort ensued to search for an optical counterpart to the GW and GRB triggers. This searches succeeded within hours in finding a thermal transient source, the first ever direct observation of a kilonova: a transient powered by the nucleosynthesis of heavy elements in the neutron-rich ejecta from the merger. Long-term monitoring of the source revealed a strongly atypical multi-wavelength afterglow signal that offered exquisite insight on the outflow launched by the merger. This afterglow is still monitored today, almost four year after merger; Recent data suggests it is currently rebrightening, possibly revealing another distinct phase of energy dissipation in the merger ejecta.

3.2 From the merger to the onset of the afterglow emission

3.2.1 The BNS merger inspiral signal

The inspiral phase GW signal was detected for the $\sim$ 3000 last orbits of the binary (Abbott et al., 2017b). The spectrogram of this signal, GW170817, as reconstructed using the data from the three interferometers, is reproduced in Fig. 3.3 (bottom).

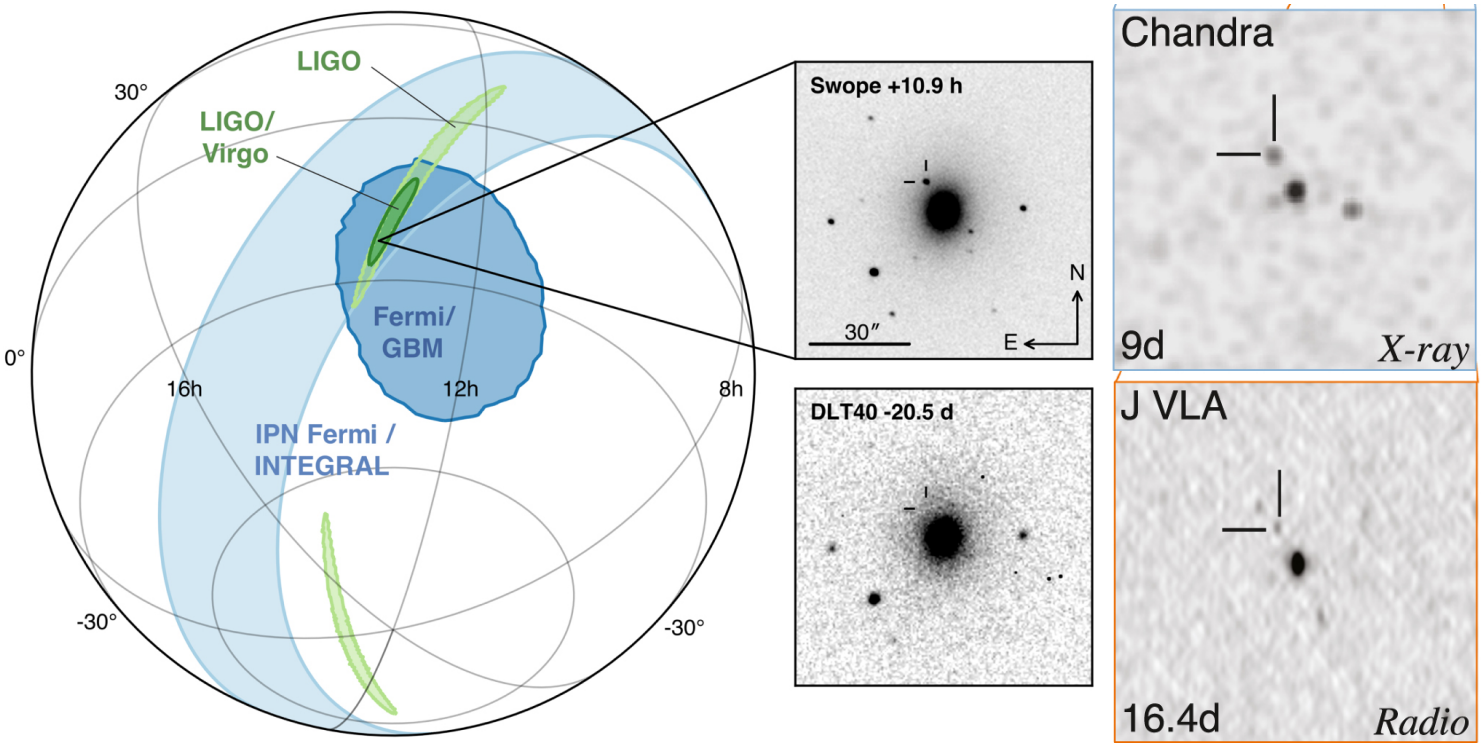


Figure 3.1: **Multi-messenger localization constraint on the source of GW170817.** Left: The GW data allowed to locate the source to the light green area with LIGO data only and dark green region with added data from Virgo. GRB170817A triggered both the *Fermi*/GBM and *INTEGRAL*/SPI-ACS, allowing to locate the source to the blue regions. Dark blue: *Fermi* only; light blue: source triangulation from GRB arrival times in the two instruments. Center: Discovery image of the kilonova transient in the NGC4993 galaxy, along with an image dated 20 days prior showing no source. Right: Discovery images of the afterglow in the X-ray and radio bands (Abbott et al., 2017e).

The signal lasted $\sim$ 100 s, in sharp contrast with the binary black hole mergers detected up to then, which last ≤ 1 s in the sensitive band (Fig. 2.2); The inspiral ended at 12:41:04.4 UTC, the GW-determined time of merger. It is the loudest gravitational wave signal detected yet, with a network signal-to-noise ratio (SNR) of $\rho_{\rm net.} = 32.4$. The GW signal infers masses in the ranges of 0.86–1.36 $M_\odot$ and 1.36–2.26 $M_\odot$ for the two initial components (at 90% confidence). Given the ranges of measured galactic black hole masses, which are substantially higher than these , and the measured masses of some galactic binary neutron stars, which are consistent with these, a binary neutron star is the most likely nature of the GW progenitor (Abbott et al., 2017b, and references therein). This is further supported by the detection of electromagnetic counterparts to this GW signal, indicating the presence of matter in the immediate environment after the merger, and thus the unlikeliness of a binary black hole progenitor. Uncertainty as to the exact nature (BNS or NSBH) partially stems from the uncertainty on the maximum mass for a neutron star (Köppel et al., 2019). In the signal of GW170817, the tidal deformabilities of both the components of the initial binary have been measured in the last orbits to be marginally consistent with a NSBH progenitor, but more largely consistent with those predicted by typical neutron star equations of state. This once again points to a BNS progenitor for GW170817.

The joint GW and GRB localization constraint is illustrated in Fig. 3.1. The two LIGO instruments constrained the source to an area of 190 deg^2 in low-latency, as shown by the pale green region. The Virgo detector did not measure a significant signal, while it should have, had the source in a more sensitive direction. Using Virgo's antenna pattern (Sec. 2.3.2), this allowed to exclude a region of the sky and improved the localization constraint to 31 deg^2 in low-latency revised to 29 deg^2 in high-latency several hours after merger; This is the dark green region. Virgo thus likely greatly decreased the duration of subsequent searches for electromagnetic counterparts. The detection of the GRB indicated the dark blue region (*Fermi*/GBM alone) and the light blue region (joint *Fermi–INTEGRAL* constraint). We see that the GRB information would have already excluded one half of the LIGO skymap.

The GW-inferred luminosity distance to the source is 40^{+8}_{-14} Mpc, making the GW and short GRB the closest such events ever detected. This distance is further refined by combining GW data with electromagnetic observations of the host galaxy to 42.9 ± 3.2 Mpc (Cantiello et al., 2018). Similarly, from a GW-only constraint on the viewing angle of $\iota \leq 55$ deg, knowing the source's sky location allows to further constrain it to ≤ 28 deg.

An essential feature of GW170817 is the non-detection of a ring-down signal, i.e the gravitational radiation emitted by the remaining central black hole during its relaxation to a stationary Kerr state. A detection of the ring-down would have allowed to constrain the nature and behavior of the transient compact object—if any—present between the merger and the final formation and relaxation of the black hole. The lifetime of this transient object can, however, be indirectly probed through the kilonova signal (Gill et al., 2019). In any case, seeing the likely mass of the final black hole (smaller than the total BNS mass of $\sim 3M_\odot$), its ring-down signal would have peaked at frequencies well above the detectors' sensitive bands.

The signal's chirp mass is extremely well constrained to $1.188 \pm 0.004\ M_\odot$, and the peak GW luminosity was $\sim 9\ M_\odot c^2$/s (Zappa et al., 2018).

3.2.2 The kilonova thermal transient

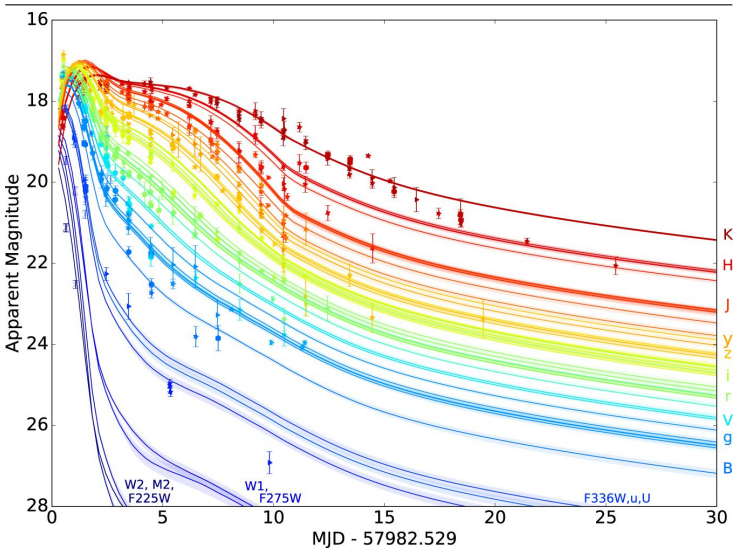


Figure 3.2: **Consolidated multi-wavelength light curve of the kilonova transient AT2017 gfo.** Photometry points cover the near-UV to infrared bands, showing the shorter evolution timescales in the bluer bands than in the redder ones. In solid lines are models with three emission components of different opacities, reproducing this feature (Villar et al., 2017).

Using the galaxy-targeting strategy, the One-Meter-Two-Hemisphere collaboration using the Swope telescope was the first to discover and announce a bright optical transient in the localization box, 11.9 hours after merger (Abbott et al., 2017e; Coulter et al., 2017). In Fig. 3.1, we show the discovery image along with an image taken 20 days prior to GW170817. The new object appears clearly; It was later called AT 2017gfo for the 4950$^{\rm th}$ astronomical transient of the year 2017. It was located in the lenticular galaxy NGC 4993 within the ESO 508 group of galaxies in Hydra, at a significant offset from the galactic center. Upon discovery, the r-band magnitude was measured to ~ 17, equivalent to an absolute magnitude of $\check{}16$: about 2–3 orders of magnitude fainter than a Type Ia supernova and brighter than a typical nova, as expected from a *kilo*nova transient (Metzger, 2019).

This kilonova dimmed out of detection limits within 30 days. In the mean time, an extensive photometric and spectral effort was produced to sample the transient with hundreds of optical points and tens of spectra. In Fig. 3.2, we report the consolidated multi-wavelength light curve of AT2017 gfo (Villar et al., 2017). The spectra of AT2017 gfo was thermal, cooling from temperatures of $\sim$ 8000 K at 0.7 days to 2500 K at 9 days

(Drout et al., 2017). These spectra feature broad absorption lines. Notwithstanding the very large uncertainties on the numerous atomic lines and transition probabilities in r-process elements (Barnes et al., 2020), some authors claimed to identify some such elements in these spectra, such as Cs and Te (Smartt et al., 2017) or Sr (Watson et al., 2019).

Kilonovae are thermal radiation from an initially optically thick mass of ejecta from the merger event. It is understood that the neutron-rich matter ejected by the merger event is the site of r-process nucleosynthesis, that is the synthesis of neutron-rich heavy nuclei by rapid neutron capture in a neutron-dense environment, and the subsequent decay of these nuclei to heavy elements such as the lanthanides and actinides (Cowan et al., 2021). The decay of the nuclei are a heat source within the ejecta, and the many open atomic f-shells of r-process elements results in an extremely opaque medium, ensuring the thermalization of the injected energy. Moreover, sudden decompression of this once-neutron-star material into the rarefied external medium drives the rapid expansion of the ejecta.

The transient's light curve (Fig. 3.2) reveals two different evolution timescales in the infrared ($\gtrsim 10$ days) and blue ($\lesssim 5$ days). This suggests the existence of two separate sites in the ejecta for nucleosynthesis (Villar et al., 2017). The faster-evolving radiation dominating the blue bands is associated with less opaque ejecta with a smaller concentration of lanthanides, resulting in a smaller diffusion time. Conversely, the slower-evolving component is associated with a higher opacity site due to greater concentration in lanthanides and thus larger opacity. This picture is supported by the two different types of ejecta expected from a BNS merger: tidal and dynamical (i.e., powered by sheer pressure forces upon object contact) ejecta, which is rich in neutrons and where lanthanides are efficiently synthesized; and a neutrino-driven wind lifted above the hot accretion disk around the final central object or from the transient compact object itself, where the neutrinos displace the β equilibrium towards a lower neutron density and a less efficient r-process (Fernández & Metzger, 2016; Metzger, 2019; Barnes, 2020). Moreover, the bluer component is faster and of rather polar direction, whereas the redder component is more isotropic and slower. Overall, the photometry of AT2017 gfo allowed to approximately constrain the blue component to carrying a mass of $2 \times 10^{-2} M_\odot$ with opacity of 0.5 cm^2/g and expanding at a speed of $0.26c$, and the red component of mass $2 \times 10^{-2} M_\odot$, opacity 3.7 cm^2/g and expansion speed $0.15c$ (Villar et al., 2017). The earlier-peaking blue component and the later-peaking red components are evident in the model fits shown in Fig. 3.2.

3.2.3 The peculiar short GRB170817A

The GRB transient event started 1.72 s after the GW-inferred merger time (Abbott et al., 2017f); Its duration was 2.0 ± 0.5 s, placing it in the class of short GRBs with a confidence level of $\sim 72\%$ (Goldstein et al., 2017; Savchenko et al., 2017). In Fig. 3.3, one can find the gamma-ray light curves in two energy bands of *Fermi*/GBM and in the *INTEGRAL*/SPI-ACS; The delay with respect to the GW-inferred time of merger is emphasized. GRB170817A's total isotropic-equivalent radiated energy was $E_{\gamma,\mathrm{iso}} \sim 6 \times 10^{47}$ erg, around 4 orders of magnitude weaker that typical short gamma ray bursts, though photons with energies as high as 185 keV were emitted. Its peak luminosity was $L_{\gamma,\mathrm{iso}} \sim 2 \times 10^{47}$ erg/s.

Moreover, it appears that the light curve of GRB170817A presents two successive components, a hard spike lasting 0.58 s, and a smoother softer component lasting 1.1 s. Spectrally, the first component is best-fit by a Comptonized spectrum with peak energy 185 ± 62 keV, and the second on is best fit by a blackbody with temperature $kT \sim 10$ keV; The spectral shape of this second component is not well constrained.

These facts immediately allow to compare GRB170817A with the other short GRBs. In Fig. 3.4, we place GRB170817A (red crosses correspond to different spectral fits) along with the S-BAT 4 flux-limited catalog of short GRBs (black crosses, (D'Avanzo et al., 2014)) in the $L_{\gamma,\mathrm{iso}}$–$E_{p,\mathrm{peak}}$ plane (Sec. 1.4). We also show the best-fit correlation found for short GRBs in this plane (Ghirlanda et al., 2009; D'Avanzo et al., 2014).

It clearly appear that GRB170817A is an outlier of this correlation: It is very hard given its low luminosity. A first hypothesis to explain this peculiarity is that the relativistic jet associated with this GRB was not aligned with our line of sight. Indeed, GRBs are associated with such jets (Sec. 1.2.4), and the closeness of this event could have allowed us to detect this GRB even without the full boosting of a relativistic jet. Given observed $E_{p,\mathrm{peak}}$ and $L_{\gamma,\mathrm{iso}}$, it is straightforward to determine the values of these quantities which would have been measured if the event had been viewed under a different angle; They can be transformed to other lines of sight using the Doppler factor:

$$\mathcal{D} = \frac{1}{\Gamma(1 - \beta \cos\theta)} \tag{3.1}$$

where Γ is the jet's Lorentz factor and θ the angle between our line of sight and the emitter's velocity. In Fig. 3.4, we predict the peak energies and luminosities that would have been observed for an aligned observer assuming we had different viewing angles for GRB170817A and with $\Gamma = 100$. Even when correcting for the possibly non-zero viewing angle, GRB170817A remains an outlier of this correlation. The singularity of this GRB also manifests when examining the $E_{\gamma,\mathrm{iso}}$ (Matsumoto et al., 2019a).

These facts call for an alternative explanation for the dimness of GRB170817A and possibly a different emission mechanism than for *bright GRBs*, i.e., the GRBs observed at cosmological distances down or very near their relativistic jets as we described in Chap. 1. In the perspective of finding an alternative mechanism for GRB170817A, its other remarkable characteristics must be taken into account, for example, that its light curve was double peaked and that its spectrum evolved from hard to soft. An archival search for other short GRBs similar to GRB170817A in these respects showed that such GRBs are rare (von Kienlin et al., 2019).

These features prompted Gottlieb et al. (2018) to suggest *ultra-relativistic shock breakout* as a mechanism for GRB170817A. Shock breakout is the phenomenon by which a hydrodynamic shock reaches the edge of its propagation

medium, or more generally a sharp discontinuity in propagation medium density. For example, when a shock wave from the collapsing of a massive star reaches the edge of the star's envelope. In an astrophysical context, shock breakout generally entails matter ejection from the breakout surface and, as seen from a distant observer, a burst of high energy radiation from the abrupt leaking of all the energy contained in the shock. This burst is followed by a longer lasting *cooling emission*, when radiation from the shocked material behind the shock front is finally revealed. Shock breakout therefore naturally predicts a hard peak and a softer tail (Nakar & Sari, 2012).

In ultra-relativistic shocks, the temperature of the freshly shocked material is high enough ($\geq$ 200 keV) that pairs are spontaneously created and a permanent, extended pair plasma is present at the head of the shock. In this case, only the high-energy radiation from the shock itself is released, and most of the shock's kinetic energy is dissipated into thermal energy downstream. Therefore, it is a highly inefficient mechanism, recalling the dimness of GRB170817A. Shock breakout was also suggested to explain other low-luminosity GRBs (Nakar & Sari, 2012).

In the case of GRB170817A, the merger of two neutron stars occurred and the GRB scenario described in Sec. 1.2.6 suggests that a relativistic jet was launched from the central compact object. At this early moment, the remnant must have been embedded in an envelope made up of the merger ejecta, that which would eventually produce the kilonova transient. As the relativistic jet was launched into this slow and dense ejecta, a strong shock formed at its head and Gottlieb et al. (2018) suggested that the gamma-rays of GRB170817A were produced in the breakout of this relativistic shock from the merger ejecta. In this picture, the gamma-rays originate from material on the observer's line of sight, and not in the central jet.

3.2.4 The afterglow signal

Whether the gamma-rays in GRB170817A were produced by a shock breakout or not, the interaction of the initial relativistic jet with the merger ejecta posed the question of the ultimate structure of the outflow from the merger. This outflow would eventually penetrate the circum-merger medium and decelerate to produce the afterglow emission through a forward-shock dissipation mechanism analogous to that of bright GRBs (Sec. 1.2.3).

The afterglow counterpart was detected in the X-ray and radio bands were detected to 9 days and 15 days respectively post-merger (Troja et al., 2017b; Hallinan et al., 2017). In Fig. 3.1 (right), we reproduce the detection images from the *Chandra X-ray Observatory* and the Very Large Array. When the kilonova signal had sufficiently decayed at 150 days post-merger, an optical band afterglow was also detected as emerging from the dimming kilonova signal (Lyman et al., 2018). The afterglow photometry points in these bands until 180 days post-merger are reported in Fig. 3.6.

A remarkable feature of these afterglow light curves are their homothetic structure, with $F_\nu \propto \nu^k$ and a time-independent spectral slope $k \sim -0.6$. Under the interpretation that the afterglow is synchrotron radiation from electrons accelerated at the forward shock of the ejecta, the particle population's spectral slope can be inferred to $p = 1 + 2k \sim 2.2$ from this sole observation (Mooley et al., 2018c,a; Troja et al., 2019a, Appendix E,); In our multi-messenger population model of Chap. 4, we will come back to the modeling of BNS merger afterglows.

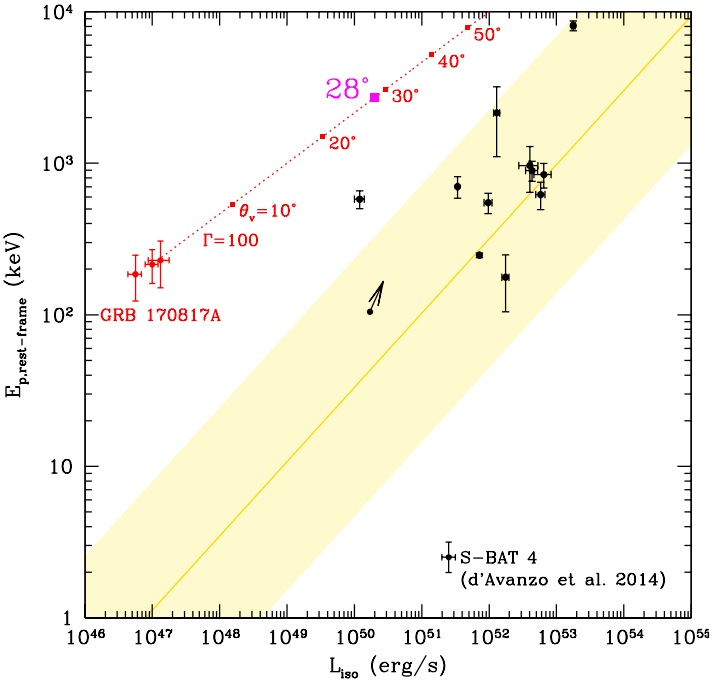


Figure 3.4: **Can misalignment of the relativistic jet explain the dimness of GRB170817A?** Yellow shaded region: correlation between the source-frame peak energy $E_{p,\rm peak}$ of short GRBs spectra and their isotropic-equivalent peak luminosity $L_{\gamma,\rm iso}$. Black points: A sample of short GRBs selected to represent the typical population of cosmic short GRBs. Red points with error bars: Observed values for GRB170817A. Assuming this GRB was actually viewed from different viewing angles, one may infer the value of peak energy and luminosity had it been seen on axis, producing the red dotted line and red squares. Whatever the supposed viewing angle, GRB170817A is an outlier. We also highlight the case of $\theta_v = 28$ deg, the maximum viewing angle consistent with the GW data.

Furthermore, these first afterglow photometry points set this event aside from other GRBs: A steady increase for a hundred days was a unique feature among GRB afterglows. It is this peculiar afterglow emission that would eventually provide detailed insight on the outflow from GW170817.

3.3 Was there a jet launched in GW170817?

The jet launched from the central engine must have interacted with the dense merger ejecta. This interaction will shape the

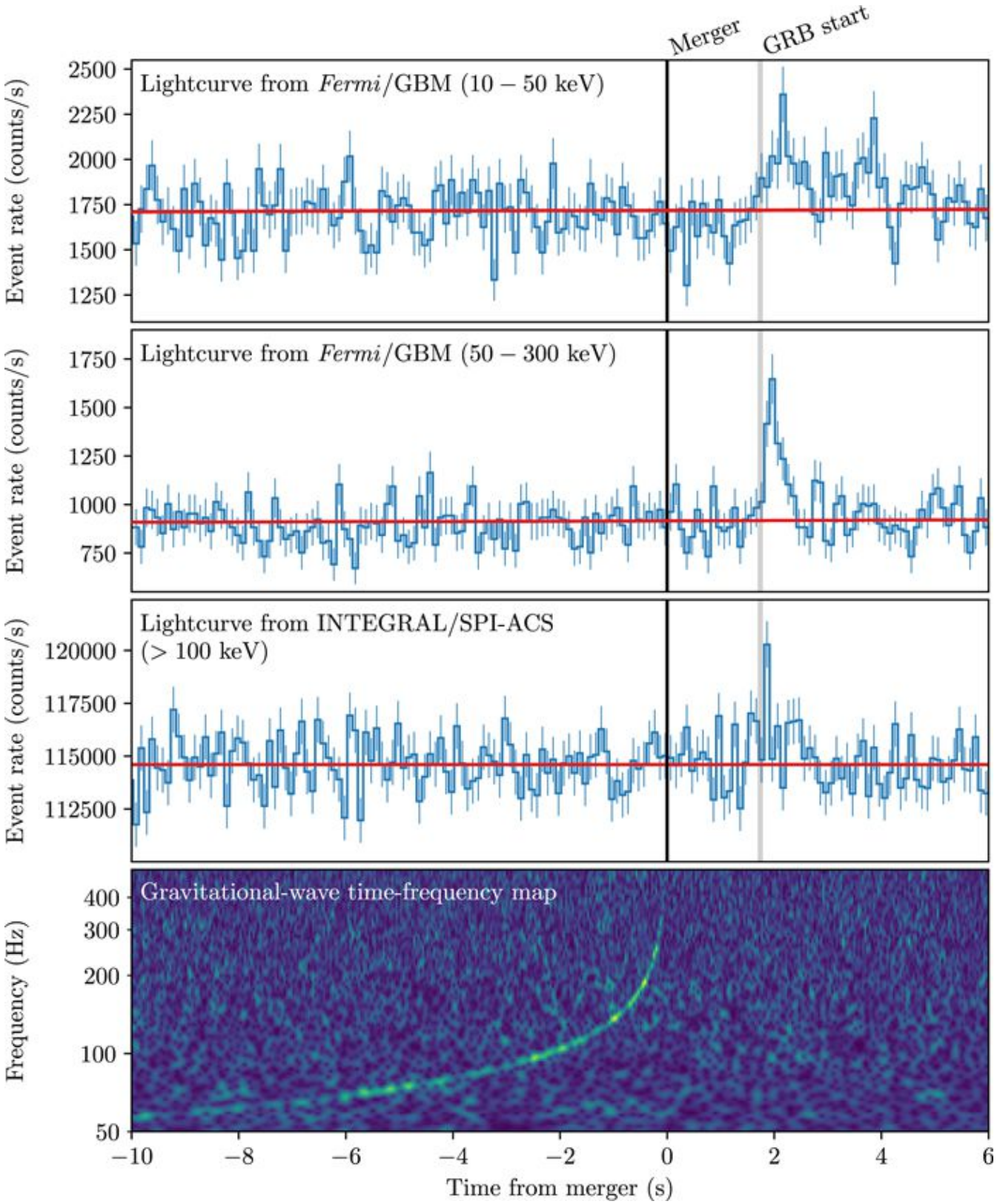


Figure 3.3: **GW and gamma-ray signals from the BNS merger GW170817.** Top: Gamma-ray light curve from *Fermi*/GBM in two different energy bands and *INTEGRAL*/SPI-ACS. Bottom: Time-frequency map obtained by combining the data from the three gravitational interferometers in the network at the time of the event: two LIGOs and Virgo. The GW-inferred time of merger is figured, along with the beginning of the GRB T_{90} interval; The delay between these two is ~ 1.72 s, see text for more details (Abbott et al., 2017f).

jet and the general structure of the outflow after the jet breaks out. Depending on the kinetic energy of the jet, its opening angle, the velocity profile of the ejecta and the details of the jet-ejecta interaction, the jet may or may not remain collimated as a relativistic jet until emerging. On the one hand, if the jet is wide and relatively unenergetic, or if its energy is efficiently lost to the merger ejecta on its flank during propagation, then the jet will energize the ejecta into a wide-angle, nearly spherical outflow with a radial gradient of velocities from the slower inner shells to the faster outer. On the other hand, if the jet is energetic enough and deposits little energy in the cocoon, it may retain its collimation throughout its propagation, and emerge as a relativistic jet. In this case, the final outflow would rather resemble the original structure of the jet, with a fast and energetic core and slower material in the wings. For GRB170817A, these two outcomes were dubbed a *choked jet* or a *successful jet*, respectively (Mooley et al., 2018a).

More generally, the expected outflow after jet-ejecta interaction is a *structured jet*, where the density of kinetic energy and the Lorentz factor in the outflow depend on the angular coordinate from the axis of the jet (angular profile, e.g., Margutti et al. 2017; Resmi et al. 2018) and on the radial coordinate from the central engine (radial structure, e.g., Kasliwal et al. 2017; Gottlieb et al. 2018). The two extreme outflows with either a strictly angular or a strictly radial structure are represented in Fig. 3.5. Once the outflow is launched, its structure will define its interaction with the circum-merger environment and therefore shape the longer-lasting afterglow emission. In actual events, the outflow naturally possesses a combination of both radial and angular structure; The afterglow will, however, be dominated by one or the other depending on whether the core is really more energetic than the wings, and if the radial stratification of the outflow remains at the time of observation. Indeed, as the outflow decelerates, the shells with different velocities will catch up to each other, progressively collapsing to a mono-kinetic outflow.

The central question after having observed GRB170817A was whether a successful jet had emerged (e.g. Mooley et al., 2018a; Margutti et al., 2018). This question was important to the link between short GRBs and BNS mergers, to whether this BNS merger was similar to other mergers producing short GRBs, what the energy budget of this event had been, and what the physical conditions in the ejeta were.

3.3.1 Early expectations from GRB theory

First, according to hydrodynamical simulations of jet-ejecta interactions and analytical estimates, one expected most GRB jets to successfully breakout from BNS merger events (Duffell et al., 2018). Furthermore, it was found that jets sufficiently energetic to reach the edge of the cocoon and produce shock breakout radiation would systematically retain its collimation after the breakout. In other words, either a jet is chocked altogether by the cocoon and produces no breakout radiation, or it is able to reach the surface and produce the gamma-rays, in which case it remains collimated and relativistic after breakout. This was found to be due to the negligible amount of thermal energy transfer between the jet and the cocoon during its propagation. This conclusion is, however, contrary to the picture given in the above-cited Gottlieb et al. (2018), who suggested both that GRB170817A was a manifestation of relativistic shock breakout and that the resulting outflow structure was a mildly-relativistic wide-angle outflow.

Second, a statistical study led by Beniamini et al. (2019) also found that most BNS mergers produced successful jets. This was found by comparing the rate of BNS mergers as found by gravitational astronomy with the rate of short GRBs known from GRB science. As mentioned in Sec. 1.2.4, the latter is uncertain because of the beaming factor of GRB jets. However, within this uncertainty, data show that most mergers result in successful jets.

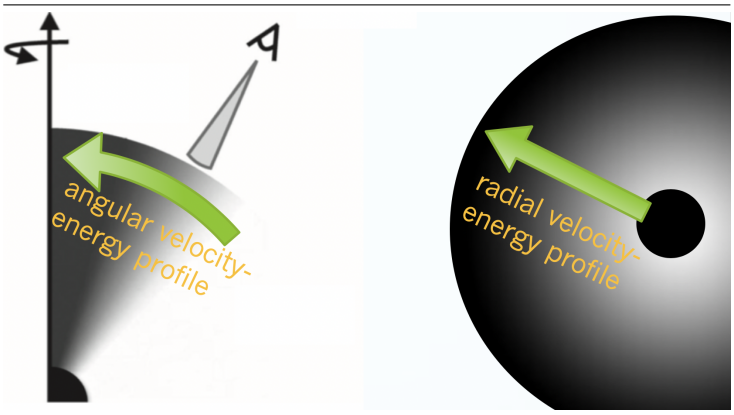


Figure 3.5: **Two outflow structures able to explain the early afterglow observations of GRB170817A.** Left: An angular structure, where the outflow velocity and energy decreases with angle from an ultra-relativistic core to slower lateral wings. Right: A radial structure, where the outflow is mildly relativistic everywhere and the outer regions are faster than the inner regions.

3.3.2 Outflow structure degeneracy in the early afterglow

The answer to the successful jet riddle in GRB170817A would come from the afterglow emission, which is the signature of the interaction of the outflow with the external medium. The first 180 days of afterglow photometry data are given in Fig. 3.6, along with synthetic light curves from best-fitting radial-structured and angular-structured outflows reported by D'Avanzo et al. (2018). The right panels show the selected structures.

In the case of an angular structure, an off-axis observer does not initially receive radiation from the core, because it is beamed away from the line of sight by relativistic effects. It is only after the core has decelerated, after interaction with the external medium, that the observer progressively sees it. The core being more energetic than the wings, this produces a slowly increasing afterglow light curve. After it is uncovered, the observer simply sees the decaying radiation from the core, resulting in the decreasing phase of the light curve. Grossly,

in this interpretation, one can deduce the slope of the angular structure and the viewing angle from the afterglow's temporal slope and time of peak (Beniamini et al., 2020b).

In the case of a mildly-relativistic radial structure, an observer sees radiation from essentially the whole outflow from the beginning. However, as the outer shells decelerate, the inner shells catch up to them, progressively injecting more energy into the forward shock, resulting in a slowly increasing light curve as well. Once the slowest, inner-most shell has caught up, energy is no longer injected into the shock, and a decreasing phase ensues. Grossly, one may deduce the slope of the radial structure and the velocity of the slowest shell from the light curve's temporal slope and time of peak.

As shown in Fig. 3.6, it is quite remarkable that both extreme models were able to fit the data at this stage; Both possible structures were almost perfectly degenerate. The light curve of the afterglow shows a distinctive shallow increasing phase from the first points at ~ 10 days up to what appears to be a peak around 180 days.

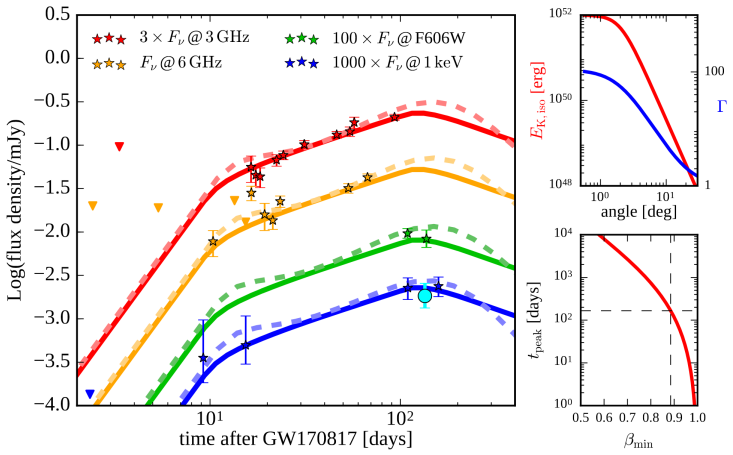


Figure 3.6: **Early afterglow light curve fitting in GRB170817A.** Left: Multi-wavelength light curve of the GRB170817A afterglow, up to around 180 days after merger. At this stage, both an angular and a radial jet structure could fit the data, as shown by the dashed and solid lines model light curves which equally fit the photometry points. Top right: Angular structure corresponding to the dashed line light curve in the left panel. Bottom right: Adjustment of the minimal velocity in the radial structure to reproduce the correct peak time of the afterglow data (D'Avanzo et al., 2018).

3.3.3 A breakthrough observation: very long baseline interferometry imagery of the remnant

In order to break the degeneracy and resolve the structure of the jet, another measurement was necessary. The light curve is not the only observable of the afterglow radiation: The remnant can also be imaged using very long baseline interferometry (VLBI). Under the assumption of a successful relativistic jet, the image is expected to be small—because only the jet's head produces radiation—, and mobile, possibly at apparent superluminal speeds. Under the assumption of a mildly-relativistic cocoon structure, the image is expected to be large—extending in all directions around the merger locus—, and static, because of the symmetry of the spherical expansion.

The combination of VLBI imagery performed 70, 205 and 230 days after merger allowed to confirm the relativistic and jet-like nature of the outflow from the merger GRB170817A (Mooley et al., 2018b; Ghirlanda et al., 2019). The VLBI image along with a synthetic image expected from a successful jet scenario are presented in Fig. 3.7. The apparent size of the emitting material was constrained to being smaller than 3 mas—0.5 pc at the distance of GW170817—, and its apparent speed was $\sim 4c$.

This measurement was a great technical challenge utilizing a global network of radio facilities, the longest baseline running over 10 000 km from South Africa to the US. Its output is invaluable in the study of GRB170817A and short GRBs in the gravitational wave era in general.

Combining all the multi-wavelength photometry and imagery data for the remnant allowed to probe the structure of the jet in great detail. Fig. 3.8 shows the structures in both kinetic energy density and Lorentz factor inferred by Ghirlanda et al. (2019); These are the kinetic energy density and Lorentz factor of the initial pre-deceleration outflow as a function of the material's latitude with respect to the jet axis. They are flat in the core of the jet, up to $\theta_j = 3.5 \pm 1$ deg and decay as power laws with slopes -5.5 ± 1.5 and -3.5 ± 1.5 for energy and Lorentz factor, respectively. Our viewing angle to the jet's axis was inferred to $\theta_v = 15 \pm 1.5$ deg. The core material's Lorentz factor is $\Gamma_0 = 250^{+540}_{-150}$ and its isotropic-equivalent kinetic energy is $E_{\rm k,iso,0} = 30^{+100}_{-24} \times 10^{52}$ erg (all 1-σ confidence intervals); There was indeed an ultra-relativistic jet launched from GW170817.

The joint afterglow photometry and imagery fit also allows to constrain the density of the circum-merger medium. It is found that the local number density is $n_{\rm ext} \leq 10^{-3}$ cm^{-3} at 1-σ confidence level. This value is supported by the less robust measurements made with the radio and X-ray continuum of the host galaxy, which also find $n_{\rm ext} \leq 10^{-2}$ cm^{-3} (Hallinan et al., 2017; Hajela et al., 2019). Such a rarefied environment for the merger and the large offset of GW170817 from its host galaxy enforces the picture of a long migration of BNS due to their long inspiral times before merging presented in Sec. 1.3.

Monitoring of the afterglow continued after the VLBI campaign and photometry confirmed the tendency of a jet-like post-peak decaying phase (Lamb et al., 2018; Mooley et al., 2018c; Hajela et al., 2020). Nonetheless, the VLBI data allowed exquisite insight into the jet structure and showed that we had had a *significantly misaligned line of sight* to this event, with $\theta_v/\theta_j \gtrsim 4$; In both respects, this is a historic first.

In Fig. 4.2, we show a structured jet fit to the radio data up to 300 days post-merger (more details in Sec. 4.4.5). The jet's structure is a power-law in energy and Lorentz factor with slopes 4.5 and 2.5 respectively, and the viewing angle is 22 deg. The red dashed and dotted lines show the contributions of the core jet (up to 4 deg) and lateral structure respectively. By

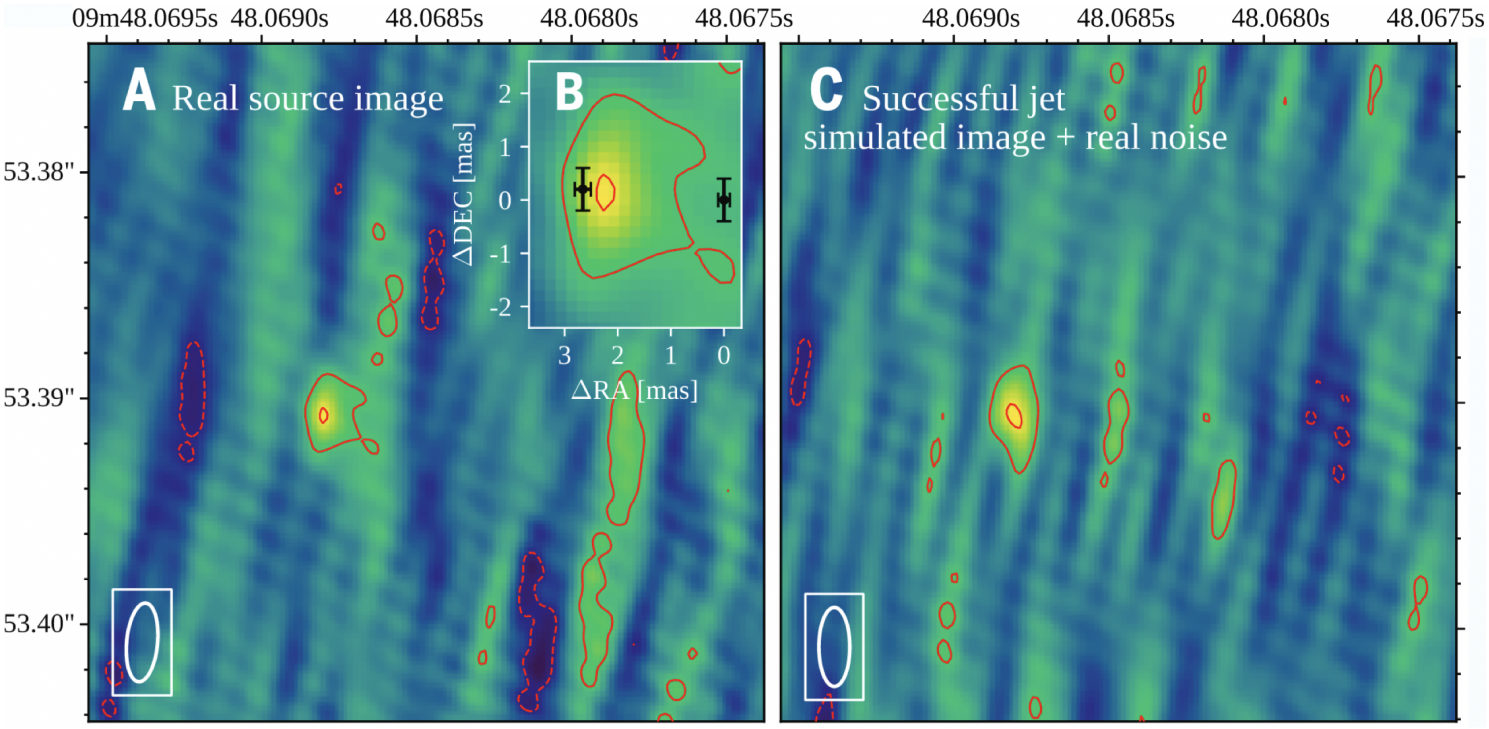


Figure 3.7: **VLBI imaging of the remnant of GW170817.** Left: 30-mas-size image of GW170817's remnant taken with the European VLBI Network in the 4.84 GHz 205 days after merger. The black crosses show the remnant's position as reported by Mooley et al. (2018b) 70 and 230 days after merger, using the High Sensitivity Array. The three successive positions show the apparent super-luminal motion of the remnant, proving its ultra-relativistic jet nature. Right: Synthetic image of the remnant assuming a relativistic Gaussian-structured jet and including synthesized instrumental noise (Ghirlanda et al., 2019).

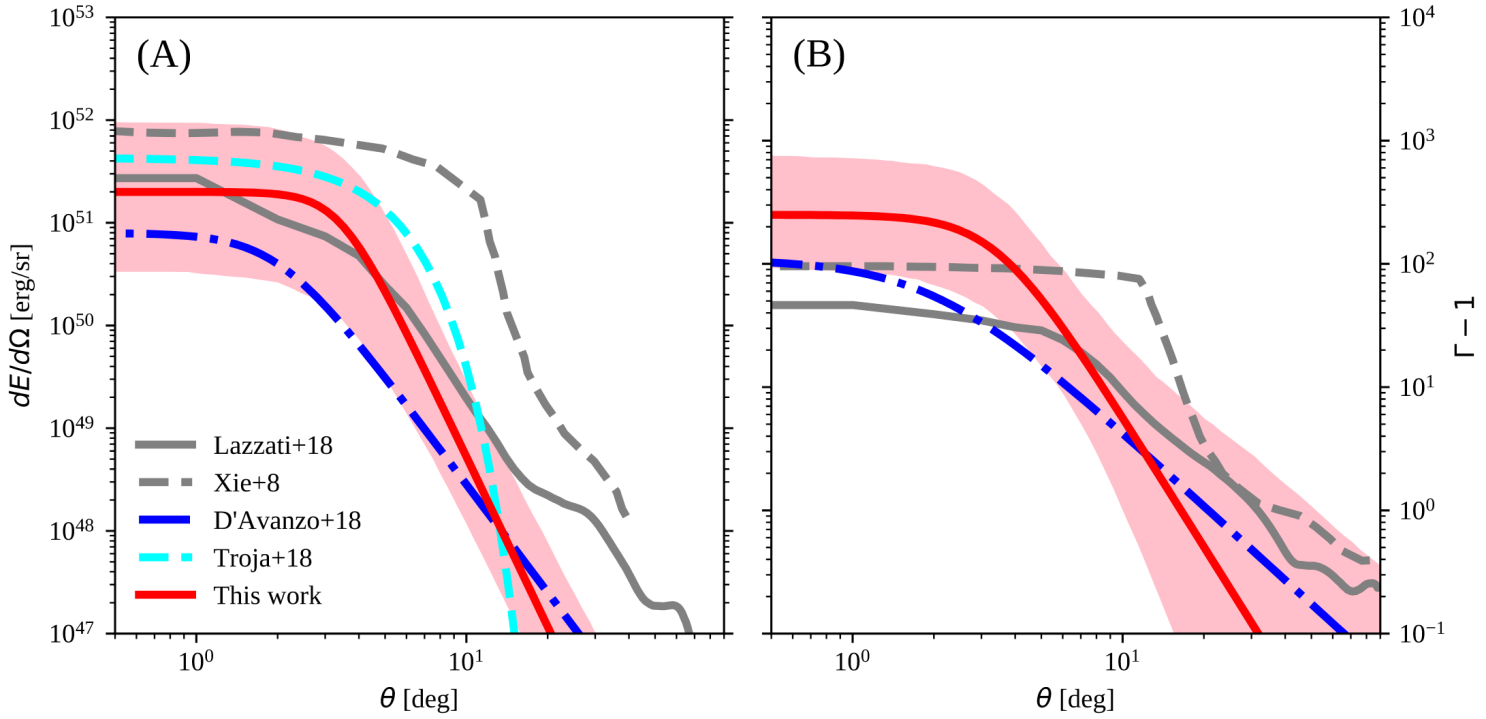


Figure 3.8: **Angular jet structure in GRB170817A.** Inferred by Ghirlanda et al. (2019) from the combined GW, afterglow light curve and remnant imagery, with 1-σ uncertainty region (red curve), compared to those from numerical simulations (gray) and some adopted in the literature to model GW170817's afterglow (cyan and blue). Left: Kinetic energy per unit solid angle. Right: Lorentz factor. Such an exquisite detail of the structure of a short GRB is truly historic.

distinguishing the contributions from the core and the wings, it is clear that the core is subdominant at early times, and only appears and dominates the radiation from the peak onward, confirming the intuition presented earlier.

We finally note that, once the relativistic nature of the outflow and our significantly misaligned line of sight was established, a thorough calculation of the compactness limits on the emitting region of GRB170817A showed that, indeed, the gamma-rays could not have come from the core of the jet (Matsumoto et al., 2019a,b). This confirmed our prior intuition that the dimness of the GRB could not be explained by its misalignment (Fig. 3.4 and discussion thereabout) and that the gamma-rays had in fact come from material near the line of sight, advocating a different emission mechanism than for bright GRBs. GRB170817A is therefore not a bright GRB, the first for which this was proved to be the case.

3.4 Latest news from GW170817

At the time of writing, the source is still detected and monitored in the radio and X-ray bands. In Fig. 3.9, we reproduce the complete afterglow light curve complete with some of the latest photometry points in these bands (1234 days post-merger, Balasubramanian et al. 2021). The figure also shows a fit of a smoothly broken power-law model for the entire photometric dataset, and the residuals in the bottom panel. It is apparent that the achromaticity of the light curve remained until ~ 1000 days post-merger; However, the X-ray band shows a significant excess in the last data point, which is not present in the radio band. An even later observation confirms the achromatic excess in the X-ray at 1258 days (Hajela et al., 2021). Whether the X-ray data show a new increasing phase or simply a flattening is still debated (Troja et al., 2021); Regardless, the latest X-ray data are in growing tension with the power-law decrease expected from the jet afterglow. This could be the first chromatic behavior of the afterglow, with an observed hardening of the broadband spectrum.

Such an excess can have many origins, either in a modification of the propagation or radiation properties of the same forward shock responsible for the whole afterglow, or in the emergence of a new radiative component from a different emission site. The "same-shock" explanations include an episode of energy injection into the outflow's forward shock—e.g., from a pulsar wind—, or ad-hoc variations of the shock microphysical conditions (Hajela et al. 2021 and references therein). The "new component" scenario includes, e.g., a novel episode of accretion of some kilonova material falling back onto the central compact object (Ishizaki et al., 2021).

However, the most favored origin is the *kilonova afterglow emission*. This is non-thermal radiation expected from the shock produced by the decelerating dense and only mildly relativistic ejecta responsible for the kilonova transient (Hotokezaka et al., 2018b; Nakar et al., 2018; Kathirgamaraju et al.,

2019). Because this outflow is only mildly relativistic ($\beta \lesssim 0.3c$, Sec. 3.2.2) and carries more mass, its afterglow emerges and peak on much longer timescales than the relativistic jet. This ejecta is likely radially stratified, and the kilonova afterglow peaks once the slowest shell has caught up to the decelerating forward shock, as mentioned above. Consequently, the peak time of the kilonova afterglow is very sensitive to the minimal velocity $\beta_{\rm min}$ present in the outflow. Similarly, the phase of increasing kilonova afterglow flux up to the peak is determined by the stratification of the outflow, captured in the index $\zeta_{\rm KN}$ such that the kinetic energy is cumulatively distributed as $E(\leq \beta\Gamma) \propto (\beta\Gamma)^{-\zeta_{\rm KN}}$ in the outflow for $\beta \geq \beta_{\rm min}$. If the observed excess in GW170817 corresponds to the emergence of the kilonova afterglow, then steep structures with $\zeta_{\rm KN} \geq 5$ are favored (Hajela et al., 2021).

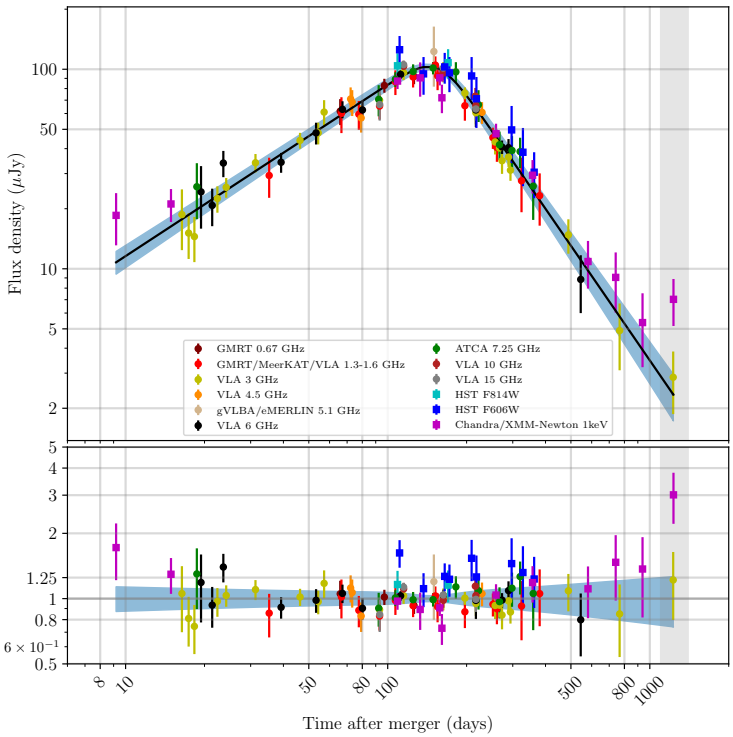


Figure 3.9: **Long-term multi-wavelength photometry of the afterglow of GRB170817A.** Top: Afterglow light curve from the radio to the X-ray band up to 1234 days post-merger. The blue shaded area is a fit of a smoothly broken power-law profile. Bottom: The fit residuals. At late times ($\gtrsim$ 1000 days), achromaticity breaks down and an excess in the X-ray appears, possibly signing a new emission component and thus a novel counterpart to this already-rich event (Balasubramanian et al., 2021).

The kilonova outflow ultimately depends on the ejecta from the inspiral and merger phases of the BNS coalescence. The sensitivity of the kilonova afterglow to the outflow structure thus makes it an interesting tool to study the dynamics of these phases, the binary parameters such as mass ratio or the neutron star equation of state (Nedora et al., 2021). It remains nonetheless that the kilonova forward shock is only mildly relativistic, and the microphysical conditions in such shocks have yet to be studied by plasma physics and probed by observations; They are indeed intermediate between supernova remnant shocks and GRB afterglow shocks (for more details, see Sec. 7.3.1).

3.5 Motivation of the doctoral thesis: questions

In summary, the BNS merger GW inspiral signal GW170817 was very rich with electromagnetic counterparts: a weak short GRB, a thermal kilonova transient, a long-lived multi-wavelength afterglow possibly showing evidence for a new kilonova afterglow contribution. A tremendous instrumental effort from the search for the kilonova to the VLBI imagery of the remnant produced a treasure trove of data on this event and afforded many major breakthroughs in physics and astrophysics.

Most importantly, this event kicked off the era of multi-messenger astronomy with gravitational waves, which continues to unfold with improving GW detectors, planned GW observing runs and ever more efficient and better organized electromagnetic follow-up. In this nascent era, more binary compact object mergers should be expected, possibly with more electromagnetic counterparts and multi-messenger datasets.

The multi-messenger dataset of GW170817 itself was already very fruitful: from an independent measurement of the Hubble constant to a unique probe of the environment of the merger. In the multi-messenger era, more such datasets are expected, with a combination of GW, GRB and afterglow insights on the unraveling of the BNS merger. These datasets should be processed with further *multi-messenger methods* to better understand this phenomenon.

From the point of view of GRB science, this event confirmed that BNS mergers can launch relativistic jets where bright GRBs are produced. Even though this particular GRB was not bright and did not originate in the core of the jet, this event proved that GRB jets possess structure and provided hitherto unseen insight on this structure.

The multi-messenger event GW170817 thus motivated the following questions that the doctoral work presented in this thesis has tried to address:

1. What electromagnetic signals should we expect to detect as counterparts to binary compact object merger gravitational-wave observations in the future?

2. What original multi-messenger methods can be developed to leverage the upcoming multi-messenger events?

3. What is the influence of the structure of relativistic jets on the prompt and afterglow phases of bright gamma-ray bursts?

In Part II, we develop a multi-messenger population model (Chap. 4) and start to answer question 1 (Chap. 5). We address question 2 by developing a multi-messenger method to better probe the environments of BNS mergers (Chap. 6) and studying the prospects for using BNS mergers and their afterglows to

contribute to multi-messenger measurements of the Hubble constant (Chap. 7). In Part III, we reintroduce structured jets in a broader astrophysical context (Chap. 8) and address question 3 by developing an interpretation based on structured jets of the plateau (Chap. 9) and flaring (Chap. 10) behavior of GRB afterglows. In Part IV, we summarize our results and draw perspectives for future GRB studies in the multi-messenger era.

Part II

The multi-messenger era: New avenues for high-energy astrophysics

Chapter 4

A multi-messenger population model for the gravitational-wave era

Abstract

The historic binary neutron star merger event GW170817 was rich with electromagnetic counterparts: a short gamma-ray burst, a kilonova, a jetted afterglow and a mildly relativistic afterglow. It was understood early on that such an association of signals was due to the particular closeness and well-tuned inclination of the system. As the gravitational interferometer network continues to observe binary neutron star merger signals, which counterparts should one expect to detect? What will be the detection rates and main features of the upcoming multi-messenger event population? In this chapter, we describe the population model developed to address these questions. The results of this population model are given in the following chapter. We focus on the kilonova, and afterglow photometry and imagery sectors. We first describe the new multi-messenger observing scenarios we expect to see in the multi-messenger era, where detecting afterglow counterparts is subject to pinpointing the source in the sky with the kilonova. We then detail the physical ingredients of our population prospects: physical models for the emission and detection of gravitational and electromagnetic signals from mergers, and prescriptions for the distributions of the source parameters, which we choose to be motivated by gamma-ray burst science.

4.1 Why make multi-messenger population prospects?

With GW170817 and counterparts the era of multi-messenger astronomy with gravitational waves opened. The sheer scientific output of this events promised bright prospects for future such events. GW observing runs planned for the after-GW170817 as well as organized follow-up structures announced a forthcoming population of multi-messenger events. In this context, we sought to contribute to this new astronomy by making population prospects for multi-messenger events.

This endeavor responds to the following needs:

1. To replace GW170817 in the context of BNS mergers with electromagnetic counterparts: GW170817 was rich with data, possibly because it was close and well inclined. How *lucky* were we to observe such a event? How likely are we to observe another equivalent event?

2. To inform the designing of trigger and follow-up instruments by describing expected targets and their features: By predicting the typical fluxes, variability and evolution timescales of electromagnetic counterparts, the designing of instruments' sensitivities, cadences and integration times can be optimized.

3. To guide multi-messenger observation campaigns by predicting trends between the observables of some counterparts and the detectability of others: In the population of multi-messenger events to be detected, there are correlations between, e.g., afterglow flux levels, time of peak and source proper displacement, which are useful to tailor searches for counterparts. Furthermore, such trends can help to identify the electromagnetic counterpart among a set of candidates.

4. To outline the data which will be available for future multi-messenger studies. Indeed, the observation of GW and electromagnetic counterparts make for new types of datasets—a GW inspiral signal with an afterglow light curve, etc.—that can be analyzed and provide new astrophysical insight. By describing which types of datasets should be available in the multi-messenger era we seek to discern in the pool of multi-messenger methods those

which will effectively be applicable.

In this chapter we will motivate and describe the multi-messenger population model developed during the doctoral studies. The results and applications of this population model will make up the remaining Chaps. 5, 6 and 7 of Part II.

4.2 Some early takes on multi-messenger prospects

In the pre-GW170817 era, population prospects for multi-messenger events drew on the only expected counterpart to have been already observed: short GRBs. The luminosity function of sGRB was inferred for example by (e.g., D'Avanzo et al., 2012; Wanderman & Piran, 2015; Ghirlanda et al., 2016). This was particularly the case of two studies on which our own population model is based: (Ghirlanda et al., 2016, , hereafter referred to as "G16") and (Wanderman & Piran, 2015, , hereafter, "WP15"). Both these studies seek to constrain the luminosity function of short GRBs, which is modeled as a broken power law; This is the only model adopted by G16 and the preferred model in WP15:

$$\phi_{\rm sGRB}(L) \propto \begin{cases} \left(\frac{L}{L_*}\right)^{-\alpha_1} & \text{for} \quad L < L_* \\ \left(\frac{L}{L_*}\right)^{-\alpha_2} & \text{for} \quad L_* < L \end{cases} \tag{4.1}$$

$\phi_{\rm sGRB}$ is defined such that the probability of the luminosity of a short GRB be in the $[L_1, L_2]$ interval is $\int_{L_1}^{L_2} {\rm d}L \phi_{\rm sGRB}(L)$. Both studies seek to infer the population indices α_i's and the break luminosity L_* and their associated uncertainties with a Monte Carlo Markov chain (MCMC) algorithm. These studies differ in the three crucial steps of MCMC population inference: the sample selection, the choice of the individual observables, and the likelihood function. These differences add to the inherent uncertainty in the collection of the short GRB sample linked to the difficulty in detecting their afterglows and thus their redshifts (Sec. 1.3)

This results in their very different results on the low-luminosity slope, which is constrained to $\alpha_1 = 0.53 \pm 0.5$ by G16 and $\alpha_1 = 2 \pm 0.1$ by WP15. These are the most extreme values of population indices among published short GRB luminosity constraints: the flatter luminosity function of G16 predicting many more luminous events than WP15. In our population model, we will consider both these luminosity functions and consider the results obtained with each of them as confidence bounds representing the uncertainty stemming from the uncertainty on the luminosities of short GRB jets.

In constraining the luminosity function, both studies constrain the redshift evolution of short GRBs and thus of their progenitor systems assumed to be binary neutron star or neutron star-black hole mergers. Assuming the then-projected interferometer range of $\sim$ 300 Mpc, they make predictions for short GRB-GW co-detections with rates of $0.01 - 0.5$ yr^{-1}. Considering the beaming factor of GRB jets $\gtrsim 10$, the expectations are higher for GW triggers from misaligned binaries, that lack a short GRB prompt signal but present an afterglow counterpart; G16 advocate for effort to follow-up GW events in the search for these "orphan afterglows". The population model we develop below considers exactly this type of multi-messenger event.

Shortly after GW170817, further prospects for GW-short GRB combinations were made, with more detailed modeling for the gamma-ray flux from structured jets or the shock breakout mechanism (e.g., Beniamini et al., 2019; Saleem, 2020). Such studies further quantified the small expected rates of joint short GRB-GW detections and pointed to the rareness of GW170817-like events in this respect.

Concerning joint GW–jet afterglow associations, Saleem et al. (2018b,a) use detailed numerical models for the multi-band afterglow and GW signals to make population prospects. Their studies are however based on uninformative (e.g., uniform or log-uniform) event parameter distributions. Moreover, the discussion therein is centered around the parameter-space constraints that can be derived from detections or non-detections of the afterglow in various electromagnetic bands, and on the conditions necessary to observe afterglows from off-axis lines of sight. A similar approach was followed by Gottlieb et al. (2019), with a focus on the environment or shock conditions constraints deduced from non-detections of afterglow counterparts.

An early population model for the optical afterglow from a structured jet can be found in Lamb & Kobayashi (2017). Here, the authors focused on the detectability of afterglows with a focus on the dependence to the core jet's opening angle and the contrast with the kilonova signal, which can outshine the early afterglow for very misaligned jets.

In the prospects for afterglow counterparts we develop below (Sec. 5.5) we will use astrophysically motivated parameter distributions and discuss, rather, the effect of the population parameters and detector configurations on the expected population of afterglows.

Finally, though the kilonova counterparts are the prime means to locate the merger in the sky and search for further counterparts such as the afterglow, these signals were not the object of as many population studies as the jet afterglow was. This might be due to the large uncertainty on any population results stemming from the current uncertainty in kilonova modeling. Nonetheless, we led a first population study on the kilonova counterpart to identify the main trends (Sec. 5.3). In this study, we integrated this modeling uncertainty, and nourished a discussion on this uncertainty.

4.3 Ingredients of a population study

A prospective population study seeks to predict the population of sources that will arise from observing campaigns. To carry out such studies, one must specify (i) the relevant physical parameters describing a source and their distribution among the sources (i.e., the population model), (ii) an emission model allowing to predict the signals from each system with these parameters, and (iii) a detection model that will emulate the observing and selection process, leading to the *detected pop-*

ulation. In theory, the detection model must include all the observational feats, including non-instrumental, human effects: Follow-up can only be carried out for a finite time, it is not systematic, etc. In practice, these effects are extremely difficult to model, and therefore we should rather describe the events our model selects as *detectable*. More on this topic is given in Sec. 5.7.

The strength of multi-messenger population studies is to draw from various branches of astrophysics for the population, emission and detection models. Thus results from, e.g., GW inspiral physics, relativistic jet physics, high-energy dissipation mechanisms, nucleosynthesis calculations are combined.

In this population model we are mostly interested in BNS mergers. On the one hand, there is more variability in NSBH binary parameters—such as component masses and spins—than in BNS systems. These parameters influence the post-merger outflows and thus the electromagnetic counterparts (e.g. Nedora et al., 2021). This adds some complexity that we could not control in our population model based on practical, readily accessible models for the electromagnetic counterparts; We thus chose to restrain our study to BNS progenitors. On the other hand, we wish to draw on short GRB science to inform our parameter distributions, for example using their luminosity function; and it is clear that short GRB progenitors include both BNS and NSBH systems. However, it seems that the current estimates for NSBH component masses and spins suggests that most NSBH systems end with a direct plunge of the NS into the BH, thus without mass ejection necessary to short GRB generation (Foucart, 2020). This is confirmed by the two first NSBH inspiral signal detections, in which the inferred component masses and spins suggest that no material was ejected in these systems ($\leq 10^{-6} M_\odot$, Abbott et al. 2021c). We can therefore safely consider the observed population of short GRBs to primarily descend from BNS systems.

Our population study covers all the signals observed for GW170817: the inspiral gravitational wave signal, the kilonova, the jet afterglow photometry and imagery, and the short GRB:

- Concerning the gravitational-wave signals, we adopted two different approaches, according to the specific application and the availability of LVKC data and software. In Sec. 4.4.1, we describe state-of-the-art GW detection modeling, making use of LVKC proprietary resources. We will only use this particular modeling in Chap. 6 on multi-messenger cosmology, where precision cosmology predictions require precise gravitational-wave modeling. In Sec. 4.4.2, we describe a simplified model for GW detection, which suits the needs of the Chap. 5, where we give the prospective results of our population study. Also, this simplified model affords more intuition and is easier to manipulate analytically.

- For the kilonova counterpart, we made the hypothesis that the searches for the kilonova in the GW skymap were not limited by the size of this skymap; We justify this hypothesis below with the recent astronomical community's performance during O3. For the actual signal, we built a simple viewing-angle-dependent model for the peak magnitude in various bands deduced from state-of-the art modeling of the ouflows from BNS mergers and radiation therefrom.

- Supposing that all BNS merger produce relativistic jets, we applied standard GRB afterglow theory to predict the jet afterglow photometry. While describing the jet afterglow emission model, we also discuss our hypothesis of systematic BNS merger–jet association.

- The proper displacement of the afterglow shock front on the sky was a breakthrough observation of GW170817 (Sec. 3.3). This will therefore also be part of the predictions of our population model. For this observable, we also consider two different models: a simple approach concerned only with the jet's proper motion at afterglow peak, and a fuller model that describes the total angular displacement of the source in the sky over the course of the afterglow.

- For the short GRB, we shall derive a criterion based only on the observer's alignment with the jet. This is only valid for bright GRBs, those observed with lines of sight close or within their jet openings, as we discussed in Sec. 3.2.3.

4.4 The emission and detection models

4.4.1 Inspiral gravitational-wave signals: detailed modeling

In GW searches an event is detected if its *detected signal-to-noise ratio* (SNR) $\hat{\rho}_{\rm det}$ exceeds a certain threshold. The detected SNR is a measure of the power of the GW signal recorded at the GW detectors. The higher the SNR, the more chance there is to recognize the signal from the detector noise at the moment of detection. It is important to note that the detected SNR $\hat{\rho}_{\rm det}$ is different from the *optimal-filter SNR* $\rho_{\rm opt}$ which is calculated taking into account the average sensitivity of the detector network instead of the noise realization at the moment of the detection, which are different because of noise fluctuations. Both $\hat{\rho}_{\rm det}$ and $\rho_{\rm opt}$ are calculated using the usual SNR definition (Eq. 2.34):

$$\rho^2 = \int \mathrm{d}f \frac{\tilde{s}(f)\overline{\tilde{s}(f)}}{S_n(f)} \tag{4.2}$$

where $\tilde{s}$ denotes the Fourier transform of the signal, and S_n is the noise power spectrum density (PSD) of the interferometer, evaluated either at the moment of detection for $\hat{\rho}_{\rm det}$ and on average for $\rho_{\rm opt}$. Up to a scaling, the noise power spectrum density is the interferometer sensitivity curve, as plotted in Fig. 2.3. Eq. 4.2 makes it obvious why a better detector has a lower sensitivity curve, and why one can tune the sensitivity minimum to the target signals with signal recycling and other optical setups.

Also, due to noise fluctuations, the $\tilde{s}$'s will be different for the detected and optimal filter signals. Therefore the measured

luminosity distance and inclination angle $\hat{D}_L$ and $\cos\hat{\iota}$ will differ from the true D_L and $\cos\iota$. The *GW likelihood model* is a function which, given true D_L and $\cos\iota$, allows to sample the $\hat{D}_L, \cos\hat{\iota}$ and the corresponding $\hat{\rho}_{\rm det}$ which would be detected in presence of the true signal in the detectors.

In this section and in Sec. 6, we use the Cutler and Flanagan (CF) approximant for the GW likelihood (Cutler & Flanagan, 1994; Poisson & Will, 1995; Chassande-Mottin et al., 2019). This approximant to the GW likelihood is accurate under the assumption that the chirp mass of the signal is well estimated, as is always the case for BNS detections (Cutler & Flanagan, 1994). By sampling from the CF approximant we can obtain a value of the detected $\cos\hat{\iota}$ and $\hat{D}_L$ which can then be used to compute the detected SNR $\hat{\rho}_{\rm det}$ with the following (Cutler & Flanagan, 1994; Chassande-Mottin et al., 2019):

$$\hat{\rho}^2_{\rm det} = \rho^2_{\rm fo}\sigma_d\left[(\chi^2_+ + \cos^2\hat{\iota}) + \epsilon_d(\chi^2_+ - \cos^2\hat{\iota})\cos(4\overline{\psi})\right], \quad (4.3)$$

where $\chi_+ = (1+\cos^2\hat{\iota})/2$, while ϵ_d, σ_d are variables which depend on the detector network and the sky-position of the GW source and $\overline{\psi}$ is the GW polarization angle. The software to sample from the CF likelihood is a typical LVCK proprietary resource which allowed the sophisticated GW detection model described here.

The variable $\rho_{\rm fo}$ is the optimal-filter SNR that the binary would have had if it had been face-on,

$$\rho^2_{\rm fo} = \frac{1}{\hat{D}^2_L}\left[\frac{5}{6\pi^{4/3}}\frac{G\mathcal{M}_c^{5/3}}{c^3}\int_{f_{\rm low}}^{f_{\rm LSO}} {\rm d}f\,\frac{f^{-7/3}}{S_{n,aver}(f)}\right], \quad (4.4)$$

with $S_{n,aver}(f)$ the harmonic mean of the noise PSDs of the interferometric detectors composing the network and $\mathcal{M}_c$ the system's chirp mass in the detector frame. In Eq. 4.4, one can recognize the leading-order post-Newtonian inspiral signal, which we describe in more detail in Appendix A in deriving the simplified GW detection model of Sec. 4.4.2.

As we show below when choosing the PSD considered for our study, all the systems we consider in our multi-messenger studies have low redshift, $z \leq 0.06$ for both Planck (Planck Collaboration et al., 2020b) and SH0ES (Riess et al., 2019a) values of H_0. Therefore, we assimilate the chirp mass and last stable orbit frequency $f_{\rm LSO}$ in the detector frame to their values in the source frame. The integral low boundary $f_{\rm low} = 20$ Hz is set to the low-frequency cut-off for ground-based GW detectors.

We seek to evaluate the GW detection probability, i.e. the function $p^{\rm GW}_{\rm det}$ such that a binary with true D_L and ι have probability $p^{\rm GW}_{\rm det}(D_L, \iota)$ of being detected. We divide the $\cos\iota$ range in 200 bins and we simulate 10000 BNS merging at fixed D_L and uniformly distributed over the celestial sphere. The BNS masses are generated from a Gaussian distribution with mean $1.35M_\odot$ and standard deviation $0.15M_\odot$, as observed for Galactic binary neutron stars (Farrow et al., 2019). For each binary we then draw a detected $\hat{D}_L$ and $\cos\hat{\iota}$ from the CF likelihood and calculate $\hat{\rho}_{\rm det}$ following Eqs. 4.3–4.4. We then count the binaries with SNR exceeding the fiducial online match-filtering threshold for detection of 14[1] and compute the detection probability as the fraction of BNS events detected.

In this approach, the binary's chirp mass and the noise fluctuations are technically not bounded, and thus there is no maximum distance to which a binary can be detected. However, to provide a distance scale to the study, we must define a maximum system distance. Following Chen et al. (2021a), we define the 0.2% *response distance* $d^r_{0.2\%}$ at which 0.2% of the simulated binaries (with isotropic distribution in the sky and orientation) will be detected by the network.

Fig. 4.1 shows the GW detection probability as a function of the BNS D_L and $\cos\iota$ marginalized over the GW polarization angle and sky-position. Another consequence of the low redshift range of our simulation, the shape of the GW detection probability functions are not affected by non-linear D_L–z relationship, and are therefore the same for all PSD hypotheses.

As it can be seen from Fig. 4.1, face-on binaries are easier to detect as the GW emission is stronger perpendicularly to the orbital plane, and they can be observed at higher luminosity distances. This is in line with the physical intuition formed on GW emission in Sec. 2.2, and motivates the derivation of a simpler GW detection model.

4.4.2 Inspiral gravitational-wave signals: simplified modeling

The GW detection model developed in the previous section is operational from a numerical point of view. It is realistic thanks to access to instrument-level data from the LVKC. It is however not intuitive as most of the physics are hidden in the evaluations of Eqs. 4.3, 4.4. Fig. 4.1 shows that, in the end, the main remaining effect is simply due to anisotropy in GW emission. We seek a handy criterion which (i) depends only on source distance and inclination angle, (ii) encapsulates only the said anisotropy and (iii) is linked to instrumental data by a single, readily accessible number.

Such a criterion is derived in Appendix A, and is reproduced here for clarity. Given a system with distance D_L and inclination angle ι, we will deem it detected by a single interferometer if it satisfies the following:

$$\sqrt{\frac{1+6\cos^2\iota+\cos^4\iota}{8}} > \frac{D_L}{\overline{H}} \quad (4.5)$$

Here, the number of parameters was reduced to two, as required. This reduction was achieved by averaging the SNR of systems over the other source parameters, such as sky location. In doing so, the distance to which an optimally-aligned (i.e., $\iota = 0$) system can be detected must be revised, leading to the definition of the sky-location-averaged horizon distance $\overline{H} = \sqrt{2/5} \times H < H$. The original horizon distance H is defined as the distance to which an optimally-aligned and optimally-located system can be detected, and is a readily accessible number for all interferometers of the LVKC.

In Fig. 4.1, we plot $p^{\rm GW}_{\rm det}$ according to this new criterion. Given the simplicity of Eq. 4.5, the similarity with the detailed

[1] Usually an SNR threshold of 8 is assumed for the detection on a single detector Abbott et al. (2020a), as the CF approximator is valid in the high SNR regime, here we assume a threshold of 14. This is equivalent on average to an SNR of about 8 in each detector.

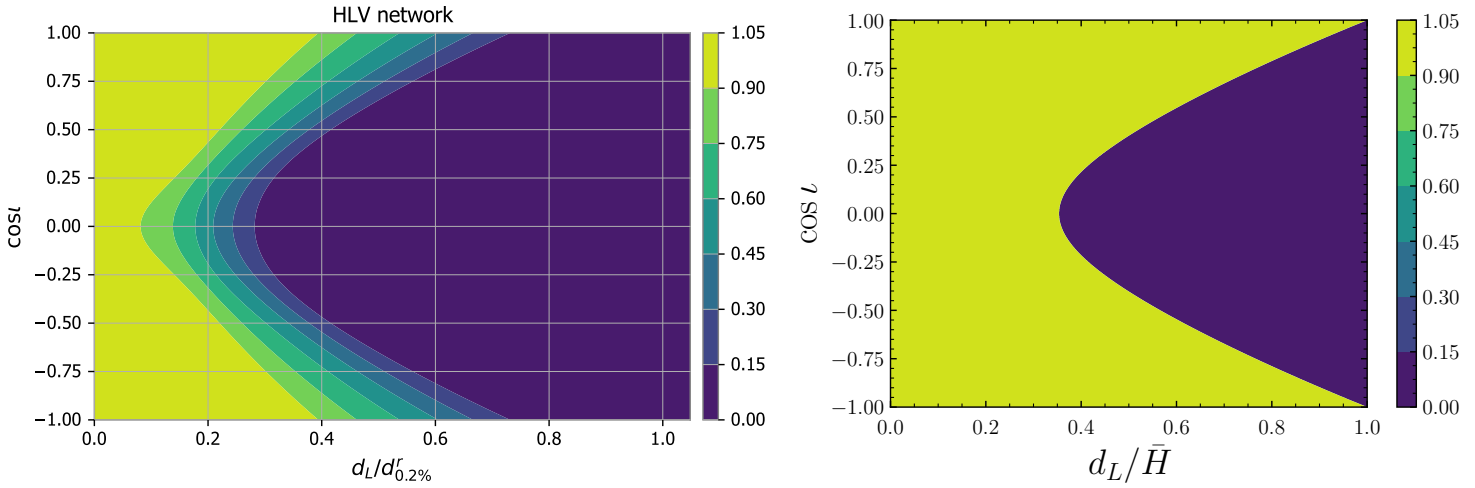


Figure 4.1: **GW detection probability for a BNS merger.** Left: GW detection probability according to the full GW detection modeling (Sec. 4.4.1), using an SNR threshold of 14. The horizontal axis is scaled to the BNS 0.2% response distance $d^r_{0.2\%}$. In the multi-messenger cosmology study (Chap. 6), these are $d^r_{0.2\%} = 107$ Mpc for O2; $d^r_{0.2\%} = 218$ Mpc for O3; $d^r_{0.2\%} = 477$ Mpc for O4. Right: GW detection probability assuming the simplified detection model (Sec. 4.4.2). In both cases, this plot scales with the horizon distance. [Mastrogiovanni et al. 2021]

GW detection model is striking. A rough comparison leads to the estimate $\overline{H} \sim 0.5 \times d^r_{0.2\%}$. Recall however that the full criterion is for a network of interferometers, whereas the one described in Eq. 4.5 is for a single instrument.

The criterion of Eq. 4.5 is valid for a detection using a single instrument. However, GW detection by the interferometer network is based on multi-instrument analysis in order to reduce false positives. Therefore, when using this criterion, we will systematically consider the horizon of the second-most sensitive instrument of the interferometer network, that is, the LIGO-Hanford instrument. In addition, as was illustrated in the case of GW170817, locating the kilonova is indispensable to access the other counterparts. A double or triple GW detection facilitates this by reducing the GW localization map. However, as we will comment in Sec. 4.4.3, the sky surface to cover should not be the limiting factor of future optical follow-up, due to the large fields of view of high-cadence instruments.

As shown in Appendix A, assuming homogeneity of the sources within the horizon and isotropy of the binary polar direction, the mean viewing angle of GW-detected events is ~ 38 deg and the fraction of GW-detected events among all mergers is $\sim 29\%$, both regardless of the horizon value. This latter fraction is, thus, an absolute maximum for the fraction of GW triggers events among all BNS mergers.

The various instances of this population model were run in preparation for, during, or after LVKC runs. Therefore, the horizons considered varied from one study to another, and in Tab. 4.1 we collect the horizons assumed for the results presented in Chap. 5 and following.

Table 4.1: **Values assumed for $\overline{H}$ in the various instances of the population model described in this section.** These were derived from the ranges cited in Abbott et al. (2020a) and correspond to 1.4+1.4 $M_\odot$ systems. For the particularly massive GW190425, we adapted the horizon value, see Sec. 5.4 for details.

Run	$\overline{H}$ [Mpc]
Secs. 5.3, 5.4 and 5.6	
O3	157
O3@GW190425	181
O4	229
O5	472
Sec. 5.5 and Chap. 7	
O2	85.8
O3	143
Design	272

4.4.3 Kilonova signals

The kilonova peak magnitude depends on the distributions of mass, velocity and composition of the ejected material and on the viewing conditions: distance and viewing angle. The ejection is anisotropic with a neutron-rich, dynamical ejecta in the equatorial plane, where the formation of lanthanides leads to a large opacity while a relatively neutron-poor wind of lower opacity is blown in the polar direction (Fernández & Metzger, 2016; Metzger, 2019; Barnes, 2020). This wind is expected to be present when a short-lived massive neutron star is formed before collapsing to a black hole, but probably not in the case of a direct collapse. The lanthanide-rich ejecta produces the "red kilonova", which peaks in the near-infrared while the neutron-

poor wind is responsible for the "blue kilonova" at optical wavelengths. The blue kilonova declines on a timescale of one day, whereas this is one week for the red component.

For the purpose of our population model, our default scenario assumes that all kilonovae have an quasi-isotropic red component and a polar blue component and we obtain the peak absolute magnitude at a given wavelength and viewing angle from the following simple parametrization

$$M_{\lambda,\theta_v} = \begin{cases} M_{\lambda,0} + \Delta M_\lambda \left(\frac{1-\cos\theta_v}{1-\cos\theta_0}\right) + \delta M_\lambda, & \theta_v \leq \theta_0 \\ M_{\lambda,0} + \Delta M_\lambda + \delta M_\lambda, & \theta_0 \leq \theta_v \end{cases} \quad (4.6)$$

where $M_{\lambda,0}$ is the peak absolute magnitude for a polar viewer, ΔM_λ is the amplitude of the polar effect and δM_λ represents the intrinsic (i.e., non-viewing-angle-related) variability assumed for kilonovae. For $\theta_0 = 60\,\mathrm{deg}$, we find that the linear-in-$\cos\theta_v$ form of Eq. 4.6 reproduces the trends of sophisticated kilonova modeling work (Wollaeger et al. 2018, Kawaguchi et al. 2020, and the "asymmetric model" of Villar et al. 2017). We chose δM_λ to be uniformly distributed in $[-1, 1]$, reproducing the expected variability in KN magnitude stemming from variability in ejecta mass, velocity and opacity (Wollaeger et al., 2018, Eq. 33). The difference in magnitude between equatorial and polar views is moderate in the infrared but increases rapidly in the visible, already reaching about magnitude 4 in the r band. This is mainly due to the stronger anisotropy of the blue component.

To calibrate this expression, we use AT 2017gfo assuming $\theta_v = 15\,\mathrm{deg}$, as derived with the VLBI data (Sec. 3.3). Corresponding values can be found in Tab. 4.2.

Table 4.2: **Parameters for the kilonova peak absolute magnitude dependence on the viewing angle,** as given in Eq. 4.6.

Band	$M_{\lambda,0}$	ΔM_λ
g	-16.3	7
r	-16.3	4
i	-16.4	3.5
z	-16.5	2.5

Calibrating Eq. 4.6 with AT 2017gfo supposes that this transient was representative of the kilonova population. This is the minimal hypothesis one can make while waiting for the number of kilonovae with robust angle measurements to increase in the future. We note that AT 2017gfo could have been brighter or dimmer than the average of the population our model seeks to encapsulate. We briefly indicate below how our results might change if this is indeed the case.

As the polar ejecta may not be produced in all mergers, depending for instance on the post-merger formation of a massive neutron star before the collapse to a black hole (see, e.g., Metzger, 2019), we also consider the possibility that a fraction of the kilonova population lacks the blue component, affecting the kilonovae brightness in the bluer bands (see a preliminary luminosity function in Ascenzi et al. 2019 and related discussions in Gompertz et al. 2018 and Kasliwal et al. 2020).

