OM SHARAN SALAFIA

SHORT GAMMA-RAY BURSTS AS ELECTROMAGNETIC COUNTERPARTS OF COMPACT BINARY MERGERS

The revolution just started

Dottorato di Ricerca in Fisica e Astronomia
Dipartimento di Fisica “G. Occhialini”
Facoltà di Scienze Matematiche, Fisiche e Naturali
Università degli Studi di Milano-Bicocca

Fall 2017 – v 1.0
Om Sharan Salafia: Short gamma-ray bursts as electromagnetic counterparts of compact binary mergers, The revolution just started, © Fall 2017

SUPERVISORS:
Gabriele Ghisellini
Monica Colpi
Giancarlo Ghirlanda

LOCATION:
Università degli Studi di Milano-Bicocca, piazza della Scienza 3, 20124 Milano, Italy

TIME FRAME:
Fall 2017
Dedico questo lavoro a Ilaria, l’unica persona capace di capirmi veramente, di sostenermi con pazienza, e di far emergere il meglio di me.
Gamma-ray bursts (GRBs) are brief flashes of photons that trigger current space-based hard X-ray and gamma-ray detectors every two or three days. During a time ranging from less than one to several thousand seconds, a highly variable photon flux with an unpredictable time structure is recorded by the detector. Fifty years have flown since the first observation of this kind, during which a long series of technological and theoretical breakthroughs paved the way for the current, widely-accepted paradigm that relates these flashes to accretion of matter on a newborn stellar-mass black hole or neutron star. Two are the natural birthplaces of such relativistic beasts: the collapse of a massive star and the coalescence of two compact objects. The latter, perhaps the most intriguing of the two, was the first to be proposed as a candidate progenitor of GRBs, but in 1998 the association of GRB 980425 with supernova 1998bw provided compelling evidence for the former. Nevertheless, no supernova has been associated so far – in some cases down to very stringent limits – to members of a particular subclass of these events, known as short gamma-ray bursts (SGRBs).

Several pieces of evidence support the idea that the progenitor of SGRBs is indeed the coalescence of two neutron stars, or of a black hole and a neutron star. If this is true, then SGRBs are also intimately related to gravitational waves (GW). The advanced network of ground-based GW detectors – which at present consists of the two Advanced LIGO interferometers in the USA and of Advanced Virgo in Italy – is especially sensitive in the frequency range of GW produced by the inspiral and merger of a stellar mass compact object binary, so that we are right in the position to start testing the SGRB–GW connection.

In August of this year, the first observation of GW from a neutron star binary coalescence, followed by the first observation of a kilonova – the UV/Optical/Infrared emission from the expanding material ejected during the merger and post-merger phases of the coalescence, powered by nuclear decay of unstable nuclei synthesized by the r-process – and an associated SGRB-like transient marked the start of a revolution, whose effect on our understanding of these subjects still needs to be completely unfolded. For this reason, in this thesis I do not to draw firm conclusions about these observations, but rather I discuss some possible interpretations and implications, leaving many questions open to future investigation.

In the first part of this thesis I show the results of a detailed modeling of the SGRB population, with a careful handling of selection effects. The model satisfies for the first time all available observational constraints. Despite the model being agnostic about the progenitor, the shape of the resulting redshift distribution reproduces well that expected in the compact binary progenitor scenario. The predicted local density of SGRBs, though, is very low, and the implied rate of joint detections of SGRBs and GWs from compact binary mergers is less than one event per decade. The low rate is in part due to the need for the SGRB jet to be aligned with the line of sight in order for the gamma-ray emission to be detected. Somewhat better prospects are found for the joint detection of GW and an SGRB orphan afterglow.
These results triggered my interest in maximizing the effectiveness of the electromagnetic follow-up of GW detections. In the subsequent chapter, I describe a method to optimize the electromagnetic follow-up of a GW event which makes use of the information about compact binary parameters that can be extracted from the GW signal through parameter estimation. The compact binary parameter posterior distributions – and their correlations – are used to build a family of possible SGRB afterglow and kilonova light curves. These are the basis to construct a time-dependent sky-position-conditional probability of detection that guides the optimization of the follow-up strategy.

The second part of the thesis is devoted to GRB170817A, the SGRB-like burst detected by Fermi/GBM and INTEGRAL/SPI-ACS in temporal and spatial association with the gravitational wave event GW170817, which has been interpreted as the spacetime perturbation due to the inspiral of a double neutron star binary. Despite the electromagnetic burst appears as a rather ordinary SGRB from a purely observational point of view, the very low luminosity and energy implied by the small distance are absolutely unprecedented. The community is still in the process of finding a consensus on the interpretation of this event. Most proposed scenarios involve the presence of a jet misaligned with the line of sight, whose emission would have been that of an ordinary SGRB if seen on-axis. The extremely low energy and luminosity of the burst (with respect to any previous SGRB for which a redshift has been measured) is either interpreted as emission from the border of the jet, or from a slower and less energetic sheath surrounding the jet, or from a “cocoon” of plasma heated by the jet while it had been excavating its way through the merger ejecta. Since it is unclear whether a jet should be always produced after this kind of merger, and not even if it can always successfully break out of the region polluted by the merger ejecta, I tried to imagine a jet-less scenario able to explain the properties of GRB170817A – knowing that the absence of a jet would affect the impact that this observation has on the rate of SGRBs. In the final part of this thesis I describe such a scenario, in which the SGRB is produced by an isotropic fireball powered by reconnection of a very strong magnetic field surrounding the merging binary, produced by amplification of the neutron star magnetic field due to magnetohydrodynamic turbulence. I set up a simple physical model to predict the properties of such emission, and I show that there is a set of parameters for which the model fits the properties of both the prompt gamma-ray emission and the late X-ray and radio data. The model predicts the future evolution of the radio and X-ray light curve, which will be soon put on test by new observations. I finally discuss some possible difficulties related to the very low electron fraction needed in the fireball according to the fit parameters.
Most of the ideas and figures in this thesis appeared previously in articles I got published on peer-reviewed journals during my three year PhD course. Nevertheless, this work is not a simple collection of previously published papers, but rather elaborates on selected works in order to provide a coherent picture of the SGRB–GW connection and of my personal view about the interpretation of GRB170817A. Here is a complete list of my publications as of February 13, 2018:


## CONTENTS

1 **INTRODUCTION** 1

I **SHORT GAMMA-RAY BURSTS** 3

2 **GAMMA-RAY BURST ASTRONOMY** 5

2.1 The discovery 5
2.2 The prompt emission 5
2.3 The afterglow 10
2.4 Presence of a jet 12
2.5 Progenitors 14

3 **TRYING TO GET A GRASP OF THE SHORT GRB POPULATION** 15

3.1 Briefing 15
3.2 Introduction 15
3.3 Selecting a good sample 17
3.4 A sensible parametrization of the \(\phi(L)\) and \(\Psi(z)\) 19
3.5 From population properties to observables 22
3.6 A too steep luminosity function is at odds with the observer-frame constraints 23
3.7 A Monte Carlo simulation of the SGRB population 24
3.8 Looking for the best fit parameters 26
3.9 The results 28
3.10 The local SGRB rate 32
3.11 Summing up 35

4 **WAW: A GW-INFORMED EM FOLLOW-UP STRATEGY** 37

4.1 The EM follow-up problem 37
4.2 Where and when to look 41
4.3 Detectability maps 49
4.4 Test of the IDW method 50
4.5 Test example: injection 56
4.6 Discussion 63
4.7 Conclusions 67

II **LIGHT AND GRAVITATIONAL WAVES FROM A SINGLE SOURCE: GW170817** 69

5 **A REVOLUTION STARTED ON THE 17TH OF AUGUST 2017** 71

5.1 The short gamma-ray burst GRB170817A 71
5.2 Observational properties of GRB170817A 72
5.3 Some possible interpretations of GRB170817A 73
5.4 Impact on the rate of SGRBs 73

6 **PROMPT EMISSION FROM AN OFF-AXIS JET** 75

6.1 Pulses: building blocks of GRB light curves 75
6.2 Pulse overlap and light curve variability 76
6.3 Pulses in the internal shock scenario 79
6.4 Pulse light curves and time dependent spectra in the shell-curvature model 80
6.5 Characteristics of pulses from an off-axis jet 93
6.6 Multi-pulse light curves 96
6.7 The number of off-axis Short GRBs seen by Fermi/GBM 99
6.8 The time delay between GW170817 and GRB170817A 101

7 THE JET-LESS FIREBALL SCENARIO 105
7.1 Set up of the isotropic fireball model 105
7.2 Comparison of the afterglow emission from the three scenarios 108
7.3 Application to GRB170817A and estimation of the model parameters 111
7.4 The physical picture that emerges from the results 113
7.5 Future light curve evolution 115
7.6 Can the fireball retain the low electron fraction up to the transparency radius? 115

8 CONCLUSIONS 117

BIBLIOGRAPHY 119
INTRODUCTION

Our knowledge of the Universe has expanded dramatically during the last century. The new theory of gravity and spacetime, born after a formidable intellectual effort by Albert Einstein and several other visionary minds, has survived several tests and is now mature. Among many other mind-boggling products, this theory predicts the existence of black holes and, when coupled with our understanding of degenerate matter from quantum mechanics, of neutron stars. When two such compact objects orbit each other, the theory predicts that the perturbation of spacetime due to their motion propagates at the speed of light and can produce measurable effects at cosmological distances: these perturbations are known as gravitational waves. During the last few decades, new techniques for solving the Einstein equations both analytically and numerically allowed theorists to build an increasingly refined understanding of these phenomena. Moreover, we are in a golden age for what concerns the development of instruments and facilities for the detection of cosmic signals, which allowed observations to keep the pace with the theoretical advancement.

The existence of compact relativistic stars, with a gravitational binding energy of the order of $10^{53}$ erg or more, is also the basis for the understanding of extremely luminous sources of radiation such as quasars and gamma-ray bursts. No process other than the extraction of gravitational energy from a black hole or a neutron star is able to reach the efficiency required to explain the observations of these kinds of sources.

For these reasons, gamma-ray bursts are intimately related to strong gravity. More specifically, their short duration and extreme luminosity both point to a cataclysmic event where a large amount of matter is accreted in a short time onto a compact object: the most natural setting is thus the birth of a black hole or a neutron star after the collapse of a massive star or the coalescence of two compact objects. Incidentally, these events also produce gravitational waves. The two phenomena are thus inseparable.

In this thesis, I explore this connection from the point of view of gamma-ray bursts, which are the physical phenomena I had the chance to study in greatest detail during my PhD. In particular, I focus on the connection between short gamma-ray bursts and the coalescence of two neutron stars. After a brief introduction to the observational features of gamma-ray bursts in general (Chapter 2), I present the results of a study of the properties of the population of short gamma-ray bursts I realized together with Giancarlo Ghirlanda and other co-authors (Chapter 3). In the chapter that follows, I describe a method I developed to use the information on the neutron star binary that can be extracted from the gravitational wave signal, combined with models of the short gamma-ray burst afterglow and of the kilonova, to optimize the electromagnetic follow-up of a gravitational wave detection of that kind of source. All the subsequent chapters (Chapters 5–7) focus on the first ever association of a gamma-ray burst (GRB170817A) to a detection of gravitational waves (GW170817), which took place just a few months ago, on
the interpretation of the properties of that particular gamma-ray burst, and on its impact on our understanding of the population of short gamma-ray bursts.
Part I

SHORT GAMMA-RAY BURSTS
GAMMA-RAY BURST ASTRONOMY

THE DISCOVERY

The first observation of a Gamma Ray Burst dates back to 1967 (Bonnell and Klebesadel, 1996), but its announcement was made only in 1973 in a letter to ApJ by R. W. Klebesadel (Klebesadel, Strong, and Olson, 1973), due to the initial difficulties in discriminating this and some other similar events from the frequent false triggers of the instruments aboard the Vela satellites - a series of US satellites inspired by a nuclear test ban treaty, which serendipitously allowed for this discovery - and to the need of more data to confirm the cosmic origin of such events. The news ignited the astronomical community, as can be evinced from Figure 1, where the citation history for the original work by Klebesadel and collaborators is shown (the figure was produced by querying the SAO/NASA Astrophysics Data System1).

The Compton Gamma Ray Observatory (CGRO), launched in 1991, collected the first large sample of gamma-ray bursts, providing strong evidence for a uniform sky distribution (see Figure 2) and showing that their cumulative fluence distribution does not follow the $-3/2$ power law expected for a homogeneous population of sources in Euclidean space (Meegan et al., 1992). This was the first compelling evidence against a Galactic origin.

No search for optical counterparts of GRBs had been successful at that time, though, so that no redshift measurement (and thus no intrinsic energy estimate) was available yet. If GRBs were extragalactic, the huge energy release and the fast variability suggested (Paczynski and Rhoads, 1993) that a slowly fading emission (called afterglow) at longer wavelengths (from X-ray down to Radio) should follow the main event. The search for such emission did not produce results until 1997, when the first X-ray afterglow was observed by BeppoSAX 3 following the burst GRB970228 (Costa et al., 1997). Soon later the afterglow of another event, GRB970508, was observed and finally led to the first redshift measurement (Reichart, 1998; Metzger et al., 1997) of a GRB, namely $z = 0.835$. This showed that GRBs are among the most powerful photon-emitters known in the Universe.

THE PROMPT EMISSION

After the discovery of the afterglow, a distinction between the initial gamma-rays and the subsequent longer wavelength counterpart became necessary. Hereafter, I adopt the name “prompt emission” to indicate the initial highly variable, short duration (less than 1000 seconds in the vast majority of cases) hard-X-ray and gamma-ray signal detected by instruments such as Swift/BAT, Fermi/GBM, CGRO/BATSE, INTEGRAL/SPI-ACS and the like. Some example light curves constructed using CGRO/BATSE data are shown in Figure 3.

---

1 http://adsabs.harvard.edu/
Spectrum

The spectrum of the GRB prompt emission is most often evidently non-thermal, featuring one or more smoothly-connected power law segments, sometimes with an exponential cut-off above some photon energy. In many cases, the time-integrated photon spectrum $\frac{dN}{dE}$ can be fit by the empirical “Band function”

$$
\frac{dN}{dE} = A \left\{ \begin{array}{ll}
\left( \frac{E}{100\text{ keV}} \right)^\alpha \exp \left( -\left(2 + \alpha\right) \frac{E}{E_p} \right) & \text{for } E < \frac{\alpha - \beta}{2 + \alpha} E_p \\
\left( \frac{\alpha - \beta}{2 + \alpha} \frac{E_p}{100\text{ keV}} \right)^\alpha \exp \left( \beta - \alpha \right) \left( \frac{E}{100\text{ keV}} \right)^\beta & \text{for } E \geq \frac{\alpha - \beta}{2 + \alpha} E_p 
\end{array} \right.
$$

introduced by Band et al., (1993). The parameter $E_p$ represents the peak of the corresponding $\nu F(\nu)$ spectrum, and the exponents $\alpha$ and $\beta$ are usually called respectively low and high energy spectral index. The amplitude $A$ is in units of photons cm$^{-2}$ s$^{-1}$ keV. An example of a prompt emission spectrum fit with this function is shown in Fig. 4. In the remaining cases, the spectrum is often well fit by a single power law with an exponential cut-off, namely

$$
\frac{dN}{dE} = A \left( \frac{E}{100\text{ keV}} \right)^\alpha \exp \left( -\left(2 + \alpha\right) \frac{E}{E_p} \right)
$$

In a relatively recent work, Nava et al., (2011a) published the results of careful spectral analysis on a sample of 438 Fermi/GBM bursts. In Figure 5 I reproduce some figures from their work, which give an idea of the typical values of the $E_p$, $\alpha$ and $\beta$ parameters found in modeling GRB time-integrated spectra. The parameters of the gaussian fits shown in the figures are reported in Table 1.
Figure 2: Sky distribution of GRBs detected by CGRO/BATSE, from https://gammaray.nsstc.nasa.gov/batse/grb/skymap/.
Figure 3: Light curves of 12 bright gamma-ray bursts detected by CGRO/BATSE. The vertical axes are in units of thousands of detector counts per second. Figure created by Daniel Perley with data from the public BATSE archive (https://gammaray.nsstc.nasa.gov/batse/grb/catalog/).
Figure 4: The datapoints (coloured crosses in the right-hand panel) represent Fermi/GBM observations of a GRB spectrum, corresponding to the photons in the hatched part of the light curve shown in the left-hand panel. The black solid line is a fit with the Band function described in the text. Reproduced from Tierney et al., (2013).