The flux criterion for kilonovae signals in a given band m is simply that their peak magnitude be below a detection threshold:

$$m_{\lambda,\theta_v} < m_{\mathrm{lim}} \quad (4.7)$$

Because these limiting magnitudes depend on the follow-up network and strategy (galaxy-targeting, tiling, survey, see Sec. 2.4.2) and are likely to evolve in the near future with the commissioning of new instruments, we did not select a particular set of limiting magnitudes. Rather, we compared the detectable population assuming various m_{lim}'s and, when comparing to the results of actual observing runs, we adopted the relevant thresholds.

4.4.4 The question of source localization: a first approach

In GW170817, it is the kilonova that allow to pinpoint the source in the sky and therefore access further counterparts such as the jet afterglow. In the future, the kilonova could be bypassed with radio surveys that detect the afterglow directly (Sec. 2.4.2). Such a scenario should however not occur before the advent of deep radio surveys such as the Square Kilometer Array Braun (2014); Dewdney (2015). Waiting for the radio array to detect the afterglow will only allow for a very late access to the kilonova signal, which would surely make the latter less interesting, especially concerning the faster-evolving blue component. A short GRB detected by a repointing satellite such as *Swift* also has the capability to locate the source, though this channel will only be effective for the unlikely aligned systems with a classical fast-rising X-ray afterglow. Thus, it is important for our population model to incorporate the localization of the kilonova by optical follow-up.

In the perspective of upcoming high-cadence and large-FOV optical facilities such as the Zwicky Transient Facility (ZTF, Bellm et al., 2019) or Vera C. Rubin Observatory (LSST, Ivezic et al., 2008), we assume the detection of the kilonova is limited only by the kilonova magnitude and sky-position, and not by the size of the GW-provided skymap that the follow-up network must cover in its searches. Indeed, we consider that these survey facilities can cover all the sky available to them within the first nights of the search, that is, before the estimated time for significant dimming of kilonovae signals.

While this can be considered an optimistic assumption, this level of performance was indeed reached during the campaigns following GW events in the recent O3 run by, e.g., the GROWTH collaboration (Kasliwal et al., 2020). In particular, for the only confirmed BNS event of the O3 run GW190425 (Coughlin et al., 2019b), the ZTF covered the $\sim 8000\mathrm{deg}^2$ of the skymap overlapping with their night sky.

Therefore, we will consider a source localized by its kilonova counterpart if it is in the sky accessible to electromagnetic follow-up. For observatories at latitudes of ~ 33 deg such as the ZTF (northern hemisphere) and the LSST (southern hemisphere), this represents a season-averaged fraction $p_{\mathrm{night}} \sim 52\%$ of the whole sky (Bellm, 2016). We therefore define a *localization criterion*, which is a simple coin toss with probability

$p_{\rm night} = 0.52$. We will only use this criterion in the multi-messenger cosmology study (Chap. 6).

4.4.5 Jet afterglow radio photometry

As we will show in Sec. 7.4 (Fig. 7.3), the radio 3 GHz band and the X-ray band are both suitable to detect the afterglows of jets from BNS mergers. This is a result of our population model, fully taking into account the sources' distributions in viewing angle, distance, etc. and the typical sensitivities of current follow-up instruments. The optical band, in contrast, provides a smaller probability of detection. However, the radio band also provides the possibility of detecting the source proper displacement, which we are also interested in predicting. Therefore, our afterglow population predictions will focus on the radio band.

The multi-wavelength afterglow of GW170817 played a central role in the scientific outputs of this event, as described at length in Sec. 3.3. The afterglow is associated with the deceleration of a structured relativistic jet emitted by the central source formed in the merger. More precisely, it is due to the synchrotron emission of electrons accelerated by the forward shock propagating in the circum-merger medium. Remarkably for GW170817, the emission was produced in the same spectral regime of the slow-cooling synchrotron process, that is, $\nu_m < \nu < \nu_c$, and the power-law distribution of shock-accelerated electrons assumed a non-evolving slope $p \sim 2.2$ until 1000 days after the merger (Sec. 3.4, Nakar et al., 2018; Mooley et al., 2018c; Troja et al., 2019a).

Once the existence of the jet was established, the observed slow rise of the afterglow in GW170817 was clear evidence of our misaligned line of sight and of the structuredness of the jetted outflow: a significant variation of the outflow's energy and Lorentz with the angle from the jet's axis. Typical fitting to the afterglow photometry use power laws for the structures in isotropic-equivalent outflow energy:

$$E_{\rm iso}(\theta_v) = E_{\rm iso,c} \times \begin{cases} 1 & \text{for} \quad \theta_v < \theta_j \\ \left(\frac{\theta_v}{\theta_j}\right)^{-a} & \text{for} \quad \theta_j < \theta_v \end{cases} \tag{4.8}$$

and for the initial outflow Lorentz factor:

$$\Gamma_0(\theta_v) = 1 + (\Gamma_0 - 1) \times \begin{cases} 1 & \text{for} \quad \theta_v < \theta_j \\ \left(\frac{\theta_v}{\theta_j}\right)^{-b} & \text{for} \quad \theta_j < \theta_v \end{cases} \tag{4.9}$$

With such prescriptions, the data for GW170817 required steep slopes of $a, b \gtrsim 2$ (Gill & Granot, 2018; Ghirlanda et al., 2019).

For our population model, we seek to faithfully reproduce afterglows of BNS mergers, and thus set out to derive an emission model from structured jets, just as GW170817. As mentioned above, the existence of a relativistic jet in GW170817 was not settled before the VLBI data were acquired. The emerging of the jet may not have been settled at all without either these or late-time observations, if the event had been more distant or less energetic, for instance. Though there is some evidence for a positive answer to the emerging of jets in all BNS mergers (e.g., Beniamini et al., 2019; Duffell et al., 2018), it is not certain that every merger produces a relativistic jet. The observation of low-luminosity GRBs and some evidence for shock-breakout emission or "choked" jets in some bursts blur the BNS–jet association (Secs. 1.4 and 3.3). The singularity of GRB170817A and its established origin outside of the jet's core only add to this blurring. Finally, though GW170817 is a strong signal in favor of this, the identification of BNS mergers with short GRBs is not completely clear, even if only because of the ill-definiteness of GRB durations and thus of the short GRB sample.

Nonetheless, our idea in these population prospects is to draw from short GRB science—its results on jet energetics and physical conditions, short GRB environments, jet afterglow emission—and we will thus assume that all BNS mergers lead to relativistic jetted outflows, with structures similar to GW170817. However as we shall soon see, the actual structure of the jet has little impact on our predictions.

An example of the radio light curves obtained for such a structured jet is plotted in Fig. 4.2 (left) using parameters typical of the various fits to the data that have been published (e.g., Gill & Granot, 2018; Troja et al., 2019a; Ghirlanda et al., 2019).

In order to calculate the radio afterglow light curves from such a structured jet, we divide the jet into a core component and a number of angular rings. The deceleration dynamics of the material in different rings is calculated assuming an adiabatic deceleration of the outflow, as derived in Appendix D. In line with the typical environments expected for short GRBs and the large offset of GW170817 from its host galaxy, we assumed a homogeneous circum-merger medium (Secs. 1.3, 3.2). The dynamics of rings of different latitudes are computed independently. This is a simplifying assumption, as the jet material is expected to interact and spread over time. The expansion of the jet can influence the afterglow properties—especially at late times—and any subsequent measurement with the data (Lamb et al., 2018; Fernández et al., 2021).

At a latitude of θ, the deceleration radius is $R_{\rm dec}(\theta) = \left(\frac{3\epsilon(\theta)}{\Gamma_0^2(\theta) n_{\rm ext} m_{\rm p} c^2}\right)^{1/3}$, which is constant at the core and slowly increases with θ outside the core: $R_{\rm dec}(\theta) \propto \theta^{(2b-a)/3}$. For $a = 4.5$ and $b = 2.5$ chosen in Fig. 4.2 below, we have $\frac{2b-a}{3} = 0.17$.

Subsequently, the synchrotron emission of forward-shock-accelerated electrons in the shock frame is assumed to follow the standard synchrotron spectrum including the self-absorption correction, which is derived in Appendix E. Finally, the observed light curve is computed by summing the contributions of all rings on equal-arrival time surfaces, taking relativistic beaming and Doppler boosting into account.

The separate contributions of the core jet and the sheath are plotted in Fig. 4.2 (left) and the emergence of the core at the peak is clearly visible. The evolution of the peak flux of the same structured jet is plotted as a function of the viewing angle in Fig. 4.2 (right), along with the ratio of total and core-only contributions to the peak flux. Interestingly for the selected jet structure, this ratio is almost constant at ~ 1.5, except for the aligned cases ($\theta_v \leq \theta_j$) where the core is more dominant.

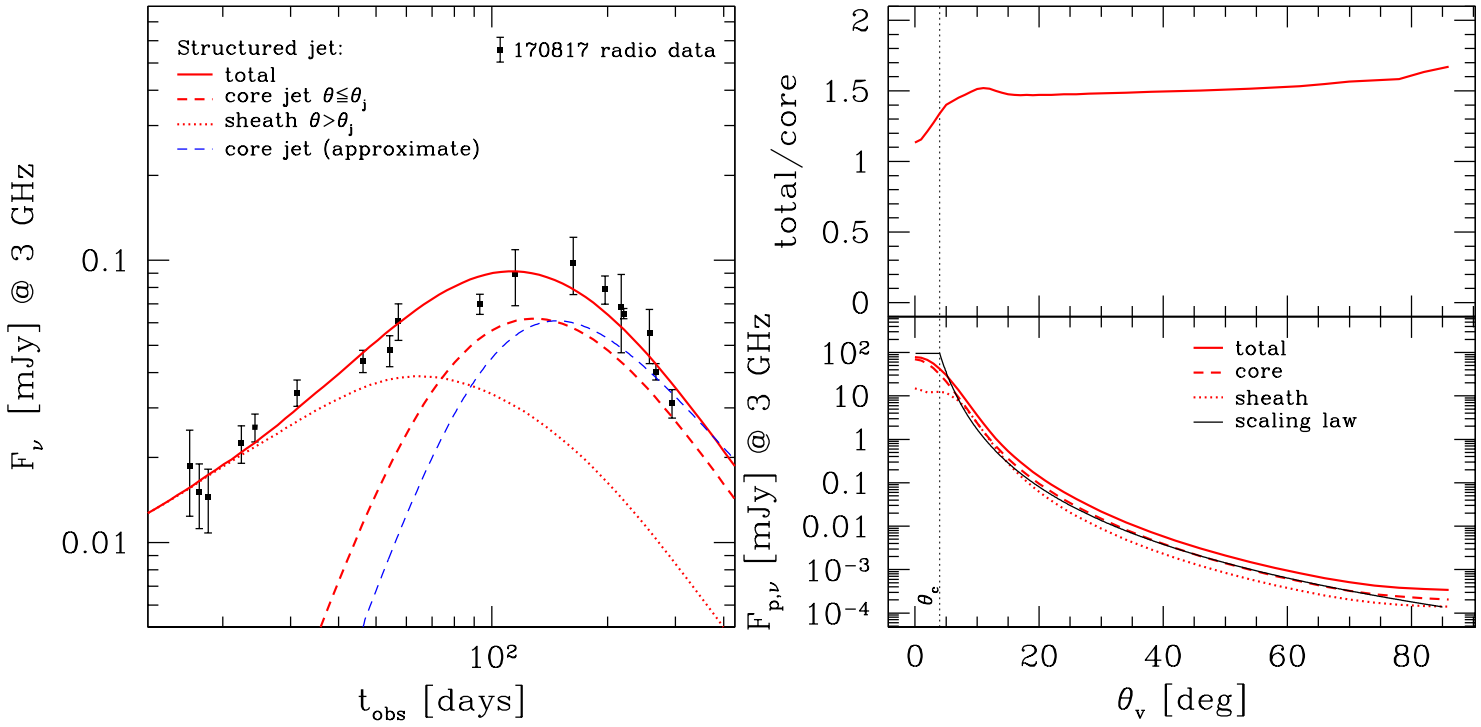


Figure 4.2: **Radio afterglow light curve from a structured jet.** Left: 3 GHz light curve of afterglow from structured jet at distance $D_L = 42$ Mpc, with sharp power-law structure ($a = 4.5$, $b = 2.5$; see Eq. 4.8-4.9), viewing angle $\theta_v = 22$ deg, external density $n_{\rm ext} = 3\,10^{-3}\,{\rm cm}^{-3}$, core jet with $\theta_j = 4$ deg, $\Gamma_{\rm c} = 100$, $E_{\rm iso,c} = 2 \times 10^{52}$ erg, and with microphysics parameters $p = 2.2$, $\epsilon_e = 0.1$ and $\epsilon_B = 10^{-4}$. The radio observations of GW170817 are also plotted and are compiled from Hallinan et al. (2017); Alexander et al. (2018); Margutti et al. (2018); Mooley et al. (2018c,c); Dobie et al. (2018). The respective contributions of the core jet ($\theta \leq \theta_j$), the sheath and the total are plotted in dashed, dotted and solid red lines. The flux of the core jet computed with the simplified treatment described by Eq. 4.11 is also plotted in dashed blue line for comparison. Bottom right: Peak flux of same structured jet as function of viewing angle θ_v (solid red line), peak flux of light curve from core jet only (dashed red line) or from sheath only (dotted red line), and peak flux of core jet as computed using scaling law in Eq. 4.13 (thin black line). Top right: Ratio of peak flux from whole outflow (core jet and sheath) to that of core jet only. [Duque et al. 2019]

As established in the aftermath of GW17017 (Beniamini et al., 2020b, see discussion in Sec. 3.3,), the core jet is expected to dominate the flux at the afterglow peak, as long as the kinetic energy and the Lorentz factor decay steeply with θ. The precise value of the total-to-core ratio at the peak may slightly vary depending on the details of the assumed lateral structure. Whether we seek to select those events that are marginally detectable—with a peak flux around the detection threshold—or with an extended light curve—peaking well above the threshold—, our detection criterion will always concern the peak flux. Therefore, it is enough for our emission model to compute only the contribution from the core jet, while keeping in mind that it may slightly underestimate the total peak flux. This simplification is useful, as it allows us to disregard the actual structure, and to focus on the emission of a *top-hat* jet.

The contribution from a top-hat core jet can be efficiently computed following these steps:

1. Compute the flux $F_\nu^{\rm iso}(t_{\rm obs})$ from a spherical ejecta with initial Lorentz factor $\Gamma_{\rm c}$ and kinetic energy $E_{\rm iso,c}$.

2. Include the jet break to determine the flux from an aligned jet with finite size:

$$F_\nu^{\rm on}(t_{\rm obs}) = F_\nu^{\rm iso}(t_{\rm obs}) \times \begin{cases} 1 & \text{if } \Gamma\theta_j \leq 1 \\ (\Gamma\theta_j)^{-2} & \text{if } \Gamma\theta_j \geq 1 \end{cases} \quad (4.10)$$

3. Transform to a misaligned line of sight with the following formulae:

$$F_\nu(t_{\rm obs}) = \begin{cases} F_\nu^{\rm on}(t_{\rm obs}) & \text{if } \theta_{\rm v} < \theta_{\rm j} \text{ (on axis)} \\ a^3 F_{\nu/a}^{\rm on}(b\, t_{\rm obs}) & \text{if } \theta_{\rm v} > \theta_{\rm j} \text{ (off axis)} \end{cases} \quad (4.11)$$

with $a = \frac{1-\beta}{1-\beta\cos(\theta_v-\theta_j)}$ and $b = \frac{1-\frac{R}{ct}}{1-\frac{R}{ct}\cos(\theta_v-\theta_j)}$ where t is the central source frame time.

The final light curve derived from this method is plotted in Fig. 4.2 (left, blue dashed line) and corresponds well with the exact calculation of the core-jet contribution.

This top-hat jet afterglow calculation is standard. Therefore analytical expressions for the afterglow properties at the peak are available. They depend slightly on the assumptions for a possible late jet lateral expansion. If lateral expansion is taken into account like in standard GRB afterglow theory, the peak flux of the radio light curve scales as (Nakar et al., 2002):

$$F_{p,\nu} \propto E_{\rm iso,c}\, \theta_j^2\, n_{\rm ext}^{\frac{p+1}{4}}\, \epsilon_e^{p-1}\, \epsilon_B^{\frac{p+1}{4}}\, \nu^{\frac{1-p}{2}}\, D_L^{-2}\, \max(\theta_j, \theta_v)^{-2p}\,, \quad (4.12)$$

as long as the spectral regime remains $\nu_m < \nu < \nu_c$.

Upon simplifying the calculation down to power-law formulae such as Eq. 4.12, we suppose a single branch of the synchrotron spectrum applies for all events, and we disregard the effect of self-absorption and inverse Compton scattering. In the results of the population study reported in Chap. 5, we shall check the validity of this hypothesis and quantify the effects we neglect here.

At a frequency of 3 GHz, we obtain:

$$F_{p,3\,\rm GHz} = 8.6\, E_{52}\, \theta_{j,-1}^2\, n_{-3}^{4/5}\, \epsilon_{e,-1}^{6/5}\, \epsilon_{B,-3}^{4/5} \times D_{100}^{-2}\, \max(\theta_{j,-1}, \theta_{v,-1})^{-4.4} \text{ mJy}\,, \quad (4.13)$$

where $E_{52} = E_{\rm iso,c}/10^{52}$ erg, $\theta_{j,-1} = \theta_j/0.1$ rad, $n_{-3} = n_{\rm ext}/(10^{-3}\text{ cm}^{-3})$, $\epsilon_{e,-1} = \epsilon_{\rm e}/10^{-1}$, $\epsilon_{B,-3} = \epsilon_B/10^{-3}$, $\theta_{v,-1} = \theta_v/0.1$ rad and $D_{100} = D_L/(100\text{ Mpc})$, and where we assume $p = 2.2$.

This scaling law is plotted in Fig. 4.2 (right, solid black line) and, again, corresponds well with the structured jet and the top-hat calculations. This further simplification avoids us using Eqs. 4.10 and 4.11 and greatly speeds up our population calculations.

In some cases it is useful to isolate the dependency of the peak flux on the observing conditions, in which case we write:

$$F_{\rm 3GHz} \sim 8.6\, \varphi\, D_{100}^{-2}\, \max(\theta_{v,-1}, \theta_{j,-1})^{-4.4}\, \text{mJy} \quad (4.14)$$

where $\varphi = E_{52}\theta_{j,-1}^2 n_{-3}^{0.8}\, \epsilon_{e,-1}^{1.2}\, \epsilon_{B,-3}^{0.8}$ collects the flux dependencies on the parameters not related to the observing conditions. In Eq. 4.14, the normalization of φ was chosen such that $\varphi = 1$ for GW170817-like afterglows (e.g., Lamb & Kobayashi, 2017; Resmi et al., 2018; Lazzati et al., 2018; Troja et al., 2019a, or references above).

In the perspective of aiding follow-up campaigns and designing new follow-up instruments and strategies, we are also interested in the time at which BNS afterglows peak.

For a misaligned line of sight, the afterglow will peak once the core has decelerated enough to include the observer in its beaming cone. The expression of the peak time is therefore linked to the deceleration dynamics of the jet and its lateral expansion. Late observations of GRB 170817A give some evidence for the lateral expansion of the core jet, for example the temporal decay index being $\sim -p$ (Lamb et al., 2018). However, it is not clear if this expansion should be as strong for a core jet embedded in a sheath as for a 'naked' top-hat jet. Therefore, it is useful to consider both cases: with and without lateral expansion. The expressions in both cases are derived in Appendix H, where we also prove that the expansion dynamics only affects the peak flux slightly.

We reproduce the equations for the peak time under the jet expansion or no-expansion hypotheses here:

$$t_{p,\text{no ex}} = 4.9 \left(\frac{E_{52}}{n_{-3}}\right)^{1/3} \theta_{v,-1}^{8/3} \text{ days} \quad (4.15)$$

and

$$t_{p,\rm ex} = 5 \left(\frac{E_{52}}{n_{-3}}\right)^{1/3} \theta_{j,-1}^{2/3}\, \theta_{v,-1}^2 \text{ days} \quad (4.16)$$

In our population studies, unless otherwise noted, we assume Eqs. 4.13 and 4.15-4.16 to estimate the peak flux and peak time of the radio afterglow from a binary neutron star merger.

As hinted to above, the criterion for detecting the radio afterglow depends on the application: One can either require only a few points around the peak—as in Chap. 7–, or an extended

light curve to carry out full model fitting—as in Chap. 6. However, for a given jet structure, the temporal slopes α_+ and α_- of the increasing and decreasing phases of the light curve are independent of the viewing angle (Beniamini et al., 2020b). Thus, even requiring to detect a certain *time interval* or a *number of photometry points* of the afterglow is equivalent to specifying that the afterglow peak be larger than a given threshold.

Therefore, in most cases, it is natural to define the radio afterglow detection criterion as simply for the peak flux to be larger than the limiting flux of the radio array available for follow-up at the time of consideration.

In the various instances of our population model, we considered three typical radio sensitivities according to the epochs for which we sought to make predictions, which we summarize in Tab. 4.3.

4.4.6 Source proper displacement

We estimate the angular displacement of the shock front as:

$$\delta\theta_r = \delta t_{\rm VLBI} \times \frac{{\rm d}\theta}{{\rm d}t}_{|\rm max} \tag{4.17}$$

where $\delta t_{\rm VLBI}$ is the total time the afterglow remains detectable by the radiotelescope array, that is, the time its flux is above threshold. Monitoring the remnant for angular displacement requires detecting the afterglow in the first place; Therefore we fix the VLBI array threshold to the same level as in the afterglow detection criterion, though imagery actually requires a larger flux to detect the source than photometry. In Eq. 4.17, ${\rm d}\theta/{\rm d}t|_{\rm max}$ is the proper motion of the remnant at the time of the afterglow peak, i.e., when $\Gamma \times \theta_v \sim 1$. At this time, it is straightforward to estimate the remnant's proper motion as:

$$\frac{{\rm d}\theta}{{\rm d}t}_{|\rm max} \sim \frac{c\beta_{\rm app}}{D_L} \tag{4.18}$$

with $\beta_{\rm app} = \frac{\beta \sin\theta_v}{1-\beta\cos\theta_v}$ the apparent velocity of the remnant, $\beta = \sqrt{1-1/\Gamma_p^2}$ and $\Gamma_p = 1/\theta_v$ the jet head Lorentz's factor at the afterglow peak. By considering the source's proper motion to be that at afterglow peak during the entire follow-up, Eq. 4.17 in fact over-estimates $\delta\theta_r$. Thus, our prospects for events with detectable angular displacement using this model are optimistic.

Also, Eq. 4.18 is not valid if the observer is within the jet's opening, i.e., $\theta_v < \theta_j$ with θ_j the half-opening angle of the jet. In this case, no jet displacement is observed, $\delta\theta_r = 0$.

In this approach, one must have access to $\delta t_{\rm VLBI}$, which depends on the details of the light curve and therefore on, e.g., the jet structure, its expansion dynamics, and the surrounding medium density profile. To simplify, we assume all jets launched from mergers have the same structure as GW170817. In this case, we mentioned above that the slopes of the increasing and decreasing phases of the light curve are the same as for GW170817's afterglow, regardless of inclination angle (Beniamini et al., 2020b). Therefore, we empirically modeled the afterglow light curves as a broken power-law with temporal slopes $\alpha_+ = 0.80$ and $\alpha_- = -2.2$ (Mooley et al., 2018c, recall Sec. 3.3) respectively before and after the peak. Using the afterglow peak flux model (Sec. 4.4.5), the time the signal is above the radio threshold can thus be analytically estimated, and thereby the total source displacement.

Then, the criterion of this displacement is simply conditioned by the angular resolution of the VLBI network: $\delta\theta_r > \Delta\theta_{\rm VLBI}$, which was ~ 2 mas during the follow-up if GW170817 (Ghirlanda et al., 2019).

4.4.7 Source peak proper motion

If one chooses to disregard the jet structure and focus on the core jet only, a more simpler displacement detection criterion can be derived. It considers only the peak proper motion ${\rm d}\theta/{\rm d}t_{|\rm max}$ (Eq. 4.18), which it compares to a fiducial limiting proper motion $\mu_{\rm lim}$.

Such a criterion only makes sense if the VLBI observations are carried out on a same time stretch for all events, and systematically around the afterglow peak.

During the follow-up of GW170817, the uncertainties on the measured peak proper motion were $\sim 3\mu$as/day (Mooley et al., 2018b; Ghirlanda et al., 2019), which we will consider as $\mu_{\rm lim}$.

In this approach, for a given value of the viewing angle, there is a maximum distance until which the measurement of the proper motion is possible. It is given by

$$D_{\rm max,VLBI}(\theta_v) = \frac{c}{\mu_{\rm lim}} f(\theta_v) \sim 50\, f(\theta_v) \;\; {\rm Mpc}\,, \tag{4.19}$$

with

$$f(\theta_v) = \frac{\sqrt{1-\theta_v^2}\sin\theta_v}{1-\sqrt{1-\theta_v^2}\cos\theta_v} \tag{4.20}$$

4.4.8 Bright short GRBs

With a minimum peak luminosity of $\sim 10^{50}\,{\rm erg/s}$, as found by both G16 and WP2015, and a peak energy of the order of 1 MeV (Nakar, 2007), short GRBs seen from aligned lines of sight are detectable at any distance below 600 Mpc (peak flux of the order of $1\,{\rm ph/cm^2/s}$) and the main limitation for detection by gamma-ray satellites is their sky coverage. Therefore, we shall consider a bright short GRB to be detected if $\theta_v < \theta_j$.

As GW170817 showed, a short GRB can be detected even for larger viewing angles, and our criterion therefore applies for *bright* GRBs, similar to cosmological GRBs (see the discussion on GRB170817 and bright GRBs in Sec. 3.2.3). Nonetheless, integrating the viewing angle dependence of such a signal assumes to choose an emission model, and we do not make this choice. Our studies have dealt with this problem in much detail (Sec. 4.2).

4.5 The population model

The actual population model prescribes the distribution of the emission model parameters in the population. At the low redshifts we are interested in, we can consider the sources to be homogeneous in space and isotropic in inclination, therefore D_L

Table 4.3: **Radio sensitivities considered in various instances of our population model.**

Instrument	3 GHz detection threshold [μJy[	Epoch	Reference
Karl Jansky Very Large Array (VLA)	15	Current	
Phase I of the Square Kilometer Array (SKA1)	3	2025+	Braun 2014; Dewdney 2015
Phase II of the SKA (SKA2) or Next Generation VLA (ngVLA)	0.3	2030+	Braun 2014; Dewdney 2015; Selina et al. 2018

is distributed as D_L^2 and θ_v as $\sin\theta_v$. D_L and θ_v are the only parameters relevant for the GW inspiral and kilonovae signals.

Concerning the afterglow counterparts, it is in the population model that we draw from short GRB science. Under the BNS merger-short GRB-relativistic jet we hypothesize, the energy available for BNS merger afterglow should be defined by the prompt energy dissipated in short GRBs, modulo a γ-ray dissipation efficiency and a typical duration of prompt dissipation. We suppose that short GRBs have an average γ-ray efficiency of $\langle f_\gamma\rangle = 20\%$; Because of the scarcity of short GRB afterglows, this value is uncertain and, in any case, very degenerate with the other parameters of luminosity function inference, such as the slopes or minimal luminosity. We will keep $\langle f_\gamma\rangle$ fixed, and consider the associated uncertainty to be held in the luminosity function slopes; A more detailed discussion on $\langle f_\gamma\rangle$ and its impact on our population study can be found in Sec. 5.5.6. The average short GRB duration in the instrument frame is $\langle T_{90,\rm obs}\rangle = 1$ s (Sec. 1.3), which we assume is $\langle T_{90}\rangle \sim 0.5$ s in the source frame.

Therefore, we distribute $E_{\rm iso,c}$ according to:

$$\phi(E_{\rm iso,c}) \propto \phi_{\rm sGRB}\left(\frac{E_{\rm iso,c}}{\langle T_{90}\rangle\left(\frac{1}{\langle f_\gamma\rangle}-1\right)}\right) \tag{4.21}$$

with $\phi_{\rm sGRB}$ the inferred luminosity function of short GRBs (Eq. 4.1).

As announced, we will study the population deduced from both Ghirlanda et al. (2016)'s and Wanderman & Piran (2015)'s luminosity function, referred to as G16 and WP15 respectively in the sequel. In any case, we adopt fixed bounds for the energy of $E_{\rm min} = 10^{50}$ erg and $E_{\rm max} = 10^{53}$ erg, consistent with both studies. The break luminosity $L_* = 2.8\,10^{52}\,{\rm erg.s^{-1}}$ leads to $E_{\rm iso,c,*} = 2\,10^{52}$ erg.

We adopted $\theta_j = 0.1 \sim 6$ deg throughout, in line with measurements on GW170817 (Gill & Granot, 2018; Troja et al., 2019b; Mooley et al., 2018c; Troja et al., 2019a) and other short gamma-ray burst studies (Fong et al., 2015; Beniamini et al., 2019).

In the case of GW170817, afterglow fitting typically leads to a circum-merger density constraint of $n_{\rm ext} \sim 10^{-3}\,{\rm cm^{-3}}$, which is in agreement with the estimation of the HI content of the host galaxy NGC 4996 (Hallinan et al., 2017) and with X-ray constraints on the average density of the galaxy (Hajela et al., 2019). The external densities observed in short GRBs cover the interval 10^{-3}–1 ${\rm cm^{-3}}$ (Berger, 2014), but this is probably biased towards high densities, which favor afterglow detection. Here we will consider $n_{\rm ext} \sim 10^{-3}$ ${\rm cm^{-3}}$ as typical but still allow for larger densities by prescribing a log-normal of mean 10^{-3} ${\rm cm^{-3}}$ with a standard deviation of 0.75.

Throughout, we set the value of ϵ_e to 0.1. This value is generally adopted in GRB afterglow modeling (Wijers & Galama, 1999; Panaitescu & Kumar, 2000; Santana et al., 2014; Beniamini & van der Horst, 2017; D'Avanzo et al., 2018). More importantly, this value is consistent with the fitting of the afterglow of GW170817 (e.g., Troja et al., 2019a; Margutti et al., 2018; Nakar et al., 2018; Xie et al., 2018). It is consistent with microphysical particle shock-acceleration simulations, which find that ϵ_e saturates to this value for ultra-relativistic weakly magnetized shocks (Sironi & Spitkovsky, 2011; Sironi et al., 2013). We also mention the small scatter in ϵ_e in GRB outflow forward shocks suggested from the clustering of *Fermi*/LAT light curves (Nava et al., 2014).

Similarly, we set p to 2.2, in agreement with shock-acceleration theory in the relativistic regime (e.g., Sironi et al., 2015) and with the value inferred from the multi-wavelength analysis of the afterglow of GW170817 (Sec. 3.3).

There is more uncertainty on ϵ_B, which is the less constrained of GRB afterglow parameters. We assume a log-normal distribution centered on 10^{-3} with a scatter of 0.75, and will later on discuss possible other choices.

Finally, we note that we will identify the binary's inclination angle—ι, defined by the system's total angular momentum—and its viewing angle–θ_v, defined by the relativistic jet's axis. While it is clear that the jet's axis is aligned with the merged object's angular momentum, the latter may be different from the binary's original momentum. This can be due to spin-orbit misalignment in the progenitor system or to unequal dissipation of spin and orbital momentum components in the merger process. Spin-orbit misalignment in BNS merger has yet to be probed by GW observing runs.

In Tab. 4.4, we summarize all the afterglow-related parameters and their distributions assumed in our population model. The kilonova related parameters can be found in Tab. 4.2.

4.6 Conclusion

In this chapter, we exhibited a multi-messenger population model for BNS mergers. This model draws from state-of-the-art modeling for the relativistic jet afterglow and kilonova signals. It is astrophysically motivated in its parameter distributions

Table 4.4: **Parameters and associated distributions for the electromagnetic emission models in our population model.**

Definition	Notation	Distribution
Angle between the line of sight and the core jet's axis	θ_v or ι	$\propto \sin\theta_v$
Luminosity distance to the source	D_L	$\propto D_L^2$ up to $\overline{H}$ or $d^r_{0.2\%}$
Number density of the homogeneous circum-merger medium	$n_{\rm ext}$	Log-normal distribution with mean $n_0 = 10^{-3}$ cm^{-3} and standard deviation 0.75
Half the opening angle of the core jet	θ_j	Fixed to 0.1 rad
Initial Lorentz factor of the core jet (useful when using Eq. 4.11)	Γ_0	Fixed to 100
Isotropic-equivalent pre-afterglow energy in the core jet	$E_{\rm iso,c}$	Broken power law (Eq. 4.1) from $E_{\rm min} = 10^{50}$ to $E_{\rm max} = 10^{53}$ erg, with a break at $E_{\rm iso,c,*} = 2\times10^{52}$ erg. The slopes are either $\alpha_1 = 0.53$ and $\alpha_2 = 3.4$ (G16) or $\alpha_1 = 1.9$ and $\alpha_2 = 3$ (WP15)
Spectral index of the population of non-thermal electrons in the forward-shocked material	p	Fixed to 2.2
Fraction of forward shock internal energy carried by the non-thermal electrons	ϵ_e	Fixed to 0.1
Fraction of forward shock internal energy carried by magnetic field	ϵ_B	Same log-normal as $n_{\rm ext}$
Temporal slope of the increasing phase of the afterglow light curve when modeled as a broken power law	α_+	Fixed to 0.8
Temporal slope of the decreasing phase of the afterglow light curve when modeled as a broken power law	α_-	Fixed to -2.2

and thus apt to make realistic predictions for future GW observing runs.

We will describe such predictions in Chap. 5. This population model will serve as a tool for our study of the viability of BNS merger afterglows for multi-messenger cosmology (Chap. 6) and as tools to probe the environments of BNS mergers (Chap. 7).

Chapter 5

Multi-messenger population prospects for the gravitational-wave era

Abstract

Based on the population model described in Chap. 4, we give population prospects for multi-messenger binary neutron star merger events. We start by replacing GW170817 in its population context, underlining how lucky we were to detect this rich event. We then give detection rates and describe the population of mergers with different electromagnetic counterpart associations. We also study the binary neutron star candidate event GW190425, and show how our population model allows to constrain its viewing angle from the fact that no kilonova was detected for this event. Finally, we discuss the sensitivity of these results to the population prescriptions and the main practical limitations in actually detecting the sample we expect to be in reach of follow-up instruments.

5.1 Introduction

Equipped with the population model of Chap. 4, we will now make population prospects for multi-messenger events in the GW era. In this chapter we will use the simplified GW criterion and proper displacement models (Secs. 4.4.2, 4.4.7). The GW horizons we adopt can be found in Tab. 4.1.

We start by replacing GW170817 in its population context, underlining how rare such a rich multi-messenger event is (Sec.5.2). We will then move to prospects, studying first the kilonova counterpart alone (Sec. 5.3), then the afterglow counterpart disregarding the requirement of the kilonova to localize the event (Sec. 5.5), then finally joint kilonova-afterglow events (Sec. 5.6). In the considerations linked to the afterglow, we shall consider the parameter distributions described in Tab. 4.4, with the short GRB luminosity function $\phi_{\rm sGRB}$ inferred by G16 (recall Sec. 4.2 for the definition of G16). Then, in Sec. 5.5.6 we will vary these distributions to discuss the impact of the energy distribution, of the jet opening angle and finally of the circum-merger density on the rates of multi-messenger events and the correlations within the population. Finally, we will summarize and discuss our results in Sec. 5.7.

5.2 How lucky were we to detect GW170817 during O2?

We recall that under the simplified GW detection criterion, there is a distance $D_0 = \overline{H}/\sqrt{8}$ such that for $D_L \leq D_0$, all sources are detected, while for $D_0 < D_L \leq \overline{H}$, they are progressively lost until, for $\overline{H}$, only those pointing directly to the observer remain.

The distribution of distances to the GW-triggered events was derived in Appendix A, it is reproduced in Fig. 5.1, showing a maximum at $D_L/\overline{H} = 0.63$. At D_0, the differential distribution of distances for the GW triggers transitions from $\propto D_L^2$ to a non-quadratic form, producing the small peak seen in Fig. 5.1. This, of course, is only a consequence of our simplified GW sky-averaged detection criterion. Near the horizon, the maximum viewing angle allowing GW-detection is $\theta_{\rm max,GW} \propto \sqrt{1 - D_L/\overline{H}}$ (Eq. A.10), which produces a strict decrease of event density. The distribution in viewing angle of these sources is represented in Fig. 5.1. It peaks at $\theta_v \sim 30\,$deg, with an average value $\langle\theta_v\rangle = 38\,$deg.

Had GW170817 occurred during the O3 run, with $\overline{H} = 157\,$Mpc, the distance to the source would have verified $D_{170817} \sim 40\,$Mpc $< D_0$. Therefore any merger at this distance would have been detectable, regardless of the inclination angle. In this case, the expected rate of binary neutron star

Table 5.1: **Percentage fraction of kilonovae associated with GW triggers in three magnitude intervals:** $m < 18$, $18 < m < 20$ and $m > 20$. Values correspond to the observing runs O3, O4 and O5.

Mag. range	< 18			$18 - 20$			> 20		
Band	O3	O4	O5	O3	O4	O5	O3	O4	O5
g	2.5	0.81	< 0.1	24	11	1.4	74	88	99
r	4.1	1.3	0.15	39	19	2.3	57	80	98
i	5.6	1.8	0.21	48	25	3.1	46	73	97
z	8.9	2.9	0.33	65	38	4.9	26	59	95

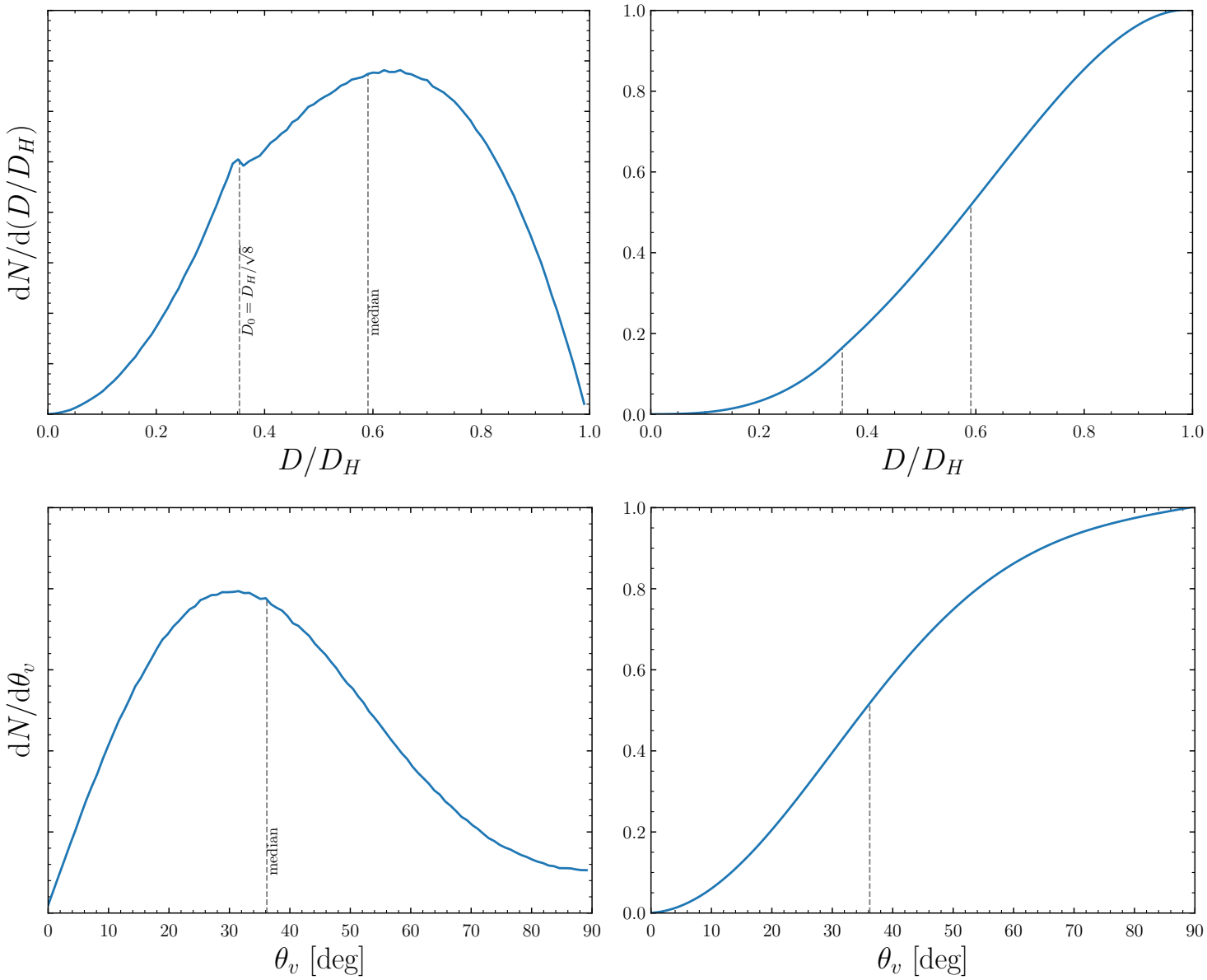


Figure 5.1: **Distance and viewing angle distributions for GW triggers.** Differential (left) and cumulative (right) distributions in distance (top) and viewing angle (bottom) of the GW triggers. We also indicate the median values and D_0, distance under which no selection in viewing angle occurs. The cusp at D_0 in the differential distribution in distance is non-physical and a consequence of the simplified nature of the adopted GW detection criterion. In this figure we used the notation D_H for $\overline{H}$. [Mochkovitch et al. 2021]

mergers up to the distance of GW170817 is simply given by

$$R = \tau_{\rm BNS} \times \frac{4\pi}{3} {D_{170817}}^3 \quad (5.1)$$

where $\tau_{\rm BNS} = 320^{+490}_{-240}$ Gpc^{-3}yr^{-1} is the local binary neutron star merger rate (Abbott et al., 2021a). This leads to an average rate of 1 event every $R^{-1} = 12^{+36}_{-7}$ yr.

GW170817 not only was a nearby event but had a low inclination angle, $\theta_v^{170817} < \theta_v^{170817,\rm max} \sim 18$ deg according to the VLBI observations (Sec. 3.3) The detection of the radio afterglow and source proper motion were possible only up to a viewing angle of $\theta_v^{\rm AG,max} \sim 40$ deg (see Fig. 5.12 below). Requiring $\theta_v \leq \theta_v^{\rm AG,max}$ to get a rich multi-messenger dataset with inspiral signal, kilonova and afterglow photometry and imagery data therefore leads to a rate of approximately

$$R' = R \times (1 - \cos\theta_v^{\rm AG,max}) \quad (5.2)$$

i.e., an average rate of one event every ${R'}^{-1} = 50^{+149}_{-31}$ yr.

The detection of the short GRB may require an even smaller viewing angle, $\theta_v \leq \theta_v^{\rm GRB,max}$ with $\theta_v^{\rm GRB,max} \simeq \theta_v^{170817,\rm max}$, as GRB170817A was detected only at the $\sim 5\sigma$ level by the Gamma-Ray Burst Monitor aboard *Fermi* (Goldstein et al., 2017). Requiring $\theta_v \leq \theta_v^{\rm GRB,max}$ to get a full GW170817-like multi-messenger dataset, including the short GRB, leads to an even lower rate $R'' = R \times (1 - \cos\theta_v^{170817,\rm max})$, i.e., one event every ${R''}^{-1} = 239^{+713}_{-146}$ yr.

During O2 ($\overline{H} \sim 86$ Mpc, Abbott et al. 2020a), where D_{170817} was not smaller than $D_0 \sim 30$ Mpc, these rates were even less. These numbers illustrate how exceptionally lucky we were to detect GW170817 so early and how long we may have to wait to observe another equivalent event.

5.3 Prospects for kilonova counterparts

5.3.1 Apparent magnitude

From the known distance and viewing angle distributions and our adopted parametrization for the magnitude (Eq. 4.6) we can readily obtain the distribution of apparent magnitudes for kilonovae associated with GW detections. It is shown in Fig. 5.2a for the g, r, i and z bands for the O4 observing run. If AT 2017gfo was in fact brighter than average, all the curves have to be shifted by the corresponding difference $\delta{\rm mag} = \langle m \rangle - m_{170817}$. Changing the GW horizon implies an interplay between maximum detection distance for the kilonova and the GW, and thus does not result in a simple shifting of the magnitude distribution.

Indeed, the number of kilonovae brighter than a limiting magnitude $m_{\rm lim}$ is proportional to:

$$\int_0^{\pi/2} {\rm d}\theta_v \, \sin\theta_v \times \min\left(D_L(m_{\rm lim},\theta_v), D_{\rm max,GW}(\theta_v)\right)^3 \quad (5.3)$$

where $D_L(m_{\rm lim},\theta_v)$ is such that $m_{\rm lim} = M_{\lambda,\theta_v} - 5\log\frac{D_L(m_{\rm lim},\theta_v)}{10\,{\rm pc}}$ and $D_{\rm max,GW}$ is the maximum distance for a GW detection at viewing angle θ_v (Eq. A.11). As the horizon changes, Eq. 5.3 evolves in a non-trivial manner. However, we found that, to a good approximation, changing from O4 to O5, the magnitude distribution is shifted by a magnitude of about $5\,{\rm Log}(\overline{H}_{O5}/\overline{H}_{O4}) = 1.6$.

The distribution of kilonovae in different magnitude ranges are summarized in Tab. 5.1 for three GW sensitivity hypotheses: O3, O4, and O5. It can be seen that there are very few kilonovae with $m < 18$ in all cases beyond O3. We note that recalibrating Eq. 4.6 assuming AT 2017gfo was one magnitude brighter than average leads to dividing the expected fractions in the < 20 magnitude ranges for O4 by 3, approximately.

The kilonova magnitudes depend on the different merger ejectas and their physical conditions. The blue kilonova component is likely linked to neutrino- or magneto-hydrodynamical-viscosity-driven winds from the transient remnant product and accretion disk around the latter (Gill et al., 2019, and references therein).

It is possible that in some systems, blue-enhancing ejection episodes are less effective—e.g., because of a short-lived merger remnant—leading to a lack of a blue kilonova component. We briefly consider this possibility, without seeking to know the fraction of these cases in the population.

If a fraction $f_{\rm red}$ of the kilonovae lack the blue component, a simple approximation consists in stating that such kilonovae will be dimmer, and thus transferred from the two brightest magnitude groups into the $m > 20$ group. This leads to the following for all bands:

$$\begin{array}{lcl} f_{<18} & \sim & f^0_{<18} \times (1 - f_{\rm red}) \\ f_{18-20} & \sim & f^0_{18-20} \times (1 - f_{\rm red}) \\ f_{>20} & \sim & f^0_{>20} \times (1 - f_{\rm red}) + f_{\rm red} \end{array} \quad (5.4)$$

where the f^0 fractions are those listed in Tab. 5.1.

We tested these approximations with our kilonova population model by emulating the absence of the blue component in a fraction $f_{\rm red}$ of the synthetic kilonovae. We did this by adopting the $\theta_v > \theta_0$ case of Eq. 4.6 for all viewing angles, as if only the red component were present. As the blue component affects more the g, r and i bands, these expressions represent reasonable approximations of the exact results while in the z band they somewhat overestimate the number of events which change from the < 20 to the > 20 magnitude groups.

The expected rates of kilonovae brighter than a given limiting r magnitude are shown in Fig. 5.2b for O3, O4 and O5 normalized to a GW neutron star coalescence detection rate of $\tau_{\rm BNS,GW} = 10\,{\rm yr}^{-1}$ for O4 (Abbott et al., 2020a). At the bright end of the distribution (i.e., $r < 19$), a fit to Fig. 5.2b shows that the rate approximately follows:

$$\log\frac{\tau_{\rm KN}}{{\rm yr}^{-1}} = 0.60 \times r_{\rm lim} - 11.6 + \log(1 - f_{\rm red}) \quad (5.5)$$

where $r_{\rm lim}$ is the limiting magnitude in the r band.

As an illustration, with $f_{\rm red} = 0$ (resp. $f_{\rm red} = 0.2$), we expect one kilonova brighter than $r = 19$ every 1.6 years (resp. 2.0 years) and one brighter than $r = 18$ every 6.3 years (resp.

7.9 years), independently of any future improvement of the sensitivity of GW detectors. On the other hand, the rate of kilonovae detectable by a follow-up with a limiting magnitude of $r_{\rm lim} = 21$ is increased by a factor ~ 3 between O3 and O4.

In Fig. 5.2b, we also show the maximum r-band magnitude of any kilonova associated with a GW trigger for O3, O4 and O5, denoted by $r_{\rm max}$. These magnitudes are the search depths required to recover 100% of the kilonovae. Because our peak magnitude dependence with viewing angle saturates at $\theta_0 = 60\,$deg, these maximum-magnitude events have $\theta_v = \theta_0$ and are placed at the largest distance to which the GW signal can be detected at this angle, i.e., at $D_L/\overline{H} = \sqrt{(1 + 6\cos^2\theta_0 + \cos^4\theta_0)/8} \sim 0.55$.

5.3.2 Distribution in viewing angle for different limiting magnitudes

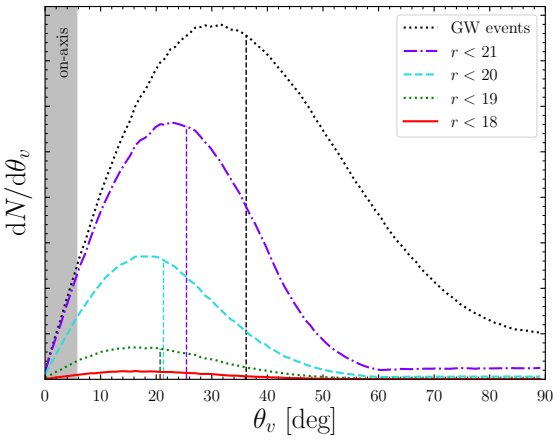


Figure 5.3: **Predicted distribution in viewing angle for detectable kilonovae associated with GW detections,** during O4 and for limiting r-band magnitudes 21, 20, 19 and 18. The vertical dashed lines represent the median. [Mochkovitch et al. 2021]

The distribution in viewing angle of the kilonovae associated with binary neutron star merger triggers that are brighter than a given limiting magnitude is shown in Fig. 5.3 for O4. As the limiting magnitude decreases, the median kilonova viewing angle—close to 36 deg in the entire population of GW triggers—significantly decreases: 26 deg for an r-band limiting magnitudes of 21, 21 deg for all $r_{\rm lim}$ smaller than 20.

In Fig. 5.4 we study the distributions in distance and viewing angles for events detected in the gravitational or optical domains. The green shade represents the density of GW triggers with a detectable kilonova counterpart; the blue shade GW triggers without a kilonova counterpart; and the red shade the "orphan kilonovae", detectable in the optical but not triggered by a GW signal. For this figure only, we removed the intrinsic variability of kilonovae introduced in Eq. 4.6 (i.e. we set δM_λ to 0) in order to clearly separate the different observing scenarios in the distance-viewing angle plane.

For limiting r-band magnitudes equal to or smaller than 20, practically all kilonovae that can be detected will follow a GW event if the interferometers are taking data at the corresponding time. Conversely, for deeper searches reaching magnitude 21 or 22, the fraction of "orphan kilonovae" without a GW alert increases and becomes dominant.

Recently, an archival study was carried out searching for kilonovae in 23 months of ZTF data. Down to a limiting magnitude of $r_{\rm lim} \sim 20.5$ for source detection, no kilonova was identified (Andreoni et al., 2020). Considering that a kilonova can be safely detected and characterized only if its peak magnitude is at least one magnitude brighter than the limit of the survey, Fig. 5.4 allows us to estimate the number of expected kilonova detections over the 23 month period. Assuming perfect identification and a sky coverage of $\sim$ 50% as appropriate for ZTF, we find between 0.4 and 2.6 detections, taking into account the uncertainty on the BNS merger rate but not that linked to the kilonova model.

Beyond the kilonova model uncertainties, an overestimated rate of BNS mergers or the limitations of the kilonova identification algorithm as discussed in Andreoni et al. 2020 could also contribute to the non detection. Future surveys and archival studies by other optical facilities (Almualla et al., 2021; Setzer et al., 2019) should clarify which of these options is the most likely.

As discussed previously, the question of the short GRB-compact object merger connection remains open after GW170817. The detection of a bright short GRB seen on-axis following GW from a binary neutron star merger would represent a direct evidence for this connection. Fig. 5.4 allows to discuss the probability of such events in the future. Assuming our typical jet opening angle $\theta_j = 0.1\,$rad, Fig. 5.4 clearly indicates that a triple association of GW, kilonova and bright short GRB seen on-axis should remain especially rare: one event every 5–20 years in the whole sky according to our calculations. This conclusion corroborates those of early joint GW–short GRB association studies (Sec. 4.2).

The association of a bright short GRB with a kilonova even without a GW detection is also a solid argument in favor of the merger connection. Fig. 5.4 shows that the rate of such double association is more optimistic if the limit magnitude in r band is at least 21, with $\sim$ 2 such events per year. However, for such bright-GRB associations, it has been noted that the optical kilonova signal should only outshine the afterglow flux in dense circum-merger media and with less energetic jets, allowing for an early-breaking or dimmer afterglow (Guessoum et al., 2018).

GRB130603B and GRB050709 were well in the parameter region allowing for the kilonova to appear (Fong et al., 2015), and still the associated claimed kilonova components (Tanvir et al., 2013; Jin et al., 2016) were only marginally brighter than the afterglow and required a follow-up duration larger than a week to be detected. Still, the potential of such sources to study BNS merger physics and the larger distances to which these can be detected encourage deep photometric follow-up of short GRBs.

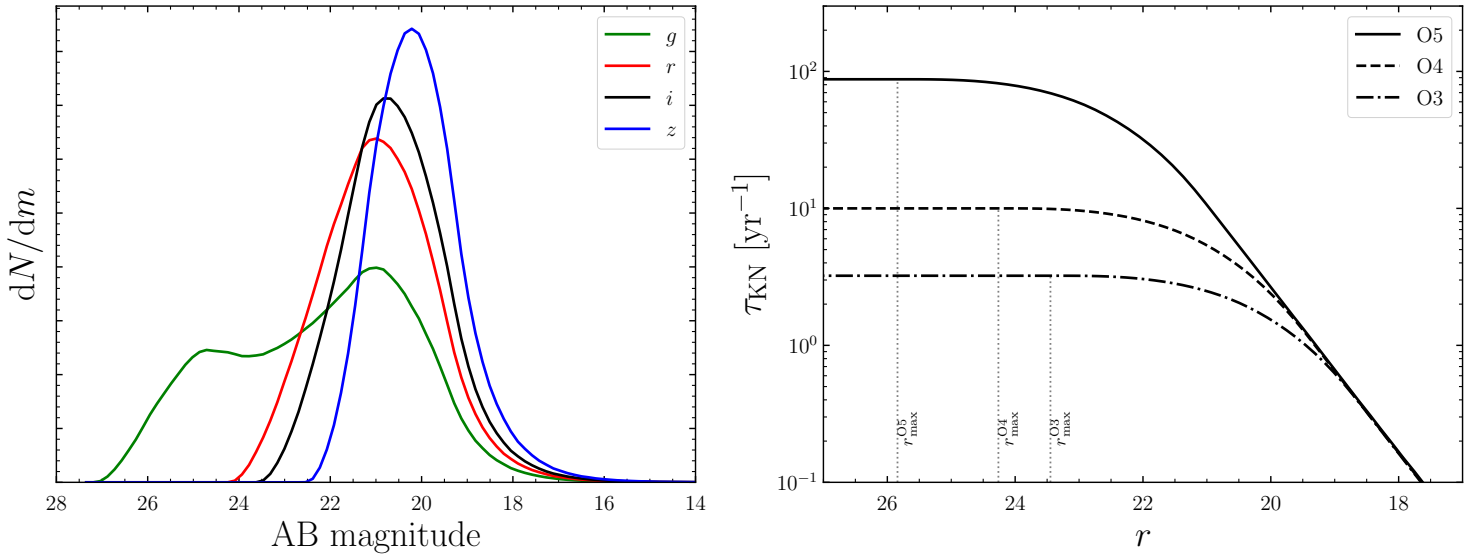


Figure 5.2: **Kilonovae magnitudes from GW triggers.** Left: Distribution of the peak magnitude in the g, r, i and z bands predicted for kilonovae associated with GW triggers during O4. Right: Rate of kilonovae brighter than a given r magnitude associated with GW detections during O3, O4 and O5, assuming a GW neutron star coalescence detection rate of $\tau_{\rm BNS,GW} = 10\,{\rm yr}^{-1}$ for O4 (Abbott et al., 2020a). The bright end of the distribution ($r < 19$) is well fitted by Eq. 5.5. [Mochkovitch et al. 2021]

5.4 What can our population model teach us about GW190425?

The GW trigger GW190425, the only confirmed binary neutron star merger during the LVKC observing run O3, was located at 159^{+69}_{-71} Mpc (Abbott et al., 2020b). No kilonova was found during the follow-up by several facilities. The deepest searches were led by ZTF to magnitude 21 in the g and r bands, covering 21% of the probability enclosed in the very large final GW skymap of nearly 7500 deg^2 (Kasliwal et al., 2019; Coughlin et al., 2019b), and by Pan-STARRS to mag 21.5 in the i band, covering 28% of the initial GW skymap (Smith et al., 2019; Coughlin et al., 2020c).

This non detection can have different origins, the most obvious one being simply that the kilonova was not located in the searched areas. But it is also interesting to explore the possibility that the kilonova was there but below the detection limit. Fig. 5.5 illustrates the resulting constraints in viewing angle-magnitude diagrams, adopting Eq. 4.6 for the kilonova magnitude. For this particularly massive event, we increased the horizon accordingly with the larger chirp mass of GW190425 with respect to a 1.4+1.4 $M_\odot$ system (see Tab. 4.1).

The viewing angle to GW190425 is bounded below by the non-detection of the kilonova, and bounded above by the detection of the GW inspiral signal. This results in the individual constraints from the g, r and i bands that can be read off Fig. 5.5. The strongest constraint comes from the i band, in which the kilonova is expected to be above typical optical follow-up thresholds for the largest viewing angle range, according to our model. The combined three-band constraint is $\theta_v^{190425} = 53.3^{+9.8}_{-12.4}$ deg, to which a systematic uncertainty due to the kilonova model should be added, see Sec. 5.7. Finally, in the case where no blue kilonova was produced in that event—possibly because the central core of the merged object directly collapsed to a black hole—no useful constraint can likely be obtained. This last possibility is indeed worth considering because of the large masses of the two neutron stars inferred for GW190425, suggesting its kilonova could have been dimmer or faster-decaying (Nicholl et al., 2021).

5.5 Prospects for radio afterglow counterparts

In this section, we describe the population of events detectable jointly in the GW and radio domains, independently of their kilonova counterpart. This allows us to focus on the afterglow-only parameters. In Sec. 5.6 we will study a more realistic scenario by requiring the kilonova to be detected before accessing the afterglow. We will start by considering the G16 short GRB luminosity function, and study the impact of the WP15 hypothesis in Sec. 5.5.6; We recall that this ingredient of the population model is the one that holds the most uncertainty.

We will refer to a jointly GW- and radio afterglow-detectable population of mergers using the names of the corresponding LVKC observing run and limiting radio facility, as in "O3–VLA combination". The corresponding radio sensitivities and GW horizons can be found in Tabs. 4.3 and 4.1.

5.5.1 Rate of joint GW and radio detectable events

Starting from the fraction $f_{\rm GW} = N_{\rm GW}/N \sim 29\%$ of binary systems detected by the LVKC (Sec. 4.4.2), we then estimated

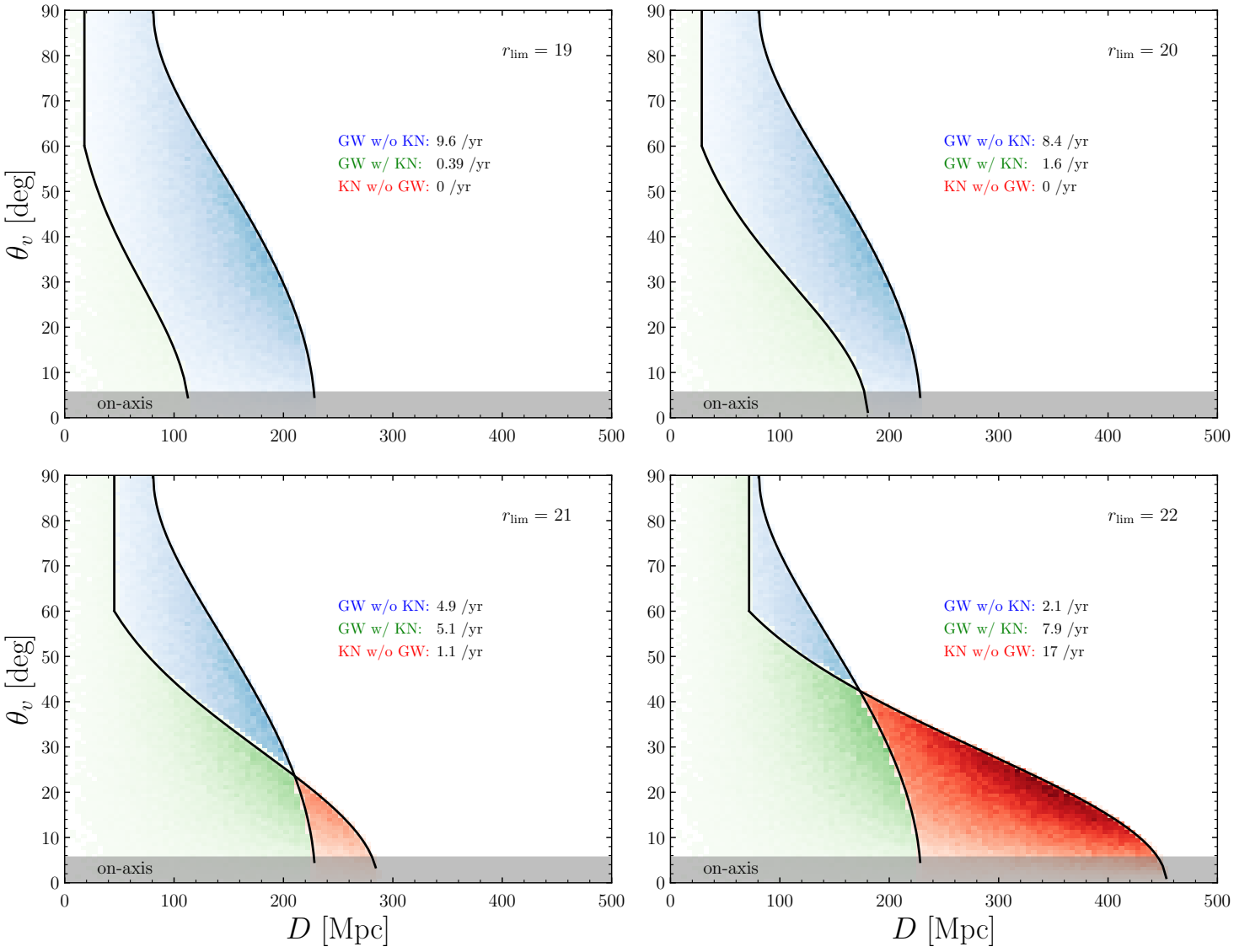


Figure 5.4: **Relative occurrence rates of signals in the distance-viewing angle plane predicted for the O4 run.** Colors indicate different detection scenarios: events detectable (i) both as GW triggers and kilonovae (green), (ii) as GW inspiral signals alone (blue), (iii) only as kilonovae ("orphan kilonovae", red). For the kilonova detection, the four diagrams correspond to r-band limiting magnitudes from 19 to 22. We also indicate the total occurrence rates in each detection scenario, assuming a GW detection rate of 10 yr^{-1} for O4. [Mochkovitch et al. 2021]

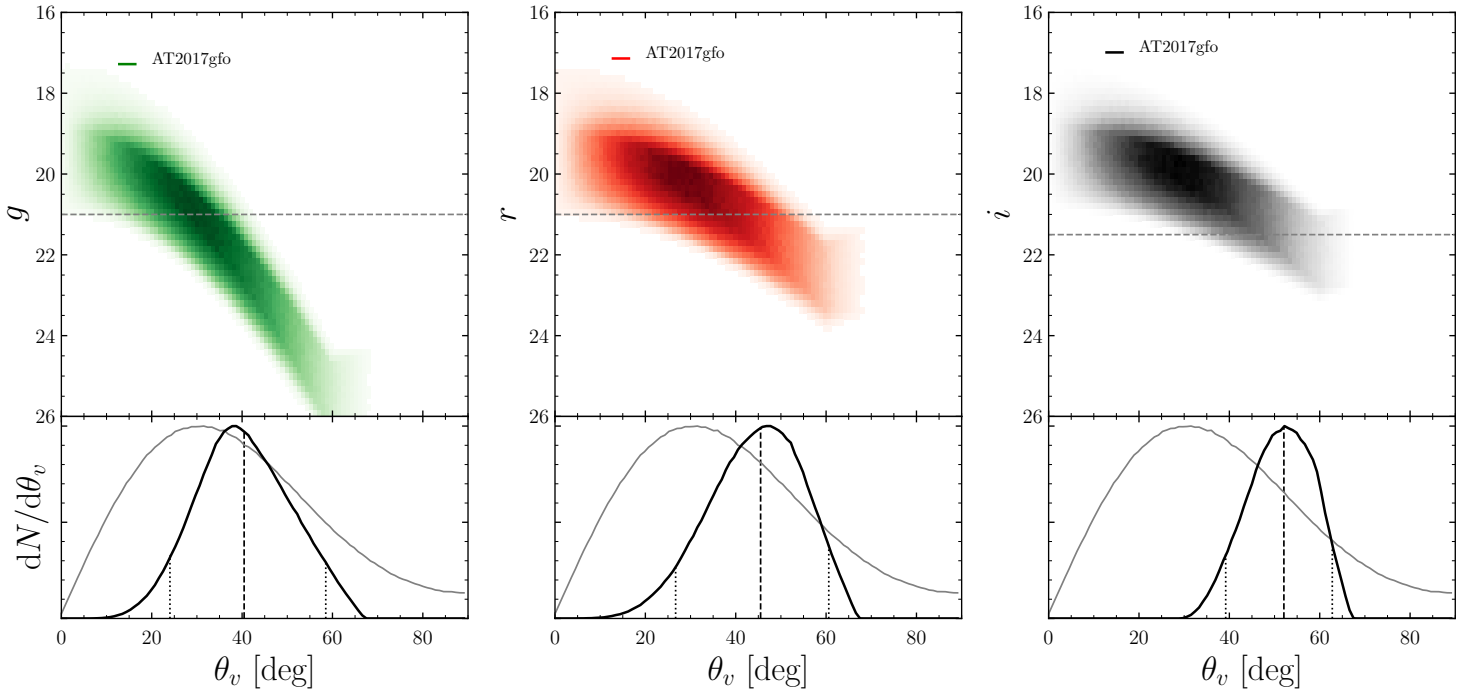


Figure 5.5: **Multi-messenger constraints on the viewing angle to GW190425.** Top: Viewing angle-magnitude diagrams in the g, r and i bands for events with distances consistent with GW190425 during O3. The horizontal lines are limiting magnitudes for ZTF (g and r bands) and Pan-STARRS (i band). Bottom: Constraints on the viewing angle to GW190425 assuming is was below the detection limits (black), compared to the viewing angle distribution of all GW triggers (gray). [Mochkovitch et al. 2021]

which fraction $N_{\rm joint}/N_{\rm GW}$ of those events would also produce a detectable radio afterglow.

This fraction is shown in Fig. 5.6 (left) as a function of the GW horizon $H = \overline{H}\sqrt{5/2}$ and for the VLA, SKA1 and SKA2/ngVLA limiting sensitivities (report to Tab. 4.3 for the corresponding sensitivities). In the design configuration of the interferometer network, 75% (resp. 2 times) more events are detectable by SKA1 (resp. SKA2/ngVLA) than by the VLA.

The number of joint event detections normalized to the O3–VLA combination is represented in Fig. 5.6 (right). The number of joint detections behaves approximately as H^γ, with $\gamma < 3$ because of the reduction in radio detection efficiency when the distance is increased (Fig. 5.6, left). We find $\gamma \simeq 2.4$ (resp. 2.6, resp. 2.8) for the VLA (resp. SKA1, resp. SKA2/ngVLA).

As illustrated in Fig. 5.6 (left), we find that the fraction of detected events decreases from 43% (O2) to 31% (O3) with the VLA limiting flux. However the absolute number of detections increases (Fig. 5.6, right) due to the rapid recession of the GW horizon. With the design–SKA1 combination, a fraction of 35% of GW events can lead to a radio detection. With regard to the much better radio sensitivities that may be reached after 2030, more than a half of GW events may become detectable in the radio band. Indeed we found a percentage of 63% of detectable joint events in the design–SKA2/ngVLA combination.

For the purposes of a comprehensive survey, the number of GW and jointly detected events per continuous year of GW network operation for future detector configurations can be found in Tab. 5.2, where we have used the G16 luminosity function hypothesis.

5.5.2 Distance and viewing angle

The distributions of distances are shown in Fig. 5.7 for the O2–VLA and O3–VLA combination joint GW-radio sources, as well as for the entire GW population. In this figure, the distributions are not normalized to unity so as to show the decrease in the fraction of joint event as the horizon increases.

It can be observed that as a result of the GW and radio detection thresholds, sources are progressively lost when the distance increases so that this distribution exhibits a near linear increase (as opposed to the $\mathrm{d}N/\mathrm{d}D_L \propto D_L^2$ of the intrinsic homogeneous population).

Fig. 5.8 shows the distribution of the viewing angles. For the joint events, the peak of the distribution takes place at small viewing angles $\theta_v \sim 15 - 20$ deg as a result of the rapid decline of the peak flux with angle ($F_p \propto \theta_v^{-2p}$), and about 39% (resp. 47%) of the events would be seen with a viewing angle $\theta_v < 20$ deg for O2–VLA (resp. O3–VLA), the angle at which the GW170817 event was likely to have been seen. As shown in Fig. 5.9 (left), the mean viewing angle of the jointly observed events strongly depends on the GW horizon and the radio sensitivity. It appears that even if the radio sensitivity is not substantially improved while the GW interferometers reach their design horizons, a significant number of events would still

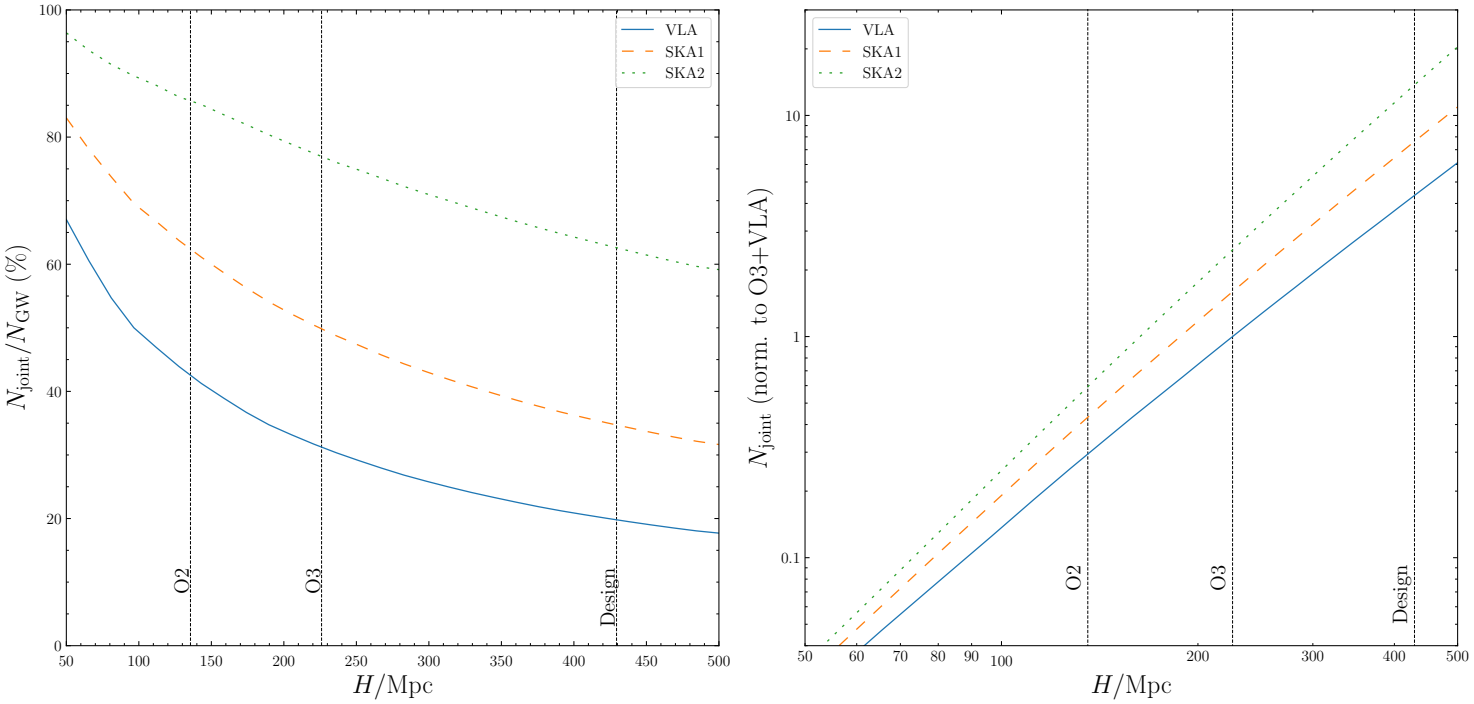


Figure 5.6: **Radio-detectable population of GW triggers.** Left: Detectable fraction of radio afterglows among gravitational wave events as a function of the horizon distance $H = \overline{H}\sqrt{5/2}$. Right: Expected number of joint detections normalized to case of O3–VLA combination. In both panels the full (resp. dashed, resp. dotted) lines correspond to the VLA (resp. SKA1, resp. SKA2/ngVLA) being the limiting radio facility. [Duque et al. 2019]

Table 5.2: **Number of jointly detectable events per continuous year of GW network operation.** The expected rate of total GW events is taken from Abbott et al. (2020a). The uncertainties here are due to the uncertainty on the merger rate in the local Universe derived from GW observations of the LVC O1 and O2 runs. The additional uncertainties related to the population model are discussed in Sec. 5.5.6. Note: Our population model determines the fraction of events detectable in the radio among all GW triggers. Here we updated the number of detectable events with the latest version of the GW detection prospects (Abbott et al., 2020a). The figures that changed with respect to those published in Duque et al. (2019) are marked with an asterisk (*).

LVC Run	Radio configuration		GW Events	Joint Events	Fraction of detectable events
	Instrument	$s_{3\rm GHz}$ [μJy]	$N_{\rm GW}$	$N_{\rm joint}$	(assuming G16 luminosity function)
O3	VLA	15	9^{+19}_{-7}	3^{+6}_{-2}	31.4%
Design	VLA	15	62^{+130}_{-48} *	12^{+26}_{-9} *	19.8%
Design	SKA1	3	62^{+130}_{-48} *	22^{+45}_{-17} *	34.7%
Design	SKA2/ngVLA	0.3	62^{+130}_{-48} *	39^{+81}_{-30} *	62.5%

be observable off-axis. The increasing and decreasing phases of the afterglows of these events would allow for an improved study of the jet structure, which is better revealed by off-axis events (Sec. 3.3). Fig. 5.9 (right) shows the fraction of on-axis events ($\theta_v \leq \theta_j$) within the joint detections. It increases from 4.0% (O2–VLA) to 5.3% (O3–VLA).

5.5.3 Radio afterglow peak times, peak fluxes and synchrotron spectral regime

The distributions of peak times and peak fluxes in the O3–VLA combination are represented in Fig. 5.11. The fraction of events with a peak time earlier than 150 days (approximately as observed in GRB170817A) is 55% without jet expansion and 81% with lateral expansion. The distribution in peak flux is shown for all sources which are detected in gravitational waves within the sky-position-averaged horizon of O2, O3 and design instruments. It appears clear that for the present VLA sensitivity, most radio afterglows cannot be detected. This explains why improving the sensitivity of the future ngVLA would have such an impact on the joint detection rates, as shown in Fig. 5.6.

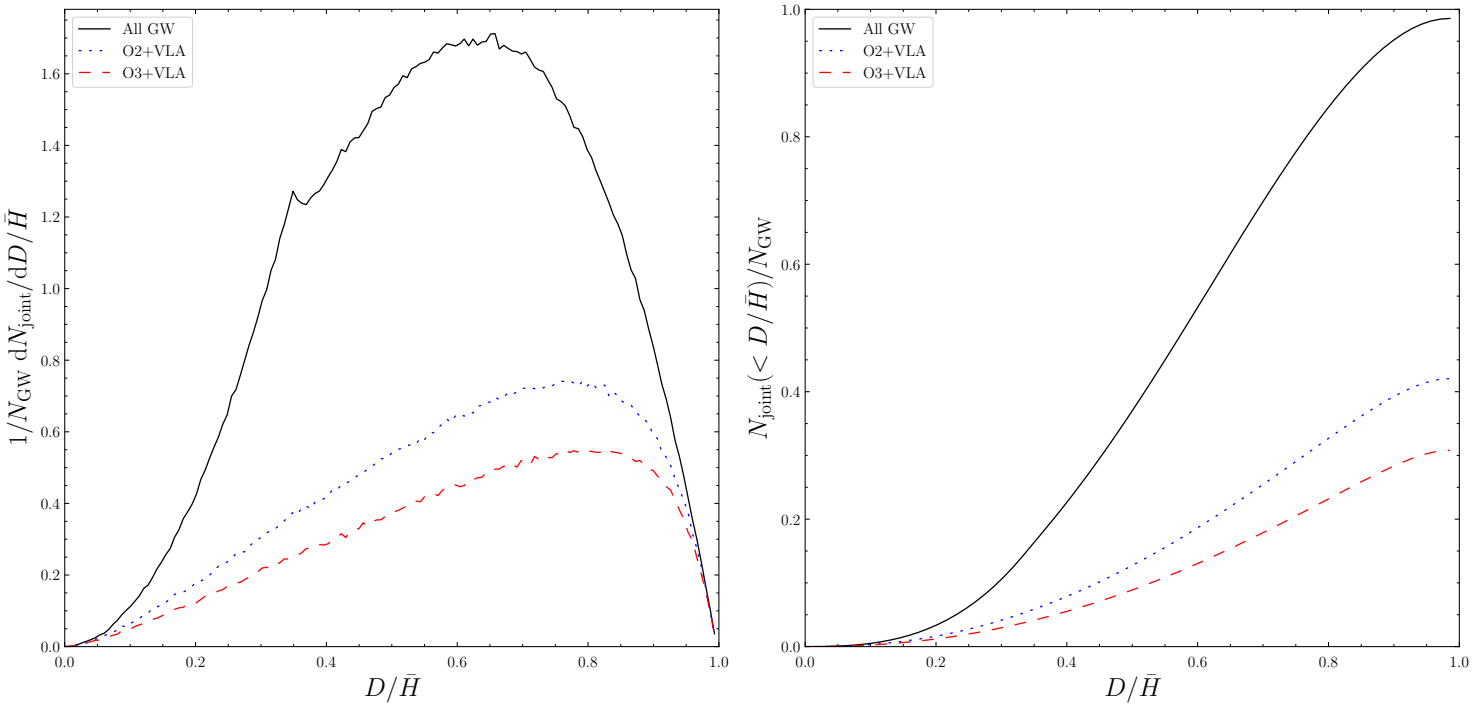


Figure 5.7: **Distance distribution to radio-detectable GW triggers.** Left: Differential distribution of distances (normalized to $\overline{H}$) of events in the total GW population (full line), and detectable jointly in O2–VLA (dotted line) and O3–VLA (dashed line) combinations. Right: Cumulative distribution of same events. [Duque et al. 2019]

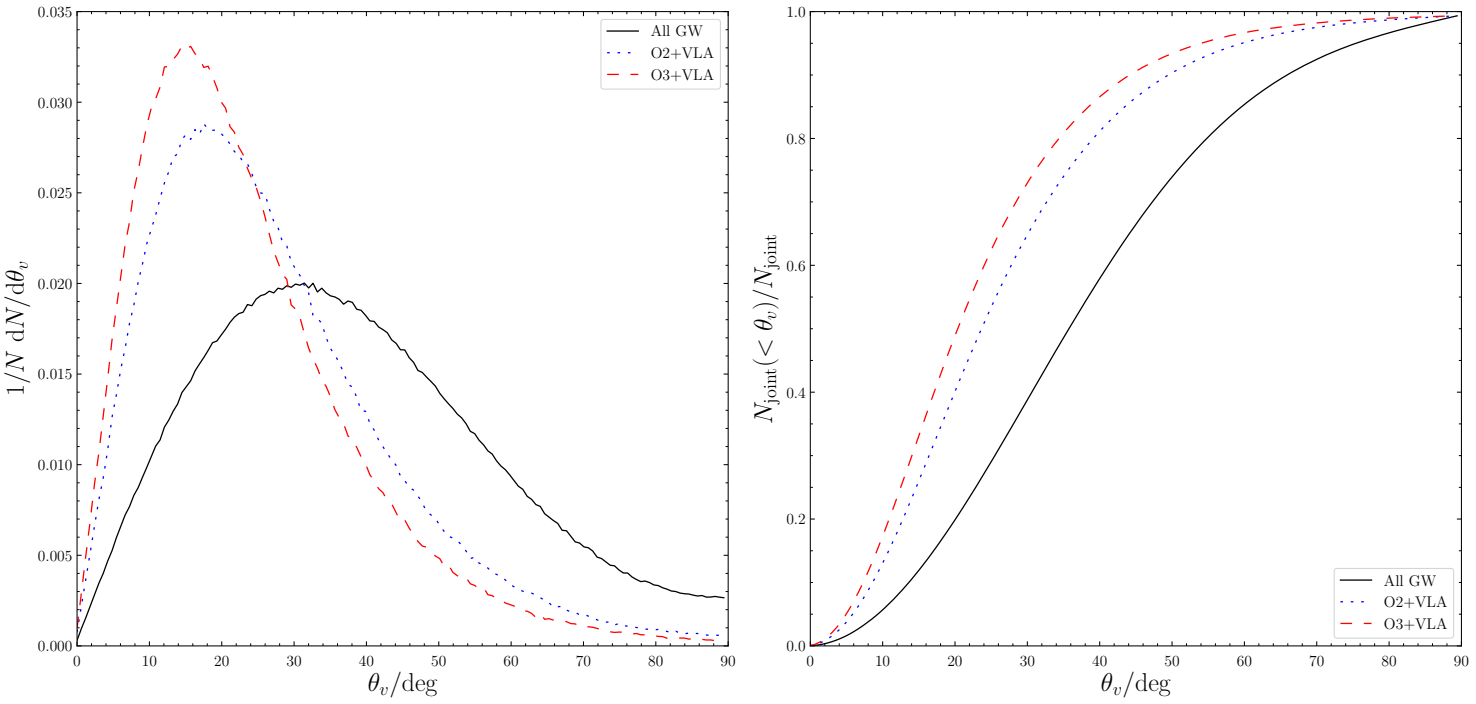


Figure 5.8: **Viewing angle distribution of radio-detectable GW triggers.** Same as Fig. 5.7, for viewing angle. Distributions are normalized. [Duque et al. 2019]

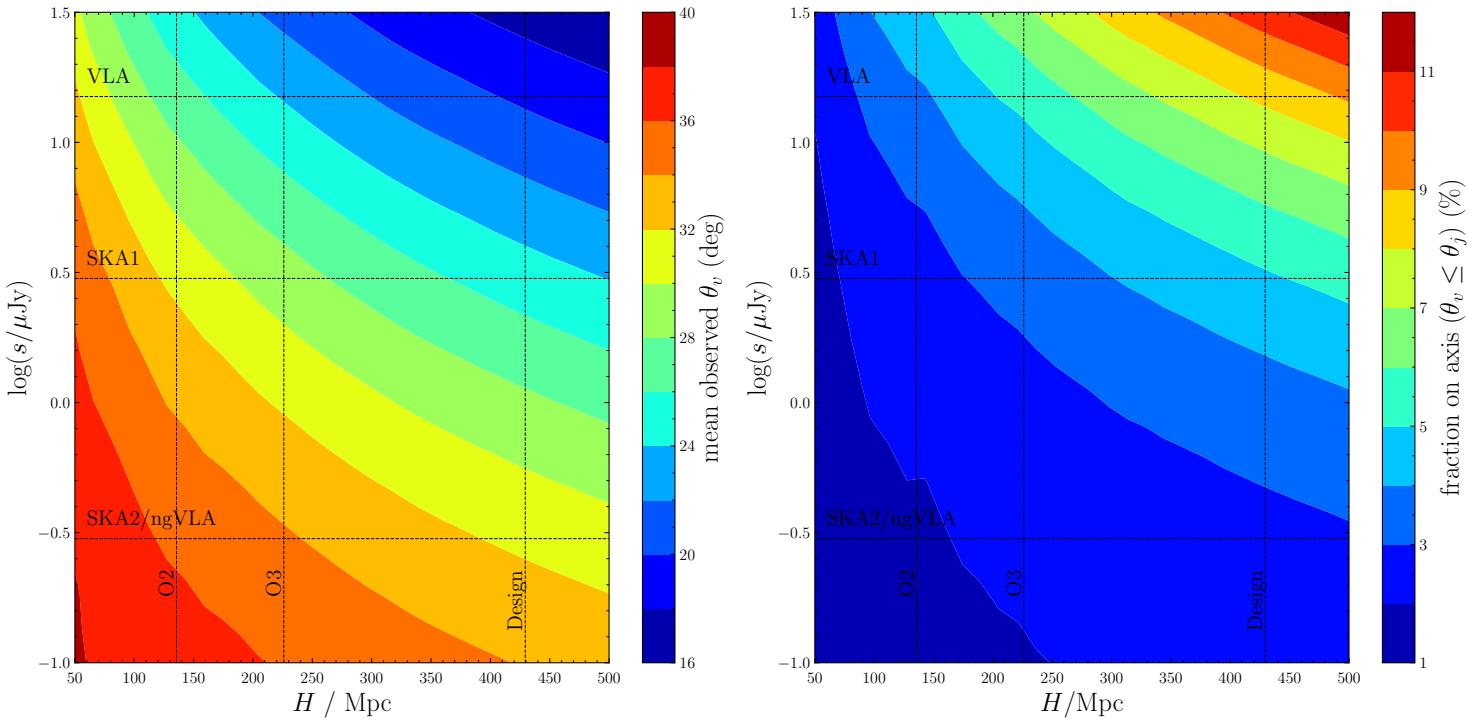


Figure 5.9: **Viewing angle statistics of radio-detectable GW triggers.** Left: Mean viewing angle of jointly detected events, as function of horizon and radio limiting flux $s_{3\mathrm{GHz}}$. Values of H for O2, O3 and design instruments, as well as $s_{3\mathrm{GHz}}$ for VLA, SKA1 and SKA2/ngVLA, are indicated as dashed lines. Right: Fraction of on-axis events (defined by $\theta_v \leq \theta_j$) among jointly detected events in same H–$s_{3\mathrm{GHz}}$ diagram. [Duque et al. 2019]

The scaling law used to compute the radio emission at the peak assumes that the observing frequency remains in the same spectral regime of the slow-cooling synchrotron spectrum, that is, $\nu_m < \nu < \nu_c$, and above the absorption frequency. Using our more detailed calculation of the radio afterglow from the core jet (Eq. 4.11 and accompanying text), we can ascertain if this condition is fulfilled for the bulk of the population.