Figure 5: Distributions of spectral parameters in a sample of 438 GRBs detected by Fermi/GBM analyzed by Nava et al., (2011a). Blue histograms refer to Long GRBs (duration longer than 2 s), red histograms refer to Short GRBs, and black histograms refer to the whole sample.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th># of GRBs</th>
<th>Central value</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log($E_p$)</td>
<td>All</td>
<td>318</td>
<td>2.27</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>Short</td>
<td>44</td>
<td>2.69</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>272</td>
<td>2.21</td>
<td>0.36</td>
</tr>
<tr>
<td>α</td>
<td>All</td>
<td>318</td>
<td>−0.86</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Short</td>
<td>44</td>
<td>−0.50</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>274</td>
<td>−0.92</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Table 1: Parameters of the gaussian fits to the distributions of Fermi/GBM burst spectral parameters as analyzed in Nava et al., (2011a).
When I learnt about the existence of gamma-ray bursts, I was amazed by their short duration. I had been told that astronomical time scales were huge compared to our lives – but these events, instead, can be as fast as lightning! A precise measurement of their duration is not trivial, because of many practical issues such as the relatively narrow band of the detectors, the uncertainty in distinguishing the signal from the background in the dim part of the lightcurve, the presence of quiescent phases. A simple and widely adopted definition of their duration is the time $T_{90}$ over which 90 percent of the detector counts above the background are recorded, leaving out the first and the last five percent. Such definition clearly depends on the instrument band, on its sensitivity, and on the background model, but it is practical in many cases, and it is a standard piece of information given in catalogues. Figure 6 shows the distribution of $T_{90}$ for all bursts detected by CGRO/BATSE. The distribution is clearly bimodal, and can be fit by a mixture of two log-normal distributions (Kouveliotou et al., 1993), which overlap at approximately 2 s. This is the historical reason of the distinction between Short GRBs ($T_{90} < 2$ s) and Long GRBs ($T_{90} > 2$ s).

THE AFTERGLOW

Since its discovery in 1997, afterglow emission has been routinely observed in the follow-up of many GRBs. The observed behaviour is rather diverse, but I will try to identify some general features in order to get to a broad description of their observational appearance.

Temporal evolution

Let me divide the afterglow evolution into four stages: early afterglow (minutes to hours after the prompt emission), normal decay phase (few days), steep decay phase (from one week to few months), late afterglow (months to years). Before the launch of Swift, the “normal decay phase” was the one most commonly observed:
Figure 7: Observed afterglows of short gamma-ray bursts. In all panels, grey triangles represent 3 sigma upper limits, while solid lines represent detections. For radio observations only the one sigma uncertainty is represented by the error bars. **Upper panel:** flux densities from spacecraft X-ray observations at $1\,\text{keV}$. **Middle panel:** flux densities and corresponding AB magnitudes from Earth-based observations in the optical and near-infrared. **Lower panel:** flux densities from radio observations between $1$ and $10\,\text{GHz}$. All data are from the catalogues of Fong et al., (2015).
not many bursts were well localized, and hours to days were usually needed to start the follow-up, during which the emission faded significantly. When observed in X-ray and in optical, this phase usually features a power law behaviour both in frequency and in time, i.e. $F_\nu \propto \nu^\alpha t^\beta$, with the temporal decay index being $\beta \sim -1$. The automatic, rapid repointing system of *Swift* (Gehrels, Ramirez-Ruiz, and Fox, 2009) enabled observations of the “early” phase, where a more complex behaviour is often found (see e.g. Kann et al., 2010), which often involves the presence of an early peak and a subsequent steeper (or sometimes shallower) decay with respect to the normal decay phase. If the afterglow is sufficiently bright, it can still be observable after a few days or weeks, after which often it shows an achromatic steepening of the temporal decay index (the “steep decay phase”). The “late afterglow” phase is uniquely observed in the Radio, and it usually shows no evident temporal evolution.

*Afterglow spectra*

GRB afterglows have been detected in a very wide range of frequencies, from GeV gamma-rays (with instruments such as LAT onboard *Fermi*) down to less than 1 GHz in radio. As stated in the preceding paragraph, GRB afterglow spectra feature a non-thermal spectrum. Broadband simultaneous spectral energy distributions (SEDs) reveal the presence of spectral breaks, showing that the spectrum is composed of multiple, smoothly connected power law branches. The spectral break frequencies are seen to evolve during time, and the peak of the spectrum to move towards lower frequencies.

*The blastwave interpretation*

The above observations are most commonly interpreted as being caused by a blastwave expanding into the interstellar medium (ISM) surrounding the GRB progenitor. After producing the prompt emission, the GRB ejecta move at relativistic speed and they expand into the ISM, sweeping it. As soon as the rest mass of the collected ISM becomes comparable to the kinetic energy of the ejecta, a strong shock wave is formed. At the shock, electrons are accelerated by the Fermi process and radiate mainly by synchrotron emission. As long as the expansion is ultra-relativistic, the structure of the blastwave and its deceleration evolve in a self-similar fashion (Blandford and McKee, 1976). When the blastwave becomes mildly relativistic (Lorentz factor $\Gamma \lesssim 3$) a slow transition to the Newtonian regime takes place, after which the system resembles a supernova remnant.

*Presence of a Jet*

Right after the discovery of the first GRB afterglow, Rhoads, (1997) argued that if the afterglow emission comes from material collimated into a jet, rather than from an isotropic explosion, then the behaviour of the light curve must change qualitatively after the reciprocal of the bulk Lorentz factor $\Gamma^{-1}$ becomes comparable to the jet half-opening angle $\theta_{\text{jet}}$. The qualitative change corresponds to an achromatic steepening of the decay of the light curve, usually called “jet break”.
Figure 8: Possible jet breaks in the X-ray afterglow light curves of a sample of *Swift* gamma-ray bursts. The break time is shown by a vertical bar. Reproduced from Racusin et al., (2009).
Quantitative predictions for the afterglow of jetted GRBs were published two years later (Rhoads, 1999), just before the first clear observation of a jet break in the optical and radio light curve of GRB990510 (Harrison et al., 1999). Despite quite convincing jet breaks have been found in several afterglow light curves in the following years (see Fig. 8 for some examples), some afterglows do not show any achromatic steepening up to several days after the prompt emission, suggesting that the distribution of jet opening angles could be wide, or that some ingredients are missing in standard afterglow modeling.

PROGENITORS

A typical Long GRB has a fluence of the order of $10^{-5}$ erg cm$^{-2}$ and it is located at a redshift $z \sim 1.5$. This implies an isotropic equivalent energy release of $10^{52}$ erg. The presence of afterglow radiation indicates that the energy radiated in the prompt emission is only a fraction of the kinetic energy of the ejecta, which can be 10 times as energetic. If the ejecta are collimated, though, the actual (collimation-corrected) kinetic energy content is lowered by a factor $\sim \theta_{\text{jet}}^2$. The result is still of the order of $10^{51}$ erg. This huge energy must be liberated in a process that produces variability on the millisecond time-scale. Accretion on a stellar black hole has a high enough efficiency and happens in a compact enough region to be compatible with the energy and variability time-scale requirements. The transient nature of GRBs, moreover, suggests that such accretion must be linked to a catastrophic event. All these pieces of information, when gathered together, suggest that GRBs are linked to the birth of a stellar black hole.

*Long versus Short, Collapsar versus Non-Collapsar*

The natural birthplaces of stellar black holes are two: the collapse of a massive star and the coalescence of two compact objects. The latter was the first to be identified as a promising GRB progenitor candidate (Eichler et al., 1989), followed a few years later by the former (Woosley, 1993). In 1998, the observation of supernova 1998bw associated to the close-by GRB980425 (Galama et al., 1998) provided compelling evidence for the collapse of a massive star (later dubbed the “collapsar scenario” MacFadyen and Woosley, 1999). In the following years, several secure associations between GRBs and supernovae have been made. On the other hand, the search for supernovae associated to nearby short GRBs always led to (sometimes very stringent) upper limits so far (Berger, 2014a). For this class of GRBs the coalescence of two compact objects is thus the favoured scenario.
TRYING TO GET A GRASP OF THE SHORT GRB POPULATION

BRIEFING

When somewhat more than a few events of a particular class have been observed, a fundamental question is whether the properties of the observed sample can inform us, in a statistical sense, about the intrinsic properties of the whole population. In astronomy there is usually no guarantee that a sample is representative, so that particular care has to be put in modeling selection effects and trying to keep them under control. Moreover, in most cases only incomplete information is available: for example, most known GRBs have no redshift measurement, thus their intrinsic luminosity or energy cannot be derived. In this chapter I will present how these problems have been dealt with, in the case of short gamma-ray bursts (SGRBs), in a recent work led by Giancarlo Ghirlanda and me (Ghirlanda et al., 2016). Our approach led us to define the first model of the SGRB population able to explain all the statistical properties of the observed population. Based on this model, we made predictions for the rate of SGRBs to be observed in association with gravitational waves (GW) from compact binary mergers in the upcoming runs of Advanced LIGO and Virgo. In the following sections, I present the approach and results. At the end of the chapter, I comment on the results in light of the recent development following the observation of GW170817 and its electromagnetic counterparts.

INTRODUCTION

The intrinsic properties of the SGRB population are still poorly understood, partly because of the small number of events with measured redshift (see e.g. Berger, 2014b; D’Avanzo, 2015, for recent reviews). Rather sparse information about the origin of these events is available, but the low density of the circum-burst medium (Fong and Berger, 2013; Fong et al., 2015a), the variety of galaxy morphologies (e.g. D’Avanzo, 2015), the lack of any associated supernova in nearby SGRBs, and the possible recent detection of a “kilonova” (Eichler et al., 1989; Li and Paczyński, 1998; Yang et al., 2015a; Yang et al., 2015b; Jin et al., 2016; Jin et al., 2015) signature (Berger, Fong, and Chornock, 2013; Tanvir et al., 2013), all hint at the merger of two compact objects (e.g. double neutron stars) rather than a single massive star collapse as in long GRBs.

On the other hand, the prompt γ-ray emission properties of SGRBs (Ghirlanda et al., 2009; Ghirlanda et al., 2015), the sustained long-lasting X-ray emission (although not ubiquitous in short GRBs; Sakamoto and Gehrels 2009) and flaring activity suggest that the central engine and radiation mechanisms are similar to those of long GRBs. Despite this is based on a few breaks in the optical light curves, it seems also that SGRBs have jets: current measures of θjet are between 3° and 15°, while lower limits seem to suggest a wider distribution (e.g. Berger, 2014b; Fong et al., 2015b).
If the progenitors are compact object binaries (made of two neutron stars – “NS-NS” – or of a neutron star and a black hole – “NS-BH”), SGRBs are among the most promising electromagnetic counterparts of GW events detectable by advanced interferometers. The rate of association of GW events with SGRBs is mainly determined by the rate of SGRBs within the relatively small horizon set by the sensitivity of aLIGO and Advanced Virgo (Abbott et al., 2016). However, estimates of the local SGRB rates vary from $0.1 - 0.6 \text{ Gpc}^{-3} \text{ yr}^{-1}$ (e.g. Guetta & Piran, 2005; Guetta and Piran, 2006; Guetta and Stella, 2009; Coward et al., 2012; Siellez, Boër, and Gendre, 2014; Wanderman and Piran, 2015) to even larger values, e.g. $40 - 240 \text{ Gpc}^{-3} \text{ yr}^{-1}$ (Nakar, Gal-Yam, and Fox, 2006; Guetta and Piran, 2006). These rates are not corrected for the collimation angle, i.e. they represent the fraction of bursts whose jets are pointed towards the Earth, whose $\gamma$–ray prompt emission can be detected.

These rate estimates depend mainly on the luminosity function $\phi(L)$ and redshift distribution $\Psi(z)$ of SGRBs. These functions are usually derived by fitting the peak flux distribution of SGRBs detected by BATSE (Guetta and Piran, 2005; Guetta and Piran, 2006; Nakar, Gal-Yam, and Fox, 2006; Hopman et al., 2006; Salvaterra et al., 2008). Owing to the degeneracies in the parameter space (when both $\phi(L)$ and $\Psi(z)$ are parametric functions), the redshift distribution is also constrained by comparison with that constructed from the few SGRBs with measured $z$.

The luminosity function $\phi(L)$ is typically modelled as a single or broken power law, and in most cases it is found to be similar to that of long GRBs (i.e. proportional to $L^{-1}$ and $L^{-2}$ below and above a characteristic break $\sim 10^{51-52} \text{ erg s}^{-1}$; Guetta and Piran, 2006; Salvaterra et al., 2008; Virgili et al., 2011; D’Avanzo et al., 2014, hereafter D14; or much steeper ($L^{-2}$ and $L^{-3}$; Wanderman and Piran, 2015, hereafter WP15). Aside from the mainstream, Shahmoradi and Nemiroff, 2015 modelled all the distributions with lognormal functions.

The redshift distribution $\Psi(z)$ (the number of SGRBs per comoving unit volume and time, as a function of redshift $z$) has always been assumed to follow the cosmic star formation rate convolved with a delay time distribution, which accounts for the time necessary for the progenitor binary system to merge. With this assumption, various authors derived the delay time $\tau$ distribution, which could be a single power law $P(\tau) \propto \tau^{-\delta}$ (e.g. with $\delta = 1 - 2$; Guetta and Piran 2005, 2006; D14; WP15) with a minimum delay time $\tau_{\text{min}} = 10 - 20 \text{ Myr}$, or a peaked (lognormal) distribution with a considerably large delay (e.g. $2 - 4 \text{ Gyr}$, Nakar and Gal-Yam 2005; WP15). Alternatively, the population could be described by a combination of prompt mergers (small delays) and large delays (Virgili et al., 2011) or to the combination of two progenitor channels, i.e. binaries formed in the field or dynamically within globular clusters (e.g. Salvaterra et al., 2008).

Many past works feature a common approach: parametric forms are assumed for the compact binary merger delay time distribution and for the SGRB luminosity function; free parameters of these functions are then constrained through the small sample of SGRBs with measured redshift, and through the distribution of the $\gamma$–ray peak fluxes of SGRBs detected by past and/or present GRB detectors. A number of other observer frame properties, though, are available and have not been used: fluence distribution, duration distribution, observer frame peak energy. The last of these has been considered in Shahmoradi and Nemiroff, (2015) which,
however, lacks a comparison with rest-frame properties of SGRBs as is done in this chapter. Another relevant issue with many previous works is the comparison of the model predictions with small and incomplete samples of SGRBs with measured $z$. Indeed, only recently D14 have constructed a flux-limited complete sample of SGRBs detected by Swift.