In Fig. 5.10 we plot the distributions of the observer-frame synchrotron characteristic frequencies and the Compton parameter at the peak of the afterglow in the jointly detectable O3–VLA population.

We find that the GHz radio band is always well below ν_c and above ν_m. For off-axis events (the bulk of the population, see Fig. 5.8), it is well above ν_m. For the rare nearly on-axis events, the peak of the afterglow occurs precisely when the ν_m frequency passes the radio band, and thus $\nu_{\mathrm{obs}} = \nu_m$. We could check that the peak in ν_m distribution present in Fig. 5.10 were precisely those events with $\theta_v < 2\theta_j$. Therefore, the bulk of the population is in the announced synchrotron regime at afterglow peak.

However, we find that the GHz band is only marginally in the non-absorbed regime, with an average absorption frequency of $\nu_a \sim 10$ GHz at afterglow peak. Similarly, Fig. 5.10 shows the inverse Compton scattering is only marginally negligible at afterglow peak, with a median value of $Y = 2.5$ in the O3–VLA population, and 95% of sources having $Y < 10$. Studying the correlations of parameters in the detectable population shows that larger absorption frequencies are strongly correlated with higher densities, and larger Y values are correlated with larger ϵ_B, as expected from the definitions of ν_a and Y (Appendices E.4 and E.5). The effects of absorption and inverse Compton scattering can therefore be considered as further uncertainty on ϵ_B and n_{ext} in the population. Once again, the goal of this chapter is an overall description of the expected population of mergers and to reveal trends within this population. Furthermore, as shown below, the dominant uncertainty in our model is the distribution of jet energies, which has a larger impact than the distribution of ϵ_B and n_{ext}. We can therefore continue to consider the population as unaffected by self-absorption and inverse Compton scattering.

For mergers in high density environments (for n_{ext} larger than 10 cm^{-3}), self-absorption will affect the flux more and the injection frequency ν_m may be larger than the radio frequency at times earlier than the peak. For an event with $n_{\mathrm{ext}} = 10$ cm^{-3} and $\theta_v \sim 40$ deg (the mean viewing angle of the GW-detected population), the radio frequency typically meets the injection break at around 30 days (resp. before 10 days). In this case, chromatic light curves are expected with different behaviors in different observing bands.

5.5.4 Source proper motion

The distribution of μ_{max} is shown in Fig. 5.11 (right) for O2, O3 and design-level detectable GW events respecting the VLBI

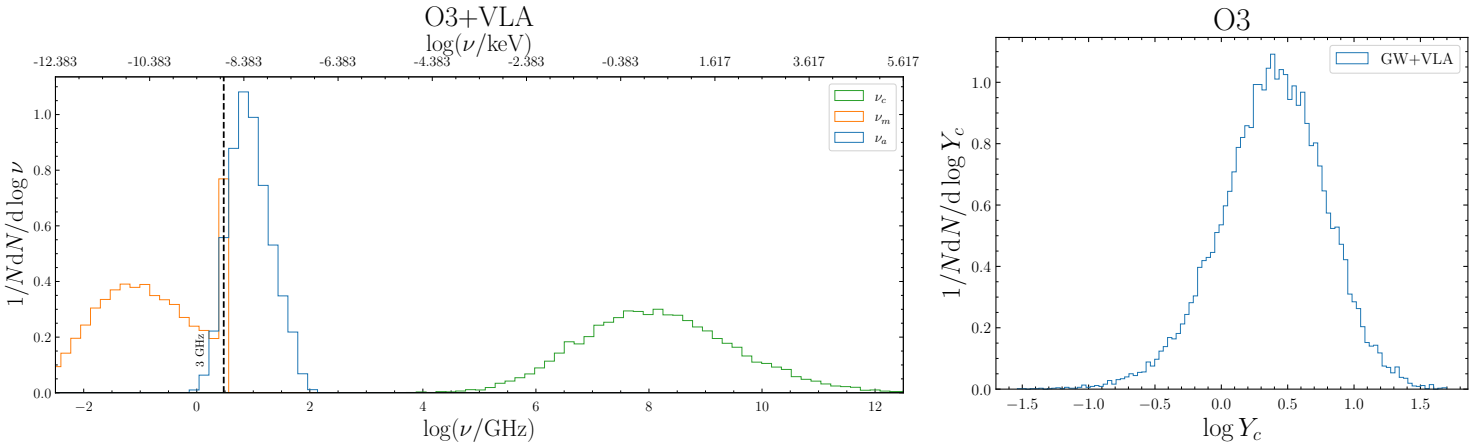


Figure 5.10: **Spectral statistics of radio-detectable GW triggers.** Left: Distribution of the observer-frame synchrotron characteristic frequencies ν_m and ν_c and the self-absorption frequency ν_a, at the peak of the afterglow. The O3–VLA combination was considered. Right: Distribution of the Compton Y parameter in the shocked material at the peak of the radio afterglow for the same population.

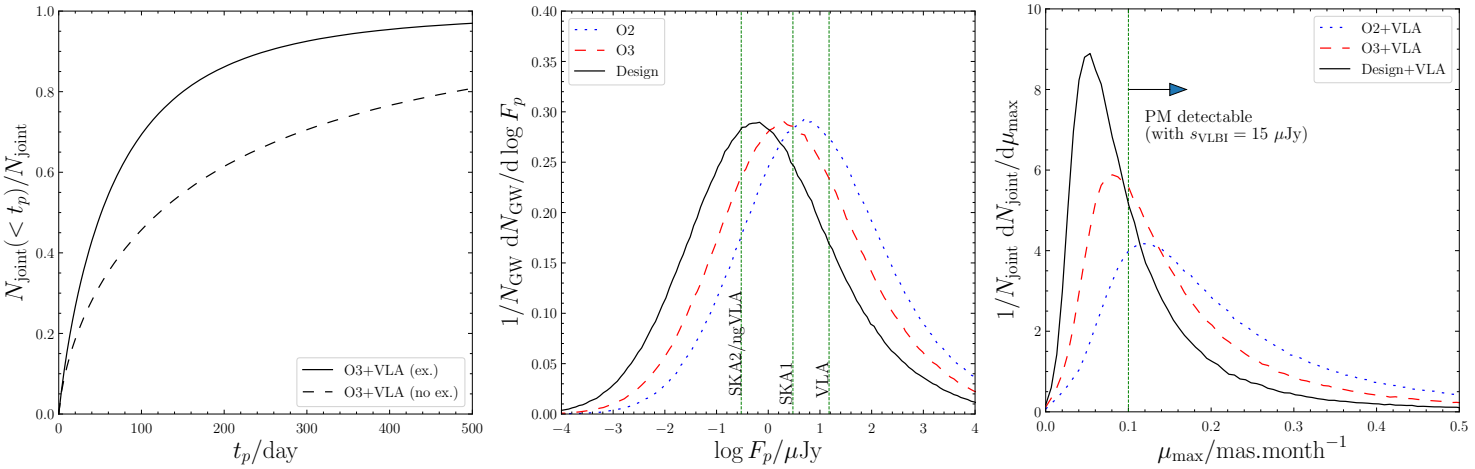


Figure 5.11: **Afterglow peak-time and proper motion statistics of GW triggers.** Left: Cumulative distributions of peak times of radio afterglows for jointly detected events in O3–VLA combination, assuming jet with (solid line) or without (dashed line) lateral expansion. Center: Differential distribution of peak fluxes of GW-detected events assuming horizon for O2 (dotted line), O3 (dashed line) and design (full line) configurations of GW interferometers. The radio sensitivities for VLA, SKA1 and SKA2/ngVLA are indicated for comparison. Right: Differential distribution of maximum proper motion for jointly detected events assuming O2–VLA (dotted line), O3–VLA (dashed line), design–SKA1 (full line) combinations. The dotted vertical line shows the lower limit for detection of the proper motion. [Duque et al. 2019]

flux criterion. Here, we have taken $s_{\rm VLBI} = 15$ μJy, which represents a limiting S/N of $\sim$ 3 with the level of noise reported in Mooley et al. (2018b); Ghirlanda et al. (2019) for the e-MERLIN and High Sensitivity Array networks. Within the O2–VLA (resp. O3–VLA, resp. design–VLA) populations, the measurement of the proper motion is possible in only 79% (resp. 64%, resp. 42%) of the jointly-detectable events.

Finally, we represent in Fig. 5.12 various plots connecting two observable quantities: (θ_v, D_L), $(\theta_v, F_{p,3\,\rm GHz})$, (θ_v, t_p) and $(D_L, \mu_{\rm max})$. The gray dots correspond to events detectable in GW and the red crosses correspond to those detectable in both GW and radio. In the (θ_v, D_L) diagram, the limiting distance for the measurement of the proper motion (Eq. 4.19) is shown. This figure illustrates the various observational biases discussed in this section, and especially the bias towards small viewing angles, mainly due to the relatively limited sensitivity of radio telescopes.

5.5.5 Kinetic energy, external density and microphysics parameters

The distributions of kinetic energy $E_{\rm iso,c}$ and external density $n_{\rm ext}$, which are not directly observable but may result from fits of the afterglow, are shown in Fig. 5.13. In each case, the distributions of jointly detected events have been compared to total GW detections. As the GW selection is independent of the kinetic energy and density, these distributions are the intrinsic distributions of our population model (Sec. 4.5).

As could be expected, the detection of mergers leading to a relativistic core jet with a large $E_{\rm iso,c}$ is favored. In jointly detectable mergers, the energy distribution approximately remains a broken power-law, but the respective indices below and above the break decrease from 0.53 to 0.2 and 3.4 to 3.2, respectively. After applying the joint detectability criterion, the external density distribution is simply shifted to higher densities, as the mean value is increased by a factor of ~ 2.2. The distribution of the microphysics parameter ϵ_B is very similar to that of the density, because both parameters appear with the same power in the peak flux (Eq. 4.13).

We find that the Lorentz factor of the jet at the time of peak flux has a median value of ~ 4. This is consistent with the approximate jet core debeaming relation $\Gamma \sim (\theta_v - \theta_j)^{-1}$ for a jet with $\theta_j = 0.1$ rad seen with the mean viewing angle $\theta_v \sim 24$ deg for the O3–VLA combination. This justifies keeping microphysics parameters representative of shock-acceleration in the ultra-relativistic regime for the whole population.

Table 5.3: **Fraction of joint events among GW events for different intrinsic density distribution central values,** for the O3–VLA detector combination, as calculated with the semi-analytical model (Eq. 4.11).

n_0 [cm^{-3}]	$N_{\rm joint}/N_{\rm GW}$
10^{-4}	22 %
10^{-3}	40 %
10^{-2}	63 %
10^{-1}	83 %
1	94 %
10^{1}	97 %

5.5.6 Impact of the population model uncertainties on the predicted population

Here we consider alternatives to our initial population model. Assuming the WP15 short GRB luminosity function, the fraction of events detectable with the VLA would typically be three times smaller: 15% for O2 and 10% for O3 compared to 43% and 31% under the G16 hypothesis. The viewing angle and peak times would also be strongly affected, showing mean θ_v's for O2 (resp. O3) at only 17 deg (resp. 14 deg), 98% (resp. 99%) of afterglows peaking before 150 days in the jet expansion hypothesis, and 90% (resp. 93%) without lateral expansion.

These results were obtained assuming a typical value $f_\gamma = 20\%$ for gamma-ray efficiency but one cannot exclude that f_γ may vary from from less than 1% to more than 20%. Estimates of f_γ from afterglow fitting have indeed been found to cover a large interval (e.g., Lloyd-Ronning & Zhang, 2004; Zhang et al., 2007). Thus, the three orders of magnitude in isotropic gamma-ray luminosity could partially result from differences in efficiency, the jet isotropic kinetic energy being restricted to a smaller interval. Fig. 5.14 shows the effect of increasing the minimum kinetic energy $E_{\rm min}$ from 10^{50} to 10^{52} erg in the energy function on the detected fraction, the mean viewing angle and the fraction of afterglows to peak before 150 days. As could be expected, when the minimum kinetic energy was increased, the results came closer to those of the G16 model which expects more energetic jets.

We also show in Fig. 5.14 the effect of changing the typical value of the jet opening angle from 0.1 to 0.05 or 0.15 rad. When $\theta_j = 0.05$ (resp. 0.15), the detected fraction decreases (resp. increases) to 19% (resp. 40%), the mean viewing angle decreases (resp. increases) by 4 deg, and the peak of the light curve generally occurs later (resp. earlier).

The afterglow peak flux depends greatly on the density of the circum-merger medium, as can be seen from Eq. 4.13. Therefore, the ability to detect the afterglows of a population of mergers depends strongly on the density of the media hosting the binaries upon merger. This is illustrated in Tab. 5.3, where we give this fraction for different external density distribution central values. These figures were calculated using the

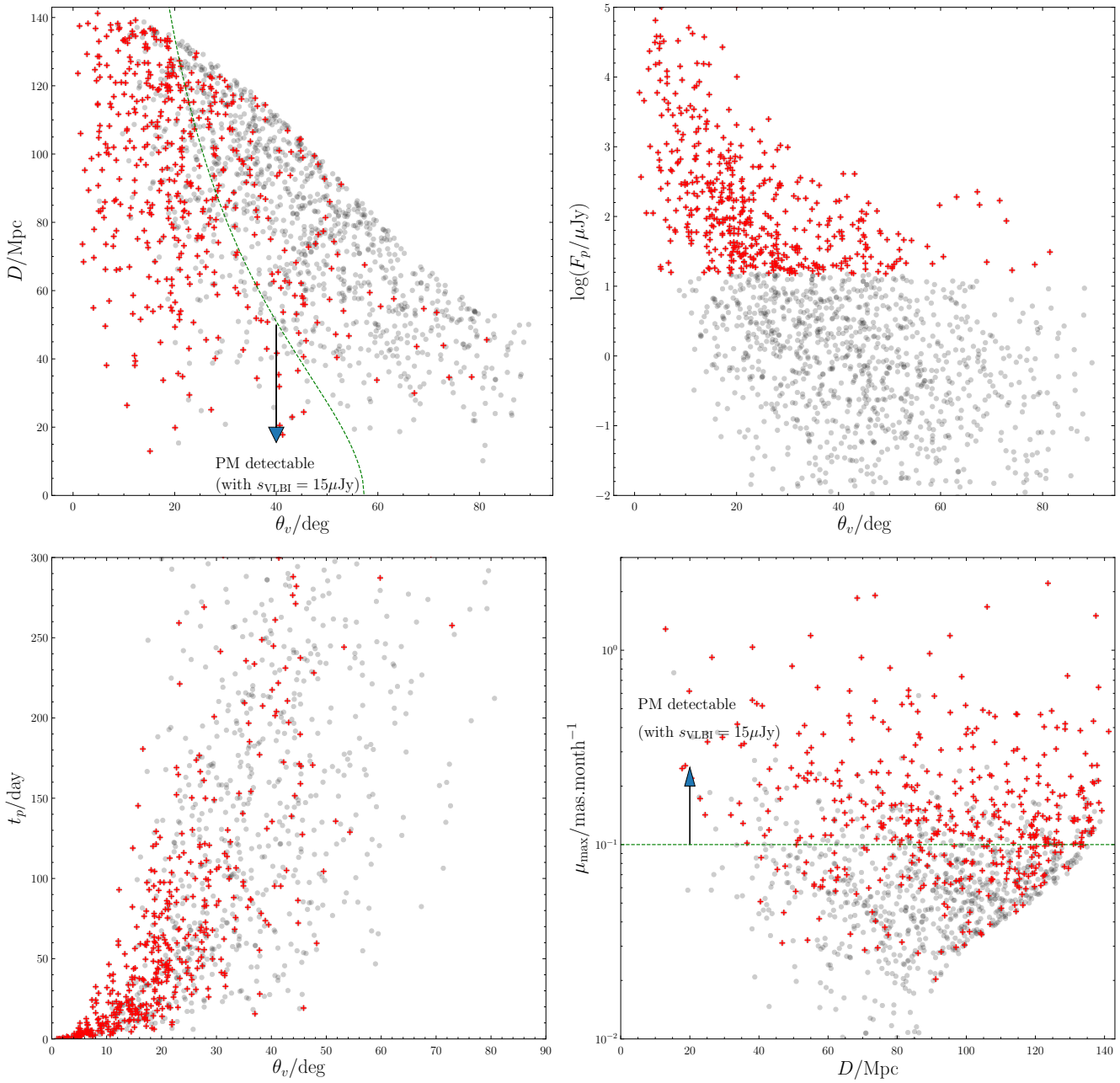


Figure 5.12: **Predicted samples of GW-detected and jointly detectable mergers in planes of various pairs of observables,** for O3–VLA combination. Grey: GW triggers; red: radio-detectable events. Top left: Distance and viewing angle, and maximum distance at which the proper motion can be measured ($D_{\rm max,VLBI}(\theta_v)$, see Eq. 4.19, green dotted line). Top right: Peak flux and viewing angle. Bottom left: Time of radio afterglow peak (assuming lateral expansion of jet) and viewing angle. Bottom right: Maximum remnant proper motion and distance, and limiting proper motion $\mu_{\rm lim}$ (green dotted line). [Duque et al. 2019]

semi-analytical model (Eq. 4.11), taking into account the full synchrotron spectrum including self-absorption.

If there is a significant population of fast-merging neutron star binaries, as proposed by Matteucci et al. (2014); Hotokezaka et al. (2015); Vangioni et al. (2016); Beniamini & Piran (2016) (see detailed discussion in Sec. 7.2), these should merge in higher density environments, producing brighter afterglows and offering the possibility to notably contribute to the observed population even if they are intrinsically fewer in number. With a density distribution centered at $n_{\rm ext} = 10$ cm^{-3}, and the other parameters taking their values from our initial population model, 99% of GW events produce detectable

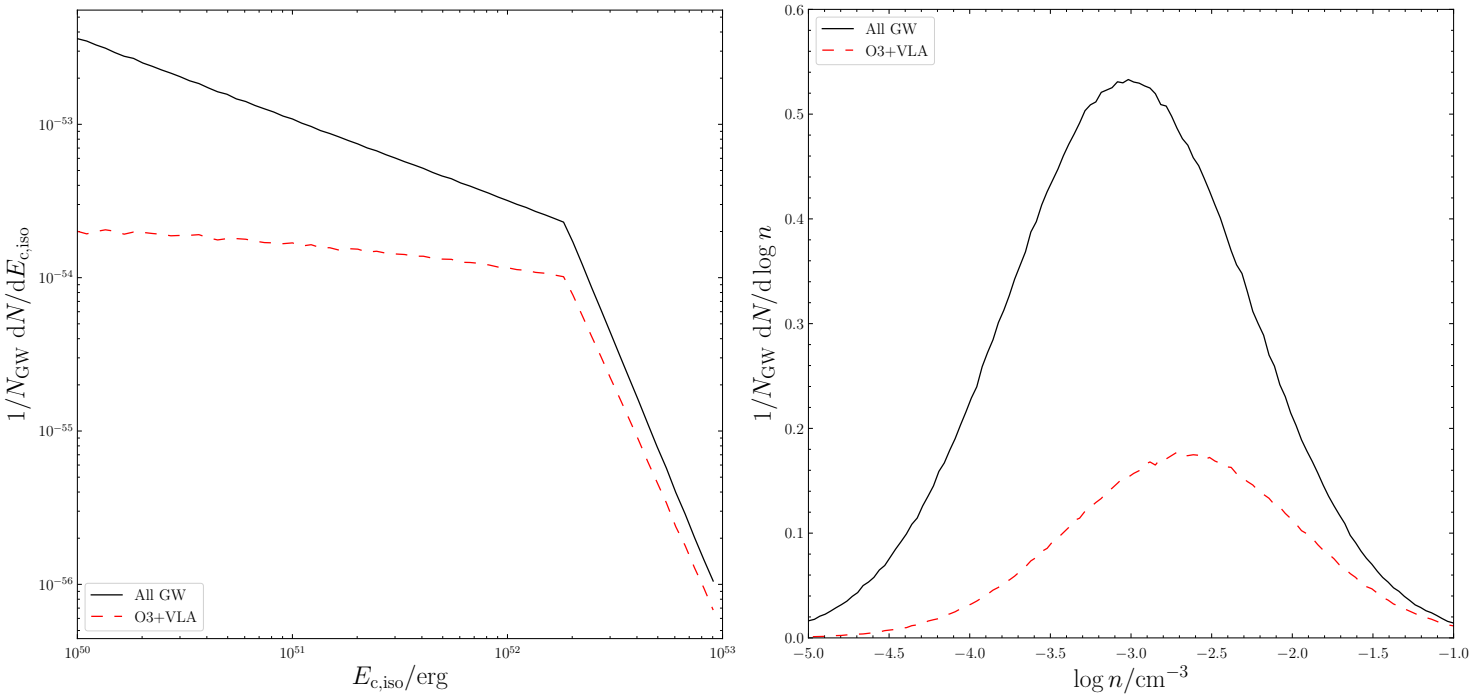


Figure 5.13: **Jet kinetic energy and circum-merger medium statistics of radio-detectable events.** Differential distribution of core jet kinetic energy (left) and external medium density (right) of entire GW population (full) and jointly detected events (dashed), in O3–VLA combination. Distributions are normalized to fraction of jointly detected events among all GW-detected events, see text for details. [Duque et al. 2019]

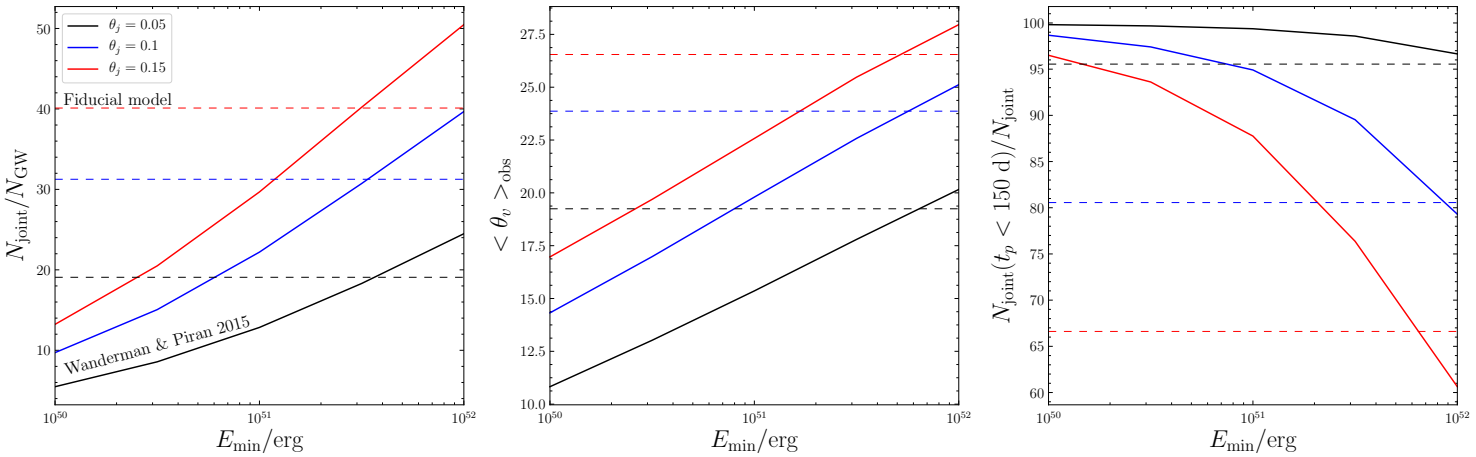


Figure 5.14: **Statistics of radio-detectable events assuming different jet kinetic energy distributions in the population of mergers.** Left: Fraction of joint events among GW events assuming the WP15 short GRB luminosity function and different values of minimum kinetic energy $E_{\rm min}$ (full lines), and for the G16 hypothesis and $E_{\rm min} = 10^{50}$ erg ("fiducial", dashed lines), in the O3–VLA combination. Results have been plotted for different values of the jet opening angle $\theta_j = 0.05$, 0.1 and 0.15 rad (red, black and blue lines respectively). Center: Mean viewing angle (same color coding). Right: Fraction of joint events with peaks earlier than 150 days post-merger, with hypothesis of jet lateral expansion (same color coding). [Duque et al. 2019]

radio afterglows in the O3–VLA combination.

Tab. 5.3 hints to an idea to probe high-density environments, based on the afterglows of BNS mergers. We will fully develop this multi-messenger method in Chap. 7, and present it as a new and independent approach to the question of BNSs with short delay times or small kicks.

As mentioned above, the microphysics parameter ϵ_B appears in the afterglow flux with the same power as the external density (Eq. 4.13). Moreover, our initial choices for the distribution of these two quantities are the same, such that changing its central value affects the detected fraction in an identical way.

5.6 Joint kilonova and afterglow counterparts

Up to now, we have considered the kilonova and afterglow counterparts independently. However, for GW170817 and in the near future, the afterglow counterpart should be detected only after monitoring the merger locus determined by the kilonova when it is found (Sec. 4.4.4). Therefore, we briefly study in this section the impact of requiring detection of both the kilonova and the afterglow on the kilonova- and afterglow-only population features we derived previously. We will compare the three-signal GW–kilonova–afterglow population with the design–VLA population described in Sec. 4.4.5 (on radio afterglows), because the design-level GW sensitivity in that section is close to the O4 sensitivity of Sec. 5.3 (on kilonovae).

Of the detectable kilonovae for O4, we select those detectable in the radio band using Eq. 4.14 and with $\varphi = 1$, thus generating afterglows close to those of our initial population model. In this case, the afterglow can be detected at the VLA 3 GHz limitinng flux of 15 μJy for 37%, 56% and 76% of kilonovae, assuming r-band limiting magnitudes of 21, 20 and 19 respectively for the kilonova search. This is 19%, 9% and $< 1\%$ of all the GW triggers. For all limiting magnitudes except the deep $R_{\rm lim} = 21$, this represents a significant decrease with respect to the 20% of detectable afterglows among all GW triggers reported in Tab. 5.2.

In order to construct an exploitable light curve, we can impose a radio flux threshold of three times the detection limit. In this case, these fractions become 12%, 6% and <1% of the total GW trigger population. In terms of absolute numbers, they respectively correspond to 1.2, 0.6 and < 0.1 joint GW-kilonova-afterglow detections per year. For particularly energetic jets or dense circum-merger environments, i.e. with the dimensionless factor $\varphi = 10$ (Eq. 4.14), the fractions of kilonovae with radio afterglows at three times the VLA limit rise to 59%, 81%, 97%, that is 2.9, 1.6 and 0.5 three-signal detections per year.

We have represented in Fig. 5.15 the distribution in viewing angle of the afterglows that can be detected with the VLA at three times the detection limit for $\varphi = 1$ and 10 and different kilonova search limiting magnitudes. Due to the very steep dependence of the afterglow flux with viewing angle ($F \propto \theta_v^{-4.4}$ for $p = 2.2$, see Eq. 4.14) the detection is possible at small viewing angles only. For $\varphi = 1$, we find a mean viewing angle of ~ 10 deg for the joint GW-kilonova-afterglow, significantly smaller than the ~ 20 deg found disregarding the kilonova (Fig. 5.9). It is only for $\varphi = 10$ that this regime of viewing angles is recovered.

5.7 Discussion

5.7.1 Summary

In this chapter, we studied the population of binary neutron star mergers to be observed through the GW inspiral signal and through the various expected electromagnetic counterparts: the kilonova, the jet afterglow and the short GRB. To do so, we have assumed a likely population of mergers inspired by the observations of short GRBs, we have made predictions on the rates of the different combinations of electromagnetic counterparts, and studied the impact of the uncertainties in the population prescriptions on our results.

Thanks to this population model, we were able to replace GW170817 in its population context, showing that such rich events are in fact very rare and that we were lucky to observe this event so early during the O2 run.

Concerning the jet afterglow, in the case of the O3 run of the LVKC, and assuming a top-heavy short GRB luminosity function (G16), we predict that $\sim 30\%$ of GW events should have a radio afterglow detectable by the VLA. These joint events should have mean viewing angles of ~ 24 deg, and should peak earlier than 150 days post-merger in $\sim 80\%$ of cases, assuming lateral expansion of the jet. The important novelty of this upcoming population of afterglows is that their jets should be viewed from *significantly misaligned* lines of sight, with $\theta_v \sim 5 \times \theta_j$. This strongly contrasts with the jets of bright GRB detected up to GW170817, which were likely observed with only *slightly misaligned* lines of sight, as we shall discuss in Chap. 8.

Concerning kilonovae, we obtained the rate of kilonovae brighter than a given limiting magnitude, see Eq. 5.5. The median of the distribution of viewing angles to detectable kilonovae—which is about 36 deg for the GW triggers—, decreases as the search becomes shallower, reaching 26 deg then 21 deg for r-band limiting magnitudes of 21 and lower than 20.

We also studied the prospects for "orphan kilonovae", not associated with GW triggers. For deep surveys reaching magnitude 21–22, the rate of such events becomes dominant, opening the way to detecting kilonovae counterparts to short GRBs, which should become much more common. The progress in understanding the merger phenomenon by leveraging such signals motivates the effort to carry out these surveys and optical follow-up of short GRBs.

Finally, we studied the case of detecting the radio afterglow after having secured the kilonova counterpart; This is the most probable multi-messenger scenario for detecting the afterglow. We showed that requiring both detections significantly decreases the fraction of events with afterglow counterparts with the respect to the afterglow-only criterion studied in Sec. 5.5.

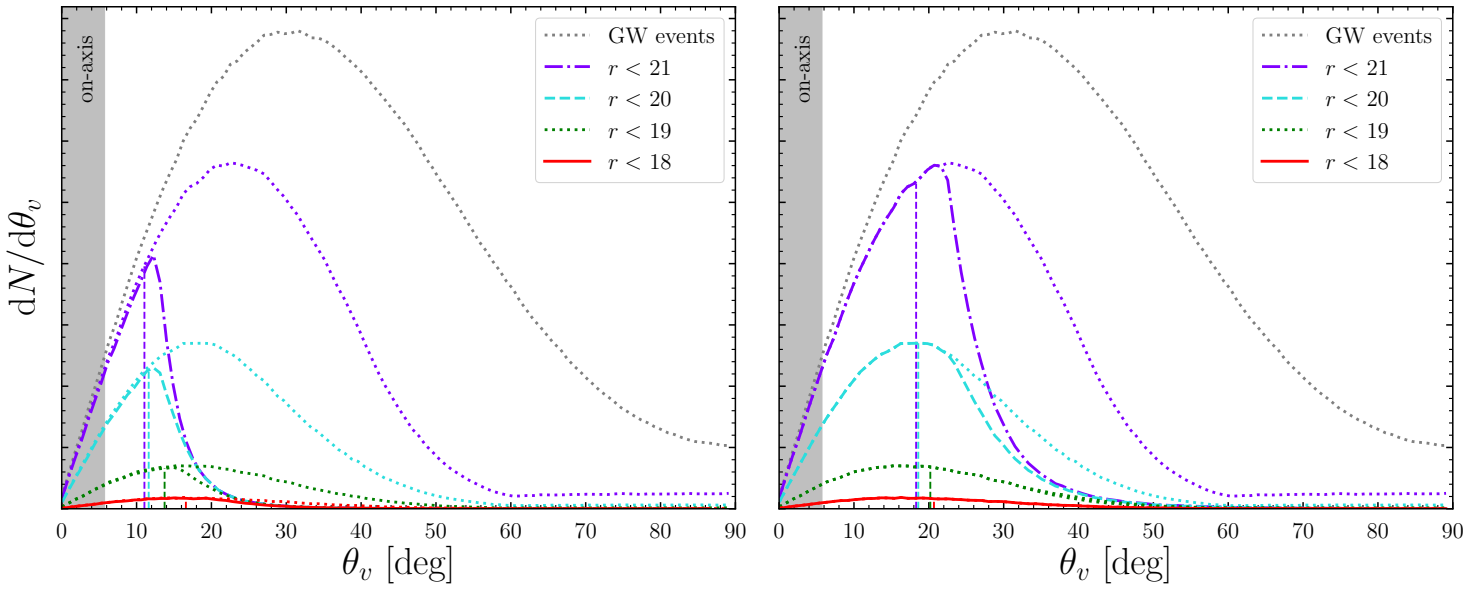


Figure 5.15: **Distributions in viewing angle of the afterglows detectable in the radio band at three times the VLA threshold** (45 μJy) following a GW-triggered kilonova detection with a limiting r-band magnitude of 18 to 21. The dotted lines represent the corresponding distributions for the kilonova sources, as in Fig. 5.3. In the left (resp. right) panel, a value $\varphi = 1$ (resp. $\varphi = 10$, corresponding to particularly energetic jets or dense circum-merger media) was adopted in Eq. 4.14. [Mochkovitch et al. 2021]

5.7.2 Detecting the detectable events

Throughout this chapter, we have applied a threshold on the GW sign-to-noise ratio, kilonova peak magnitude and radio afterglow peak flux to label an event as detected. Thus our study concerns the class of *detectable events*. The fraction of these detectable events that can actually be detected heavily depends on the capability to localize the kilonova. For small field-of-view instruments, this is limited by the large GW skymaps to cover. For large-field-of-view high-cadence instruments such as ZTF or the LSST, the limitation is rather the search magnitude and the position of the source in the sky available to the instrument. The highly unlikely case of a short GRB counterpart could also ease the pinpointing of the source in the sky.

During the O3 run, the follow-up force was largely small field-of-view telescopes, and the limitations due to the size of the GW skymaps showed drastically. Nonetheless, the advent of additional GW interferometers such as KAGRA (Kagra Collaboration et al., 2019) will largely improve the median gravitational-wave localization skymap down to ~ 40 deg^2 (Abbott et al., 2020a), paving the way to more effective follow-up by smaller field-of-view instruments and therefore better-sampled kilonova light curves. This will be complemented with the first light of the LSST, providing deep, high-cadence capabilities to the southern sky.

Moreover, there are other difficulties related to the search for the kilonova: kilonova-host galaxy contrast at large distances, possible large offsets, availability of a photometric and spectroscopic follow-up of the candidates, recognizing kilonovae among a myriad of optical transients, even with quality spectroscopic or color evolution observations. In this context, it is clear that the number of detected kilonovae will remain below those predicted in Sec. 5.3.

Even assuming the localization has been acquired thanks to the detection of a kilonova, a continuous monitoring of the remnant up to ~ 150 days may be necessary to detect the peak of the radio afterglow, in the case of marginally detectable events. Even then, only events with peaks somewhat larger than the radio threshold should yield observations of astrophysical interest because extended observations of the rise, peak and decay of the afterglow are necessary to resolve the structure and dynamics (e.g., expansion) of the ultra-relativistic jet, as illustrated by the case of GRB170817A.

As also illustrated in the case of GRB170817A, VLBI measurements are instrumental in assessing the presence and resolution of the jet structure. Therefore, an even more restrictive criterion for full event characterization might be the detectability of the remnant proper motion. As we demonstrated in Sec. 5.5.4, this would significantly decrease the fraction of events, especially in the context of design-level GW detectors.

Finally, future facilities such as the SKA may bring detections of orphan radio afterglows through deep radio surveys. These would not be subject to the GW criterion and may probe another sub-population of mergers, bringing yet another class of constraints upon these phenomena.

5.7.3 Leveraging BNS mergers with counterparts

As the GW horizon increases, more events will be seen on-axis. In the case of an opening angle of 0.1 rad and in the likely short-term evolution of the detector configuration from O4–VLA (current) to design–VLA (mid 2020s), we predict that 5% to 10% of events should be seen on-axis, increasing the probability of a GRB counterpart in addition to the GW and afterglow. This small figure is fully consistent with those found in early studies of GW–bright short GRB associations (Sec. 4.2. In fact, as a GW–short GRB association is equivalent to an aligned line of sight, one may measure the opening angle of typical GRB jets by considering the ratio of GW events with afterglow counterpart only to events with both afterglow and GRB, thus exploiting the new possibility of observing afterglows without detecting the GRB.

More generally, a population of observed events corresponds to an intrinsic population of mergers. Thus, a comparison of a statistical number of joint event observations to our predicted distributions is a means of measuring fundamental parameters of the population of mergers, and of constraining GRB quantities, such as their luminosity function.

Moreover, population models are useful to constrain individual events, as we showed for GW190425, the only confirmed binary neutron star merger event since GW170817. No associated kilonova was detected, which could simply be a consequence of the poor localization, which limited the search to less than 30% of the GW skymap. If, however, a kilonova was in the searched area but too weak to be detected, a constraint on the viewing angle can be obtained, we find that the viewing angle must have been 53 ± 10 deg, assuming there was an AT 2017gfo-like blue component. Because of the obvious possibility that the source was outside the regions covered by follow-up, our constraints on GW190425 should be seen as proof of concept. This method will reveal most useful in the case of genuine non-detections when the future, smaller GW skymaps will effectively be fully covered by follow-up. For future GW triggers, one could use the GW-measured progenitor properties (such as component masses and tidal deformability) to tailor the kilonova modeling to the specific events (Nicholl et al., 2021). In case of a non-detection, the viewing angle constraints thus obtained would be more robust because informed with the complete multi-messenger dataset.

Finally, as we will show in Chap. 6, kilonovae and even mild associated viewing angle measurements seem to be the only means for electromagnetic modeling to contribute to multi-messenger cosmology and the resolution of the Hubble tension. The effort to collect a kilonova sample and study their variability and viewing angle properties thus appears even more desirable in this regard.

5.7.4 Modeling uncertainty and consequences

We have studied the sensitivity of the expected distributions of observables to the population parameters, as well as the electromagnetic and GW detector configurations. In particular, we have shown the uncertainties on the rates and typical viewing angles of these events stemming from the uncertainty on the luminosity function of short GRBs and of the density of the media hosting the mergers.

For example, adopting the WP15 hypothesis for the short GRB luminosity function decreases the afterglow detectable fraction by a factor of 3 and decreases the expected viewing angle of events. We also discussed the combined influence of increasing the radio and GW detector sensitivities. Improvements in both sectors would be beneficial as the predicted rate in the current configuration is well below 50% of joint events among GW events. However, it is likely that the GW horizon will recede faster than the reach of radio arrays in the future. Therefore, it is expected that the evolution of the detection configuration will favor smaller viewing angles for radio afterglow counterparts.

In our kilonova model, there is double uncertainty: some linked to the polar-to-equatorial view contrast (ΔM_λ in Eq. 4.6), and some linked to the calibration of the polar magnitude (M_λ). The former was fit onto theoretical expectations from sophisticated modeling, and the latter from calibration on GW170817. Both aspects of this uncertainty should improve in the coming years, with the detection and observation of even a limited sample of kilonovae following GW signals, allowing to explore both their intrinsic diversity and their properties under different viewing angles. When the burst afterglow is also detected, information on the external density and a better estimate of the viewing angle can be obtained, which might be completed, on a longer term, with the possible observation and leveraging of the kilonova afterglow (Hotokezaka et al., 2018b; Nakar et al., 2018; Kathirgamaraju et al., 2019).

5.8 Conclusion

In conclusion, regardless of the evolution of GW and electromagnetic detectors, kilonovae and multi-wavelength afterglows will remain instrumental in the study of binary neutron star mergers as windows on both their environment and their role as progenitors of short GRBs. They will bring invaluable insight at the level of the population of mergers and on an event-to-event basis.

In Chaps. 6 and 7, we will show how upcoming BNS mergers with electromagnetic counterparts can contribute to the measurement of the Hubble constant and the understanding of short GRB environments. In particular, we will show how to exploit the upcoming afterglows from the significantly misaligned jets that the multi-messenger era will see. These are two multi-messenger prospective studies that only hint on the fruitful role that BNS mergers can play in astrophysics.

In Chap. 11, we propose to extend our population model to other electromagnetic observables such as afterglow radio polarization. Once more, these would be useful to understand the multi-messenger dataset we expect to be available for future events.

Chapter 6

What role will compact object merger afterglows play in multi-messenger cosmology?

Abstract

The Hubble constant H_0 is a fundamental cosmological parameter of the local Universe. There exist many probes to measure H_0, relying on phenomena having occurred either in the early or in the local Universe. Early- and late-Universe measurements of H_0 currently are in tension. The gravitational-wave signals from merging binary compact objects allow to measure the distance to the source, i.e., they are *standard sirens*. Gravitational-wave signals together with redshift information from such sources are therefore excellent candidates to make a measurement of H_0 that is independent of any cosmic ladder. GW170817 opened up this era of gravitational-wave cosmology. However using the bare gravitational data and source redshift from a single event only produces a fairly poor constraint on H_0, because of fundamental degeneracy between source distance and inclination angle in the gravitational-wave signal. The electromagnetic counterparts to these sources together with emission models can provide further knowledge on the source's inclination angle, and therefore improve the measurement of H_0. GW170817, with its full-house of counterparts, showcased this *multi-messenger cosmology*, with afterglow-derived inclination-angle information reducing the H_0 uncertainty threefold. Could future binary neutron star mergers with multi-messenger datasets help to resolve the Hubble tension? For events endowed with electromagnetic counterparts to significantly accelerate the tension resolution, such events must occur often enough and their inclination angle constraints must be tight and unbiased. Coupling our population models for electromagnetic counterparts to binary neutron star mergers to realistic estimates of inclination angle uncertainties, we quantify the benefit of exploiting radio afterglows to make multi-messenger measurements of H_0 during upcoming gravitational-wave observing runs. Our statistical framework in this calculation includes the selection effects linked to using only the brightest afterglows. We find that, while each afterglow event greatly improves the measurement of H_0, these should prove too rare to significantly accelerate the Hubble tension resolution, with respect to events without afterglow counterparts. This conclusion, together with the potential biases linked to using uncertain electromagnetic modeling to measure H_0 when afterglows are available, relativizes the use of afterglows for multi-messenger cosmology. We discuss how multi-messenger events can still contribute to precision cosmology, leveraging kilonova signals for example, which should prove more readily accessible though currently more poorly understood than afterglows.

6.1 The Hubble constant and the Hubble tension

The definition of the Hubble constant stemmed from the observations of a linear relation between the distances and radial velocities of nearby galaxies by Hubble (1929). In a physical cosmology context, it is defined as the current and local value of the expansion rate of the Universe as predicted by general-relativistic cosmological models (Lemaître, 1927). From a conceptual standpoint, the Hubble constant, denoted H_0, is the most fundamental cosmological parameter of any isotropic Universe. We will briefly follow an elementary derivation by Peacock (1999, Chap. 3) which reveals the fundamental place of H_0 in any cosmological model.

Denote by $\overrightarrow{v}(\overrightarrow{r})$ the velocity field of matter at position $\overrightarrow{r}$. In the local Universe, for matter closer than a certain length scale which is undefined for now, $\overrightarrow{v}$ can be expanded linearly in $\overrightarrow{r}$ as $\overrightarrow{v}(\overrightarrow{r}) = M \cdot \overrightarrow{r}$ for a certain matrix M. As for all matrices,

M can be decomposed into symmetrical and antisymmetrical parts[1] $M = S + A$. Now, antisymmetical A can be expressed as a rotation: $A \cdot \overrightarrow{r} = \Omega \wedge \overrightarrow{r}$. However, by isotropy of the Universe there cannot exist a privileged direction and therefore $\Omega = 0$. What's more, symmetrical S can be diagonalized in a certain basis $S \sim \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$. However, by isotropy again all directions are equivalent, and therefore $S = H \times \mathbb{1}$ for a certain constant H. If follows that, locally:

$$\overrightarrow{v}(\overrightarrow{r}) = H\overrightarrow{r} \tag{6.1}$$

which is known as the *Hubble–Lemaître law*.

The only local motion allowed by isotropy is pure radial expansion, and the expansion rate H, with units km/s/Mpc appears as the most fundamental cosmological parameter.

The above derivation relies solely on isotropy and assumes no particular physics framework. Thus Eq. 6.1 is perfectly scale-free and one has no prior knowledge of the physical length scale under which this linear expansion is valid. Adopting a relativistic setting introduces c as a universal speed scale, v/c as the expansion parameter and the *Hubble length* c/H as the validity region for the linear expansion of the expansion flow. Currently, this is ~ 4 Gpc.

During the last twenty years, cosmology has become a precision science, with relative uncertainty on H_0 changing from 100% in 1994 to 1% in 2020 (Montanet et al., 1994; Zyla et al., 2020). This impressive progress is the result of ever-improving instruments, data quality and analysis techniques, as well as physical understanding of cosmological probes. The improving precision in different H_0 measurements methods has revealed a significant tension between the outputs of different methods to measure H_0. This *Hubble tension* is a salient puzzle of contemporary astrophysics, and the context to the rise of *multi-messenger cosmology*, which will concern us in this chapter. Before delving into multi-messenger cosmology, we will make a brief account of pre-gravitational-waves cosmology.

6.2 Classical-probe cosmology

There are many ways to measure H_0, relying on cosmological probes from different eras of the Universe, early or late. Most prominently, current "early-Universe" measurements rely on the Cosmic Microwave Background (CMB) and assume the Λ-Cold Dark Matter cosmological model, while "late-Universe" measurements rely on type Ia supernovae and various local cosmic distance ladder calibrators.

6.2.1 CMB cosmology

The CMB is faint thermal radiation observed in all directions with a quasi-uniform temperature of $\sim$ 2.7 K. It was accidentally discovered by Penzias and Wilson in 1964, confirming predictions of early expanding-Universe cosmological models. In such models, the CMB is made of the first free-streaming photons of the Universe, that is, photons which last scattered onto matter when the matter-radiation decoupling occurred. The CMB spectrum currently peaks as millimeter wavelengths. This places the last scattering surface at a redshift of $z_* \sim 1000$, assuming an emitted spectrum peaking around the Rydberg energy ($\sim$ 13 eV), as appropriate for photons emitted from the recombination of electrons with protons to form hydrogen.

The COBE satellite launched in 1992 discovered a quadrupolar feature in the CMB temperature skymap. This was the first intrinsic anisotropy identified, as the dipole component can be attributed to a boost with respect to the CMB emitter's rest-frame.

Successive space missions, up to Planck launched in 2016, set out to measure the CMB ever more precisely in order to study the angular scales of the anisotropies of the temperature map. These anisotropies are on the order of $\Delta T/T \sim 10^{-5}$. They trace anisotropies in the cosmic plasma at or before the epoch of recombination (primary anisotropies) or interaction of the CMB during its propagation to us (secondary anisotropies).

In Fig. 6.1, we show the final CMB temperature skymap obtained by Planck along with its multipole spectrum. The multipole spectrum is obtained by decomposing the temperature 2-point function on spherical harmonics, thereby studying the primary anisotropies on different scales at the epoch of recombination.

Under the Λ-CMD cosmological model, the description of the cosmic plasma at this epoch requires a number of parameters. Certain possess "natural" values (such as the dark energy equation of state $w = -1$) or values obtained from other observations and are found not to affect the final measurement of H_0 (such as the curvature of the Universe, fixed to $k = 0$). Beyond these, there remain six independent parameters, which determine the various physical processes at the origin of the primary anisotropies, which are imprinted on the CMB spectrum. A fit to the CMB spectrum allows to measure these parameters. Then, the Λ-CDM model allows to propagate the cosmological parameters up to the present day. This constitutes the method to measure H_0 with CMB observations.

The Planck mission measured the multipole spectrum up to $l \sim 2500$, all decently fit with the six Λ-CDM parameters. Note that anisotropies are also present in other CMB observables, such as the radiation polarization. These anisotropies alone and their correlations with temperature anisotropies make for *cross spectra*, which are also fit. Using this procedure, the Planck Collaboration states a pure-CMB measurement of $H_0 = 67.27 \pm 0.60$ km/s/Mpc (Planck Collaboration et al. 2020c, 68% confidence interval)[2].

6.2.2 Type Ia supernovae cosmology

Type Ia supernovae (SNIa) are optical transient sources powered by runaway thermonuclear fusion in a white dwarf member of a binary system, after accretion from the main sequence companion star (*single-degenerate* progenitor scenario) or upon

[1] These are of course $(M + M^\intercal)/2$ and $(M - M^\intercal)/2$ respectively.

[2] This value is different from the often-cited $H_0 = 67.66 \pm 0.42$ km/s/Mpc from the same reference, which makes use of non-CMB probes, notably galaxy structure and lensing experiments.

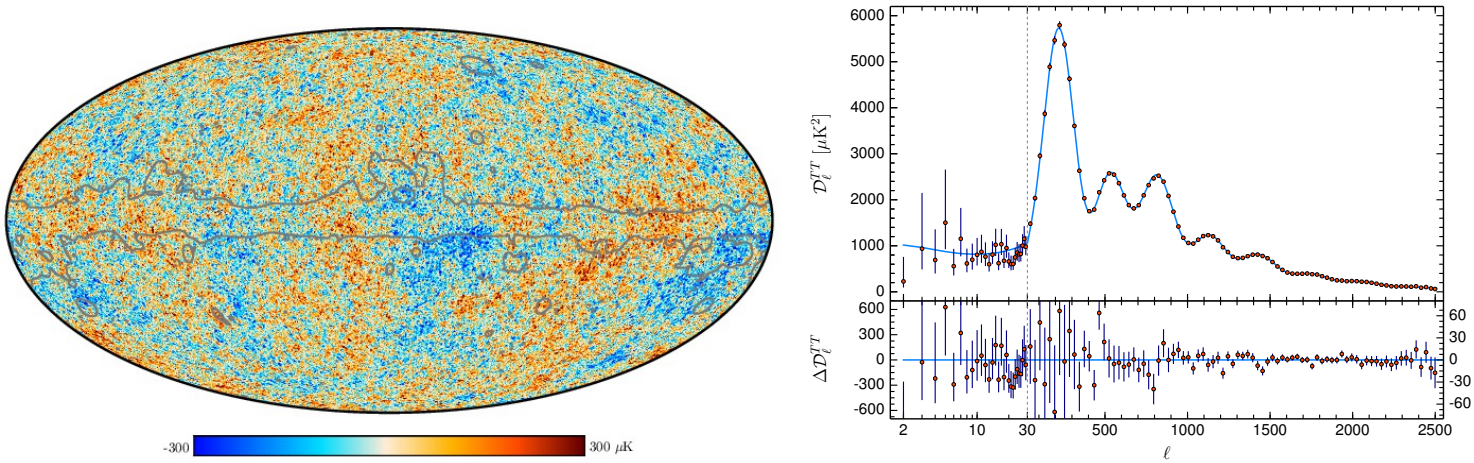


Figure 6.1: **Cosmic Microwave Background cosmology.** Left: Planck CMB temperature map (Planck Collaboration et al., 2020a). Right: Planck CMB temperature multipole spectrum, along with best-fit Λ-CDM cosmological model (this particular fit also contains information from non-CMB probes, Planck Collaboration et al., 2020c).

merger in the case of a double white dwarf (*double-degenerate* progenitor). Upon formation, the white dwarf is sustained by electron degeneracy pressure, at a temperature too low to fuse its core elements, most often carbon or oxygen. Through continued accretion from the companion, the weight of the accreted material may increase the core density and temperature enough to ignite fusion. As the star is made of degenerate matter and dominated by temperature-independent pressure, the additional pressure from the core fusion reactions does not lead to expansion and temperature regulation. Thus the core temperature increases boundlessly, eventually leading to a runaway fusion of the entire star. The energy release, typically 10^{51} erg, is enough to unbind the star.

In Milky-Way-like galaxies, SNIa occur a few times over 100 years. There are currently $O(10^5)$ known SNIa explosions, and their bright peak magnitude allows to detect them up to cosmic distances (Guillochon et al., 2017). The *Hubble Space Telescope* has observed SNIa up to redshifts of 1.9.

The ignition of the explosion mechanism from nearly the same initial conditions for all systems render SNIa's almost uniform in luminosity, with absolute V magnitudes consistently at $M_V = -19.3$. This uniformity is likely due to the mass of nickel-56 synthesized in SNIa's being standard, as spectral data shows that the decay of nickel-56 is the dominant source of radiation at the SNIa peak, ~ 20 days after explosion.

Regardless, there is stills some variability in the peak luminosity of SNIa's, which are therefore not exactly *standard candles*. However, they are *standardizable*, through the so-called Phillips relation. This is a tight empirical correlation observed between SNIa peak magnitude and post-peak decay rate: brighter SNIa decay slower after peak. The Phillips relation is most often stated in the B-band as a correlation between the peak magnitude M_B and the magnitude drop 15 days after peak $\Delta m_{B,15}$. Assuming this relation holds exactly for all systems, one may assign a standardized peak luminosity to SNIa's using the measured $\Delta m_{B,15}$, which does not depend on the source distance. Fig. 6.3 shows the standardization procedure, by which scaling time and magnitude axes to the Phillips relation allows all SNIa light curves to assume the same, standard, shape. In turn, comparing the peak apparent magnitude to the standardized peak luminosity allows to measure the luminosity distance to the SN.

The Phillips relation must be calibrated on sources with independent SNIa distance measurements. These are deduced from other empirical correlations of closer-by sources—on a lower scale of the *cosmic distance ladder*—which overlap with SNIa in terms of distance. The historical choice for these are Cepheid variable stars, with a luminosity-period correlation. Finally, the latter relation can be calibrated for Cepheids lying in nearby galaxies, on the lowest, so-called *geometrical* scale of the cosmic distance ladder, where parallax provides the basic distance measurement.

In Fig. 6.2, we reproduce the unified cosmic distance ladder, with three scaled figured: geometrical, Cepheid and SNIa. Empirical correlations allow to step from one scale to another, and sources lying in two scales at once allow to calibrate such relations: Cepheids lying in local galaxies, and SNIa lying in Cepheid-bearing galaxies.

The procedure to measure H_0 with SNIa's is therefore to collect a sample of sources, with associated distances using the Phillips relation, and cosmic flow velocity from host galaxy redshift measurements. Using the Hubble–Lemaître flow definition of H_0 (Eq. 6.1), a late-probe measurement of H_0 follows.

In practice, a purely Cepheid-calibrated SNIa measurement of H_0 does not yield a precise result, and state-of-the-art measurements also include other geometric distance indicators, such as detached eclipsing binaries or galactic masers. Also note that other calibrator sources than Cepheid variables can be used, for instance stars at the tip of the red giant branch (TRGB, Freedman et al. 2019 or highly-evolved low-mass stars (Mira stars, Huang et al. 2020), the luminosity of which can be inferred from stellar physics.

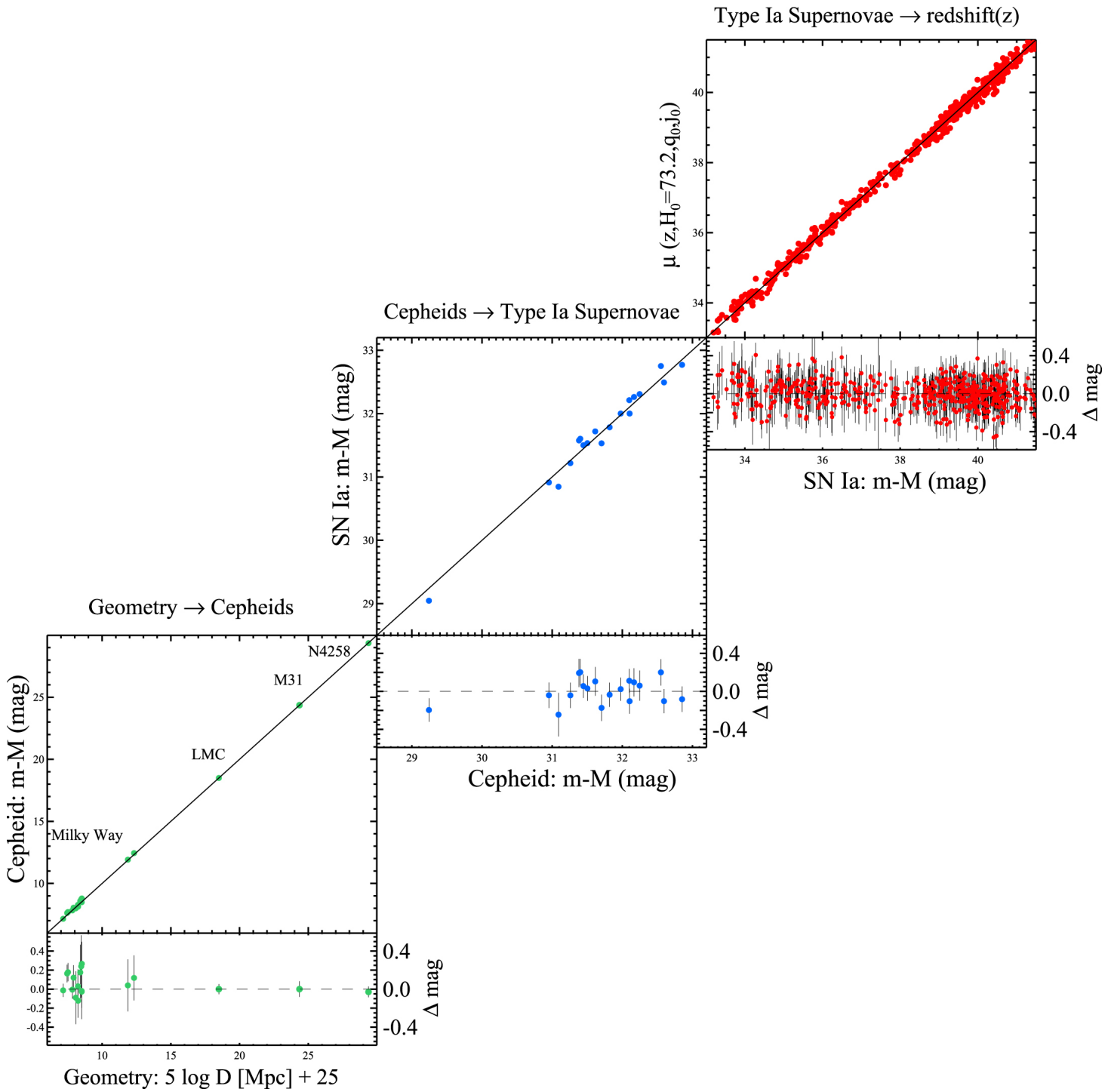


Figure 6.2: **Cosmic distance ladder,** where geometrical distance measurements (in the Milky-Way, the Large Magellanic Cloud, M31, and NGC4258) allow to calibrate Cepheid period-luminosity relations, which in turn allow to calibrate the Phillips relation and determine the distance to SNIa's (Riess et al., 2016).

The current best late-probe measurement of the Hubble constant is $H_0 = 74.03 \pm 1.42$ km/s/Mpc (Riess et al., 2019b).

Before this precise determination of H_0, the large redshift range up to which SNIa's are observed already allowed to reconstruct the local z-dependence of H. This led to the first evidence for a currently accelerated expansion of the Universe, and the introduction of dark energy as a dominant component (Riess et al., 1998).

A fundamental limitation of SNIa cosmology is the question of whether these sources are standard, and systematics in the standardization can spoil the Hubble constant measurement. To a lesser extent, local-probe cosmology is limited by the motion of the sources due to physical interaction—their *peculiar motion*—which is indistinguishable from the motion due to the Hubble–Lemaître expansion flow. Cosmology using very close sources—such as compact object mergers—is particularly sensitive to this effect. We will discuss this in more detail in Sec. 6.8.

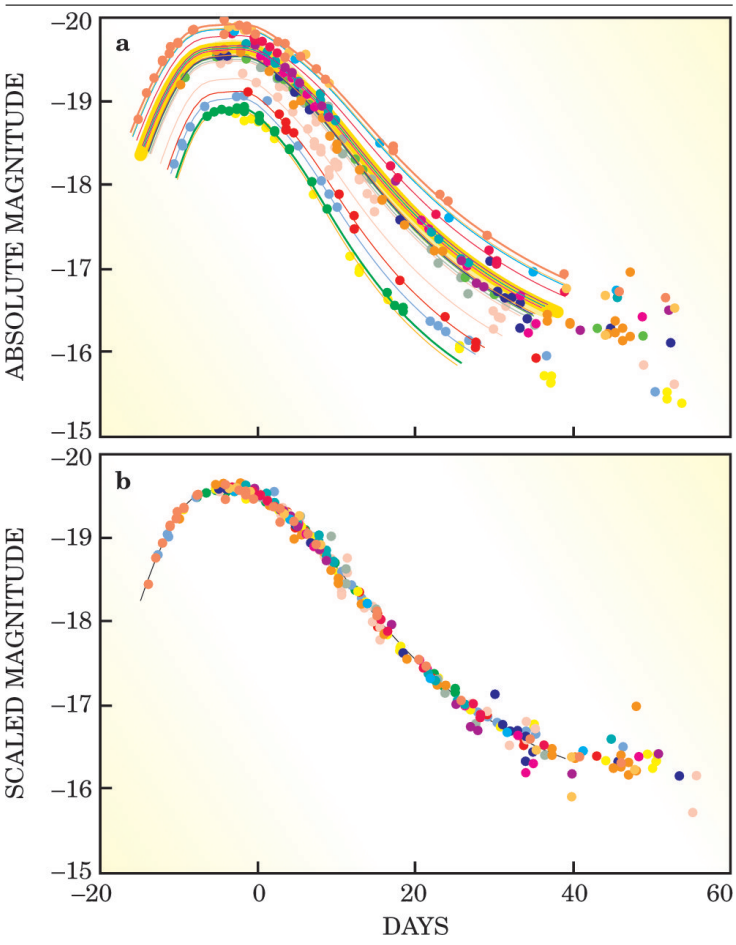


Figure 6.3: **Standardization of SNIa light curves.** Scaling the time axis for individual light curves (top) to a standard 15-day decay $\Delta m_{B,15}$ and scaling the magnitude axis by a factor so as to satisfy the Phillips relation reveals the standard nature of SNIa light curves (bottom) and allow to define a standard SNIa luminosity (Perlmutter, 2003).

6.2.3 On possible origins of the Hubble tension

The two estimates of the Hubble constant derived from CMB and type Ia supernovae are in tension at the ≥ 4-σ level, which is significant (Feeney et al., 2018). There are many possible origins of this discrepancy, and a rich literature that has explored these avenues. We will simply mention that these can be roughly categorized as follows:

- New physics in the early Universe: The Λ-CDM model may overlook physics that influence the cosmology in the young Universe. This could be either on the side of the standard model for particle physics, or the description of gravity. For a review of these *theoretical solutions*, see Di Valentino et al. 2021.
- Methodology biases in the late Universe: While the measurement of the CMB by satellite is relatively straightforward, photometry of SNIa's is prone to many limiting effects such as crowding or extinction which may be uncontrolled (e.g. Riess et al., 2020).
- Cosmic variance: At the scales where the standard candles are collected, the homogeneity of the Universe may not be exact, such that the H_0 measured in our local Universe is different from that in the early Universe, or that on average at the present cosmic time (e.g. Zehavi et al., 1998).
- Cosmic distance ladder statistical systematics: Either within the samples of sources used to calibrate a given step of the cosmic ladder or within the sample of SNIa's used to finally measure H_0, there could be selection effects. For example, an eventual dependence of the SNIa's luminosity to its galactic environment would require a careful selection of the sources for calibration samples (e.g. Rigault et al., 2013).

To date and to my knowledge, there is no clear indication as to where the tension lies.

6.3 Standard-siren and multi-messenger cosmology

In the context of the Hubble tension, multi-messenger cosmology based on compact binary coalescences may be a game-changer. Measuring the local Hubble constant requires essentially a source with known distance and redshift. In this regard, the *self-calibrating* nature of the inspiral signal from binary compact objects is fundamental. Indeed, the distance to the source can in theory be directly deduced from the time-frequency map of an inspiral signal (Sec. 2.2); For this reason, they are dubbed "standard candles" and were early understood as valuable independent probes for cosmology (Schutz, 1986; Nissanke et al., 2013).

However, this distance measurement remains degenerate with other source parameters, notably the inclination angle ι

that appears in the GW amplitude (Eq. 2.23) and which is difficult to infer from the GW data (Sec. 2.3.2; It is precisely in this chapter that we will deal with such degeneracy in the GW data.

Acquiring the source's redshift can follow many routes, according to the availability of electromagnetic counterparts:

- Completely disregarding the electromagnetic domain, the redshift can be measured from the GW data itself, using higher-level approximations of the signal. Indeed, the tidal interaction between the binary components imprints the inspiral signal at the fifth order in post-Newtonian expansion (Henry et al., 2020). Supposing knowledge of the equation of state of neutron star matter, this tidal interaction appears at predictable frequencies in the source frame, which are redshifted in the observer frame (Messenger & Read, 2012; Del Pozzo et al., 2017). This "GW-only" method suffers from uncertainty on the equation-of-state. Currently, the quality of the GW data forbids using this method in real observing data.

- In the case where no electromagnetic counterpart can be identified, the GW skymaps together with galaxy catalogs informs us on the galaxies compatible with hosting the source. Weighing in all the redshifts of those galaxies allows to make a statistical measurement of the source redshift(Fishbach et al., 2019). This method was first effectively applied during the O3 run for the binary black hole merger GW170814 (Soares-Santos et al., 2019; Gray et al., 2020). This "dark-siren" method suffers from incompleteness in galaxy catalogs.

- In the case where the source can be attributed to a host galaxy thanks to an electromagnetic counterpart, the redshift of the source can be secured with much increased precision. This of course was one of the breakthroughs of GW170817, with an independent measurement of the Hubble constant of $H_0 = 70^{+12}_{-8}$ km/s/Mpc (Abbott et al., 2017d). Such a "basic multi-messenger" method is limited by the above-mentioned degeneracy of the distance with the other source parameters, and especially the inclination angle ι. Improving the multi-messenger H_0 constraint requires to break this degeneracy.

- One promising solution is to measure ι from anisotropic electromagnetic counterparts emitted after the merger. The most stringent constraint comes from the photometry and imaging of the afterglow. For GW170817, these observations provided inclination angle measurements precise to ~ 12 deg (Guidorzi et al., 2017; Troja et al., 2019a; Hajela et al., 2019, e.g.) and ~ 5 deg respectively (Mooley et al., 2018b; Coughlin et al., 2019a; Ghirlanda et al., 2019; Ascenzi et al., 2021) and refined the H_0 measurement to $H_0 = 70.3^{+5.3}_{-5.0}$ km/s/Mpc (Hotokezaka et al., 2019); This "fully multi-messenger method" provided a three-fold improvement with respect to the basic method.

Therefore, further afterglow counterparts could drastically improve H_0 measurements and play a leading role in multi-messenger cosmology. However precise the H_0 measurement can become using the afterglow, we showed in Chap. 5 that these signals are faint and difficult to detect for distant or very inclined binaries. They should therefore prove rare in the future (especially the afterglow, recall the $\leq 20\%$ fraction of detectable afterglows after kilonova detection). It is thus important to study whether the rareness of these sources will prevail before their capability to precisely measure H_0 on individual events, or whether they will play a leading role in multi-messenger cosmology.

Furthermore, using electromagnetic measurements of ι entails selection effects which, if they are not correctly understood, can significantly bias the subsequent measurement of H_0 and must be carefully taken into account (Chen, 2020).

In this chapter, we will equip ourselves once again with our population model to study the potential for the afterglows of upcoming BNS mergers to measure the Hubble constant with fewer events than relying only on the GW data to measure the distance and on the kilonova to provide the redshift.

6.4 Which electromagnetic counterparts can we use to measure H_0?

In this section, we describe the observing scenario leading to a fully multi-messenger measurement of the Hubble constant.

The multi-messenger scenario naturally starts with the GW trigger. In this chapter, we will use the detailed modeling of the GW detection process (Sec. 4.4.1), which relies on proprietary LVKC resources; We will use three noise power spectrum distributions (PSDs) for the HLV network, composed of the LIGO Hanford, LIGO Livingston and Virgo detectors. The first PSD is indicative of the detectors sensitivity during O2 (Abbott et al., 2017b), the second is indicative of the sensitivity reached during the first 3 months of O3 (Abbott et al., 2020a), while the third one is a projection for the O4 run sensitivity (Biwer et al., 2019).

Detection of the kilonova counterpart and thereby acquiring the system's redshift through its host galaxy is the minimal scenario required for a multi-messenger measurement of H_0. We will refer to this scenario as "Level 1". It is subject to detection of the kilonova, for which we adopt the model described in Sec. 4.4.3; We consider the r band, and a limiting magnitude of $r_{\rm lim} = 21$, as appropriate for, e.g., ZTF (Bellm et al., 2019). In order to make the precise predictions necessary to precision cosmology, we will also consider the criterion for the localization of the kilonova, as in Sec. 4.4.4.

In this scenario, the information on D_L is provided solely by the GW data, without any contribution from electromagnetic counterparts. Indeed, any direct distance information from an electromagnetic counterpart would require using the cosmic distance ladder, which is of course excluded in the perspective of measuring H_0.

In principle, the kilonova signal could indirectly contribute through the measurement of ι from color-evolution considerations (Kashyap et al., 2019; Dhawan et al., 2020). However,

these signals currently crucially lack modeling and observing history. Inclination angle measurements from kilonova data are very model-dependent and lack robustness (Doctor, 2020; Heinzel et al., 2021). It is only when the kilonovae sample will have grown that the potential impact of kilonovae-derived angle constraints will be appreciated. Therefore, we will exclude any contribution to ι from kilonova data.

As expected from the observation of short gamma-ray bursts and evidenced in the case of GW170817, relativistic jets are launched from BNS mergers (Mooley et al., 2018b; Ghirlanda et al., 2019). The interaction of this jet with the circum-merger environment leads to the long-lived afterglow emission.

The afterglow photometry can provide an independent measurement of inclination angle. In such measurements from off-axis jet afterglow data fitting, there is unavoidable degeneracy with the jet opening angle (Nakar & Piran, 2021). However, when combined with prior estimates for short gamma-ray burst jet opening angles (see references in Sec. 4.5), afterglow data can still lead to inclination angle information. It can therefore indirectly inform on D_L by breaking the D_L–ι degeneracy in the GW data. We will refer to the scenario where Level 1 is realized and a ι measurement from the jet afterglow photometry is made as "Level 2". It is subject to the realization of Level 1 and to detection criteria on the jet afterglow light curve.

Detecting the afterglow at its peak does not suffice to make an inclination angle measurement. This requires an extended and well-sampled light curve on which to fit afterglow models, as was extensively done for GW170817 (e.g., Lamb & Kobayashi, 2017; Resmi et al., 2018; Lazzati et al., 2018; Troja et al., 2019a). We therefore define the criterion for ι measurement with the afterglow light curve as:

$$F_p > 10 \times F_{\rm lim} \tag{6.2}$$

where F_p is the peak flux of the afterglow light curve as determined by our afterglow model (Eq. 4.13), and $F_{\rm lim}$ is the limiting flux of the follow-up facility. As in Chap. 5, we consider the 3 GHz band and the Very Large Array (VLA) as the limiting radio facility, with $F_{\rm lim} = 15$ μJy. In this chapter, we will continue to consider both the G16 and the WP15 population hypotheses in order to study their impact on our results; For a reminder on the definitions of G16 and WP15, report to Sec. 4.2.

Note that GW170817 does not exactly qualify for our Level 2 scenario. Indeed it had $\log F_p/F_{\rm lim} \sim 0.9 < 1$ (Mooley et al., 2018c). It seems that much of the uncertainty in the measurement of ι with GW170817 is held in the very early phases of the afterglow, where the fitting models most diverge (Ghirlanda et al., 2019). Had the afterglow been brighter—at the level of our Level 2 scenario—and these earlier points observed, a better measurement of ι would have certainly ensued. Nonetheless, as we detail in Sec. 6.6, we consider the case of GW170817 as representative of the ι measurements possible in the Level 2 scenario.

Additionally, the relativistic nature of the jetted outflow can be revealed by VLBI observations (Mooley et al., 2018b; Ghirlanda et al., 2019), that evidenced an apparent superluminal motion of the jet head. Detecting this image centroid displacement is possible for events that are particularly close or bright or under specific inclination angle conditions (Sec. 5.5). Doing so further constrains ι and narrows down the measurement of H_0, as shown in the case of GW170817. We call "Level 3" the scenario where such a constraint on ι can be extracted from afterglow VLBI imaging, in addition to those of afterglow photometry. This final level of ι constraint is the most informative on ι and H_0, but also the most difficult to obtain.

For this chapter, we use the detailed model for the source angular displacement, described in Sec. 4.4.6, with an angular resolution of $\Delta\theta_{\rm VLBI} = 2$ mas, as appropriate for the European VLBI network used during the follow-up of GW170817 (Ghirlanda et al., 2019).

Other electromagnetic counterparts could potentially provide further independent measurements of ι. These are the short gamma-ray burst and the rebrightening in the source's multi-wavelength signal due to the emergence of emission from the forward shock of the decelerating mildly relativistic ejecta responsible for the kilonova emission called the "kilonova afterglow" (Sec. 3.4 Hotokezaka et al., 2018b; Kathirgamaraju et al., 2019). However, the short gamma-ray burst should prove extremely rare in future events (Ghirlanda et al., 2016; Beniamini et al., 2019) and robust modeling lacks for gamma-ray signals, especially for very inclined events. Furthermore, the singularity of GRB170817A with respect to other short gamma-ray bursts has cast more uncertainty on the modeling of gamma-ray emission from BNS mergers (Sec. 3.2.3, Kasliwal et al., 2017; Nakar et al., 2018). Modeling of the kilonova afterglow is still uncertain, rendering any angle measurement difficult; For more discussion on this topic, see Sec. 7.3.1. To sum up, we will consider neither the gamma-ray burst nor the kilonova afterglow as viable to measure ι and thus H_0.

Thanks to our population model, we numerically determine the probability of occurrence of Level 1, 2 and 3 scenarios as a function of the source luminosity distance D_L and inclination angle ι. We denote these functions as $p_{\rm det}^{L_i}(D_L, \cos\iota)$ for $i = 1, 2, 3$. Higher levels require more counterparts and we therefore naturally have $p_{\rm det}^{L_3} \leq p_{\rm det}^{L_2} \leq p_{\rm det}^{L_1} \leq 1$ for all D_L and $\cos\iota$.

6.5 Selection effects in multi-messenger cosmology

Before we study the prospects for multi-messenger cosmology, it is instructive to anticipate selection effects in this endeavor. As noted in Sec. 6.3, uncontrolled selection effects could be at the origin of the current Hubble tension; We must thus control them as much as possible in this new multi-messenger cosmology.

In Appendix F, we derive the statistical framework appropriate for measuring population parameters under selection effects. This is the framework we will use in Secs. 6.6 and following.

In the present case, the population parameter λ to infer is H_0, and the individual event properties θ_i's are z and $\cos\iota$. The

posterior distribution on H_0 obtained after observing a single-event multi-messenger set d of GW and electromagnetic data is (Eq. F.13):

$$p(H_0|d) = \frac{1}{\beta_S(H_0)} \frac{p_0(d|H_0)\pi(H_0)}{p(d)}. \tag{6.3}$$

where $\pi(H_0)$ is prior information before the measurement and $p_0(d|H_0)$ is the likelihood of the data in the absence of selection effects (Eq. F.11).

All the potential selection effects are included in the *selection function* $\beta_S(H_0)$. It accounts for the bias arising from the fact that the probability to observe an event with properties θ_i–and use it to measure λ–is not the same for all θ_i's .

For a given electromagnetic information level L_i, the selection function reads (Eq. F.9):

$$\beta_S(H_0) = \int \mathrm{d}\cos\iota\, \mathrm{d}z p_{\mathrm{det}}^{L_i}(z, \cos\iota|H_0) p_{\mathrm{pop}}(z, \cos\iota|H_0) \tag{6.4}$$

where $p_{\mathrm{det}}^{L_i}(z, \cos\iota|H_0)$ is the probability of making the measurement at level L_i on source located at redshift z and inclination ι, and $p_{pop}(z, \cos\iota|H_0)$ is the distribution of the intrinsic population of sources in redshift and inclination. In these two functions, the conditioning on H_0 simply indicates that a value of H_0 is assumed for the correspondence between D_L and z.

As explicit in Eq. 6.3, the H_0 posterior will only be impacted by selection effects if the selection function is not flat, i.e., has a significant variation with H_0. In our case, the detection probabilities are best expressed in terms of D_L, which is the preferred parameter for the GW and electromagnetic models:

$$\beta_S(H_0) = \int_0^{+\infty} \mathrm{d}D_L \int_{-1}^{+1} \mathrm{d}\cos\iota \frac{\mathrm{d}z}{\mathrm{d}D_L} p_{\mathrm{det}}^{L_i}(D_L, \cos\iota) \times p_{pop}(z(D_L, H_0), \cos\iota) \tag{6.5}$$

Considering the simplest hypothesis of the formation of binaries uniformly in comoving volume up to a boundary redshift z_0 and linear cosmology, we have $p_{pop}(z, \cos\iota) = 3z^2/z_0^3$, such that $\int_0^{z_0} \mathrm{d}z p_{pop}(z, \cos\iota) = 1$.

Our electromagnetic and gravitational-wave detection models neglect the effects of redshift on the detected chirp mass and last stable circular orbit for the GW part, and the electromagnetic wavelengths for counterpart searches. This is justified because all the GW triggers detectable with the current generation of interferometers have redshift ≤ 0.1, for both the CMB- and SN1a-deduced values of H_0; In Sec. 6.8 we discuss the next generation further-reaching instruments. Therefore, the detection probabilities p^{L_i} we consider have no explicit dependence on the source redshift. In these circumstances, the cosmology is $\mathrm{d}z = H_0/c\,\mathrm{d}D_L$, and the integrand in $\beta_S(H_0)$ will be:

$$p_{\mathrm{det}}^{L_i}(D_L, \cos\iota) \frac{\mathrm{d}z}{\mathrm{d}D_L} p_{pop}(z(D_L, H_0), \cos\iota) \tag{6.6}$$

$$= p_{\mathrm{det}}^{L_i}(D_L, \cos\iota) \frac{H_0}{c} \times 3 \left(\frac{c}{H_0 D_{L,0}}\right)^3 \left(\frac{H_0 D_L}{c}\right)^2 \tag{6.7}$$

leaving no dependence on H_0, whatever the boundary redshift z_0 or distance $D_{L,0}$.