The aim of this chapter is to present a model of the SGRB population, whose redshift distribution $\Psi(z)$ and luminosity function $\phi(L)$ are constrained using all the available observational constraints of the large sample of bursts detected by Fermi/GBM. These constraints are the (1) peak flux, (2) fluence, (3) observer frame duration and (4) observer frame peak energy distributions. Additionally, I also consider as constraints (5) the redshift distribution, (6) the isotropic energy, and (7) the isotropic luminosity of a complete sample of SGRBs detected by Swift (D14). This is the first time that the $\phi(L)$ and $\Psi(z)$ of SGRBs are constrained using 2–4 and 6–7. In the formulation of the model, I do not assume any delay time distribution for SGRBs, but I assume a quite general parametric form and derive directly its parameters.

In §3.3 I describe a sample of SGRBs (without measured redshifts) detected by Fermi/GBM, which provides the observer-frame constraints 1–4, and the complete (though smaller) sample of Swift SGRBs of D14, which provides the rest-frame constraints 5–7. In §3.6 I show that that a steep $\phi(L)$ is excluded when all the available constraints (1–7) are considered. In §3.7 I show a Monte Carlo approach to derive the parameters describing the $\phi(L)$ and $\Psi(z)$ of SGRBs. In §3.9 the results on the $\phi(L)$ and $\Psi(z)$ of SGRBs are presented and discussed. I assume a standard flat $\Lambda$CDM cosmology with $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$ and $\Omega_m = 0.3$ throughout this chapter.

**SELECTING A GOOD SAMPLE**

As stated in the preceding section, the luminosity function and redshift distribution of SGRBs have been derived by many authors by taking into account the following two constraints:

1. the peak flux distribution of large samples of SGRBs detected by CGRO/BATSE or Fermi/GBM;
2. the redshift distribution of the SGRBs with measured $z$.

However, a considerable amount of additional information on the prompt $\gamma$-ray emission of SGRBs can be extracted from the BATSE and GBM samples. In particular, we can learn more about these sources by considering the distributions of

3. the peak energy $E_{p,0}$ of the observed $\nu F_{\nu}$ spectrum;
4. the fluence $F$;
5. the duration $T_{90}$.

Moreover, for the handful of events with known redshift $z$, we have also access to the$^1$

---

$^1$ To avoid too much redundancy, throughout this chapter I will sometimes drop the “iso” subscript, so that $L_{iso}$ and $E_{iso}$ will be equivalently written as $L$ and $E$. For the same reason, the peak energy
6. isotropic luminosity $L_{\text{iso}}$;

7. isotropic energy $E_{\text{iso}}$.

**Observer-frame constraints: a flux-limited Fermi/GBM sample**

For the distributions of the observer frame prompt emission properties (constraints 1, 3, 4, 5) I consider the sample of 1767 GRBs detected by Fermi/GBM (from 080714 to 160118) as reported in the online spectral catalogue\(^2\). It contains most of the GRBs published in the second spectral catalogue of Fermi/GBM bursts (relative to the first four years) (Gruber et al., 2014), plus events detected by the satellite in 2015 and 2016. 295 events in the sample are SGRBs (i.e. with $T_{90} \leq 2$ s). According to Bromberg et al., 2013, for both the Fermi and CGRO GRB populations, this duration threshold should limit the contamination from collapsar-GRBs to less than 10\% (see also WP15).

I only select bursts with a peak flux (computed on 64ms timescale in the 10-1000 keV energy range) larger than 5 ph cm\(^{-2}\) s\(^{-1}\) in order to work with a well-defined sample, less affected by the incompleteness close to the detector limiting flux. With this selection, the sample reduces to 211 SGRBs, detected by Fermi/GBM in 7.5 years within its field of view of ~70\% of the sky.

I consider the following prompt emission properties of the bursts in the sample to be used as constraints of the population synthesis model:

- the distribution of the 64ms peak flux $P_{64}$ (integrated in the 10-1000 keV energy range). This is shown by black symbols in the top left panel of Fig. 9;

- the distribution of the observed peak energy of the prompt emission spectrum $E_{p,o}$ (black symbols, bottom left panel in Fig. 9);

- the distribution of the fluence $F$ (in the 10–1000 keV energy range) (black symbols, bottom middle panel in Fig. 9);

- the distribution of the duration $T_{90}$ of the prompt emission (black symbols, bottom right panel in Fig. 9);

Short GRB spectra have a typical observer frame peak energy $E_{p,o}$ distribution (e.g. Ghirlanda et al., 2009; Nava et al., 2011b; Gruber et al., 2014) centred at relatively large values (~ 0.5 – 1 MeV), as is also shown by the distribution in the bottom left panel of Fig. 9. For this reason, I adopt here the peak flux $P_{64}$ and fluence $F$ computed in the wide 10–1000 keV energy range as provided in the spectral catalogue of Fermi bursts rather than the typically adopted 50–300 keV peak flux (e.g. from the BATSE archive), which would sample only a portion of the full spectral curvature.

The distributions of the peak flux, fluence, peak energy, and duration are shown in Fig. 9 with black symbols. Error bars are computed by resampling each measurement ($P$, $F$, $E_{p,o}$, and $T_{90}$) from a normal distribution with a sigma equal to the measurement uncertainty. For each bin, the vertical error bars represent the standard deviations of the bin heights of the resampled distributions.

---

\(^2\) https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermigbrst.html
Rest-frame constraints: the Swift SBAT4 sample

For the redshift distribution and the rest frame properties of SGRBs (constraints 2, 6, and 7) I consider the sample published in D14. It consists of bursts detected by Swift, selected with criteria similar to those adopted for long GRBs in Salvaterra et al., 2012, with a peak flux (integrated in the 15–150 keV energy range and computed on a 64 ms timescale) $P_{64} \geq 3.5$ photons cm$^{-2}$ s$^{-1}$. This corresponds to a flux which is approximately four times larger than the Swift–BAT minimum detectable flux on this timescale; for this reason, I call this sample SBAT4 (Short BAT 4). The redshift distribution of the SBAT4 sample is shown in the top right panel of Fig. 9 (solid black line). Within the SBAT4 sample I consider the 11 GRBs with known $z$ and determined $L_{\text{iso}}$ and $E_{\text{iso}}$ (the distributions of these quantities are shown in the inset of Fig. 9, top right panel, with black and grey lines respectively). The grey shaded region shows how the distribution changes when the five SGRBs in the sample with unknown $z$ are all assigned the minimum or the maximum redshift of the sample.

A sensible parametrization of the $\phi(L)$ and $\Psi(z)$

Given the incompleteness of the available SGRB samples, particularly with measured $z$, no direct method as for the population of long GRBs; see e.g. Pescalli et al., 2016 can be applied to derive the shape of the SGRB luminosity function $\phi(L)$ and redshift distribution $\Psi(z)$ from the observations. The typical approach in this case consists in assuming some simple analytical shape for both functions, with free parameters to be determined by comparison of model predictions with observations.

For the luminosity function, a power law

$$\phi(L) \propto L^{-\alpha}$$

or a broken power law

$$\phi(L) \propto \begin{cases} (L/L_b)^{-\alpha_1} & L < L_b \\ (L/L_b)^{-\alpha_2} & L \geq L_b \end{cases}$$

normalized to its integral is usually assumed. I will assume the latter.

If SGRBs are produced by the merger of compact objects, their redshift distribution should follow a “retarded” star formation,

$$\Psi(z) = \int_z^\infty \psi(z')P[t(z) - t(z')] \frac{dt}{dz'} dz'$$

where $\psi(z)$ represents the formation rate of SGRB progenitors per comoving Gpc$^{-3}$ yr$^{-1}$, and $P(\tau)$ is the delay time distribution, i.e. the probability density function of the delay $\tau$ between the formation of the progenitors and their merger (which produces the SGRB). Adopting the point of view that SGRBs are produced by the coalescence of a NS-NS (or BH-NS) binary, one can assume a delay time distribution and convolve it with a $\psi(z)$ of choice to obtain the corresponding SGRB formation rate $\Psi(z)$. Theoretical considerations and population
Figure 9: Black dots show the distributions obtained from the Fermi/GBM and Swift SBAT4 samples (§3.3). Horizontal error bars are the bin widths, while vertical error bars are 1σ errors on the bin heights accounting for experimental errors on single measurements. The results of the Monte Carlo population synthesis code are shown by solid red lines (for the model in which $E_p - L_{iso}$ and $E_p - E_{iso}$ correlations are assumed to hold) and by triple dot-dashed orange lines (for the model with no correlations). Predictions based on the models of D14 and WP15 are shown by dashed blue and dot-dashed cyan lines, respectively (the latter only in the first three panels; see the text). These are obtained by analytical methods of §3.5. **Top left panel:** Distribution of the peak flux $P$ of the Fermi/GBM sample. **Top right panel:** Normalized cumulative redshift distribution of the SBAT4 sample. The grey shaded area represents the change in the distribution if the remaining bursts with unknown $z$ are all assigned the largest or the lowest $z$ of the sample. The inset shows the cumulative distributions of the isotropic luminosity $L_{iso}$ (solid black line) and energy $E_{iso}$ (solid grey line) of the same sample. **Bottom panels:** From left to right, distributions of peak energy $E_{p,o}$, fluence, and duration of SGRBs of the Fermi/GBM sample.
3.4 A sensible parametrization of the \( \phi(L) \) and \( \Psi(z) \)

synthesis (Portegies Zwart and Yungelson, 1998; Schneider et al., 2001; Belczynski et al., 2006; O’Shaughnessy, Belczynski, and Kalogera, 2008; Dominik et al., 2013) suggest that compact binary coalescences should typically follow a delay time distribution \( P(\tau) \propto \tau^{-1} \) with \( \tau \geq 10 \) Myr. Equation 5 is actually a simplification, in that it implicitly assumes that the fraction of compact binaries with respect to all stars formed does not depend on redshift. The actual fraction very likely depends on metallicity and on the initial mass function, and thus indirectly on redshift. Moreover, the star formation history itself is affected by uncertainties, which affect the result of the convolution. To make the analysis as general as possible, I thus prefer to adopt a generic parametric form for the redshift distribution \( \Psi(z) \) of SGRBs. A posteriori, the delay time distribution (in the compact binary progenitor scenario) can be recovered by direct comparison of the result with the star formation history of choice. I parametrize the \( \Psi(z) \) following Cole et al., 2001, namely

\[
\Psi(z) = \frac{1 + p_1 z}{1 + (z/z_p)^{p_2}}
\]

which rises and then decays (for \( p_1 > 0, p_2 > 1 \)), with a peak roughly⁴ corresponding to \( z_p \).

Two past works

Let me now consider the works of D14 and WP15 in more detail, which will be useful as a comparison.

D’Avanzo et al. (2014) assume a power law shape for both the \( \phi(L) \) and the delay time distribution \( P(\tau) \), and they adopt the parametric form of Cole et al., 2001 for the cosmic star formation history, with parameter values taken from Hopkins and Beacom, 2006. They assume that SGRBs follow the \( E_p - L_{iso} \) correlation \( E_{\text{peak}} = 337 \text{keV} (L_{iso}/2 \times 10^{52} \text{ergs}^{-1})^{0.49} \) and that their spectrum is a Band function (Band et al., 1993) with low and high energy photon spectral indices -0.6 and -2.3, respectively. They constrain the free parameters by fitting the BATSE peak flux distribution and the redshift distribution of bright Swift short bursts with measured \( z \). They find \( \phi(L) \propto L^{-2.17} \) between \( 10^{49} \text{ erg s}^{-1} \) and \( 10^{55} \text{ erg s}^{-1} \), and \( P(\tau) \propto \tau^{-1.5} \) with a minimum delay of 20 Myr. The blue dashed lines in Fig. 9 are obtained through Eq. 4 and Eq. 5 using the same parameters as D14: their model (limited to \( P_{\text{lim}} \geq 5 \text{ ph cm}^{-2} \text{ s}^{-1} \)) in order to be compared with the sample selected in this work) reproduces correctly the peak flux distribution (top left panel of Fig. 9) of Fermi SGRBs and the redshift distribution of the bright SGRBs detected by Swift (top right panel).

The preferred model for \( \phi(L) \) in WP15 is a broken power law, with a break at \( 2 \times 10^{52} \text{ erg s}^{-1} \) and pre- and post-break slopes of \( -1.9 \) and \( -3.0 \), respectively. Their preferred models are either a power law delay time distribution \( P(\tau) \propto \tau^{-0.81} \) with a minimum delay of 20 Myr or a lognormal delay time distribution with central value 2.9 Gyr and sigma \( \leq 0.2 \). Differently from D14, rather than assuming the \( E_p - L_{iso} \) correlation they assign to all SGRBs a fixed rest frame

---

⁴ The exact peak is not analytical, but a good approximation is \( z_{\text{peak}} \approx z_p \left( \frac{1 + p_2}{1 + p_1 z_p} \right) \) where the exact peak is not analytical, but a good approximation is \( z_{\text{peak}} \approx z_p \left( \frac{1 + p_2}{1 + p_1 z_p} \right) \).
E_{\text{p,rest}} = 800 \text{ keV}. The dot-dashed cyan lines in Fig. 9 are the model of WP15 (for the lognormal P(\tau) case).

In the following I show how the results of WP15 and D14, both representative of a relatively steep luminosity function, compare with the other additional constraints (bottom panels of Fig. 9) that I consider in this chapter.

FROM POPULATION PROPERTIES TO OBSERVABLES

Given the two functions \( \phi(L) \) and \( \Psi(z) \), the peak flux distribution can be derived as

\[
N(P_1 < P < P_2) = \frac{\Delta \Omega}{4\pi} \int_0^\infty \frac{dV(z)}{dz} \frac{\Psi(z)}{1+z} \int_{L(P_1,z)}^{L(P_2,z)} \phi(L) dL,
\]

where \( \Delta \Omega/4\pi \) is the fraction of sky covered by the instrument/detector (which provides the real GRB population with which the model is to be compared) and \( dV(z)/dz \) is the differential comoving volume. The flux \( P \) corresponding to the luminosity \( L \) at redshift \( z \) is

\[
P(L,z,E_{\text{peak}}, \alpha) = \frac{L}{4\pi d_L(z)^2} \int_{\epsilon_1 z}^{\epsilon_2 z} N(E|E_{\text{peak}}, \alpha) dE \int_0^\infty E N(E|E_{\text{peak}}, \alpha) dE,
\]

where \( d_L(z) \) is the luminosity distance at redshift \( z \) and \( N(E|E_{\text{peak}}, \alpha) \) is the rest frame photon spectrum of the GRB. The photon flux \( P \) is computed in the rest frame energy range \( [(1+z)\epsilon_1, (1+z)\epsilon_2] \), which corresponds to the observer frame \( [\epsilon_1, \epsilon_2] \) band.

The SGRB spectrum is often assumed to be a cut-off power law, i.e. \( N(E|E_{\text{peak}}, \alpha) \propto E^{-\alpha} \exp(-E(\beta - \alpha)/E_{\text{peak}}) \), or a Band function (Band et al., 1993). Typical parameter values are \( \alpha \sim 0.6 \) (i.e. the central value of the population of SGRBs detected by BATSE and Fermi - Ghirlanda et al., 2009; Nava et al., 2011b; Goldstein and Preece, 2010; Gruber et al., 2014) and, for the Band function, \( \beta \sim 2.3 - 2.5 \). The peak energy is either taken as fixed (e.g. 800 keV in WP15) or derived assuming that SGRBs follow a \( E_{\text{p}} - L_{\text{iso}} \) correlation in analogy to long bursts (e.g. D14; Virgili et al., 2011). Recent evidence supports the existence of such a correlation among SGRBs (see e.g. D14; Calderone et al., 2015; Tsutsui et al., 2013; Ghirlanda et al., 2009) with similar parameters to those present in the population of long GRBs (Yonetoku et al., 2004).

In order to compare the model peak flux distribution obtained from Eq. 7 with the real population of GRBs, only events with peak flux above a certain threshold \( P_{\text{lim}} \) are considered. The integral in Eq. 7 is thus performed over the \( (L,z) \) range where the corresponding flux is larger than \( P_{\text{lim}} \).