Therefore, assuming (i) linear cosmology, (ii) uniform-in-comoving-volume system formation rate, and (iii) no explicit redshift dependence in the detection process, the selection effects on the multi-messenger measurement of H_0 are null. This fact was already underlined in Mandel et al. (2019).

If, however, any of these hypotheses fails, such simplifications do not occur and selection effects can appear. Note the intricate relation between selection effects and the redshift dependence of the source population or of the detection probability.

There is naturally another source of selection effects, which is simply the uncertainty on p_{det} itself. Measuring H_0 accounting for selection effects requires knowledge of β_S, which itself depends on p_{det}. Therefore, mismodeling of the emission and detection process results in biases on the ultimate measurement of H_0. The electromagnetic sector is likely more prone to such mismodeling. If the afterglows does not prove effective sources for multi-messenger cosmology, this mismodeling and its potential biases could be a reason to disqualify these counterparts altogether.

Finally, note that measuring H_0 requires knowledge of β_S, which itself paradoxically requires a cosmological model to compute, unless the sources are so close that linear cosmology is enough. In principle, this can be avoided by introducing both H_0 and, e.g., the deceleration parameter q as population parameters, supposing the adequate $D_L(z)$ function in the expression for β_S (Eq. 6.5) and making a joint measurement of both parameters with the standard sirens and the same statistical framework.

In order to illustrate the effects of a general, non-linear cosmology, in Fig. 6.4 we plot the selection function for our various observing scenarios assuming a flat Λ-CDM cosmology with current dark matter density $\Omega_m = 0.308$.

In the circumstances of measuring H_0 with data from GW170817, that is, a Level 3 scenario during the O2 run, the selection effects are less than 2.0% over the 60–80 km/s/Mpc range, according to Fig. 6.4. This is well below the 14% precision claimed by studies making this measurement (Hotokezaka et al., 2019). Thus, no selection effects significantly impacted the measurement of H_0 with the GW170817 data.

However, the selection effect is not negligible compared to the 4% precision required to resolve the H_0 tension (Freedman, 2017; Feeney et al., 2018), particularly with events detected in O3- and O4-type runs, where selection effects reach 2% and 4% in the 60–80 km/s/Mpc range, respectively. Then, careful consideration of selection effects becomes necessary.

6.6 Method to determine the Hubble constant convergence rate

We now set to quantify the benefit of electromagnetic-provided information in measuring H_0. We adopt a uniform-in-comoving-volume density for the sources, which is justified in

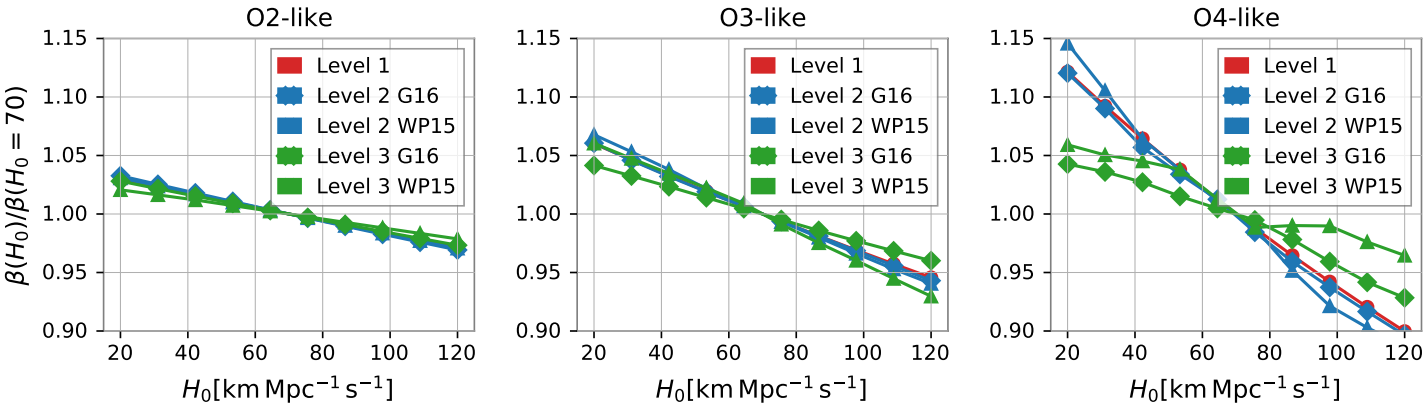


Figure 6.4: **Selection function** $\beta_S(H_0)$, normalized to an arbitrary nominal value of $H_0 = 70$ km/s/Mpc, for the different observing scenarios we considered. [Mastrogiovanni et al. 2021]

the local Universe (Mortlock et al., 2019a; Mastrogiovanni et al., 2020) and linear cosmology.

For a single event, we can rewrite the H_0 posterior as in Eq. 6.3 with the following likelihood:

$$p_0(d|H_0) = \int \mathrm{d}z\, \mathrm{d}\cos\iota\, p_{\rm GW}(x_{\rm GW}|D_L(H_0,z),\cos\iota) \times p_{\rm EM}(x_{\rm EM}|\cos\iota,z) p_{pop}(z,\cos\iota|H_0). \quad (6.8)$$

In the above, $p_{pop}(z,\cos\iota|H_0)$ is the assumed population distribution in redshift and inclination for the entire BNS population.

The likelihood encodes the statistical uncertainties of the GW and electromagnetic data measurements $x_{\rm GW}$ and $x_{\rm EM}$. It is in this function that we can include the precision with which ι can be inferred from the electromagnetic data in different scenario levels.

The function $p_{\rm GW}(x_{\rm GW}|D_L(H_0,z),\cos\iota)$ is the GW likelihood, which provides the distributions for the GW-detected source properties $\hat{D_L}$ and $\cos\hat{\iota}$. We adopt the Cutler-Flanagan approximation for consistency with our GW detection model (Sec. 4.4.1).

Likewise, $p_{\rm EM}(x_{\rm EM}|\cos\iota,z)$ is the likelihood for electromagnetic measurement of source parameters $\cos\iota$ and z. This is informative on z only for a Level 1 scenario and on both z and $\cos\iota$ for higher-level scenarios. It is not informative on D_L. We decompose this electromagnetic likelihood as $p_{\rm EM}(\hat{z}|z)p_{\rm EM}(\cos\hat{\iota}|\cos\iota)$ by supposing the electromagnetic measurements of redshift and angle are independent. This is reasonable as the redshift information is deduced from the host galaxy alone while the ι information is expected to be provided by the jet itself.

We assume the redshift measurement is unbiased and set $p_{\rm EM}(\hat{z}|z)$ to a Gaussian distribution centered on the true event redshift with standard deviation $5\cdot10^{-4}$ for all our scenarios. This is the same accuracy measured for GW170817's redshift and is dominated by the uncertainty on the source's peculiar motion (Abbott et al., 2017d); For events further than GW170817, it is thus an overestimate of the uncertainty on z. For Level 1 scenarios, the electromagnetic counterpart is uninformative on $\cos\iota$ and we set $p_{\rm EM}(x_{\rm EM}|\cos\iota)$ to a flat function.

For Level 2 scenarios, we assumed that one can obtain an unbiased ι constraint at the level of that deduced from GW170817's afterglow light curve alone. That is, a Gaussian constraint with a 12deg 1-σ uncertainty (Troja et al., 2019a; Hajela et al., 2019) for $p_{\rm EM}(\cos\hat{\iota}|\cos\iota)$.

For Level 3 scenarios, we based our predicted constraints on those of GW170817 and set $p_{\rm EM}(\cos\hat{\iota}|\cos\iota)$ to an unbiased Gaussian constraint with a width of 4deg (Mooley et al., 2018b; Ghirlanda et al., 2019; Hotokezaka et al., 2019) for all the events. We discuss the validity and impacts of these assumptions in Sec. 6.8.

Using Eqs. 6.3, 6.5 and 6.8, we can simulate the H_0 measurement for a number of events drawn from the intrinsic population and supposedly observed with electromagnetic information Levels 1, 2 or 3. This allows us (i) to directly compare the electromagnetic levels in terms of how fast the Hubble converges when the different levels of information are available and (ii) considering the probabilities to observe each level and therefore the actual occurrence rates of each level, to determine if higher-level events will be too rare or if they will statistically allow us to measure H_0 with fewer events.

Technically, we adopt a virtual value of 70 km/s/Mpc for the Hubbble constant, and simulate the H_0 measurement process of 500 binary systems in all three observing scenarios and all three GW sensitivity hypotheses: O2, O3 and O4. For the individual measurements, we systematically assumed a prior on H_0 uniform in [40, 120] km/s/Mpc .

We then combine the measurements of the first 100 events to emulate a thread of multi-messenger events. We repeated the combining step after reordering the 500 events to reproduce different possible time orderings of the events. This allowed us to study the reconstruction of H_0 by the multi-messenger measurements, and in particular the speed of convergence.

6.7 Forecast on multi-messenger H_0 measurements

6.7.1 Bulk comparison of observing scenarios

Breaking the $D_L - \cos\iota$ degeneracy is fundamental for measuring H_0. Let us show this by inferring H_0 using *only* Level 2 or 3 scenarios in comparison with Level 1 scenarios.

Fig. 6.5 shows the H_0 posteriors obtained by combining 10 BNS events in different observing scenarios. From the plot, we can see that, when the knowledge of $\cos\iota$ is refined by electromagnetic observations, the estimation of H_0 improves, as predicted. We can also observe that the H_0 posterior reaches Gaussian convergence after under 10 events.

The bottom panel of Fig. 6.6 shows the relative uncertainty $\Delta H_0/H_0$ with 1-σ confidence intervals for the H_0 estimation as more events are detected. The uncertainty corresponds to the different population realizations of the detected events. In Fig. 6.6, one can read that a single Level 3 event during an O2-type run results on average in the uncertainty of 14% on the estimation of the Hubble constant, as observed for GW170817 (Hotokezaka et al., 2019).

It is clear that (i) the precision on H_0 improves as more events are combined, (ii) the convergence is faster when $\cos\iota$ is more constrained from the electromagnetic emission, (iii) above about 10 events, the combined H_0 posterior becomes Gaussian, i.e. $\Delta H_0/H_0 \sim \Theta/\sqrt{N}$ where N is the number of events, allowing us to define Θ as an effective single-event H_0 estimation standard deviation.

In Fig. 6.7 (left) we show the values of the average effective single-event standard deviation Θ of the different scenarios, as fit to the curves in Fig 6.6. There is a clear boost in the H_0 convergence speed when considering the information in $\cos\iota$ from the electromagnetic counterparts. We find that a $\cos\iota$ precision of 4 deg (Level 3) provides a 1.4-times faster convergence than one of 12 deg (Level 2), which is itself about 1.5 times faster when there is no angle information at all from electromagnetic counterparts. Roughly, it means that the H_0 accuracy reached combining ten Level 2 events is equivalent to that reached by combining five Level 3 events.

We also find that detectors with better sensitivities will be able to better constrain H_0, even without $\cos\iota$ measurements from electromagnetic counterpart. This is due to the higher redshifts of the events used to infer H_0, as we consider a constant uncertainty in redshift measurements. Considering events without electromagnetic contributions to $\cos\iota$ for O4-type runs, we find an average effective single-event standard deviation of $\Theta = 14\%$, consistent with previous simulations (Chen et al., 2018; Gray et al., 2020).

In the top panel of Fig. 6.6, we show the expected number of years of continuous observing required to detect the number of events read on the bottom horizontal axis, according to our population model. From the bottom panel of Fig. 6.6, the number of Level 2 events required to resolve the H_0 tension assuming an O2-like sensitivity is ~ 25. According to the top panel, one would need 100 to 600 years of observation to collect these events, depending on the population model. This of course is unrealistic; In Sec. 6.7.2 we will study how these durations change with better GW sensitivity and more realistic detection rates of electromagnetic counterparts.

Counterparts to GW events detected with different GW sensitivities probe different regions of the distance and inclination parameter space. In particular, the fraction of events detectable in the electromagnetic domain is very sensitive to the value of the GW horizon (Chap. 5). Therefore, one cannot deduce the number of events and a given electromagnetic level or observing time required to resolve the H_0 tension for an O4-type run from those for an O2-type run by simply comparing the GW detection rates. In the next section, we derive the number of events required to resolve the H_0 tension assuming a long-lived O4-type run.

Note that for a given GW sensitivity, there is no difference in this first approach between G16 and WP15 population models, as these only impact the probabilities of detecting the electromagnetic counterparts. This is clear in Figs. 6.6, 6.7.

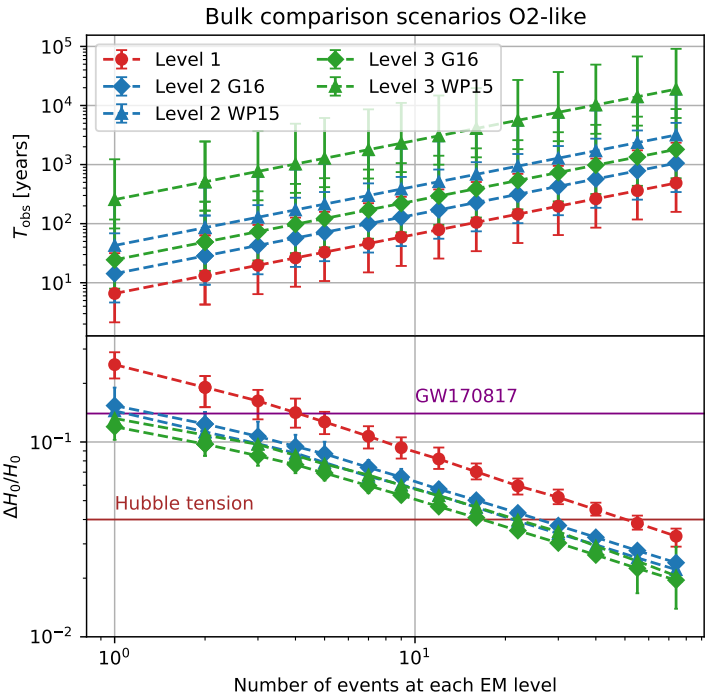


Figure 6.6: **Precision on H_0 as a function of the number of events.** Bottom: Precision on the estimation of H_0 with 68.3% confidence intervals as a function of the number of events detected at each electromagnetic level. Here, all events are assumed to be at a given electromagnetic information level, as denoted in different colors. We also indicate the tension in the Hubble constant and the precision on H_0 obtained with GW170817 and counterparts. Top: Expected number of years of continuous observation required to detect events in various multi-messenger scenarios. [Mastrogiovanni et al. 2021]

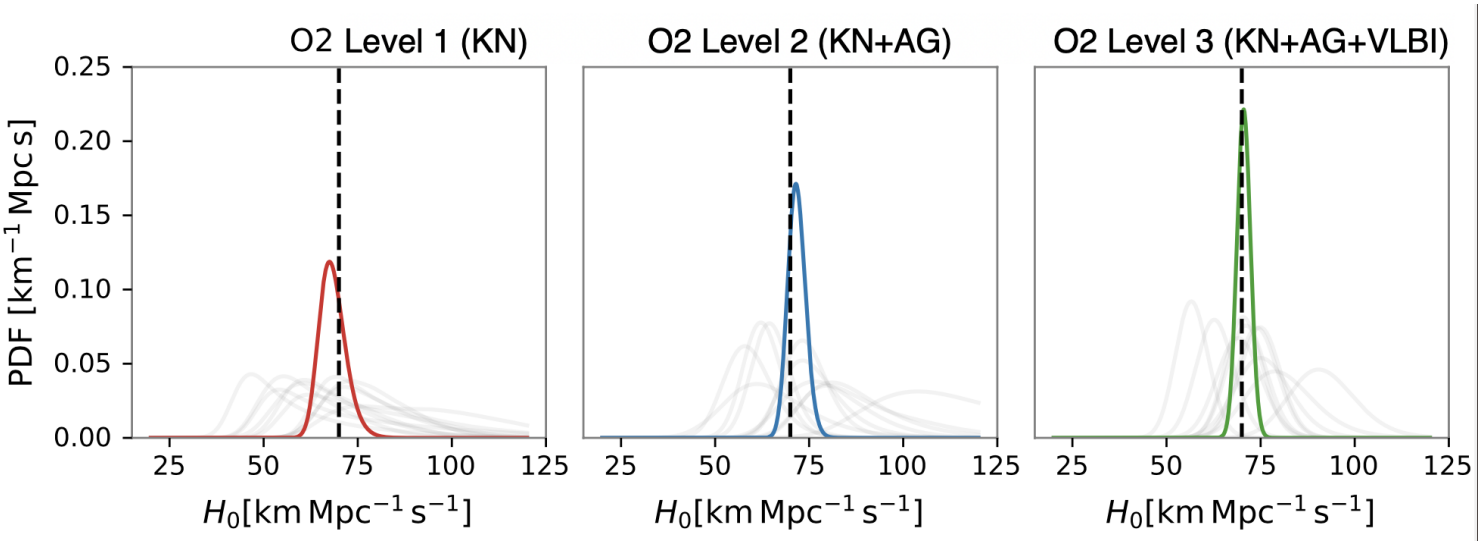


Figure 6.5: **Posterior on H_0 obtained for 10 BNS events observed with different level of electromagnetic information** on ι in an O2-type run. Left: Level 1 (no electromagnetic information on ι), Center: Level 2 (electromagnetic-based ι precision of 12 deg), Right: Level 3 (electromagnetic-based ι precision of 4 deg). [Mastrogiovanni et al. 2021]

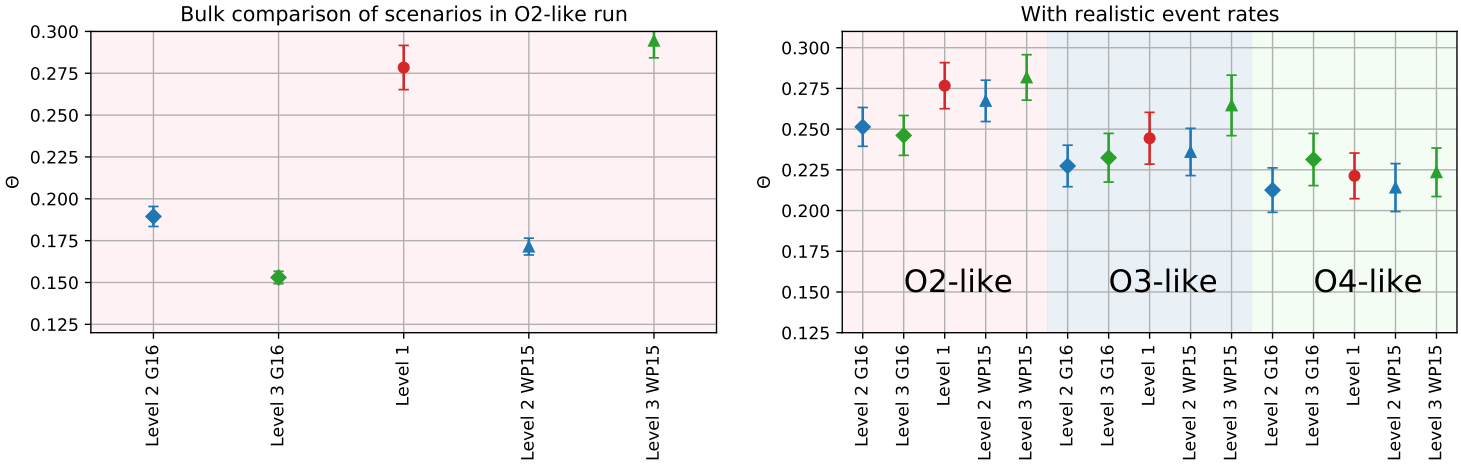


Figure 6.7: **Effective single-event standard deviation Θ in different observing scenarios.** Left: Assuming all events are at a given electromagnetic information level, in an O2-type GW run. Right: Assuming a realistic rate of electromagnetic counterpart detection. [Mastrogiovanni et al. 2021]

6.7.2 Considering realistic detection rates

In a real observing run, not all the GW events with a redshift estimation will have $\cos\iota$ measurements from electromagnetic counterparts. In a second approach, we estimate the H_0 convergence by including the relative detection rates of the different electromagnetic counterparts. More precisely, we generated threads of events as in the direct level comparison of the previous section, but successively allowing for only Level 1 events, then up to Level 2, then up to Level 3, to quantify the acceleration of H_0 convergence each electromagnetic level allows.

We then simulated the H_0 measurement for each of them, and combined their measurements throughout the first 100 events. We repeat this process 500 times with varying event time orderings to simulate different realizations of the subset of 100 detections.

Fig. 6.8 shows the H_0 convergence as a function of the total number of GW detections, while Fig. 6.7 (right) shows the single-event standard deviation Θ for all scenarios.

According to Sec. 5.5, about 30% of O2 events with associated redshift would have had Level 2 information on ι, and about 5% Level 3 (neglecting the requirement of the kilonova localization). As seen in Fig. 6.8, this detection fraction is enough to somewhat improve the H_0 convergence, i.e., the convergence speed allowing for Level 2 and 3 events is larger than with kilonova events alone.

The situation changes drastically when we consider the multi-messenger events in O3- and O4-type observing runs. In this case, the fraction of Level 2 or Level 3 events are so small that on average they bring no additional improvement. Indeed, we find that the precision on H_0 obtained combining a given number of GW events does not change whether we allow for afterglow counterparts or if we do not. In particular, for an O3-like run, only allowing for Level 3 events and assuming the optimistic G16 population prescription could provide a slight acceleration in the H_0 narrowing-down, while for all the other cases the improvement is negligible. In O4-type runs, neither Level 2 or Level 3 events should statistically speed up the convergence of H_0.

During an O4-type run, Level 2 or 3 events are too rare to significantly improve the H_0 convergence and shorten the time needed resolve the H_0 tension. Relying on Level 1 events only is just as fast. In Sec. 6.8, we argue that, in this case, discarding measurements of the inclination angles from afterglows prevent further biases in the H_0 estimate.

For O4-like runs, our population model for kilonovae (Sec. 5.3) shows that one Level 1 event is detected every 6 months on average. Fig. 6.8 shows that, to resolve the H_0 tension, 30 such events are required. Thus, with an O4-like sensitivity, 15 years of data taking are necessary to collect the number of detections with measured redshift.

6.8 Discussion

In this chapter, we studied the prospects of measuring the Hubble constant with GW standard sirens coupled to inclination angle measurements from merger afterglow counterparts. We first studied the potential impact of selection effects in multi-messenger cosmology and showed that these were negligible in the H_0 measurement reported after GW170817. We illustrated how selection effects increase with the GW sensitivity. For events in future GW observing runs, selection effects will be important and should be taken into account.

In the future, third-generation GW detectors such as the Einstein Telescope (ET, Punturo et al., 2014) and Cosmic Explorer (Reitze et al., 2019) will open up the detection range of larger redshifts, where these selection effects should prove very important. For larger redshifts, follow-up observations will be limited by the sky localization, with an average resolution of 200 deg^2 for BNS at $z = 0.1$ with ET. This coverage issue combined with the unprecedented dimness of the counterparts at these redshifts may call for totally different follow-up strategies for the high-cadence large field-of-view survey facilities. Supposing the source is identified, photometric and spectroscopic follow-up would still be limited to $z = 0.5$ and $z = 0.3$ respectively, for the largest optical telescopes such as the Extremely Large Telescope (Maggiore et al., 2020). Therefore, access to those events at cosmic redshifts should rely on the observation of counterparts other than kilonova, such as the short gamma-ray burst, the detection of which could be facilitated by the GW inspiral signal being present in the ET band hours before merger. In this case, however, it is not certain that a robust measurement of ι can ensue, as it could be that short GRBs can still be observed at significant viewing angles in this intermediate redshift range.

We studied whether the observation of merger afterglow signals and subsequent measurements of $\cos\iota$ will significantly accelerate the narrowing-down of H_0 when combined with GW detections in the future. We considered only the afterglow signal—its photometry and imaging—as potential providers of ι measurements. We deemed the other counterparts such as the kilonova and gamma-ray burst unfit for such a measurement, for their still large modeling uncertainties.

We considered an optimistic measurement model in which all events with an afterglow counterpart contribute a ι measurement with an accuracy comparable to GW1710817, for both afterglow photometry and imaging. This is an optimistic assumption as the uncertainty on ι depends on the number of photometric points detected from the light curve, and thereby on the event distances and density of the follow-up. Also, only a subset of the follow-up campaigns is expected to provide such detailed multi-wavelength photometric data. Taking the variability of the follow-up scope and data quality into account is a possible extension to this work.

Furthermore, our analysis assumes that $\cos\iota$ estimated from electromagnetic is accurate, which is a simplifying assumption. For both the afterglow photometry and imaging analysis, the leading uncertainty in the electromagnetic modeling is the treatment of the jet lateral expansion (Ghirlanda et al., 2019). The jet expansion affects both the time of afterglow peak flux (Sec. 4.4.5) and the dynamics of the VLBI image (Fernández et al., 2021), possibly biasing and widening the electromagnetic

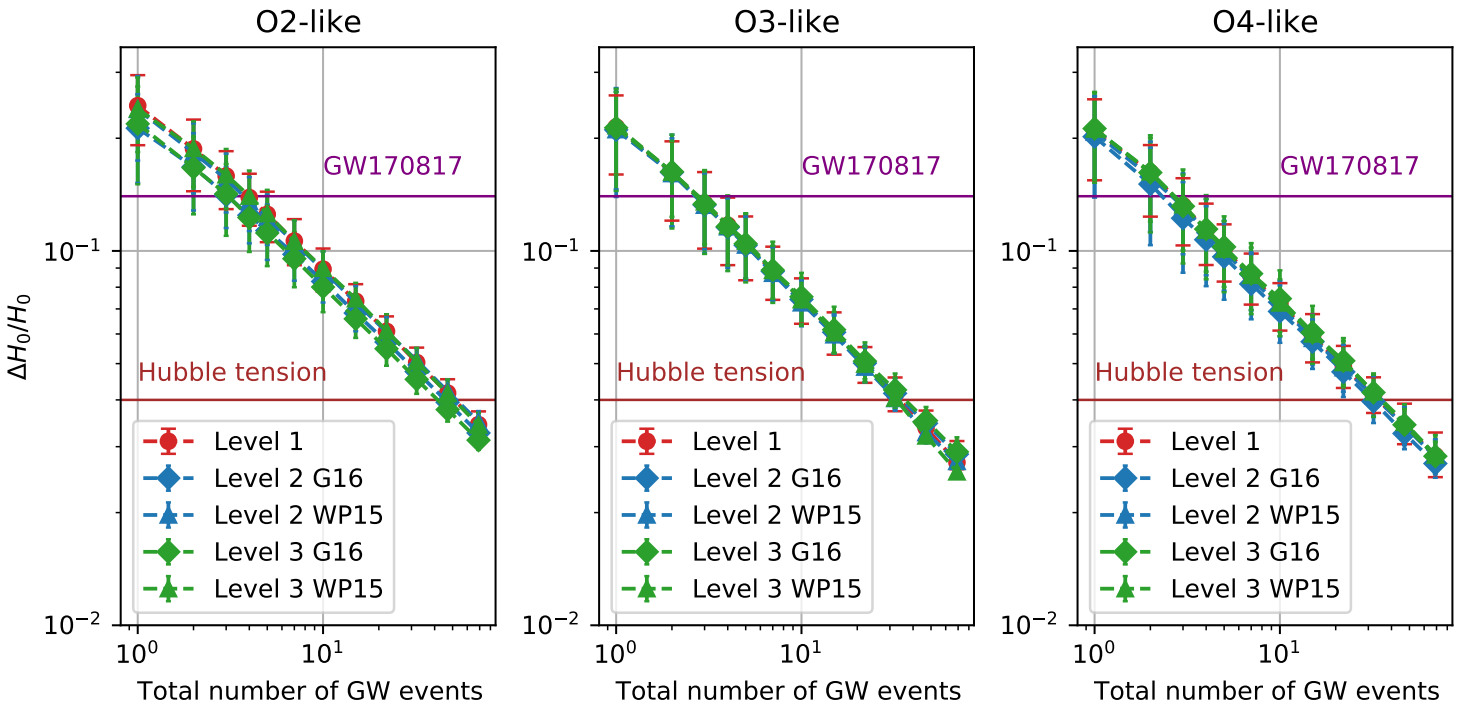


Figure 6.8: **Precision on the Hubble constant as a function of the number of events observed, with realistic detection rates.** Same as Fig. 6.6 (bottom panel), with the bottom axis counting the total number of GW events, regardless of the nature of the electromagnetic counterpart. Therefore, this accounts for the realistic detection rate of events with different electromagnetic counterparts. [Mastrogiovanni et al. 2021]

posterior on ι more than assumed here.

From the point of view of observations, most of the indeterminations in measuring ι from GW170817's afterglow lie in the very early phases of the light curve and the late phases of the source displacement curve (see the posterior sampling in Ghirlanda et al., 2019). Acquiring early photometric data with deeper searches would have provided a better estimate of ι from this source. Recently, radio points were acquired from GW170817 to flux levels much deeper than the early radio monitoring of GW170817 (Balasubramanian et al., 2021, , see Sec. 3.4), proving that this is in reach of current radio facilities. We thus advocate for such early deep searches.

Also, we assumed perfect kilonova detection and identification above a magnitude threshold over the accessible sky, thus overlooking limitations already discussed in Sec. 5.7. Furthermore, our expression of the source angular displacement during follow-up in Eq. 4.17 is clearly an overestimate, leading to over-predict the number of sources with detectable proper displacement.

Finally, the effect of peculiar velocities can bias the estimation of H_0. Galaxies' peculiar motions can be as high as $v_p = 300$ km/s, with an associated error of their measurement of the same order (Mukherjee et al., 2021). This corresponds to a redshift correction (and additional uncertainty on H_0) of 10% for events such as GW170817 and it is fundamental to take into account. With future GW detectors, that will detect BNS up to redshift ~ 0.1 this type of correction will be negligible but important in the case of close-by, high-SNR events.

For all these reasons, our hypotheses are optimistic. Even so, we have found that, for all GW runs after O3, events with afterglow counterparts should prove so rare that, statistically, using the afterglow counterparts when available for the multi-messenger measurement of H_0 on individual events will not bring any acceleration to the measurement of H_0.

We found that, for the electromagnetic measurement of $\cos\iota$ to significantly increase the H_0 convergence, the GW and electromagnetic detection probabilities should be comparable, or the understanding of electromagnetic emission from compact binary mergers should drastically improve, to the point where a degree-level precision on ι is accessible from a typical afterglow light curve. Even then, care should be taken with systematic effects, e.g., in angle measurements from VLBI imaging, to not create a second tension on H_0.

The current state of kilonova modeling does not allow for robust measurements of the inclination angle. This may change in the future as the model uncertainties will probably reduce after more signals are observed. Given the above-mentioned rates of kilonova associations with GW, we estimate that kilonovae could accelerate the narrowing-down of H_0 if a $\sim 10\%$ calibration can be reached between kilonova data (such as light curves or color evolution) and inclination angle. In this respect, the advent of wide field-of-view, high-cadence optical facilities is an asset as they are expected to collect a large sample of kilonovae detected both serendipitously and as GW counterparts. With

tens of well-sampled kilonova light curves, the 10% accuracy level for the light curve vs. inclination relation may be within reach, especially if these sources are standardizable (Kashyap et al., 2019; Coughlin et al., 2020b,a). This perspective would truly allow multi-messenger cosmology to develop.

We proved that the electromagnetic-provided $\cos\iota$ measurements will likely not drive the H_0 narrowing-down. Therefore, direct biases to H_0 through electromagnetic mismeasurements of $\cos\iota$ should not be feared. However, as the detection probabilities of the electromagnetic counterparts should dominate the selection effect for joint GW–electromagnetic standard sirens, incorrectly modeled dependence of the kilonova signal on the inclination can lead to errors on the detection probability and eventually to H_0 biases through uncontrolled selection effects, as shown in Sec. 6.5. Correct modeling of the kilonova signal to control the selection effects in follow-up campaigns should be a point of care for future endeavors in multi-messenger cosmology.

Once these selection biases are dealt with, the limiting uncertainty in multi-messenger cosmology should be the GW data calibration. This calibration uncertainty is currently at the level of 5–8% (Abbott et al., 2021c) and projected to reach $\sim 1\%$ at the design configuration of the current generation of interferometers (Karki et al., 2016). This level is below the Hubble constant tension and therefore calibration uncertainty should not impede the resolution of the tension by multi-messenger cosmology when combining a low number of events in the future. However, a systematic effort on calibration uncertainties when combining a large number of events should be performed.

As the afterglow counterparts should not accelerate the measurement of H_0, we can state that the number of multi-messenger events necessary to resolve the H_0 tension is still that given by Chen et al. (2021b); Mortlock et al. (2019b) considering only the kilonova counterpart, that is 20–50. This represents about fifteen years of continuous O4-level GW observation. We expect this sample of events to be collected in less than this time; Indeed, the O4 run is projected to last for one year and a half and lead to further instrument improvement and thus higher detection rates for the following runs (Abbott et al., 2020a).

6.9 Conclusion

The afterglow counterparts of binary neutron star mergers represent viable means to measure the inclination angle of sources, and thereby to improve the standard-siren measurement of the Hubble constant. Afterglows could therefore play the role of narrowing down H_0 and possibly resolve the Hubble tension with fewer events than by leveraging only the gravitational-wave data and source redshift. To quantify how much faster afterglow-enhanced H_0 measurements could solve the Hubble tension, we carried out a realistic population model considering that every future afterglow counterpart could provide a constraint on the source inclination angle at the same level as GW170817. We found that, while each afterglow allows for a jump in H_0 precision, events with afterglow counterparts should prove very rare, to the point that allowing for afterglow-enhanced measurements should not statistically make any difference in the number of events required. Once models have improved, kilonova light curves could be viable for inclination angle measurements and, as these should be much more frequently acquired, kilonovae could play the leading role in multi-messenger cosmology. Whether for kilonova or afterglow counterparts, one must treat selection effects with care so as to not produce yet another tension, because upcoming gravitational-wave observing runs will probe distances where selection effects are important.

Chapter 7

New insights on binary neutron star environments and delay-time distribution

Abstract

The binary neutron star merger GW170817 occurred in a rarefied medium with a density smaller than 10^{-2} cm^{-3}. Since kicks are imparted to neutron star binaries upon formation, and due to their long inspiral times before merger, such low-density circum-merger media are generally expected. However, there is some indirect evidence for fast-merging or low-kick binaries, which would coalesce in denser environments. Nonetheless, present astronomical data are largely inconclusive on the possibility of these high-density mergers. As shown by GRB170817A and our subsequent population model, the gravitational-wave era will bring the observation of afterglows from misaligned jets. We describe a method to directly probe this hypothetical population of high-density mergers exploiting multi-messenger observations of binary neutron star mergers, including these misaligned afterglows. This method is based on a sample of merger afterglows that has yet to be collected. Its constraining power is large, even with a small sample of events. We discuss the method's limitations and applicability. In the upcoming era of third-generation gravitational wave detectors, this method's potential will be fully realized as it will allow us to probe mergers that occurred soon after the peak of cosmic star formation, provided the follow-up campaigns are able to locate the sources.

7.1 Introduction

Upon the second supernova leading to their formation, binary neutron stars (BNS) are kicked away from their dense star-forming birth regions (Blaauw, 1961; Boersma, 1961; Fryer & Kalogera, 1997), allowing them to migrate to a different environment before merging (Portegies Zwart & Yungelson, 1998). These kicks are attested, e.g., by the large measured proper motions of Galactic binary pulsars (average speed of ~ 400 km/s, Hobbs et al. 2005). The duration of this migration is the inspiral time of the newly formed BNS, and defined by its initial separation a and eccentricity e (Eq. 2.20, normalized to Husle-Taylor-like systems):

$$\tau_c = 14 \text{ Gyr} \left(\frac{M}{3\ M_\odot}\right)^{-2} \left(\frac{\mu}{0.75\ M_\odot}\right)^{-1} \left(\frac{a}{5\ R_\odot}\right)^{4} (1-e)^{7/2} \tag{7.1}$$

Because this time is long for typical neutron star masses, this migration is expected to lead to mergers far from the dense formation region of the binary, possibly in a rarefied medium (e.g. Bloom et al., 1999). This was the case for GW170817, which occurred in a medium with density $n_{\rm ext} \lesssim 10^{-2}$ cm^{-3} (Hallinan et al. 2017; Hajela et al. 2019, and see Secs. 3.3 and 7.3.1 below).

However, because of the steep dependence of $t_{\rm ins}$ on the initial separation a, long migrations may not be systematic. For instance, assuming a scale-free distribution of initial separations in the population of newly formed BNS's ($\mathrm{d}N/\mathrm{d}a \propto a$), the distribution of inspiral times is $\mathrm{d}N/\mathrm{d}t_{\rm ins} \propto 1/t_{\rm ins}$, which is also scale-free, thus favoring neither long nor short inspiral times.

In any case, for systems with low kick velocities or short delay times, we expect the distances covered during migration to be shorter than for the rest of the population, leading to the possibility of binaries merging in environments that are much denser than those encountered by systems with long migrations. We refer to these events, with densities $n_{\rm ext} \gtrsim 1$ cm^{-3}, as *high-density mergers*. As we will show in Sec. 7.2, there is some motivation for such a population of high-density mergers.

The link with the merger environment itself is further complicated by the migration velocity of the binaries, which stem from the second supernova of the system. These supernova kicks are poorly constrained in the general picture (Podsiadlowski et al., 2005) and may be variable from a system to

another (e.g., Podsiadlowski et al., 2004). In addition to the system's velocity, the supernova kicks also affect the initial separation and eccentricity and in turn its inspiral time (Brandt & Podsiadlowski, 1995; Kalogera, 1996; Belczyński & Bulik, 1999).

The inspiral phase is the longest of the system's binary life; It follows the nuclear phase spanning from the formation of the binary main sequence star to the formation of the BNS. The nuclear phase is a small fraction of the total lifetime of the system, the *delay time*. However, the behavior of the system during this phase greatly conditions the eventual BNS's orbital properties and kick.

The distribution of delay times in the population is important to many branches of astrophysics: galactic chemical enrichment, the history of galaxy formation, and any use of BNS merger signatures—such as GRBs—for cosmology. For example, the rate of BNS mergers is tightly linked to the cosmic star formation rate through the distribution of delay times.

This delay time distribution (DTD) is essentially the inspiral times of BNSs. The DTD can be approached through a number of ways: through the observation of Galactic systems as pulsars, through the statistics of short GRBs, from the observation of r-process elements produced in BNS mergers, and through binary stellar evolution studies seeking to understand the system's behavior during the nuclear phase. In Sec. 7.2, we provide a concise review of the state of the art on the DTD from all these approaches; We will show that it is mostly inconclusive of the prominence of long delays in the BNS population.

From the point of view of binary stellar evolution, the main sources of uncertainty in the system's behavior in the pre-BNS phase are (i) the mass transfers and winds from the main sequence star to the NS in the pre-BNS binary, (ii) the *common envelope phase* during which the NS shares the companion's envelope leading to uncertain transfers between orbital and the envelope's thermal and gravitational energy, (iii) the asymmetry and mass ejection of the second supernova. All of these effects condition the final BNS's orbital parameters and migration velocity.

From the point of view of GRBs, afterglow modeling is one probe of the environments of BNS mergers (e.g., Fong et al., 2015), and thus indirectly of the DTD and binary kicks. Unfortunately, because of the poor localization of most short GRBs and of the relative faintness of their afterglows, the X-ray afterglow of only a small fraction have been found, and less than a handful have detected afterglows in the radio band (D'Avanzo, 2015).

In this chapter, we prompt yet another way of using BNSs to indirectly constrain the DTD, again through the environments of BNS mergers. In the multi-messenger era, we have shown in Chap. 5 that the afterglows of misaligned jets would be observed. We will prove below that such observations in a multi-messenger context can allow to constrain the environment of a BNS merger. Moreover, we showed that the fraction of afterglows detectable in the radio band sharply increases with the density $n_{\rm ext}$ of the medium hosting the mergers (Tab. 5.3). This is due to the fact that (i) radio frequencies ν_R are expected to fall between the injection and cooling frequencies ν_m and ν_c of the synchrotron slow-cooling regime for the bulk of the population, and (ii) in this regime, the afterglow peak flux scales as $F_p \propto n_{\rm ext}^{\frac{p+1}{4}}$. Thus, should there be mergers in high-density environments, these would be over-represented in the afterglow population with respect to their actual number. In other words, the radio afterglow acts as an amplifier for these higher density mergers.

Given a statistical flux-limited sample of BNS merger afterglow counterparts endowed with sufficient completeness in circum-merger density estimates, one can determine the apparent fraction of high-density mergers. Starting from this number, by estimating the amplification factor related to the high-density-selection effect from population models, one can constrain the intrinsic fraction of mergers in high-density media. This is the principle of the new method we propose in order to study the class of high-density mergers.

As we develop later on, this method should allow us to constrain the number of high-density mergers, even after a small number of GW events with afterglow counterpart. The exact link between the rate of high-density events and the distribution of delay times and kick velocities is not clear, in particular because of the aforementioned uncertainty on the supernova kicks. Nonetheless, the method we suggest here is a first step toward studying the delay-time distribution of BNSs from their merger multi-messenger datasets.

7.2 Indirect evidence regarding mergers in dense media

Theoretically, mechanisms exist that lead to fast-merging or low-kick systems. Among these are (i) an efficient common envelope phase, that reduces initial separation (e.g., Ivanova et al., 2003; Dominik et al., 2012) and merger time, (ii) a favorable supernova kick, that causes high eccentricity and thus rapid merger or a small migration velocity (e.g., Kalogera, 1996), (iii) the formation of the BNS by dynamical capture in a migration trap within an active galactic nucleus disk (Secunda et al., 2019), or (iv) the interaction of the BNS with another compact object therein (Liu & Lai, 2018; Fernández & Kobayashi, 2019). The frequency with which these actually occur is still unclear.

Over the years, a body of indirect evidence on high-density mergers has emerged. However, as we show here, current data is inconclusive regarding the importance of this class of mergers.

First, some population synthesis studies suggest the existence of a 'fast' channel for BNS mergers, and, thus, a delay-time distribution featuring a peak around time-scales as short as 20 Myr (Perna & Belczynski, 2002; Ivanova et al., 2003; Belczynski et al., 2006). These correspond to tight binaries that undergo a third mass transfer episode, and merge while still within star-forming regions in dense environments. These conclusions are corroborated by population study predictions on, for example, r-process element abundances in the Milky Way (Côté et al., 2017) or the redshift distribution of short GRBs (D'Avanzo et al., 2014). The two latter studies suggest a delay-

time distribution with a slope $\lesssim -1$, favoring a population of fast mergers, and therefore possibly mergers in dense external media. However, it has been pointed out that the conclusions of population synthesis studies are somewhat sensitive to the assumptions on the physics of the common envelope phase (Dominik et al., 2012) or the distribution of natal kicks (Safarzadeh & Côté, 2017).

A second approach is the study of the delay times and kick velocities of Galactic systems. This approach is limited by statistics and by the uncertainty in estimating these from observations. However, finding short delay times or weak natal kicks can imply that a significant fraction of double neutron star mergers should occur in regions where star formation may still be significant, and in turn, the densities are large too. Recently, Beniamini & Piran (2019) have shown that at least $10-20\%$ of Galactic systems are born with delay times of less than 100 Myr between formation and merger. Furthermore, Beniamini & Piran (2016) have shown that the majority of the observed BNSs received relatively weak kicks at birth ($v_{\rm kick} \lesssim 30$ km/s, see also Tauris et al. 2017).

Another approach is to consider the nature of short GRB host galaxies. On the one hand, these are found to be star-forming two to three times more often than they are found to be elliptical galaxies (Berger, 2014). This suggests higher density media for a significant fraction of mergers. This is particularly noteworthy since up to a redshift $z \lesssim 1$, that is, where short GRB hosts can be seen, elliptical and star-forming galaxies share roughly equal fractions of the cosmic stellar mass (Bell et al., 2003). This suggests that short GRBs are preferentially found in lower mass galaxies, and thus experience larger external densities on average (Zheng & Ramirez-Ruiz, 2007).

Also, the observed host galaxy offset distribution has a median value of 1.5 half-light radii, with $\sim 20\%$ of objects lying outside five half-light radii and $\sim 20\%$ within one half-light radius (Fong & Berger, 2013; Berger, 2014). This favors higher density environments for the most centered $\sim 20\%$ of systems. However, host-galaxy completeness of typical samples is small. Moreover, the offset distribution relies on a correct identification of the host galaxy, and may be grossly overestimating the true offset if, for example, the true host is a fainter, unobserved galaxy of lower mass or higher redshift (e.g., Behroozi et al. 2014).

Insight into short GRBs occurring in dense environments also comes from GRB afterglow observations. On the one hand, Nysewander et al. (2009) have shown that (i) short and long GRBs present a similar correlation between X-ray flux and gamma-ray fluence, (ii) above a gamma-ray fluence threshold of 10^{-7} erg cm^{-2}, optical afterglows are detected in almost all short GRBs and (iii) short and long GRB afterglows have similar radio-to-X-ray flux ratios. These results prompted Nysewander et al. (2009) to suggest that short GRBs have similar or larger external densities to long GRBs, with typical values that may be as large as 1 cm^{-3}. For a selected sample of short GRB early afterglows, O'Connor et al. (2020) have found that less than 16% of events took place at densities smaller than 10^{-4} cm^{-3}, suggesting that few short GRBs occur in very rarefied media. On the other hand, short GRB afterglow catalogs such as Fong et al. (2015) or Berger (2014) do not exhibit a population of high-density afterglows. Similarly, these studies are limited by poor afterglow sampling, parameter degeneracy in photometry fitting and, often, by a lack of the synchrotron self-Compton cooling component in the radiation modeling. In recent years, with the detection of long-lived emission from GRBs with the *Fermi*/LAT (Ajello et al., 2018), the synchrotron self-Compton cooling channel has been realized to be an important ingredient of the physical picture. As the Compton parameter affects the position of the cooling frequency, using the cooling break in the X-ray band to estimate the density while disregarding the synchrotron self-Compton effect can particularly bias the result (Beniamini et al., 2015). These caveats may impede a reliable estimation of the circum-burst density and explain this apparent contradiction.

Finally, an independent approach to short merger binaries comes from r-process abundance studies. The arguments in favor of short merger times, and therefore possibly mergers in dense environments, have recently been summarized in some detail in Hotokezaka et al. (2018a) and Beniamini & Piran (2019). A prevalence of short merger times is implied by (i) observations of r-process enriched stars in ultra-faint dwarf galaxies (Beniamini et al., 2016), (ii) the large scatter of r-process abundances in extremely metal-poor stars in the Milky Way halo (Argast et al., 2004; Tsujimoto & Shigeyama, 2014; Wehmeyer et al., 2015; Vangioni et al., 2016; Dvorkin et al., 2016; Beniamini et al., 2018; Dvorkin et al., 2021), (iii) the declining rate of deposition of radioactive ^{244}Pu and ^{247}Cm on Earth (Hotokezaka et al., 2015; Wallner et al., 2015; Beniamini & Hotokezaka, 2020) and (iv) the declining rate of [Eu/Fe] as a function of [Fe/H] observed in Milky Way stars for [Fe/H] $\gtrsim -1$ (Matteucci et al., 2014; Côté et al., 2016; Komiya & Shigeyama, 2016; Hotokezaka et al., 2018a; Simonetti et al., 2019). However, these conclusions rely on knowledge of the rates and r-process yields of BNS mergers, and the iron yields of core-collapse and thermonuclear supernovae, all of which are still a matter of debate (see Cowan et al. 2021 and Hotokezaka et al. 2018a for reviews respectively on the r-process in general and on BNS mergers as its astrophysical site).

7.3 Determining the apparent fraction of high-density mergers from afterglow observations

We now describe the method we suggest to directly probe the class of high-density mergers. Our method relies on a sample of afterglow counterparts to GW triggers from BNS mergers, which would have a sufficient completeness in density above a certain limiting afterglow flux. Population models such as ours (Chap. 4) or Gottlieb et al. (2019) apply criteria based on afterglow flux levels, and thus provide predictions on detectable events. Therefore, applying a flux cut to a sample of detected afterglows ensures that the sample actually represents all the detectable events above the threshold. This in turn allows one

to safely use the predictions from population models to compensate for the density-selection effect and infer the intrinsic fraction of high-density events $f_{\rm HD}$ from the apparent fraction $f_{\rm HD}^{\rm obs}$, that is, the one observed in the sample.

In this section, we describe how to estimate $f_{\rm HD}^{\rm obs}$ for a sample of afterglow counterparts to BNS mergers. This can be done by inferring the densities of individual events from multi-messenger observations, or directly on the level of the entire sample.

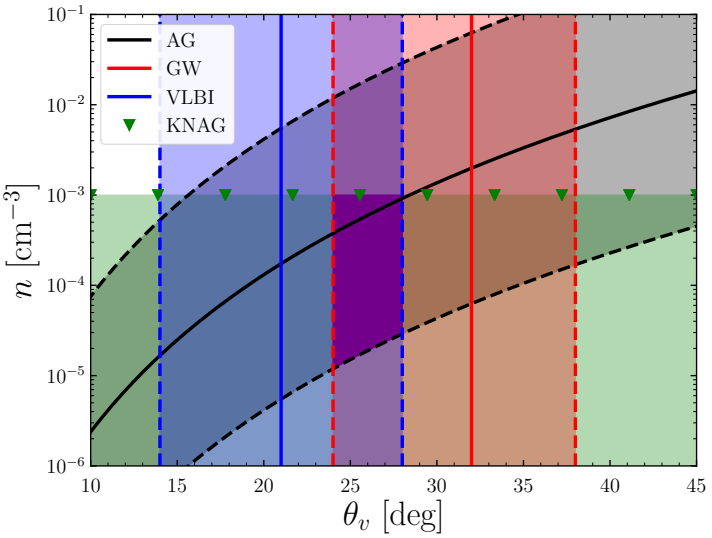


Figure 7.1: **Multi-messenger determination of the viewing angle θ_v and circum-merger medium density $n_{\rm ext}$ in the case of GRB170817A.** We present 1-σ confidence regions (solid line: median; dashed line: 68% confidence limits) obtained from the GW data assuming the source localization (red), the radio afterglow's properties around its peak (black, see Eq. 7.2) and very long baseline interferometry imaging measurements (blue). Green triangles show the upper limit on $n_{\rm ext}$ deduced from the (as yet) undetected kilonova afterglow. The preferred region for θ_v and $n_{\rm ext}$ is highlighted in purple. The text gives details and references. [Duque et al. 2020]

7.3.1 Measuring the viewing angle and density for a single merger event

Combining the GW and electromagnetic information channels allows one to place individual events quite accurately in the θ_v–$n_{\rm ext}$ plane, as has been done in Fig. 7.1 for the case of GRB170817A.

First, in Fig. 7.1 we present the constraints on θ_v obtained from the GW data using the information on the event localization from the electromagnetic counterpart, as was found by Finstad et al. (2018). These are marked in Fig. 7.1, and are representative of three-interferometer constraints that can be obtained in the favorable case where the source is pin-pointed thanks to the detection of the kilonova or early afterglow.

Second, we plot the constraint arising from the properties of the light curve of the radio afterglow around its peak. This constraint is derived in Appendix I and relies on the observation of the turnover in the afterglow light curve from the increasing phase to the decay phase. The duration of the turnover Δt depends both on the viewing angle θ_v and the jet's opening angle θ_j. Combining the corresponding expression with the equation for the 3 GHz afterglow peak flux F_p and peak time t_p as a function of the jet parameters—supposing that the radio band lies in the $[\nu_m, \nu_c]$ portion of the synchrotron spectrum—finally leads to the following relation between observable quantities (left-hand side) and the jet parameters (right-hand side):

$$\left(\frac{F_p}{8.6\ \text{mJy}}\right)\left(\frac{t_p}{4.9\ \text{d}}\right)^{-3}\left(\frac{D_L}{100\ \text{Mpc}}\right)^{2}\times\begin{cases}(\kappa\eta)^2 & \text{no ex.}\\ 1 & \text{ex.}\end{cases} = \theta_{v,-1}^{-6-2p}\, n_{-3}^{\frac{p+5}{4}}\, \epsilon_{e,-1}^{p-1}\, \epsilon_{B,-3}^{\frac{p+1}{4}}, \tag{7.2}$$

where D_L is the luminosity distance to the event, ϵ_e and ϵ_B are the usual shock microphysics parameters, and κ is such that the forward shock Lorentz factor during deceleration is $\Gamma \propto t^{-\kappa}$. For a jet plowing through a uniform medium, κ equals 3/8 for a non-expanding jet, and 1/2 for a jet with sound-speed lateral expansion (Rhoads, 1999). We define the afterglow's *shape factor* $\eta = \Delta t/t_2$ with the turnover time Δt and the onset time of the decay phase t_2; Both are measurable from the afterglow light curve data near its peak (Appendix I). The numerical normalization values on the left-hand side of Eq. 7.2 are valid for $p = 2.2$.

We provide these relations in both the expanding and non-expanding jet hypotheses, which are extreme options regarding the jet lateral dynamics. The actual dynamics should lie somewhere in between, and the discrimination between both can be done on the basis of the post-peak afterglow temporal slope (e.g., Lamb et al. 2018). We note that, in the case of an expanding jet, the θ_v–$n_{\rm ext}$ relation no longer depends on the turnover time, which may prove difficult to measure in the poorly sampled afterglows of marginally detectable events.

Fortunately, the strongest dependencies here are in the measurable quantities t_p, F_p and D_L, rather than on the uncertain ϵ_e and ϵ_B, allowing us to obtain a thin uncertainty region in the θ_v–$n_{\rm ext}$ plane. This constraint, which requires only data on the afterglow around its peak, is shown in Fig. 7.1, where we have taken the values of afterglow observables for GRB170817A from Mooley et al. (2018c). Here, the width of the uncertainty region is obtained by propagating the 1-σ uncertainties on t_p, F_p, D_L and adding an uncertainty of 0.3 (resp. 2) on $\log \epsilon_e$ (resp. $\log \epsilon_B$), deduced from the scatter of its value in GRB jet forward shocks (Beniamini & van der Horst, 2017; Nava et al., 2014; Santana et al., 2014).

Third, we include the viewing angle constraints from the very long baseline interferometry (VLBI) imagery of the radio remnant. By comparing high-resolution imagery of the remnant to synthetic images based on jet models, Mooley et al. (2018b) and Ghirlanda et al. (2019) were able to constrain the viewing angle to the region shown in blue in Fig. 7.1.

Finally, we add the constraint that comes from the non-detection of the kilonova afterglow. This is expected radiation

from the forward shock formed by the mildly relativistic material responsible for the kilonova signal on the external medium (Hotokezaka et al., 2018b; Nakar et al., 2018; Kathirgamaraju et al., 2019). Due to the small Lorentz factor and smooth velocity structure of this ejecta, this afterglow component is expected to peak within a decade post-merger in the case of GRB170817A (Kathirgamaraju et al., 2019). Though we could currently be witnessing the emergence of this component of the afterglow (Sec. 3.4), the flux increase associated with this emergence is still not firmly established (Troja et al., 2021). Taking into account the uncertainty on the physical conditions in the corresponding shock, the absence of this counterpart to date allows us to obtain $n_{\rm ext} \lesssim 10^{-3}$ cm^{-3} (Kathirgamaraju et al. 2019, Fig. 3).

If however the emergence of the kilonova afterglow is confirmed, an actual measurement of the density ensues, and not only an upper limit. However, we note that, in both cases, the constraint depends on the kilonova outflow structure and the assumed value for ϵ_e in the corresponding shock, which is still uncertain for mildly relativistic shocks. Allowing this parameter to assume values suggested for such shocks ($\epsilon_e \lesssim 10^{-2}$, Crumley et al. 2019) by particle-in-cell simulations and by observations of young supernova remnants (Morlino & Caprioli, 2012) loosens the bound on $n_{\rm ext}$. Therefore, we advise prudence on the use of the kilonova afterglow for measurements of the density. More details on this last point may be found in Sec. 7.5.

As seen in Fig. 7.1, the combination of the constraints from the GW, the afterglow light curve and VLBI measurement and the kilonova afterglow leads to $\theta_v \in [24, 28]$ deg and $\log n_{\rm ext}/{\rm cm}^{-3} \in [-5, -3]$ (all 1-σ confidence intervals) for GRB170817A. Disregarding the kilonova afterglow constraint because of the aforementioned uncertainty on ϵ_e in the corresponding shock, the range of inferred densities becomes $\log n_{\rm ext}/{\rm cm}^{-3} \in [-5, -2]$.

It is remarkable how these multi-messenger datasets combine to constrain the density: From the 7 parameters of the radio afterglow $(n_{\rm ext}, \theta_v, E, D_L, \theta_j, \epsilon_e, \epsilon_B)$, measuring 3 electromagnetic observables (F_p, t_p, η) and the GW-determined distance decreases the effective number of degrees of freedom to 3. Drawing on short GRB science, accounting for the uncertainty or variability in the ϵ's reduces to 1 degree of freedom: a quasi-horizontal band in the θ_v–$n_{\rm ext}$ plane. Finally, the GW and VLBI data cut this band almost orthogonally, resulting in the ultimate constraint. We note that the misalignment of the jet—allowed by the fact that these events are GW trigger counterparts—is essential in providing the afterglow observables. Indeed, for aligned jets, the η constraint finds an equivalent in observing the jet break, but the peak flux and time have no equivalents. In Sec. 7.5.1, we will comment further on the benefit of using counterparts to GW triggers in our method.

Such a combination of constraints is only obtained if all the possible multi-messenger observations are made. Using these after a number of events, an estimate of $f_{\rm HD}^{\rm obs}$ can be obtained. It is clear from Fig. 7.1 that GW and VLBI data crucially narrow down the constraint on θ_v. Unfortunately, VLBI remnant imagery will likely become impossible in most cases as the GW horizon increases and we expect its contribution to vanish for most events as of the O3 run (Sec. 5.5.4). In the future, this may be compensated for by some improvement in the GW constraint as more interferometers come online, though it will probably be modest (Veitch et al., 2012; Ghosh et al., 2016a).

An advantage of this multi-messenger estimation of $n_{\rm ext}$ is the use of Eq. 7.2, which requires the properties of the radio afterglow around its peak only and thus is applicable even for faint or poorly-sampled afterglows. Also, it can easily be adapted to other bands, such as the optical, provided they lie between ν_m and ν_c and the afterglow is not outshined by the kilonova. However, Eq. 7.2 is valid only for small densities, when the effects of synchrotron self-absorption in the forward shock are negligible. As illustrated later in Fig. 7.3, this is no longer the case as soon as $n_{\rm ext} \gtrsim 10 - 100$ cm^{-3}, depending on the distribution of jet kinetic energies of the population. Nonetheless, from Fig. 7.3, one expects that at these densities, the X-ray afterglow will be readily accessible and $n_{\rm ext}$ can be estimated from fully-fledged afterglow fitting, containing more physics than Eq. 7.2.

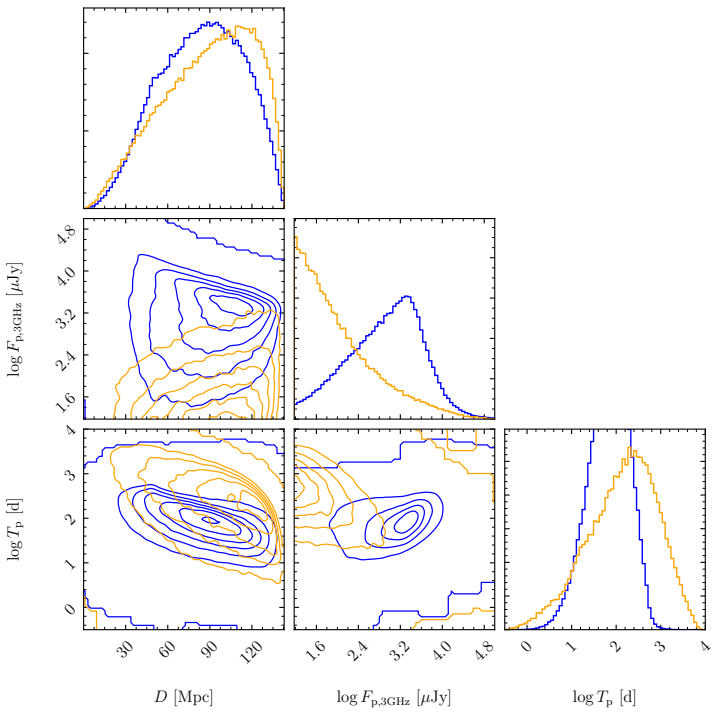


Figure 7.2: **Corner plot of luminosity distance, 3 GHz afterglow peak flux and time of peak of two populations of mergers:** one in density of 10^{-3} cm^{-3} (yellow), and another in 1 cm^{-3} (blue). Shown here are synthetic populations for radio-GW jointly detectable events as expected from the population model of Chap. 4 for the O3 run and taking the Very Large Array as the limiting radio instrument, with a 3 GHz limiting flux of 15 μJy. [Duque et al. 2020]

7.3.2 Using $n_{\rm ext}$–θ_v correlations in the sample of merger afterglows

If such follow-up observations are not done and the only available data are GW and afterglow photometry, $f_{\rm HD}^{\rm obs}$ can still be retrieved at the level of the observed sample thanks to important density-dependent correlations in the afterglow peak properties.

In Fig. 7.2, we plot the distributions of the distance, 3 GHz afterglow peak flux and peak time for two populations of mergers, in high- or low-density media. These are the distributions for the mergers predicted to be detectable jointly in GW and in the radio band by the VLA (with a limiting flux of 15 μJy) for the O3 run according to the population model described in Chap. 4 under the G16 short GRB luminosity function hypothesis. The events were placed in media with unique high ($n_{\rm ext} = 1$ cm^{-3}) or low ($n_{\rm ext} = 10^{-3}$ cm^{-3}) densities. In this case, the afterglow radiation was computed using the full synchrotron spectrum (Eq. 4.11). Synchrotron self-Compton effects were ignored in this analysis.

In particular for t_p and F_p, the distributions are qualitatively different. The low-density mergers accumulate around the limiting flux, showing that the bulk of the population is undetectable, whereas the high-density mergers present a peak at the mJy level. The combination of these population-level correlations with an adequate statistical treatment of afterglow observations should allow one to estimate $f_{\rm HD}^{\rm obs}$ for the sample. For example, one could apply a joint fit of the total sample of afterglow peak fluxes, seeking to retrieve two underlying subpopulations of high- and low-density events, each with peak fluxes distributed as in Fig. 7.2.

7.4 Constraining high-density mergers with $f_{\rm HD}^{\rm obs}$

We now illustrate our method of constraining high-density mergers starting from their apparent fraction $f_{\rm HD}^{\rm obs}$ obtained from multi-messenger follow-up campaigns, as shown in Sec. 7.3.

For the sake of illustration, suppose mergers occur in two different types of media: high-density (n_2) and low-density ($n_1 \leq n_2$). We are interested in inferring from multi-messenger BNS merger observations the intrinsic fractions $f_{\rm HD}$ and $f_{\rm LD} = 1 - f_{\rm HD}$ of mergers occurring respectively in media of densities n_2 and n_1.

For a certain electromagnetic band B, let $r_B(n_{\rm ext})$ denote the *afterglow recovery fraction* at density $n_{\rm ext}$, meaning the fraction of mergers occurring at density $n_{\rm ext}$ to produce a detectable afterglow in the B band. This is provided in Fig. 7.3 for the X-ray (1 keV), optical (r), and radio (3 GHz) bands, assuming detection limits respectively of 10^{-15} erg/s/cm^2 (50 ks exposure of *Chandra* in 0.5–8 keV band), magnitude 24 (space telescope routine observation) and 15 μJy (18 ks exposure of VLA in 2–4 GHz band). The plotted $r_B(n_{\rm ext})$ were determined from populations synthesized for the O3 run as in Fig. 7.2, but placed in media with densities that are constant within a population but varying from one population to another. Also, we give the multi-wavelength afterglow recovery fraction $r_{\rm M\lambda}(n_{\rm ext})$, which accounts for events detectable in at least one of the three bands.

Synchrotron self-absorption tends to decrease $r_{\rm 3GHz}(n_{\rm ext})$ as of $n_{\rm ext} \gtrsim 10-100$ cm^{-3}, which appears clearly in Fig. 7.3. This leads us to consider other bands (and most prominently the X-ray) for the estimation of $n_{\rm ext}$ in individual events. Therefore, we shall consider $r_{\rm M\lambda}$ as the relevant recovery fraction in what follows.

As explained in Sec. 7.1, because of the strong dependence of the afterglow peak flux to the circum-merger density ($F_p \propto n_{\rm ext}^{\frac{p+1}{4}}$), we have $r(n_1) \ll r(n_2)$. Therefore, mergers in high-density media should be over-represented in the observed population with respect to their intrinsic fraction $f_{\rm HD}$. This establishes a method to effectively constrain the latter following the observation of only a few of these high-density events.

The probability of observing a high-density merger is

$$p_{\rm HD} = \frac{r(n_2) f_{\rm HD}}{r(n_1) f_{\rm LD} + r(n_2) f_{\rm HD}}. \tag{7.3}$$

Furthermore, after observing N afterglow counterparts to GW, the likelihood that a fraction $f_{\rm HD}^{\rm obs}$ will be found to occur in a high-density medium is that of a binomial process with success probability $p_{\rm HD}$ and N tries[1]:

$$p(f_{\rm HD}^{\rm obs} | f_{\rm HD}, N) = \binom{N}{f_{\rm HD}^{\rm obs} N} p_{\rm HD}^{f_{\rm HD}^{\rm obs} N} (1 - p_{\rm HD})^{(1 - f_{\rm HD}^{\rm obs}) N}. \tag{7.4}$$

Finally, since according to Bayes' theorem with no prior information on $f_{\rm HD}$ we have $p(f_{\rm HD} | f_{\rm HD}^{\rm obs}, N) \propto p(f_{\rm HD}^{\rm obs} | f_{\rm HD}, N)$, a constraint on $f_{\rm HD}$ follows. Given the high sensitivity of the fraction $r(n_{\rm ext})$ to the density, we expect these constraints to be tight even with a small number of events.

This is clear in Fig. 7.4, where we have chosen $n_1 = 10^{-3}$ cm^{-3}, $n_2 = 1$ cm^{-3}, and we show the constraints that could be obtained from ten events among which one, three or five are in a high-density medium. We observe that the constraints do not center around $f_{\rm HD}^{\rm obs}$ and are tighter than if the bias towards high-density events were ignored, as can be seen by comparing the solid blue curves with the dotted blue curves. This illustrates the 'magnifying effect' of the selection by the afterglow.

The slope of the jet energy function is steeper for WP15 than for G16. This implies that, overall, G16 predict more high-energy events than WP15. This explains why $r(n_{\rm ext})$ is systematically larger for G16 than for WP15, at least in the regime where $F_p \propto E n_{\rm ext}^{\frac{p+1}{4}}$, that is, before the onset of the self-absorption suppression. This also implies that the rate at which afterglows are recovered by increasing the density is greater for WP15 than for G16. In terms of recovery fraction, this is expressed by saying that the contrast $\mu_{2/1} = r(n_2)/r(n_1)$ is larger for WP15 (where $\mu_{2/1} \sim 7$) than for G16 (where $\mu_{2/1} \sim 4$), which naturally leads to tighter constraints, as is clear from Fig. 7.4.

[1] Here we denote the binomial coefficient $\binom{b}{a} = \frac{b!}{a!(b-a)!}$.

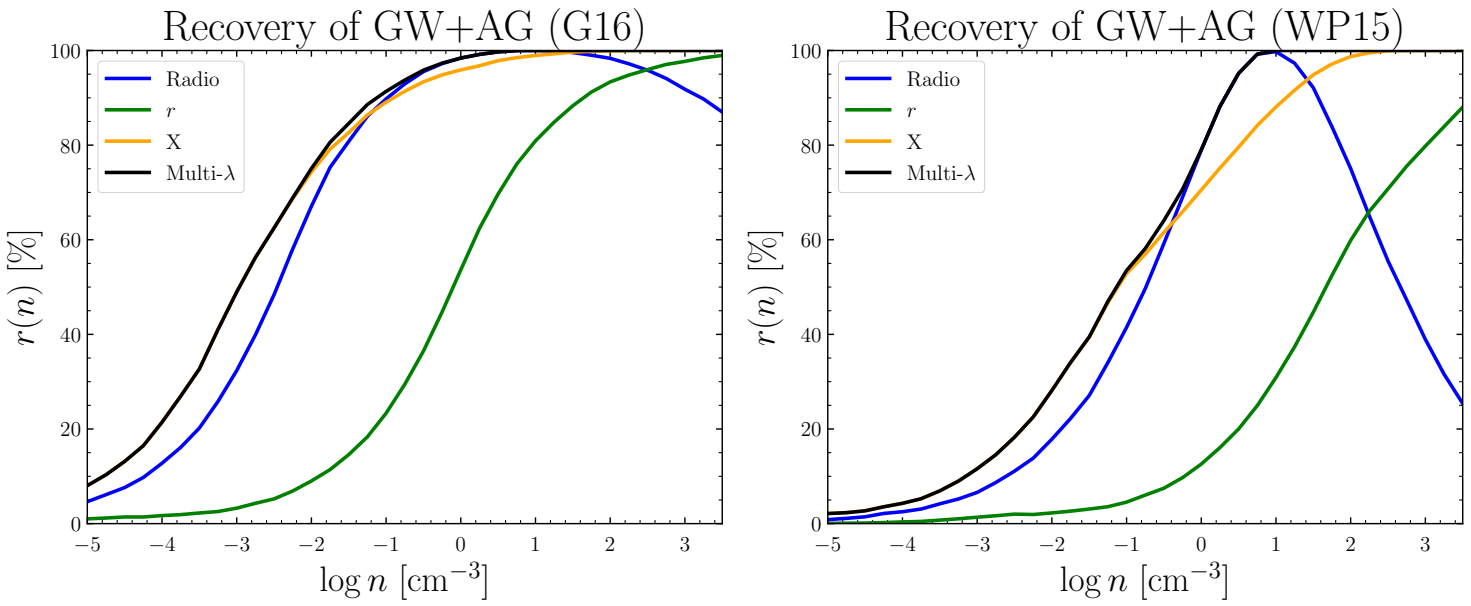


Figure 7.3: **Afterglow recovery fraction in X-ray, optical, and radio bands as function of circum-merger medium density,** for a population with energy distribution function deduced from G16 (left) or WP15 (right). We note the effect of synchrotron self-absorption on the recovery fraction in the radio band as of $n_{\rm ext} \gtrsim 10$ cm^{-3}. [Duque et al. 2020]

In the case where no high-density events are observed, upper limits on the intrinsic fraction $f_{\rm HD}$ can be deduced. This is done in Table 7.1, where we report the 95%-confidence level upper limits deduced from the observation of N events, all in low-density media. It appears that the observation of only five low-density events (e.g., observing exclusively low-density events during 18 months of an O3-type run) suffices to constrain $f_{\rm HD}$, at the 95%-confidence level, to being smaller than 18.5% (resp. 9.4%), assuming the short GRB luminosity function of G16 (resp. WP15).

Table 7.1: **95%-confidence level upper limits on $f_{\rm HD}$ obtained after observing no high-density events among N afterglows,** in two short GRB energy function distribution hypotheses.

N	1	5	10	20	50
G16	70 %	18.5 %	8.5 %	4 %	1.5%
WP15	64 %	9.4%	3.9%	1.7%	0.7%

7.5 Discussion

We have presented a method of effectively constraining the class of BNS mergers that occur in high-density media. It is based on the observation of their afterglow counterparts. We will now discuss the limitations, conditions for application and possible extensions of this method.

7.5.1 On the benefit of the GW detection criterion in our method

We have already commented that the misaligned nature of the jets expected to be observed in the multi-messenger is vital to our method, as it allows to measure the afterglow peak time, peak flux and shape factor observables necessary to Eq. 7.2. In fact, the misaligned nature of the jets also drives the physical motivation of our method, i.e., the magnifying effect of the afterglow. In Fig. 7.5, we plot the afterglow recovery fraction in various bands under the G16 population hypothesis (just as in Fig. 7.3, left), but imposing some additional restrictions on the observed population.

In the first case (left), we restrict the joint GW-radio afterglow events to being aligned, with $\theta_v = 0$. In this case, because the GW horizon is relatively close, essentially all the events are detected, regardless of the circum-merger medium density. Therefore, the contrast $\mu_{2/1} = r(n_2)/r(n_1) \sim 1$ and there is no magnifying effect at all. Without misaligned sources, the method would thus not be particularly suited to study high-density mergers.

In the second case, we drop the GW trigger constraint to study classic bright short GRBs. To this effect, we allow sources up to a distance of 6 Gpc ($z \sim 1$), all with aligned jets. In this case, we observe that radio afterglows become extremely rare, even before the onset of self-absorption. We also observe that the contrast $\mu_{2/1} \sim 1.3$ is smaller than when requiring the GW trigger and allowing for misaligned events.

Thus, studying the afterglow counterparts of GW triggers is beneficial to our method, both by allowing to measure the afterglow peak observables and by allowing to probe misaligned sources, thus increasing the high-density–low-density contrast

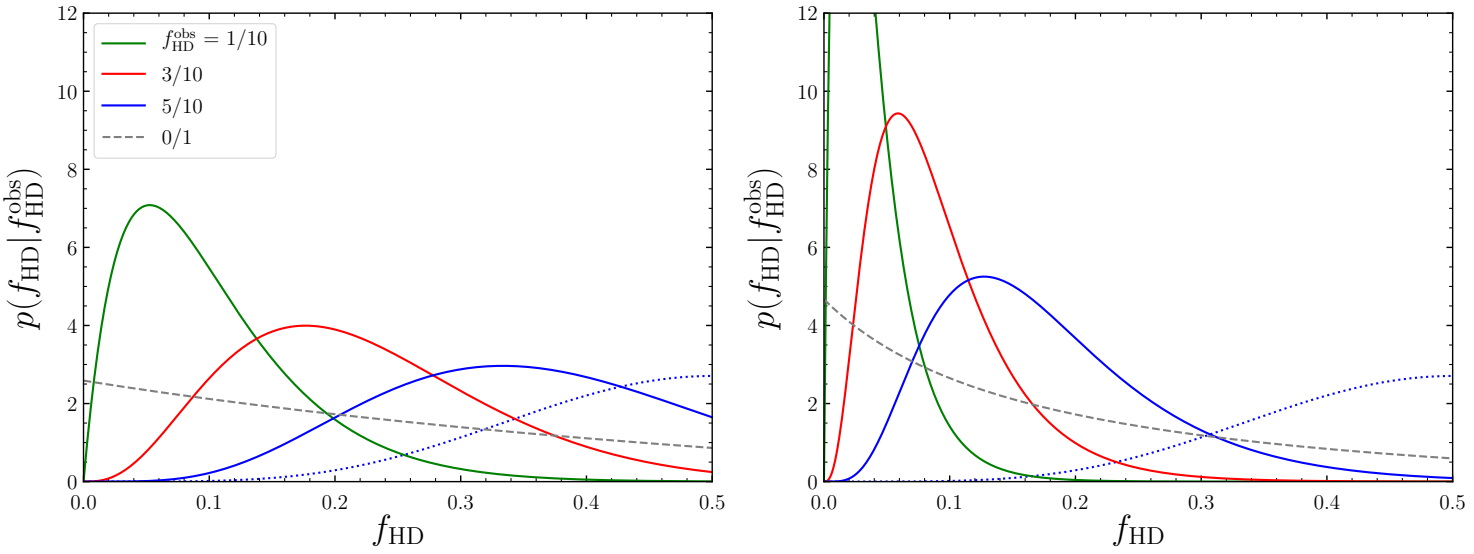


Figure 7.4: **Posterior probability density function of $f_{\rm HD}$ obtained after having observed fraction $f_{\rm HD}^{\rm obs}$ of high-density mergers among ten events,** for varying $f_{\rm HD}^{\rm obs}$. The dashed line shows the current constraint, obtained after the single low-density event GRB170817A. The dotted blue line shows the constraint obtained with $f_{\rm HD}^{\rm obs} = 5/10$, but ignoring the selection effect, i.e., with $r(n_1) = r(n_2) = 1$. Left: Assuming the short GRB luminosity function of G16. Right: Same, for that of WP15. [Duque et al. 2020]

and driving the magnifying effect.

7.5.2 Method limitations and applicability

A first limitation of the method presented here is the requirement that the sample be density-complete above a certain afterglow flux. In other words, it requires the certitude that all detectable afterglows with fluxes above a limit were effectively detected. Only in this case can the model-determined recovery fraction $r(n_{\rm ext})$ be used to infer $f_{\rm HD}$ from $f_{\rm HD}^{\rm obs}$. As the observational biases resulting in practical limitations to these detections were discussed in Sec. 5.7, we do not repeat them here. We only mention that the difficulty in following-up GW events linked to the size of the localization sky-maps should be met by large-field facilities such as the Zwicky Transient Facility (Bellm et al., 2019), and by future high-cadence survey instruments such as the Large Synoptic Survey Telescope (Ivezic et al., 2008). In practice, density-completeness will be difficult to obtain, and an uncertainty on $f_{\rm HD}^{\rm obs}$ must be taken into account in applying this method.

Furthermore, there is a selection bias towards high-density mergers for reasons unrelated to the afterglow flux. For instance, afterglows of mergers occurring in denser media should peak at earlier times, favoring their detection during follow-up, regardless of their flux level. Consequently, the flux-related selection bias we quantified here in $r(n_{\rm ext})$ actually underestimates the bias toward high-density events.

Similarly, there is a selection bias towards bright afterglows regardless of the events' circum-merger density. For instance, events closer or brighter in gamma-rays should be better localized by the GW or GRB data, easing their follow-up, regardless of the circum-merger density. These density-unrelated biases towards afterglow detection actually correlate positively with afterglow flux and thus, statistically, with density. Therefore, once again, the bias towards high-density events we consider here is underestimated.

This method is not applicable to the population of cosmological short GRBs for which densities have been estimated, for three main reasons. The first is that the densities claimed for this population are deduced from uncertain fits, as argued in Sec. 7.2, and that only a small fraction of GRBs have a claimed density. Thus, the resulting $f_{\rm HD}^{\rm obs}$ would be quite uncertain. The second is that, for these regular short GRBs, the afterglow detectability depends more on factors that are not density-related, such as (i) the availability of sufficiently rapid follow-up observations and other human factors, or (ii) the quality of the localization of the GRB, which is linked to its prompt properties and not to its afterglow. Three, for regular GRBs, we showed the contrast of recovery fraction is much smaller and thus our method would be less efficient.

In Sec. 7.3, we mentioned the kilonova afterglow as an alternative means of measuring the merger environment density regardless of the viewing angle, as allowed by the quasi-isotropy of this signal. Nonetheless, we caution against the feasibility and robustness of such a measurement. First of all, as shown by particle-in-cell simulations and the observation of young supernova remnants (Crumley et al., 2019; Morlino & Caprioli, 2012), mildly or non- relativistic shocks are expected to be poor electron accelerators, with ϵ_e up to orders of magnitude lower than in relativistic shocks. Therefore, seeing as the afterglow flux scales with ϵ_e, the kilonova afterglow should be significantly fainter than the jet's afterglow and unlikely to be

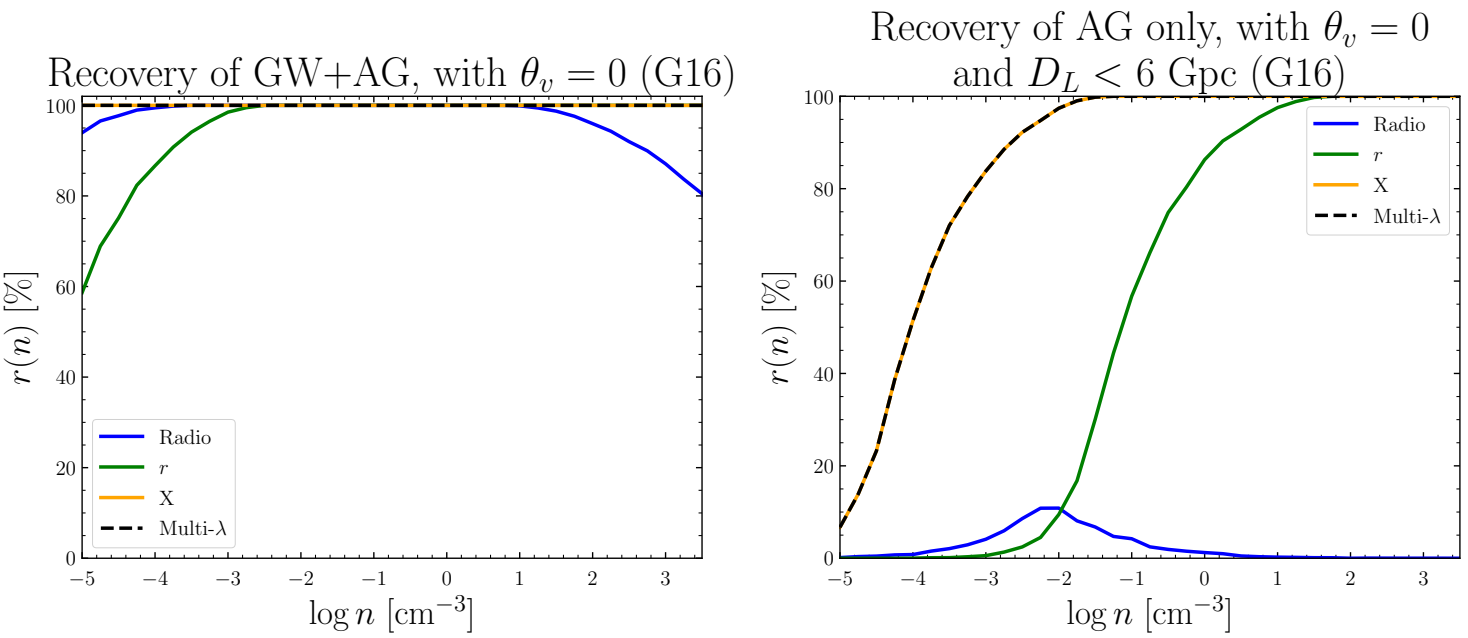


Figure 7.5: **Afterglow recovery fraction;** analogous to Fig. 7.3, but imposing the system to be aligned (left) or allowing systems up to cosmological distances (right). In the latter case, no criterion for GW detection was applied. In both cases, the contrast $\mu_{2/1} = r(n_2)/r(n_1)$ is smaller than imposing the GW detection and allowing for all viewing angles. The G16 population model was used.

detectable in most cases. Furthermore, in the typical case of a low-density medium, this signal is expected to peak up to a decade post-merger, posing some challenge to its detection in follow-up campaigns. Finally, the kilonova afterglow light curve depends on the minimal velocity of the merger ejecta and on its radial velocity structure, both of which are still uncertain for lack of modeling and observation history. Therefore, although the kilonova afterglow signal's quasi-isotropy dismisses the degeneracy between the density and the viewing angle, its use introduces some uncertainty to the measurement, which is thus rendered not robust.