In D14 the assumption of the correlation \( (E_{\text{p}} - L_{\text{iso}}) \) between the isotropic luminosity \( L_{\text{iso}} \) and the rest frame peak energy \( E_{\text{p}} \) also allows one to derive, from Eq. 7, the expected distribution of the observer frame peak energy \( E_{\text{p,o}} \).

\[
N(E_{\text{1,p,o}} < E < E_{\text{2,p,o}}) = \int_0^\infty dC(z) \int_{L(E_{\text{1,p,o},z})}^{L(E_{\text{2,p,o},z})} \phi(L) dL,
\]

The assumption of a spectrum is required to convert the bolometric flux into a characteristic energy range for comparison with real bursts.
where \( E_{p,o} \) is the peak energy of the observed \( \nu F(\nu) \) spectrum, and I set \( C(z) = [\Delta \Omega/4\pi]\Psi(z)/(1+z)[dV(z)/dz] \). The limits of the luminosity integral are computed by using the rest frame correlation \( E_p = Y L^m \), namely

\[
L(E_{p,o}, z) = \left( \frac{E_p}{Y} \right)^{1/m_y} = \left( \frac{(1+z)E_{p,o}}{Y} \right)^{1/m_y}.
\]

In order to compare the distribution of \( E_{p,o} \) with real data, the integral in Eq. 9, similarly to Eq. 7, is performed over values of \( L(E_{p,o}, z) \) corresponding to fluxes above the limiting flux adopted to extract the real GRB sample (e.g. \( 5 \) ph cm\(^{-2}\) s\(^{-1}\) for SGRBs selected from the Fermi sample).

Similarly, by assuming a \( E_p - E_{iso} \) correlation to hold in SGRBs (see D14; Tsutsui et al., 2013; Amati, 2006; Calderone et al., 2015), i.e. \( E_p = A E^{m_a} \), one can derive a relation between luminosity and energy \( (L_{iso}-E_{iso}) \), which reads

\[
L(E) = \left( \frac{A}{Y} \right)^{1/m_y} E^{m_a/m_y}.
\]

This can then be used to compute the fluence distribution, where the fluence is related to the isotropic energy as \( F = E/(1+z)/4\pi d_L(z)^2 \),

\[
N(F_1 < F < F_2) = \int_0^\infty dz C(z) \int_{L(E_1)}^{L(E_2)} \phi(L) dL,
\]

again by limiting the integral to luminosities which correspond to fluxes above the given limiting flux.

Finally, since the light curves of SGRBs are usually single spikes, one can assume a triangular shape and thus let \( 2E/L \propto T \) in the rest frame of the source. Therefore, it is possible to combine the \( E_p - E_{iso} \) and \( E_p - L_{iso} \) correlations to derive the model predictions for the distribution of the duration to be compared with the observed distribution,

\[
N(T_{1,o} < T < T_{2,o}) = \int_0^\infty dz C(z) \int_{L(T_{1,o}, z)}^{L(T_{2,o}, z)} \phi(L) dL,
\]

where

\[
L(T_o, z) = \left( \frac{Y}{A} \right)^{1/m_a} \frac{2(1+z)}{T_o} \frac{1}{(1-m_y/m_a)}.
\]

A TOO STEEP LUMINOSITY FUNCTION IS AT ODDS WITH THE OBSERVER-FRAME CONSTRAINTS

The bottom panels of Fig. 9 show the distributions of peak energy \( E_{p,o} \) (left), fluence \( F \) (middle), and duration \( T_{90} \) (right) of the sample of short Fermi GRBs described in §3.3 (black symbols). Predictions using the same parameters as in D14 are shown by dashed blue lines in Fig. 9: while the \( P \) and \( z \) distributions are correctly reproduced (top panels of Fig. 9), the model is inconsistent with the distributions of peak energy \( E_{p,o} \), fluence \( F \), and duration (bottom panels of Fig. 9). For the D14 model, I assumed the \( E_p - E_{iso} \) correlation as reported in that paper.
to derive the fluence and (in combination with the $E_p - L_{iso}$ correlation) the duration distribution. Since WP15 assume a unique value of the peak energy $E_{p,o}$, it is not possible to derive the fluence and duration of their model unless independent functions for these parameters are assumed. Therefore, the model of WP15 (dot-dashed cyan line in Fig. 9) is compared only with the peak flux, redshift (top panels), and observed peak energy (bottom left panel of Fig. 9) distributions.

The figure shows that a steep $\phi(L)$ combined with either a power law distribution of delay times favouring short delays (as in D14) or a nearly unique long delay time (as in the log–normal model of WP15) correctly reproduce the observer frame peak flux distribution of Fermi GRBs and the redshift distribution of Swift bright short bursts. However, they do not reproduce the peak energy, fluence, and duration distributions of the same population of Fermi SGRBs.

Motivated by these results, Giancarlo Ghirlanda and I implemented a Monte Carlo (MC) code aimed at deriving the $\phi(L)$ and $\Psi(z)$ of SGRBs which satisfy all the constraints (1–7) described above. The reason to choose a MC method is that it allows for an easy implementation of the dispersion of the correlations (e.g. $E_p - L_{iso}$ and $E_p - E_{iso}$) and of any distribution assumed (which are less trivial to account for in an analytic approach as that shown above).

A MONTE CARLO SIMULATION OF THE SGRB POPULATION

In this section I describe the Monte Carlo (MC) approach adopted in Ghirlanda et al., (2016) to generate the model population. The approach is based on the following choices:

1. Customarily, Eq. 5 has been used to compute the redshift distribution $\Psi(z)$ of SGRBs from an assumed star formation history $\psi(z)$ and a delay time distribution $P(\tau)$. As stated in §3.4, this approach implies simplifications we wanted to avoid, so the more general form given in Eq. 6 was assumed;

---

5 Here I consider as a constraint the population of Fermi/GBM GRBs. Nava et al., 2011 showed that the BATSE SGRB population has similar prompt emission properties as Fermi SGRBs (peak flux, fluence, and duration distribution).
2. In order to avoid inducing spurious correlations, it is convenient to extract $E_p$ from an assumed probability distribution and then use the correlations to associate to it a luminosity and an energy. We considered a broken power law shape for the $E_p$ distribution:

$$
\phi(E_p) \propto \begin{cases} 
    (E_p/E_{p,b})^{-a_1} & \text{if } E_p \leq E_{p,b} \\
    (E_p/E_{p,b})^{-a_2} & \text{if } E_p > E_{p,b}
\end{cases}
$$

(15)

Through the $E_p - L_{iso}$ and $E_p - E_{iso}$ correlations, also accounting for their scatter, one can then associate to $E_p$ a luminosity $L_{iso}$ and an energy $E_{iso}$. The luminosity function of the population is then constructed as a result of this procedure;

3. We assumed the correlations $E_p - L_{iso}$ and $E_p - E_{iso}$ written as

$$
\log_{10}(E_p/670 \text{ keV}) = q_Y + m_Y \log_{10}(L/10^{52} \text{ erg s}^{-1})
$$

(16)

and

$$
\log_{10}(E_p/670 \text{ keV}) = q_A + m_A \log_{10}(E_{iso}/10^{51} \text{ erg}).
$$

(17)

For each GRB, after sampling $E_p$ from Eq. 15, we associated a luminosity (resp. energy) sampled from a lognormal distribution whose central value is given by Eq. 16 (resp. 17), with $\sigma = 0.2$. There are still too few SGRBs with known redshift to measure the scatter of the corresponding correlations. We thus assumed the same scatter found in the correlations holding for the population of long GRBs (Nava et al., 2012);

4. For each GRB, a typical Band function prompt emission spectrum was assumed, with low and high photon spectral index $-0.6$ and $-2.5$, respectively. We kept these two parameters fixed after checking that our results were unaffected by sampling them from distributions centred around these values.

For each synthetic GRB, the scheme in Fig. 10 was followed: a redshift $z$ is sampled from $\Psi(z)$ and a rest frame peak energy $E_p$ is sampled from $\phi(E_p)$; through the $E_p - L_{iso}$ (resp. $E_p - E_{iso}$) correlation a luminosity $L_{iso}$ (resp. energy $E_{iso}$) with lognormal scatter is assigned; using the redshift and luminosity (energy), the peak flux $P$ (fluence $F$) in the observer frame energy range $10^{-1000} \text{ keV}$ is derived via the assumed spectral shape. The observer frame duration $T$ is obtained as $2(1+z)E/L$, i.e. the light curve is approximated with a triangle. This scheme reflects the procedure followed to compute the observer frame quantities in “model (a)”.

The minimum and maximum values of $E_p$ admitted are $E_{p,\text{min}} = 0.1 \text{ keV}$ and $E_{p,\text{max}} = 10^5 \text{ keV}$. These limiting values correspond to a minimum luminosity $L_{\text{min}}$ and a maximum luminosity $L_{\text{max}}$ which depend on the $E_p - L_{iso}$ correlation. While the maximum luminosity is inessential (in all solutions the high luminosity slope $\alpha_2 \gtrsim 2$), the existence of a minimum luminosity might affect the observed distributions. We thus implemented an alternative scheme (“model (b)”) in which the minimum luminosity $L_{\text{min}}$ is a parameter, and values of $E_p$ which correspond to smaller luminosities are rejected.

---

We also made sure that our results were not sensitive to a slightly different choice of the spectral parameters, i.e. low and high energy spectral index $-1.0$ and $-3.0$, respectively.
In order to investigate the dependence of the results on the assumption of the $E_p - L_{iso}$ and $E_p - E_{iso}$ correlations, we also implemented a third MC scheme (“model (c)”) where independent probability distributions (i.e., independent from the peak energy) were assumed for the luminosity and duration. A broken power law

$$P(L) \propto \begin{cases} (L/L_b)^{-\alpha_1} & L \leq L_b \\ (L/L_b)^{-\alpha_2} & L > L_b \end{cases}$$

was assumed for the luminosity distribution, and a lognormal shape

$$P(T_r) \propto \exp \left[ -\frac{1}{2} \left( \frac{\log(T_r) - \log(T_c)}{\sigma_{T_c}} \right)^2 \right]$$

was assumed for the rest frame duration $T_r = T/(1+z)$ probability distribution. Again, the energy of each GRB was computed as $E = LT_r/2$, i.e. the light curve was approximated with a triangle.

**Looking for the best fit parameters**

In model (a) there are ten free parameters: three ($p_1, z_p, p_2$) define the redshift distribution (Eq. 6), three ($a_1, a_2, E_{p,b}$) define the peak energy distribution (Eq. 15), and four ($q_Y, m_Y, q_A, m_A$) define the $E_p - L_{iso}$ and $E_p - E_{iso}$ correlations (Eqs. 16 and 17). Our constraints are the seven distributions shown in Fig. 9 (including the insets in the top right panel).

In order to find the best fit values and confidence intervals of our parameters, we employed a Markov chain Monte Carlo (MCMC) approach based on the Metropolis-Hastings algorithm (Hastings, 1970). At each step of the MCMC

- we displace each parameter $p_i$ from the last accepted value. The displacement is sampled from a uniform distribution whose maximum width is carefully tuned in order to avoid the random walk remaining stuck in local maxima;

- we compute the Kolmogorov-Smirnov (KS) probability $P_{KS,j}$ of each observed distribution to be drawn from the corresponding model distribution;

- we define the goodness of fit $\mathcal{G}$ of the model as the sum of the logarithms of these KS probabilities, i.e. $\mathcal{G} = \sum_{j=1}^{7} \log P_{KS,j}$;

- we compare $g = \exp(\mathcal{G})$ with a random number $r$ sampled from a uniform distribution within 0 and 1: if $g > r$ the set of parameters is “accepted”, otherwise it is “rejected”.

We performed tests of the MCMC with different initial parameters, to verify that a unique global maximum of $\mathcal{G}$ could be found. Once properly set up, 200,000 steps of the MCMC were run. After removing the initial burn in, the posterior density

7 For parameters corresponding to slopes, like $m_Y$ and $m_A$, we actually displace the corresponding angle $\phi = \arctan(m)$, otherwise a uniform sampling of the displacement would introduce a bias towards high (i.e. steep) slopes.
Figure 11: Marginalized densities of sampled parameters in model (a) (i.e. with correlations and no minimum luminosity). Red lines indicate the means of the marginalized distributions.
distribution of each parameter (and the joint distribution of each couple of parameters) was extracted with the getDist python package\(^8\). The resulting 1D and 2D marginalized distributions are shown in Fig. 11, where red lines indicate the position of the mean of the marginalized density of each parameter. The contours represent the 68% and 95% probability areas of the joint density distributions. The means and 68% probability intervals of the 1D marginalized distributions are summarized in Table 2.a, where the corresponding luminosity function parameters are also reported.

For the solution represented by the mean values in Table 2.a, the minimum luminosity is \(L_{\text{min}} \sim 10^{47} \text{ erg s}^{-1}\). For comparison, we tested case (b) fixing \(L_{\text{min}} = 10^{50} \text{ erg s}^{-1}\). This is the highest minimum luminosity that can be assumed, since the lowest SGRB measured luminosity in the Swift sample considered is \(L = 1.2 \times 10^{50} \text{ erg s}^{-1}\) (D14). Table 2.b summarizes the results of the analysis after 200,000 MCMC steps. The two cases are consistent within one sigma. The best fit luminosity function in case (b) is slightly shallower at low luminosities (i.e. there is a slight decrease in \(\alpha_1\)) than in case (a), and it remains much shallower than in D14 and WP15.

Finally, we tested model (c) performing 200,000 MCMC steps. In this case, there are 11 free parameters: three \((p_1, z_p, p_2)\) for \(\Psi(z)\) and three \((\alpha_1, \alpha_2, E_{p,b})\) for \(\phi(E_p)\) as before, plus three \((\alpha_1, \alpha_2, L_b)\) for the luminosity function (Eq. 18) and two \((T_c, \sigma_{T_c})\) for the intrinsic duration distribution (Eq. 19). Consistently with model (a) and model (b), we assumed two broken power laws for \(\phi(E_p)\) and \(\phi(L)\). The results are listed in Table 2.c. We found that if no correlations are assumed between the spectral peak energy and the luminosity or energy, the luminosity function and the peak energy distributions become peaked around characteristic values. This result is reminiscent of the findings of Shahmoradi and Nemiroff, 2015 who assumed lognormal distributions for these quantities.

**THE RESULTS**

**Luminosity function**

In model (a) we found that the luminosity function is shallow \((\alpha_1 = 0.53^{+0.47}_{-0.14}\) and flatter than 1.0 within the 68% confidence interval) below a break luminosity \(\sim 3 \times 10^{52} \text{ erg s}^{-1}\) and steeper \((\alpha_2 = 3.4^{+0.3}_{-1.7})\) above this characteristic luminosity. The minimum luminosity \(\sim 5 \times 10^{47} \text{ erg s}^{-1}\) is set by the minimum \(E_p\) coupled with the \(E_p - L_{iso}\) correlation parameters (see §3.7). Similar parameters for the \(\phi(L)\) are obtained in model (b), where a minimum luminosity was introduced, thus showing that this result is not strongly dependent on the choice of the minimum luminosity of the \(\phi(L)\).

Relaxing the assumption about the correlations (model (c)), we found that the distributions of the peak energy and luminosity are peaked. However, the 68% confidence intervals of some parameters, common to cases (a) and (b), are larger in case (c). In particular, the slope \(\alpha_1\) of the luminosity function below the break

---

\(^8\) getDist is a python package written by Antony Lewis of the University of Sussex. It is a set of tools to analyse MCMC chains and to extract posterior density distributions using Kernel Density Estimation (KDE) techniques. Details can be found at [http://cosmologist.info/notes/GetDist.pdf](http://cosmologist.info/notes/GetDist.pdf).
Table 2: Summary of Monte Carlo Markov Chain results. C.I. = confidence interval. $E_{\text{peak,b}}$, $L_b$, and $T_c$ are in units of keV, $10^{52}$ erg s$^{-1}$ and s, respectively.

(a) model with correlations and no minimum luminosity

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>68% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>0.53</td>
<td>(0.2, 1)</td>
</tr>
<tr>
<td>$a_2$</td>
<td>4</td>
<td>(1.9, 4.4)</td>
</tr>
<tr>
<td>$E_{\text{peak,b}}$</td>
<td>1400</td>
<td>(880, 2000)</td>
</tr>
<tr>
<td>$m_Y$</td>
<td>0.84</td>
<td>(0.58, 0.88)</td>
</tr>
<tr>
<td>$m_A$</td>
<td>1.1</td>
<td>(0.76, 1.2)</td>
</tr>
<tr>
<td>$q_Y$</td>
<td>0.034</td>
<td>(-0.069, 0.18)</td>
</tr>
<tr>
<td>$q_A$</td>
<td>0.042</td>
<td>(-0.061, 0.13)</td>
</tr>
<tr>
<td>$p_1$</td>
<td>2.8</td>
<td>(0.59, 3.7)</td>
</tr>
<tr>
<td>$z_p$</td>
<td>2.3</td>
<td>(1.7, 3.2)</td>
</tr>
<tr>
<td>$p_2$</td>
<td>3.5</td>
<td>(0.94, 4.4)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.53</td>
<td>(0.39, 1.0)</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>3.4</td>
<td>(1.7, 3.7)</td>
</tr>
<tr>
<td>$L_b$</td>
<td>2.8</td>
<td>(0.91, 3.4)</td>
</tr>
</tbody>
</table>

(b) model with correlations and minimum luminosity

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>68% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>0.39</td>
<td>(-0.15, 0.8)</td>
</tr>
<tr>
<td>$a_2$</td>
<td>3.5</td>
<td>(1.9, 3.7)</td>
</tr>
<tr>
<td>$E_{\text{peak,b}}$</td>
<td>1400</td>
<td>(730, 1700)</td>
</tr>
<tr>
<td>$m_Y$</td>
<td>0.88</td>
<td>(0.61, 0.97)</td>
</tr>
<tr>
<td>$m_A$</td>
<td>1.1</td>
<td>(0.77, 1.2)</td>
</tr>
<tr>
<td>$q_Y$</td>
<td>0.045</td>
<td>(-0.039, 0.17)</td>
</tr>
<tr>
<td>$q_A$</td>
<td>0.043</td>
<td>(-0.037, 0.14)</td>
</tr>
<tr>
<td>$p_1$</td>
<td>3.1</td>
<td>(1.4, 2)</td>
</tr>
<tr>
<td>$z_p$</td>
<td>2.5</td>
<td>(1.9, 3.3)</td>
</tr>
<tr>
<td>$p_2$</td>
<td>3</td>
<td>(0.9, 3.1)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.38</td>
<td>(0.34, 0.98)</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>3</td>
<td>(1.7, 3.2)</td>
</tr>
<tr>
<td>$L_b$</td>
<td>2.3</td>
<td>(0.71, 2.8)</td>
</tr>
</tbody>
</table>

(c) model with no correlations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>68% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>-0.61</td>
<td>(-0.73, -0.41)</td>
</tr>
<tr>
<td>$a_2$</td>
<td>2.8</td>
<td>(2.1, 2.9)</td>
</tr>
<tr>
<td>$E_{\text{peak,b}}$</td>
<td>2200</td>
<td>(1900, 2500)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-0.15</td>
<td>(-1.5, 0.81)</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>2.0</td>
<td>(1.2, 2.8)</td>
</tr>
<tr>
<td>$L_b$</td>
<td>0.63</td>
<td>(0.32, 1.6)</td>
</tr>
<tr>
<td>$T_c$</td>
<td>0.11</td>
<td>(0.084, 0.13)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.91</td>
<td>(0.79, 1.0)</td>
</tr>
<tr>
<td>$p_1$</td>
<td>3.1</td>
<td>(0.51, 4.1)</td>
</tr>
<tr>
<td>$z_p$</td>
<td>2.5</td>
<td>(2.0, 3.3)</td>
</tr>
<tr>
<td>$p_2$</td>
<td>3.6</td>
<td>(1.1, 3.7)</td>
</tr>
</tbody>
</table>
is poorly constrained, although this cannot be steeper than 0.81 (at the 68% confidence level). We believe that the larger uncertainty on the best fit parameters in case (c) is due to the higher freedom allowed by the uncorrelated luminosity function, peak energy distribution, and duration distribution.

Redshift distribution

Figure 12 shows a comparison of our predicted redshift distributions (case (a): red solid line; case (c): orange triple dot-dashed line; mean values adopted) with the following redshift distributions:

- the convolution of the Madau and Dickinson, (2014, hereafter MD14) star formation history (SFH) with the delay time distribution \( P(\tau) \propto \tau^{-1} \) with \( \tau > 20 \text{Myr} \), grey dashed line (the normalization is arbitrary);

- the redshift distribution of NS–NS mergers as predicted by Dominik et al., 2013 (we refer to the standard binary evolution case in the paper) based on sophisticated binary population synthesis, assuming two different metallic-
ity evolution scenarios: high-end (pink solid line) and low-end (pink dotted line);

- the SGRB redshift distribution found by D14, which is obtained convolving the SFH by Hopkins and Beacom, 2006 with a delay time distribution $P(\tau) \propto \tau^{-1.5}$ with $\tau > 20$ Myr, blue dashed line;

- the SGRB redshift distribution found by WP15, which is obtained convolving an SFH based on Planck results (“extended halo model” in Planck Collaboration et al., 2014) with a lognormal delay time distribution $P(\tau) \propto \exp\left[-(\ln \tau - \ln \tau_0)^2 / (2\sigma^2)\right]$ with $\tau_0 = 2.9$ Gyr and $\sigma < 0.2$ (we used $\sigma = 0.1$), cyan dot-dashed line.

The redshift distribution by D14 peaks between $z \sim 2$ and $z \sim 2.5$, i.e. at a higher redshift than the MD14 SFH (which peaks at $z \sim 1.9$). This is due to the short delay implied by the delay time distribution assumed in D14, and because the (Hopkins and Beacom, 2006) SFH peaks at higher redshift than the MD14 SFH. On the other hand, the redshift distribution by WP15 peaks at very low redshift ($\sim 0.8$) and predicts essentially no SGRBs with redshift $z \gtrsim 2$ because of the extremely large delay implied by their delay time distribution.

Assuming the MD14 SFH (which is the most recent SFH available) to be representative, our result using model (a) seems to be compatible with the $P(\tau) \propto \tau^{-1}$ delay time distribution (grey dashed line), theoretically favoured for compact binary mergers. For model (c), the redshift distribution we find seems to be indicative of a slightly smaller average delay with respect to model (a). Since the cosmic SFH is still subject to some uncertainty, and since the errors on our parameters $(p_1, z_p, p_2)$ are rather large, no strong conclusion about the details of the delay time distribution can be drawn.

E$_p$ – L$_{iso}$ and E$_p$ – E$_{iso}$ correlations

Our approach allowed us, in cases (a) and (b), to derive the slope and normalization of the intrinsic E$_p$ – L$_{iso}$ and E$_p$ – E$_{iso}$ correlations of SGRBs. For the E$_p$ – E$_{iso}$ and E$_p$ – L$_{iso}$ correlations of SGRBs, Tsutsui et al., 2013 finds slope values of $0.63 \pm 0.05$ and $0.63 \pm 0.12$, respectively. Although our mean values for $m_Y$ and $m_A$ (Table 1) are slightly steeper, the 68% confidence intervals reported in Tab. 1 are consistent with those reported by Tsutsui et al., 2013. In order to limit the free parameter space, we assumed a fixed scatter for the correlations and a fixed normalization centre for both (see Eq. 14 and Eq. 15). This latter choice, for instance, introduces the small residual correlation between the slope and normalization of the E$_p$ – L$_{iso}$ parameters (as shown in Fig. 11).

Inspection of Fig. 11 reveals another correlation in the MCMC chain between the parameters $q_Y$ and $q_A$ of the E$_p$ – L$_{iso}$ and E$_p$ – E$_{iso}$ correlations. This is expected, as can be seen by taking the difference between Eqs. 17 and 16

$$q_Y - q_A = \log\left(\frac{E_{mA}^{\text{fix}}}{L_{mY}^{\text{fix}}}\right) + 52m_Y - 51m_A. \quad (20)$$

Since $E_{mA}^{\text{fix}}$ and $L_{mY}^{\text{fix}}$ are both proportional to $E_p$, this induces a linear correlation between $q_A$ and $q_Y$. 
THE LOCAL SGRB RATE

The local rate of SGRBs is clearly of great importance for the connection with gravitational wave events to be detected by the advanced interferometers (Advanced LIGO - LIGO Scientific Collaboration et al. 2015; Abbott et al. 2016; Advanced Virgo - Acernese et al. 2015). The work done in Ghirlanda et al., (2016) enabled us to make new (and somewhat surprising) predictions about the local SGRB rate and the prospects for association with GW from compact binary mergers. The local rate $R_0$ of SGRBs, defined as the number of events with a jet pointing towards the Earth per comoving Gpc$^3$ and yr at redshift zero, is found by imposing that the total number of SGRBs in the flux-limited Fermi/GBM sample is equal to the number of events with the corresponding characteristics predicted by the model, namely

$$N_{\text{GBM}}/T_{\text{GBM}} = R_0 \int_0^\infty \left\{ C(z) \int_{E_{p,\text{min}}}^{E_{p,\text{max}}} \phi(E_p) \int_{L_{\text{lim}}(E_p,z,P_{\text{lim}})}^\infty \mathcal{P}(L|E_p) \, dL \, dE_p \right\} \, dz \quad (21)$$

where:

- $N_{\text{GBM}} = 211$ is the number of events in our GBM sample;
- $T_{\text{GBM}} = 7.5$ yr is the time over which the events have been observed;
- the simultaneous sky coverage $\Delta\Omega/4\pi$ in the definition of $C(z)$ (see §3.5) is 70 percent;
- $L_{\text{lim}}$ is the luminosity that corresponds (at a given redshift and for a given $E_p$) to the flux cut $P_{\text{lim}} = 5 \text{ph cm}^{-2}\text{s}^{-1}$ of the sample;
- $\mathcal{P}(L|E_p)$ in model (a) is a lognormal distribution centered at the luminosity $L(E_p)$ given by the $E_p - L_{\text{iso}}$ relation, with $\sigma = 0.2$, while in model (c) it is the luminosity distribution given in Eq. 15.

The integral can be computed both analytically or via the same Monte Carlo approach described in §3.7, in which case it corresponds to the fraction of SGRBs in the synthetic sample for which the photon flux in the Fermi/GBM band is $P > P_{\text{lim}}$. To find out how the uncertainty on the model parameters propagates to the rate estimate, one can simply compute the local rate for (a randomly selected subsample of) the posterior samples from the MCMC, and derive the uncertainty from the resulting distribution of local rates. The result is that both model (a) and model (c) predict a quite low local rate, namely $R_0 = 0.36^{+0.53}_{-0.17} \text{Gpc}^{-3}\text{yr}^{-1}$ for the former and $R_0 = 0.32^{+0.98}_{-0.17} \text{Gpc}^{-3}\text{yr}^{-1}$ for the latter (one sigma uncertainties). This is about one order of magnitude lower than WP15 and D14, which seems daunting for the GW-SGRB association prospects, as can be seen by the corresponding number of events per year expected within the aLIGO detection volume shown in Table 3. A somewhat more promising prediction comes from the fact that the luminosity function is shallower than previous ones, which implies that on average jets

---

9 These estimates are updated with respect to those given in Ghirlanda et al., (2016), where we used only the best fit parameters of the luminosity function and redshift distribution, and we only took into account the uncertainty on a subsample of the parameters. The results are compatible with those given in that paper, but now the uncertainties are estimated in a more rigorous way.
Table 3: Short GRB rates in yr$^{-1}$ (68% errors) within the volume corresponding to different distances: $R =$ limiting distance for binary inspiral detection by aLIGO, averaged over sky location and binary inclination (often dubbed “range”), $D =$ limiting distance for a face–on binary, averaged on sky location. Limiting distances are obtained considering the aLIGO design sensitivity to NS–NS or NS–BH inspirals (top and bottom portions of the table, respectively). If SGRBs have a jet, the correct distance for the comparison is closer to $D$ than to $R$.

<table>
<thead>
<tr>
<th></th>
<th>$R$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS–NS</td>
<td>$\leq 200$ Mpc</td>
<td>$\leq 300$ Mpc</td>
</tr>
<tr>
<td>Model (a)</td>
<td>$0.01^{+0.02}_{-0.004}$</td>
<td>$0.04^{+0.06}_{-0.02}$</td>
</tr>
<tr>
<td>Model (c)</td>
<td>$0.01^{+0.03}_{-0.005}$</td>
<td>$0.04^{+0.11}_{-0.02}$</td>
</tr>
<tr>
<td>NS–BH</td>
<td>$\leq 410$ Mpc</td>
<td>$\leq 615$ Mpc</td>
</tr>
<tr>
<td>Model (a)</td>
<td>$0.10^{+0.15}_{-0.05}$</td>
<td>$0.35^{+0.52}_{-0.17}$</td>
</tr>
<tr>
<td>Model (c)</td>
<td>$0.10^{+0.27}_{-0.06}$</td>
<td>$0.31^{+0.95}_{-0.16}$</td>
</tr>
</tbody>
</table>

are more powerful and thus have a stronger afterglow. This in turn should impact on the rate of orphan afterglows (i.e. afterglows of jets that do not point towards the Earth, see next chapter), but I am still working out the quantitative details.

There is a considerable number of other predictions for the rate of SGRBs within the horizon of GW detectors in the literature. The wide range, extending from 0.1 Gpc$^{-3}$ yr$^{-1}$ to $> 200$ Gpc$^{-3}$ yr$^{-1}$ (e.g. Guetta and Piran, 2005; Guetta and Piran, 2006), can be tested and further constrained by forthcoming GW-SGRB associations (Coward et al., 2014; Branchesi, Ligo Scientific Collaboration, and Virgo Collaboration, 2012). If SGRBs have a jet, one must account for the collimation factor, i.e. multiply the rate by $f_b = \langle (1 - \cos \theta_{\text{jet}})^{-1} \rangle$, and by an additional factor $f_{\text{jet}}$ which represents the fraction of mergers that produce a jet, in order to compare such predictions with the compact binary merger rate.

Fig. 13 shows the rate of SGRBs within a given redshift $z$ predicted by our models. The blue dashed curve is obtained using the formation rate $\Psi(z)$ and luminosity function $\phi(L)$ from WP15, while the results of model (a) are represented by the solid red line, with the yellow area showing the 90% uncertainty.