However, if we were able to better understand the structure of the kilonova afterglow and the physical conditions in its forward shock, then the kilonova afterglow counterpart could effectively contribute to constraining the environments of BNS mergers. In Chap. 11, we propose a numerical simulations project to study the dependence of the kilonova outflow structure to the progenitor's parameters. It would be a first step to obtaining a handle on the kilonova afterglow as a BNS merger counterpart.

7.5.3 From mergers in dense environments to fast-merging and low-kick binaries

In Sec. 7.1, we presented this method of determining $f_{\rm HD}$ as a first step towards constraining the population of fast-merging binaries required to explain various astrophysical data, as summed-up in Sec. 7.2.

First of all, an astrophysically interesting constraint on the densities of circum-merger media should be given as a continuous distribution of densities within the population, and not only as a fraction of high- and low-density mergers as we have shown here for simplicity. A continuous (parametric) distribution of densities does not pose any mathematical problems and could be included in this method.

Second, constraining the distribution of merger delay times from that of the merger environment densities is non-trivial, because the medium hosting the merger effectively depends on the locus of the second supernova in the galaxy, on the kick it imparts to the binary system, on the galactic potential, and on the galactic density profile. All of these are uncertain or variable from system to system. In Chap. 11, we propose a project to tackle these issues on the level of population synthesis.

Third, our method relies on observing the afterglow counterparts to GW inspiral signals, and thus can only inform us on the high-density mergers within the horizon of the GW instruments. However, the fast-merging binary population suggested by the r-process element observations mentioned in Sec. 7.2 must have formed and enriched their hosts shortly after the peak of cosmic star formation, that is, at $z \sim 2$. Thus, this method will remain ineffective with regards to this particular population, as long as we rely on second-generation GW instruments. However, with the prospect of detecting inspiral signals from systems at $z \gtrsim 1$ with third-generation interferometers such as the Einstein Telescope (Punturo et al., 2014) or Cosmic Explorer (Reitze et al., 2019), the constraining power of this method becomes larger and extends to the redshifts where fast-merging binaries are a matter of debate. In this context, a complete description should require a redshift-varying fraction $f_{\rm HD}(z)$, the addition of which is a possible extension of our

method. At these redshifts, however, detection of the kilonova may reveal challenging and the localization of the source needed for multi-wavelength follow-up should be ensured directly by detection of the afterglow by wide-field X-ray instruments such as Theseus (Amati et al., 2018) or radio survey facilities such as the Square Kilometer Array (Braun, 2014; Dewdney, 2015).

7.6 Conclusion

We have described a method of directly probing the binary neutron stars that merge in dense environments, based on the expected observation of afterglow counterparts to misaligned binary neutron star mergers and exploiting the high sensitivity of these to the circum-merger medium density. Its constraining power is large and, since high-density mergers are naturally associated with fast-merging or low-kick binaries, this method is a first step toward a new independent approach to the delay-time and kick velocity distributions based on binary neutron stars.

Part III

Back to gamma-ray burst physics: Consequences of misaligned structured jets

Chapter 8

Introduction to structured relativistic jets

Abstract

A relativistic outflow's physical quantities—typically kinetic energy density or Lorentz factor—may vary with the angle from the outflow's axis. These outflows are referred to as *structured relativistic jets*. Such relativistic jets are ubiquitous in high-energy astrophysics settings on a large range of length scales: from micro-quasars to active galactic nuclei. The afterglow observations of GW170817 revealed that such a jet was launched from the merger, and analyzing the multi-messenger data allowed the most exquisite insight yet on the structure of the outflow from a binary neutron star merger. Structured relativistic jets are now an essential ingredient of modeling efforts for gamma-ray bursts and other high-energy events. In this chapter, we reintroduce relativistic jets in a broader context, outlining the physical processes that shape their outflows. In the case of gamma-ray bursts, we summarize observational evidence for the existence of these jets and reproduce an argument showing that the bright bursts are observed at most at slightly misaligned lines of sight to their jets. We discuss the possible origins of the structure of gamma-ray burst jets, leading us to study the interaction of the jet launched by the central engine with the thick layer of material enshrouding the engine: the collapsar's envelope for long bursts and the merger ejecta for short bursts. Finally, we derive the analytic framework to determine the flux from a radiating structured jet as a function of the outflow structure in emitted energy and Lorentz factor.

8.1 Introduction

In this part of the thesis, we turn to the second class of studies that GW170817 motivated in my PhD work: drawing from the lessons of GW170817 to develop new models for GRBs. One important lesson that we will focus on throughout Part III is that the relativistic jets launched in BNS mergers—and likely also in long GRBs from collapsars—possess structure. How this lesson came about in the unraveling of GW170817 was described in detail in Sec. 3.3; The robustness of this conclusion and its integration in our population model was discussed in Sec. 4.4.5.

We define a *structured relativistic jet* as a relativistic outflow of matter and energy that possesses a preferred direction—the jet axis—along which the out-flowing material has the most extreme energy and velocity properties of the outflow. The material's properties depend on the angular distance to the axis, and are significantly lower around than along the axis. While this definition excludes some outflow structures found in numerical simulations—with, e.g., non-decreasing energy from the axis outwards, Sec. 8.3—, it applies to the case of GW170817, and we will show that the formation of the jet structure naturally favors a decrease of material properties from the axis. This definition formally contrasts with that of a *radially stratified outflow*, in which the outflow is better described as spherically symmetric, with however a radial gradient of material velocity.

This type of structure was an antagonist to the structured relativistic jet hypothesis in the early interpretation of the afterglow of GW170817. It was suggested that both could have occurred, according to whether the jet launched by the central engine was "choked" by or successfully broke out from the merger ejecta (Sec. 3.2.3). The imagery of the remnant eventually strongly favored a structured relativistic jet (Sec. 3.3); The complete photometry now supports this. Naturally, astrophysical outflows possess a combination of radial and angular structures, and mildly relativistic spherical outflows could be favored in some cases, e.g., for low-luminosity GRBs (Sec. 1.4). In this chapter, we will focus on structured relativistic jet outflows, study their diverse manifestations in high-energy astrophysics and open the way to using them for GRB modeling.

It is likely that in astrophysical structured jets, the outflow properties do not decline right from the jet axis, but remain somewhat constant over an extended region, referred to as the *core jet*—vocabulary already introduced in Chap. 4. The transition from the core jet to the actual structure manifests as the

jet break for an aligned observer (Sec. 1.2). For a very misaligned observer, the core jet imprints the afterglow light curve with the extended turnover time interval from the increasing to the decreasing phase (Sec. 7.3.1, Appendix I).

Because of the relativistic motion and presence of the preferred direction in the outflow, the observation of radiation from the outflow crucially depends on the position of the observer with respect to this direction. Beyond the actual emission mechanism in the outflow, the observed radiation will be shaped by the purely geometrical effects which are Doppler boosting and beaming, and light travel time delays from different regions of the outflow.

As detailed in Sec. 3.3, these effects were at work in GW170817, which was observed at an angle of about five times the central core's opening angle (Nakar & Piran, 2021): At early times, the beaming of the core disallows to observe radiation therefrom; As the outflow decelerates, regions closer and closer to the core progressively debeam and contribute to the flux; Because these are more and more energetic due to the structure, the flux increases; When the core material has slowed down and debeamed, it appears; It is revealed with a smaller flux than for an aligned observer precisely because of the debeaming; After the core is revealed, essentially the entire outflow contributes and the flux declines because the intrinsic emitted power decreases, just as for an aligned observer.

In this Part III, we will study how these same geometrical effects can combine with the structure of GRB jets in order to explain recurring features of the X-ray afterglows of bright GRBs: plateaus in Chap. 9 and flares in Chap. 10. Though the geometrical effects in these new models for plateaus and flares are the same as witnessed in GW170817, they will apply in a different setup where the observer is only slightly misaligned with the jet and at a cosmological distance. Before we present this new modeling, the present Chap. 8 motivates and presents structured relativistic jets in a broader astrophysical context.

In Sec. 8.2, we summarize pre-GW170817 evidence for the existence of structured relativistic jets in GRBs and other sources. In Sec. 8.3, we overview current knowledge on the possible origins of these structured outflows. In Sec. 8.4, we lay down the analytical framework to determine the radiation detected by a misaligned observer as a function of the radiation emitted from the outflow.

8.2 Observational manifestations of structured jets

8.2.1 Signatures from jet–ejecta interactions in GRB–supernova associations

Long GRBs were early associated with supernova (SN) explosions through their host galaxy types and strong star-forming activities and their rate evolution with redshift; This connection was confirmed by historic observations of GRB–SN associations (Sec. 1.3). When observed, supernova counterparts to long GRBs are systematically of the Ic broad-line type (Woosley & Bloom, 2006): core-collapse explosions that exhibit mildly relativistic expansion velocities (15 000–30 000 km/s; e.g., Modjaz et al. 2016) and kinetic energies up to ten times larger than most core-collapse SN (Mazzali et al., 2017). A tentative interpretation of these facts is the signature of the relativistic jet interacting with the SN outflow. As we will discuss in Sec. 8.3, such an interaction is destined to produce a structured outflow, whether the jet successfully breaks out from the SN material or not. However, all Ic broad-line SN do not present a long GRB counterpart, and the concurring effects of the beaming in the case of successful jets and the possibility that some jets fail to emerge from the SN hinder a clear identification of broad-line Type Ic SN and long GRBs to the same phenomenon (Japelj et al., 2018).

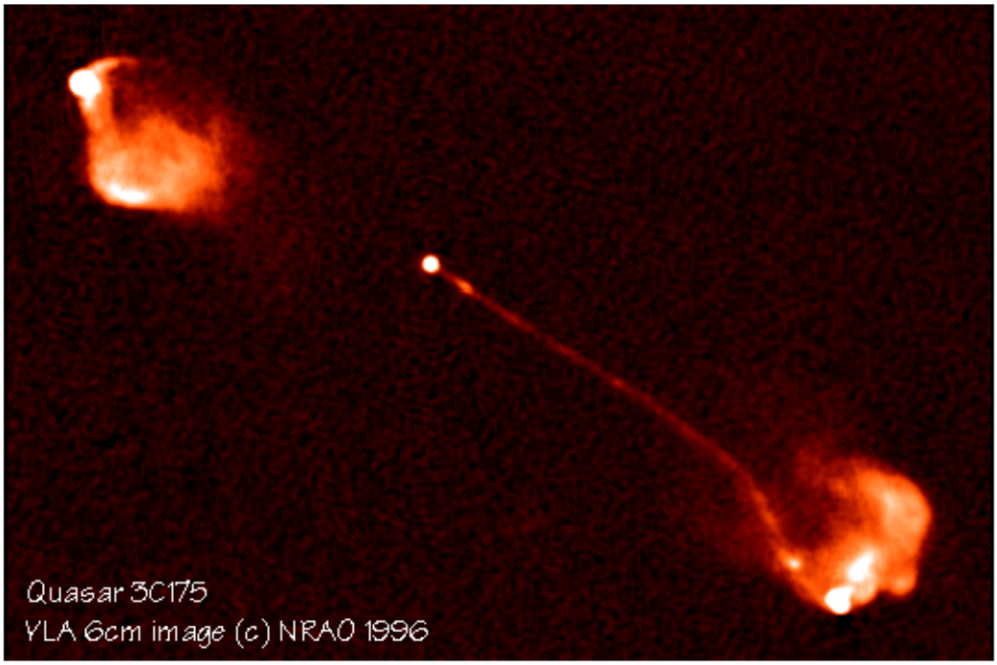
8.2.2 Structured jets in micro-quasars and AGNs

Relativistic jets are present in many astrophysical contexts. In micro-quasars, jets are launched from the accretion disk around a compact object, which is fed by a companion star. The interaction of the jetted outflow with the companion's wind structures the jet in a non-trivial manner, leading to observable signatures in the high-energy emission from the source (e.g., Molina et al., 2019). This interaction leads to *bent jets*, as they are influenced by unequal pressure on each side in the wind outflow. High-mass X-ray binaries are particularly subject to these jet–wind interactions.

This structure is further complicated by the orbital motion of the compact object and the eventual precession of the accretion disk, which can imprint a helical shape to the jet, as in SS433 (Monceau-Baroux et al., 2014). In some sense, these jets possess structure in addition to a variable preferred direction, on the contrary to the transient jets of GRBs. These complex geometries pose a radiation-transfer challenge in predicting the orbital modulation of the light curve or spectra of these sources. Moreover, the flux of these sources is composed not only of radiation from the jet, but also of the up-scattering of the companion's radiation field to high energies (Dubus et al., 2010), adding further complexity to the problem. Fortunately, these complications due to source binarity do not occur in GRB physics.

On much larger scales, jets from Active Galactic Nuclei (AGN) are rather subject to instabilities in the outflow itself or interaction with the immediate propagation medium (Hardee, 2008). Most prominently, these are the Kelvin-Helmholtz instability at the steep velocity gradient on the edge of the jet, and current-driven instabilities in the case of highly magnetized jets. These contribute to decollimating the jet and can imprint a helical structure to the outflow Hsu & Bellan (2002). AGN jets are observed to remain collimated on very large distances (FR II types, such as 3C175; Fig. 8.1, left; Bridle et al. 1994), or on the contrary to fail to instabilities and decollimate promptly (FR I types, such as 3C31; Fig. 8.1, right; Laing et al. 2008).

The various morphologies of AGN jets show the role that these instabilities can play in the physics of relativistic jets; In fact, the shape of a given jet can be traced to its degree of resis-



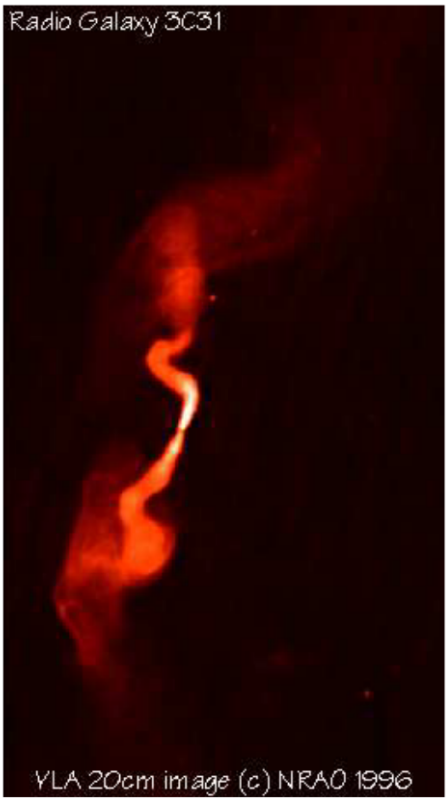


Figure 8.1: **Two morphologies in AGN jets.** Left: 3C175, displaying the FR II morphology that maintains collimation on a large length scale. Right: 3C31, with FR I morphology, decollimates because of jet–external medium interface instabilities (Bridle et al., 1994; Laing et al., 2008).

tance to these instabilities and therefore be used to constrain its physical conditions. The Kelvin-Helmholtz and magneto-hydrodynamical (MHD) instabilities are only starting to be accounted for in the GRB field (see Sec. 8.3).

In both microquasar and AGN jets, the large-scale structures are in part determined by the physics in the smaller scales of the jet launching mechanism: accretion onto the compact object, jet collimation and development of disk instabilities.

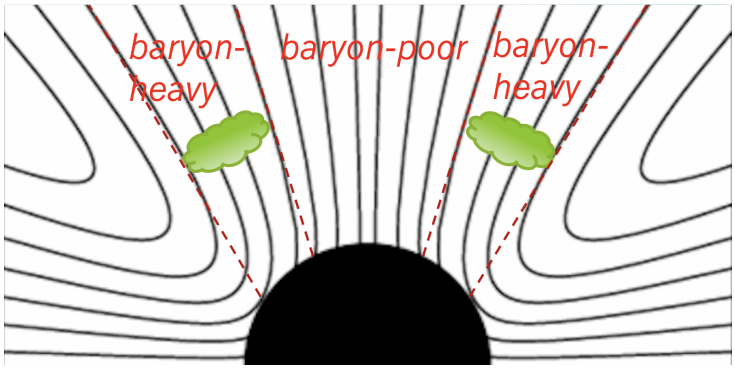


Figure 8.2: **An early example of jet structure in GRB modeling.** The baryon loading puzzle is solved by the threading of magnetic field lines through the central black hole. The resulting outflow is structured in baryon content.

8.2.3 Structured jets in GRBs prior to GRB170817A

Historically, structured relativistic jets can be traced—though unnamed—to the early days of GRB science, even before the first afterglow observation. For example, Levinson & Eichler (1993) give a solution to the baryon loading puzzle (Sec. 1.2) by the threading of baryon-loaded magnetic field line through the central engine of GRBs, a black hole. In this picture schematized in Fig. 8.2, the central field lines of the outflow are baryon-poor as required while the surrounding material is less pure because it does not thread the black hole. The outflow is structured in baryon content.

Shortly after the discovery of afterglows, Mészáros et al. (1998) explored the expected diversity in this signal assuming the possible misaligned lines of sight to structured jets. At a time when the forward-shock origin of the prompt emission was not excluded, they suggested light travel delays in a misaligned-observer scenario to explain the offset between prompt emission and afterglow.

Structured jets and their geometrical effects were regularly introduced to explain both the dim prompt and afterglow features of low-luminosity GRBs (Sec. 1.4). The cases of GRB980425 (Waxman, 2004) or GRB031203 (Ramirez-Ruiz et al., 2005) are emblematic and prompted the idea that a wide variety of high-energy phenomena from X-ray flashes, to these low-luminosity GRBs could be manifestations of a same object—relativistic structured jets—seen from different viewing angles (Granot et al., 2005; Ramirez-Ruiz et al., 2005; Granot & Ramirez-Ruiz, 2010). These unification schemes however meet with statistical issues due to the much larger occurrence rates of low-luminosity events than expected (Sec. 1.4, Soderberg et al. 2006).

Some authors go even further in suggesting a single *universal structure* for the jet in all these events, attributing statistical properties of GRBs such as the luminosity function or correlations to the exploration of this structure at different viewing angles (e.g., Salafia et al., 2015). The explanation of correlations by the exploration of viewing-angle space is a recurring theme of the structured relativistic jet scenario in GRB modeling; In our own models of Chaps. 9 and 10 we shall illustrate this.

GRB unification schemes based on angle effects are elegant and have analogs in other branches of astrophysics—prominently for AGNs. Also, they corroborate our earlier discussion on the ill-definiteness of the GRB sample by replacing these events in a continuum of phenomena. However, recent developments on the physics of the prompt emission from misaligned lines of sight show that the transformation from a viewing angle to another is not simply geometrical; Recall, e.g., the discussion of the shock breakout mechanism and the misaligned version of the compactness puzzle in Sec. 3.4.

The discovery of the peculiar behaviors of GRB afterglows after the launching of *Swift* (Sec. 1.2) further motivated modeling based on structured jets. Eichler & Granot (2006) suggested an interpretation of plateaus in X-ray afterglows based on *slightly misaligned* lines of sight to structured jets. In this model, geometrical effects—similar to those involved in GW170817's afterglow—allow the observer to explore the structure up to the core, producing the plateau phase. In Chap. 9, we will show how the work accomplished during the PhD enhances this model with analytical estimates of plateau features such as flux level and duration, as well as a confrontation to statistical data from the numerous catalog of plateaus observed by *Swift*.

Afterglow counterparts from significantly misaligned jets are announced for the multi-messenger era (Sec. 5.7) and such a possibility has sparked much interest in the influence of the jet structure in this setting. However, when approaching bright GRBs, only slightly misaligned lines of sight can be considered. Indeed, Beniamini & Nakar (2019) have shown evidence that the outflows from these GRBs can only efficiently produce gamma-rays up to angles $\theta_\gamma \lesssim 2\theta_j$. Here, θ_j is the half-opening angle of the core and θ_γ is the angle up to which efficient gamma-ray production occurs, such that bright GRBs are all observed from angles $\theta_v \leq \theta_\gamma$. They considered the ratio between the energy emitted in the early X-ray afterglows of long GRBs and the energy emitted in their prompt phase. In the selected sample of GRBs, this ratio is found to vary little from a burst to another. Having made the observation on the early X-ray to prompt ratio, the authors study which structure in outflow energy and Lorentz factor between θ_j and θ_γ are consistent; Since they are interested in a structure, either the material's Lorentz factor or energy must significantly decline in this interval.

First, it could be that the Lorentz factor declines signif-

icantly. This would imply that the material down the observer's line of sight has a large deceleration radius and a small Lorentz boost. This would lead to the observation of this material's afterglow before it decelerates, and thus very extended near-flat-flux phases in afterglows, which are not observed. By very extended, we mean longer than days, i.e. more than the thousand-second near-flat fluxes phases that are indeed observed—plateaus, see Chap. 9. The smaller boost of the material would also lead to a smaller early afterglow flux, in contrast with the observed clustering in early-X-ray–prompt flux ratio. Thus, the Lorentz factor structure cannot be steep between θ_j and θ_γ.

Second, it could be that the material's energy declines significantly. Since the Lorentz factor structure is shallow, all the material in the θ_j–θ_γ interval is ultra-relativistic and the observed prompt energy from a given viewing angle is essentially that emitted at the same latitude (await Sec. 8.4 for proof). In this case however, the angular exploration of the structure would over-produce weak GRBs in comparison with inferred luminosity functions. Bearing in mind the large uncertainties in these luminosity functions—for short GRBs especially but also for long GRBs—, a somewhat robust conclusion that the energy structure cannot be steep also holds.

Excluding these two possibilities, Beniamini & Nakar (2019) thus conclude that *there is no steep structure* between θ_j and θ_γ: i.e., $\theta_\gamma \sim \theta_j$. The quantitative analysis leads to the constraint that $\theta_\gamma \lesssim 2\theta_j$ for long bright GRBs.

This analysis is limited by uncertainty in the long GRB luminosity function and by the underlying assumption that the efficiency in gamma-ray dissipation itself does not vary from θ_j to θ_γ. Also, it does not hold for short GRBs, afterglow data for which is very scarce. Nonetheless, this reasoning provides some evidence that all bright GRBs are observed from only slightly misaligned lines of sight: θ_v less than a few θ_j. We will adopt this hypothesis in our modeling of bright GRBs afterglow plateaus and flares, in Chaps. 9 and 10.

8.2.4 Structured jets in GRBs since GRB170817A

Before moving to the origin of jet structure, we simply note that GW170817 has sparked a renewed interest in studying the influence of jet structure on GRB jet afterglow observations.

Beniamini et al. (2020b) classified the possible morphologies of afterglow light curves from significantly misaligned structured jets. For GW170817, the afterglow light curve was single-peaked, as predicted from the reasoning we presented in Sec. 8.1. Essentially, the authors outlined the parameter space of viewing angle, energy and Lorentz factor structures where such a reasoning holds and single-peaked light curves are expected.

For shallow Lorentz factor structures or large viewing angles, the material directly down the observer's line of sight can produce its own complete afterglow signal before the core jet is revealed. In such configurations, the observer detects an afterglow light curve with an increasing phase, a peak and the beginning of a declining phase before receiving radiation from the decelerating structure and eventually from the core jet: The light curve is thus double-peaked. For steeper Lorentz factor structures, the observer receives radiation from material closer to the core from the start, and the exploration of the jet up to the core is not delayed: The light curve is single-peaked.

The morphology of an afterglow light curve can be predicted by determining the critical angle θ_* defined by the lowest latitude of the material that is *initially* debeamed from the observer: If $\theta_* \sim \theta_v$, the afterglow is predicted to be single-peaked; If $\theta_* \ll \theta_v$, the afterglow should be double-peaked. For the viewing angles to upcoming events suggested by our population model (Sec. 5.6) and the structures observed for GW170817 (Sec. 3.3), single-peaked afterglows are expected.

Other examples include Lamb et al. (2021), who study the combined influence of jet structure and core lateral spreading on the exact shape of afterglows at the jet break transition for aligned observers. They quantify the sharpness of the break as a function of the smoothness of the core-structure connection, and apply their model to constrain the smoothness of the core jet edge for a sample of GRBs.

From the point of view of analytical studies, exploring the influence of jet structure on GRB prompt emission has stayed on the level of prescribing the angular dependence of the parameters of a given emission model (e.g., for shock breakout radiation, Beniamini et al. 2019) or simply prescribing the angular dependence of gamma-ray dissipated energy, with no particular mention of the underlying emission mechanism (Sec. 8.4). We have yet to study the influence of the jet structure on the actual physical conditions in the jet and thus on the emission mechanism. In Chap. 11 we will propose to explore these questions.

8.3 Origin of jet structure

When interested in the radiation observed from a structured jet, the relevant structure is the one at time of dissipation. While apparently trivial, this statement pertains to the fact that the structure of the jet naturally evolves with time. Beyond the already noted evolution due to core jet spreading and differential deceleration of the structure (Sec. 4.4.5), the formation and early evolution of the jet leads to the question of the origin of the jet structure: Was the outflow *launched with structure* by the central engine or did the outflow *acquire the structure* through some inherent process or through interaction before energy dissipation?

Once again, the electromagnetic observables of prompt or afterglow radiation depend only on the final structure of the jet. However, the evolution of the outflow bridges the gap between the observed radiation and the jet launching mechanism near the central engine and thus to the prior phases of the event: stellar collapse or compact object merger. Studying the physical origin of the jet structure is therefore essential to establish a global picture of the GRB phenomenon.

We refer to the outflow launched by the central engine as the *incipient jet*; It is the result of the launching phase occurring after formation of the compact engine and defined by

the MHD conditions thereabout. This incipient jet will evolve through mechanisms that we will briefly overview in the next paragraphs into the *emerging jet*, the structure of which determines the observed radiation.

8.3.1 Jet collimation and jet emergence

In the context of GRBs, the first evolution phase of the incipient jet will involve interaction with a slowly moving, dense layer of material: the massive progenitor's envelope for long GRBs; the merger ejecta for short GRBs. Before studying how this interaction imprint the structure of the jet, it is important to figure out whether the incipient jet will be collimated while penetrating this layer and upon emerging. A collimated jet is a monodirectional outflow, with sides defined by a cylinder; A decollimated jet has adopted a conical structure or, more extremely, has started to expand laterally. Intuitively, a luminous or thin incipient jet will lose little energy to the layer which will be unable to pressure the jet from a conical to a collimated shape. This will favor the ultimate emergence of a relativistic core similar to the incipient jet, surrounded by a lateral structure shaped during the traversal: a *successful jet*. Conversely, a less luminous or wider incipient jet has more interaction cross section with the layer. It will loose more energy and can be collimated before emergence, loosing its initial conical structure. For very dim jets, the final outflow can end up only mildly relativistic, with only mild dependence of outflow properties to the angular distance to the outflow axis: a *chocked jet*.

One early analytical approach to determine whether a jet is collimated in crossing the dense layer overhead can be found in Bromberg et al. (2011), of which we reproduce the physical setup in Fig. 8.3. The incipient, conical, supersonic jet pushes a forward shock where layer material is shocked, and a reverse shock runs up the jet towards the central engine (Fig. 8.3, right). The over-pressure in both the shocked regions lead shocked material to evacuate laterally, forming inner and outer cocoons; These in turn exert pressure on the jet's sides. Depending on the jet's luminosity, the layer's density and the incipient jet's opening angle, the pressure balance (and causal connectedness) between the jet, the shocked regions and the cocoons can lead to the collimation of the jet (left) or not (center). The collimation condition will define the position of the *collimation shock* (dotted line within the jet): If collimation is effective, the cocoon–jet pressure balance imprints a cylindrical shape to the jet. Then, the convergence of the flow lines in the jet flow will increase the flow velocity, up to a sonic point within the jet, where the collimation shock forms. If collimation is not effective, the flow line convergence is not enough to form a sonic point within the jet, and the collimation shock is not closed within the jet.

Quantitatively, these regimes are separated by the following collimation criterion (Bromberg et al., 2011, Eq. 27):

$$\frac{L_j}{\pi z_h^2 \rho_a c^3} \leq \theta_0^{2/3} \tag{8.1}$$

where L_j is the incipient jet's luminosity, z_h is the thickness of the dense material layer—in other words, the distance the jet must cover inside the dense material before breaking out—and ρ_a its mass density. θ_0 is the incipient jet's opening angle.

This analysis is valid only for non-magnetized jet, and these analytical scalings are now supported by hydrodynamical simulations (e.g., Gottlieb et al., 2021). Specific analytical work and numerical simulations were applied to the conditions for the emergence of jets from BNS mergers and, in the case of non-emergence, the breaking out of the shock driven by the jet in the ejecta (e.g., Duffell et al., 2018). Such estimates played a important role in early takes on the emergence of a jet in GW170817 (Sec. 3.3).

8.3.2 Imprints from jet–cocoon interaction

Once it has emerged from the dense ejecta, the structured jet penetrates the far less dense circum-burst medium, in which we have studied the jet's deceleration and spreading (Sec. 4.4.5 and associated appendices). The jet structure upon emergence is thus defined primarily by the jet's prior interaction with the dense material.

GW170817 sparked a renewed interest in numerical simulations to determine the impact of this interaction on the final jet structure. For hydrodynamic jets, it is found that the asymmetry and jet-edge instability features mentioned above for microquasars and AGN jets (Sec. 8.2.2) can play and important role for GRB jets, at smaller length scales.

Gottlieb et al. (2021) identified the role of the Rayleigh-Taylor instability at the jet-cocoon interface in mixing material from the two structures. This leads to loading of the jet and energy losses to the cocoon. Further, a trend was found of a systematic power-law structure on the outflow properties in the immediate vicinity of the core of the emerged jet, lending some credence to the power law structures we will adopt in Chaps. 9 and 10.

By launching a top-hat jet (i.e., a bare core jet) into a setup resulting from a hydrodynamical simulation of the merger of two BNSs, Pavan et al. (2021) have also identified the Kelvin-Helmholtz instability at the jet-cocoon interface. Their simulations profit from a full 3D setup to find that these instabilities alone can imprint departure from axisymmetry of the emerging structured jet. This effect combines with an imprint of the ambient medium's asymmetry onto the jet's propagation, resulting in significant asymmetry of the final structure: From one side of the jet to the other, at the same latitude below twice the core, the Lorentz factor or energy density can be within a factor $\gtrsim 2$. In one case, they find that the outflow energy is not even a strictly decreasing function of the latitude. While such features should be confirmed by further work, they advise caution when using axisymmetric jet structures in GRB modeling.

The presence of an ambient magnetic field can significantly change the physical processes at play, and its topology adds a further degree of freedom. Gottlieb et al. (2020) found that a toroidal magnetic field stabilizes the jet against the jet-external medium interface instabilities mentioned above. As these instabilities influence the final structure and thus its afterglow, they suggest to use GRB observations as probes of the ambient magnetic fields around the central source, especially for long GRBs

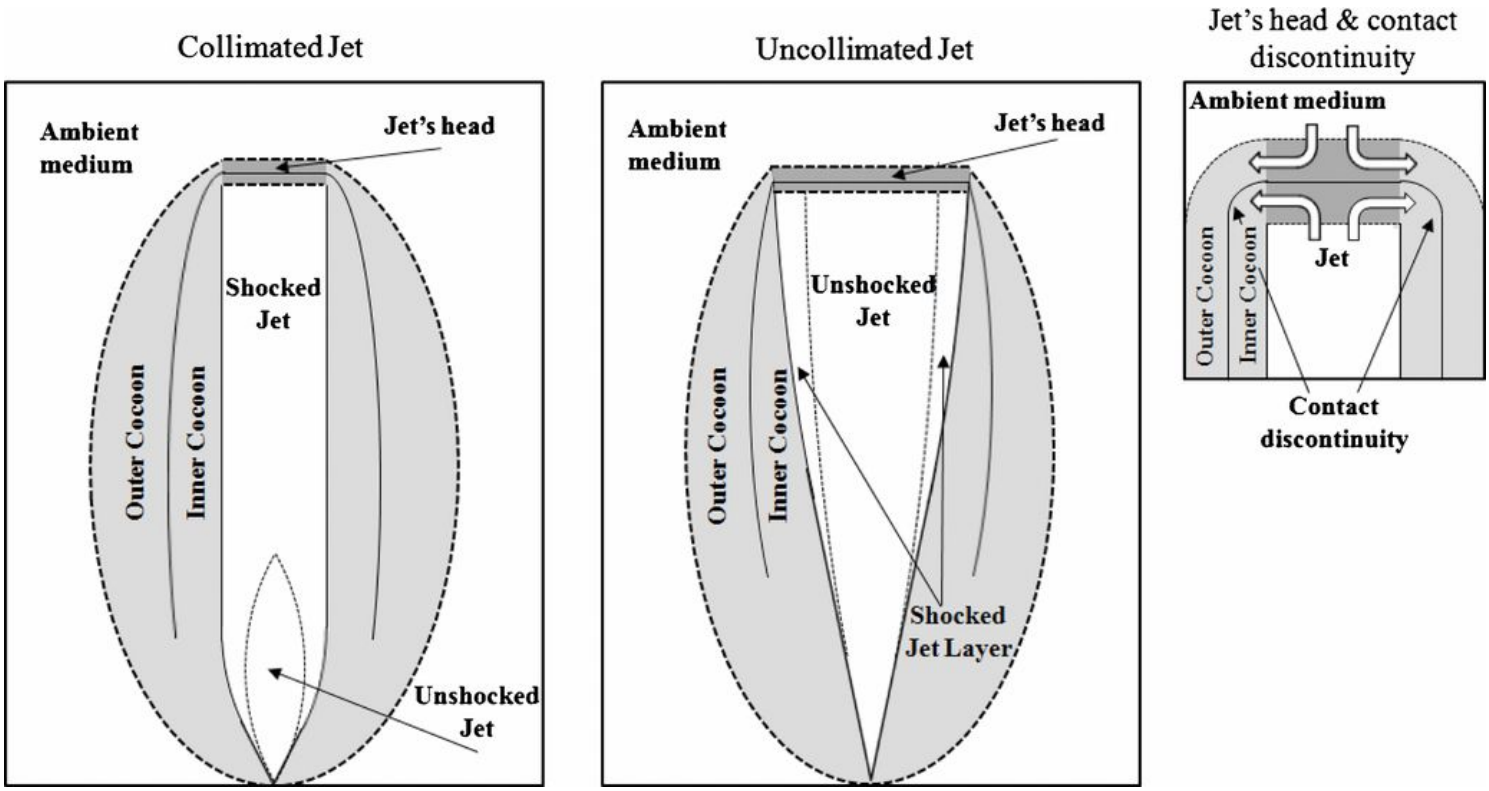


Figure 8.3: **Physical setup for jet collimation while penetrating the dense material around the central engine.** The overpressure at the jet's head pushes shocked jet and ambient matter to the sides, forming cocoon regions. These exert pressure to the jet, and can possibly collimate it depending to the causal connectedness and pressure balances between the jet and the cocoon (Bromberg et al., 2011).

for which the jet-cocoon mixing is stronger.

Nathanail et al. (2021, 2020) analyzed the emerging jet structure for a broader set of magnetic field topologies and with jets launched self consistently from the central engine-accretion torus setup and letting the magneto-rotational instability develop. They evidenced the formation of a *hollow core* structure, in which the central half of the core jet sees a relative decrease in outflow energy and a significant decrease in Lorentz factor, hence the "hollow" denomination. This feature is distinctive of MHD setups with respect to purely hydrodynamical jets and appears robustly for various initial conditions, and confirms prior findings (see references in Nathanail et al. 2020). In these structures the core jet is up to two times larger than those from hydrojets. The core jet is the dominant contributor to the afterglow at its peak, and a hollow core should influence the flux at this point; An eventual "dip" is not present in the afterglow light curves from Nathanail et al. (2021) and should be further discussed.

8.4 Intrinsic and apparent jet structures: how outflows shape observations

In this section, we will lay down the analytical framework to calculate the observed behavior of a structured jet as seen from a misaligned line of sight.

8.4.1 Isotropic-equivalent prompt energy

In Sec. 4.4.5, we derived the method to calculate the afterglow from a decelerating structured jet. We will now describe the method to transform the distribution of emitted energy in the structured jet to the observed flux from any line of sight.

Consider a jet in which the material's kinetic, magnetic, thermal energy per unit solid angle as measured in the central engine frame and Lorentz factor at latitude θ are $\epsilon(\theta)$ and $\Gamma_0(\theta)$ before prompt dissipation. These are constant within a core of half opening angle θ_j. The jet is axisymmetric about its axis.

Part of the energy described by $\epsilon(\theta)$ powers the prompt gamma-rays while the rest remains as kinetic energy for the blast-wave during the afterglow phase. Following our discussion in Sec. 8.2.3, we suppose that the efficiency of gamma-ray dissipation is constant, equal to η_γ within an angle $\theta_\gamma \lesssim 2\theta_j$, and zero elsewhere. Therefore, the post-prompt isotropic-equivalent kinetic energy of the blast wave is:

$$E_{\mathrm{k,iso}}(\theta) = 4\pi \left(1 - \eta_\gamma \Theta(\theta_\gamma - \theta)\right) \epsilon(\theta), \tag{8.2}$$

and the isotropic-equivalent contribution to the emitted gamma-ray energy along θ is:

$$\epsilon_{\gamma,\mathrm{em,iso}}(\theta) = \eta_\gamma 4\pi \epsilon(\theta) \Theta(\theta_\gamma - \theta), \tag{8.3}$$

where Θ denotes the Heaviside function.

In Appendix G, we prove that the observed isotropic-equivalent gamma-ray energy at a viewing angle θ_v is then given by (Eq. G.4):

$$E_{\gamma,\mathrm{obs,iso}}(\theta_v) = \frac{1}{4\pi} \int \mathrm{d}\Omega \frac{\epsilon_{\gamma,\mathrm{em,iso}}(\theta)}{\Gamma_0(\theta)^4 \left(1 - \beta_0(\theta) \cos\chi\right)^3} \tag{8.4}$$

where the spherical angle coordinates for the integration are θ, ϕ with angle element $\mathrm{d}\Omega = \sin\theta \mathrm{d}\theta \mathrm{d}\phi$, and χ is the angle between the line of sight and the direction of motion of the emitting material: $\cos\chi = \cos\theta_v \cos\theta + \sin\theta_v \sin\theta \cos\phi$.

This fundamental equation relates the emitted energy by the material in the structure at latitude θ to the observed energy by an observer at a viewing angle θ_v; We refer to these two respectively as the *intrinsic* jet structure and the *apparent* jet structure. This vocabulary was introduced by Salafia et al. (2015), who explored the relation between the functional forms of the intrinsic and the apparent structures. For example, they find that Gaussian intrinsic structures ($\epsilon_{\gamma,\mathrm{em,iso}}(\theta) \propto \exp -\theta^2/2\theta_j^2$) robustly lead to power-law apparent structures ($E_{\gamma,\mathrm{obs,iso}}(\theta_v) \propto (\theta_v/\theta_j)^{-a}$).

The numerical integration of Eq. 8.4 is readily done, and we shall do so in our application to plateau modeling in Chap. 9. However, for the sake of intuition and analytical endeavors, it is useful to distinguish different regimes in the dominant contribution to $E_{\gamma,\mathrm{obs,iso}}$ as a function of the viewing angle. The reasoning for this is done in Appendix G; We briefly reproduce it here.

As we are interested in the total observed energy, the only geometrical effect at play is Doppler beaming, and not light travel delays. At any given viewing angle, the main contribution to the observed energy can be either the *line-of-sight material*—the material directly down the line connecting the observer to the central engine—or the core. The line-of-sight material's contribution is $\epsilon_{\gamma,\mathrm{em,iso}}(\theta_v)$. The core's contribution is $\epsilon_{\gamma,\mathrm{em,iso}}(0)$, however transformed by boosting to the misaligned observer's frame.

As the only effect is beaming, the relevant angular scale is the core's initial Lorentz factor $\Gamma_j = \Gamma_0(0)$, leading to introduce $q = |\theta_v - \theta_j|\Gamma_j$. We now note that:

- For an aligned observer ($\theta_v \leq \theta_j$), the core's contribution is simply $\epsilon_{\gamma,\mathrm{em,iso}}(0)$;
- For a slightly misaligned observer ($\theta_j \leq \theta_v \leq 2\theta_j$), the core contribution is produced only by a patch of size $1/\Gamma_j$ within the core, and deboosted by a Doppler factor ratio of $\sim 1/(1+q^2)$ with respect to an aligned observer;
- For a very misaligned observer ($\theta_v \leq 2\theta_j$), the whole core contributes with a solid angle of Ω_j with the same change in Doppler boosting.

Collecting all these regimes, and systematically taking the greatest contribution between the line-of-sight material and the

core, we finally have (Eq. G.11):

$$E_{\gamma,\mathrm{obs,iso}}(\theta_v) = \tag{8.5}$$

$$\begin{cases} \epsilon_{\gamma,\mathrm{em,iso}}(0) & \theta_v \leq \theta_j \\ \max\left[\epsilon_{\gamma,\mathrm{em,iso}}(\theta_v), \frac{1}{(1+q^2)^2}\epsilon_{\gamma,\mathrm{em,iso}}(0)\right] & \theta_j \leq \theta_v \leq 2\theta_j \\ \max\left[\epsilon_{\gamma,\mathrm{em,iso}}(\theta_v), \frac{1}{q^2(1+q^2)^2}(\Gamma_j\theta_j)^2\epsilon_{\gamma,\mathrm{em,iso}}(0)\right] & 2\theta_j \leq \theta_v \end{cases} \tag{8.6}$$

In Fig. 8.4, we illustrate the intrinsic–apparent jet structure relationship using both the exact integration of Eq. 8.4 and the approximation of Eq. 8.6. We directly prescribe the intrinsic structure $\epsilon_{\gamma,\mathrm{em,iso}}(\theta)$ and show the corresponding apparent structure $E_{\gamma,\mathrm{obs,iso}}(\theta_v)$ on the same plot. We also show the Lorentz factor structure $\Gamma_0(\theta)$. The prescriptions for the intrinsic structures are power-laws up to $\theta = \pi/2$ with varying parameters (Eqs. 4.8 and 4.9), as found in the caption. We label the result of the full integration as "full" and that of the approximate calculation as "approx".

For shallow energy structures or large core Lorentz factors (top left), the line-of-sight material dominates and $E_{\gamma,\mathrm{obs,iso}}$ follows $\epsilon_{\gamma,\mathrm{em,iso}}$; The observed gamma-ray energy is dominated by the energy emitted by the line-of-sight material. This is still the case for moderately steep structures (top right), where $E_{\gamma,\mathrm{obs,iso}}$ still follows $\epsilon_{\gamma,\mathrm{em,iso}}$. It is only for very steep energy structures or smaller core Lorentz factors (bottom) that the core contribution dominates the observed radiation for misaligned lines of sight and $E_{\gamma,\mathrm{obs,iso}}$ departs significantly from $\epsilon_{\gamma,\mathrm{em,iso}}$.

8.4.2 Characteristic dynamical times

Because of light travel effects, the onset of a given material's deceleration will appear at a different time than for an aligned observer. We briefly derive the relevant equations here.

As the blast wave pushes into the external medium it starts decelerating. Seen from an angle θ_v, the Lorentz factor of material at position θ, ϕ remains roughly constant up to an observer time:

$$t^{\theta_v}_{\mathrm{dec}}(\theta,\phi) = (1-\beta_0\cos\chi) \tag{8.7}$$

$$\times \begin{cases} \left(\frac{17E_{\mathrm{k,iso}}}{8\pi n_{\mathrm{ext}} m_p c^5}\right)^{1/3} \beta_0^{-5/3}\Gamma_0^{-2/3} & \text{uniform} \\ \frac{9E_{\mathrm{k,iso}}}{16\pi A\beta_0^3 c^3\Gamma_0^2} & \text{wind} \end{cases} \tag{8.8}$$

where $E_{\mathrm{k,iso}}$, β_0 and Γ_0 are evaluated at θ. We recognize here the travel time delayed version of the deceleration time (Appendix D.6). The first line holds for a uniform external medium of density n_{ext}, and the second line holds for a stellar wind external medium, in which we define the usual wind parameter A. For material moving along the line of sight, $(1-\beta_0\cos\chi) \propto \Gamma_0^{-2}$ and one obtains $t_{\mathrm{dec}} \propto \Gamma_0^{-8/3}$ for a uniform medium and $t_{\mathrm{dec}} \propto \Gamma_0^{-4}$ for a wind. We denote these special cases as $t_{\mathrm{d,los}}(\theta) = t^{\theta}_{\mathrm{dec}}(\theta, \phi = 0)$.

After $t^{\theta_v}_{\mathrm{dec}}(\theta,\phi)$, the Lorentz factor at position θ, ϕ decreases. At a later time $t > t^{\theta_v}_{\mathrm{dec}}(\theta,\phi)$, Eq. 8.7 still holds when replacing $t^{\theta_v}_{\mathrm{dec}}$ by t, hence providing an implicit equation for the Lorentz factor of any portion of the jet at observer time t. This way, one can determine the Lorentz factor of any material in the structured jet at any misaligned observer time. For material moving along the line of sight, this yields simply $\Gamma(\theta,t)/\Gamma_0(\theta) = (t/t_{\mathrm{d,los}}(\theta))^{-3/8}$ for a uniform environment and $(t/t_{\mathrm{d,los}}(\theta))^{-1/4}$ for a wind, again as in Appendix D.6.

8.4.3 Core spreading for slightly misaligned lines of sight

As it decelerates, energy density gradients within the jet will lead to lateral spreading. For the near-core lines of sight we consider here, this spreading may affect afterglow predictions, especially when considering steep jet structures. In Appendix D.8, we estimate the core lateral spreading until the core is revealed to the observer, which is the time of interest in the model we develop in Chap. 9 and generally when studying afterglows from misaligned jets. Our estimates show that it is negligible in this case, allowing us to disregard core spreading for the rest of this Part.

8.5 Conclusion

Structured relativistic jets are ubiquitous in phenomena implicating compact central sources such as micro-quasars, AGNs and GRBs. The variety of length and time scales represented in these phenomena allow to outline the various physical mechanisms at play in these outflows: source orbital motion, collimation by dense material surrounding the central engine, jet-external medium interactions, inherent hydrodynamical or MHD instabilities.

In GRBs, structured jets manifest in GRB–SN associations and generally in low-luminosity events, hinting to misaligned lines of sight to these structures. The structuredness of short GRB jets manifested spectacularly in GW170817. This event showcased the geometrical effects of Doppler beaming and light travel delays that can shape the afterglow radiation from such an outflow. These effects are at play on long time scales for very misaligned lines of sight like we had in GW170817.

However, the ultra-relativistic nature of the cores of GRB jets suggest that these geometrical effects can also occur on only slightly misaligned lines of sight, with which bright GRBs are observed. Following this idea, we will now develop two models to explain the plateaus and flares in GRB X-ray afterglows, both based on slightly misaligned lines of sight to structured jets. The first (Chap. 9) was first laid down by Eichler & Granot (2006), and we will enhance it with an analytical formulation linking plateau properties to the jet structure and with a statistical study of plateau properties. The second (Chap. 10) is original and opens the way to a unified picture for both plateaus and flares in GRB afterglows.

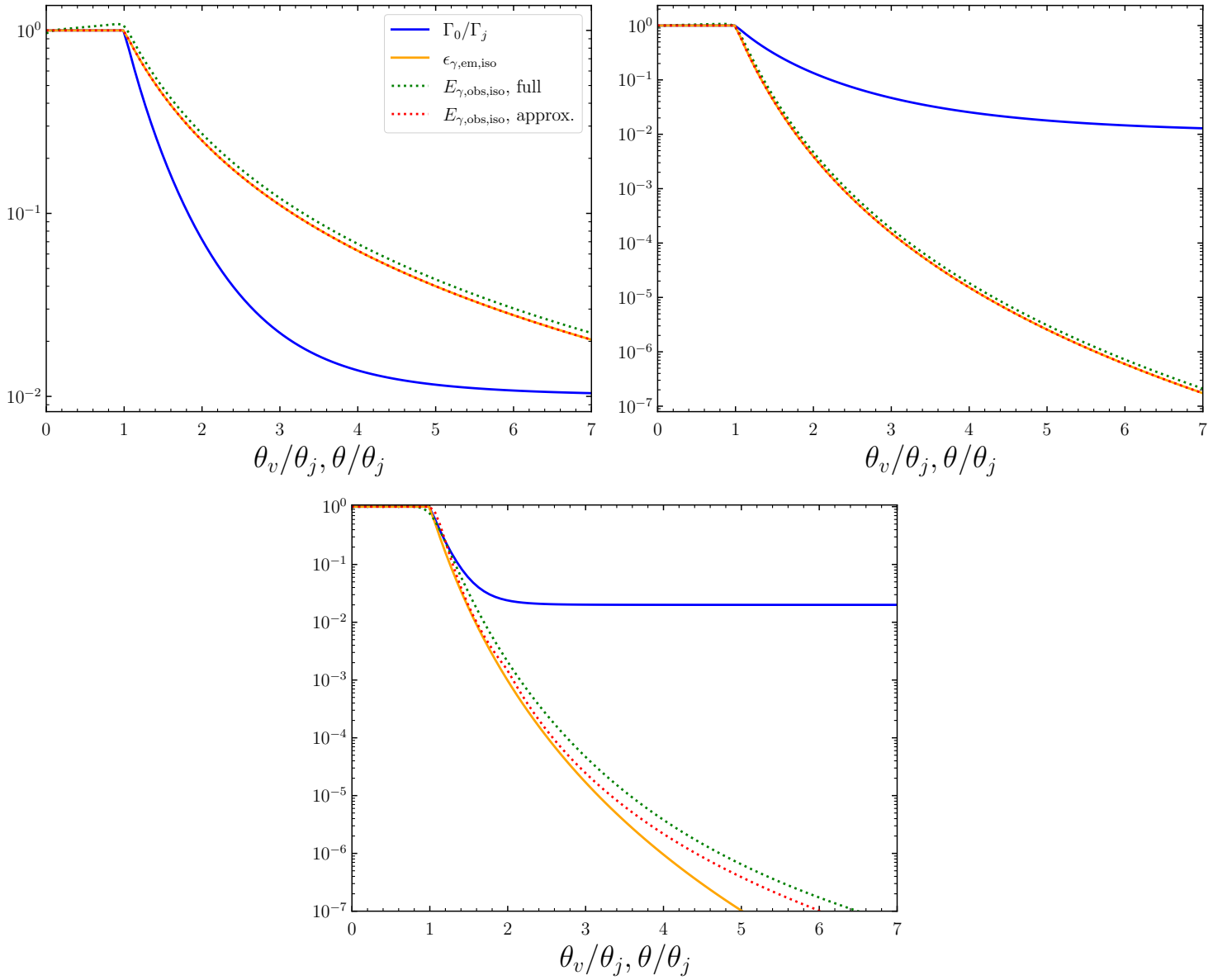


Figure 8.4: **Apparent energy structures arising from power-law intrinsic energy structures.** On the same axis for the latitude θ in the structure and the viewing angle θ_v of the observer, we plot the emitted energy $\epsilon_{\gamma,\mathrm{em,iso}}(\theta)$ and the observed energy $E_{\gamma,\mathrm{obs,iso}}(\theta_v)$, both normalized to the core-dissipated energy $\epsilon_{\gamma,\mathrm{em,iso}}(\theta = 0)$. These are known as the intrinsic and apparent jet structures. We also show the Lorentz factor structure $\Gamma_0(\theta)$. The emitted energy and Lorentz factor structures are both power-laws with varying parameters. The apparent structure is determined either using the full integration over the jet ("full", Eq. 8.4) or using the approximation ("approx", Eq. 8.6). Top left: Shallow structure, with power-law parameters $a = 2$, $b = 4$ and $\Gamma_j = 100$. Top right: Moderately steep structure, with $a = 8$, $b = 3$ and $\Gamma_j = 100$. Bottom: Steep structure, with $a = 10$, $b = 8$ and $\Gamma_j = 50$.

Chapter 9

A refreshed model for plateaus in gamma-ray burst X-ray afterglows

Abstract

Plateaus are recurring features of the afterglows of gamma-ray bursts. They present a rich phenomenology, in particular correlations between their durations and luminosities, and with the properties of their gamma-ray burst prompt emission. In this chapter, we develop a model to interpret plateaus as a purely geometrical effect arising when a decelerating structured jet is observed from a slightly misaligned line of sight, as is the case for bright gamma-ray bursts. This model was suggested early after the discovery of plateaus and we enhance it here by providing analytical formulae for the plateau properties and showing that these remarkable correlations naturally arise in our geometrical setup. This interpretation presents many advantages, as it does not require late-time energy injection, a feature of many plateau models that, as we show here, is less natural given the observed correlations between plateau and prompt properties. Further, we note that afterglow plateau behavior is often associated with flares, motivating a unified picture for plateaus and flares in the context of structured jets and slightly misaligned lines of sight.

9.1 Introduction

Plateaus are recurring features of the early afterglows of GRBs. They were only discovered after the launching of *Swift*, which allowed to acquire afterglow data earlier than 1000 seconds after the trigger (Sec. 1.2). Plateaus are defined by phases of shallow increase or decrease in afterglow flux, starting from the early steep decay phase and lasting for hundreds to tens of thousands of seconds (Nousek et al., 2006). They then suddenly steepen and connect to the regular decay phase (Sec. 1.4).

Many interpretations of plateaus have been suggested in the literature over the years. Some examples are late-time energy injection from the central engine, shining through an internal process (Ghisellini et al., 2007; Beniamini & Mochkovitch, 2017) or as the fresh material joins the external shock (Nousek et al., 2006; Zhang et al., 2006), forward shock emission from an inhomogeneous jet (Toma et al., 2006), forward shock emission with time-dependent microphysical parameters (Granot et al., 2006; Ioka et al., 2006; Panaitescu et al., 2006), contributions from a long-lived reverse shock (Uhm & Beloborodov, 2007; Genet et al., 2007; Hascoët et al., 2014), external shock emission in the thick-shell regime (Leventis et al., 2014) and delayed afterglow deceleration (Granot & Kumar, 2006; Kobayashi & Zhang, 2007; Shen & Matzner, 2012; Duffell & MacFadyen, 2015). More recently, Oganesyan et al. (2020) have suggested another structured-jet-related interpretation, in which plateaus are the result of the prompt emission photons produced at a large angular distance from the observer (more than $1/\Gamma$ from the line of sight) and received by the observer at later times.

Early on after the discovery of X-ray plateaus, Eichler & Granot (2006) suggested that plateaus could be the result of structured jets viewed at latitudes beyond the jets' cores. The idea of this model is that for misaligned observers, relativistic beaming implies that material outside of an angle $\sim 1/\Gamma_j$ from the line of sight contributes very little to the observed radiation, where Γ_j is the core's Lorentz factor. For observers at lines of sight outside the jet core ($\theta_v > \theta_j$), where the energy content of the jet is smaller and the deceleration time longer, this implies a weaker prompt signal and an initially weak afterglow. As time goes by, the jet slows down and the observer starts receiving radiation from the more energetic material along the jet's core. Under certain conditions that we explore in this chapter, this can lead to a plateau-like phase in the X-ray light curve.

The jet structure revealed by GW170817 motivates new GRB modeling with structured jets; In particular, it motivates a re-discussion of this model for plateaus. In this chapter, we

+ Beniamini et al. 2020a, 1

come back to this model in the light of GW170817 and enhance it with analytical derivations to obtain the plateau properties as a function of the jet structure. We will show how this picture naturally reproduces plateau behavior and some associated features.

Since the initial discovery of X-ray plateaus by *Swift*, many more plateaus have been observed and their statistics and correlations with other burst properties studied in detail (Dainotti et al., 2008; Margutti et al., 2013a; Dainotti et al., 2017; Tang et al., 2019). We will show that the forward shock emission of GRBs viewed beyond their jet cores can naturally account for these observed correlations without any need to invoke late time energy injection, which is challenging from the point of view of the central engine and, as we show here, less natural given the observed correlations. Although some plateaus end with a very rapid temporal decline that is clearly inconsistent with an external shock origin (Zhang et al., 2006; Liang et al., 2006a; Troja et al., 2007; Beniamini & Mochkovitch, 2017), there are less than a handful of such cases. The vast majority of plateaus are compatible with the geometrical interpretations we adopt here. Furthermore, the fraction of bursts with plateaus puts strong constraints on the region within which prompt gamma-rays are efficiently produced and their typical durations restrict the allowed structure of energy and Lorentz factor beyond the jets' cores, we discuss these points below.

9.2 Plateaus from slightly misaligned structured jets

In the scenario presented here, the light curve is dominated by material close to the core of the jet. In order to clearly enforce the slightly misaligned setup, we focus on the case where the edge of the core is separated by an angle larger than Γ_j^{-1} from the line of sight. In this regime, the most energetic part of the jet is initially beamed away from the observer due to relativistic beaming. The beaming decreases over time, until eventually the entire jet becomes visible. This can result in a shallow plateau-like phase, assuming that $\Gamma_j^{-1} \ll \Delta\theta = \theta_v - \theta_j \lesssim 0.5\theta_j$ and a sufficiently large $b \gtrsim 8$ to ensure that the contribution from the core dominates that from material moving along the line of sight; See Sec. 9.3.2 for a detailed discussion of the allowable jet structures in our model.

The duration of the plateau in this case is dictated by the time it takes the core to become visible to the observer, i.e., when $\Gamma(\theta_j) \sim \Delta\theta^{-1}$ or

$$t_{\rm p} = t_{\rm d,los}(0) q^{1+2\epsilon}\epsilon \tag{9.1}$$

$$= \begin{cases} 1700 E_{j,53}^{1/3} n_{\rm ext}^{-1/3} (\Delta\theta/0.02)^{8/3}\ {\rm s} & \text{uniform}, \\ 970 E_{j,53} A_{*,-1}^{-1} (\Delta\theta/0.02)^{4}\ {\rm s} & \text{wind}, \end{cases} \tag{9.2}$$

where $E_j = E_{\rm k,iso}(0)$ is the isotropic-equivalent post-prompt core kinetic energy, ϵ defines the deceleration dynamics by $\Gamma \propto R^{-\epsilon}$ and we used Eq. 8.7. We recall the definition $q = \Gamma_j|\theta_v - \theta_j|$.

The strong dependence on $\Delta\theta$ makes it easy to explain a wide range of plateau durations with little change in the viewing angle. Indeed, the observed distribution of $t_{\rm p}$ spans about three orders of magnitude (see the correlation plots below, Figs. 9.2). Such a span, assuming all other parameters are fixed, requires values of $\Delta\theta$ to vary by a factor of at most 13 for a uniform external medium, and 6 for a wind. This is very reasonable given that the lowest value of $\Delta\theta$ in this scenario is $\Gamma_j^{-1} \sim 0.003$ and the largest is roughly $\theta_j/2 \sim 0.05$. If, in addition, one allows for variation in the core energy and ambient density, the same span of plateau durations can be reproduced with an even smaller range of $\Delta\theta$.

We turn next to calculate the luminosity at the end of the plateau phase. The luminosity is somewhat reduced as compared to the standard on-axis case, in which the isotropic-equivalent energy of the jet's core is visible to the observer. This is because at $t = t_{\rm p}$, there is still a sizable fraction of the jet that lies beyond an angle of Γ_j^{-1} from the observer, and its emission is therefore strongly suppressed. Since Γ evolves slower in a wind environment, the effect is slightly more pronounced in that case. Naturally, regardless of the surrounding medium, at $t \gg t_{\rm p}$, when emission from the entire jet becomes visible, the luminosity seen by off-core observer matches that seen by on-axis observers. The luminosity in the X-ray band (defined here as 0.3–30 keV) at the end of the plateau is therefore

$$L_{\rm p} \sim 10^{46} f (1+z)^{\frac{2+p}{4}} \epsilon_{e,-1}^{p-1} \epsilon_{B,-2}^{\frac{p-2}{4}} E_{j,53}^{\frac{2+p}{4}} t_{\rm p,3}^{\frac{2-3p}{4}} \left(\frac{4}{1+Y}\right) {\rm erg\ s^{-1}}$$

$$\sim \begin{cases} 7\times 10^{46} (1+z)^{\frac{2+p}{4}} \epsilon_{e,-1}^{p-1} \epsilon_{B,-2}^{\frac{p-2}{4}} E_{j,53}^{\frac{2}{3}} \times \\ n_{\rm ext}^{\frac{3p-2}{12}} (\Delta\theta/0.02)^{\frac{4-6p}{3}} \frac{4}{1+Y}\ {\rm erg\ s^{-1}} & \text{uniform}, \\ 1.5\times 10^{46} (1+z)^{\frac{2+p}{4}} \epsilon_{e,-1}^{p-1} \epsilon_{B,-2}^{\frac{p-2}{4}} E_{j,53}^{\frac{2-p}{2}} \times \\ A_{*,-1}^{\frac{3p-2}{4}} (\Delta\theta/0.02)^{2-3p} \frac{4}{1+Y}\ {\rm erg\ s^{-1}} & \text{wind}, \end{cases} \tag{9.3}$$

where in the first line f is a normalization that is 10 for a uniform medium and 1.5 for a wind. We have taken here the synchrotron spectrum branch above the cooling and injection frequencies ν_m and ν_c, where the X-rays reside for typical jet parameters (e.g. Nava et al., 2014; Santana et al., 2014). Also, $1+Y$ accounts for flux suppression due to Inverse Compton cooling (Appendix E); We have normalized it by the value obtained for the canonical parameter values chosen here: $\epsilon_e = 0.1, \epsilon_B = 0.01$. For these values of p, one then finds that approximately $L_{\rm p} \propto t_{\rm p}^{-1}$, hinting to the observed correlations in plateaus which we will discuss in Sec. 9.3.1.

Light curves arising from this scenario for a given set of physical parameters and changing values of $\Delta\theta$ are shown in Fig. 9.1. We chose power-law structures in both energy and Lorentz factor:

$$\epsilon(\theta) = \frac{dE}{d\Omega} = \epsilon_j \begin{cases} 1 & \theta < \theta_j\ , \\ \left(\frac{\theta}{\theta_j}\right)^{-a} & \theta \geq \theta_j\ , \end{cases} \tag{9.4}$$

and

$$\Gamma_0(\theta) = 1 + (\Gamma_j - 1) \begin{cases} 1 & \theta < \theta_j\ , \\ \left(\frac{\theta}{\theta_j}\right)^{-b} & \theta \geq \theta_j\ , \end{cases} \tag{9.5}$$

with parameters found in the caption.

In addition to the jet afterglow, we add the early steep decay component (Sec. 1.2), which we interpret according to the classical vision as originating from high-latitude emission of the material producing the prompt phase (e.g. Zhang et al., 2006). Due to relativistic beaming, it is typically dominated by material that is within several $\Gamma_0(\theta_v)^{-1}$ from the line of sight. We empirically model it in the following way:

$$L_{\rm ESD,iso} = \frac{E_{\gamma,\rm iso}(\theta_v)}{T_{90}}\left(\frac{t}{T_{90}}\right)^{-3}, \quad (9.6)$$

where T_{90} is the duration of the prompt emission phase. In what follows, we use a typical value of $T_{90} = 20\,$s. In Eq. 9.6, we recognize the t^{-3} behavior of the bolometric luminosity of high-latitude emission (Appendix C), as well as the average luminosity during the GRB prompt emission $E_{\gamma,\rm iso}/T_{90}$. This form is a good approximation so long as the γ-ray energy is dominated by the line-of-sight material.

Finally, the energy at the core of the jet can be related to the observed gamma-ray energy in the prompt phase. Assuming $\theta_v < \theta_\gamma$ and that the observed gamma-rays are dominated by line-of-sight material (i.e., the structure is not extremely steep, Sec. 8.4.1), we have:

$$E_{\gamma,\rm iso} = \frac{\eta_\gamma}{1-\eta_\gamma}E_{\rm k,iso}(\theta_v) = \frac{\eta_\gamma}{1-\eta_\gamma}E_j\left(\frac{\theta_v}{\theta_j}\right)^{-a}. \quad (9.7)$$

where we assumed the power-law structure.

Writing $\theta_v = \theta_j + \Delta\theta$ and using the relation between $\Delta\theta$ and $t_{\rm p}$ (Eq. 9.2) we plug the previous expression into Eq. 9.3 to obtain:

$$\frac{L_{\rm p}}{\rm erg/s} \sim \frac{4\times10^{46}}{1+Y}(1+z)^{\frac{2+p}{4}}\epsilon_{e,-1}^{p-1}\epsilon_{B,-2}^{\frac{p-2}{4}}\left(\frac{1-\eta_\gamma}{\eta_\gamma}E_{\gamma,\rm iso,53}\right)^{\frac{2+p}{4}} t_{\rm p,3}^{\frac{2-3p}{4}} \times \begin{cases} 7\left[1+0.16t_{\rm p,3}^{3/8}\left(\frac{1-\eta_\gamma}{\eta_\gamma}E_{\gamma,\rm iso,53}\right)^{-1/8} n_{\rm ext}^{1/8}\theta_{j,-1}^{-1}\right]^{\frac{(2+p)a}{4}} & \text{unif.} \\ 1.5\left[1+0.2t_{\rm p,3}^{1/4}\left(\frac{1-\eta_\gamma}{\eta_\gamma}E_{\gamma,\rm iso,53}\right)^{-1/4} A_{*,-1}^{1/4}\theta_{j,-1}^{-1}\right]^{\frac{(2+p)a}{4}} & \text{wind} \end{cases} \quad (9.8)$$

The term in the bracket is the leading order approximation of $1+\left(\frac{t_{\rm p}}{t_j}\right)^{\frac{\epsilon}{1+2\epsilon}}$, where $t_j = t_{\rm d,los}(\theta_j)[\theta_j\Gamma_j]^{\frac{1+2\epsilon}{\epsilon}}$ is approximately the jet break time and $t_{\rm p}/t_j = (\Delta\theta/\theta_j)^{\frac{1+2\epsilon}{\epsilon}}$.

Writing the equation in this way makes it clear that since $\Delta\theta < \theta_j$, we have $t_p < t_j$. This means that the evolution immediately after the plateau still follows the normal pre-jet-break decline phase of GRB afterglows. For longer plateaus the two time-scales start approaching each other, leading to a shorter "regular decline" phase. In principle, a measurement of t_p, t_j from observations of a given burst would lead to a direct estimate of $\Delta\theta/\theta_j$ that is independent of any of the other physical parameters. However, as the viewing angle becomes larger, the jet break transition tends to become smoother or even changes morphology (Lamb et al., 2021), and so in practice it may prove quite challenging to extract this information from observations.

Eq. 9.8 provides a relation between the three observable quantities $E_{\gamma,\rm iso}$, $L_{\rm p}$, $t_{\rm p}$ that is largely independent of the energy and Lorentz factor profile beyond the core. The correlation between $L_{\rm p}/E_{\gamma,\rm iso}$ and t_p, as well as the correlation between $E_{\gamma,\rm iso}$ and $L_{\rm p}$ are depicted in Fig. 9.2 as compared with observations. Note that the latter correlation does depend on the structure beyond the core.

It appears that the observed correlations can be readily reproduced. We stress that we do not attempt here any detailed fitting of the model, as there is clearly some degeneracy between some of the parameters which will hinder the usefulness of such an approach. The purpose of this figure is simply to demonstrate that correlations similar to the observed ones can naturally be reproduced by this model with very reasonable choices of the physical parameters.

We end this description by noting that these plateaus will exist even in the idealized scenario of purely top hat jets, where there is no gamma-ray and afterglow production by material beyond the core. In this case, the plateau properties remain the same as discussed above. However, in order for the gamma-rays to remain detectable, the observation angle has to be somewhat closer to the core: $\Delta\theta \lesssim 5\Gamma_j^{-1}$. $E_{\gamma,\rm iso}$ in Eq. 9.7 is then obtained with the right hand side $(1+q^2)^{-2}$ term in Eq. 8.6.

9.3 Discussion

9.3.1 Relationships between plateau and prompt properties

In any interpretation where there is a strong correlation between the prompt gamma-ray energy and the kinetic energy used to power the plateau which is largely independent of the jet structure, there will be a specific relationship between the three observables: the isotropic-equivalent γ-ray energy ($E_{\gamma,\rm iso}$) the duration of the plateau ($t_{\rm p}$) and the luminosity at the end of the plateau ($L_{\rm p}$). Indeed, let us assume that $E_{\gamma,\rm iso} \propto E_{\rm k,iso}$, the kinetic energy used to power the plateau. Under the usual Blandford-Mckee blast wave evolution, $E_{\rm k,iso}$ is tapped to radiation mainly through the forward shock as the blast wave interacts with the surrounding medium. As stated in Sec. 9.2, the X-rays reside above ν_c, ν_m for typical burst parameters. Thus, the luminosity scales with the kinetic energy and the time as: $L_{\rm p} \propto E_{\rm k,iso}^{(2+p)/4}t_{\rm p}^{(2-3p)/4} \propto E_{\gamma,\rm iso}^{(2+p)/4}t_{\rm p}^{(2-3p)/4}$. For $p \sim 2$, this leads to $L_{\rm p} \propto E_{\gamma,\rm iso}t_{\rm p}^{-1}$, which is close to the observed relation.

Some small modifications to the relation above are expected due to the effects of, e.g., deviations from the linear relation between $E_{\gamma,\rm iso}$ and $E_{\rm k,iso}$ or Inverse Compton cooling effects, causing a slightly shallower evolution of the luminosity with time.

Note, however, that this correlation is much less natural in the common interpretation of plateaus that associates them with large amounts of energy injection onto the external shock at late times. In the latter interpretation, the available energy at the time of gamma-ray production is much smaller than, and not necessarily correlated with, the kinetic energy of the blast-wave at the end of the energy injection phase, and the reasoning

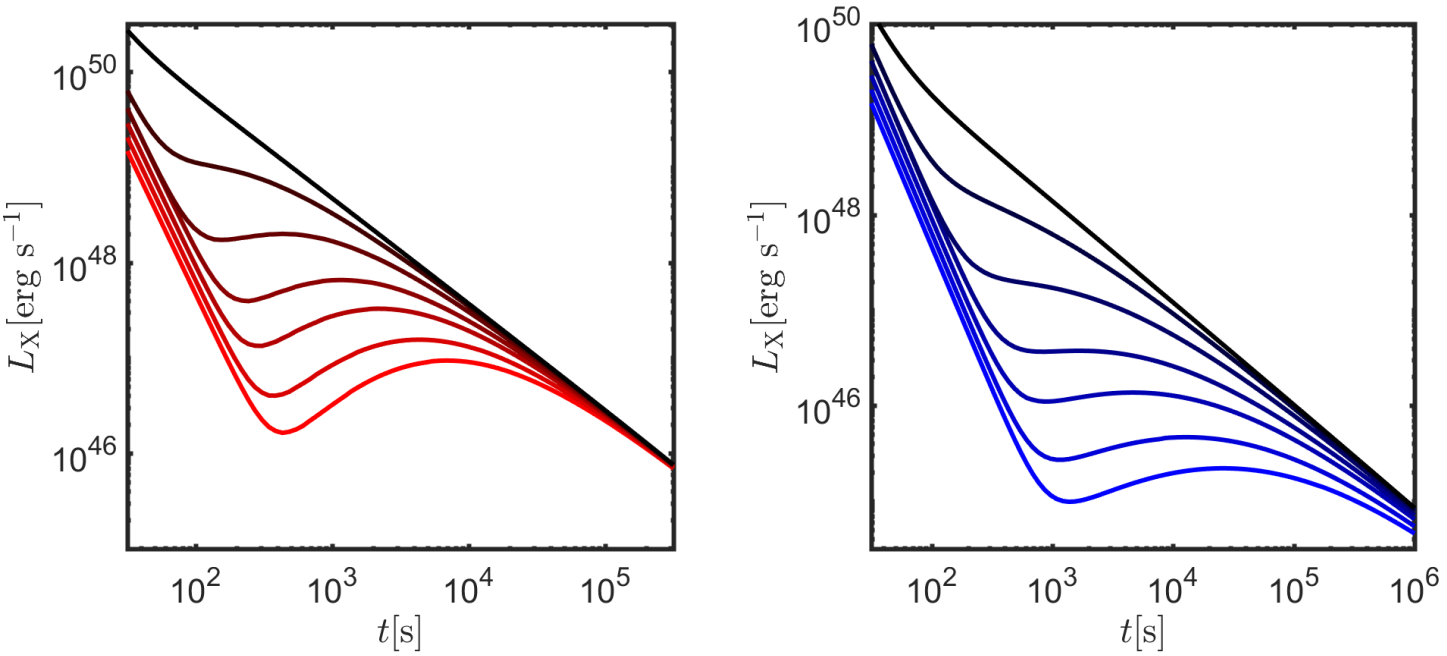


Figure 9.1: **Plateaus arising from structured jets with slightly misaligned observers:** X-ray light curves for a structured jet with $a = 8$, $b = 10$ and different observation angles; from top to bottom: $\Delta\theta = 0 - 0.03$ in steps of 0.005. The X-rays are initially dominated by high-latitude emission, and at later times by the forward shock afterglow. Results are shown for an uniform medium (left) with $n_{\rm ext} = 1$ cm^{-3} and a wind medium (right) with $A_* = 0.1$. We have also taken here: $4\pi\epsilon_j = 10^{54}$ erg, $\theta_j = 0.1$, $\Gamma_j = 400$, $\eta_\gamma = 0.1$, $\epsilon_e = 0.1$, $\epsilon_B = 0.01$, $p = 2.2$. [Beniamini et al. 2020a]

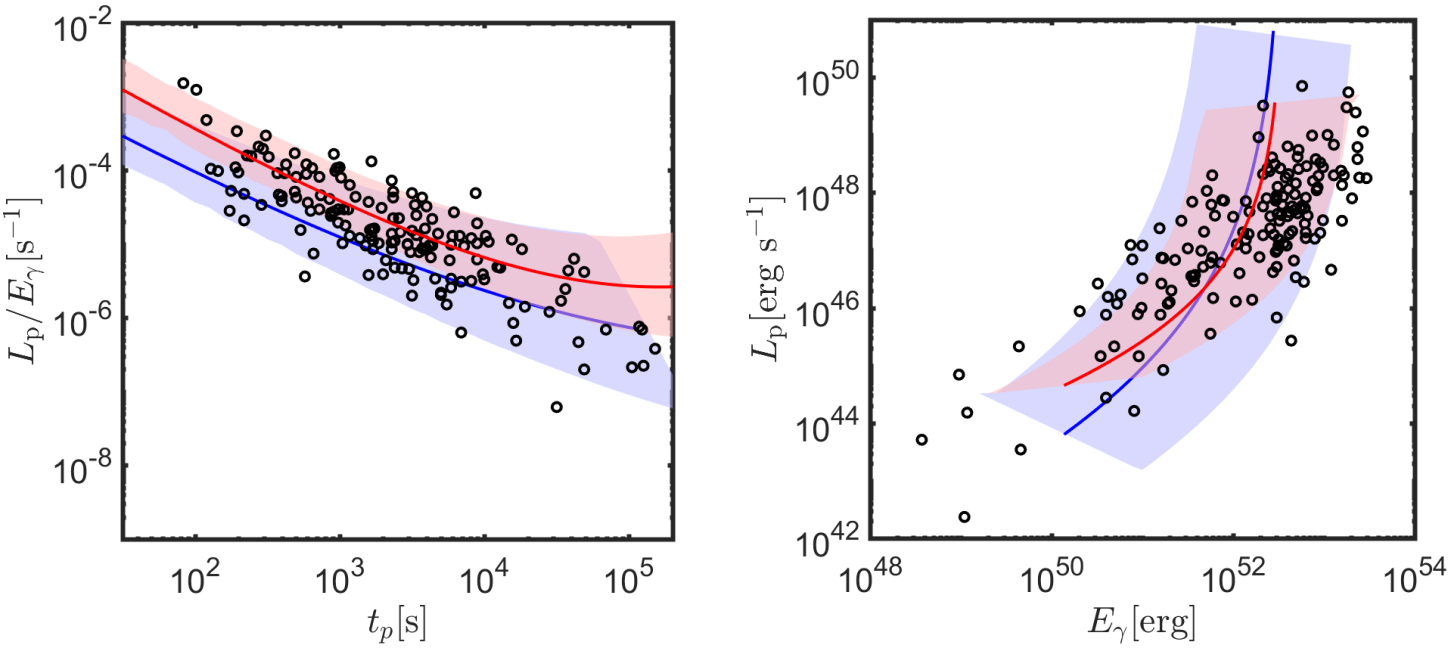


Figure 9.2: **Correlations between plateau properties expected from the structured-jet model:** $L_{\rm p}/E_{\gamma,{\rm iso}}$–$t_p$ (left) and $L_{\rm p}$–$E_{\gamma,{\rm iso}}$ (right) as deduced from Eqs. 9.2, 9.3, 9.8 for uniform (red) and wind environments (blue). Results are shown for $a = 8$, $b \gg 1$, $\theta_j = 0.1$, $\eta_\gamma = 0.05 - 0.2$, $p = 2.2$, $4\pi\epsilon_j = 10^{53-54}$ erg, $\Gamma_j = 400$, $\epsilon_e = 0.1$, $\epsilon_B = 0.01$ as well as $n_{\rm ext} = 0.1 - 1$ cm^{-3} for uniform and $A_* = 0.1 - 1$ for wind. The solid lines depict the median choice of parameters in both cases, varying only the viewing angle and leaving all other parameters fixed. Circles mark observed GRB data, adapted from (Tang et al., 2019). [Beniamini et al. 2020a]

above will no longer hold.

9.3.2 Compatible jet structures

The debeamed core model can result in a range of behaviours, from slowly declining plateau phases to ones that exhibit a shallow bump. Indeed this kind of behaviour is observed in some cases, on the order of a few percent of the entire population. Some examples within *Swift* bursts are: GRBs 081028, 090205, 100901A, 110213A, 120118B, 120215A, 120224A, 150911A, 170202A, 170822A, 181110A, 190422A.

The debeamed core model requires relatively small values of $\Delta\theta$ as well as a Γ_0 profile that falls rapidly beyond the core to avoid the afterglow from the line-of-sight material from dominating over the core contribution. Typically, $b \gtrsim 8$ is required. However, this steepness does not necessarily turn off the emission from line-of-sight material along viewing angles which are considered here. For example for $\Gamma_j = 400, b = 8, \Delta\theta = 0.03$, this material still has a sizeable initial Lorentz factor of $\Gamma_0(\theta_v) \gtrsim 50$ and contributes to the afterglow.

In the new era of GW detections with short GRB counterparts, we will have the possibility to observe and measure large viewing angles of GRB jets (Chap. 5). We thus may be able to resolve these different possibilities for the jet structure and prompt emission at large latitudes, by collecting statistical data on the properties of the prompt and afterglow emission of such structured jets. It should be noted, however, that plateaus are more often observed in long GRBs (see Margutti et al. 2013a and Sec. 9.3.5 below), and it remains an open question whether or not the structure of short and long GRB jets are similar.

In the case of GRB170817A, currently the event with the most detailed insight into the structure of any GRB jet, afterglow photometry points towards a somewhat more shallow jet structure, with $a, b \gtrsim 2$ (Sec. 3.3. This, however, does not contradict the aforementioned requirement of our debeamed core plateau model for rather large b-values. Indeed, on the one hand, the ability to produce core debeamed plateaus is sensitive to the jet's structure only up to near-core angles (recall that, here, we consider $\Delta\theta < \theta_j \sim 0.1$ rad). On the other hand, the afterglow photometry of far-core events, such as GRB170817A, is sensitive to the overall structure, up to $\theta \gtrsim 25$ deg. In other words, a jet structure with a sharp drop ($a, b \gtrsim 8$) between θ_j and $1.5\theta_j$, and then a shallow decrease ($a, b \gtrsim 2$) for larger angles is compatible with both a plateau phase for near-core observers and a GRB170817A-type afterglow for far-core observers.

9.3.3 Plateau statistics

It is illuminating to consider also that the fraction of bursts that exhibit plateaus is ~ 0.5 (Kumar & Zhang, 2015). Under the debeamed core model interpretation, this can easily be related to the maximal angle at which cosmological bursts can typically be viewed, θ_γ. Such a limiting observation angle can exist due to a strong reduction in the gamma-ray production efficiency beyond the core of the jet.

Unless η_γ has decreased significantly by θ_γ, the fraction of bursts with plateaus is roughly proportional to the solid angle of aligned and misaligned observable bursts, i.e.,

$$\frac{\theta_\gamma^2 - \theta_j^2}{\theta_\gamma^2} \sim 0.5\,, \tag{9.9}$$

or $\theta_{\rm max} \sim 1.4\theta_j$.

In other words, since there is a significant fraction of bursts with no plateaus, the maximum angle at which cosmological bursts can be detected cannot be much larger than the jet opening angle, θ_j. This is consistent with the aforementioned discussion that $\theta_\gamma \lesssim 2\theta_j$ (Sec. 8.2.3). Furthermore, note that this argument becomes even more stringent if some of the plateaus are not due to the misaligned observer setup.

9.3.4 Spectral features and chromaticity in plateaus

Observationally, there is usually no evidence for a change in spectrum between the plateau phase and the following X-ray emission (Kumar & Zhang, 2015). For scenarios in which the plateau is produced internally, i.e., from material dissipating at radii smaller than the external shock—see Sec. 9.1 for examples—, this requires fine tuning and should therefore be a source of concern regarding their viability for producing the majority of the observed plateaus. In the scenario proposed here, the cause for the end of the plateau is geometric or dynamical in nature, and therefore there is no change of spectrum associated with the plateau's demise.

Extending beyond the X-rays, it is interesting to consider optical observations simultaneous to X-ray plateaus. As it turns out, the observed situation is somewhat complex (Panaitescu et al., 2006; Li et al., 2012; Liang et al., 2013). In some cases, there are simultaneous plateaus in optical and X-rays, while in others the optical band exhibit a distinct temporal behavior to the X-rays. In our picture, the optical may either mimic the X-rays or not, depending on the location of the injection (ν_m) and cooling (ν_c) frequencies at the time of the plateau. With reasonable variations in the microphysical parameters, it is quite possible for the optical band to be, in some cases, in between ν_m, ν_c during an X-ray plateau, while in others, to be above both frequencies.

A side-by-side comparison of X-ray and optical light curves that will be seen for given GRBs, with different physical parameters and viewing angles is shown in Fig. 9.3. Here we can see the existence of chromaticity in the light curves, especially in the wind case. The time indices in different bands are simply due to the different relevant branches of the synchrotron spectrum, and can change in non-trivial ways during the afterglow.

9.3.5 Plateaus in short GRBs

Although less frequent, plateaus are also observed in short GRBs. An examination of the *Swift* database (Gehrels, 2004) suggests that when a plateau is seen in a short burst it often has a short duration. Indeed, from the plateau duration-luminosity

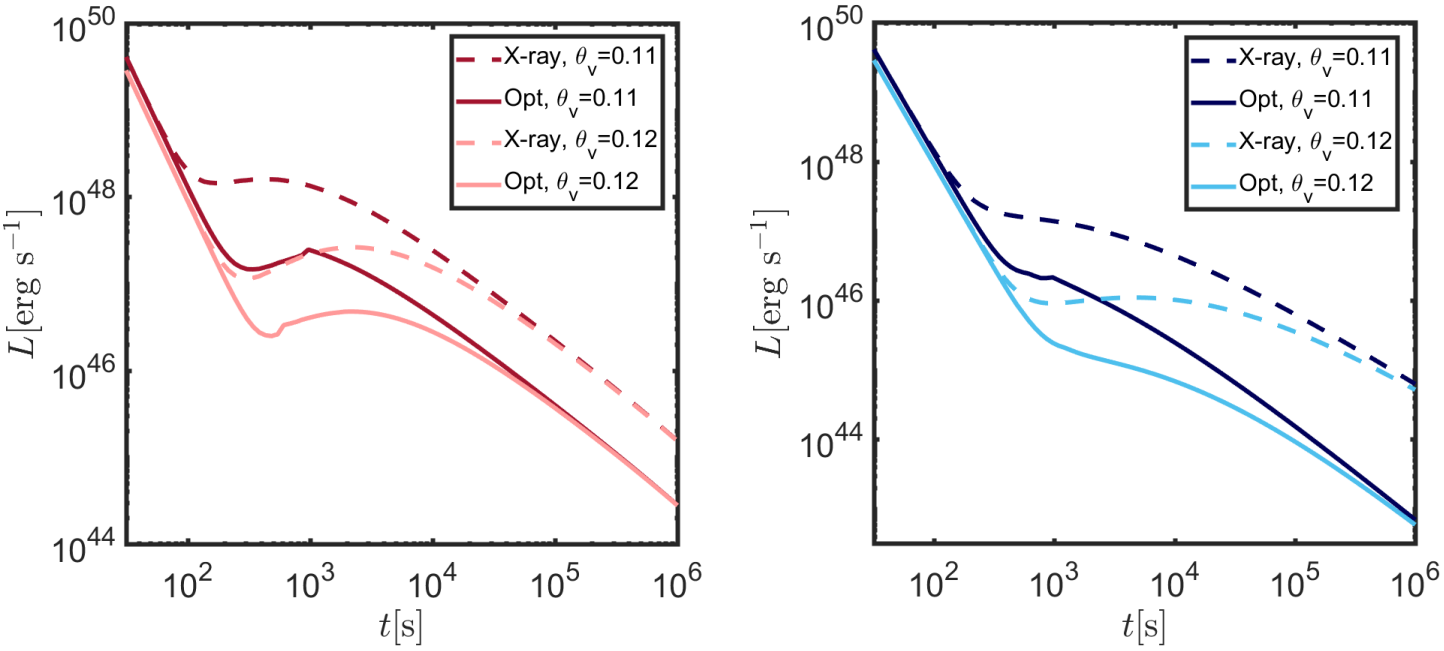


Figure 9.3: **Chromatic features in plateaus expected from the structured-jet model:** X-ray (solid) and optical (dashed) light curves for a structured jet with $a = 8$, $b = 10$, $\theta_j = 0.1$, $\eta_\gamma = 0.1$, $\Gamma_j = 400$, $4\pi\epsilon_j = 10^{54}$ erg, $\epsilon_e = 0.1$, $\epsilon_B = 0.01$, $p = 2.2$ as well as $n_{\rm ext} = 1$ cm^{-3} for uniform (red) and $A_* = 0.1$ for wind (blue) external media. [Beniamini et al. 2020a]

relation, short plateaus have a larger luminosity and therefore long plateaus might be observationally discriminated against because they are weaker. Furthermore, assuming typical values for the isotropic-equivalent kinetic energy and external density in long vs. short GRBs we find that, for the same duration, the plateau luminosity is weaker in short GRBs. To illustrate the latter point, consider $E_{\rm j} = 10^{51}$ erg ($E_{\rm j} = 10^{53}$ erg) and $n_{\rm ext} = 0.1$ cm^{-3} ($A_* = 0.1$) as typical values for short and long bursts respectively. Then, Eq. 9.2 results in comparable durations,

$$t_{\rm p,S} = 800 E_{S,51}^{1/3} n_{-1}^{-1/3} (\Delta\theta/0.02)^{8/3} \text{ s},$$
$$t_{\rm p,L} = 970 E_{L,53} A_{*,-1}^{-1} (\Delta\theta/0.02)^{4} \text{ s}, \quad (9.10)$$

where the subscript S (resp. L) denotes short (resp. long) GRBs. Using Eq. 9.3 we can obtain the ratio of the plateau luminosities for the same parameters

$$\frac{L_{\rm p,S}}{L_{\rm p,L}} = 0.08 E_{\rm S,51}^{2/3} E_{\rm L,53}^{1/10} n_{-1}^{0.38} A_{*,-1}^{-1.15} (\Delta\theta/0.02)^{1.53}, \quad (9.11)$$

where we adopted $p = 2.2$.

Eq. 9.11 demonstrates that the ratio is small for typical values of the burst parameters, making plateaus of a given duration more faint in short bursts as compared with long bursts. Notice that if the external density in the vicinity of short GRB explosions is weaker (as may be expected for double neutron stars mergers with strong kicks or delays between formation and merger, Chap. 7), the conclusion regarding the luminosity ratio becomes even stronger. Naturally, one should also take into account the difference in typical distances between short and long bursts. Since short GRBs are likely to on average be closer than long GRBs, the ratio of the observed fluxes might be somewhat closer to unity as compared to the luminosity ratio. Still, this is unlikely to qualitatively change the conclusion. Indeed, changing the typical redshift between $z = 1$ for short bursts to $z = 2$ for long bursts, corresponds to a modification by a factor of $\lesssim 6$ between the luminosity ratio in Eq. 9.11 and the corresponding flux ratio. Overall, in our picture, short GRB plateaus, and especially the longer ones, are expected to be harder to detect than those of long GRBs.

9.4 Conclusion

We have presented here an interpretation of GRB afterglow plateaus, linking them to forward shock emission viewed by observers on lines of sight very slightly beyond the GRB jet's core in structured jets. Depending on the jet structure, such observers may see a plateau in the early X-ray afterglow light curve that is due to late debeamed emission from the core coming gradually into view.

Due to the strong dependence on viewing angle, our interpretation can reproduce the large span of observed plateau durations and luminosities with very modest variations in the viewing angle between bursts. Furthermore, they can naturally reproduce the observed correlations between the isotropic-equivalent gamma-ray energy, the duration of the plateau and the luminosity at the end of the plateau.

Generalizing beyond this particular model, we have shown that the observed correlations arise in any model where $E_{\gamma,\rm iso}$

is roughly linearly correlated with, and represents a large fraction of, the blast wave kinetic energy tapped during the plateau phase. The most common interpretation for the plateau, involving significant injection onto the external shock at late times does not naturally reproduce these properties.

Due to the geometric and dynamical interpretations associated with these plateau models, no spectral change is expected between the plateau and post-plateau emission. In the optical band, during the plateau phase, due to interplay with the characteristic synchrotron frequencies, complex chromatic behavior is possible. This feature is consistent with the observation that during the time of X-ray plateaus, the optical light curves of the same GRBs are in some instances also flat, but in others are not.

The fraction of bursts that exhibit plateaus, and the statistics of their durations can be related in this model to the exploration of viewing angle space. Indeed, the fact that only ~ 0.5 of bursts have an X-ray plateau is consistent with the interpretation that cosmological bursts are viewed at most only slightly off-core. The latter point is consistent with the conclusion on the fact that $\theta_\gamma \lesssim 2\theta_j$ (Sec. 8.2.3) and is natural in various prompt emission models that lead to very inefficient gamma-ray production at angles beyond the core, where the energy or the Lorentz factor have significantly decreased.

The strong ties of this work with the structures of GRB jets open perspectives to develop this model. These are provided in detail in Chap. 11; They include making fits to X-ray data to explore the parameter space and determine the compatible jet structures, and an in-depth study of the statistics of plateaus to confront these with expectations from an angular exploration of jet structure.

The plateau correlations and other natural consequences of this geometrical setup show that slightly misaligned lines of sight to structured jets are a fruitful idea in GRB modeling. Motivated by this, we will in Chap. 10 seek to carry further this interpretation and model the other salient feature of GRB afterglows: flares.

Chapter 10

Flares in gamma-ray burst afterglows as prompt emission from slightly misaligned structured jets

Abstract

In this chapter, we develop a model to explain the flaring activity in gamma-ray burst X-ray afterglows within the picture of slightly misaligned lines of sight to structured jets. We suggest that these flares could be the manifestation of prompt dissipation within the core of the jet, appearing to a misaligned observer in the X-ray band due to less favorable Doppler boosting and during the afterglow phase because of core–observer light travel delays. In this picture, this observer's prompt comes from material down their line of sight, in the lateral structure of the jet. We start by laying down the basic analytical framework to determine the flare's characteristics as a function of those of the gamma-ray pulse an aligned observer would have seen. We show that, for typical flare observing times and luminosities, there is indeed some parameter space to explain flares in this way. We then analytically explore this model and show that it naturally produces flares with small width, a salient property of flares. We make fits of our model to two flares representing two different types of morphology, to show that our model can capture both. These fits reveal that the duration of the ejection activity in the core jet is a sensitive parameter of the model. While the fits do not require late-time ejection activity, they do require ejection times longer than the duration of the line-of-sight prompt emission, posing the question of variation of central engine activity lifetime from a direction to another. Finally, we discuss the possibility of joint plateau–flare behavior, as the plateau model exposed in the previous chapter is set in the same geometrical picture as this flare model.

10.1 Introduction

Flares[1] are sudden rebrightenings observed in the afterglow phases of GRBs, primarily in the X-ray band (Nousek et al., 2006). Flares occur in around one third of observed GRB afterglows. Just like plateaus (Chap. 9), they were discovered thanks to the early afterglow monitoring capabilities of *Swift*. Most observed flares occur less than 1000 s after prompt trigger (e.g. Chincarini et al., 2010; Yi et al., 2016). From the first catalogs of flares in *Swift*/XRT light curves, flares showed the salient feature of having small and tightly distributed *aspect ratios* (i.e., the ratio of their width to their arrival time, Chincarini et al. 2007). It was also noted that flare morphologies were quite diverse—with fast and slow rising and decay phases—and that they mimicked GRB prompt pulses, with a very similar distribution of rising-to-decay-time ratios (Chincarini et al., 2010) and other remarkably analogous phenomenology such as spectral lag, lag-luminosity correlations, width-energy band relations features analogous to prompt pulses (Margutti et al., 2010).

These temporal and spectral similarities with prompt pulses suggest a common origin for flares and prompt emission. For example, the central engine of GRBs could have a second episode of activity, explaining the delay between prompt emission and X-ray flares (e.g. Burrows et al., 2005; Liang et al., 2006b). Causes for the engine delayed restart can, for example, be fragmentation then accretion of a collapsing star (King et al., 2005), instability-induced variability in accretion around the central object (Perna et al., 2006) or magnetic activity of the young post-merger pulsar (Dai et al., 2006). However, in these late-engine-activity models, the emission in the X-ray rather than the gamma-ray bands requires an explanation. Furthermore, producing small aspect ratios often requires to tune the second activity's duration to the time of quiescence between the

[1]The work presented in this chapter will shortly be submitted for publication.

two episodes in an unnatural way. Other pictures for flares include emission from the reverse shock propagating in a stratified ejecta (Hascoët et al., 2017; Lamberts & Daigne, 2018; Ayache et al., 2020) or Compton up-scattering of photons from the reverse shock when crossing the forward shock (Kobayashi et al., 2007). For a compact review of possible origins of afterglow variability, see Ioka et al. (2005).

In Chap. 9, we presented a successful model for plateaus in GRB afterglows. While we know of no quantitative assessment of this fact, scrolling the *Swift*/XRT repository reveals that plateau activity seems to be correlated with flaring activity. We thus set out to interpret flares in GRB afterglows within the same physical setup: slightly misaligned lines of sight to a structured jet. Motivated by the aforementioned similarities between X-ray flares and GRB prompt pulses, we also posit a common origin for the two. However, as for the plateau, we will explain the delayed occurrence of the flares not by their delayed emission, but rather by the light travel time between the flare production site within the core and the misaligned observer: We suggest that flares in GRB X-ray afterglows are the manifestation of prompt dissipation in the core of the jet, as seen from slightly off-axis lines of sight. Because of relativistic effects, this radiation appears delayed, dimmer and downshifted in energy. In other words, X-ray flares are *deboosted* versions of gamma-ray pulses from prompt energy dissipation in the core.

We present our unified picture for plateaus and flares in Fig. 10.1: For an aligned viewer (A), the prompt emission comes from the core jet shining in gamma-rays (green), and the afterglow phase contains the early steep decay and radiation from the decelerating forward shock; All other jet regions are too weak and not enough boosted to contribute to the aligned observer's signal. For a misaligned observer (B), the prompt emission and early steep decay come from the material down their line of sight (red), as we showed is most often the case in Chap. 8. Progressively the structured jet decelerates, giving rise to the plateau phase; In the mean time, prompt photons from the core (green) travel to the observer, and reach them as X-ray flares, i.e., deboosted and dimmer than they would have reached the aligned observer.

Note that we consider a single central engine activity episode, and the delay in flare occurrence is a geometrical effect. We anticipate Sec. 10.2 in mentioning that the arrival time for flares in such a picture is bounded to $\lesssim 1000$ s after prompt trigger. We are therefore dealing with *early flares*. Remarkably, these early flares seems to constitute a distinct (and largely statistically dominant) class of flares, as shown by their distinctive temporal behavior and flare-to-continuum contrasts ($\Delta F/F$) with respect to late flares ($\gtrsim 1000$ s, Margutti et al. 2011; Bernardini et al. 2011). A different origin for these two classes, and our model's natural restriction to the early class is a further motivation to explore this picture.

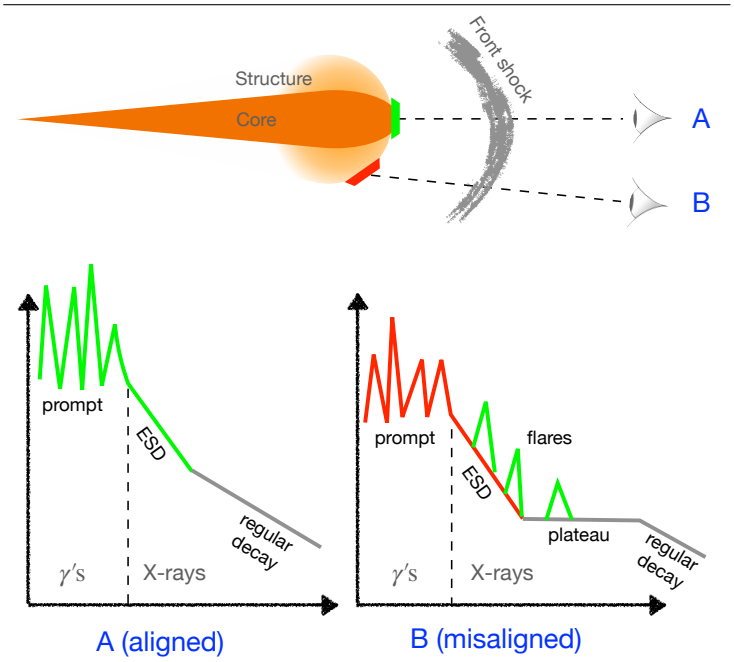


Figure 10.1: **Schematic description of our picture for flares.** The colors of the elements of the light curves correspond with the colors of the emitting regions in the jet: core (green) or lateral structure (red). Each component of the prompt and afterglow phases of aligned (A) and misaligned (B) viewers comes from a different region, according to our picture. For the misaligned observer, the emission from the core matter appears as flares in the X-ray band, atop the ESD and the plateau phase.

10.2 Model outline and first properties

We consider a shell of ultra-relativistic matter with Lorentz factor Γ ejected at a time $t_{\rm ej}$ from the central engine, within the core of the jet. At an emission time t_e, this shell reaches a dissipation radius $R_e = \beta\,(t_e - t_{\rm ej})$ and radiates energy which, for an aligned observer appears as gamma rays. For this aligned observer, this radiation is observed at time:

$$t_{\rm on} = t_{\rm ej} + \frac{1-\beta}{\beta}\frac{R_e}{c} \tag{10.1}$$

where $\beta \sim 1$ is the shell's velocity.

We now consider a misaligned observer, lying at a viewing angle θ_v from the jet's core, with $\theta_v > \theta_j$, and θ_j is the core's half-opening angle. For this observer, the first photons from this shell's radiation arrives at time $t_{\rm off} = t_{\rm ej} + \frac{1-\beta\cos(\theta_v-\theta_j)}{\beta}\frac{R_e}{c}$. This is:

$$t_{\rm off} = \mathcal{S} t_{\rm on} - (\mathcal{S}-1) t_{\rm ej} \tag{10.2}$$

where we have denoted:

$$\mathcal{S} = \frac{1-\beta\cos\left(\theta_v - \theta_j\right)}{1-\beta}, \tag{10.3}$$

referred to as the *stretch factor*. This factor is the ratio of the Doppler boosts between the aligned and misaligned observers.

Furthermore, the peak energy of the detected spectrum transforms as $\mathcal{S}^{-1}$ between the two observers' rest frames:

$$e_{\rm off} = \mathcal{S}^{-1} e_{\rm on} \tag{10.4}$$

Finally, considering a slightly misaligned line of sight ($0 < \Delta\theta \ll 2\theta_j$), we show in Appendix C.2 that the peak bolometric luminosity of the detected radiation transforms according to (Eq. C.27):

$$L_{\rm off} = f_{\rm geo} \mathcal{S}^{-3} L_{\rm on} \tag{10.5}$$

where $f_{\rm geo} \leq 1/2$ is a numerical factor accounting for the transverse angular size of the core jet as seen from the slightly off-axis line of sight; For larger viewing angles, $f_{\rm geo}$ will progressively decrease. For the typical $\theta_v/\theta_j \sim 1.3$ that we considered in Chap. 9, one finds $f_{\rm geo} \sim 0.14$ to best reproduce numerical calculations (Sec. 10.3.3).

These relations show that what is seen as a prompt pulse by an observer aligned with the jet core transforms, for a slightly misaligned observer, into a signal that is delayed, softer and less luminous. In this picture, we will use the term *pulse* for the signal detected by a core-aligned observer, and *flare* for the deboosted emission a misaligned observer detects. A gamma-ray pulse turns into an X-ray flare when seen by an off-axis observer.

For example, let us consider $\Gamma = 150$ and $\theta_v - \theta_j = 0.03\,{\rm rad} \sim 5/\Gamma$, a viewing angle that could typically lead to plateau behavior in the afterglow once the structure decelerates (Fig. 9.1), according to the misaligned-observer interpretation developed in Chap. 9. This configuration leads to $\mathcal{S} = 21$. We take the following typical GRB pulse characteristics: a pulse observed at $t_{\rm on} = 8\,{\rm s}$, coming from a shell ejected at $t_{\rm ej} = 3\,{\rm s}$ with peak bolometric luminosity $L_{\rm on} = 5 \times 10^{52}\,{\rm erg/s}$ and peak energy $e_{\rm on} = 300\,{\rm keV}$. Such a pulse would appear as a flare at $t = 110\,{\rm s}$ with a peak bolometric luminosity of $L_{\rm off} = 3 \times 10^{48}\,{\rm erg/s}$. The flare peak energy would be $e_{\rm off} = 14\,{\rm keV}$, in the X-ray band. As the pulse and flare peak energies are respectively in the gamma-ray and X-ray bands, the peak bolometric luminosity is equivalent to the peak luminosity in the respective bands. It seems therefore that this mechanism can explain the flaring activity in GRB afterglows.

Let us generalize this order-of-magnitude estimate. We denote by $L_{\rm flare}$ the peak bolometric luminosity of an observed flare and $t_{\rm flare}$ its peak time. We seek under which conditions can this flare be interpreted as deboosted core prompt emission. We denote by $\tau = R_e/2\Gamma^2 c$ the shell's angular timescale. Assuming the duration of the emission is negligible—i.e., instantaneous dissipation of energy in the shell—,the duration of the pulse as seen by an aligned observer is well approximated by the delay between the arrivals of photons from their beaming cone's axis and edge:

$$\Delta t_{\rm on} = \frac{R_e}{c}\left(1 - \cos 1/\Gamma\right) \tag{10.6}$$

$$\sim \frac{R_e}{2c\Gamma^2} \tag{10.7}$$

$$= \tau \tag{10.8}$$

Assuming the dissipation is instantaneous simplifies much of the derivation below. We will adopt this hypothesis as this is a first exploration of this model and it allows to carry out the analytical development further. In practice, the relationship between τ and $\Delta t_{\rm on}$ depends on the prompt dissipation mechanism. Different mechanisms would affect the light curve profiles of the flares we will obtain below, but not the general features of the model. In Sec. 10.5.2, we will discuss this point relative to GRB emission mechanisms in more detail

Therefore, writing $\Delta t_{\rm on} = \tau \sim (1-\beta)R_e/c$ and using Eq. 10.1 and 10.2, we arrive at this new form for the flare peak time:

$$t_{\rm flare} = \tau S + t_{\rm ej} \tag{10.9}$$

Furthermore, the bolometric luminosity observed by the aligned observer is $L_{\rm on} \sim E_{\rm iso}/\Delta t_{\rm on}$, where $E_{\rm iso}$ is the isotropic-equivalent source-frame dissipated energy in the shell. Using Eq. 10.5, we finally obtain:

$$L_{\rm flare} = f_{\rm geo} S^{-3} \frac{E_{\rm iso}}{\tau} \tag{10.10}$$

In our picture, the shell's ejection occurs during the central engine activity, which lasts for a duration depending on the smaller-scale physics around the central engine and in the accretion disk. The duration $T_{\rm CE}$ of this central engine activity can however be estimated by the duration of GRB prompt phases, through the T_{90}. In fact, T_{90} naturally underestimates $T_{\rm CE}$, by a factor depending on the GRB light curve profile. Nonetheless, for a given GRB, the ejection times that we can consider are bounded approximately by the prompt duration, $t_{\rm ej} \lesssim T_{90} \leq 100$ s for most GRBs. Similarly, the durations of pulses in GRBs are generally less than a few seconds (e.g., Hakkila et al., 2018). Therefore, the $\Delta t_{\rm on}$ (or τ) we can consider are also constrained, to being less than a few seconds. Finally, for the prompt gamma-ray pulses to transform to X-ray flares, the stretch factor must typically be $\mathcal{S} \lesssim 100$, as mentioned in the numerical example. As we shall soon see, keeping reasonable values for $E_{\rm iso}$ also implies that $\mathcal{S} \lesssim 100$, due to the flare flux suppression by $\mathcal{S}^{-3}$ (Eq. 10.5). Therefore, we conclude from Eq. 10.9 that the flares explained in our off-axis mechanism cannot appear much later than $t_{\rm flare} \sim 1000$ s. As mentioned in Sec. 10.1, there seems to be a dichotomy both in post-peak decay slopes and in the $\Delta F/F$ distribution between early ($t_{\rm flare} < 1000$ s) and late ($t_{\rm flare} > 1000$ s) flares (Bernardini et al., 2011). Early flares may thus be a distinct subclass of flares produced by a specific mechanism and the natural production of such arrival times by our model further motivates to pursue this picture. Moreover, these early flares are by far the most numerous in the observed population: In the source rest frame (i.e., redshift-corrected), 80% of X-ray flares occur less than 260 s after the prompt trigger (Yi et al., 2016). See however in Sec. 10.5.4 a discussion on the nature of these early flares and possible pollution by prompt emission

The typical values required for $\mathcal{S} \lesssim 100$ in our picture also allow us to discuss the core shell's Lorentz factor to expect. As the lines of sight are confined to slight misalignment for classical bright GRBs ($\theta_v \lesssim 2\theta_j$, Sec. 8.2.3), the condition on $\mathcal{S}$ translates to $\Gamma \lesssim 200$.

Inverting Eqs. 10.9 and Eqs. 10.10, we conclude that, for observed flare properties $L_{\rm flare}$ and $t_{\rm flare}$ and assuming the shell's ejection time $t_{\rm ej}$, the isotropic-equivalent dissipated energy required to produce the flare is given as a function of τ by:

$$E_{\rm iso}(\tau) = \tau^{-2} \frac{L_{\rm flare}}{f_{\rm geo}} \left(t_{\rm flare} - t_{\rm ej}\right)^3 \tag{10.11}$$

Assuming these $E_{\rm iso}$ and τ indeed reproduce the flare, the corresponding stretch factor is:

$$\mathcal{S} = \left(\frac{\tau L_{\rm flare}}{f_{\rm geo} E_{\rm iso}} \right)^{-1/3} \tag{10.12}$$

In Fig. 10.2 (colored lines), we plot $E_{\rm iso}(\tau)$ assuming ejections times from 0 to $2 \times T_{\rm CE}$ and a typical GRB central engine activity duration of $T_{\rm CE} \sim T_{90} = 40$ s (average duration of *Fermi*/GBM long bursts, von Kienlin et al. 2020) and with $f_{\rm geo} = 1/2$. In Fig. 10.2, we chose the median source-frame (i.e., redshift-corrected) $t_{\rm flare}$ and $L_{\rm flare}$ in the *Swift*/XRT afterglow flare sample: $\overline{t_{\rm flare}} = 85$ s and $\overline{L_{\rm flare}} = 1.8 \times 10^{49}$ erg/s (Chincarini et al., 2010; Yi et al., 2016). We also plotted isocontours of the corresponding $\mathcal{S}$-factor. This figure allows us to outline which model parameter values are required to reproduce the typical X-ray flare and if these values are consistent with our physical setup and typical GRB quantities. Our parameter space is constrained by the following conditions:

- Shell ejection must occur during the primary central engine activity, thus $t_{\rm ej} \lesssim T_{90}$;
- We must maintain a slightly misaligned line of sight, thus $\mathcal{S} \gtrsim 10$;
- Prompt gamma-rays must end up in the X-ray band, thus $\mathcal{S} \lesssim 100$;
- An aligned observer should detect a classical prompt pulse with duration typically between 0.1 s and 5 s (e.g. Hakkila et al., 2008): $0.1 \text{ s} \lesssim \Delta t_{\rm on} \sim \tau \lesssim 5$ s;
- The shell must dissipate an energy in line with what is observed for single pulses in typical GRBs. For entire GRBs, the dissipated energies are in the range $E_{\rm iso}^{\rm GRB} = 10^{53\pm1}$ erg (e.g., Amati 2006). Assuming a few pulses per GRB, we can estimate that each pulse dissipates an energy on the order of $E_{\rm iso} = 10^{52.3\pm1}$ erg. We can thus consider this range for the shell's single dissipation episode.

The region in parameter space respecting all of these conditions is colored in purple in Fig. 10.2: For any $(\tau, E_{\rm iso})$ pair in this zone, there corresponds an ejection time $t_{\rm ej}$ during the central engine activity and an $\mathcal{S}$-factor of a misaligned line of sight such that the flare appears at $\overline{t_{\rm flare}}$ with peak luminosity $\overline{L_{\rm flare}}$. This figure shows that there is some parameter space available to our model to explain typical flares as deboosted core prompt dissipation.

Naturally, one should use plots like Fig. 10.2 to find solutions to flares in actual GRBs, adapting $t_{\rm flare}$, $L_{\rm flare}$ and T_{90}. For brighter or later flares, Eq. 10.11 shows that larger $E_{\rm iso}$ are required, and this available parameter space will shrink. This equation also shows that allowing for larger ejection times eases the constraint on $E_{\rm iso}$. In Fig. 10.2 we bounded $t_{\rm ej}$ with a generic estimate for central engine activity duration, $\lesssim 40$ s. In fact, this bound should be the duration of central engine activity *on directions within the core*, $T_{\rm CE}^{\rm c}$. This duration may be different than the duration of the central engine activity on the misaligned observer's line of sight $T_{\rm CE}^{\rm LOS}$, measured by the actual GRB in which the flare we seek to explain appeared. Here, we will suppose that the central engine activity has the same duration on all directions and discuss this hypothesis in Sec. 10.5.2.

For a given flare, one must also consider the actual temporal profile of the flare and seek solutions within this available parameter space that correctly fit the light curve. We will do so for two typical flare shapes in Sec. 10.4.

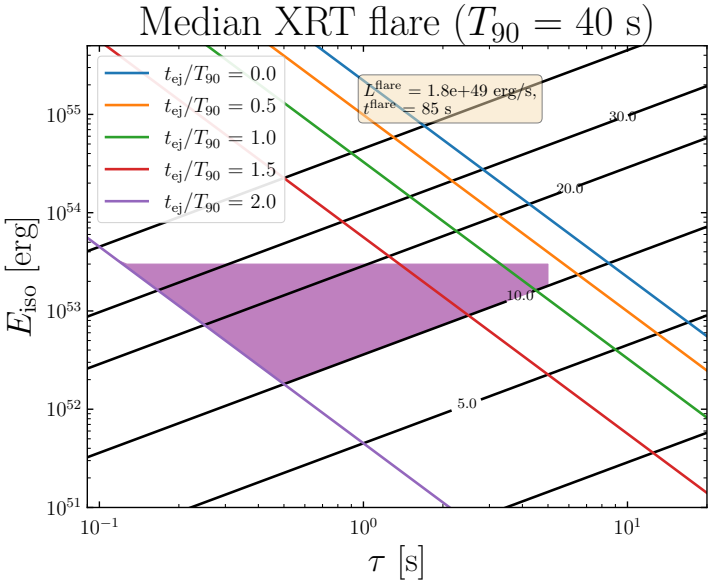


Figure 10.2: **Misaligned structured-jet flare model parameter space exploration to reproduce the median XRT flare.** The colored lines represent the $E_{\rm iso}$ energy required to produce a typical flare peaking at $t_{\rm flare} = 85$ s and with peak luminosity $L_{\rm flare} = 1.8 \times 10^{49}$ erg/s as a function of the shell's geometrical timescale τ and for different ejections times from 0 to 2 times the typical long GRB T_{90}. The black lines are contours of the corresponding required $\mathcal{S}$-factor, values are indicated on the lines. The region in this parameter space consistent with the off-axis geometrical setup of our model and with expected ranges for shell dissipated energies and GRB pulse durations is marked in purple. We find that there is available parameter space for our model to explain the typical flare.

Table 10.1: **Parameters for all flares** with light curves represented in Figs. 10.3. $t_{\rm ej}$: ejection time of the core shell responsible for the flare; $\tau^{\rm c}$: angular timescale $R_e/2\Gamma^2 c$ of the shell, where Γ is the shell's Lorentz factor; shell diameter: angular diameter of the shell, given in units of $1/\Gamma$. $t_{\rm flare}$: peak time of the flare; w: width of the flare, measured as the interval between the two times when the flux is $1/e$ times the peak flux; $L^{\rm flare}/L^{\rm ESD}$: ratio of the peak flux of the flare to the ESD flux at flare peak; $L^{\rm prompt}_{p,{\rm BAT,on}}$: peak pulse luminosity that an observer aligned with the shell would detect in the BAT band.

#	$t_{\rm ej}$ [s]	$\tau^{\rm c}$ [s]	shell diameter [units of $1/\Gamma$]	$t_{\rm flare}$ [s]	$w/t_{\rm flare}$	$L^{\rm flare}/L^{\rm ESD}$	$L^{\rm prompt}_{p,{\rm BAT,on}}$ [erg/s]
A	70	0.5	15	76	0.10	9.8	4.9e+52
B	50	1	7	62	0.24	1.5	2.5e+52
C	80	1	6	92	0.16	8.5	2.5e+52
D	50	2	3	74	0.33	1.2	1.2e+52

10.3 Early flare visibility

10.3.1 Conditions for flare visibility

The majority of X-ray flares occur early in the X-ray afterglow, during the early steep decay (ESD) phase. Within the context of a structured jet, we will now outline the conditions for the flares produced by our mechanism to appear visible above the ESD.

We suppose the observer lies at an angle θ_v from the core. For this observer, the ESD will be produced by high-latitude emission from the last shell that flashed on their line of sight. Introducing the dissipated energy $E^{\rm LOS}_{\rm iso}$ of this last shell and its decay time scale $\tau^{\rm LOS}$, this ESD phase will have an approximate isotropic-equivalent bolometric luminosity of:

$$L^{\rm LOS}_{\rm ESD}(t) = \frac{E^{\rm LOS}_{\rm iso}}{\tau^{\rm LOS}}\left(\frac{t}{\tau^{\rm LOS}}\right)^{-3} \tag{10.13}$$

for times larger than the end of the prompt emission.

Using the same notations as in Sec. 10.2 adding the superscript $^{\rm c}$ for the shell in the core responsible for the flare, the flare's peak luminosity is still $L_{\rm flare} = f_{\rm geo}\mathcal{S}^{-3}E^{\rm c}_{\rm iso}/\tau^{\rm c}$ and the arrival time $t_{\rm flare} = \mathcal{S}\tau^{\rm c} + t_{\rm ej}$. We now introduce the jet structure, prescribing that, on average, the dissipated energies of shells in the core and on the θ_v line of sight are linked by:

$$\frac{E^{\rm LOS}_{\rm iso}}{E^{\rm c}_{\rm iso}} = \left(\frac{\theta_v}{\theta_j}\right)^{-a} \tag{10.14}$$

where we chose a power-law for ease of the analytic development that will follow. The energy and Lorentz factor structures are chosen so that the observer's GRB is dominated by line of sight material. As shown in Fig. 8.4 (top right), this will be the case for the values of $a = 8$, $b = 3$ and the $\Gamma_j = 100$ that we adopt in our numerical exploration of Sec. 10.3.3.

We therefore have the following flare–ESD contrast at the time of the flare:

$$\frac{L_{\rm flare}}{L^{\rm LOS}_{\rm ESD}(t_{\rm flare})} = f_{\rm geo}\mathcal{S}^{-3}\frac{\tau^{\rm c}}{\tau^{\rm LOS}}\left(\frac{\theta_v}{\theta_j}\right)^{a}\left(\frac{t_{\rm flare}}{\tau^{\rm c}}\right)^{3} \tag{10.15}$$

$$= f_{\rm geo}\frac{\tau^{\rm c}}{\tau^{\rm LOS}}\left(\frac{\theta_v}{\theta_j}\right)^{a}\left(\frac{\mathcal{S}^{-1}t_{\rm flare}}{\tau^{\rm c}}\right)^{3} \tag{10.16}$$

$$= f_{\rm geo}\left(\frac{\tau^{\rm c}}{\tau^{\rm LOS}}\right)^{2}\left(\frac{\theta_v}{\theta_j}\right)^{a}\left(1-\frac{t_{\rm ej}}{t_{\rm flare}}\right)^{-3} \tag{10.17}$$

where we used $t_{\rm flare} = \mathcal{S}\tau^{\rm c} + t_{\rm ej} \Leftrightarrow \mathcal{S}^{-1}t_{\rm flare} = \tau^{\rm c}\left(1 - t_{\rm ej}/t_{\rm flare}\right)^{-1}$.

First, it appears from Eq. 10.17 that a steeper structure favors the appearance of flares during the ESD. While this is true—it is simply that the ESD is dimmer for a steep structure—, one must bear in mind that a steeper structure also suppresses the misaligned prompt emission, and thus hinders the GRB detection and flare observation altogether. A finer calculation considering the likeliness of observing the GRB taking the structure into account is thus called for, see Sec. 10.5.3 for details.

Second, we find that, for fixed τ's, a later ejection time favors bright flares. This was already noted in our discussion of Fig. 10.2 and Eq. 10.11: One always has $t_{\rm flare} = \mathcal{S}\tau^{\rm c} + t_{\rm ej} \geq t_{\rm ej}$. A larger $t_{\rm ej}$ favors a small $\mathcal{S}$ and thus a smaller suppression of flare flux by the $\mathcal{S}^{-3}$ factor. In the context of a flare during the ESD phase, it is clear that attributing a larger portion of the flare arrival time to ejection delay simply allows the flare to appear when the ESD flux is lower, as is clear in Eq. 10.17.

Finally, the appearance of the flare is dependent on the ratio of the core and line of sight τ's. The τ's depend on the material's Lorentz factor, dissipation radii and possibly composition. These properties are not uncorrelated in most GRB prompt emission mechanisms. On two different directions, it is not clear how they should generally be ordered.

On the one hand, if generally $\tau^{\rm c} < \tau^{\rm LOS}$, a bright flare can only be obtained if the ejection time is close to the time of the flare, as we just discussed. This may appear as a strong condition and leads us to discuss once more whether the central engine activity must be the same on all directions. If this activity is more durable in the core than on misaligned directions, there is more flexibility for large $t_{\rm ej}$ and bright flares, even if $\tau^{\rm c} < \tau^{\rm LOS}$. On the other hand, if $\tau^{\rm c} > \tau^{\rm LOS}$ is possible, this constraint does not apply and a flare can be produced without imposing a strict constraint on the duration of the central jet activity. The central engine activity duration is thus a sensitive point in our model, and we will discuss more in Sec. 10.5.2.

It should finally be noted that the above considerations are valid for the average pulse. Depending on the actual pulse shape and luminosity, Eq. 10.17 may under- or over-estimate the actual contrast by a factor of a few. Also, spectral effects, not included in this section, could affect the flare visibility but are expected to remain moderate since both the ESD and the

flares shine in the X-rays.

10.3.2 A natural mechanism for narrow flare production

As mentioned in the introduction, a salient property of X-ray flares is their temporal aspect ratio $w/t_{\rm flare}$, where w is the flare width. The flare width has been measured in a number of ways: adopting for w the width of a Gaussian fit to the flare light curve (e.g. Chincarini et al., 2007); a smoothly broken power-law profile (Yi et al., 2016); or using the *Norris profile* (Norris et al., 2005), in which the width w is naturally given by the time span between the two points before and after the peak when the flare flux is a factor of e below the peak flux (Chincarini et al., 2010; Bernardini et al., 2011). Using this last definition, the aspect ratio $w/t_{\rm flare}$—which is independent of redshift—is found to be tightly distributed around $w/t_{\rm flare} = 0.23$ with a standard deviation of 0.14 (Chincarini et al., 2010). In other words, flares' aspect ratios are small and nearly constant in the population. Any flare model must reproduce this fact.

We will now exhibit a natural mechanism built into our misaligned observer picture for flares by which the brighter flares tend to be thin. While we will later estimate $w/t_{\rm flare}$ for actual flare light curves in our model (Sec. 10.3.3), we will temporarily focus on the temporal slope of flares in their post-peak phase, denoted by $\sigma = |{\rm d}\log L(t)/{\rm d}\log t|$. Analytically, we have a better grasp on σ than on w, and for any given profile, larger σ is related to smaller w, allowing us to focus on σ for now.

We introduce the time-dependent stretch factor $S(t)$ (note the different font) which is simply the $\mathcal{S}$-factor of the core material the radiation of which is received at time t by the misaligned observer. The calculation leading to Eq. 10.5 is still valid when applied only to a strip of the shell from which the observer receives radiation at time t. The $f_{\rm geo}$ term changes with time, as the strip will progressively scan the geometry of the shell as seen from the observer's stand point. However, for a generic shell shape, $f_{\rm geo}$ will not change significantly, and we will thus have $L_{\rm flare}(t) \sim f_{\rm geo} L_{\rm on} S(t)^{-3}$. Finally, writing $S(t) = (t - t_{\rm ej})/\tau^{\rm c}$ as for Eq. 10.9, we arrive at the following approximate flare time behavior:

$$L_{\rm flare}(t) = f_{\rm geo} L_{\rm on} \left(\frac{t - t_{\rm ej}}{\tau^{\rm c}}\right)^{-3} \tag{10.18}$$

Taking the logarithmic derivative of Eq. 10.18 at $t = t_{\rm flare}$, we find that the initial post-peak decay index for the flare is:

$$\sigma = \left|-3 - \frac{3t_{\rm ej}}{\mathcal{S}\tau^{\rm c}}\right| \tag{10.19}$$

$$= \frac{3}{1 - \frac{t_{\rm ej}}{t_{\rm flare}}} \tag{10.20}$$

Therefore, coming back to Eq. 10.17, we find that the flare–ESD contrast is linked to the initial decay index by:

$$\frac{L_{\rm flare}}{L^{\rm LOS}_{\rm ESD}(t_{\rm flare})} = f_{\rm geo} \left(\frac{\tau^{\rm c}}{\tau^{\rm LOS}}\right)^2 \left(\frac{\theta_v}{\theta_j}\right)^a \left(\frac{\sigma}{3}\right)^3 \tag{10.21}$$

This last equation shows that, under our interpretation of flares during the ESD, brighter flares—or simply those that appear above the continuum—tend to decay faster, and therefore be thinner, explaining the low values observed for $w/t_{\rm flare}$.

Moreover, the width of flares in our picture is $w \propto \mathcal{S}\Delta t_{\rm on} \sim \mathcal{S}\tau^{\rm c}$, because the prompt pulse duration transforms like the photon arrival times to the off-axis line of sight. The proportionality constant will depend on how exactly the width is measured. It therefore follows that the aspect ratio will be:

$$\frac{w}{t_{\rm flare}} \propto \frac{\mathcal{S}\tau^{\rm c}}{\mathcal{S}\tau^{\rm c} + t_{\rm ej}} \tag{10.22}$$

Ignoring $t_{\rm ej}$, one would find that the aspect ratio is the same for all flares. The ejection time introduces some diversity in the aspect ratio, which we expect to be small as long as the arrival time is mostly determined by the angular effect ($\mathcal{S}\tau^{\rm c}$) and not by the ejection time. In addition, the different values of $\mathcal{S}$ explored for different bursts will introduce some scatter in the distribution of aspect ratios. The explanation of the small scatter in aspect ratio is thus tightly linked to the expected range in $t_{\rm ej}$ and $\mathcal{S}$, which we discuss in Sec. 10.5.2.

10.3.3 X-ray flares from core material prompt dissipation

For more concreteness we will now prescribe the emission physics for the core shell and study the production of flares for the misaligned observer. We place the observer at $\theta_v = 0.13$ rad, at the same place as in our previous computation of plateau emission (Fig. 9.1). The shell of material in the jet core instantaneously radiates a source-frame energy of $E^{\rm c}_{\rm iso} = 10^{53}$ erg while at a Lorentz factor of $\Gamma = 100$, placing us in the hypotheses of our previous calculations. We prescribed a circular shape to the shells with a varying diameter; All the shells are placed tangent to the close edge of the core, as shown in Fig. 10.4. The relevant equations for the luminosity observed from this shell can be found in Appendix C. We assign to this shell $t_{\rm ej}$'s and $\tau^{\rm c}$'s that we will vary. All the light curves that we will show are integrated in the *Swift*/XRT band (0.2–10 keV).

In order to produce the ESD radiation, we consider a shell on the observer's line of sight geometrically covering the entire jet from the line of sight to the core, which dissipates an energy $E^{\rm LOS}_{\rm iso} = 1.2 \times 10^{52}$ erg, consistent with the core-dissipated energy mentioned above and a power-law jet energy structure with $a = 8$. The shell producing the ESD is supposed to be the last flashing shell, to which we assign an ejection time of $t^{\rm LOS}_{\rm ej} = 30$ s (the average duration of long GRBs, Sec. 1.1, von Kienlin et al. 2020) and a pulse duration of $\tau^{\rm LOS} = 2$ s. The line-of-sight shell has a Lorentz factor of 50, consistent with the core shell's Lorentz factor given above and a power-law Γ_0 structure with $b = 3$.

For all shells, we adopt a comoving emission photon spectrum of broken power-law shape, with low- and high-energy slopes $\alpha_{\rm ph} = -1.1$ and $\beta_{\rm ph} = -2.2$ (average slopes found in the *Fermi*/GBM GRB prompt phases, Sec. 1.1, Poolakkil et al. 2021) and a shell-frame peak energy $E'_p = 1$ keV, corresponding

to the average observed E_p of ~ 200 keV for an aligned observer and a Lorentz factor of 100. With their $\mathcal{S} = 12$, all the flares would thus have a peak energy of 17 keV.

In Fig. 10.3, one can find the resulting flare light curves for the core shells (colored lines) and the line of sight shell (ESD, black line). For completeness, we also added the level of a plateau predicted by our misaligned plateau model (Eq. 9.3), with parameters exactly as in Fig. 9.1, except for the uniform external medium density, which we brought down to 10^{-3} cm^{-3} as expected for short GRBs (Chap. 7). With these values adopted, it is as if the core shell dissipated $\sim 10\%$ of the initial available kinetic energy in the core, and the rest served to produce the plateau.

The parameters for the flares can be found in Tab. 10.1: We provide the $t_{\rm ej}$ and τ, the diameter of the shell in units of $1/\Gamma$, the peak flare time $t_{\rm flare}$, the aspect ratio $w/t_{\rm flare}$ as determined with the two points with a flux a factor of e smaller than the peak flux, the contrast $L_{\rm flare}/L_{\rm ESD}$ and, finally, the peak luminosity in the *Swift*/BAT band (15–150 keV) that an observer would detect if aligned with the shell.

First, Fig. 10.3 shows that our model is capable of producing thin flares with occurrence times and luminosities consistent with typically observed flares. Second, Tab. 10.1 shows that, as seen on axis, these flares would produce peak luminosities in the BAT band on the order of 10^{52} erg/s, therefore their properties from the point of view of gamma-ray emission are consistent. Third, an inspection of the table shows that we find aspect ratios slightly smaller than observed: This could be due to our instantaneous-dissipation hypotheses, as any intrinsic duration would naturally broaden the flares leading to more typical flare widths. The flare widths is also linked to the shape of the shells, we will discuss this topic in more detail when fitting *Swift*/XRT data in Sec. 10.4. Therefore, these first results are encouraging.

Furthermore, we find that, at given $\tau^{\rm c}$ (e.g., flares B and C), thinner flares tend to be more contrasted, as discussed in Sec. 10.3.2, though we find the effect to be more pronounced than expected. Similarly, for a same flare-ESD contrast (e.g., flares A and C), it appears that a larger $\tau^{\rm c}$ results in a larger flare, in line with Eq. 10.21 (recall that a larger σ is equivalent to a thinner flare). We also find that the larger $t_{\rm ej}/t_{\rm flare}$ (still ≤ 1 of course), the steeper the post-decay phase and the thinner the flare, as expected from Eq. 10.20. The extreme case here being flare A, with $t_{\rm ej}/t_{\rm flare} \sim 92\%$ (because of its small $\tau^{\rm c}$) resulting in an extremely steep initial decay, with $\sigma \gtrsim 6$. In conclusion, this instance of the model is consistent with the analytical results and observed trends in X-ray flares.

Up to now, we did not discuss the actual size of the shells in the core. In principle, the angular size of the shells are not constrained by the physics in the jet, apart from causality arguments implying their angular size cannot be smaller than $1/\Gamma$. In our model, we find that their size has little influence. In particular, the intuition that smaller shells produce thinner flares is not verified (e.g., flares A and D); It is in fact the contrary, as flare width is mostly determined by $\mathcal{S}$, τ and $t_{\rm ej}$. The shell sizes would only appear in the light curve once radiation from the entire shell has reached the observer: Then the flux would suddenly drop. In all cases however, this occurs at a flux level much lower than the underlying continuum. Finally, note that as their influence on the radiation is minimal, different shell sizes can allow to decrease the energy budget in the shell dissipation.

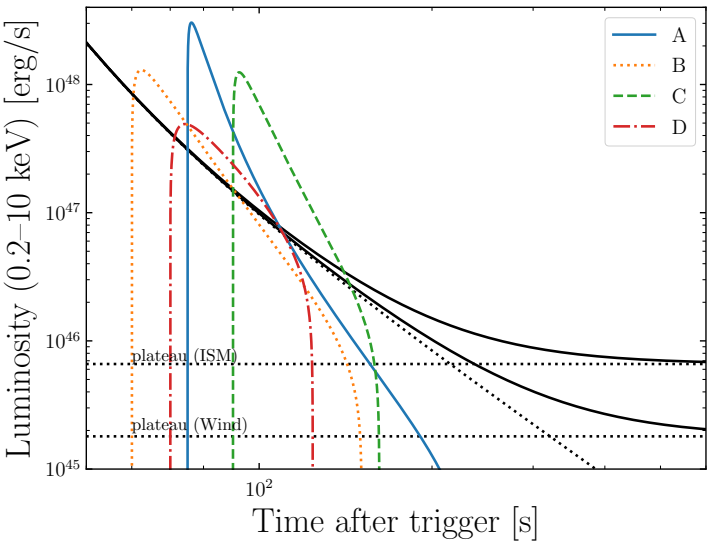


Figure 10.3: **X-ray luminosity expected from flashing core shells, interpreted as flares in our picture**. Colored lines: luminosity from the core shells. Black lines: ESD signal, produced by material on the line of sight, along with typical plateau levels predicted by our misaligned observer plateau model (Chap. 9). The parameters for the shells are reported in Tab.10.1.

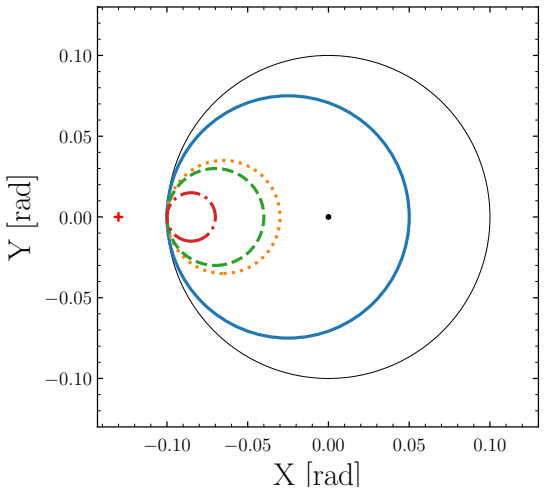


Figure 10.4: **Positions and sizes of the core shells with respect to the misaligned observer.** Red cross: observer's position. Black point: jet axis. Colored circles: core shells, with same color-coding as in Fig. 10.3.

Table 10.2: **Best-fit parameters for XRT data of flares in GRB060719 and GRB100816A with our misaligned-observer interpretation of flares.** We present parameters both constraining the ejection time to $\leq T_{90}$ or letting it free. Notations for flare parameters are the same as in Tab. 10.1.

GRB	060719 ($z = 1.532$, $T_{90,\mathrm{RF}} = 26.4 \pm 4.5$)		100816A ($z = 0.8034$, $T_{90,\mathrm{RF}} = 1.6 \pm 0.3$)	
	t_{ej} free	$t_{\mathrm{ej}} < T_{90,\mathrm{RF}}$	t_{ej} free	$t_{\mathrm{ej}} < T_{90,\mathrm{RF}}$
E_{iso} [erg]	2.0×10^{52}	5.9×10^{52}	3.0×10^{52}	7.1×10^{52}
Γ	78	90	106	75
$\Delta\theta$ [rad]	0.035	0.039	0.026	0.044
t_{ej} [s]	61	31	48	0
τ [s]	1.7	2.7	2.5	5
$\mathcal{S}$	9	9	9	9

10.4 Flare morphology

Beyond the width of flares, it appeared in the first catalog (Chincarini et al., 2007) that flares presented a variety of morphologies, with rising and decay phases being fast or slow, i.e., exponential or power-law profiles. In terms of actual rise and decay times, virtually all flares decay in a longer time than they rise (Chincarini et al., 2010). However, the temporal profiles near flare peak define either very localized peaks in the case of fast rise and decay, or rounder flares in the case of slow rise and decay. Here, we will further assess the capabilities of our misaligned-observer interpretation of flares by studying which flare morphology it is able to capture.

We choose GRB060719 (at $z = 1.532$) and GRB100816 (at $z = 0.8034$), which feature flares peaking at $t_{\mathrm{flare}} \sim 90$ s in the source frame in both cases. The first presents a slow rise followed by a fast peak, while the second has a rounder peak and a slower decline, thus representing the variety in flare morphology. In Fig. 10.5, one can find the XRT data points for these two flares corrected for redshift. We applied the redshift correction by transforming times with $t_{\mathrm{RF}} = t_{\mathrm{obs}}/(1+z)$ and determining flux from luminosity by $F_{[\nu_{1,\mathrm{obs}},\nu_{2,\mathrm{obs}}]} = (1+z)L_{[(1+z)\nu_{1,\mathrm{obs}},(1+z)\nu_{2,\mathrm{obs}}]}/4\pi D_L(z)^2$. We determined the luminosity distance $D_L(z)$ using a generic flat world model with $H_0 = 70$ km/s/Mpc, $\Omega_{\mathrm{M}} = 0.3$, $\Omega_\Lambda = 0.7$.

We proceed by fitting the source-frame XRT light curve by the sum of a broken power law representing the continuum component and our flare model from core shells representing the excess. For a reminder, the off-axis shell model has the following parameters: isotropic-equivalent dissipated energy E_{iso}, shell Lorentz factor Γ, angular distance between the observer and the shell's close edge $\Delta\theta$, shell ejection time t_{ej} and decay time scale τ. The emitted spectrum adopted is the same as in Sec. 10.3.3.

We seek the best-fit model as measured by a χ^2 statistic under the same parameter constraints as in Sec. 10.2: E_{iso} in the range $10^{52.3\pm1}$ erg, τ in the 0.1–5 s range, $\mathcal{S}$ must be $\gtrsim 10$. Concerning t_{ej}, we consider two different conditions: Either we let it vary freely up to t_{flare} or we bound it by T_{90}. Coming back to our discussion on central engine activity, the second condition hypothesizes that this activity's duration is the same on all directions to the central engine, therefore having measured it through T_{90} on the line of sight constrains it in the core; The first condition does not make this hypothesis. In Fig. 10.5 we show the best fits under both these conditions; Best-fit parameters can be found in Tab. 10.2.

First, it seems that satisfying fits can be found to the light curve with reasonable parameter values: In both flares, the energies remain within the allowed region, the Lorentz factors are on the order of 100 as anticipated in Sec. 10.2 and the $\mathcal{S}$-factors are ~ 10, within the slightly misaligned regime. As a result, the ejection times are always less than half the flare occurrence time, meaning both the delayed ejection and the light travel time effects are at play.

Second, it is clear that the hypothesis of instantaneous dissipation in the comoving frame leads to a sharp rise in the flares, and therefore allows our model to better fit fast-rising morphologies such as in GRB100816A than slow-rising ones like in GRB060719. In GRB060719, it seems indeed that an intermediary XRT point at the start of the rise would invalidate our models that predict a deep dip before the peak. While instantaneous dissipation was practical for the analytical calculations above, it is not a requirement of our model. An intrinsic duration of dissipation will change the flare profile during the increasing phase, however the properties of the declining phase—notably, the initial decay slope discussed in Sec. 10.3.2—should remain, as this phase is dominated by the angular exploration of the shell to latitudes further and further from the observer. In particular, the transformation of the duration of a pulse to the duration of the flare (Eq. 10.9) is still valid, when restricted to the declining phase of the pulse and flare.

Similarly, a shell shape different than circular or a unequally bright shell would change the profile of the rising phase. The circular shape we adopted is not particularly physically motivated, and the uniform shell brightness allowed to simplify some derivations; Different prescriptions would not change the main features of our model. We thus conclude that, beyond slightly increasing the flare width to more typical values as mentioned in Sec. 10.3.3, an intrinsic dissipation duration and other shell shapes would allow to better capture slow-rising flares such as in GRB060719. However, the declining phase in both examples is well captured by the model, whether steep (GRB060719) or slow (GRB100816A).

Third, it is obvious that letting the ejection time run free make for much better fits. This is true for GRB060719, where the $t_{\mathrm{ej}} \leq T_{90,\mathrm{RF}}$ restriction seems to not allow the flare to peak

at the right time. A similar issue, though less pronounced, occurs for GRB100816A. We did not statistically compare the goodness of fit of the free-$t_{\rm ej}$ and restricted-$t_{\rm ej}$ models. Nonetheless these two examples show that this hypothesis plays an important role in our model for flares, and we discuss it in more details in Sec. 10.5.2.

10.5 Discussion

10.5.1 Summary

Motivated by our model for plateaus in GRB afterglows in the physical setting of a slightly misaligned observer to a structured jet (Chap. 9), we have presented a new model to interpret flares in the same setup. We suggest that flares can be produced by prompt dissipation in the jet's core that appears in the X-ray band rather than the gamma-rays because the core is less boosted to a misaligned observer than to a on-axis observer. In our picture, the delay in the flare observation with respect to the prompt emission is a combination of both a purely geometrical effect, linked to the photon travel time from the core to the misaligned line of sight and an intrinsic effect linked to the finite duration of the central engine activity, resulting in delays in ejection of different shells.

Writing down the transformations of photon arrival time, spectrum and luminosity from an aligned to a misaligned line of sight, we found that the typical X-ray flare could indeed be explained as deboosted core jet prompt emission (Sec. 10.2). In doing so, we outlined the typical properties required for the core shells responsible for the flares and the expected properties for the resulting flares: Shells have rather low Lorentz factors $\Gamma \sim 100$, carry energies typically on the higher end of those dissipated in single pulses of GRBs $E_{\rm iso} \gtrsim 10^{52}$ erg; The flares thus produced naturally occur early in the afterglow ($t_{\rm flare} \leq 1000$ s), when the majority of flares are observed in XRT light curves.

In the early afterglow phase, the continuum is dominated by the ESD. We therefore analytically studied the conditions for appearance of flares during the ESD, assuming it was produced by high-latitude emission from the last flashing shell on the misaligned observer's line of sight (Sec. 10.3). We found that flare visibility is favored by large shell ejection times $t_{\rm ej}$ and large shell decay timescales. We also exhibited a mechanism present in our model by which brighter flares—or simply those that are able to appear above the continuum—tend to be narrower, echoing the observation that most flares have aspect ratios $w/t_{\rm flare} \lesssim 0.5$. Furthermore, synthetic light curves confirmed these trends between brightness and aspect ratio, the prominent role of the ejection time in flare visibility and, overall, proved that typical flare widths are reproduced by our model (Sec. 10.3.3).

Finally, we made fit of our model to two actual flares observed in XRT afterglows (Sec. 10.4). We chose these two examples so as to represent two different morphologies found in flares: slow-rise-fast-decay (GRB060719) and fast-rise-slow-decay (GRB100816A). We found satisfactory fits with reasonable parameter values and a setup within the slightly misaligned regime, as measured by $\mathcal{S} \sim 1+|\Delta\theta\Gamma|^2 \sim 10$. The model would better capture the slow rises assuming an intrinsic duration of the shell dissipation—our analytical work considered it instantaneous for simplicity.

These fits confirmed the role of the ejection time in defining the flare arrival time and the trade-off between shell energy and ejection time anticipated in Sec. 10.2. Indeed, flare arrival time increases with both $\mathcal{S}$ (more misalignment) and $t_{\rm ej}$ (later ejection), flare luminosity however drastically decreases with $\mathcal{S}$. Thus, for a given flare, increasing $t_{\rm ej}$ allows to decrease $\mathcal{S}$ and the shell energy.

10.5.2 Admissible prompt dissipation mechanisms

In principle, our explanation of flares and our analytical results are compatible with any prompt dissipation mechanism, as long as the declining phases of gamma-ray pulses are due to the angular exploration of the shell, i.e., high-latitude emission. This will be the case for prompt mechanisms with negligible dissipation duration compared to the angular timescale, such as internal shocks (Daigne & Mochkovitch, 1998, 2000). It could also be the case for models with smaller dissipation radii, such as photospheric models, if the central engine turns off sufficiently rapidly to mimic a steep decline as in high-latitude emission. All the other prompt-related parameters of the model such as energy and shell Lorentz factor are generic.

One sensitive point of our model is the admissible range for the ejection time of the core shells. As we mentioned, larger $t_{\rm ej}$ allows for brighter, thinner flares. The shell ejection must naturally occur during the central engine activity. Having observed a GRB as a misaligned observer informs on the central engine activity duration on one's line of sight, on which the dominating material lies; This duration has no reason to be the same in the jet core where we posit the flare-producing shells lie. The example of GRB100816A (Fig. 10.5, right) shows that a better fit is found by allowing $t_{\rm ej}$ to reach typical GRB T_{90} durations of 40 s which, however, are much longer than the actual GRB's source-frame T_{90} of 1.6 s. GRB060719 (Fig. 10.5, left) provides a less drastic example.

These results ask the question of whether the central engine activity can be shorter or longer lasting depending on the ejection direction. Such variation in the activity duration around the central engine is prescribed upstream, by the physical conditions near the compact source.

Because of the significant interaction of the incipient jet with material near the compact object (Sec. 8.3), it could be that the time interval during which relativistic material effectively emerges from the system depends on the latitude. One could speculate that the jet breaks out early and continues to eject matter through the core while the line-of-sight material experiences more interaction with the cocoon and thus relativistic material emergence is shorter-lived. The comparison of relativistic ejection of different lines of sight remain to be studied in numerical simulations, and more fits of our model to XRT data must be done to determine whether the requirement

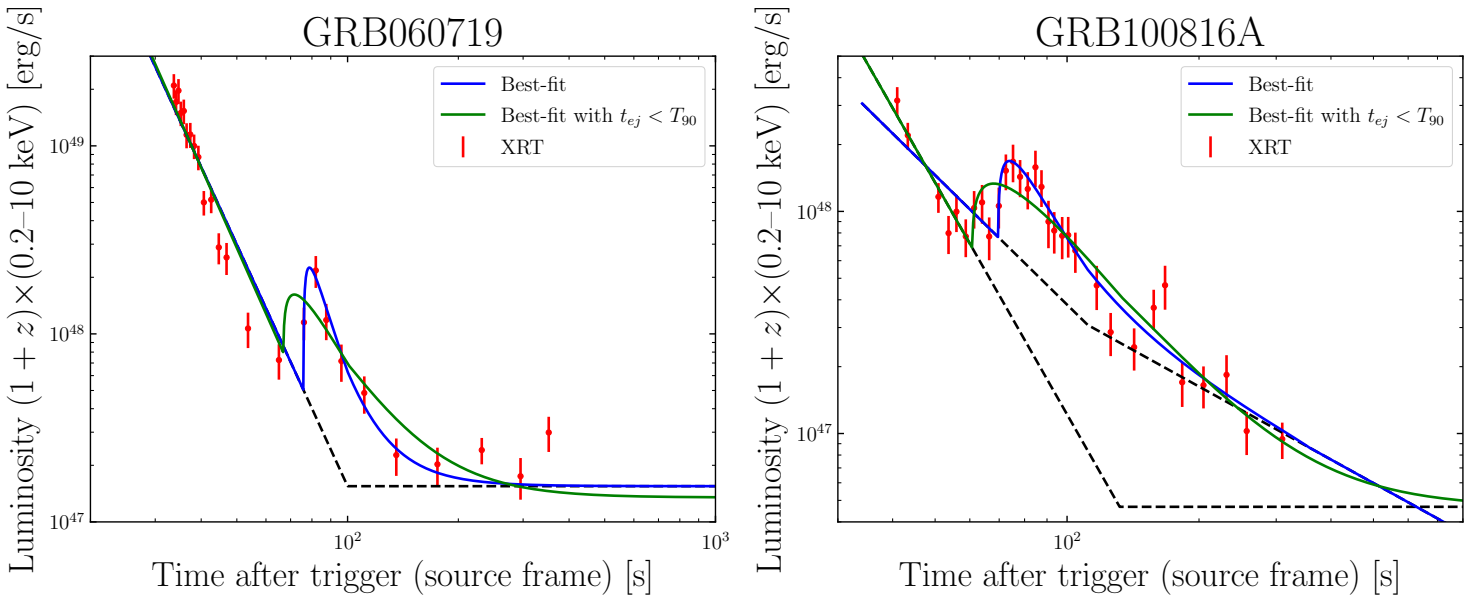


Figure 10.5: **Joint continuum and excess fitting to XRT data for two flares with different morphologies,** using our misaligned core shell dissipation model. Best fits are shown both constraining the shell ejection time to $\leq T_{90}$ (green) or not (blue). The dashed black line shows the best-fit continuum, which differs between the free-$t_{\rm ej}$ and constrained-$t_{\rm ej}$ fits only for GRB100816A (right).

of $t_{\rm ej} \geq T_{90}$ is a general feature.

Generally, a constant $w/t_{\rm flare}$ cannot be a natural consequence of a *composite model*, in which w and $t_{\rm flare}$ are set by unrelated causes. For example, in genuine late central engine activity, w and $t_{\rm flare}$ are determined respectively by the duration and onset time of the second engine episode, the tuning of which to a same value of the aspect ratio from a system to another is not clear, especially seeing the diversity of durations in the first episode, as suggested by GRB durations. In models with changes in the forward shock propagation medium, w and $t_{\rm flare}$ are determined by the size and the propagation time up to the external medium accident, tuning these also seems unnatural.

However, a constant $w/t_{\rm flare}$ is a natural consequence of *single-episode models* such as ours. In these cases, the aspect ratio is determined by a single transformation of prompt emission, in our case by geometrical effects, i.e., photon travel time. The fact that the distribution of rise-to-decay time ratios for flares closely follows that of GRB prompt pulses Chincarini et al. (2010) further encourages such models. In our picture, we showed in Sec. 10.3.2 that a small scatter in flare aspect ratio is obtained by diversity in $t_{\rm ej}$ and $\mathcal{S}$: We expect the geometrical effect ($\mathcal{S}\tau^{\rm c}$) to be dominant in shaping flare aspect ratio, thus producing a tight distribution, and the subdominant $t_{\rm ej}$ to introduce a small scatter.

Another remarkable property of our model is that the flares naturally appear in the X-rays, even if gamma-rays are produced for an aligned observer. It is not obvious why the central engine should shine in the X-rays in other models, especially late central engine activity scenarios in which the activity should be the same as the first. In fact, because of the misaligned nature of the observer in our picture, no simultaneous higher-energy counterpart (e.g., gamma-rays) is expected; We checked this was the case for the two examples treated above in Sec. 10.4. Indeed, it seems that any super-gamma-ray photons which would eventually appear as gamma-rays to the misaligned observer should be suppressed by optical depth to pair production, scattering on pairs and on electrons (Matsumoto et al., 2019a,b); The X-rays however are not affected. However, a thorough establishment of this feature in our geometrical setup remains to be done.

10.5.3 Observed and expected model consequences

There are rather strong conditions for successful flaring in our model. We discussed the role of the ratio of pulse decay times for the line-of-sight and core material (Sec. 10.3), one which we have little intuition. We showed that the energy budget prohibited very large Lorentz factors, such that we do not expect all the shells in the core to produce flares. Thus selecting the potential flare-producing shells, we therefore do not expect many flares in each GRB. Adding the slightly-misaligned-line-of-sight condition, we do not expect many GRBs with flares in general; The observation is that about one third of *Swift* GRBs exhibit noticeable flares (Yi et al., 2016). To discuss such statistics implies many parameters: the allowed range $t_{\rm ej}$; the jet structure, which conditions both the GRB trigger for off-axis observers and the flare-to-ESD contrast (Eq. 10.17); the Lorentz factors and sizes of the shells within the core. In the perspectives listed in Chap. 11, we will propose to lead such a statistical study to further root our model into flare observations; As the physi-

cal setup is essentially the same, statistics of both flaring and plateau behavior in GRB afterglow would shed further light on our models.

An interesting consequence of our model is the natural dichotomy between early and late flares. With $t_{\rm ej}$ limited to $\lesssim$ 100 s (i.e., typical central engine activity), flares with $t_{\rm flare} \geq 1000$ s are only possible in our picture with large $\mathcal{S}$. However with $L_{\rm flare} \propto \mathcal{S}^{-3} E_{\rm iso}$, we do not expect these flares to be visible. This fact remarkably echoes the observation of a dichotomy in temporal behavior and $\Delta F/F$ distributions between early and late flares (Bernardini et al., 2011): A different origin for late flares is therefore reasonable.

It is also interesting that, for these early flares, Bernardini et al. (2011) found the relation $L_{\rm flare} \propto t_{\rm flare}^{-2.7\pm0.1}$ between flare arrival times and peak fluxes. Considering that $L_{\rm flare} \sim \mathcal{S}^{-3} L_{\rm on}$ and $t_{\rm flare} = \mathcal{S}\tau^{\rm c} + t_{\rm ej}$, the correlation spanned by varying $\mathcal{S}$ should be $L_{\rm flare} \propto t_{\rm flare}^{-3}$ exactly in the absence of shell ejection delays; These delays allow however for later flares with the same luminosity, such that the slope is in fact slightly shallower than –3, as found. Finally, for GRBs with many flares, we expect the later ones to have larger $\mathcal{S}$'s generally, considering the other parameters fixed. We thus expect them to be dimmer and softer; These trends are indeed found in the few GRBs with more than one flare (Chincarini et al., 2010).

If our picture for flares is correct, there are further consequences that we could check in the population. First, while the question of dependence of $T_{\rm CE}$ on the line of sight is delicate, we should generally expect that if T_{90} is large in a given GRB, it should also be large on other lines of sight to the same GRB. As large ejection times favor flares, we expect long T_{90} to be correlated with flaring activity. Second, we describe early flares and plateaus as the consequence of misaligned lines of sight, thus we expect the jet structure to introduce a trend between flaring activity and dim prompt emission. Third, if there is a qualitative or quantitative difference in the prompt dissipation mechanism between the core and the lateral structure, it should appear when comparing prompt emission of GRBs with and without flaring or plateau activity. Indeed in our picture we expect the emitting regions in these two cases to be located in different parts of the jet structure. With this last point, we understand that flares and plateaus could provide key insight into GRB physics, the structure of GRB jets and the difference between long and short GRBs in this respect; In Chap. 11 we will suggest some leads to explore this avenue.

10.5.4 On the nature of flares and the flare sample

Our model posits a common origin for prompt emission and early X-ray flares, and is supported by some similarities between these phenomena. Consequently, it poses the question of the definition of flares and sample contamination in X-ray flares. All X-ray flare samples we mentioned in this chapter (Falcone et al., 2007; Chincarini et al., 2007, 2010; Margutti et al., 2011; Bernardini et al., 2011; Yi et al., 2016) are selected only after visual inspection of the XRT light curves, in search for excess flux over a continuum. While this is justified for late flares (such as the late flare sample of Bernardini et al. 2011), the sample of early flares thus selected must be contaminated by prompt emission as well. Generally, a more physical definition of flares is warranted to better isolate this activity and thus define what exactly models should seek to reproduce.

In our picture, it seems that flares should be characterized by an absence of a counterpart in the higher-energy bands such as the BAT and a significantly different hardness ratio than the underlying continuum emission. This definition excludes, for example, the flares in GRB050820A and GRB110801A (both around 400 s post-prompt), which show a simultaneous BAT excess. Such conditions exclude any flares produced by aligned-observer mechanisms—i.e., the bulk of flare models—and should not be generally enforced in a definition of flares.

Nonetheless, we suggest a more physical definition for the flare sample, requiring that the flaring occurs well after the end of the prompt emission, thus clearly defining two episodes of source activity or a time interval for light travel time. We also suggest to select systems where the hardness ratio is significantly different during the flaring activity than in the underlying continuum, thus enforcing a distinct origin for the flare. These conditions would restrict the flare sample and contribute to better defining what models should seek to reproduce.

Positing a common origin for GRB prompt emission and X-ray flares further adds to the discussion of the possibility that long GRBs, low-luminosity GRBs, X-ray flashes, etc. are the manifestations of the same system viewed from different orientations (Sec. 8.2.3). That fact that the low-luminosity GRB031203, GRB100316D and GRB171205A all show some phases of flat X-ray flux (respectively, Ramirez-Ruiz et al. 2005; Margutti et al. 2013b; D'Elia et al. 2018) supports the idea that they could be misaligned events, in light of our plateau model. Similarly, the investigation of X-ray flashes as misaligned core prompt dissipation with a model as presented in this chapter could shed more light on the nature of these events. In any case, both a statistical study and light curve fits should be carried out.

10.6 Conclusion

In the same physical setup as our plateau model from Chap. 9, we exhibited a novel interpretation for X-ray flares in GRB afterglows. It relies on slightly misaligned lines of sight to a structured relativistic jet, in which the core's prompt dissipation is deboosted to the X-ray band for the off-axis observer and appears during the afterglow—typically, during the early steep decay—because of both the light travel time from the core to the observer and the intrinsic duration of the central engine ejection activity. From order-of-magnitude considerations to actual fits to *Swift*/XRT data, we showed that this model is capable of explaining typical flares. Further, we showed how this model favors flares with small and tightly distributed aspect ratios, a salient property of X-ray flares. Overall, though a thorough statistical study is called for, it appears that there are many trends found in GRB afterglows flares that our picture

naturally produces.

The work presented in this chapter was the last development of the doctoral studies, and it opens up many new avenues for research, which we detail in Chap. 11. First of all, we should make fits of our plateau model to XRT data, and make further fits to flares. This would allow a thorough exploration of parameter space to reveal potential degeneracy and understand the internals of the models. Then, it is important to further assess the conformity of our model with established trends in flare and plateau behavior through statistical study. Furthermore, we should explore the existing flare and plateau sample in search for eventual trends between these features and prompt emission that our model suggests: presence of flares correlated with long T_{90}'s, with dim GRBs, etc. Finally, we could exploit our misaligned observer geometry to constrain the structure of GRB jets through their manifestations as flares and plateaus, and in particular explore the differences between short and long GRBs.

Part IV

Perspectives for multi-messenger high-energy astrophysics

Chapter 11

Thesis summary and proposals for future study

Abstract

In the preceding chapters, we described developments from prospects for multi-messenger astronomy to original gamma-ray burst modeling. In this chapter, we summarize the results of this doctoral thesis and make some proposals for further study. When possible, we describe methodology, expectable results and possible outcomes. These proposals concern: 1. Extending our population model to other afterglow observables; 2. Revisiting standard gamma-ray burst mechanisms in the context of structured relativistic jets with misaligned lines of sight.; 3. Preparing to leverage mildly relativistic afterglows to study merger environments by elucidating the outflow structure; 4. Consolidating our suggestion of using afterglow counterparts to study the delay-time distribution by assessing the link between circum-merger density, natal kicks and delay times; 5. Developing our geometrical interpretation of gamma-ray burst plateaus and flares by studying joint plateau-flaring activity and confronting the model to the phenomenon's statistics.

11.1 Summary of the doctoral work

The binary neutron star (BNS) merger gravitational-wave (GW) signal GW170817 and its electromagnetic counterparts opened up the era of multi-messenger astronomy with gravitational waves. This event featured a GW inspiral signal, a dim short GRB, a kilonova transient, a long-lived multi-wavelength afterglow, and could presently be showing the signature of a kilonova afterglow. The combined leveraging of these signals led to groundbreaking advancement in astrophysics. Legitimate questions after this event are: Should we expect similar combinations of these counterparts to BNS mergers in the future? What new multi-messenger methods can we devise to exploit these upcoming events? GW170817 proved that BNS mergers—and most likely other GRBs—launch structured relativistic jets, what are the consequences of this structure on GRB observations?

In order to answer these questions, we first developed a multi-messenger population model covering all the above-mentioned electromagnetic counterparts (except for the kilonova afterglow, Chap. 4). While GW and GRB physics allow for detailed modeling of the inspiral signal and the afterglow, the newness of kilonova signal observations only allow for quite uncertain modeling; Nonetheless, our study was based on state-of-the-art kilonova modeling calibrated on observations from GW170817. Similarly, the latent uncertainty on the physics of short GRBs observed far from the outflow's axis—such as GRB170817A—only allowed us to consider bright GRBs in our model, i.e., those observed down or near their jet axes. Concerning the population model, we used astrophysically motivated distributions for all physical parameters; in particular the jet energy, which we distributed accordingly with inferred short GRB luminosity functions. The detection model we adopted permits us to study how the detected population evolves with the sensitivity of the GW and electromagnetic instruments and thus to replace our results in the context of future instrumental development and observational strategies.

Among the salient results of this population study (Chap. 5), we found that GW170817 was indeed a very lucky event: It was extremely unlikely to observe such a plethora of electromagnetic counterparts. As the GW horizon recedes thanks to instrumental development, more compact object merger GW triggers are expected; These should however be followed by electromagnetic counterparts in a smaller and smaller fraction of cases, unless the sensitivity of the electromagnetic domain strongly increases, representing a challenge to the astronomical community. For GW triggers further and further away, the first electromagnetic observables to become extremely unlikely to detect are naturally the short GRB—disregarding the unlikely case of an aligned system—and the jetted outflow's angular displacement measured by VLBI imagery. At design sensitivity of the global network of gravitational interferometers,

only the kilonova transient should remain accessible in a significant fraction ($\sim 50\%$) of the $\lesssim 100$ yearly BNS triggers. The jet afterglow light curve should be detectable in about one third of these kilonova-endowed events, however with patience: Most of these afterglows should emerge above the radio threshold only hundreds of days after trigger. Though accessible to follow-up instruments, actually detecting these events posed a challenge during the O3 run of the LVK Collaboration, due to large GW sky maps to search for transients and the sheer number of such transients to categorize and sample in any patch of the sky. The limitation due to large GW skymaps should be met both by more GW interferometers coming online—providing smaller localization maps—and by the arrival of high-cadence large field-of-view optical instruments in both hemispheres such as the ZTF and LSST. Beyond GW triggers, we found that the surveys of such instruments could probe a numerous population of orphan kilonovae and a population of short GRB–kilonova associations; These events could play an important role in improving a kilonova models and further assessing the BNS merger-short GRB connection.

Though rare, the jet afterglows should reveal extremely interesting, because seen most often from a significantly misaligned line of sight, like GW170817. This allows a detailed study of the jet structure and interesting applications to measure the Hubble constant with greater precision (Chap. 6) and to constrain the environments of BNS mergers (Chap. 7).

Inspiral GW signals are fundamentally different from electromagnetic signals because the distance to the source can be measured directly from the waveform. In such measurements, the main degeneracy of the distance is with the orbital inclination angle of the system. In trying to measure the Hubble constant on a given system using the source's redshift acquired from its host galaxy by a kilonova counterpart, this degeneracy intrinsic to GW data spoils the measurement. Incidentally, the system's inclination angle is information one can also constrain thanks to an afterglow dataset. This suggests jet afterglows as means to lift the distance–inclination degeneracy and make for better Hubble constant measurements on an event-to-event basis; This was the case with GW170817's afterglow, which improved the measurement threefold. Jet afterglows are, however, unlikely to be detected after most GW triggers. In Chap. 6, we showed that, though every single afterglow provides a very-improved measurement of the Hubble constant, the rareness of these counterparts largely compensates, such that, overall, afterglows will not allow to narrow-down the constant faster than with GW data and kilonova-provided redshifts alone. While this disqualifies afterglows for multi-messenger cosmology endeavors, we showed that the more frequent kilonova signals could play a role, provided models improve to the point where a single kilonova light curve allows to measure the inclination angle with ten-degree precision. This stage should be reached thanks to the accumulation of a large sample of kilonovae with different viewing angles. Then, care must be taken in the treatment of selection effects when using kilonovae to measure the Hubble constant. Indeed, these are not negligible in the regime of redshifts probed by design-level GW interferometers.

Because of the long delay times between formation and merger, it is generally expected that BNS mergers—and short GRBs—occur in rarefied media. This was the case of GW170817, with inferred circum-merger number densities of $\lesssim 10^{-2}$ cm^{-3}. Over the years, a large set of indirect probes of the environments of BNS mergers has accumulated: analysis of the scarce set of short GRB afterglows; offsets of short GRBs from their host galaxies; abundances of r-process elements in various galactic environments. Taken together, this evidence is inconclusive on the existence of a population of high-density mergers, occurring in regions with densities above 1 cm^{-3}. While some binary evolution studies suggest that some systems could merge in times much less than generally expected, possibly in higher-density media, such a population of BNS mergers remains elusive. In Chap. 7, we devised a multi-messenger method based on GW data and the observation of misaligned afterglow light curves that can allow to constrain the circum-merger density of given events. While the modeling of classical GRBs' afterglow produces strong degeneracy between parameters, the combination of GW data with afterglow peak flux and peak time information allows to narrow down the density sufficiently to categorize events as low- or high-density. This method would permit to study a hypothetical population of high-density events: in the local universe using current interferometers, and around the peak of star-formation with third generation instruments. What's more, afterglow flux is strongly dependent on density, such that, if their is indeed a population of high-density mergers, these should be over-represented in detected samples to the point that only a few events should reveal them—or, if none is observed, strongly constrain them. In this method, it is the requirement of prior GW trigger that drives both the density measurement and the statistical effect that this population, if it exists, should be revealed early in the multi-messenger era.

The GW170817 event sparked a renewed interest in the structure of GRB outflows; It showcased how relativistic structured jets—an energetic core surrounded by less extreme material—can imprint the afterglow observed from lines of sight significantly misaligned with the jet's axis. Such afterglows bear signatures of geometrical effects linked to light travel delays between the core and the observer and delayed opening of the core's beaming cone. In Chap. 8, we motivated the study of structured jets and relativistic jets in general in various astrophysical contexts. We exhibited their main characteristics and the physical processes that shape these outflows. In the context of GRBs, the outflow structure is mostly determined by the interaction of the jet launched by the central engine with the layer of dense material surrounding the engine: the collapsar's envelope for long bursts, and the merger ejecta for short bursts. We summarized the known effects of the relativistic nature of these jets on the physics of GRBs. In particular, we discussed the fact that the viewing angles up to which bright cosmological GRB jets are observed cannot be larger than about twice the opening angle: On the contrary to the jet in GW170817—a very close event—, classical GRB jets are viewed at most at slightly misaligned lines of sight. Nonetheless, on these near-

core lines of sight, the geometrical and relativistic effects still play a predominant role in shaping the prompt and afterglow emission. In the physical setup of slightly misaligned observers to structured jets, we developed two models to explain each of the recurring features of GRB X-ray afterglows: plateaus (Chap. 9) and flares (Chap. 10).