These curves represent the rate of SGRBs detectable in $\gamma$–rays by current flying instruments. At redshifts as low as those shown in Fig. 13, even bursts populating the lowest end of the luminosity function can be observed above the flux limits of the available GRB detectors (e.g. the Fermi/GBM). The $\Psi(z)$ that we derive (see Fig. 12) rises, below the peak, in a way similar to those adopted in the literature (e.g. D14 and WP15). The lower rates predicted by our models with respect to those of D14 and WP15 are thus due to the flatter $\phi(L)$. 
Figure 13: Event rates within redshift $z$. The solid red line represents the SGRB rate for model (a); the filled yellow area represents the corresponding 90% uncertainty. The results for model (c) are very similar. The SGRB rate according to the best fit model of WP15 is also shown (dashed blue line). The rate of NS-NS mergers based on GW170817 (Abbott et al., 2017a) is shown by the pink hatched region. The vertical stripes represent the ranges of aLIGO for the detection of NS–NS mergers in O2 and O3 (i.e. the second and third scientific runs of the aLIGO detectors, Abbott et al. 2016), and for the design sensitivity. The turquoise rectangle represents the sky-averaged limiting distance for the detection of a face-on NS-NS binary inspiral with the aLIGO design sensitivity.
In Ghirlanda et al., (2016), we derived the luminosity function $\phi(L)$, redshift distribution $\Psi(z)$, and local rate of SGRBs. In a way similar (to some extent) to previous works present in the literature, we fitted the properties of a synthetic SGRB population, described by the parametric $\phi(L)$ and $\Psi(z)$, to a set of observational constraints derived from the population of SGRBs detected by Fermi and Swift. Any acceptable model of the SGRB population must reproduce their prompt emission properties and their redshift distributions. Our approach featured a series of improvements with respect to previous works present in the literature:

- (observer frame) constraints: we extended the classical set of observational constraints (peak flux and – for a few events – redshift distribution) requiring our model to reproduce the peak flux $P$, fluence $F$, peak energy $E_{p, o}$, and duration $T$ distributions of 211 SGRBs with $P_{64} \geq 5\text{ ph s}^{-1}\text{ cm}^{-2}$ detected by the GBM instrument on board the Fermi satellite. The uniform response of the GBM over a wide energy range ($10\text{ keV} – \text{ few MeV}$) ensures a good characterization of the prompt emission spectral properties of the GRB population and, therefore, of the derived quantities, i.e. the peak flux and the fluence;

- (rest frame) constraints: we also required our model to reproduce the distributions of redshift, luminosity, and energy of a small sample (11 events) of Swift SGRBs with $P_{64} \geq 3.5\text{ ph s}^{-1}\text{ cm}^{-2}$ (selected by D14). This sample is 70% complete in redshift and therefore it ensures a less pronounced impact of redshift-selection biases in the results;

- method: we derived our results assuming the existence of intrinsic $E_{p} - L_{iso}$ and $E_{p} - E_{iso}$ correlations in SGRBs (model (a)), similarly to what has been observed in the population of long GRBs. However, since evidence of the existence of such correlations in the population of SGRBs is still based on a limited number of bursts, we also explored the case of uncorrelated peak energy, luminosity and energy (model (c)).

Our main results are as follows:

1. The luminosity function of SGRBs in both models (a) and (c) is shallow, with a low luminosity slope $\alpha_1 \lesssim 0.5$ (68% confidence interval) below a break luminosity $L_b \sim 3 \times 10^{52}\text{ erg s}^{-1}$ and falls steeply above the break. The assumption of a minimum luminosity does not change the qualitative results (as shown with model (b));

2. The redshift distribution of SGRBs $\Psi(z)$ peaks at $z \sim 1.5$ and falls rapidly above the peak. This result is intermediate between those reported in the literature which assume either a constant large delay or a power law distribution favouring small delays. Our $\Psi(z)$ is consistent with the MD14 SFH retarded with a power law delay time distribution $\propto \tau^{-1}$;

3. We estimated the rate of SGRBs as a function of $z$, finding a rather low local rate $R_0 \sim 0.3\text{ Gpc}^{-3}\text{ yr}^{-1}$, which corresponds to a very low number of events

---

**SUMMING UP**

In Ghirlanda et al., (2016), we derived the luminosity function $\phi(L)$, redshift distribution $\Psi(z)$, and local rate of SGRBs. In a way similar (to some extent) to previous works present in the literature, we fitted the properties of a synthetic SGRB population, described by the parametric $\phi(L)$ and $\Psi(z)$, to a set of observational constraints derived from the population of SGRBs detected by Fermi and Swift. Any acceptable model of the SGRB population must reproduce their prompt emission properties and their redshift distributions. Our approach featured a series of improvements with respect to previous works present in the literature:

- (observer frame) constraints: we extended the classical set of observational constraints (peak flux and – for a few events – redshift distribution) requiring our model to reproduce the peak flux $P$, fluence $F$, peak energy $E_{p, o}$, and duration $T$ distributions of 211 SGRBs with $P_{64} \geq 5\text{ ph s}^{-1}\text{ cm}^{-2}$ detected by the GBM instrument on board the Fermi satellite. The uniform response of the GBM over a wide energy range ($10\text{ keV} – \text{ few MeV}$) ensures a good characterization of the prompt emission spectral properties of the GRB population and, therefore, of the derived quantities, i.e. the peak flux and the fluence;

- (rest frame) constraints: we also required our model to reproduce the distributions of redshift, luminosity, and energy of a small sample (11 events) of Swift SGRBs with $P_{64} \geq 3.5\text{ ph s}^{-1}\text{ cm}^{-2}$ (selected by D14). This sample is 70% complete in redshift and therefore it ensures a less pronounced impact of redshift-selection biases in the results;

- method: we derived our results assuming the existence of intrinsic $E_{p} - L_{iso}$ and $E_{p} - E_{iso}$ correlations in SGRBs (model (a)), similarly to what has been observed in the population of long GRBs. However, since evidence of the existence of such correlations in the population of SGRBs is still based on a limited number of bursts, we also explored the case of uncorrelated peak energy, luminosity and energy (model (c)).

Our main results are as follows:

1. The luminosity function of SGRBs in both models (a) and (c) is shallow, with a low luminosity slope $\alpha_1 \lesssim 0.5$ (68% confidence interval) below a break luminosity $L_b \sim 3 \times 10^{52}\text{ erg s}^{-1}$ and falls steeply above the break. The assumption of a minimum luminosity does not change the qualitative results (as shown with model (b));

2. The redshift distribution of SGRBs $\Psi(z)$ peaks at $z \sim 1.5$ and falls rapidly above the peak. This result is intermediate between those reported in the literature which assume either a constant large delay or a power law distribution favouring small delays. Our $\Psi(z)$ is consistent with the MD14 SFH retarded with a power law delay time distribution $\propto \tau^{-1}$;

3. We estimated the rate of SGRBs as a function of $z$, finding a rather low local rate $R_0 \sim 0.3\text{ Gpc}^{-3}\text{ yr}^{-1}$, which corresponds to a very low number of events
per year within the explorable volume of advanced LIGO, as quantified in Table 3.
According to the results described in the preceding chapter, the rate of SGRBs with a jet pointing towards the Earth that can be associated to GW events from NS-NS inspirals should be very low. On the other hand, the same results suggest that SGRB jets are rather energetic on average, so there could be better prospects for the detection of an orphan afterglow associated to such a GW event. For these reasons, in the following months I tried to find a method to set up an optimal strategy for the follow-up of GW events, aimed especially at discovering the orphan afterglow. The idea of the “where and when” method (WAW for short) came as an epiphany when I was half-sleeping on an old, slow bus traveling from the Kotor bay to the Durmitor mountains in Montenegro. The subsequent work resulted in an article, published by the Astrophysical Journal (Salafia et al., 2017b), whose results and methods are described in this chapter.

THE EM FOLLOW-UP PROBLEM

The first detection (The LIGO Scientific Collaboration and the Virgo Collaboration, 2016b) of gravitational waves (GW hereafter) from the inspiral and merger of a black hole binary suddenly turned these fascinating, theoretical objects into real astronomical sources.

When such a compact binary coalescence is detected, analysis of the GW signal and comparison with carefully constructed templates of the waveform (The LIGO Scientific Collaboration and the Virgo Collaboration, 2016a; Abbott et al., 2016b) enables the extraction of precious information about the parameters of the binary and of the remnant. The identification of an electromagnetic (EM) counterpart would increase further the scientific outcome of the detection, e.g. by enabling the identification of the host galaxy, by providing hints about the environment surrounding the merger, by constraining theoretical models of EM counterparts and by reducing degeneracies in the GW extrinsic parameter space (Pankow et al., 2016).

Several observatories, covering a large fraction of the EM spectrum, recently developed dedicated programs for the EM follow-up of GW events. The present main limitation for the detection of a possible EM counterpart is the large uncertainty on the sky localization of the GW source (see e.g. Singer et al. 2014; Berry et al. 2014). The problem might be alleviated by targeting bright galaxies within the localization uncertainty region (Nuttall and Sutton, 2010; Abadie et al., 2012), and the selection of target galaxies can also take into account the sky-position-conditional posterior distribution of the source luminosity distance (Hanna, Mandel, and Vousden, 2013; Nissanke, Kasliwal, and Georgieva, 2013; Gehrels et al., 2016; Singer et al., 2016a). The aim of this work is to propose an additional way to use information encoded in the GW signal to optimize the follow-up strategy.
for each single event, namely to combine posterior distributions of the compact binary parameters and available models of the EM emission to predict the best timing for the observation of different parts of the GW skymap. Such an approach can be applied in cases when a model of the expected EM counterpart is available, and it is especially useful when the lightcurve predicted by the model depends on (some of) the compact binary parameters.

The first electromagnetic follow-ups

The observation campaigns that followed up the first detections of GWs were very extensive. Hundreds of square degrees within the GW sky localization were covered by wide-field telescopes (Abbott et al., 2016a). Target areas were selected in order to maximize the contained GW source posterior sky position probability, incorporating telescope visibility constraints (e.g. Kasliwal et al., 2016). In some cases, models of the expected EM counterpart emission were used to estimate the optimal search depth (e.g. Soares-Santos et al., 2016); other searches combined the posterior sky position probability map with the areal density and luminosity of nearby galaxies to select the best target fields (e.g. Evans et al., 2016; Díaz et al., 2016). Observations were concentrated during the first days after the events and repeated weeks to months later to search for both rapid and slowly evolving possible counterparts.

Candidate EM counterparts

It is not clear whether an EM counterpart should be expected in the case of a binary of black holes (BH-BH), due to the unlikely presence of matter surrounding the binary (but see Yamazaki, Asano, and Ohira 2016; Perna, Lazzati, and Giacomazzo 2016; Loeb 2016); on the other hand, if the merger involves a black hole and a neutron star (BH-NS) or two neutron stars (NS-NS), there are solid reasons to believe that EM emission should take place. The most popular mechanisms for such an emission in both BH-NS and NS-NS cases include prompt (gamma-ray) and afterglow (panchromatic) emission from a short gamma-ray burst (SGRB) jet, and “macronova” (optical/infrared) emission from ejecta launched during and after the merger, powered by the decay of unstable heavy nuclei resulting from r-process nucleosynthesis taking place within the neutron-rich ejecta during the early expansion phase.

Many other promising EM counterparts have been proposed, e.g. the long lasting radio transient (Nakar and Piran, 2011) arising from the deceleration of the dynamical ejecta due to interaction with the interstellar medium (ISM), the jet cocoon emission (Lazzati et al., 2017b; Gottlieb, Nakar, and Piran, 2018) or the spindown-powered emission described by Siegel and Ciolfi, (2016) in the case when a (meta-)stable neutron star is left after the merger. To keep the discussion as simple as possible, in this chapter I will only consider the (Optical and Radio) SGRB afterglow and the dynamical ejecta macronova as examples, leaving the possibility to apply the present approach to other EM counterparts to future works.
The SGRB afterglow

The detectability of the SGRB prompt emission depends crucially on the jet viewing angle $\theta_v$, i.e. the angle between the jet axis and our line of sight. If the viewing angle is larger than the jet half opening angle $\theta_{\text{jet}}$ (in other words, if the jet points away from the Earth), the prompt emission flux received by an observer on Earth is severely suppressed (e.g. Salafia et al., 2016) due to relativistic beaming (by the compactness argument, the bulk Lorentz factor in GRB jets must be comparable to or larger than one hundred – e.g. Lithwick and Sari 2001 – and estimates based on observations are sometimes even larger than a thousand – as in the short burst GRB090510, see Ghirlanda, Ghisellini, and Nava 2009; Ackermann et al. 2010). Since the typical half opening angle $\theta_{\text{jet}}$ is somewhere between $5^\circ$ and $15^\circ$ (e.g. Berger, 2014a), the prompt emission goes undetected in the majority of cases (for an isotropic population, the probability that $\theta_v < 15^\circ$ is less than 2 percent). Soon after producing the prompt emission, the jet starts interacting significantly with the ISM, and a shock develops (Meszaros and Rees, 1996). Electrons in the shocked ISM produce synchrotron radiation, giving rise to a fading afterglow (observed for the first time by Beppo-SAX, Costa et al., 1997). Since the consequent deceleration of the jet reduces the relativistic beaming, an off-axis observer (who missed the prompt emission) could in principle detect the afterglow before it fades (Rhoads, 1997): in this case, the afterglow is said to be orphan. No convincing detection of such a transient has been claimed to date, consistently with predictions for current and past surveys (Ghirlanda et al., 2015; Ghirlanda et al., 2014), but future deep surveys (e.g. MeerKAT in the Radio – Booth et al. 2009, LSST in the Optical – Ivezić et al. 2008, eROSITA in the X-rays – Merloni et al. 2012) are anticipated to detect tens to thousands of such events per year.

Given the large uncertainty on the expected rate of NS-NS and BH-NS detections by the aLIGO and Advanced Virgo facilities in the near future (LIGO Scientific Collaboration et al., 2010; Kim, Yoon, and Koo, 2015; Dominik et al., 2015; Mink and Mandel, 2016; Abbott et al., 2016) and the rather low expected fraction of GW events with an associated SGRB jet pointing at the Earth (Ghirlanda et al., 2016; Patricelli et al., 2016; Wanderman and Piran, 2014; Metzger and Berger, 2011), the inclusion of orphan afterglows as potential counterparts is of primary importance to test the SGRB-compact binary coalescence connection.

The dynamical ejecta macronova

Despite the idea dates back to almost twenty years ago (Li and Paczyński, 1998), the understanding of the possible macronova emission following a compact binary merger has been expanded relatively recently, as a result of the combined effort of researchers with expertise in a wide range of areas. A non-exhaustive list of the main contributions should include:

- numerical simulations of the merger dynamics (relativistic simulations by many groups using different approaches – e.g. Dietrich et al. 2017; Ciolfi et al. 2017; Radice et al. 2016; Sekiguchi et al. 2016; Ruiz et al. 2016; Giacomazzo et al. 2015; Bauswein and Stergioulas 2015; Just et al. 2016; East et al. 2015; Sekiguchi et al. 2015; Wanajo et al. 2014; Kiuchi et al. 2014; Tanaka
and Hotokezaka 2013; Rezzolla et al. 2010 – and non relativistic simulations, especially by Stephan Rosswog and collaborators – e.g. Rosswog et al. 2014);

- studies to assess the efficiency of r-process nucleosynthesis and the consequent heating rate due to heavy element decay in the various ejecta (Freiburghaus, Rosswog, and Thielemann, 1999; Rosswog, Freiburghaus, and Thielemann, 2000; Korobkin et al., 2012; Wanajo et al., 2014; Lippuner and Roberts, 2015; Hotokezaka et al., 2015; Eichler et al., 2016; Rosswog et al., 2016);

- atomic structure modeling which revealed the role of lanthanides in the ejecta opacity evolution (Kasen, Badnell, and Barnes, 2013);

- simulations including neutrino physics to model the neutrino-driven wind and the associated macronova (Dessart et al., 2008; Martin et al., 2015; Perego, Yasin, and Arcones, 2017);

Results (especially for the dynamical ejecta) from various research groups begin to converge, and the dependence of the emission features on the parameters of the binary is in the process of being understood. Both analytical and numerical models capable to predict the lightcurve have been developed recently (Barnes and Kasen, 2013; Grossman et al., 2013; Kawaguchi et al., 2016; Dietrich and Ujevic, 2017; Barnes et al., 2016; Rosswog et al., 2016). The emission from the dynamical ejecta is generally thought to be isotropic, which is an advantage with respect to the SGRB afterglow (which is instead beamed) from the point of view of the EM follow-up. The energy reservoir is the ejected mass \( M_{\text{ej}} \), which depends most prominently on the mass ratio \( q = M_1/M_2 \) of the binary and on the neutron star compactness, which in turn reflects the mass of the neutron star and its equation of state (EoS).