The first model is in fact an enhancement of a proposal made shortly after the discovery of X-ray plateaus in GRB afterglows. During the early phases of the afterglow, the core jet is too beamed for its radiation to reach a misaligned observer; Initially, they only see radiation from the less energetic material near their line of sight. As the structure progressively decelerates, the beaming cone from material closer and closer to the core opens so as to include the observer's line of sight. Because of the structure of the jet, the material that is progressively revealed is more intrinsically energetic, however less Doppler-boosted. This mechanism can produce long-lasting phases of nearly flat flux, i.e., plateau behavior. Once the core is revealed, no more material has yet to join the observer's line of sight, and the light curve thus transitions to an ordinary decay phase, just as for an aligned observer. In Chap. 9, we thoroughly explored this model showing both synthetic light curves with plateau behavior and analytical developments allowing to predict the plateau's duration and flux level as a function of the jet structure and the observer's position. Furthermore, these analytical developments showed that some robust observed correlations between plateau and prompt properties in GRBs with plateaus are natural consequences of our geometrical interpretation: Notably, brighter plateaus are shorter-lived and associated with brighter prompt emission.

Motivated by the success of the plateau model, and acknowledging apparent trends between plateau and flaring behavior in GRB X-ray afterglows, we proposed a new interpretation for these flares in the same physical setup: a slightly misaligned observer to a structured jet. In Chap. 10, we suggested that X-ray afterglow flares are in fact the signatures for slightly misaligned observers of prompt dissipation occurring in the core jet, while for these same observers the prompt emission comes from material down the line of sight, in the lateral structure. While an aligned observer detects core prompt pulses in the gamma-ray band, slightly off-axis observers detect this emission in the X-rays because of less favorable Doppler boosting; They naturally detect them later because of light travel time from the core. Positing a common origin for flares and prompt pulses naturally provided the observed similarities between these two phenomena such as temporal profiles and spectral lag. Order-of-magnitude calculations show that the typical flare can indeed be reproduced in this mechanism assuming total dissipated energies in such episodes consistent with what we know on prompt dissipation in GRBs. Our picture applies to the class of early flares—occurring typically during the early steep decay phase or early plateau—, which are statistically dominant in the population and likely have an origin distinct from late flares. A refined analysis showed that our model favors the emergence of thin flares and suggests that the widths of flares should be tightly distributed, both remarkable established properties of X-ray afterglow flares. Finally, light curve fits to actual events showed the model is capable of reproducing the diverse morphologies of flares with reasonable parameter values and maintaining a genuinely slightly misaligned line of sight. They also revealed the sensitiveness of the model to the time interval during which one allows the shells responsible for the flares to be ejected, i.e., the central engine activity. Better fits are found allowing for an ejection activity in the core longer than the GRB's T_{90}, though much less than the occurrence time of the flare. The duration of the central engine activity is linked to the smaller-scale physics around this engine, and the T_{90} is only an underestimate of this duration. What's more, as the misaligned observer's prompt emission come from material on the line of sight, T_{90} is only an estimate of the ejection duration on this direction. Nonetheless, the apparent requirement of an activity lasting for longer than the T_{90} poses the question of the variation of the engine's ejection activity on different directions. Note that both our models for plateaus and flares are set in the same physical setup, and apply to both long and short GRBs.

11.2 Proposals for future study

We will now describe some possible developments motivated by the results of this doctoral work.

11.2.1 Extending the population model to other afterglow observables

Our population model (Chap. 4) covers the main observables of the multi-wavelength afterglow: photometry and source angular displacement; These were the most determining in the astrophysical analysis of GW170817. While the proper motion should reveal extremely difficult to detect for upcoming events, there is another observable which could stay accessible: radio polarization. In the case of GW170817, an upper limit on the linear polarization of the afterglow near its peak was established (Corsi et al., 2018). While this feat was less conclusive than the VLBI imagery, it did somewhat constrain the outflow at a time when the jetted outflow was not fully established: The outflow could not be a relativistic jet unless the shock-frame magnetic field had a significant component in the direction orthogonal to the shock. It is of course the large uncertainty in the magnetic field geometry near the shock that rendered the polarization constraint weak. In the perspective of a rarefaction of source proper motion measurements on BNS merger afterglows, it seems legitimate to treat the polarization in a population-study approach, as we did for the other observables.

For polarization measurements, the structure of the jet plays an important role, and must be taken into account (Gill & Granot, 2018). Naturally, the geometry of the magnetic field is also fundamental and should become a system parameter of the population model. Apart from this additional parameter, the calculation of radio polarization is essentially the same as for photometry: The integration over the jet structure can be led in a similar manner as in our afterglow light curve calculations

described in Sec. 4.4.5; It is simply the integrand which should be adapted accordingly with the well-known physics of synchrotron radiation polarization (e.g. Granot & Königl, 2003). The observable could simply be the linear polarization of the radio afterglow radiation at afterglow peak.

The aforementioned constraint on GW170817 together with the current knowledge that a relativistic jet did indeed emerge from this system suggests that the magnetic field had a significant component perpendicular to the shock. Such a geometry could be starting point for a population study, with an eventual variation allowing some non-zero parallel component, in order to quantify the uncertainties linked to the field geometry, just as we did by adopting two extreme jet energy distribution functions in our population study.

Such an afterglow polarization population study could quantify how often a polarization measurement can be expected to be made and whether this observable—which received little echo in the case of GW170817—could play the role of constraining BNS merger outflow structures when the VLBI measurements will become inaccessible.

11.2.2 Analytical studies for the gamma-ray dissipation mechanism in the lateral structure of GRBs

The short GRB counterpart to GW170817 was particularly dim for its hardness, in the light of other short GRBs. This motivated alternative mechanisms for the origin of GRB170817A, such as relativistic shock breakout of the incipient jet from the merger ejecta; We commented on this in Sec. 3.2.3. This mechanism occurs especially for the very off-axis lines of sight to the system, as was the case for GW170817. This mechanism is qualitatively different that those advanced for bright GRBs, which are observed down or near their relativistic jets; We briefly outlined these mechanisms in Sec. 1.2.6: internal shocks, magnetic reconnection, etc. It is thus legitimate to explore whether these "classical" GRB mechanisms can still occur in the lateral structure of GRB jets.

For very high latitudes, the very small Lorentz factors likely hinder these mechanisms: For GW170817, the joint imagery–photometry fit suggests $\Gamma \sim$ a few at the viewing angle we had. However, for the intermediate latitudes with sizable Lorentz factors, we should study if the conditions to produce gamma-rays in these mechanisms are still met. Indeed, in the lateral structure, the material's energy and Lorentz factor depend on the latitude and so do, e.g., the optical depth to pair creation, the photospheric radius of the outflow. Depending on the structure, the escaping of gamma-rays or production of non-thermal spectra could no longer be possible very far from the core.

Up to now, only a very off-axis line of sight—in GRB170817A—and slightly misaligned lines of sight—classical GRBs—have been explored in observations (for low-luminosity GRBs, the line of sight is not certain). However, further short GRB counterparts to compact object mergers could explore a larger range, possibly revealing the signatures of a diversity of emission mechanisms, from shock breakout to quantitatively different (e.g., softer) versions of the classical mechanisms.

11.2.3 Quantifying diversity in kilonova outflows and their afterglows

The multi-messenger method we devised in Chap. 7 to constrain the circum-merger density relies on GW data and basic merger afterglow information: essentially the afterglow peak flux and time of peak. We mentioned the eventual density constraint that could arise from the observation of the peak flux and time of peak of the kilonova afterglow. As a reminder, this is non-thermal radiation similar to the jet afterglow, arising however in the mildly relativistic shock formed as the much slower kilonova outflow decelerates in the cicrum-merger environment. For GW170817, we could be currently witnessing the emergence of this counterpart in the afterglow (Sec. 3.4). Leveraging this counterpart to improve our method would be all the more profitable that kilonova afterglows are less sensitive to the viewing angle because of their only mildly relativistic nature: They could directly provide a density constraint.

Beyond the circum-merger density, the peak features of this afterglow are most sensitive to the microphysical conditions in the forward shock, and the structure of the outflow. Concerning the microphysical parameters, they are very uncertain, as these kilonova shocks probe a regime intermediate between the slower supernova remnant shocks ($\beta\Gamma \ll 1$) and the faster mildly relativistic shocks of, e.g., internal shock GRB prompt mechanisms ($\beta\Gamma = 1.2$–1.5); Determining the electron-energy and magnetic redistribution parameters in these shocks should await more observations or advancements in plasma physics. The uncertainties in these parameters can nonetheless be taken into account empirically by broadening the subsequent density constraint, as we did in the afterglow constraint of Fig. 7.1.

The other sensitive unknowns of the kilonova afterglow are the outflow's minimal velocity and stratification index. Indeed, the kilonova outflow is likely radially stratified, with a kinetic energy cumulative distribution as $E(\leq \beta\Gamma) \propto (\beta\Gamma)^{\zeta}$ for $\beta\Gamma \geq (\beta\Gamma)_{\rm min}$, where ζ is the stratification index and $(\beta\Gamma)_{\rm min}$ is the outflow's minimal velocity (e.g. Hotokezaka et al., 2018b). The peak of the afterglow is reached when the slowest shell has caught up to the forward shock, and thus defined by the minimal velocity. The temporal slope of the afterglow light curve is defined by the rate of energy injection into the forward shock, and thus by the stratification index. These two parameters are thus crucial to interpret kilonova afterglows.

Nedora et al. (2021) pioneered the study of kilonova afterglows starting from outflows resulting from ab-initio numerical relativity simulations of BNS mergers. Using a selected sample of nuclear equations of state and binary component mass ratios and corresponding simulated outflows, they calculated a number of plausible kilonova afterglow light curves, thus probing expected diversity in these. They conclude that, broadly, the rebrightening in GW170817's afterglow is consistent with this set of light curves.

In doing so, the authors jump directly from the binary parameters (e.g., mass ratio and tidal deformability) and numer-

ical outflows to the afterglow light curve. However, for actual kilonova afterglow light curve fitting or simply analytical manipulations, these heavy numerical simulations are not tractable. For this reason, we could suggest an intermediate approach where we use a sample of merger numerical simulations with varying binary parameters in order to *measure* the resulting sensitive outflow parameters $(\beta\Gamma)_{\rm min}$ and ζ in the simulations. This approach would allow to study their dependence to the binary parameters, to quantify the expected diversity in these outflow parameters and, possibly, obtain tabulated values for them or fit a empirical functional dependence to the binary parameters. Such an approach would allow to maintain a semi-analytical method for kilonova afterglow fitting and parameter estimation—such as density—while being informed by detailed outflow simulations.

Such an approach has already been used in the context of the jet afterglow, where ab-initio simulations allow to measure such parameters as the jet opening angle or energy and Lorentz factor structures. Once these are measured, a semi-analytical method is used to do the actual afterglow study (Sec. 8.3 and references thereabout).

11.2.4 Assessing the link between BNS circum-merger densities and their delay-time distribution

In the review of the accumulating evidence on the environments of BNS mergers (Chap. 7, Sec. 7.2), many are indirect. Notably, the methods relying on r-process abundances are more constraining on the delay time distribution of BNS than on the actual environments of the mergers. There is an intuitive link between the delay time and the merger medium of a binary: Longer delays allow a longer migration and possibly a rarefied or even extra-galactic merger locus. This link—delay time to merger medium density—is not currently quantified, and is influenced by many parameters: the natal kicks of the binaries, their dynamics in the galactic potential, the density profile of the galaxy, the typical size of the high-density regions in the galaxy.

To assess the link between delay time and merger density, we could try to answer the following question: Given a newly formed BNS that is kicked away from its dense formation medium, can we statistically describe its merger medium? In order to answer this question, we could propose very simple dynamical simulations where we launch kicked systems in a galactic potential endowed with a galactic density profile and dense HI or molecular gas regions. One could then determine the probability of a high-density merger as a function of delay time and kick, galactic model, etc. One could answer such questions as: What is the maximal kick velocity or delay time so that the binary merges in its own dense formation region? In other words: After how long a migration do systems forget about their formation environments? Coming back to our perspective of r-process abundance constraints, this would allow to assess to what extent rarefied merger media require longer delay times, and whether it is consistent to observe both early r-process-element enrichment and most short GRBs with large offsets from their hosts.

Such Newtonian-gravity simulations were performed after GW170817 in Abbott et al. (2019a). Their analysis shows that GW170817 is consistent with many formation scenarios involving very short delay times (constrained median delay ≤ 10 Myr), thus encouraging a statistical study of the merger locus for these fast-merging systems.

11.2.5 Developing the geometrical interpretation of gamma-ray burst plateaus and flares

The models developed in Chaps. 9 and 10 to interpret plateau and flaring activity in GRB afterglows call for many developments.

First, we should further ground our models in observations by making more fits to actual afterglow light curves with our new model. Concerning plateaus, a single fit has yet to be done: We have only studied the plateau-prompt correlations until now. The plateau data is not the only that our model should seek to reproduce: In our picture, the prescribed jet structure is also responsible for the misaligned observer's prompt emission and the early steep decay, attributed to high-latitude emission from the last flashing shell on the observer's line of sight. A full fit should therefore prescribe a jet structure in Lorentz factor and in energy, such that the plateau is reproduced and the prompt emission and early steep decay are consistent with having dissipated only a fraction of the total line-of-sight energy—it is only the last flashing shell—, which in turn must be consistent with the jet structure and the kinetic energy remaining for the afterglow phase. Further richness could be brought by plateaus observed in the optical as well, as the natural outcome of our model is chromatic light curves and spectral features would greatly constrain the model. A similar perspective exists for flares, with their own model parameters such as core shell ejection time. Numerous fits for both models would contribute to the legitimacy of the models and reveal possible degeneracy and unconstrained degrees of freedom.

In flares for instance, we discussed the general degeneracy between shell ejection time and observer viewing angle in our model. It could be that fitting many flares reveals that flare width often forces a certain regime of ejection times, thus removing an apparent degree of freedom in the model. Naturally, such restrictions in the model's parameter space greatly influences the statistical predictions of the model, that we discuss below.

Second, in our picture, the appearance of plateaus and early flares are both tied to the slightly misaligned lines of sight that are probed in classical GRBs. It is thus natural to expect combined plateau and flaring activity in GRB afterglows. Beyond the challenge of consistent joint plateau–flare fits to X-ray data, we should confront our model predictions to the statistics of plateau, flaring and combined activity in X-ray afterglows. Here, the essential ingredient is the jet structure, which alone defines many quantities such as: the number of systems that

lead to a prompt trigger when exploring the structure in viewing angle; the number of systems in which a flare emerges from the early steep decay; the distribution of plateau durations—the further from the core the structure allows to go, the more low-luminosity and long plateaus are expected—; for nearer-core lines of sight where the plateau starts early, the number of systems in which flares emerge from the plateau. We immediately understand that the structure partially defines the four main high-level observables: the number of bursts with flares, those with plateaus, those with both and those with neither.

The other main prescription of such a study is the allowed parameter space for the model. In the flare model, fits suggest that the duration of central engine activity could be longer in the core jet than on lateral lines of sight, though still on the order of long GRB typical durations. Clearly, allowing for longer activity increases the number of emerging flares, and the permitted interval for shell ejection thus appears as a sensitive point of the study.

Before any statistical inference, a careful selection of the sample must be done. In our picture, the greater opacity to pair creation from the core to misaligned lines of sight should suppress any gamma-ray counterpart to X-ray flares, though a thorough examination of this effect should be led beforehand. Some early flares do present such counterparts and are likely contamination from prompt emission and not due to the same mechanism, these should be removed from the sample. Finally, note that, as the jet structure is at the heart of both a statistical study or and light curve fitting to flares and plateaus, these developments are an open door to study the differences between jet structures in short and long GRBs.

Third, again because the appearance of plateaus and flares is tied to the slightly off-axis line of sight to the jet, our model predicts a dichotomy between GRBs with plateaus or flares and GRBs without these features. Indeed, we predict that GRBs with and without flares or plateaus probe different regions of the jet—because the prompt gamma-rays arise from the line-of-sight material in any case. For example, GRBs with these features should be less energetic on average, and possibly softer. We could lead a specific study to assess this dichotomy in prompt phase characteristics between GRBs with and without afterglow features. Beyond the prompt energetics or spectral features, this dichotomy could in fact reveal further particularities, for example in duration or pulse profile. This would open a perspective to study the gamma-ray dissipation mechanism specifically outside the core of relativistic jets. The outcome of such an endeavor would allow us to better understand the still elusive mechanism for gamma-ray dissipation.

11.3 Conclusion

The doctoral work presented in this manuscript explored many avenues of high-energy astrophysics research opened with the multi-messenger era of gravitational-waves: predicting upcoming populations of multi-messenger events, seeking how to leverage these events to the profit of understanding the environments of BNS mergers or precision cosmology, and exploring the consequences of the first lessons of this new era on the physics of GRBs.

In this chapter, we described possible developments, to further invest the vibrant field of high-energy multi-messenger astrophysics. These proposals are multidisciplinary: classical GRB afterglow physics, hydrodynamical simulations, galactic structures, statistical inference, etc. For each of the projects, we sought to provide first steps to enter the study, the main sensitive points to consider, expectable results and some references for inspiration.

Postface

This concludes the doctoral work presented in this manuscript. This work was motivated by the historic observation of the binary neutron neutron star merger GW170817 in both the gravitational and electromagnetic sector; in other words, the opening of a new multi-messenger era.

In the course of this work, the O3 run of the global interferometer network took place. The gravitational-wave triggers produced very large areas to search for electromagnetic counterparts, challenging astronomical follow-up, and eventually none was found. The near future after theese doctoral studies will hold the O4 run, with even more triggers and much improved estimates of source sky location, thanks to new interferometers; It is likely that this run will see electromagnetic counterparts to gravitational-wave signals anew. This gravitational-wave observing run will be all the more interesting that it should coincide with the launch of novel electromagnetic facilities such as *SVOM* and generally with better-organized follow-up.

In the doctoral work presented here, I have made predictions for the observations of these upcoming runs, and developed tools to interpret the events. These tools rely on the combination of gravitational-wave and electromagnetic information, they are multi-messenger methods. The development of such methods should gain momentum in the future; We are only at the dawn of this new multi-messenger era. From the point of view of gamma-ray burst science, this era and the new insight it brings constitutes another revolution in the field. There is now much work to do, both coming back to the existing sample of bursts with new ideas and collecting new samples tied with data from other messengers.

In the last chapter, I made some proposals for future study on the modeling side. This plan both naturally falls into the domain of theoretical high-energy astrophysics and is an occasion to open up to new methodologies and new astrophysical communities.

Bibliography

Abadie J., et al., 2010, Nuc. Instr. and Meth. in Phys. Res. Sec. A, 624, 223

Abbott B., et al., 2004, Nuc. Instr. and Meth. in Phys. Res. Sec. A, 517, 154

Abbott B. P., et al., 2016a, Phys. Rev. Lett., 116, 061102

Abbott B. P., et al., 2016b, Phys. Rev. Lett., 116, 241102

Abbott B. P., et al., 2017a, Phys. Rev. Lett., 119, 141101

Abbott B. P., et al., 2017b, Phys. Rev. Lett., 119, 161101

Abbott B. P., et al., 2017c, Phys. Rev. Lett., 119, 161101

Abbott B. P., et al., 2017d, Nature, 551, 85

Abbott B. P., et al., 2017e, ApJ, 848, L12

Abbott B. P., et al., 2017f, ApJ, 848, L13

Abbott B. P., et al., 2019a, Phys. Rev. X, 9, 011001

Abbott B. P., et al., 2019b, Phys. Rev. X, 9, 031040

Abbott B. P., et al., 2020a, Liv. Rev. in Rel., 23, 3

Abbott B. P., et al., 2020b, ApJ, 892, L3

Abbott R., et al., 2021a, Phys. Rev. X, 11, 021053

Abbott R., et al., 2021b, ApJ, 913, L7

Abbott R., et al., 2021c, ApJ, 915, L5

Abdikamalov E., Pagliaroli G., Radice D., 2020, arXiv e-prints, p. arXiv:2010.04356

Abdo A. A., et al., 2009, ApJ, 707, 580

Accadia T., et al., 2012, Journal of Instrumentation, 7, P03012

Ajello M., et al., 2018, ApJ, 863, 138

Akerlof C., et al., 1999, Nature, 398, 400

Albert A., et al., 2017, ApJ, 850, L35

Alexander K. D., et al., 2018, ApJ, 863, L18

Almualla M., et al., 2021, MNRAS, 504, 2822

Amati L., 2006, MNRAS, 372, 233

Amati L., et al., 2002a, Astron. Astrophys., 390, 81

Amati L., et al., 2002b, A&A, 390, 81

Amati L., et al., 2018, Advances in Space Research, 62, 191

Andreoni I., et al., 2020, ApJ, 904, 155

Antier-Farfar S., 2016, PhD thesis, Université Paris-Sud, http://www.theses.fr/2016SACLS467

Argast D., Samland M., Thielemann F.-K., Qian Y.-Z., 2004, A&A, 416, 997

Ascenzi S., et al., 2019, MNRAS, 486, 672

Ascenzi S., Oganesyan G., Branchesi M., Ciolfi R., 2021, Journal of Plasma Physics, 87, 845870102

Atteia J. L., et al., 2017, ApJ, 837, 119

Ayache E. H., van Eerten H. J., Daigne F., 2020, MNRAS, 495, 2979

Baiotti L., Rezzolla L., 2017, Reports on Progress in Physics, 80, 096901

Balasubramanian A., et al., 2021, ApJ, 914, L20

Band D., et al., 1993, ApJ, 413, 281

Barbieri C., Salafia O. S., Colpi M., Ghirlanda G., Perego A., 2020, arXiv e-prints, p. arXiv:2002.09395

Barnes J., 2020, Frontiers in Physics, 8, 355

Barnes J., Zhu Y. L., Lund K. A., Sprouse T. M., Vassh N., McLaughlin G. C., Mumpower M. R., Surman R., 2020, arXiv e-prints, p. arXiv:2010.11182

Behroozi P. S., Ramirez-Ruiz E., Fryer C. L., 2014, ApJ, 792, 123

Belczyński K., Bulik T., 1999, A&A, 346, 91

Belczynski K., Perna R., Bulik T., Kalogera V., Ivanova N., Lamb D. Q., 2006, ApJ, 648, 1110

Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289

Bellm E. C., 2016, PASP, 128, 084501

Bellm E. C., et al., 2019, PASP, 131, 018002

Beloborodov A. M., Mészáros P., 2017, Space Sci. Rev., 207, 87

Beloborodov A. M., Stern B. E., Svensson R., 2000, ApJ, 535, 158

Beniamini P., Hotokezaka K., 2020, MNRAS, 496, 1891

Beniamini P., Mochkovitch R., 2017, A&A, 605, A60

Beniamini P., Nakar E., 2019, MNRAS, 482, 5430

Beniamini P., Piran T., 2016, MNRAS, 456, 4089

Beniamini P., Piran T., 2019, MNRAS, 487, 4847

Beniamini P., van der Horst A. J., 2017, MNRAS, 472, 3161

Beniamini P., Nava L., Duran R. B., Piran T., 2015, MNRAS, 454, 1073

Beniamini P., Hotokezaka K., Piran T., 2016, ApJ, 829, L13

Beniamini P., Dvorkin I., Silk J., 2018, MNRAS, 478, 1994

Beniamini P., Petropoulou M., Barniol Duran R., Giannios D., 2019, MNRAS, 483, 840

Beniamini P., Duque R., Daigne F., Mochkovitch R., 2020a, MNRAS, 492, 2847

Beniamini P., Granot J., Gill R., 2020b, MNRAS, 493, 3521

Berger E., 2014, Ann. Rev. in Astron. and Astrophys., 52, 43

Bernardini M. G., Margutti R., Chincarini G., Guidorzi C., Mao J., 2011, A&A, 526, A27

Bernardini M. G., et al., 2015, MNRAS, 446, 1129

Biwer C. M., Capano C. D., De S., Cabero M., Brown D. A., Nitz A. H., Raymond V., 2019, PASP, 131, 024503

Blaauw A., 1961, Bull. Astron. Inst. Netherlands, 15, 265

Blanchet L., Damour T., 1989, Annales de L'Institut Henri Poincaré Section (A) Physique Théorique, 50, 377

Blandford R. D., Znajek R. L., 1977, MNRAS, 179, 433

Bloom J. S., Sigurdsson S., Pols O. R., 1999, MNRAS, 305, 763

Bloser P., Ryan J., 2008, Spie Newsroom

Boersma J., 1961, Bull. Astron. Inst. Netherlands, 15, 291

Boër M., Gendre B., Stratta G., 2013, ApJ, 800

Brandt N., Podsiadlowski P., 1995, MNRAS, 274, 461

Braun R., 2014, SKA Technical Document, SKA-TEL.SCI-SKO-SRQ-001

Bridle A. H., Hough D. H., Lonsdale C. J., Burns J. O., Laing R. A., 1994, AJ, 108, 766

Briggs M. S., et al., 1999, ApJ, 524, 82

Briggs M. S., Tajima H., Dermer C. D., 2009, Science, 323, 1688

Bromberg O., Nakar E., Piran T., Sari R., 2011, ApJ, 740, 100

Burrows D. N., et al., 2005, Science, 309, 1833

Cantiello M., et al., 2018, ApJ, 854, L31

Casanueva Diaz J., 2018, Control of the Gravitational Wave Interferometric Detector Advanced Virgo. Springer Theses, doi:10.1007/978-3-319-96014-2_2

Chassande-Mottin E., Leyde K., Mastrogiovanni S., Steer D. A., 2019, Phys. Rev. D, 100, 083514

Chen H.-Y., 2020, Phys. Rev. Lett., 125, 201301

Chen H.-Y., Fishbach M., Holz D. E., 2018, Nature, 562, 545

Chen H.-Y., Holz D. E., Miller J., Evans M., Vitale S., Creighton J., 2021a, Class. and Quant. Gravity, 38, 055010

Chen H.-Y., Holz D. E., Miller J., Evans M., Vitale S., Creighton J., 2021b, Class. and Quant. Gravity, 38, 055010

Chincarini G., et al., 2007, ApJ, 671, 1903

Chincarini G., et al., 2010, MNRAS, 406, 2113

Christensen N., 2019, Reports on Progress in Physics, 82, 016903

Cobb B. E., Bailyn C. D., van Dokkum P. G., Natarajan P., 2006, ApJ, 645, L113

Corsi A., et al., 2018, ApJ, 861, L10

Côté B., Ritter C., O'Shea B. W., Herwig F., Pignatari M., Jones S., Fryer C. L., 2016, ApJ, 824, 82

Côté B., Belczynski K., Fryer C. L., Ritter C., Paul A., Wehmeyer B., O'Shea B. W., 2017, ApJ, 836, 230

Coughlin M. W., Dietrich T., Margalit B., Metzger B. D., 2019a, MNRAS, 489, L91

Coughlin M. W., et al., 2019b, ApJ, 885, L19

Coughlin M. W., et al., 2020a, Phys. Rev. Research, 2, 022006

Coughlin M. W., et al., 2020b, Nature Communications, 11, 4129

Coughlin M. W., et al., 2020c, MNRAS, 492, 863

Coulter D. A., et al., 2017, Science, 358, 1556

Cowan J. J., Sneden C., Lawler J. E., Aprahamian A., Wiescher M., Langanke K., Martínez-Pinedo G., Thielemann F.-K., 2021, Rev. of Modern Phys., 93, 015002

Crumley P., Caprioli D., Markoff S., Spitkovsky A., 2019, MNRAS, 485, 5105

Curran P. A., Starling R. L. C., van der Horst A. J., Wijers R. A. M. J., de Pasquale M., Page M., 2011, Advances in Space Research, 47, 1362

Cutler C., Flanagan É. E., 1994, Phys. Rev. D, 49, 2658

D'Avanzo P., 2015, J. of HE Astrop., 7, 73

D'Avanzo P., et al., 2012, MNRAS, 425, 506

D'Avanzo P., et al., 2014, MNRAS, 442, 2342

D'Avanzo P., et al., 2018, A&A, 613, L1

D'Elia V., et al., 2018, A&A, 619, A66

Dai Z. G., Wang X. Y., Wu X. F., Zhang B., 2006, Science, 311, 1127

Daigne F., Mochkovitch R., 1998, MNRAS, 296, 275

Daigne F., Mochkovitch R., 2000, A&A, 358, 1157

Dainotti M. G., Cardone V. F., Capozziello S., 2008, MNRAS, 391, L79

Dainotti M. G., Hernandez X., Postnikov S., Nagataki S., O'brien P., Willingale R., Striegel S., 2017, ApJ, 848, 88

De Pasquale M., et al., 2003, ApJ, 592, 1018

Del Pozzo W., Li T. G. F., Messenger C., 2017, Phys. Rev. D, 95, 043502

Della Valle M., et al., 2006, ApJ, 642, L103

Dewdney P., 2015, in , SKA1 SYSTEM BASELINEV2 DESCRIPTION. SKAO, https://www.skatelescope.org/wp-content/uploads/2014/03/SKA-TEL-SKO-0000308_SKA1_System_Baseline_v2_DescriptionRev01-part-1-signed.pdf

Dhawan S., Bulla M., Goobar A., Sagués Carracedo A., Setzer C. N., 2020, ApJ, 888, 67

Di Valentino E., et al., 2021, Class. and Quant. Gravity, 38, 153001

Dobie D., et al., 2018, ApJ, 858, L15

Dobie D., Murphy T., Kaplan D. L., Hotokezaka K., Ataides J. P. B., Mahony E. K., Sadler E. M., 2021, MNRAS, 505, 2647

Doctor Z., 2020, ApJ, 892, L16

Dominik M., Belczynski K., Fryer C., Holz D. E., Berti E., Bulik T., Mandel I., O'Shaughnessy R., 2012, ApJ, 759, 52

Drout M. R., et al., 2017, Science, 358, 1570

Dubus G., Cerutti B., Henri G., 2010, MNRAS, 404, L55

Ducoin J. G., Corre D., Leroy N., Le Floch E., 2020, MNRAS, 492, 4768

Duffell P. C., MacFadyen A. I., 2015, ApJ, 806, 205

Duffell P. C., Quataert E., Kasen D., Klion H., 2018, ApJ, 866, 3

Duque R., Mochkovitch R., Daigne F., 2019, in The New Era of Multi-Messenger Astrophysics (Asterics2019). p. 25 (arXiv:1907.11945), doi:10.22323/1.357.0025

Duque R., Beniamini P., Daigne F., Mochkovitch R., 2020, A&A, 639, A15

Dvorkin I., Vangioni E., Silk J., Petitjean P., Olive K. A., 2016, MNRAS, 458, L104

Dvorkin I., Daigne F., Goriely S., Vangioni E., Silk J., 2021, MNRAS, 506, 4374

Dyson F., 1963, Instellar Communication

Eichler D., Granot J., 2006, ApJ, 641, L5

Eichler D., Livio M., Piran T., Schramm D. N., 1989, Nature, 340, 126

Einstein A., 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), pp 844–847

Einstein A., 1916, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), pp 688–696

Einstein A., 1918, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), pp 154–167

Falcone A. D., et al., 2007, ApJ, 671, 1921

Farrow N., Zhu X.-J., Thrane E., 2019, ApJ, 876, 18

Feeney S. M., Mortlock D. J., Dalmasso N., 2018, MNRAS, 476, 3861

Fernández J. J., Kobayashi S., 2019, MNRAS, 487, 1200

Fernández R., Metzger B. D., 2016, Ann. Rev. of Nuc. and Part. Sciences, 66, 23

Fernández J. J., Kobayashi S., Lamb G. P., 2021, arXiv e-prints, p. arXiv:2101.05138

Finn L. S., Chernoff D. F., 1993, Phys. Rev. D, 47, 2198

Finstad D., De S., Brown D. A., Berger E., Biwer C. M., 2018, ApJ, 860, L2

Fishbach M., et al., 2019, ApJ, 871, L13

Fong W., Berger E., 2013, ApJ, 776, 18

Fong W., et al., 2012, ApJ, 756, 189

Fong W., Berger E., Margutti R., Zauderer B. A., 2015, ApJ, 815, 102

Ford L. A., et al., 1995, ApJ, 439, 307

Foucart F., 2020, Frontiers in Astronomy and Space Sciences, 7, 46

Frail D. A., Kulkarni S. R., Nicastro L., Feroci M., Taylor G. B., 1997, Nature, 389, 261

Freedman W. L., 2017, Nature Astronomy, 1, 0169

Freedman W. L., et al., 2019, ApJ, 882, 34

Fryer C., Kalogera V., 1997, ApJ, 489, 244

Galama T. J., et al., 1998, Nature, 395, 670

Gehrels N., 2004, in Fenimore E., Galassi M., eds, American Institute of Physics Conference Series Vol. 727, Gamma-Ray Bursts: 30 Years of Discovery. pp 637–641 (arXiv:astro-ph/0405233), doi:10.1063/1.1810924

Gehrels N., Cannizzo J. K., Kanner J., Kasliwal M. M., Nissanke S., Singer L. P., 2016, ApJ, 820, 136

Gendre B., et al., 2013, ApJ, 766, 30

Genet F., Daigne F., Mochkovitch R., 2007, MNRAS, 381, 732

Ghirlanda G., Ghisellini G., Celotti A., 2004a, A&A, 422, L55

Ghirlanda G., Ghisellini G., Lazzati D., 2004b, ApJ, 616, 331

Ghirlanda G., Nava L., Ghisellini G., Celotti A., Firmani C., 2009, A&A, 496, 585

Ghirlanda G., Bernardini M., Calderone G., D'Avanzo P., 2015, J. of HE Astrop., 7, 81

Ghirlanda G., et al., 2016, A&A, 594, A84

Ghirlanda G., et al., 2019, Science, 363, 968

Ghisellini G., Ghirlanda G., Nava L., Firmani C., 2007, ApJ, 658, L75

Ghosh A., Del Pozzo W., Ajith P., 2016a, Phys. Rev. D, 94, 104070

Ghosh S., Bloemen S., Nelemans G., Groot P. J., Price L. R., 2016b, A&A, 592, A82

Giannios D., Spruit H. C., 2005, A&A, 430, 1

Gill R., Granot J., 2018, MNRAS, 478, 4128

Gill R., Nathanail A., Rezzolla L., 2019, ApJ, 876, 139

Goldstein A., Preece R. D., Mallozzi R. S., Briggs M. S., Fishman G. J., Kouveliotou C., Paciesas W. S., Burgess J. M., 2013, ApJS, 208, 21

Goldstein A., et al., 2017, ApJ, 848, L14

Gompertz B. P., et al., 2018, ApJ, 860, 62

Gottlieb O., Nakar E., Piran T., Hotokezaka K., 2018, MNRAS, 479, 588

Gottlieb O., Nakar E., Piran T., 2019, MNRAS, 488, 2405

Gottlieb O., Bromberg O., Singh C. B., Nakar E., 2020, MNRAS, 498, 3320

Gottlieb O., Nakar E., Bromberg O., 2021, MNRAS, 500, 3511

Granot J., Königl A., 2003, ApJ, 594, L83

Granot J., Kumar P., 2006, MNRAS, 366, L13

Granot J., Piran T., 2012, MNRAS, 421, 570

Granot J., Ramirez-Ruiz E., 2010, preprint, (arXiv:1012.5101)

Granot J., Ramirez-Ruiz E., Perna R., 2005, ApJ, 630, 1003

Granot J., Königl A., Piran T., 2006, MNRAS, 370, 1946

Gray R., et al., 2020, Phys. Rev. D, 101, 122001

Guessoum N., Zitouni H., Mochkovitch R., 2018, A&A, 620, A131

Guetta D., Piran T., 2006, A&A, 453, 823

Guidorzi C., et al., 2017, ApJ, 851, L36

Guillochon J., Parrent J., Kelley L. Z., Margutti R., 2017, ApJ, 835, 64

Guiriec S., et al., 2010, ApJ, 725, 225

Hajela A., et al., 2019, ApJ, 886, L17

Hajela A., et al., 2020, Research Notes of the AAS, 4, 68

Hajela A., et al., 2021, arXiv e-prints, p. arXiv:2104.02070

Hakkila J., Giblin T. W., Norris J. P., Fragile P. C., Bonnell J. T., 2008, ApJ, 677, L81

Hakkila J., Horváth I., Hofesmann E., Lesage S., 2018, ApJ, 855, 101

Hallinan G., et al., 2017, Science, 358, 1579

Hardee P., 2008, Journal of Physics: Conference Series, 131, 012052

Hascoët R., Daigne F., Mochkovitch R., Vennin V., 2012, MNRAS, 421, 525

Hascoët R., Daigne F., Mochkovitch R., 2014, MNRAS, 442, 20

Hascoët R., Beloborodov A. M., Daigne F., Mochkovitch R., 2017, MNRAS, 472, L94

He J., et al., 2020, Astrophysics and Space Science, 365

Heinzel J., et al., 2021, MNRAS, 502, 3057

Henry Q., Faye G., Blanchet L., 2020, Phys. Rev. D, 102, 044033

Hjorth J., et al., 2003, Nature, 423, 847

Hobbs G., Lorimer D. R., Lyne A. G., Kramer M., 2005, MNRAS, 360, 974

Hotokezaka K., Piran T., Paul M., 2015, Nature Physics, 11, 1042

Hotokezaka K., Beniamini P., Piran T., 2018a, Intern. J. of Modern Phys. D, 27, 1842005

Hotokezaka K., Kiuchi K., Shibata M., Nakar E., Piran T., 2018b, ApJ, 867, 95

Hotokezaka K., Nakar E., Gottlieb O., Nissanke S., Masuda K., Hallinan G., Mooley K. P., Deller A. T., 2019, Nature Astronomy, 3, 940

Hsu S. C., Bellan P. M., 2002, MNRAS, 334, 257

Huang C. D., et al., 2020, ApJ, 889, 5

Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168

Hulse R. A., Taylor J. H., 1975, ApJ, 195, L51

Ioka K., Kobayashi S., Zhang B., 2005, ApJ, 631, 429

Ioka K., Toma K., Yamazaki R., Nakamura T., 2006, A&A, 458, 7

Ishizaki W., Ioka K., Kiuchi K., 2021, ApJ, 916, L13

Ivanova N., Belczynski K., Kalogera V., Rasio F. A., Taam R. E., 2003, ApJ, 592, 475

Ivezic Z., et al., 2008, Serbian Astronomical Journal, 176, 1

Iyer B., 2015, in APS April Meeting Abstracts. p. J13.001

Japelj J., Vergani S. D., Salvaterra R., Renzo M., Zapartas E., de Mink S. E., Kaper L., Zibetti S., 2018, A&A, 617, A105

Jin Z.-P., et al., 2016, Nature Communications, 7, 12898

Kagra Collaboration et al., 2019, Nature Astronomy, 3, 35

Kalogera V., 1996, ApJ, 471, 352

Karki S., et al., 2016, Review of Scientific Instruments, 87, 114503

Kashyap R., Raman G., Ajith P., 2019, ApJ, 886, L19

Kasliwal M. M., et al., 2017, Science, 358, 1559

Kasliwal M. M., et al., 2019, GRB Coordinates Network, 24191, 1

Kasliwal M. M., et al., 2020, ApJ, 905, 145

Kathirgamaraju A., Giannios D., Beniamini P., 2019, MNRAS, 487, 3914

Kawaguchi K., Shibata M., Tanaka M., 2020, ApJ, 889, 171

King A., O'Brien P. T., Goad M. R., Osborne J., Olsson E., Page K., 2005, ApJ, 630, L113

Klebesadel R. W., Strong I. B., Olson R. A., 1973, ApJ, 182, L85

Klingler N. J., et al., 2019, ApJS, 245, 15

Kobayashi S., 2000, ApJ, 545, 807

Kobayashi S., Zhang B., 2007, ApJ, 655, 973

Kobayashi S., Piran T., Sari R., 1997, ApJ, 490, 92

Kobayashi S., Zhang B., Mészáros P., Burrows D., 2007, ApJ, 655, 391

Komiya Y., Shigeyama T., 2016, ApJ, 830, 76

Köppel S., Bovard L., Rezzolla L., 2019, ApJ, 872, L16

Kouveliotou C., Meegan C. A., Fishman G. J., Bhat N. P., Briggs M. S., Koshut T. M., Paciesas W. S., Pendleton G. N., 1993, ApJ, 413, L101

Kulkarni S. R., Perley D. A., Miller A. A., 2018, ApJ, 860, 22

Kumar P., Zhang B., 2015, Phys. Rep., 561, 1

LIGO Scientific Collaboration 2018, LIGO Algorithm Library - LALSuite, free software (GPL), doi:10.7935/GT1W-FZ16

Laing R. A., Bridle A. H., Parma P., Feretti L., Giovannini G., Murgia M., Perley R. A., 2008, MNRAS, 386, 657

Lamb G. P., Kobayashi S., 2017, MNRAS, 472, 4953

Lamb G. P., Mandel I., Resmi L., 2018, MNRAS, 481, 2581

Lamb G. P., Kann D. A., Fernández J. J., Mandel I., Levan A. J., Tanvir N. R., 2021, MNRAS, 506, 4163

Lamberts A., Daigne F., 2018, MNRAS, 474, 2813

Lazzati D., Perna R., Morsony B. J., Lopez-Camara D., Cantiello M., Ciolfi R., Giacomazzo B., Workman J. C., 2018, Phys. Rev. Lett., 120, 241103

Le Floc'h E., et al., 2003, A&A, 400, 499

Leaver E. W., 1985, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 402, 285

Lemaître G., 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Leventis K., Wijers R. A. M. J., van der Horst A. J., 2014, MNRAS, 437, 2448

Levinson A., Eichler D., 1993, ApJ, 418, 386

Li L., et al., 2012, ApJ, 758, 27

Liang E. W., et al., 2006a, ApJ, 646, 351

Liang E. W., et al., 2006b, ApJ, 646, 351

Liang E., Zhang B., Virgili F., Dai Z. G., 2007, ApJ, 662, 1111

Liang E.-W., et al., 2013, ApJ, 774, 13

Lien A., et al., 2016, ApJ, 829, 7

Lithwick Y., Sari R., 2001, ApJ, 555, 540

Liu B., Lai D., 2018, ApJ, 863, 68

Lloyd-Ronning N. M., Zhang B., 2004, ApJ, 613, 477

Lyman J. D., et al., 2017, MNRAS, 467, 1795

Lyman J. D., et al., 2018, Nature Astronomy, 2, 751

MAGIC Collaboration et al., 2019, Nature, 575, 455

MacLachlan G. A., et al., 2013, MNRAS, 432, 857

Maggiore M., 2007, Gravitational Waves: Volume 1: Theory and Experiments. Oxford University Press, doi:10.1093/acprof:oso/9780198570745.001.0001

Maggiore M., et al., 2020, J. Cosmology Astropart. Phys., 2020, 050

Mandel I., Farr W. M., Gair J. R., 2019, MNRAS, 486, 1086

Mangano V., et al., 2006, Nuovo Cimento B Serie, 121, 1297

Margutti R., Guidorzi C., Chincarini G., Bernardini M. G., Genet F., Mao J., Pasotti F., 2010, MNRAS, 406, 2149

Margutti R., Bernardini G., Barniol Duran R., Guidorzi C., Shen R. F., Chincarini G., 2011, MNRAS, 410, 1064

Margutti R., et al., 2013a, MNRAS, 428, 729

Margutti R., et al., 2013b, ApJ, 778, 18

Margutti R., et al., 2017, ApJ, 848, L20

Margutti R., et al., 2018, ApJ, 856, L18

Mastrogiovanni S., Steer D. A., Barsuglia M., 2020, Phys. Rev. D, 102, 044009

Mastrogiovanni S., Duque R., Chassande-Mottin E., Daigne F., Mochkovitch R., 2021, A&A, 652, A1

Matsumoto T., Nakar E., Piran T., 2019a, MNRAS, 483, 1247

Matsumoto T., Nakar E., Piran T., 2019b, MNRAS, 486, 1563

Matsumoto T., Kimura S. S., Murase K., Mészáros P., 2020, MNRAS, 493, 783

Matteucci F., Romano D., Arcones A., Korobkin O., Rosswog S., 2014, MNRAS, 438, 2177

Mazzali P. A., Sauer D. N., Pian E., Deng J., Prentice S., Ben Ami S., Taubenberger S., Nomoto K., 2017, MNRAS, 469, 2498

Messenger C., Read J., 2012, Phys. Rev. Lett., 108, 091101

Mészáros P., Rees M. J., Wijers R. A. M. J., 1998, ApJ, 499, 301

Metzger B. D., 2019, Liv. Rev. in Rel., 23, 1

Metzger M. R., Djorgovski S. G., Kulkarni S. R., Steidel C. C., Adelberger K. L., Frail D. A., Costa E., Frontera F., 1997, Nature, 387, 878

Metzger B. D., Quataert E., Thompson T. A., 2008, MNRAS, 385, 1455

Mochkovitch R., Hernanz M., Isern J., Martin X., 1993, Nature, 361, 236

Mochkovitch R., Daigne F., Duque R., Zitouni H., 2021, A&A, 651, A83

Modjaz M., Liu Y. Q., Bianco F. B., Graur O., 2016, ApJ, 832, 108

Molina E., del Palacio S., Bosch-Ramon V., 2019, A&A, 629, A129

Monceau-Baroux R., Porth O., Meliani Z., Keppens R., 2014, A&A, 561, A30

Montanet L., et al., 1994, Phys. Rev. D, 50, 1173

Mooley K. P., et al., 2018a, Nature, 554, 207

Mooley K. P., et al., 2018b, Nature, 561, 355

Mooley K. P., et al., 2018c, ApJ, 868, L11

Morlino G., Caprioli D., 2012, A&A, 538, A81

Mortlock D. J., Feeney S. M., Peiris H. V., Williamson A. R., Nissanke S. M., 2019a, Phys. Rev. D, 100, 103523

Mortlock D. J., Feeney S. M., Peiris H. V., Williamson A. R., Nissanke S. M., 2019b, Phys. Rev. D, 100, 103523

Mueller A., 2020, WordCloud for Python, 1.8.1

Mukherjee S., Lavaux G., Bouchet F. R., Jasche J., Wandelt B. D., Nissanke S., Leclercq F., Hotokezaka K., 2021, A&A, 646, A65

Nakar E., 2007, Phys. Rep., 442, 166

Nakar E., Piran T., 2002, MNRAS, 330, 920

Nakar E., Piran T., 2021, ApJ, 909, 114

Nakar E., Sari R., 2012, ApJ, 747, 88

Nakar E., Piran T., Granot J., 2002, ApJ, 579, 699

Nakar E., Gottlieb O., Piran T., Kasliwal M. M., Hallinan G., 2018, ApJ, 867, 18

Nathanail A., Gill R., Porth O., Fromm C. M., Rezzolla L., 2020, MNRAS, 495, 3780

Nathanail A., Gill R., Porth O., Fromm C. M., Rezzolla L., 2021, MNRAS, 502, 1843

Nava L., et al., 2014, MNRAS, 443, 3578

Nedora V., Radice D., Bernuzzi S., Perego A., Daszuta B., Endrizzi A., Prakash A., Schianchi F., 2021, MNRAS,

Nemiroff R. J., 1994, Comments on Astrophysics, 17, 189

Nicholl M., Margalit B., Schmidt P., Smith G. P., Ridley E. J., Nuttall J., 2021, MNRAS, 505, 3016

Nicuesa Guelbenzu A., et al., 2011, A&A, 531, L6

Nissanke S., Holz D. E., Dalal N., Hughes S. A., Sievers J. L., Hirata C. M., 2013, arXiv e-prints, p. arXiv:1307.2638

Norris J. P., Bonnell J. T., 2006, ApJ, 643, 266

Norris J. P., Nemiroff R. J., Bonnell J. T., Scargle J. D., Kouveliotou C., Paciesas W. S., Meegan C. A., Fishman G. J., 1996, ApJ, 459, 393

Norris J. P., Bonnell J. T., Kazanas D., Scargle J. D., Hakkila J., Giblin T. W., 2005, ApJ, 627, 324

Nousek J. A., et al., 2006, ApJ, 642, 389

Nysewander M., Fruchter A. S., Pe'er A., 2009, ApJ, 701, 824

O'Connor B., Beniamini P., Kouveliotou C., 2020, MNRAS, 495, 4782

Oganesyan G., Ascenzi S., Branchesi M., Sharan Salafia O., Dall'Osso S., Ghirlanda G., 2020, ApJ, 893, 88

Paciesas W. S., et al., 1999, ApJS, 122, 465

Paczynski B., 1986, ApJ, 308, L43

Paczyński B., 1998, ApJ, 494, L45

Palmerio J. T., Daigne F., 2021, A&A, 649, A166

Palmerio J. T., et al., 2019, A&A, 623, A26

Panaitescu A., Kumar P., 2000, ApJ, 543, 66

Panaitescu A., Mészáros P., Burrows D., Nousek J., Gehrels N., O'Brien P., Willingale R., 2006, MNRAS, 369, 2059

Pang P. T. H., Hannuksela O. A., Dietrich T., Pagano G., Harry I. W., 2020, MNRAS, 495, 3740

Pavan A., Ciolfi R., Vijay Kalinani J., Mignone A., 2021, MNRAS, 506, 3483

Peacock J., 1999, Cosmological Physics. Cambridge Astrophysics, Cambridge University Press, https://books.google.fr/books?id=t8O-yylU0j0C

Peng Z. Y., Ma L., Zhao X. H., Yin Y., Fang L. M., Bao Y. Y., 2009, ApJ, 698, 417

Perlmutter S., 2003, Physics Today, 56, 53

Perna R., Belczynski K., 2002, ApJ, 570, 252

Perna R., Armitage P. J., Zhang B., 2006, ApJ, 636, L29

Pihlström Y., Taylor G., Granot J., Doeleman a., 2008, ApJ, 664, 411

Piran T., 2004, Rev. of Modern Phys., 76, 1143

Piron F., 2016, Comptes Rendus Physique, 17, 617

Planck Collaboration et al., 2020a, A&A, 641, A1

Planck Collaboration et al., 2020b, A&A, 641, A6

Planck Collaboration et al., 2020c, A&A, 641, A6

Podsiadlowski P., Langer N., Poelarends A. J. T., Rappaport S., Heger A., Pfahl E., 2004, ApJ, 612, 1044

Podsiadlowski P., Pfahl E., Rappaport S., 2005, Neutron-Star Birth Kicks. Rasio, Fred A. and Stairs, Ingrid H., p. 327

Poincaré H., 1905, Compte rendu de l'Académie des sciences

Poisson E., Will C. M., 1995, Phys. Rev. D, 52, 848

Poolakkil S., et al., 2021, ApJ, 913, 60

Portegies Zwart S. F., Yungelson L. R., 1998, A&A, 332, 173

Prodi G., 2018, in 6th International Conference on New Frontiers in Physics.

Punturo M., Lück H., Beker M., 2014, in Bassan M., ed., Astrophysics and Space Science Library Vol. 404, Advanced Interferometers and the Search for Gravitational Waves. p. 333, doi:10.1007/978-3-319-03792-9_13

Ramirez-Ruiz E., Granot J., Kouveliotou C., Woosley S. E., Patel S. K., Mazzali P. A., 2005, ApJ, 625, L91

Rau A., Kienlin A. V., Hurley K., Lichti G. G., 2005, A&A, 438, 1175

Rees M. J., 1966, Nature, 211, 468

Rees M. J., Meszaros P., 1994, ApJ, 430, L93

Reitze D., et al., 2019, in BAAS. p. 35

Resmi L., et al., 2018, ApJ, 867, 57

Rhoads J. E., 1999, ApJ, 525, 737

Riess A. G., et al., 1998, AJ, 116, 1009

Riess A. G., et al., 2016, ApJ, 826, 56

Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019a, ApJ, 876, 85

Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019b, ApJ, 876, 85

Riess A. G., Yuan W., Casertano S., Macri L. M., Scolnic D., 2020, ApJ, 896, L43

Rigault M., et al., 2013, A&A, 560, A66

Rybicki G. B., Lightman A. P., 1979, Radiative processes in astrophysics. New York, Wiley-Interscience, 1979. 393 p.

Safarzadeh M., Côté B., 2017, MNRAS, 471, 4488

Salafia O. S., Ghisellini G., Pescalli A., Ghirlanda G., Nappo F., 2015, MNRAS, 450, 3549

Saleem M., 2020, MNRAS, 493, 1633

Saleem M., Resmi L., Misra K., Pai A., Arun K. G., 2018a, MNRAS, 474, 5340

Saleem M., Pai A., Misra K., Resmi L., Arun K. G., 2018b, MNRAS, 475, 699

Santana R., Barniol Duran R., Kumar P., 2014, ApJ, 785, 29

Savaglio S., Glazebrook K., Le Borgne D., 2009, ApJ, 691, 182

Savchenko V., et al., 2017, ApJ, 848, L15

Schulze S., et al., 2014, A&A, 566, A102

Schutz B. F., 1986, Nature, 323, 310

Secunda A., Bellovary J., Mac Low M.-M., Ford K. E. S., McKernan B., Leigh N. W. C., Lyra W., Sándor Z., 2019, ApJ, 878, 85

Selina R. J., et al., 2018, in Ground-based and Airborne Telescopes VII. p. 107001O (arXiv:1806.08405), doi:10.1117/12.2312089

Setzer C. N., Biswas R., Peiris H. V., Rosswog S., Korobkin O., Wollaeger R. T., LSST Dark Energy Science Collaboration 2019, MNRAS, 485, 4260

Shen R., Matzner C. D., 2012, ApJ, 744, 36

Simonetti P., Matteucci F., Greggio L., Cescutti G., 2019, MNRAS, 486, 2896

Singer L. P., Price L. R., 2016, Phys. Rev. D, 93, 024013

Sironi L., Spitkovsky A., 2011, ApJ, 726, 75

Sironi L., Spitkovsky A., Arons J., 2013, ApJ, 771, 54

Sironi L., Keshet U., Lemoine M., 2015, Space Sci. Rev., 191, 519

Smartt S. J., et al., 2017, Nature, 551, 75

Smith K. W., et al., 2019, GRB Coordinates Network, 24210, 1

Soares-Santos M., et al., 2019, ApJ, 876, L7

Soderberg A. M., et al., 2006, Nature, 442, 1014

Tang C.-H., Huang Y.-F., Geng J.-J., Zhang Z.-B., 2019, ApJS, 245, 1

Tanvir N. R., Levan A. J., Fruchter A. S., Hjorth J., Hounsell R. A., Wiersema K., Tunnicliffe R. L., 2013, Nature, 500, 547

Tauris T. M., et al., 2017, ApJ, 846, 170

Toma K., Ioka K., Yamazaki R., Nakamura T., 2006, ApJ, 640, L139

Troja E., et al., 2007, ApJ, 665, 599

Troja E., Rosswog S., Gehrels N., 2010, ApJ, 723, 1711

Troja E., et al., 2017a, Nature, 547, 425

Troja E., et al., 2017b, Nature, 551, 71

Troja E., et al., 2019a, MNRAS, p. 2169

Troja E., et al., 2019b, MNRAS, 489, 1919

Troja E., et al., 2021, arXiv e-prints, p. arXiv:2104.13378

Tsujimoto T., Shigeyama T., 2014, A&A, 565, L5

Uhm Z. L., Beloborodov A. M., 2007, ApJ, 665, L93

Ukwatta T. N., et al., 2010, ApJ, 711, 1073

Vangioni E., Goriely S., Daigne F., François P., Belczynski K., 2016, MNRAS, 455, 17

Veitch J., et al., 2012, Phys. Rev. D, 85, 104045

Villar V. A., et al., 2017, ApJ, 851, L21

Virgo Collaboration 2009, The Gravitational Voice, Special Anniversary Edition

Wallner A., et al., 2015, Nature Communications, 6, 5956

Wanderman D., Piran T., 2015, MNRAS, 448, 3026

Watson D., et al., 2019, Nature, 574, 497

Waxman E., 2004, ApJ, 602, 886

Wehmeyer B., Pignatari M., Thielemann F.-K., 2015, MNRAS, 452, 1970

Wei J., et al., 2016, arXiv e-prints, p. arXiv:1610.06892

Weisberg J. M., Huang Y., 2016, ApJ, 829, 55

Wijers R. A. M. J., Galama T. J., 1999, ApJ, 523, 177

Wollaeger R. T., et al., 2018, MNRAS, 478, 3298

Woosley S. E., 1993, ApJ, 405, 273

Woosley S. E., Bloom J. S., 2006, ARA&A, 44, 507

Xie X., Zrake J., MacFadyen A., 2018, ApJ, 863, 58

Yi S.-X., Xi S.-Q., Yu H., Wang F. Y., Mu H.-J., Lü L.-Z., Liang E.-W., 2016, ApJS, 224, 20

Yonetoku D., Murakami T., Nakamura T., Yamazaki R., Inoue A. K., Ioka K., 2004, ApJ, 609, 935

Zappa F., Bernuzzi S., Radice D., Perego A., Dietrich T., 2018, Phys. Rev. Lett., 120, 111101

Zehavi I., Riess A. G., Kirshner R. P., Dekel A., 1998, ApJ, 503, 483

Zhang B., Fan Y. Z., Dyks J., Kobayashi S., Mészáros P., Burrows D. N., Nousek J. A., Gehrels N., 2006, ApJ, 642, 354

Zhang B., et al., 2007, ApJ, 655, 989

Zheng Z., Ramirez-Ruiz E., 2007, ApJ, 665, 1220

Zyla P., et al., 2020, PTEP, 2020, 083C01

van Paradijs J., et al., 1997, Nature, 386, 686

von Kienlin A., et al., 2019, ApJ, 876, 89

von Kienlin A., et al., 2020, ApJ, 893, 46

List of Figures

List of Tables

Appendices

Appendix A

Simplified inspiral gravitational-wave signal detection criterion

A.1 General criterion

The detection of an inspiral signal by an orthogonal gravitational interferometer depends on the luminosity distance D_L, orientation (ι, ξ), and sky-position (θ, ϕ) of the source: ι is the inclination angle of the line of sight to the binary's angular momentum, ξ is the angle between the projections of the angular momentum and of one of the instrument's arm on the tangent plane, and (θ, ϕ) are given in a spherical coordinate system with $\theta = 0$ corresponding to the zenith of the interferometer. This *wave coordinates system* is described in more detail in Fig. 2.5. The signal-to-noise ratio (SNR) of an inspiral signal is given by Finn & Chernoff (1993), Eq. 3.16:

$$\rho^2 = A^2(\iota, \xi, \theta, \phi)\mathcal{M}^{5/3} S_I \tag{A.1}$$

where $\mathcal{M}$ is the chirp mass of the binary, S_I is a quantity depending only the sensitivity profile of the interferometer, and A is the amplitude of the signal, which is expressed as:

$$A^2 = \frac{\Theta^2(\iota, \xi, \theta, \phi)}{D_L^2} \tag{A.2}$$

where the dependence on the angles is collected in:

$$\Theta^2(\iota, \xi, \theta, \phi) = 4\left[F_+^2(\xi, \theta, \phi)(1 + \cos^2 \iota)^2 + 4F_\times^2(\xi, \theta, \phi)\cos^2 \iota\right] \tag{A.3}$$

where we see that the instrument's antenna patterns F_+ and $F_\times$ appear. The antenna patterns are given explicitly in Eq. 2.38.

Θ^2 admits a global maximum of $\Theta_M^2 = 16$ corresponding to an optimally oriented ($\iota = 0$) and positioned ($\theta = 0$) binary. The fiducial detection threshold for the LVK network is an SNR of $\rho_0 = 8$. For this optimal binary, the maximum distance to which it can be detected is known as the *horizon* of the interferometer H, and is a common figure-of-merit along with the *range*. The range is the average distance to which an inspiral signal can be detected, on average on the orientation and sky positions; The range R is linked to the horizon H by $H = 2.26R$.

By definition of the horizon, an optimally oriented system at the horizon will produce the threshold SNR, therefore:

$$\rho_0^2 = \Theta_M^2 \mathcal{M}^{5/3} S_I / H^2 \tag{A.4}$$

and thus $H = \sqrt{\frac{16 S_I \mathcal{M}^{5/3}}{\rho_0^2}}$. The horizon was respectively 218 Mpc, 107 Mpc and 58 Mpc for LIGO-Livingston, LIGO-Hanford, and Virgo at the moment of GW170817 (Abbott et al., 2017c) for a standard 1.4-1.4 $M_\odot$ neutron star binary. For an arbitrary binary inspiral signal, we may rewrite the criterion as:

$$\rho^2 > \rho_0^2 \Leftrightarrow \Theta^2(i, \xi, \theta, \phi) > 16 \left(\frac{D_L}{H}\right)^2 \tag{A.5}$$

A.2 Sky-position-averaged criterion

In a full population study, one would have to draw from distributions on all four angles i, ξ, θ, ϕ and evaluate this criterion in every case. We choose to reduce the number of parameters to two: D_L and i, and must thus review this criterion by averaging

the value of Θ^2 on sky-position (θ, ϕ) and orientation angle ξ. This is readily done from Eq. A.4 and it is found analytically that:

$$\langle \Theta^2 \rangle_{\xi,\theta,\phi} = \frac{1}{4\pi 2\pi} \int d\xi d\Omega \Theta^2 \tag{A.6}$$

$$= \frac{4}{5} \left[1 + 6\cos^2 i + \cos^4 i \right] \tag{A.7}$$

Hence, the detection criterion *on average* for a binary of inclination i and distance D_L is:

$$\langle \Theta^2 \rangle_{\xi,\theta,\phi} > 16 \left(\frac{D_L}{H} \right)^2 \tag{A.8}$$

which is:

$$\sqrt{\frac{1 + 6\cos^2 i + \cos^4 i}{8}} > D_L / \overline{H} \tag{A.9}$$

where we have denoted $\overline{H} = \sqrt{\frac{2}{5}} H \sim H/1.58$ the *sky-position-averaged* horizon. The left hand side of Eq. A.9 is in the interval $\left[1/\sqrt{8}, 1\right]$, signifying that $\overline{H}$ is the maximum distance to which an optimally oriented binary inspiral signal can be detected on average in the sky.

The criterion Eq. A.9 is valid for the detection using a single instrument. In fact, GW detection by the interferometer network is based on two-instrument coherent analysis, and is thus more complicated than that described by our criterion. Furthermore as was illustrated in the case of 170817, true joint detections require appropriate localization of the source, and our criterion should incorporate this. Localization of sources by the IFO network is complex, and there is a rich variety of possible localization maps. Even the simplest of predictions of the localization zone extent in the sky, based on timing accuracy of the IFOs involves the relative positions the IFOs in the network, which we do not wish to incorporate in our study. We thus simplify the detection-localization criterion to that of Eq. A.9, but taking for H the quadratic mean of the horizons of the two most sensitive IFOs of the network, namely LIGO-Livingston and LIGO-Hanford. In Tab. 4.1, one can find the values for $\overline{H}$ assumed in the different instances of the population model used in this work.

While the criterion in Eq. A.9 can seem simplified, it well reproduces the main feature of anisotropy of the detection probability of inspiral signals. In Fig. 4.1, we compare this criterion to a complete calculation of the detection probability for BNS systems for a network of interferometers. The similarity is striking, justifying the use of Eq. A.9.

A.3 Analytic results and handy distributions

According to Eq. A.9, we note the existence of a characteristic distance $D_0 = \overline{H}/\sqrt{8}$; It is such that all sources with $D_L < D_0$ are detected, regardless of the inclination angle. For $D_L > D_0$, the maximum inclination angle to which the signal can be detected is given by:

$$\cos\theta_{\mathrm{max,GW}} = \sqrt{-3 + \sqrt{8 + 8\frac{D_L^2}{\overline{H}^2}}}. \tag{A.10}$$

$\cos\theta_{\mathrm{max,GW}}$ is figured in Fig. A.1 (left) as a function of D_L.

Conversely, the maximum distance to which a signal with inclination angle θ_v can be detected is simply:

$$D_{\mathrm{max,GW}} = \overline{H} \sqrt{\frac{1 + 6\cos^2\theta_v + \cos^4\theta_v}{8}} \tag{A.11}$$

Fig. A.1 (right) represents the distribution in distance of GW-detected events, normalized to $\overline{H}$, assuming they are intrinsically distributed homogeneously in space.

The analytical form of this distribution is:

$$\frac{\mathrm{d}N}{\mathrm{d}D_L} = D_L^2 \times \int_{\theta=0}^{\theta_{\mathrm{max,GW}}} \mathrm{d}\theta \sin\theta \tag{A.12}$$

$$= D_L^2 \times \begin{cases} 1 & \text{if} \quad D_L < D_0 \\ 1 - \sqrt{-3 + \sqrt{8 + 8\frac{D_L^2}{\overline{H}^2}}} & \text{if} \quad D_0 < D_L \end{cases} \tag{A.13}$$

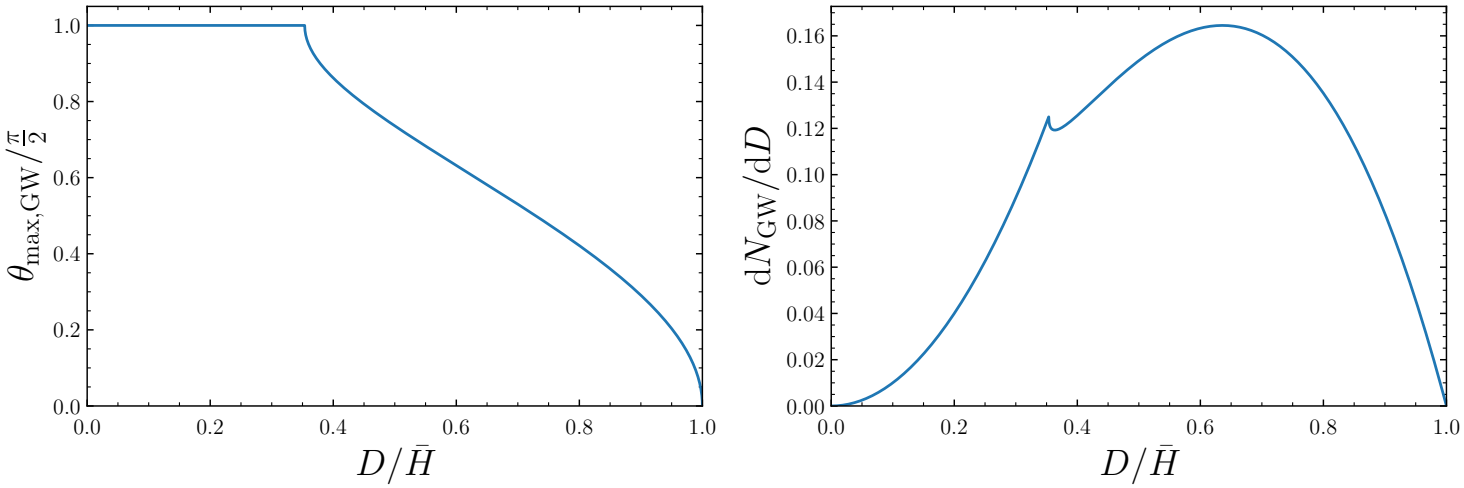


Figure A.1: **Distribution of distances to the GW-detectable events according to our simple detection criterion.** Left: $D_{\rm max,GW}$, maximum angle to which a GW event can be detected, as a function of the distance. Right: Distribution of distances of GW-detected events, normalized to the horizon.

Close to $\overline{H}$, we have $\frac{\mathrm{d}N}{\mathrm{d}D_L} \sim \overline{H}(\overline{H} - D_L)$ and $\theta_{\rm max,GW} \sim \frac{\pi}{2}\sqrt{2}\sqrt{1 - D_L/\overline{H}}$.

The mean distance to the detectable events is approximately $0.58 \times \overline{H}$, independently of the value of the horizon.

The mean angle of a GW-detected is independent of the horizon it is $\langle\theta_v\rangle_{\rm GW} \sim 38$ deg.

Assuming a uniform density of sources, the fraction of events within the horizon detected in GW is independent of the horizon. It is:

$$\frac{N_{\rm GW}}{N_{\rm pop}} = \frac{2}{3}\int_0^{\pi/2} d\theta \sin\theta \left(1 + 6\cos^2\theta + \cos^4\theta\right)^{2/3} \tag{A.14}$$

$$\sim 29\% \tag{A.15}$$

which is thus an absolute maximum on the fraction of events with any counterpart detected among all mergers.

Appendix B

Luminosity from a distant, optically thin, fast-varying relativistic source

Suppose there is a field of relativistic emitters, defined by a spectral emission coefficient $j(\nu, \overrightarrow{x_e}, t_\mathrm{e}, \overrightarrow{u})$ (Rybicki & Lightman, 1979, Chap. 1). This is defined such that the energy emitted by a volume $\mathrm{d}V$ about $\overrightarrow{x_\mathrm{e}}$ during a time interval $\mathrm{d}t$ about t_e into a solid angle $\mathrm{d}\Omega_\mathrm{em}$ about $\overrightarrow{u}$ in the frequency band $\mathrm{d}\nu$ about ν is $\mathrm{d}E = j(\nu, \overrightarrow{x_e}, t_\mathrm{e}, \overrightarrow{u})\mathrm{d}V\mathrm{d}t\mathrm{d}\Omega_\mathrm{em}\mathrm{d}\nu$. Units of j are: erg/s/cm^3/Hz/sr.

By definition, an observer lying at $\overrightarrow{x_\mathrm{obs}}$ will observe at t_obs and frequency ν_obs a spectral flux on a surface with normal $\overrightarrow{n}$ given by:

$$F(\overrightarrow{x_\mathrm{obs}}, t_\mathrm{obs}, \nu, \overrightarrow{n}) = \int \mathrm{d}^2\overrightarrow{n_\theta} I(x_\mathrm{obs}, t_\mathrm{obs}, \nu, \overrightarrow{n_\theta})\,(\overrightarrow{n_\theta} \cdot \overrightarrow{n}) \tag{B.1}$$

where I denotes the specific intensity in direction $\overrightarrow{n_\theta}$, and we integrate $\overrightarrow{n_\theta}$ on all unitary directions crossing the surface.

In turn, for an optically thin field of emitters, the specific intensity in direction $\overrightarrow{n_\theta}$ at observer position-time $\overrightarrow{x_\mathrm{obs}}, t_\mathrm{obs}$ carries contributions from all emitters on the corresponding line of sight $\mathrm{LOS}(\overrightarrow{n_\theta})$, accounting for light travel delay. We denote by $t_{ar}(\overrightarrow{x_\mathrm{obs}}, \overrightarrow{x_\mathrm{e}}, t_\mathrm{e})$ the arrival time at position $\overrightarrow{x_\mathrm{obs}}$ of light emitted at position-time $\overrightarrow{x_\mathrm{e}}, t_\mathrm{e}$.

Therefore, we have:

$$I(\overrightarrow{x_\mathrm{obs}}, t_\mathrm{obs}, \nu, \overrightarrow{n_\theta}) = \int \mathrm{d}t_\mathrm{e} \int_{\mathrm{LOS}(\overrightarrow{n_\theta})} \|\mathrm{d}\overrightarrow{x_\mathrm{e}}\|\, j(\nu, \overrightarrow{x_\mathrm{e}}, t_\mathrm{e}, \overrightarrow{n_\theta}) \delta\left(t_{ar}(\overrightarrow{x_\mathrm{obs}}, \overrightarrow{x_\mathrm{e}}, t_\mathrm{e}) - t_\mathrm{obs}\right) \tag{B.2}$$

Finally, the flux writes:

$$F(\overrightarrow{x_\mathrm{obs}}, t_\mathrm{obs}, \nu, \overrightarrow{n}) = \int \mathrm{d}^2\overrightarrow{n_\theta} \int \mathrm{d}t_\mathrm{e} \int_{\mathrm{LOS}(\overrightarrow{n_\theta})} \|\mathrm{d}\overrightarrow{x_\mathrm{e}}\|\, j(\nu, \overrightarrow{x_\mathrm{e}}, t_\mathrm{e}, \overrightarrow{n_\theta})\,(\overrightarrow{n_\theta} \cdot \overrightarrow{n})\,\delta\left(t_{ar}(\overrightarrow{x_\mathrm{obs}}, \overrightarrow{x_\mathrm{e}}, t_\mathrm{e}) - t_\mathrm{obs}\right) \tag{B.3}$$

This is best expressed in spherical coordinates centered on the observer. We denote (r, θ, ϕ) such coordinates as usual, with the spherical axis perpendicular to the detection surface, as in Fig. B.1. We suppose the emitters are entirely contained on one side of this surface.

Then the spectral flux is simply:

$$F(t_\mathrm{obs}, \nu_\mathrm{obs}) = \int \mathrm{d}t_e \int \mathrm{d}\Omega \cos\theta \int \mathrm{d}r j(\nu, r, \theta, \phi, t_\mathrm{e}, \theta) \delta(t_{ar}(r, \theta, \phi, t_\mathrm{e}) - t_\mathrm{obs}) \tag{B.4}$$

where we specified the angle θ as both the position coordinate and the emitting direction to consider, as these are identical in this case.

Evaluating the arrival time as $t_{ar}(r, \theta, \phi, t_\mathrm{e}) = t_\mathrm{e} + r/c$ allows further simplification:

$$F(t_\mathrm{obs}, \nu_\mathrm{obs}) = \int \mathrm{d}\Omega \cos\theta \int \mathrm{d}r j(\nu, r, \theta, \phi, t_\mathrm{obs} - r/c, \theta) \tag{B.5}$$

While Eq. B.5 is attractive with simplicity, these spherical coordinates are not the most adapted to most of our practical cases, where matter outflows radially from a point different from the observer's station.

We now turn to spherical coordinates centered on a point distant from luminosity distance D_L from the observer, and with axis pointed towards the observer, as in Fig. B.1. As we will see below, it is in these (R, α, ψ) that GRB-related emission coefficients are the most naturally expressed.

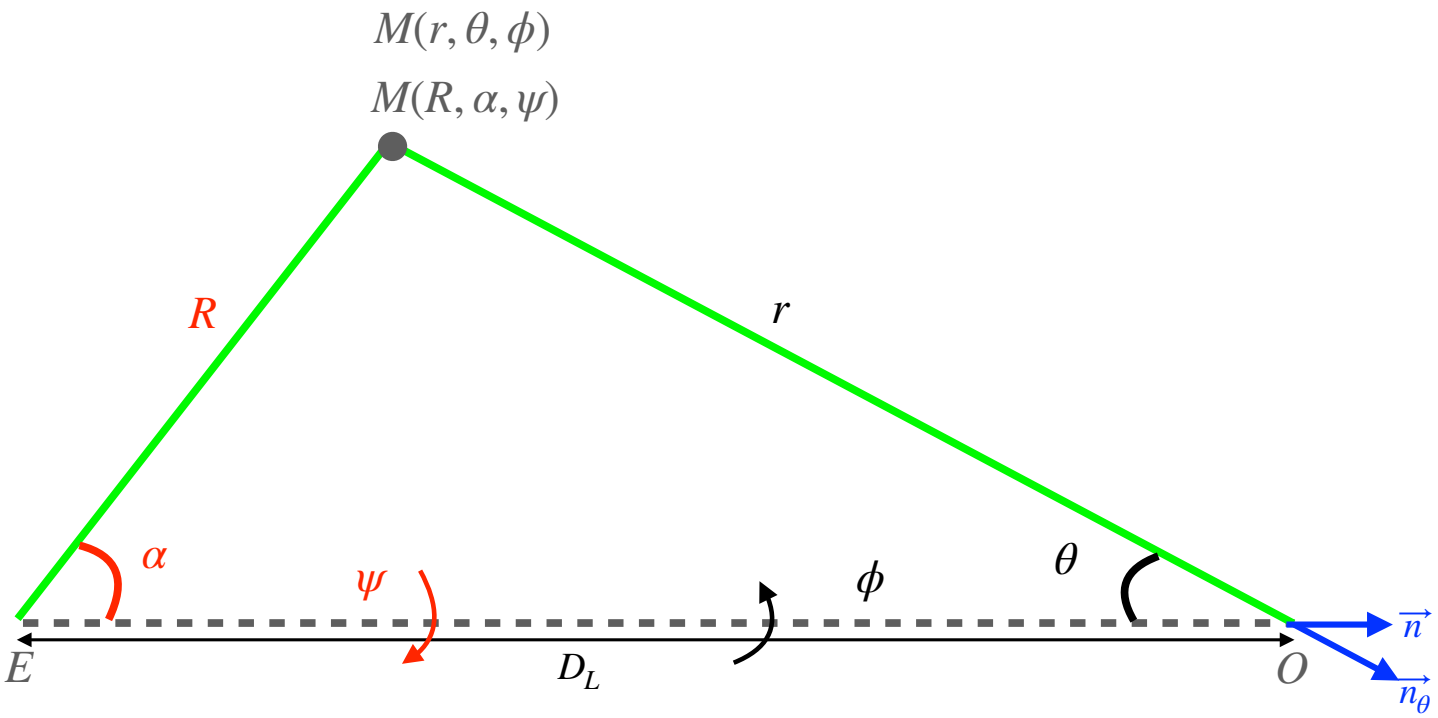


Figure B.1: **Geometrical definition of spherical coordinates,** both observer-centered (black) and central-engine-centered (red).

We seek to transform Eq. B.5 into these coordinates. Simple geometry leads to:

$$\left\{\begin{array}{ccc} \phi & = & \psi \\ D_L & = & R\cos\alpha + r\cos\theta \\ R\sin\alpha & = & r\sin\theta \end{array}\right. \tag{B.6}$$

For our purposes, the sources are very distant with respect to their spatial extension, therefore we will consider R/D_L as a small parameter. This is necessary to exactly invert Eqs. B.6, leading to the following coordinate transformation:

$$\left\{\begin{array}{ccc} r & = & D_L - R\cos\alpha \\ \sin\theta & = & R\sin\alpha/D_L \\ \phi & = & \psi \end{array}\right. \tag{B.7}$$

with Jacobian function:

$$\left|\frac{\partial(r,\theta,\phi)}{\partial(R,\alpha,\psi)}\right| = \left|\begin{array}{ccc} -\cos\alpha & R\sin\alpha & 0 \\ \sin\alpha/D_L & R\cos\alpha/D_L & 0 \\ 0 & 0 & 1 \end{array}\right| = \frac{R}{D_L} \tag{B.8}$$

In these coordinates, the flux is:

$$F(t_{\rm obs},\nu_{\rm obs}) = \int \mathrm{d}\psi \int \mathrm{d}\alpha \int \mathrm{d}R \left(\frac{R}{D_L}\right)^2 \sin\alpha j\left(\nu, R, \alpha, \psi, t_{\rm obs} + \frac{R\cos\alpha}{c}, \frac{R}{D_L}\sin\alpha\right) \tag{B.9}$$

where we dropped a global photon flight time from the source D_L/c, thus defining observer time $t_{\rm obs} = 0$ upon detection of light emitted at $t_e = R = 0$.

Denoting $\mathrm{d}\Omega_E$ the solid angle differential element in these coordinates, this is simply:

$$F(t_{\rm obs},\nu_{\rm obs}) = \frac{1}{D_L^2}\int \mathrm{d}\Omega_E \int \mathrm{d}R R^2 j\left(\nu, R, \alpha, \psi, t_{\rm obs} + \frac{R\cos\alpha}{c}, \frac{R}{D_L}\sin\alpha\right) \tag{B.10}$$

From such a source, and observer would measure a isotropic-equivalent spectral luminosity:

$$L_{\rm iso}(t_{\rm obs},\nu_{\rm obs}) = 4\pi D_L^2 F(t_{\rm obs},\nu_{\rm obs}) = 4\pi \int \mathrm{d}\Omega_E \int \mathrm{d}R R^2 j\left(\nu, R, \alpha, \psi, t_{\rm obs} + \frac{R\cos\alpha}{c}, \frac{R}{D_L}\sin\alpha\right) \tag{B.11}$$

In the case of a relativistic source, it is often more natural to specify the emission properties in the emitter's source frame, where quantities are denoted with primes ''. Let $\overrightarrow{\beta}(\overrightarrow{x_{\rm e}}, t_{\rm e})$ and $\Gamma(\overrightarrow{x_{\rm e}}, t_{\rm e})$, velocity and Lorentz factor of the material at position-time $\overrightarrow{x_{\rm e}}, t_{\rm e}$. We write $\mathcal{D}(\overrightarrow{x_{\rm e}}, t_{\rm e})$ for the corresponding Doppler factor.

Using the relativistic invariance of $\mathcal{D}^2 \times j$ (Rybicki & Lightman, 1979, Chap. 4), the isotropic luminosity is thus:

$$L_{\rm iso}(t_{\rm obs}, \nu_{\rm obs}) = 4\pi \int {\rm d}\Omega_E \int {\rm d}R R^2 \mathcal{D}^2(R, \alpha, \psi) j'(\nu'(\nu_{\rm obs}), (R', \alpha', \psi', t'_{\rm e})(R, \alpha, \psi, t_{\rm obs} + R\cos\alpha/c), \overrightarrow{u}') \tag{B.12}$$

where care must be taken in the determination of the comoving direction $\overrightarrow{u'}$ corresponding to the observer's line of sight. Here, Doppler shifting is simply $\nu'(\nu_{\rm obs}) = \nu_{\rm obs}/\mathcal{D}$.

Appendix C

Luminosity from a flashing relativistic cone

C.1 Luminosity from an arbitrary line of sight

We are interested in the observed luminosity from a flashing portion of a sphere, as in our geometrical interpretation of X-ray flares. As in Appendix B, we use spherical coordinates with origin the center of the emitting sphere (i.e., the position of the central engine, see Fig. B.1).

We consider a conical patch consisting of an infinitely thin spherical shell with axis passing through the point $(\alpha_c, 0)$ and half-opening angle r. It has a purely radial motion at velocity β and Lorentz factor Γ. At observer time T_E and radius R_E, it instantaneously and isotropically emits an energy E'_0 measured in the central source's rest frame rest frame. The light is emitted with source-frame spectrum s', such that $\int s'(\nu')\mathrm{d}\nu' = 1$. We denote by ${\Omega_p}'$ the solid angle occupied by the patch. Given the instantaneous and release of the energy, and the thinness of the shell, we suggest the following ansatz for the source-frame emission coefficient:

$$j'(\nu', R', \alpha', \psi', t'_\mathrm{e}) = A \times \delta(t'_\mathrm{e} - T'_E)\delta(R' - R'_E)\mathbf{1}_{{\Omega_p}'}(\alpha', \psi')s'(\nu') \tag{C.1}$$

for some constant A to determine.