Exciting claims of the detection of possible macronova signatures in the afterglows of few short GRBs (Tanvir et al., 2013; Yang et al., 2015; Jin et al., 2015; Jin et al., 2016) are in the process of being tested by intensive observational campaigns. All this makes the macronova emission an extremely interesting candidate EM counterpart.

Outline of this chapter

In §4.2.1 I introduce the idea of a follow-up strategy as a collection of observations that partially fill a “search volume” (search sky area \( \times \) typical transient duration), stressing that the GW “skymap” (the sky position probability density) gives information about where to observe, but not about when. In §4.2.1.1 I show how a priori information about the EM counterpart can be used to quantitatively define how likely the detection of the EM emission is if the observation is performed at time \( t \), thus providing some information about how to explore the temporal dimension of the “search volume”. In §4.2.2 I suggest that the same approach can be extended to use a posteriori information extracted from the GW signal, provided that we have a way to link the properties of the inspiral to those of the EM counterpart (as shown in §4.2.2.2). In §4.2.2.4 I go one step further by introducing the idea that the information on the inspiral parameters that we can extract from the GW signal has a dependence on sky position, and thus the clues (that we obtain from the GW signal analysis) about when to observe can also depend on the sky
position. In §4.3 I introduce a method to extract such information from the “posterior samples” obtained from the analysis of a GW signal, and in §4.5 I apply it to a synthetic example to show how it optimizes the follow-up strategy. Finally, I discuss the results in §4.6 and I draw some conclusions in §4.7.

WHERE AND WHEN TO LOOK

A sketch of the design of a follow-up observation strategy

A short time after the detection of a compact binary coalescence signal, the LIGO Scientific Collaboration and Virgo Collaboration share information about the event with a network of astronomical facilities interested in the EM follow-up. The most fundamental piece of information for the follow-up is the so-called “skymap”, i.e. the posterior sky position probability density, which I denote as \( P(\alpha | S) \). It represents the probability per unit solid angle that the source is at sky position \( \alpha \), say \( \alpha = (\text{RA}, \text{Dec}) \), given the GW signal \( S \) detected by the interferometers (\( S \) here represents all information contained in the strain amplitudes measured by all interferometers in the network). In what follows, I will most often call this probability density “skymap probability”. Imagine the EM counterpart appears at the GW position right after the event and never turns off. Assume that it can be found by comparison with previously available images of the sky, and that it can be easily identified by its spectrum or by another method. An ultra-simplified sketch of the obvious follow-up strategy would then be the following:

1. find the smallest sky area \( A_\omega \) containing a large fraction \( \omega \) (say \( \omega = 90\% \)) of the skymap probability \( P(\alpha | S) \);
2. divide such area into patches of size \( A_{\text{FoV}} \) corresponding to the field of view of the instrument;
3. observe the patches in decreasing order of skymap probability\(^1\);
4. for each patch:
   a) identify the new sources by comparison with archive images;
   b) perform a set of operations, including e.g. cross-matching with catalogues and spectral characterization, to discard known variable sources and unrelated transients in order to identify the counterpart.

The expected EM counterparts are transients, thus a first modification to the above sketch must take into account the time constraints coming from our \textit{a priori} knowledge of the transient features. If we have a physical or phenomenological model of the transient and we have some hint about the distribution of the parameters of such model, we can construct a prior probability \( P(F(t) > F_{\text{lim}}) \) that the transient flux (in a chosen band) \( F \) is above some limiting flux \( F_{\text{lim}} \) at a given time \( t \) after the GW event. Hereafter, I will call such quantity “\textit{a priori} detectability”.

\(^1\) to keep the discussion as simple as possible, I am neglecting the limitations due to observing conditions.
Figure 14: Graphical representation of the follow-up strategy as a “volume filling” problem. \( A_f \) represents the sky region that contains a fraction \( \omega \) of the sky position probability. The “search volume” is defined as the set of points \( V_\omega = \{ (\alpha, t) \mid \alpha \in A_\omega \text{ and } t \in [0, T_{tr}] \} \). Observations are sets that intersect the search volume, defined by a field of view \( A_{\text{FoV}}(\alpha_i) \) centered about sky position \( \alpha_i \), an exposure time \( T \) and an observation time \( t_{\text{obs},i} \) so that \( V_{\text{obs},i} = \{ (\alpha, t) \mid \alpha \in A_{\text{FoV}}(\alpha_i) \text{ and } t \in (t_{\text{obs},i}, t_{\text{obs},i} + T) \} \). Observations made by the same instrument cannot overlap on the time axis (unless the instrument can see more than one field at the same time). For the detection to be successful, the EM counterpart must be located within one of the \( A_{\text{FoV}}(\alpha_i) \) and its lightcurve must be above the detection threshold during the corresponding exposure time.

The probability of detecting the transient at time \( t \) by observing a sky position \( \alpha \) with an instrument with field of view \( A_{\text{FoV}} \) and limiting flux \( F_{\text{lim}} \) is then

\[
P(\text{det} \mid t, \alpha, \text{FoV}) \sim A_{\text{FoV}} P(\alpha \mid S) \times P(F(t) > F_{\text{lim}})
\]  

(22)

where I am assuming a relatively small field of view in order to consider \( P(\alpha \mid S) \) constant over its area. This is nothing more than saying that the best place to look for the transient is the point of maximum skymap probability, at the time of highest a priori detectability. The probability of detection decreases both moving away from the point of maximum skymap probability and observing at a time when \( P(F(t) > F_{\text{lim}}) \) is smaller.

Let me work in the simplifying assumption that all observations have the same exposure \( T \) and the same limiting flux \( F_{\text{lim}} \). Let me denote by \( T_{tr} \) the most conservative (i.e. largest) estimate of the transient duration, and let me define the “search volume” \( V_\omega = A_\omega \times T_{tr} \) (I refer to this set as a “volume” because it is 3-dimensional, even though the dimensions are solid angle \( \times \) time – see Figure 14). The follow-up strategy can then be thought of as an optimization problem, where one wants to (partially) fill the search volume \( V_\omega \) with \( N \) observations \( V_{\text{obs},i} = A_{\text{FoV}}(\alpha_i) \times (t_{\text{obs},i}, t_{\text{obs},i} + T) \) (where \( \alpha_i \) and \( t_{\text{obs},i} \) are respectively the sky coordinates of the center of the field of view and the starting time of the \( i \)-th observation) in order to maximise the detection probability \( P(\text{det} \mid \text{strategy}) \), which can be written as

\[
P(\text{det} \mid \text{strategy}) = \sum_{i=1}^{N} \int_{A_{\text{FoV}}(\alpha_i)} P(\alpha \mid S) \, d\alpha \times P(F(t_{\text{obs},i}) > F_{\text{lim}})
\]  

(23)

with the constraint \( NT \leq T_{tr} \) (see Figure 14).
The above paragraphs are essentially a formal description of the most basic follow-up strategy one can think of, which can be reduced to the principle “try to arrange the observations in order to cover the largest possible fraction of the GW skymap around the time when the flux is expected to be high enough for a detection”. In this approach, the proper construction of the a priori detectability $P(F(t) > F_{\text{lim}})$ is the key: it defines the time span within which the observations are to be performed, while the posterior sky position probability density defines the search area.

*How to construct the a priori detectability*

In order to construct the a priori detectability $P(F(t) > F_{\text{lim}})$, one must assume some prior probability density of the model parameters. Let me consider a simple, illustrative example. First, I construct a synthetic population of NS-NS inspirals whose properties roughly reproduce those expected for the population detected by Advanced LIGO; then I associate to each of them a jet afterglow and a macronova, under some assumptions. The detectable fraction of lightcurves in a given band, at a given time, will then constitute our estimate of the a priori detectability for this particular case. For the jet afterglow, I assume\(^2\) that all SGRB jets have an isotropic kinetic energy $E_K = 10^{50}$ erg and a half-opening angle $\theta_{\text{jet}} = 0.2$ radians (11.5 deg), and that they are surrounded by a relatively tenuous interstellar medium with constant number density $n_{\text{ISM}} = 0.01$ cm\(^{-3}\). I fix the microphysical parameters\(^3\) so that the only remaining parameters needed to predict the afterglow lightcurve of the SGRB are the distance $d_L$ and the viewing angle $\theta_v$. I will link the viewing angle to the binary orbit inclination, and the distance will be obviously set equal to that of the binary. Assuming two opposite jets launched perpendicular to the binary orbital plane, we have that

$$\theta_v(t) = \begin{cases} 
\pi - t & 0 \leq t < \pi/2 \\
\pi/2 - t & \pi/2 \leq t < \pi 
\end{cases} \tag{24}$$

where $t$ is the angle between the normal to the orbital plane and the line of sight.

For the dynamical ejecta macronova, I evaluate the disk mass and the ejecta velocity using the fitting formulas of Dietrich and Ujevic, (2017), and I use them as inputs to compute the lightcurve following Grossman et al., (2013) (using a constant grey opacity $\kappa = 10$ cm\(^2\) g\(^{-1}\)), assuming a blackbody spectrum with effective temperature equal to that of the photosphere. The input compact binary parameters in this case are the masses $M_1$ and $M_2$. To determine the compactness and the baryon mass of the neutron stars, which are necessary to associate the dynamical ejecta mass $M_{\text{ej}}$ and velocity $v_{\text{ej}}$ to the merger through the fitting formulas of Dietrich and Ujevic, (2017), I assume the H\(_4\) equation of state (Lackey, Nayyar, and Owen, 2006; Glendenning and Moszkowski, 1991) which has a mid-range stiff-

\(^{2}\) these are typical reference values for short GRBs, though they suffer from the still limited number of reliable measurements available.

\(^{3}\) I refer here to the standard synchrotron afterglow model, and I set the microphysical parameters $p = 2.5$, $\epsilon_e = 0.1$ and $\epsilon_B = 0.01$. Such values, typical of Long GRBs (e.g. Panaitescu and Kumar, 2002; Ghisellini et al., 2006; Ghirlanda et al., 2015), seem to be representative for SGRBs as well (Fong et al., 2015), despite the much smaller sample of broadband lightcurves available.
ness among those which are compatible with the observational constraints (Özel and Freire, 2016).

First, we need to derive the proper distributions of distance and orbital plane inclination of the inspiral population detected by our interferometer network. For simplicity, I neglect the dependence of the network sensitivity on sky position and on the binary polarization angle $\psi$, and I assume that the maximum luminosity distance $d_{L,\text{max}}$ out to which a NS-NS inspiral can be detected depends only on the binary plane inclination $\iota$ with respect to the line of sight, namely

$$d_{L,\text{max}}(\iota) = d_{L,\text{max}}(0) \sqrt{\frac{1}{8} \left( 1 + 6 \cos^2 \iota + \cos^4 \iota \right)}$$

(25)

where $d_{L,\text{max}}(0)$ is the maximum luminosity distance out to which our network can detect a face-on inspiral. This expression accounts for the fact that gravitational radiation from a compact binary inspiral is anisotropic (Schutz, 2011). Assuming that NS-NS mergers are uniformly distributed in space and have isotropic orientations, their distance and inclination distributions are $P(d_L) \propto d_L^2$ and $P(\iota) \propto \sin \iota$.

By the above assumptions, the probability that a binary with luminosity distance $d_L$ and inclination $\iota$ is detected is

$$P(\text{det} | d_L, \iota) = \begin{cases} 1 & \text{if } d_L < d_{L,\text{max}}(\iota) \\ 0 & \text{otherwise} \end{cases}$$

(26)

By Bayes’ theorem, the probability distribution of distance and inclination of a detected NS-NS inspiral is then $P(d_L, \iota | \text{det}) \propto P(\text{det} | d_L, \iota) \times P(d_L) \times P(\iota)$, which gives

$$P(d_L, \iota | \text{det}) \propto \begin{cases} d_L^2 \sin \iota & \text{if } d_L < d_{L,\text{max}}(\iota) \\ 0 & \text{otherwise} \end{cases}$$

(27)

The corresponding probability distribution of inclination for detected inspirals (which is obtained by marginalisation of Eq. 27 over $d_L$) is then the well known

$$P(\iota | \text{det}) = 7.6 \times 10^{-2} \left( 1 + 6 \cos^2 \iota + \cos^4 \iota \right)^{3/2} \sin \iota$$

(28)

For what concerns the distribution of masses $M_1$ and $M_2$, I simply assume a normal distribution with mean $1.35 M_\odot$ and sigma $0.1 M_\odot$ for both of them, which reproduces the mass distribution of known galactic NS-NS binaries (Özel and Freire, 2016).

Now, to construct the a priori detection probability $P(F(t) > F_{\text{lim}})$ I adopt the following Monte Carlo approach:

1. I construct a synthetic population of $N$ inspirals sampling distances and inclinations from $P(d_L, \iota | \text{det})$ and masses $M_1$ and $M_2$ from the assumed normal distribution;

2. I compute the flux $F_i = F_A(t, d_{L,i}, \theta_v(t_i))$ of the jet afterglow or $F_i = F_M(t, d_{L,i}, M_{1,i}, M_{2,i})$ of the macronova in the chosen band for each sample;

3. I estimate $P(F(t) > F_{\text{lim}})$ as the fraction of the $F_i$’s that exceed $F_{\text{lim}}$. 
Figure 15: “A priori detectability”, i.e. a priori probability that the EM counterpart of a compact binary inspiral is detected if the observation is performed at a time $t$ after the merger, for observations in Radio at 1.4 GHz, in Infrared (IR) in the $J$ band, and in Optical in the $r$ band, with limiting fluxes of 0.5 mJy in Radio and 22.4 AB magnitude in IR and Optical. The Radio and Optical probabilities account only for the jet afterglow, while the IR probability accounts only for the dynamical ejecta macronova. Based on a series of simplifying assumptions, see §4.2.1.1.
Figure 15 shows the a priori detectability computed with the above method for \(d_{\text{L, max}}(0) = 100\) Mpc (which corresponds roughly to the sky-position averaged aLIGO range for an optimally oriented NS-NS inspiral with the sensitivity of the first Advanced LIGO observing run, see Abbott et al., 2016) for Radio, Infrared and Optical observations (see the caption for details). The sensitivity of Optical observations has been chosen to match the limiting magnitude of the VST follow-up of GW150914, see Abbott et al. 2016a. The flux of the jet afterglow has been computed using BOXFIT v. 1.0 (Eerten, Horst, and MacFadyen, 2012).

A more accurate a priori detectability would require us to use astrophysically motivated priors on the other model parameters, such as the kinetic energy \(E_K\), the ISM number density \(n_{\text{ISM}}\), etc. Moreover, the actual intrinsic mass distribution of neutron stars that merge within the frequency band of GW detectors might differ significantly from the assumed one. The curves shown in Fig. 15, thus, must be taken as illustrative.

The definitely higher detectability of the macronova is due to the fact that its emission is assumed to be isotropic, while the jet is fainter for off-axis observers.

Two steps further: how to improve the strategy using posterior information about other parameters of the binary

The full posterior probability density in parameter space

Parameter estimation techniques applied to a compact binary coalescence signal \(S\) result in a posterior probability density \(P(\xi | S)\), where \(\xi \in \mathbb{R}^n\) is a point in the \(n\)-dimensional parameter space. The “skymap probability”, i.e. the posterior sky position probability density \(P(\alpha | S)\), is essentially the \(P(\xi | S)\) marginalised over all parameters but the sky position. Much more information is contained in the full posterior probability density, though, and some of it can be used to improve the design of the EM follow-up strategy.