It must be that the total dissipated source-frame energy is:

$$E'_0 = \int \mathrm{d}t'_\mathrm{e} \int \mathrm{d}V' \int \mathrm{d}\nu' \int \mathrm{d}\Omega'_\mathrm{em} j' \tag{C.2}$$

This finally leads to:

$$j'(\nu', R', \alpha', \psi', t'_\mathrm{e}) = \frac{E'_0}{4\pi\Omega'_p R'^2_E}\delta(t'_\mathrm{e} - T'_E)\delta(R' - R'_E)\mathbf{1}_{{\Omega_p}'}(\alpha', \psi')s'(\nu') \tag{C.3}$$

Anticipating on Eq. B.12, we determine $\mathcal{D} \times j'$ and express it in observer coordinates. We have the following relations:

$$\left\{ \begin{array}{rcl} \mathrm{d}t'_\mathrm{e} & = & \mathrm{d}t_\mathrm{e}/\Gamma \\ \Omega'_p R'^2 \mathrm{d}R' & = & \Gamma \times \Omega_p R^2 \mathrm{d}R \\ \mathrm{d}E' & = & \mathrm{d}E/\Gamma \end{array} \right. \tag{C.4}$$

Therefore, we have:

$$\mathcal{D}^2 j' = \mathcal{D}^2 \frac{[E_0/\Gamma]}{4\pi}[\Gamma\delta(t_\mathrm{e} - T_E)] \left[\frac{1}{\Gamma}\frac{1}{\Omega_p R_E^2}\delta(R - R_E)\right] \mathbf{1}_{\Omega_p}(\alpha, \psi) \tag{C.5}$$

$$= \mathcal{D}^2 \frac{E_0}{4\pi\Gamma\Omega_p R_E^2}\delta(t_\mathrm{e} - T_E)\delta(R - R_E)\mathbf{1}_{\Omega_p}(\alpha, \psi) \tag{C.6}$$

where we used $\delta(ax) = \frac{1}{|a|}\delta(x)$ in the Dirac functions.

Because of the radial expansion, we have the Doppler factor:

$$\mathcal{D}(R, \alpha, \psi) = \frac{1}{\Gamma(1 - \beta\cos\alpha)} \tag{C.7}$$

Thus, according to Eq. B.12, the isotropic-equivalent spectral luminosity at observer time t_obs and frequency ν_obs is:

$$L_\mathrm{iso}(t_\mathrm{obs}, \nu_\mathrm{obs}) = 4\pi \int \mathrm{d}R \int \mathrm{d}\Omega R^2 \mathcal{D}^2 \frac{E_0}{4\pi\Gamma\Omega_p R_E^2}\delta(t_\mathrm{obs} + R\cos\alpha/c - T_E)\delta(R - R_E)\mathbf{1}_{\Omega_p}(\alpha, \psi)s'\left(\frac{\nu_\mathrm{obs}}{\mathcal{D}}\right) \tag{C.8}$$

The δ-function directly select radius R_E, and thus one is led to:

$$L_{\rm iso}(t_{\rm obs},\nu_{\rm obs}) = \frac{E_0}{\Gamma\Omega_p}\int_{(\alpha,\psi)\in\mathbf{\Omega}_p} {\rm d}\Omega\mathcal{D}^2 s'\left(\frac{\nu_{\rm obs}}{\mathcal{D}}\right)\delta\left(T_E - \frac{R_E\cos\alpha}{c} - t_{\rm obs}\right) \tag{C.9}$$

Using again $\delta(ax) = \frac{1}{|a|}\delta(x)$, one obtains:

$$L_{\rm iso}(t_{\rm obs},\nu_{\rm obs}) = \frac{E_0 c}{\Gamma\Omega_p R_E}\int_{\alpha_m}^{\alpha_M} {\rm d}(\cos\alpha)\int_{\psi_m(\alpha)}^{\psi_M(\alpha)} {\rm d}\psi\mathcal{D}^2 s'\left(\frac{\nu_{\rm obs}}{\mathcal{D}}\right)\delta\left(\cos\alpha - \frac{c(T_E - t_{\rm obs})}{R_E}\right) \tag{C.10}$$

where we have introduced the following notation:

- α_m and α_M are respectively the minimum and maximum α-coordinate of any point on the patch. If the line of sight intersects the patch (the *on-axis* case), i.e., if $\alpha_c < r$, then $\alpha_m = 0$. In the *off-axis* case, $\alpha_m = \alpha_c - r$. In any case, one has:

$$\begin{cases} \alpha_m = \max(0, \alpha_c - r) \\ \alpha_M = \min(\pi, \alpha_c + r) \end{cases} \tag{C.11}$$

- $\psi_m(\alpha)$ and $\psi_M(\alpha)$ are respectively the minimum and maximum ψ-coordinate of the intersection of the patch with the parallel defined by constant α. For our calculation, given the symmetry of the patch about the $\psi = 0$ meridian, we need only $\Delta\psi(\alpha) = \psi_M(\alpha) - \psi_m(\alpha)$. This function is found to be:

$$\alpha_c < r\ ({\rm on-axis}) : \Delta\psi(\alpha) = \begin{cases} 2\pi & \text{if} \quad \alpha < r - \alpha_c \\ 2\arccos\left(\frac{\cos\alpha - \cos\alpha_c\cos\alpha}{\sin\alpha_c\sin\alpha}\right) & \text{if} \quad r - \alpha_c < \alpha < \alpha_c + r \\ 0 & \text{if} \quad \alpha_c + r < \alpha \end{cases} \tag{C.12}$$

and

$$\alpha_c > r\ ({\rm off-axis}) : \Delta\psi(\alpha) = \begin{cases} 0 & \text{if} \quad \alpha < \alpha_c - r \\ 2\arccos\left(\frac{\cos\alpha - \cos\alpha_c\cos\alpha}{\sin\alpha_c\sin\alpha}\right) & \text{if} \quad \alpha_c - r < \alpha < \alpha_c + r \\ 0 & \text{if} \quad \alpha_c + r < \alpha \end{cases} \tag{C.13}$$

Thus, the isotropic-equivalent spectral luminosity is:

$$L_{\rm iso}(t_{\rm obs},\nu_{\rm obs}) = \frac{E_0 c}{\Gamma\Omega_p R_E}\int_{\alpha_m}^{\alpha_M} {\rm d}(\cos\alpha)\Delta\psi(\alpha)\mathcal{D}^2 s'\left(\frac{\nu_{\rm obs}}{\mathcal{D}}\right)\delta\left(\cos\alpha - \frac{c(T_E - t_{\rm obs})}{R_E}\right) \tag{C.14}$$

The δ-function will select $\cos\alpha = \frac{c(T_E - t_{\rm obs})}{R_E}$ if $\cos\alpha \in [\cos\alpha_M, \cos\alpha_m]$, and contribute 0 elsewhere.

Therefore, replacing the Doppler factor by its value for the material contributing at $t_{\rm obs}$, i.e., $\cos\alpha(t_{\rm obs}) = \frac{c(T_E - t_{\rm obs})}{R_E}$, we can conclude that:

$$L_{\rm iso}(t_{\rm obs},\nu_{\rm obs}) = \frac{E_{0,\rm iso} c}{4\pi\Gamma R_E}\Delta\psi(t_{\rm obs})\mathcal{D}^2(t_{\rm obs}) s'\left(\frac{\nu_{\rm obs}}{\mathcal{D}(t_{\rm obs})}\right)\mathbb{1}_{[t_m,t_M]}(t_{\rm obs}) \tag{C.15}$$

or in a more explicit, yet brutal, form:

$$L_{\rm iso}(t_{\rm obs},\nu_{\rm obs}) = \frac{E_{0,\rm iso} c}{4\pi\Gamma R_E}\frac{\Delta\psi\left(\alpha = \arccos\frac{(T_E - t_{\rm obs})c}{R_E}\right)}{\Gamma^2\left(1 - \frac{\beta c(T_E - t_{\rm obs})}{R_E}\right)^2} s'\left(\nu_{\rm obs}\Gamma\left(1 - \frac{\beta c(T_E - t_{\rm obs})}{R_E}\right)\right)\mathbb{1}_{[t_m,t_M]}(t_{\rm obs}) \tag{C.16}$$

where we have noted $E_{0,\rm iso} = 4\pi E_0/\Omega_p$ the isotropic-equivalent dissipated energy in the shell, and we have introduced the following times:

$$t_{m/M} = T_E - \frac{R_E\cos\alpha_{m/M}}{c} \tag{C.17}$$

These times are the first and last times the radiation from the patch is observed.

Starting from Eq. C.15, we can check our expression for j' (Eq. C.3). Indeed, for an entire pulsating sphere, we have the following total observed bolometric dissipated energy:

$$\int_{t_m}^{t_M} {\rm d}t_{\rm obs}\int {\rm d}\nu_{\rm obs} L(t_{\rm obs},\nu_{\rm obs}) = \frac{E_{0,\rm iso} c}{4\pi\Gamma R_E}\int {\rm d}t_{\rm obs}\int {\rm d}\nu_{\rm obs} 2\pi\times s'\left(\frac{\nu_{\rm obs}}{\mathcal{D}(t_{\rm obs})}\right)\mathcal{D}^2(t_{\rm obs}) \tag{C.18}$$

where we have replaced $\Delta\psi$ by 2π as convenient for a sphere.

We have $\int \mathrm{d}\nu s'(\nu/\mathcal{D}) = \mathcal{D}$. Therefore, this is:

$$\frac{E_{0,\mathrm{iso}}c}{2\Gamma R_E}\int_{t_m}^{t_M}\mathrm{d}t_{\mathrm{obs}}\mathcal{D}^3(t_{\mathrm{obs}}) = \frac{E_{0,\mathrm{iso}}}{2\Gamma}\int_{-1}^{1}\mathrm{d}\cos\alpha\frac{1}{\Gamma^3(1-\beta\cos\chi)^3} \tag{C.19}$$

where we used $\cos\alpha(t_{\mathrm{obs}}) = \frac{c(T_E - t_{\mathrm{obs}})}{R_E}$ as before.

Changing to $u = \beta\cos\chi$, we have:

$$\frac{E_{0,\mathrm{iso}}}{2\Gamma^4\beta}\int_{-\beta}^{\beta}\frac{\mathrm{d}u}{(1-u)^3} \tag{C.20}$$

Evaluating the dimensionless integral yields:

$$\int_{-\beta}^{\beta}\frac{\mathrm{d}u}{(1-u)^3} = 2\beta\Gamma^4 \tag{C.21}$$

Therefore, the total dissipated energy is $E_{0,\mathrm{iso}}$, as expected.

In the case of a pulsing sphere and defining $R_E = \beta c T_E$ (i.e., counting time from the ejection of the sphere), one obtains $L_{\mathrm{iso}}(t_{\mathrm{obs}},\nu_{\mathrm{obs}}) \propto t_{\mathrm{obs}}^{-2}s'\left(\nu_{\mathrm{obs}}\Gamma\frac{t_{\mathrm{obs}}}{T_E}\right)$. Thus, by writing $L^{\mathrm{iso}} \propto t_{\mathrm{obs}}^a\nu_{\mathrm{obs}}^b$ and assuming a power-law emitted spectrum, one obtains $a = b-2$, which is the *consistency relation* for high-latitude emission.

C.2 Transformation of luminosity from aligned to misaligned lines of sight

In the previous section, we derived the luminosity for a flashing cone starting from the central source frame dissipated energy. It is also useful to transform the observed luminosity from a given line of sight to the cone to another, as required for example in Chap. 10. Here we will derive such a transformation for different viewing angle regimes.

We start from the definition for the spectral luminosity:

$$L_\nu(t) = \int\frac{\mathrm{d}L_\nu(t)}{\mathrm{d}\Omega}\mathrm{d}\Omega \tag{C.22}$$

Where $\frac{\mathrm{d}L_\nu(t)}{\mathrm{d}\Omega} = \frac{\mathrm{d}E}{\mathrm{d}t\mathrm{d}\nu\mathrm{d}\Omega}$ is the emitted energy per unit time, frequency and emitting region solid angle and we identify observer time and emission time, disregarding light travel time delays. Under frame change from the emitter's frame to the observer's frame, $\mathrm{d}E$ transforms as $\mathcal{D}$ and $\mathrm{d}t$ and $\mathrm{d}\nu$ transform as $\mathcal{D}^{-1}$, where $\mathcal{D} = \frac{1}{\Gamma(1-\beta\cos\theta)}$ is the Doppler factor.

Therefore, we have:

$$L_\nu(t) = \int\mathcal{D}^3L'_{\nu'}(t)\mathrm{d}\Omega \tag{C.23}$$

Different lines of sight to the source change the material dominating the received emission, i.e. the solid angles with the largest Doppler factor among those which radiate. We denote by θ_0 the total angular size of the emitting region and $\delta\theta$ the angular distance between the line of sight and the edge of the emitting region. We therefore have the following regimes:

- Aligned line of sight ($\delta\theta = 0$): The observer is dominated by a ring with $\theta < 1/\Gamma$, thus $\mathrm{d}\Omega \sim 2\pi 1/2\Gamma^2$, the Doppler factor is $\mathcal{D} \sim 2\Gamma$ and:
$$L_\nu^{\mathrm{on}} \sim 8\pi\Gamma L'_{\nu'} \tag{C.24}$$

- Slightly misaligned line of sight ($1/\Gamma \ll \delta\theta \ll \theta_0$): The flux is dominated by regions within the emitting region that are the most boosted, with $\delta\theta < \theta < 2\delta\theta$. In addition, the emitting region is limited to a transverse angular size of $\Delta\phi \sim \pi$ of the emitting region which occupies nearly a half plane in the observer's field. Thus $\mathrm{d}\Omega \sim 3\delta\theta^2\Delta\phi/2$, the Doppler factor is $\mathcal{D} = \frac{1}{\Gamma(1-\beta\cos\delta\theta)} \sim 2\Gamma/(1+\Gamma^2\delta\theta^2)$ and :
$$L_\nu^{\mathrm{off}} \sim \frac{8\Gamma^3}{(1+\Gamma^2\delta\theta^2)^3}\frac{3\Delta\phi\delta\theta^2}{2}L'_{\nu'} \sim \frac{\Delta\phi}{2\pi}S^{-2}L_\nu^{\mathrm{on}} \tag{C.25}$$
where we have used $S = \frac{1-\beta\cos\delta\theta}{1-\beta} \sim 1+\Gamma^2\delta\theta^2$ and $\Gamma\delta\theta \gg 1$. Therefore $L_\nu^{\mathrm{off}}/L_\nu^{\mathrm{on}} \sim f_{\mathrm{geo}}S^{-2}$, with $f_{\mathrm{geo}} = \Delta\phi/2\pi \sim 1/2$.

- Significantly misaligned ($\delta\theta > \theta_0$): The whole emitting region has nearly the same Doppler factor and still $\Gamma\delta\theta \gg 1$. Thus $\mathrm{d}\Omega \sim \theta_0^2$ and one finds $L_\nu^{\mathrm{off}}/L_\nu^{\mathrm{on}} \sim (\Gamma\theta_0)^2S^{-3}$. This case does not occur in our setup of Chap. 10.

Therefore, defining $S = \frac{1-\beta\cos\delta\theta}{1-\beta}$, the spectral luminosity transforms as the following for slightly misaligned lines of sight:

$$\frac{L_\nu^{\rm off}}{L_\nu^{\rm on}} \sim f_{\rm geo} S^{-2} \tag{C.26}$$

With a similar reasoning, we have for bolometric luminosities:

$$\frac{L^{\rm off}}{L^{\rm on}} \sim f_{\rm geo} S^{-3} \tag{C.27}$$

Appendix D

Dynamics for the deceleration of relativistic ejecta

In this appendix we derive the kinematics for the deceleration of relativistic ejecta through an external medium.

D.1 Preliminary: dissipation of kinetic energy to internal energy of the shocked matter

Consider a mass M with Lorentz factor Γ impacting a smaller mass m at rest. Upon collision, the two masses acquire a common Lorentz factor Γ'. In the mean time, a shock propagates in the smaller mass such that it is heated to an internal Lorentz factor γ_i. In an external rest frame, the conservation of energy-momentum reads:

$$\begin{cases} \Gamma M c^2 + mc^2 = \Gamma' M + \Gamma' \gamma_i m c^2 \\ \Gamma \beta M + 0 = \Gamma' \beta' M + \Gamma' \beta' \gamma_i m \end{cases} \tag{D.1}$$

The post-collision velocity is then:

$$\beta' = \beta \frac{\Gamma M}{\Gamma M + m} \tag{D.2}$$

Supposing a small mass ratio $m/M \ll 1$ and a large initial Lorentz factor, we write $\frac{\Gamma M}{\Gamma M + m} \sim 1 - m/\Gamma M$ and $\beta \sim 1 - 1/2\Gamma^2$. This leads to:

$$\frac{\Gamma}{\Gamma'} = 1 + \Gamma \frac{m}{M} \tag{D.3}$$

and finally:

$$\gamma_i = \frac{\Gamma M + m - \Gamma' M}{\Gamma' M} \tag{D.4}$$

$$\sim \left(\frac{\Gamma}{\Gamma'} - 1 \right) \frac{M}{m} \tag{D.5}$$

$$\sim \Gamma' \tag{D.6}$$

Therefore, energy conservation is put in the following form:

$$\Gamma M + m = \Gamma'(M + m) + \Gamma'(\Gamma' - 1)m \tag{D.7}$$

where we evidenced the joint mass's rest and kinetic energy $\Gamma'(M + m)$ on one side and the shocked material's internal energy $\Gamma'(\Gamma' - 1)$ on the other.

D.2 Adiabatic deceleration

We now consider the continuous deceleration of the ejecta in the external medium. We denote by Γ_0 its initial Lorentz factor and $M_{\rm ej}$ its mass. Until we consider any transverse spreading of the ejecta in Appendix D.8, we will deal will isotropic-equivalent ejecta masses and energies. The initial energy of the ejecta is $E = \Gamma_0 M_{\rm ej} c^2$.

The deceleration laws we will derive below must be understood as applying to the dynamically relevant species, in our case these will be the protons, which collect the most mass both in the ejected matter and in the external medium.

As the ejecta penetrates the external medium at supersonic speed, a shock forms that accumulates and heats matter. We denote by $\Gamma(r)$ the Lorentz factor of the shock front once it has reached radius r and $m(r)$ the mass accumulated at that radius. Let us introduce the *density profile parameter* of the external medium s. It is defined such that the mass density at radius r is $\rho_{\rm ext} \propto r^{-s}$, and therefore $m(r) \propto r^{3-s}$. For a homogeneous medium, $s = 0$ and for a stellar wind, $s = 2$.

If the shocked matter is radiatively inefficient, the energy in the shock will be the internal energy of all the accumulated matter, at at internal Lorentz factor of $\gamma_i = \Gamma(r) - 1$, as shown in Appendix D.1. Therefore, the energy balance equation is:

$$\Gamma_0 M_{\rm ej} + m(r) = \Gamma(r)(M_{\rm ej} + m(r)) + \Gamma(r)(\Gamma(r) - 1)m(r) \tag{D.8}$$

This equation is readily solved as:

$$\Gamma(r) = \Gamma_0 \frac{-1 + \sqrt{1 + 4\eta(r) + 4\frac{\eta(r)^2}{\Gamma_0^2}}}{2\eta(r)} \tag{D.9}$$

where $\eta(r) = \frac{m(r)}{M_{\rm ej}/\Gamma_0}$.

Such deceleration dynamics exhibit three phases:

1. For $\eta(r) \ll 1/4$, in the *coasting phase*, no deceleration occurs before the critical mass $M_{\rm dec} = M_{\rm ej}/\Gamma_0$ is swept-up:

$$\Gamma(r \ll R_{\rm dec}) \sim \Gamma_0 \tag{D.10}$$

   This phase continues until the *deceleration radius* $R_{\rm dec}$, at which the accumulated mass equals the critical mass.

2. For $1/4 \ll \eta(r) \ll \Gamma_0^2$, the shock front decelerates. During this *deceleration phase*, the shocked matter mass progressively reaches $\Gamma_0 M_{\rm ej}$:

$$\Gamma(R_{\rm dec} \ll r \ll R_N) \sim \frac{\Gamma_0}{\sqrt{\eta(r)}} \tag{D.11}$$

$$\sim \Gamma_0 \left(\frac{r}{R_{\rm dec}}\right)^{-(3-s)/2} \tag{D.12}$$

3. Finally, once the *Newtonian radius* $R_N = R_{\rm dec}\Gamma_0^{2/(3-s)}$ is reached, the shock is no longer relativistic, and the balance equation Eq. D.8 is no longer valid. Then, we have:

$$\Gamma(R_N < r) \sim 1 \tag{D.13}$$

For a homogeneous medium of particle density $n_{\rm ext}$, we have:

$$\begin{cases} \rho(r) = n_{\rm ext} m_P \\ m(r) = 4\pi n_{\rm ext} m_P r^3/3 \\ R_{\rm dec} = \left(\frac{3E}{4\pi n_{\rm ext} m_P \Gamma_0^2 c^2}\right)^{1/3} \\ \Gamma(r) \propto r^{-3/2} \text{ (dec. phase)} \end{cases} \tag{D.14}$$

For a stellar wind of parameter A, we have:

$$\begin{cases} \rho(r) = A r^{-2} \\ m(r) = 4\pi A r \\ R_{\rm dec} = \frac{M_{\rm ej}}{4\pi\Gamma_0 A} \\ \Gamma(r) \propto r^{-1/2} \text{ (dec. phase)} \end{cases} \tag{D.15}$$

In Fig. D.1 we show the $\Gamma(r)$ function in the adiabatic regime for both a homogeneous and a stellar wind external medium.

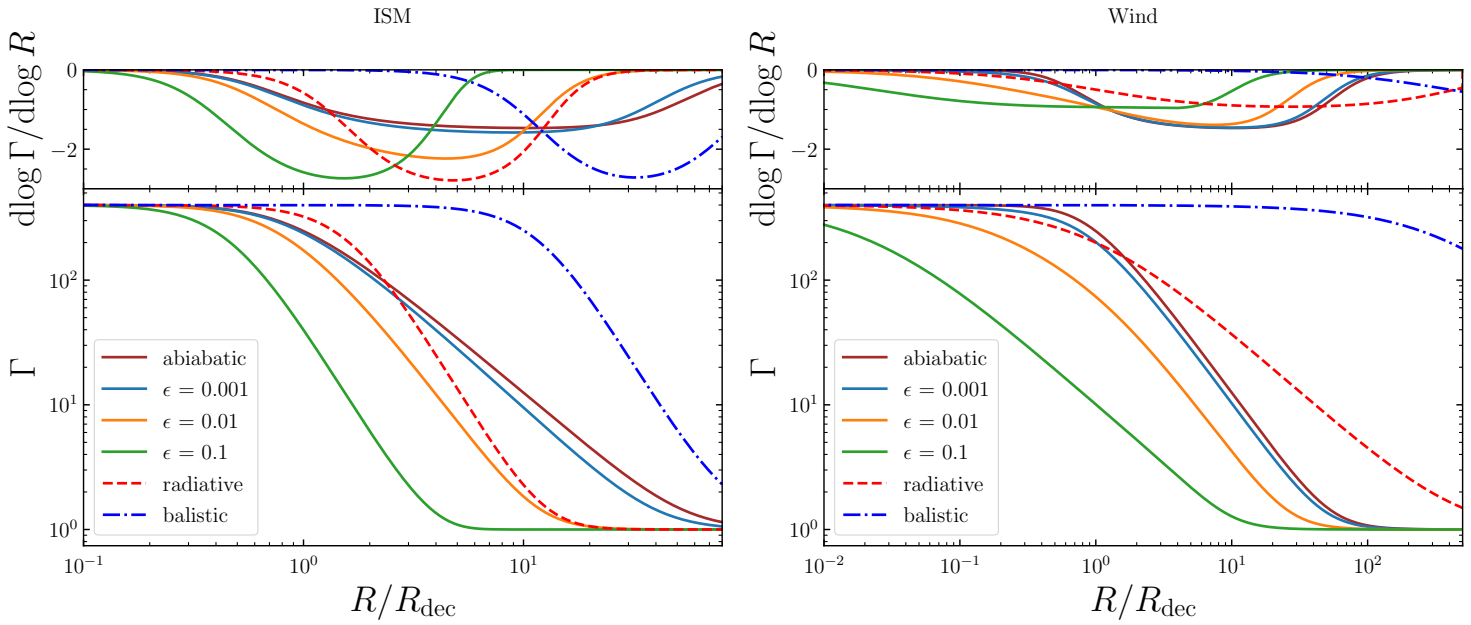


Figure D.1: **Deceleration dynamics in various radiation regimes and external density profiles.** In all cases we have $\Gamma_0 = 400$ and $E = 10^{53}$erg. Left: Homogeneous medium with $n_{\rm ext} = 1$ cm^{-3}, and thus $R_{\rm dec} = 0.015$ pc. Right: Stellar wind medium with $A = 5 \times 10^{11}$ g/cm and thus $R_{\rm dec} = 3.6 \times 10^{-5}$ pc. Top: Logarithmic derivative of $\Gamma(r)$ evidencing the deceleration phases and power laws.

D.3 Weakly radiative deceleration

The balance equation Eq. D.8 supposes no energy leaks from the shock, thus the internal energy term is $\Gamma(r)(\Gamma(r)-1)m(r)$ throughout the dynamics.

Suppose that a fraction ϵ of the internal energy of the shocked matter is radiated upon its shocking. In this case, all the radiated energy from the matter shocked in the past must be removed from the shock energy at radius r. Thus, the balance equation writes:

$$\Gamma_0 M_{\rm ej} + m = \Gamma(r)(M_{\rm ej} + m(r)) + \Gamma(r)(\Gamma(r)-1)m(r) - \epsilon \int_{r'=0}^{r} {\rm d}r' \frac{{\rm d}m}{{\rm d}r} \Gamma(r')(\Gamma(r')-1) \tag{D.16}$$

Eq. D.16 supposes some energy is left on the shock after dissipation and is therefore valid only for small ϵ. In Fig. D.1 we show the deceleration dynamics for various values of ϵ.

D.4 Fully radiative deceleration

In the case where the shock is radiatively efficient, the shocked material radiates all of its internal energy upon shocking. In this case, there is no internal energy left stocked in the shocked and the balance equation writes:

$$\Gamma_0 M_{\rm ej} + m(r) = \Gamma(r)(M_{\rm ej} + m(r)) - \int_{r'=0}^{r} {\rm d}r' \frac{{\rm d}m}{{\rm d}r} \Gamma(r')(\Gamma(r')-1) \tag{D.17}$$

where the internal energy term at r disappeared and we removed all the dissipated energy from the energy budget at r.

In this regime, the remnant decelerates approximately as $\Gamma(r) \propto 1/m(r) \propto r^{-(3-s)}$. In Fig. D.1 we show $\Gamma(r)$ for this regime as well.

D.5 Balistic regime

For completeness, we also mention the balistic regime, where the shocked medium remains completely cold ($\gamma_i = 1$) and thus the collisions are elastic. This regime however does not conserve the mass of the ejecta, and its physical sense is thus not clear. Conservation of energy writes:

$$\Gamma_0 M_{\rm ej} + m(r) = \Gamma(r)(M_{\rm ej} + m(r)) \tag{D.18}$$

As no energy is stored nor dissipated in the shock, this regime has the slowest deceleration, as shown in Fig. D.1.

D.6 Jet deceleration and break times for aligned observers

The deceleration dynamics have up to now been expressed in terms of radial coordinate. However, for a GRB observer the relevant coordinate is observer time, denoted by $t_{\rm obs}$. In the setup where a central engine launches an outflow towards the observer, $t_{\rm obs}$ is defined by the arrival time of photons emitted in the outflow at radius R and time t. For material on the observer-central engine line, this is:

$$t_{\rm obs} = t - \frac{R}{c} \tag{D.19}$$

such that $t_{\rm obs} = 0$ for photons emitted at time 0 from the central source itself.

The outflow's deceleration phase will start to be observed at:

$$t_{\rm obs,dec} = \frac{R_{\rm dec}}{\beta_0 c} - \frac{R_{\rm dec}}{c} \tag{D.20}$$

$$\sim \frac{R_{\rm dec}}{2c\Gamma_0^2} \tag{D.21}$$

where we used the ultra-relativistic approximation $1 - \beta_0 \sim 1/2\Gamma_0^2$.

Eq. D.21 can be completed with the expressions for $R_{\rm dec}$ in the different deceleration regimes and external density profiles, e.g., Eqs. D.14 and D.15.

In the deceleration phase, we often introduce the index ϵ such that we have $\Gamma(r) \propto r^{-\epsilon}$. For an adiabatic remnant, we have $\epsilon = (3-s)/2$, equal to $3/2$ (uniform medium) or $1/2$ (stellar wind).

In terms of $t_{\rm obs}$, the deceleration phase can be written as:

$$\frac{{\rm d}\Gamma}{{\rm d}t_{\rm obs}} = \frac{{\rm d}\Gamma}{{\rm d}r}\frac{{\rm d}r}{{\rm d}t_{\rm obs}} \tag{D.22}$$

$$\propto r^{-3\epsilon-1} \tag{D.23}$$

$$\propto t_{\rm obs}^{\frac{-3\epsilon-1}{2\epsilon+1}} \tag{D.24}$$

such that we finally obtain:

$$\Gamma(t_{\rm obs}) = \Gamma_0 \left(\frac{t_{\rm obs}}{t_{\rm obs,dec}}\right)^{-\frac{\epsilon}{2\epsilon+1}} \tag{D.25}$$

which is $t_{\rm obs}^{-3/8}$ for a uniform medium and $t_{\rm obs}^{-1/4}$ for a stellar wind.

The deceleration dynamics from Eq. D.25 readily allow to predict the observer jet break time (Sec. 1.2). For a jet with half-opening angle θ_j observed down its axis, the jet break occurs at time t_b such that $\Gamma(t_b)\theta_j \sim 1$, which allows to easily find:

$$t_b = t_{\rm obs,dec}\left(\Gamma_0\theta_j\right)^{2+\frac{1}{\epsilon}} \tag{D.26}$$

D.7 Shocked region radial width

In some contexts it is interesting to know the radial extent ΔR of the shocked region. For the ultra-relativistic shocks we consider, the shock-frame particle number density is:

$$n'_s = (4\Gamma + 3)n_{\rm ext} \tag{D.27}$$

In the external rest frame, this is:

$$n_s = \Gamma n'_s \sim 4\Gamma^2 n_{\rm ext} \tag{D.28}$$

All the mass swept up by the remnant is stocked in the shocked region, and therefore:

$$m(r) = 4\pi r^2 \Delta R n_s \tag{D.29}$$

which leads to the following expressions:

$$\begin{cases} \Delta R = \frac{r}{12\Gamma^2} & \text{(homogeneous)} \\ \Delta R = \frac{r}{4\Gamma^2} & \text{(stellar wind)} \end{cases} \tag{D.30}$$

We thus understand that the shocked material is extremely compressed and the shocked region's radial width is very thin.

D.8 Transverse spreading of the relativistic jet

We seek to estimate the lateral spreading of a structured jet's core over the course of an afterglow observation, for any viewing angle. We denote by $\theta_{j,0}$ the initial opening angle of the core, and $\theta_{j,f}$ its value at the time of interest. This time is, e.g., the peak of the afterglow light curve for a misaligned jet (Sec. 4.4.5) or the end of the plateau phase in the slightly misaligned line of sight interpretation (Chap. 9).

For the evolution of θ_j we adopt the prescription of Granot & Piran (2012), Eq. 13 with $a = 1$. This spreading law is derived from a consistent description of the shocked material of the strong forward shock and provides a satisfactory fit of the numerical simulations. This is:

$$\frac{d\theta_j}{dR} \sim \frac{1}{R\Gamma^2\,\theta_j}\,. \tag{D.31}$$

After the onset of deceleration of the core, we can write $\Gamma \propto R^{-\epsilon}$ for the deceleration dynamics under the adiabatic remnant hypothesis, with $\epsilon = 3/2$ and $1/2$ for a uniform circum-burst medium or a wind medium, respectively. These scalings, which were derived for a spherical outflow, are justified only if lateral spreading is ultimately found to be negligible. Using these relations we integrate Eq. D.31 to:

$$\theta_{j,f}^2 - \theta_{j,0}^2 = \frac{1}{\epsilon}\left(\frac{1}{\Gamma_p^2} - \frac{1}{\Gamma_j^2}\right) \tag{D.32}$$

where we have denoted by Γ_j the core's bulk Lorentz factor at the afterglow onset and Γ_p the core's bulk Lorentz factor at the time of interest.

For both the afterglow peak time and the end of the plateau phase, Γ_p is given by the core's beaming condition: $\Gamma_p^{-1} = \theta_{j,f} - \theta_v$, so that:

$$\theta_{j,f}^2 - \theta_{j,0}^2 = \frac{1}{\epsilon}\left[(\theta_v - \theta_{j,f})^2 - \frac{1}{\Gamma_j^2}\right] \tag{D.33}$$

which in turn leads to:

$$x_f = \frac{-x_v + \sqrt{x_v^2 + (\epsilon - 1)\left(x_v^2 + \epsilon - \frac{1}{\Gamma_j^2}\right)}}{\epsilon - 1} \tag{D.34}$$

where we have introduced $x_f = \theta_{j,f}/\theta_{j,0}$ and $x_v = \theta_v/\theta_{j,0}$ the final core opening angle and viewing angle in units of the initial opening angle. Both are larger than 1 by construction.

In Fig. D.2, we plot x_f as a function of x_v for both slightly misaligned lines of sight ($x_v \sim 1$) and very off-axis cases ($x_v \leq 2$). It is clear that jet expansion is negligible for slightly misaligned setups. For more off-axis events ($x_v \sim 4$ for GW170817), we predict that the core jet expanded by a factor of ~ 2. While this can seem significant, the calculation is valid for a bare core and core expansion is expected to be weaker for an actual structured jet.

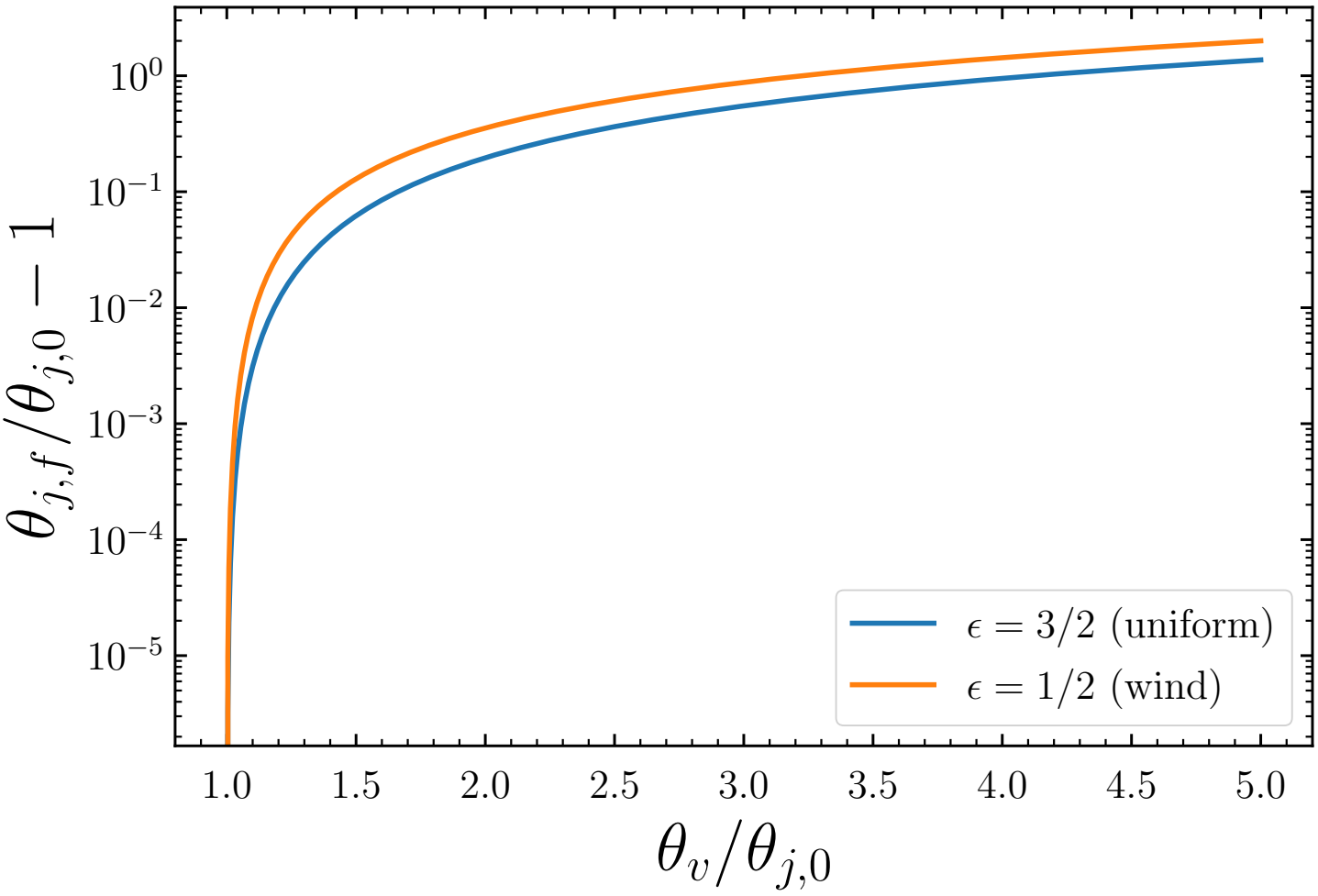


Figure D.2: **Final jet opening angle after having laterally expanded over the course of an afterglow, as a function of the viewing angle.** The final opening angle is determined when the core jet reveals. An initial Lorentz factor of $\Gamma_j = 400$ was chosen. The curve are undistinguished for any $\Gamma_j \geq 50$.

Appendix E

Emission spectrum from the shock-accelerated electrons

We seek to derive the emission spectrum from a population of electrons accelerated at the forward shock of decelerating matter. In practical implementations, the emission can only can calculated over a discrete time step $t_{\rm num}$. Seeing the often power-law dependences of Γ on R (e.g., eq D.12), the discretization is often chosen to be logarithmic in R: $\mathrm{d}R/R = \text{const.}$. In terms of the time coordinate, this is const. $= \mathrm{d}t/t_{\rm ex}$, where we have introduced the *expansion timescale*:

$$t_{\rm ex} = R/\beta c \tag{E.1}$$

as given in the lab frame. This is the time needed for the ejecta to double its size at the current expansion speed. This is also the timescale for adiabatic cooling of matter contained in the shock, the *dynamical timescale* $t_{\rm dyn} = t_{\rm ex}$.

We will thus derive the average emission spectrum from electrons over $t_{\rm ex}$. We shall consider the synchrotron emission mechanism, and two additional processes that are relevant at the observing frequencies considered in this work: synchrotron self-absorption and inverse Compton scattering.

E.1 Synchrotron emission spectrum from a single electron

We denote all quantities in the comoving frame of the shocked material with primes. In this frame, the dynamical timescale is:

$$t'_{\rm dyn} = \frac{R}{\Gamma\beta c} \tag{E.2}$$

Through the synchrotron process, a single electron of Lorentz factor γ'_e emits a spectral power varying approximately as $\nu'^{1/3}$ until a maximum frequency of (Rybicki & Lightman, 1979, , Chap. 6):

$$\nu'_s(\gamma'_e) = \gamma'^2_e \frac{eB'}{2\pi m_e c} \tag{E.3}$$

where we denote by B' the magnetic field magnitude.

The spectral power at ν'_s is:

$$P'_{\rm max} = \frac{m_e c^2 \sigma_T B'}{3e} \tag{E.4}$$

The total emitted synchroton power is $P' \sim P'_{\rm max}\nu'_s$ and thus the timescale for synchrotron cooling is therefore:

$$t'_{\rm syn} = \frac{\gamma'_e m_e c^2}{P'} \tag{E.5}$$

$$= \frac{6\pi m_e c}{\sigma_T \gamma'_e B'^2} \tag{E.6}$$

We denote by γ'_c the Lorentz factor of an electron that would cool by synchrotron emission in a dynamical timescale:

$$\gamma'_c = \frac{6\pi m_e c}{\sigma_T B'^2 t'_{\rm dyn}} \tag{E.7}$$

The magnetic field is prescribed by the hydrodynamical and microphysical conditions in the forward shock, which we consider constant on a dynamical timescale to simplify the discussion. The characteristic γ'_c therefore also varies little on this timescale.

In this notation, the synchrotron power rewrites:

$$P' = \gamma_e'^2 m_e c^2 \frac{1}{\gamma'_c t'_{\rm dyn}} \tag{E.8}$$

such that, starting at γ'_0 at $t' = 0$, an electron will cool according to the following differential equation in $\gamma'_e(t')$:

$$\dot{\gamma'_e} = -\frac{\gamma_e'^2}{\gamma'_c t'_{\rm dyn}} - \frac{\gamma'_e}{t'_{\rm dyn}} \tag{E.9}$$

where the second term represents the electron's cooling by expansion of the ejecta.

Eq. E.9 can be solved in:

$$\gamma'_e(t) = \frac{\gamma'_c}{\left(1 + \frac{\gamma'_c}{\gamma'_0}\right) e^{t'/t'_{\rm dyn}} - 1} \tag{E.10}$$

If $\gamma'_0 \ll \gamma'_c$, then the electron cools in a longer time than the expansion timescale, and we have:

$$\gamma'_e(t') \sim \gamma'_0 e^{-t'/t'_{\rm dyn}} \sim \gamma_0 \tag{E.11}$$

Over a dynamical timescale, the Lorentz factor of such electrons essentially does not change. These are said to be *slow cooling* electrons.

If $\gamma'_0 < \gamma'_c$, then the electron will cool to γ'_c in a time $t'_c < t'_{\rm dyn}$ such that:

$$\gamma'_e(t'_c) = \gamma'_c \Leftrightarrow t'_c = t'_{\rm dyn} \frac{\gamma'_0 - \gamma'_c}{\gamma'_0 + \gamma'_c} < t'_{\rm dyn} \tag{E.12}$$

where we used Eq. E.10 and the approximation $e^{t'_c/t'_{\rm dyn}} \sim 1 + t'/t'_{\rm dyn}$.

Such *fast cooling* electrons will therefore have the following Lorentz factor evolution:

$$\gamma'_e(t') = \begin{cases} \frac{1}{\frac{1}{\gamma'_0} + \frac{t'}{t'_{\rm dyn}}\left(\frac{1}{\gamma'_c} + \frac{1}{\gamma'_0}\right)} & \text{for} \quad t' < t'_c \\ \gamma'_c & \text{for} \quad t'_c < t' < t'_{\rm dyn} \end{cases} \tag{E.13}$$

The most fast cooling electrons, with $\gamma'_0 \gg \gamma'_c$, have $t'_c = t'_{\rm dyn}$ and therefore Eq. E.13 simplifies to:

$$\gamma'_e(t') = \gamma'_c t'_{\rm dyn}/t' \tag{E.14}$$

We denote by ν'_c the synchrotron frequency of an electron with Lorentz factor γ'_c, i.e. $\nu'_c = \nu'_s(\gamma'_c)$.

For a slow cooling electron, we will have $\nu'_e(t') < \nu'_c$ at all times, where we use the shorthand $\nu'_e(t') = \nu'_s(\gamma'_e(t'))$, therefore the average spectral power from such an electron over a dynamical timescale will be:

$$\langle s' \rangle_{t_{\rm dyn}}(\nu') = \frac{1}{t'_{\rm dyn}} \int_0^{t'_{\rm dyn}} {\rm d}t' P'_{\rm max} \left(\frac{\nu'}{\nu'_e(t')}\right)^{1/3} \mathbb{1}_{[0,\nu'_e(t')]}(\nu') \tag{E.15}$$

$$= P'_{\rm max} \left(\frac{\nu'}{\nu'_0}\right)^{1/3} \mathbb{1}_{[0,\nu'_0]}(\nu') \tag{E.16}$$

because the electron barely cools over $t'_{\rm dyn}$ (see Eq. E.11). The spectrum thus consists of the ordinary single branch up to the initial synchrontron frequency.

For a fast cooling electron, the instantaneous emission spectrum evolves because of the electron's cooling from γ'_0 to γ'_c over the dynamical timescale.

If $\nu' > \nu'_0$, the instantaneous spectrum never contains ν and therefore $\langle s' \rangle_{t_{\rm dyn}}(\nu_0 < \nu) = 0$.

If $\nu' < \nu'_c$, we have $\nu' < \nu'_e(t')$ at all times, and therefore using Eq. E.14 we obtain:

$$\langle s' \rangle_{t_{\rm dyn}} (\nu' < \nu'_c) = \frac{1}{t'_{\rm dyn}} \int_0^{t'_{\rm dyn}} {\rm d}t' P'_{\rm max} \left(\frac{\nu'}{\nu'_e(t)} \right)^{1/3} \tag{E.17}$$

$$= \frac{1}{t'_{\rm dyn}} \int_0^{t'_{\rm dyn}} {\rm d}t' P'_{\rm max} \left(\frac{\nu'}{\nu'_c} \frac{t'^2}{t'^2_{\rm dyn}} \right)^{1/3} \tag{E.18}$$

$$= P'_{\rm max} \left(\frac{\nu'}{\nu'_c} \right)^{1/3} \frac{1}{t'_{\rm dyn}} \int_0^{t'_{\rm dyn}} {\rm d}t' \left(\frac{t'}{t'_{\rm dyn}} \right)^{2/3} \tag{E.19}$$

$$\sim P'_{\rm max} \left(\frac{\nu'}{\nu'_c} \right)^{1/3} \tag{E.20}$$

Finally, if $\nu'_c < \nu' < \nu'_0$, then there will be a time t'_ν such that $\nu'_e(t'_\nu) = \nu'$. In this case, the instantaneous spectra must be integrated only from $t' = 0$ to t'_ν. This time is given by $t'_\nu = t_{\rm d}\sqrt{\nu'_c/\nu'}$ (Eq. E.14). We then have:

$$\langle s' \rangle_{t_{\rm dyn}} (\nu' < \nu'_c) = \frac{1}{t'_{\rm dyn}} \int_0^{t'_\nu} {\rm d}t' P'_{\rm max} \left(\frac{\nu'}{\nu'_e(t')} \right)^{1/3} \tag{E.21}$$

$$= P'_{\rm max} \left(\frac{\nu'}{\nu'_c} \right)^{1/3} \frac{1}{t'_{\rm dyn}} \int_0^{t'_\nu} {\rm d}t' \left(\frac{t'}{t'_{\rm dyn}} \right)^{2/3} \tag{E.22}$$

$$\sim P'_{\rm max} \left(\frac{\nu'}{\nu'_c} \right)^{1/3} \left(\frac{t'_\nu}{t'_{\rm dyn}} \right)^{5/3} \tag{E.23}$$

$$\sim P'_{\rm max} \left(\frac{\nu'}{\nu'_c} \right)^{-1/2} \tag{E.24}$$

To conclude, modulo factors close to one, the average synchrotron power spectrum emitted by a single electron with initial Lorentz factor γ'_0 over a dynamical timescale is:

$$\langle s' \rangle_{t_{\rm dyn}} (\nu', \gamma'_0) = P'_{\rm max} \times \begin{cases} \left(\frac{\nu'}{\nu'_c} \right)^{1/3} & \text{for} \quad \nu' < \nu'_c \\ \left(\frac{\nu'}{\nu'_c} \right)^{-1/2} & \text{for} \quad \nu'_c < \nu' < \nu'_0 \\ 0 & \text{for} \quad \nu'_0 < \nu' \end{cases} \tag{E.25}$$

E.2 Synchrotron emission spectrum from a power-law-distributed population

We suppose the population of electrons is initially accelerated to a power-law distribution with spectral index $p > 2$. We denote the minimal Lorentz factor in the population as γ'_m. The initial normalized population distribution then reads:

$$\mathcal{N}_0(\gamma'_e) = \begin{cases} 0 & \text{for} \quad \gamma'_e < \gamma'_m \\ \frac{p-1}{\gamma'_m} \left(\frac{\gamma'_e}{\gamma'_m} \right)^{-p} & \text{for} \quad \gamma'_m < \gamma'_e \end{cases} \tag{E.26}$$

The expression in Eq. E.26 is valid as the initial distribution of electron energies if the timescale $t_{\rm acc}$ for accelerating the electrons is much smaller than the dynamical timescale and the synchrotron cooling timescale. If this is not the case, the population will never assume such a distribution, because is cools while still being accelerated. In practice, this approximation is largely valid for the emitted frequency ranges we are interested in, up to hard X-rays.

The total spectrum emitted by the accelerated electron population is obtained by summing the individual emission spectra of the electrons with initial Lorentz factor distributed according to Eq. E.26:

$$S'_{\rm tot}(\nu) = \int_{\gamma'_m}^{\infty} {\rm d}\gamma'_e \mathcal{N}_0(\gamma'_e) \langle s' \rangle_{t_{\rm dyn}} (\nu', \gamma'_e) \tag{E.27}$$

We denote by $\nu'_m = \nu'_s(\gamma'_m)$ the synchrotron frequency associated with γ'_m. Evaluating Eq. E.27 leads one to consider two regimes:

1. $\nu'_m < \nu'_c$: The bulk of the electron population is in slow cooling. We call this case the *slow cooling regime*. Then one finds:

$$S'_{\rm tot}(\nu') = P'_{\rm max} \times \begin{cases} \left(\frac{\nu'}{\nu'_m}\right)^{1/3} & \text{for} \quad \nu' < \nu'_m \\ \left(\frac{\nu'}{\nu'_m}\right)^{-(p-1)/2} & \text{for} \quad \nu'_m < \nu' < \nu'_c \\ \left(\frac{\nu'_c}{\nu'_m}\right)^{-(p-1)/2} \left(\frac{\nu'}{\nu'_c}\right)^{-p/2} & \text{for} \quad \nu'_c < \nu' \end{cases} \tag{E.28}$$

2. $\nu'_c < \nu'_m$: All the electrons are fast cooling; This case is the *fast cooling regime*. Then:

$$S'_{\rm tot}(\nu') = P'_{\rm max} \times \begin{cases} \left(\frac{\nu'}{\nu'_c}\right)^{1/3} & \text{for} \quad \nu' < \nu'_c \\ \left(\frac{\nu'}{\nu'_c}\right)^{-1/2} & \text{for} \quad \nu'_c < \nu' < \nu'_m \\ \left(\frac{\nu'_m}{\nu'_c}\right)^{-1/2} \left(\frac{\nu'}{\nu'_m}\right)^{-p/2} & \text{for} \quad \nu'_m < \nu' \end{cases} \tag{E.29}$$

The peak of the emission spectrum is always $P'_{\rm max}$ at $\min(\nu'_m, \nu'_c)$.
The initial total energy in the electron population is:

$$U'_e = \langle \gamma'_e \rangle_0 \, m_e c^2 \tag{E.30}$$

$$= \int {\rm d}\gamma'_e \mathcal{N}_0(\gamma'_e) \gamma'_e m_e c^2 \tag{E.31}$$

$$= \frac{p-1}{p-2} \gamma'_m m_e c^2 \tag{E.32}$$

The total synchrotron emitted energy over $t_{\rm dyn}$ is:

$$U'_{\rm syn} = t'_{\rm dyn} \int {\rm d}\nu' S'_{\rm tot}(\nu') \tag{E.33}$$

The radiative efficiency is found to be:

$$\eta_{\rm syn} = \frac{U'_{\rm syn}}{U'_e} \sim \begin{cases} 1 & \text{(fast cooling)} \\ \left(\frac{\gamma'_m}{\gamma'_c}\right)^{p-2} & \text{(slow cooling)} \end{cases} \tag{E.34}$$

Typically, $\gamma'_m/\gamma'_c < 10^{-4}$ at the peak of an off-axis afterglows and thus $\eta_{\rm syn} \lesssim 15\%$. To obtain to total kinetic-to-radiation conversion efficiency ϵ of the relativistic jet, one must multiply $\eta_{\rm syn}$ by the fraction of shock internal energy injected in the electron population $\epsilon_e \lesssim 0.1$. Thus $\epsilon \lesssim 1\%$, for which the radiation does not impact the deceleration dynamics, as shown in Appendix D. This justifies the hypothesis of an adiabatic deceleration of the remnant adopted throughout this work.

Through radiation, the population cools unequally and therefore does not maintain a power-law distribution as in Eq. E.26. Denote by $\mathcal{N}_{t'}$ the electron population distribution at time t'. $\mathcal{N}_{t'}$ is linked to $\mathcal{N}_0$ by:

$$\mathcal{N}_{t'}(\gamma') = \mathcal{N}_0(\gamma'_0(\gamma')) \frac{{\rm d}\gamma'_0}{{\rm d}\gamma'} \tag{E.35}$$

Where we have introduced the function $\gamma'_0(\gamma')$ which gives the initial Lorentz factor of an electron that has cooled to γ' at time t'.

The average electron population distribution over $t'_{\rm dyn}$ is defined by:

$$\langle \mathcal{N} \rangle_{t_{\rm dyn}} (\gamma'_e) = \frac{1}{t_{\rm dyn}} \int_{t'=0}^{t'_{\rm dyn}} \mathcal{N}_{t'}(\gamma'_e) \tag{E.36}$$

After a very cumbersome calculation, Eq. E.35 together with the evolution equations Eqs. E.11 and E.13 allow to derive the following expression in the slow cooling regime:

$$\langle \mathcal{N} \rangle_{t_{\rm dyn}} (\gamma'_e) = \begin{cases} 0 & \text{for} \quad \gamma'_e < \gamma'_m \\ \frac{1}{\gamma'_m} \left(\frac{\gamma'_e}{\gamma'_m}\right)^{-p} & \text{for} \quad \gamma'_m < \gamma'_e < \gamma'_c \\ \frac{1}{\gamma'_m} \frac{\gamma'_c}{\gamma'_m} \left(\frac{\gamma'_e}{\gamma'_m}\right)^{-p-1} & \text{for} \quad \gamma'_c < \gamma'_e \end{cases} \tag{E.37}$$

and in fast cooling:

$$\langle \mathcal{N} \rangle_{t_{\rm dyn}} (\gamma'_e) = \begin{cases} 0 & \text{for} \quad \gamma'_e < \gamma'_c \\ \frac{1}{\gamma'_c} \left(\frac{\gamma'_e}{\gamma'_m} \right)^{-2} & \text{for} \quad \gamma'_c < \gamma'_e < \gamma'_m \\ \frac{1}{\gamma'_m} \frac{\gamma'_c}{\gamma'_m} \left(\frac{\gamma'_e}{\gamma'_m} \right)^{-p-1} & \text{for} \quad \gamma'_m < \gamma'_e \end{cases} \tag{E.38}$$

E.3 Synchrotron characteristic frequencies in a decelerating remnant

We will now put this synchrotron radiation in the context of the forward shock of a decelerating remnant. We introduce the *redistribution parameters* ϵ_e and ϵ_B which prescribe the fractions of shocked material internal energy held in the accelerated electron population and in the entangled magnetic field. The shock-frame internal energy density is (see Appendix D):

$$U'_{\rm int} = n'_s m_P c^2 (\Gamma - 1) \tag{E.39}$$

The magnetic field magnitude is defined as:

$$U'_{\rm mag} = \frac{B'^2}{8\pi} = \epsilon_B U'_{\rm int} \tag{E.40}$$

and the electron population's energy as:

$$U'_e = \epsilon_e U'_{\rm int} \tag{E.41}$$

It is through these equations that the forward shock's dynamics prescribe the emitted radiation. ϵ_e and ϵ_B parametrize the microphysical conditions in the shock and are usually inferred from observations or predicted from plasma theory and simulations. In general, they depend on the shock's Lorentz factor.

Using Eq. E.7 and E.26, we find:

$$\gamma'_m = \epsilon_e \frac{m_P}{m_e} \frac{p-2}{p-1} (\Gamma - 1) \tag{E.42}$$

and

$$\gamma'_c = \frac{3 m_e}{4 m_P c \sigma_T t'_{\rm dyn}} \frac{1}{\epsilon_B n'_s (\Gamma - 1)} \tag{E.43}$$

which allow to calculate the synchrotron emission spectrum from the shocked material at any point in the remnant's evolution.

E.4 Self-absorption of synchrotron radiation

We consider the inverse process of synchrotron radiation: self-absorption. This process will be a correction to the synchrotron spectrum given by $S'_{\rm tot}(\nu')$. To treat this problem we shall introduce the spectral energy density of the photon field $u'(\nu, t)$ in the shocked region, in erg/Hz/cm^3. In the shocked region, photon are produced by the synchrotron process with a rate $\dot{u}'(\nu') \mid_{\rm syn} = n'_s S'_{\rm tot}(\nu')$, and absorbed by the electrons on a timescale $t'_a(\nu')$, such that we have:

$$\frac{\partial u'(\nu, t')}{\partial t'} = \dot{u}'(\nu') \mid_{\rm syn} - \frac{u'(\nu', t')}{t'_a(\nu')} \tag{E.44}$$

Assuming constant $\dot{u}'(\nu') \mid_{\rm syn}$ and $t'_a(\nu')$, the formal solution to Eq. E.44 is:

$$u'(\nu', t') = \dot{u}'(\nu') \mid_{\rm syn} t'_a(\nu') \left(1 - e^{-t'/t'_a(\nu)} \right) \tag{E.45}$$

$$= \dot{u}'(\nu') \mid_{\rm syn} t' \times \frac{t'_a(\nu')}{t'} \left(1 - e^{-t'/t'_a(\nu')} \right) \tag{E.46}$$

This is:

$$u'(\nu', t') = u'(\nu', t') \mid_{\rm syn} \times \begin{cases} 1 & \text{for} \quad t' \ll t'_a(\nu') \\ t'_a(\nu')/t' & \text{for} \quad t'_a(\nu') \ll t' \end{cases} \tag{E.47}$$

where we have denoted by $u'(\nu', t') \mid_{\rm syn}$ the solution to Eq. E.44 in the absence of self-absorption.

Therefore, self-absorption appears as a correction to the average synchrotron emission spectrum, with $S'_{\rm tot,SA}(\nu') = S'_{\rm tot}(\nu')$ at frequencies such that $t'_a(\nu) \ll t'_{\rm dyn}$, and $S'_{\rm tot,SA} = \frac{t'_a(\nu)}{t'_{\rm dyn}} \times S'_{\rm tot}(\nu')$ for $t'_{\rm dyn} \ll t'_a(\nu')$.

Thus, self-absorption introduces an additional break in the emission spectrum, at the self-absorption frequency ν'_a defined by $t'_a(\nu_a) = t'_{\rm dyn}$.

The self-absorption timescale at frequency ν' is given by $ct'_a(\nu')n'_s\sigma_{\rm SA,pop}(\nu') = 1$, where $\sigma_{\rm SA,pop}$ denotes the cross section for absorption of a frequency ν' photon by all the electrons in the population:

$$\sigma_{\rm SA,pop} = \int_{\gamma'_m}^{\infty} {\rm d}\gamma'_e \mathcal{N}_0(\gamma'_e)\sigma_{\rm SA}(\gamma'_e,\nu') \tag{E.48}$$

and $\sigma_{\rm SA}(\gamma_e,\nu)$ denotes the cross section for absorption by a single electron of Lorentz factor γ_e. Self-absorption affects mostly photons at low frequency, radiated by low-energy electrons which are not affected by the synchrotron cooling. Therefore, we can use the initial electron energy distribution $\mathcal{N}_0$ in the calculation of $\sigma_{\rm SA,pop}$.

As an inverse process, the self-absorption cross section is linked to the single-electron emission spectrum $\langle s'\rangle_{t_{\rm dyn}}(\nu',\gamma'_e)$ through (Rybicki & Lightman, 1979, , Eq. 6.50):

$$\sigma_{\rm SA}(\gamma'_e,\nu') = -\frac{1}{8\pi m_e \nu'^2}\frac{\langle s'\rangle_{t_{\rm dyn}}(\nu',\gamma'_e)}{\mathcal{N}_0(\gamma'_e)}\gamma'_e\frac{\partial}{\partial\gamma'}\left(\frac{\mathcal{N}_0}{\gamma^2}\right) \tag{E.49}$$

In practice, we determine the self-absorption timescale at the frequency of interest with Eq. E.25, E.26 and E.49, and deduce the absorbed spectrum as:

$$S'_{\rm tot,SA}(\nu') = S'_{\rm tot}(\nu')\times\min(1, t'_a(\nu')/t'_{\rm dyn}) \tag{E.50}$$

Because of the broken-power-law nature of $\langle s'\rangle_{t_{\rm dyn}}$, $S'_{\rm tot}$ and $\mathcal{N}_0$, the $t'_a(\nu)/t'_{\rm dyn}$ factor is also readily expressed as a broken power law.

E.5 Inverse Compton scattering

The synchrotron photons can be up-scattered by the relativistic electrons in the shocked medium. The cross section for this inverse Compton scattering is the Thomson cross section σ_T as long as the photon's energy is $h\nu \ll m_e c^2$ in the electron's rest frame. Scattering of higher-energy photons is suppressed by the Klein-Nishina effect. The Klein-Nishina frequency can be estimated as:

$$h\nu'_{KN} = m_e c^2\langle\gamma'_e\rangle \tag{E.51}$$

which is larger than 10 MeV for typical cases of off-axis afterglows studied in this work. Therefore, we shall consider that all the synchrotron photons are in the Thomson regime.

For ultra-relativistic electrons, the average fractional energy gain for a single up-scattering of a photon on an electron is (Rybicki & Lightman, 1979, , Eq. 7.37):

$$\frac{\Delta\nu'}{\nu'} = \frac{4}{3}\gamma'^2_e \tag{E.52}$$

The optical depth to inverse Compton scattering in the shocked region is:

$$\tau'_{IC} = n'_s\sigma_T\Delta R' \tag{E.53}$$

Therefore, the average fractional energy gain for a photon escaping from the shocked region, the so-called Compton parameter is:

$$Y = \frac{4}{3}n'_s\sigma_T\langle\gamma'_e\rangle^2_{t_{\rm dyn}}\Delta R' \tag{E.54}$$

which is the same for all photons in the Thomson regime. The Y parameter can be used to determine whether a photon will significantly change its energy in traversing the shocked medium. For the case of the inverse Compton correction, scatterings on high-energy electrons are most important, and we thus use the averaged electron distribution $\langle\mathcal{N}\rangle_{t_{\rm dyn}}$.

Inverse Compton scattering has three consequences on the emission spectrum: (i) a depletion of the synchrotron component, which is decreased by a factor of $(1+Y)$, (ii) the appearance of a second spectral component of up-scattered photons, identical to the synchrotron component but boosted by $4\langle\gamma'^2_e\rangle/3$ and with a total relative power of $Y/(1+Y)$, (iii) the faster cooling of electrons due to the losses in the scatterings. We will restrain this paragraph to the slow cooling regime.

The faster cooling of the electrons changes γ'_c such that:

$$\gamma'_{c,\rm IC} = \frac{\gamma'_{c,\rm syn}}{1+Y} \tag{E.55}$$

We use the estimate $\Delta R' \sim r'/\Gamma^2 \sim ct'_{\rm ex} \sim ct'_{\rm dyn}$, so that:

$$Y = \frac{4}{3} \langle \gamma_e'^2 \rangle_{t_{\rm dyn}} n'_s c\sigma_T t'_{\rm dyn} \tag{E.56}$$

$$= \frac{4}{3} \langle \gamma_e'^2 \rangle_{t_{\rm dyn}} \frac{6\pi m_e c^2}{8\pi \epsilon_B (\Gamma - 1) m_P c^2 \gamma'_{c,\rm syn}} \tag{E.57}$$

$$= \langle \gamma_e'^2 \rangle_{t_{\rm dyn}} \frac{\epsilon_e}{\epsilon_B} \frac{p-2}{p-1} \frac{1}{\gamma'_m \gamma'_{c,\rm syn}} \tag{E.58}$$

where we used Eqs. E.42 and E.43.

Finally, one has:

$$\langle \gamma_e'^2 \rangle_{t_{\rm dyn}} = \int_{\gamma'_e = \gamma'_m}^{\infty} {\rm d}\gamma'_e \, \langle \mathcal{N} \rangle_{t_{\rm dyn}} (\gamma'_e) \gamma_e'^2 \tag{E.59}$$

$$= \frac{p-1}{(p-2)(3-p)} \gamma_m'^{p-1} \gamma_{c,\rm syn}'^{3-p} \tag{E.60}$$

such that Y is the solution to the following equation:

$$Y(1+Y)^{3-p} = \frac{\epsilon_e}{\epsilon_B} \frac{1}{3-p} \left(\frac{\gamma'_m}{\gamma'_{c,\rm syn}} \right)^{p-2} \tag{E.61}$$

In practice, we solve Eq. E.61 numerically to determine the value of Y and apply the corresponding inverse Compton correction. We finally note that the up-scattered photons are above the Klein-Nishina limit, according to the above estimate. Therefore, second inverse-Compton scatterings are negligible.

Appendix F

Bayesian population parameter inference under selection effects

We seek to derive the statistical framework for the inference of population parameters under two effects: measurement uncertainty on individual sources and selection effects. We follow Mandel et al. (2019).

We collectively denote by λ the *population parameters*, which parametrize the probability distributions of the *individual properties* of the sources, denoted by θ. These are such that $p_{\rm pop}(\theta|\lambda)$ is the probability that a random source has the properties θ. We seek to infer λ with a sample of independent uncertain measurements of the θ's of a sample of sources, given that we only detected the sources that verify a certain known detection criterion depending on the true sources' θ's.

Without measurement uncertainties, we can assume that we directly measure the properties themselves, and the selection effects apply to these. In this case, we can introduce a *detection probability* $p_{\rm det}$ of the parameters, indicating the probability that a source with properties θ will be detected.

The probability to detect a source with properties θ_i is therefore:

$$p(\theta_i|\lambda) \propto p_{\rm pop}(\theta_i|\lambda)p_{\rm det}(\theta_i) \tag{F.1}$$

Normalizing Eq. F.1 leads to:

$$p(\theta_i|\lambda) = \frac{p_{\rm pop}(\theta_i|\lambda)p_{\rm det}(\theta_i)}{\int {\rm d}\theta' p_{\rm pop}(\theta'|\lambda)p_{\rm det}(\theta')} \tag{F.2}$$

In practice, we do not measure θ_i directly, but data d_i with a certain associated likelihood to be measured given θ_i, denoted by $p(d_i|\theta_i)$. The selection effects apply to the data, which we encompass in a strict detection threshold $d_{\rm lim}$. In this case, we have:

$$p_{\rm det}(\theta_i) = \int_{d|d>d_{\rm lim}} {\rm d}d p(d|\theta_i) \tag{F.3}$$

We can now link the observed data d_i to the parameter of the underlying population by introducing all the source property values that could have produced the data:

$$p(d_i|\lambda) \propto \int {\rm d}\theta p(d|\theta)p_{\rm pop}(\theta|\lambda) \tag{F.4}$$

Once again, the normalization of Eq. F.4 is the selection function:

$$\beta_S(\lambda) = \int {\rm d}d p(d|\lambda) \tag{F.5}$$

Since the data is observed, the integration domain for Eq. F.5 is all data such that $d > d_{\rm lim}$:

$$\beta_S(\lambda) = \int_{d|d>d_{\rm lim}} {\rm d}d p(d|\lambda) \tag{F.6}$$

$$= \int_{d|d>d_{\rm lim}} {\rm d}d \int {\rm d}\theta p(d|\theta)p_{\rm pop}(\theta|\lambda) \tag{F.7}$$

$$= \int {\rm d}\theta \left[\int_{d|d>d_{\rm lim}} {\rm d}d p(d|\theta)\right] p_{\rm pop}(\theta|\lambda) \tag{F.8}$$

$$= \int {\rm d}\theta p_{\rm det}(\theta)p_{\rm pop}(\theta|\lambda) \tag{F.9}$$

where we used Eq. F.3 in the last equality.

Thus, in the presence of both measurement uncertainty and selection effects, the likelihood for data d_i is:

$$p(d_i|\lambda) = \frac{1}{\beta_S(\lambda)} \int \mathrm{d}\theta p(d_i|\theta) p_{\mathrm{pop}}(\theta|\lambda) \tag{F.10}$$

$$= \frac{1}{\beta_S(\lambda)} p_0(d_i|\lambda) \tag{F.11}$$

where we defined $p_0(d_i|\lambda)$ as the likelihood of the data in absence of selection effects, with $p_{\mathrm{det}} = 1$.

We can know apply Bayes' theorem to write the posterior on λ after observing d_i with prior information $\pi(\lambda)$:

$$p(\lambda|d_i) = \frac{p(d|\lambda)}{p(d_i)} \pi(\lambda) \tag{F.12}$$

$$= \frac{1}{\beta_S(\lambda)} \frac{p_0(d_i|\lambda)}{p(d_i)} \pi(\lambda) \tag{F.13}$$

Appendix G

Isotropic-equivalent energy from a flashing structured relativistic jet

Here, we seek to establish the relation between the intrinsic structure of a structured relativistic jet and its apparent structure (Sec. 8.4). We consider a structured jet that flashes, where the material at latitude θ has Lorentz factor $\Gamma(\theta)$ at the moment of the flash. The energy emitted per unit solid angle measured in the central source frame is denoted by $\epsilon(\theta)$. We place the observer at a viewing angle θ_v to the structure and wish to calculate the isotropic-equivalent detected energy from the flash.

This situation is very similar to that of Appendix C, by defining a multitude of elementary cones to cover the structure. We are interested in the total isotropic-equivalent observed energy. We can therefore define arbitrary emission time and radius for the flashing.

We start from Eq. C.9, replacing E_0/Ω_p by ϵ, so that the spectral luminosity is:

$$L_{\rm iso}(t_{\rm obs},\nu_{\rm obs}) = \int {\rm d}\Omega \mathcal{D}^2 s' \left(\frac{\nu_{\rm obs}}{\mathcal{D}}\right) \frac{\epsilon(\theta)}{\Gamma(\theta)} \delta\left(T_E - \frac{R_E\cos\chi}{c} - t_{\rm obs}\right) \tag{G.1}$$

where the spherical angles are θ, ϕ with differential solid angle element ${\rm d}\Omega = \sin\theta {\rm d}\theta {\rm d}\phi$. χ still denotes the angle between the line of sight and the material at θ, ϕ. The integration is, this time, over the entire structure.

Integrating over frequency trivially leads to:

$$\int {\rm d}\nu_{\rm obs} L_{\rm iso}(t_{\rm obs},\nu_{\rm obs}) = \int {\rm d}\Omega \mathcal{D}^3 \frac{\epsilon(\theta)}{\Gamma(\theta)} \delta\left(T_E - \frac{R_E\cos\chi}{c} - t_{\rm obs}\right) \tag{G.2}$$

and finally, integrating over time to obtain the total isotropic-equivalent energy removes the δ-function as the photons from any latitude will eventually reach the observer:

$$E_{\rm iso}(\theta_v) = \int {\rm d}\Omega \mathcal{D}^3 \frac{\epsilon(\theta)}{\Gamma(\theta)} \tag{G.3}$$

Writing the Doppler factor in terms of the angular distance to the line of sight χ, we finally have:

$$E_{\rm iso}(\theta_v) = \int {\rm d}\Omega \frac{\epsilon(\theta)}{\Gamma_0(\theta)^4 \left(1-\beta_0(\theta)\cos\chi\right)^3} \tag{G.4}$$

It is interesting to develop Eq. G.4 in various regimes of viewing angle. We denote the core's opening angle by θ_j and its Lorentz factor at flash by Γ_j. The natural variable to describe the beaming of the core is $q = |\delta\theta|\Gamma_j$, where $\delta\theta = \theta_v - \theta_j$ denotes the angular distance form the line of sight to the edge of the core.

When observing the structure from a angle θ_v, the radiation is dominated either by the material on the line of sight or the core. The line-of-sight material's contribution is simply $E_{\rm iso}(\theta_v)_{|\rm LOS} = 4\pi\epsilon(\theta_v)$. The core's contribution $E_{\rm iso}(\theta_v)_{|\rm core}$ will be obtained by integrating Eq. G.4 over the core only, and being careful to determine which region of the core dominates the emission, i.e., which region is the most boosted to the observer. A similar reasoning was done in Appendix C.2 and we repeat it here:

- For an aligned line of sight ($\theta_v < \theta_j$), it is clear that the core's contribution is simply $4\pi\epsilon(0)$.
- For a slightly misaligned line of sight ($\theta_j + 1/\Gamma_j \ll \theta_v \ll 2\theta_j$), the flux is dominated by regions within the emitting region that are the most boosted, with $\delta\theta < \theta < 2\delta\theta$. In addition, the emitting region is limited to a transverse angular size

of $\Delta\phi \sim \pi$, as the core occupies nearly a half plane in the observer's field. Thus $\mathrm{d}\Omega \sim 3\delta\theta^2\Delta\phi/2$, the Doppler factor is $\mathcal{D} = \frac{1}{\Gamma_j(1-\beta\cos\delta\theta)} \sim 2\Gamma_j/(1+\Gamma_j^2\delta\theta^2) = 2\Gamma_j/(1+q^2)$ and thus:

$$E_{\rm iso}(\theta_v)_{|\rm core} = \frac{3\delta\theta^2\Delta\phi}{2}\left(\frac{2\Gamma_j}{1+q^2}\right)^3\frac{\epsilon(0)}{\Gamma_j} \tag{G.5}$$

$$= 4\pi\frac{q^2}{(1+q^2)^3}\epsilon(0) \tag{G.6}$$

$$\sim 4\pi\frac{1}{(1+q^2)^2}\epsilon(0) \tag{G.7}$$

where we used $q \gg 1$.

- Significantly misaligned ($\delta\theta > \theta_j$): The whole core has nearly the same Doppler factor, that of the near edge. Thus $\mathrm{d}\Omega \sim \theta_j^2/2$, $\mathcal{D} \sim 2\Gamma_j/(1+q^2)$ and thus:

$$E_{\rm iso}(\theta_v)_{|\rm core} = 2\pi\frac{\theta_j^2}{2}\left(\frac{2\Gamma_j}{1+q^2}\right)^3\frac{\epsilon(0)}{\Gamma_j} \tag{G.8}$$

$$= 4\pi(\theta_j\Gamma_j)^2\frac{1}{(1+q^2)^3}\epsilon(0) \tag{G.9}$$

$$\sim 4\pi(\theta_j\Gamma_j)^2\frac{1}{q^2}\frac{1}{(1+q^2)^2}\epsilon(0) \tag{G.10}$$

where we again used $q \gg 1$. This last form of $E_{\rm iso}(\theta_v)_{|\rm core}$ shows that we lose a factor of $(\delta\theta/\theta_j)^2$ from slightly to very misaligned lines of sight.

Defining the observed energy as the largest of the line-of-sight or the core material's contribution, we conclude that:

$$E_{\rm iso}(\theta_v) = 4\pi \times \begin{cases} \epsilon(0) & \text{if} \quad \theta_v \le \theta_j \\ \max\left[\epsilon(\theta_v), \frac{1}{(1+q^2)^2}\epsilon(0)\right] & \text{if} \quad \theta_j \le \theta_v \le 2\theta_j \\ \max\left[\epsilon(\theta_v), \frac{1}{q^2(1+q^2)^2}(\Gamma_j\theta_j)^2\epsilon(0)\right] & \text{if} \quad 2\theta_j \le \theta_v \end{cases} \tag{G.11}$$

Appendix H

Jet break and jet afterglow peak times

In this appendix we seek to derive the peak time for the afterglow of a misaligned bare core jet. As it is the case for all the afterglow studies of Part II, we will consider a uniform external medium. We will follow a simple analytical reasoning.

For an aligned observer, the post-jet-break afterglow light curve behaves as $t_{\rm obs}^{-3p/4}$ in the case of no jet spreading and $t_{\rm obs}^{-p}$ in the case of sound speed expansion of the core jet (Rhoads, 1999). In either case, if we neglect the total jet expansion between launching and jet break, the jet break time t_b and flux at the jet break F_b will be the same, thus:

$$F(t) = \begin{cases} F_b \left(\frac{t}{t_b}\right)^{-3p/4} & \text{no expansion} \\ F_b \left(\frac{t}{t_b}\right)^{-p} & \text{expansion} \end{cases} \tag{H.1}$$

By definition of the jet break time, the Lorentz factor evolves with observer time as:

$$\Gamma(t) = \begin{cases} \theta_j^{-1} \left(\frac{t}{t_b}\right)^{-3/8} & \text{no expansion} \\ \theta_j^{-1} \left(\frac{t}{t_b}\right)^{1/2} & \text{expansion} \end{cases} \tag{H.2}$$

where we used Eq. D.25 with $\epsilon = 3/2$ as we proved for a non-expanding jet. For a sound-speed expanding jet, there is no defined power-law $\Gamma(r)$ expression (it is exponential), however one still finds $\Gamma(t) \propto t^{-1/2}$ (Rhoads, 1999).

Now by the core beaming condition for off-axis afterglow peak at time t_p, we have $\Gamma(t_p)\theta_v \sim 1$, leading to conclude that:

$$\begin{cases} t_{p,\rm ex} = t_b \left(\frac{\theta_v}{\theta_j}\right)^{8/3} & \text{no expansion} \\ t_{p,\text{no ex}} = t_b \left(\frac{\theta_v}{\theta_j}\right)^{2} & \text{expansion} \end{cases} \tag{H.3}$$

We have $t_{p,\rm ex}/t_{p,\text{no ex}} = (\theta_j/\theta_v)^{2/3} < 1$, as suggested by the fact that a spreading core reveals earlier.

Using Eq. D.26 for the jet break time, we finally obtain the following expressions:

$$t_{p,\text{no ex}} = 4.9 \left(\frac{E_{52}}{n_{-3}}\right)^{1/3} \theta_{v,-1}^{8/3} \ \text{ days} \tag{H.4}$$

and

$$t_{p,\rm ex} = 5 \left(\frac{E_{52}}{n_{-3}}\right)^{1/3} \theta_{j,-1}^{2/3}\, \theta_{v,-1}^{2} \ \text{days} \tag{H.5}$$

which are Eqs. 4.15 and 4.16.

Finally, using Eq. H.1, we can determine the influence of jet spreading on the afterglow peak flux:

$$\frac{F_{p,\text{no ex}}}{F_{p,\rm ex}} = (t_{p,\text{no ex}})^{-3p/4} (t_{p,\rm ex})^{p} t_b^{3p/4-p} \tag{H.6}$$

$$= \theta_v^{2p} \theta_j^{2p/3} \theta_v^{-2p} \theta_j^{2p/3} \tag{H.7}$$

$$= 1 \tag{H.8}$$

Therefore, expansion does not affect the peak flux significantly, justifying the use of Eq. 4.13 throughout.

Appendix I

Constraints on $n_{\rm ext}$ and θ_v deduced from observing the peak of a misaligned afterglow

In this Appendix, we seek to establish Eq. 7.2, which is the basis of our multi-messenger constraints in Chap. 7 on the circum-merger density $n_{\rm ext}$ and the viewing angle θ_v to a misaligned BNS merger. We reproduce the equation below, without the normalization factors:

$$F_p t_p^{-3} D_L^2 \times \begin{cases} (\kappa\eta)^2 & \text{no ex.} \\ 1 & \text{ex.} \end{cases} = \theta_v^{-6-2p} n_{\rm ext}^{\frac{p+5}{4}} \epsilon_e^{p-1} \epsilon_B^{\frac{p+1}{4}} \tag{I.1}$$

where F_p and t_p are the afterglow peak flux and time of peak and $\eta = (t_2 - t_1)/t_2$ is the shape factor of the afterglow peak. The times t_1 and t_2 mark respectively the end of the afterglow's increasing phase and the start the decreasing phase. These times must be estimated from the light curve after fitting a functional form, see details below. With these definitions, $\Delta t = t_2 - t_1$ is the duration of the transition from one phase to the other. Finally, κ is defined such that the forward shock's deceleration law is $\Gamma \propto t_{\rm obs}^{-\kappa}$. It depends on whether the jet expands at sound speed ('ex.', $\alpha = 1/2$) or does not expand ('no ex.', $\alpha = 3/8$); We proved these relations in Appendix D.

Since the afterglow light curve is not a mathematical broken power law, the times t_1 and t_2 are ill-defined. In practice, a smoothly broken power law is fit to the light curve, such as the following (Mooley et al., 2018c):

$$F(t_{\rm obs}) = 2^{1/s} F_p \left(\left(\frac{t_{\rm obs}}{t_p} \right)^{-s\alpha_+} + \left(\frac{t_{\rm obs}}{t_p} \right)^{-s\alpha_-} \right)^{-1/s} \tag{I.2}$$

with time slopes α_+ and α_- for the increasing and decreasing phases and smoothness parameter s.

Once this model fit, one defines t_1 and t_2 such that the factors $(t_{\rm obs}/t_p)^{-s\alpha_i}$ dominate one another, by a fiducial factor of, e.g., 20 (as in Mooley et al. 2018c); The final value of η depends little on this factor.

From a physical point of view, the increasing phase corresponds to the exploration of the jet from the line-of-sight material up to the core jet, and the decreasing phases starts once the whole core has been revealed. Therefore, we have the beaming conditions:

$$(\theta_v - \theta_j)\Gamma(t_1) = 1 \tag{I.3}$$

and

$$(\theta_v + \theta_j)\Gamma(t_2) = 1 \tag{I.4}$$

Finally, using the deceleration law, one obtains:

$$\eta = \frac{t_2 - t_1}{t_2} \tag{I.5}$$

$$= 1 - \left(\frac{\theta_v - \theta_j}{\theta_v + \theta_j} \right)^{1/\kappa} \tag{I.6}$$

$$\sim \frac{2}{\kappa} \frac{\theta_j}{\theta_v} \tag{I.7}$$

which is valid for $\theta_j \ll \theta_v$.

Eq. I.7 allows to link the observed η with the two angle parameters.

Finally, by using Eq. 4.12 for the afterglow peak flux and Eqs. H.4 and H.5 for the peak time in both jet expansion regimes, one can eliminate the jet energy E and opening angle θ_j to obtain Eq. I.1.

To choose the relevant expansion dynamics, one can observe the late-time behavior of the afterglow, as mentioned in Sec. 7.3.1. In any case, the constraints obtained in the non- and fully-expanding regimes can be seen as confidence bounds on the parameters.