Relevant parameters in our case

In §4.2.1.1 we already made use of two extrinsic parameters of the compact binary inspiral which are relevant for the SGRB afterglow and the macronova, namely the luminosity distance \(d_L\) and the binary inclination\(^5\) \(i\).

Recent works based on numerical simulations of NS-NS and BH-NS mergers (e.g. Foucart, 2012; Giacomazzo et al., 2013; Hotokezaka and Piran, 2015; Kawaguchi et al., 2016; Dietrich and Ujevic, 2017) seem to indicate that the amount of matter in the remnant disk and in the dynamical ejecta, plus some other properties of the latter such as the velocity profile, depend in a quite simple way (once an equation of state is assumed) on the parameters of the binary prior to the merger, especially the masses \(M_1\) and \(M_2\) and the effective spin \(\chi_{\text{eff}}\) of the black hole in the BH-NS

\(^4\)This translates into mid-range values of the corresponding \(M_{\text{ej}}\) and \(v_{\text{ej}}\).

\(^5\)the inclination of the orbital plane with respect to the line of sight can also be relevant for the neutrino-driven wind (Martin et al., 2015) and the disk wind (Kasen, Fernandez, and Metzger, 2014) macronovas, due to their axial geometry and to the possibility that the dynamical ejecta act as a “lanthanide curtain” obscuring their optical emission (Rosswog et al., 2016) if the binary is observed edge-on.
In order to predict the lightcurve of the SGRB and/or of the dynamical ejecta macronova associated to the merger:

1. the luminosity distance $d_L$ and the associated redshift $z$;
2. the orbital plane inclination $\iota$ with respect to the line of sight;
3. the component masses $M_1$ and $M_2$;
4. the effective spin $\chi_{\text{eff}}$ of the black hole in the BH-NS case.

Figure 16 represents a sketch of how the above parameters influence the properties of the SGRB afterglow and the dynamical ejecta macronova associated to the merger. The same approach can be adopted to link the properties of other EM counterparts (such as the long lasting radio transient described by Nakar and Piran 2011 or the X-ray spindown-powered transient described by Siegel and Ciolfi 2016) to those of the binary, whose distributions can be constrained by the GW signal.

In §4.5, I will use the fitting formulas provided in Dietrich and Ujevic, (2017) to compute the posterior ejecta mass distribution associated to the example NS-NS inspiral treated in that section. I will refrain from deriving the SGRB jet energy from disk mass as suggested in Fig. 16, though, because that would require a
detailed discussion about the proper disk mass energy conversion efficiency to be used, which is outside the scope of this work. I leave such a discussion to a future work.

*A posteriori detectability*

In §4.2.1 I introduced the idea of “a priori detectability” \( P(F(t) > F_{\text{lim}}) \), which can be regarded as the basic tool to set the timing of observations for the EM follow-up if no specific information about the source is available. Once the GW signal \( S \) is observed, information it carries can be used to construct a *posterior* probability \( P(F(t) > F_{\text{lim}} | S) \) to better plan such observations. I will call it “a posteriori detectability”. If the *a priori* detectability \( P(F(t) > F_{\text{lim}}) \) is constructed using the *prior* distributions of the parameters of the EM transient model, the *a posteriori* detectability \( P(F(t) > F_{\text{lim}} | S) \) is obtained exactly the same way (as exemplified in §4.2.1.1), but using the *posterior* distributions of the relevant parameters.

*Detectability maps*

Several parameters of a compact binary inspiral are degenerate to some degree, i.e. the same signal \( S \) can be produced by different combinations of the parameter values. These combinations, though, are not just uniformly distributed in some subset of the parameter space, but rather they follow fundamental relations which depend both on the nature of the source (the binary inspiral) and on the properties of the detector network (the locations and orientations of the interferometers, their antenna patterns, the noise power spectrum). In particular, distance, inclination, polarization angle, chirp mass and sky position of the binary share a certain degree of degeneracy: the same signal \( S \) can be produced by different combinations of values of these parameters and different realizations of the detector noises, which is the obvious reason why the sky position uncertainty is so large. For this reason, if we restrict the posterior probability density in parameter space to a certain point of the skymap, i.e. we take the *sky-position-conditional* posterior distribution of the physical parameters of the binary, in principle it will depend on the chosen sky-position. Knowing the sky-position-conditional posterior probability distribution \( P(d_{\text{L}}, i, M_1, M_2, ... | \alpha, S) \) of the relevant binary parameters at sky position \( \alpha \), we can thus derive the corresponding distribution of the properties of the EM counterpart at that particular sky position, which means that we can construct a sky-position-conditional posterior detectability \( P(F(t) > F_{\text{lim}} | \alpha, S) \) which can be used as the basis of the EM follow-up strategy. I call this quantity “detectability map”.

*Recap*

It is useful to summarize here the steps of increasing complexity that led us to the definition of the detectability maps:

- I started by assuming an unrealistic model of the EM counterpart: a source that turns on at the GW time and never turns off. In this case, no timing information is needed for the follow-up strategy, which simply consists of
scanning the localization uncertainty area, starting from the most probable sky location, until the source is found;

- If a model of the counterpart is available and prior distributions of the model parameters can be assumed (based on available astrophysical data or on an educated guess), the “a priori detectability” $P(F(t) > F_{\text{lim}})$ can be constructed, as shown in §4.2.1.1. This is the best follow-up timing information that can be constructed based on priori knowledge only;

- After a GW signal $S$ is detected and parameter estimation has been performed, prior distributions of the model parameters can be (partly) replaced with posterior distributions derived from the signal: the “a posteriori detectability” $P(F(t) > F_{\text{lim}} | S)$ can be constructed. This exploits information contained in the GW signal, but it is still independent of the sky position;

- If the counterpart is assumed to be located at a certain sky position, the corresponding sky-position-conditional posterior distributions can be used in place of the full posterior distributions. Indeed, given a signal $S$, compact binary inspiral parameters compatible with $S$ and a particular sky position are in general different from those compatible with $S$ and another sky position. By varying the assumed sky position on a grid that covers the whole skymap, one can then construct the “detectability map” $P(F(t) > F_{\text{lim}} | \alpha, S)$.

Let me now introduce a method to compute the detectability maps and apply it to a practical example.

HOW TO CONSTRUCT AND HOW TO USE THE DETECTABILITY MAPS

Extraction of the sky-position-conditional posterior distributions using a simple method based on “inverse distance weighting”

The extraction of the sky-position-conditional posterior distribution requires some multi-dimensional kernel density estimation (KDE) technique, to be applied to the posterior samples obtained from a parameter estimation pipeline run on the gravitational wave signals recorded by the detectors. Since the aim of this work is to propose a new approach in the design of the EM follow-up, rather than to discuss the technical subtleties of such multi-dimensional KDE, I adopt the following simple and intuitive method, which can be replaced with a more accurate one in a possible application of our approach to a real case.

Our simplified method to extract the sky-position-conditional posterior distribution $P(q | \alpha, S)$ of a quantity $q$ at sky position $\alpha$ is based on the concept of “inverse distance weighting” (Shepard, 1968): I assume that each posterior sample $\{\alpha_i, q_i, d_{L,i}, t_i, \ldots\}$ contributes to the $P(q | \alpha, S)$ with a weight which is a decreasing function of the angular distance $\delta(\alpha, \alpha_i)$ between the posterior sample and the sky position $\alpha$. In particular, I assign a Gaussian weight to each posterior sample

$$w_i \propto \exp \left[ -\frac{1}{2} \left( \frac{\delta(\alpha, \alpha_i)}{\sigma(\alpha)} \right)^2 \right]$$

(29)
where the bandwidth $\sigma(\alpha)$ is taken as

$$\sigma(\alpha) = \sqrt{\frac{\sum_{i=1}^{N} [\delta(\alpha, \alpha_i) - \langle \delta \rangle]^2}{N-1}} \times N^{-1/5} \tag{30}$$

where $\langle \delta \rangle$ is the arithmetic mean of the $\delta(\alpha, \alpha_i)$. The normalization of the weights is given by $\sum_{i=1}^{N} w_i = 1$.

The ideas behind this method are simply that the closer the posterior sample is to sky-position $\alpha$, the more it contributes to the conditional posterior distribution at that sky position, and that the influence of the posterior sample decreases as a Gaussian with increasing angular distance. The choice of the bandwidth (Eq. 30) is just “Silverman’s rule of thumb” (Silverman, 1982) for Gaussian KDE in a one dimensional parameter space (namely, the angular distance space).

The mean of $q$ at sky position $\alpha$ is thus computed as

$$\langle q \rangle_\alpha = \sum_{i=1}^{N} w_i q_i \tag{31}$$

and similarly the variance

$$\text{Var}_\alpha(q) = \left( \sum_{i=1}^{N} w_i q_i \right)^2 - \sum_{i=1}^{N} w_i q_i^2 \tag{32}$$

More generally, the sky-position-conditional posterior distribution of $q$ at sky position $\alpha$ is approximated as

$$P(q | \alpha, \delta) \sim \sum_{i=1}^{N} w_i K \left( \frac{q - q_i}{\sigma_q} \right) \tag{33}$$

where $K(x)$ is some kernel function, and $\sigma_q$ is its bandwidth.

In the next section, I describe tests I performed to ensure that the above method yields consistent results. As one might expect, the results are accurate in sky regions where the distribution of posterior samples is sufficiently dense.

**Testing our inverse-distance-weighting-based method of sky-position-conditional density estimation**

*Test 1: reconstruction of the position-conditional mean and standard deviation of a known distribution*

As a first test, I constructed a set of 10000 mock posterior samples whose underlying probability distribution in parameter space has a known analytic form. The parameter space is 4-dimensional, the parameters being right ascension RA, declination Dec, luminosity distance $d_L$ and a fourth quantity $q$ with no physical meaning. The RA, Dec and $d_L$ parameters are independent. The right ascension is normally distributed with mean 12h and sigma 30min; the declination is uniformly distributed between $-7^\circ30'$ and $+7^\circ30'$; the luminosity distance is normally distributed with mean 100 Mpc and sigma 20 Mpc; the distribution of quantity $q$
depends on right ascension: its position-conditional distribution is normal, with mean and sigma given by

\[
\begin{align*}
\mu_q(RA) &= \sin^2 \left( 8\pi \left( \frac{RA}{12h} - 1 \right) \right) \\
\sigma_q(RA) &= \frac{1}{5} + \frac{2}{3} \left( \frac{RA}{12h} - 1 \right) \tag{34}
\end{align*}
\]

The 90% position probability area of the posterior samples is about 435 deg^2 wide and its shape is approximately a rectangle extending in right ascension from 11 to 13 hours and in declination from −7°5′ to +7°5′. Figure 17 shows the reconstructed position-conditional means and standard deviations computed in a set of points along the Dec = 0° axis. Both moments are reconstructed with an acceptable accuracy within the 90% sky position probability area.

Test 2: comparison with the “Going the Distance” study

The “Going the Distance” study (Singer et al., 2016a, GTD hereafter) and especially the related Supplement (Singer et al., 2016b) represent an important practical step towards the use of posterior distributions of parameters other than the sky position to inform and improve the electromagnetic follow-up. In their approach, distance information encoded in the signal is used in conjunction with galaxy catalogues as the basis for a follow-up strategy based on pointing candidate host galaxies to maximise the counterpart detection probability. In the Supplement, Singer and
collaborators show a step-by-step procedure to download and visualize the sky-position-conditional posterior distribution of the luminosity distance of injection 18951 from the “First two years of electromagnetic follow-up with Advanced LIGO and Virgo” study (Singer et al., 2014, F2Y hereafter). I took that procedure as a starting point, and used it to compare the sky-position-conditional mean and standard deviation of luminosity distance derived with our method to those of the GTD study. Figure 18 shows the relative difference between the quantities computed with the two methods. Again, the difference is very small except for regions where the density of posterior samples is small, i.e. at the borders and outside the 90% sky-position confidence region.

The detectability map

By the above method, we can thus define the sky-position-conditional posterior detectability estimate (i.e. the detectability map) as

\[
P(F(t) > F_{\text{lim}} | \alpha, S) = \sum_{i=1}^{N} w_i \int_{F_{\text{lim}}}^{\infty} K \left( \frac{F - F_i(t)}{\sigma_F} \right) \, dF
\]

where \(F_i(t)\) represents the flux (in the chosen band) at time \(t\) of the lightcurve computed using the \(i\)-th posterior sample parameter values, \(F_i(t) = F(t, d_{L,i}, \iota, \ldots)\). If we approximate the kernel functions with delta functions \(K(x) \sim \delta(x)\), the expression becomes

\[
P(F(t) > F_{\text{lim}} | \alpha, S) \sim \sum_{i=1}^{N} w_i H(F_i(t) - F_{\text{lim}})
\]

where \(H(x)\) is the Heaviside function.
The earliest, best and latest detection time maps

By the above method, information encoded in the GW signal is used to estimate the detectability of the EM counterpart at a given time $t$, if it is at a certain sky position $\alpha$. By setting a minimum required detectability $\lambda$ one can define (for each sky position $\alpha$) a time interval during which $P(F(t) > F_{\text{lim}} \mid \alpha, S) \geq \lambda$. If the detectability map $P(F(t) > F_{\text{lim}} \mid \alpha, S)$ never reaches $\lambda$ for a certain sky position, that position is "hopeless", i.e. there is too little chance of detecting the EM counterpart if it is located there. The earliest and latest detection time maps are then defined respectively as

$$
\begin{align*}
t_{E,\lambda}(\alpha) &= \inf \{t \mid p(t, \alpha) \geq \lambda \} \\
t_{L,\lambda}(\alpha) &= \sup \{t \mid p(t, \alpha) \geq \lambda \}
\end{align*}
$$

(37)

where I set $p(t, \alpha) \equiv P(F(t) > F_{\text{lim}} \mid \alpha, S)$ for ease of reading. Irrespectively of $\lambda$, the best detection time map can be defined as

$$
t_{B}(\alpha) = \arg \max \{p(t, \alpha)\}
$$

(38)

These maps are the simplest piece of information about “where and when to observe” that can be constructed using the detectability map. A follow-up strategy should then try to arrange observations so that a field centered at $\alpha$ is observed at a time as close as possible to $t_{B}(\alpha)$, and in any case not earlier than $t_{E,\lambda}(\alpha)$ or later than $t_{L,\lambda}(\alpha)$. Let me now introduce a simple algorithm to construct such a follow-up strategy, after which I will be able to show a practical example (§4.5).

A follow-up strategy construction algorithm

In order to perform a first test of the approach outlined in the preceding sections, I will apply it to a simulated event and I will construct a “simulated follow-up strategy” based on it. To this end, I use an unambiguous algorithm to define the strategy for a given event and a given observing facility. To keep the discussion as simple as possible, I work in an idealized setting where all points of the skymap are observable by our facility during some pre-defined time windows. I assume...
that each observation covers an area $A_{\text{FoV}}$ at observing frequency $\nu_{\text{obs}}$ and that the limiting flux $F_{\text{lim}}$ for detection is independent of sky position and is always reached after an integration time $T_{\text{int}}$. The outline of the algorithm is the following:

- I divide the skymap in patches, each representing a potential field to be observed;

- I define a list of available time slots (i.e. possible observing time windows) on the time axis, starting 24 hours after the event (posterior samples are typically obtained after several hours or even days, so an earlier start would be unrealistic. Note that this does not mean that I discourage an earlier follow-up – which is of great importance especially for the Optical and X-ray afterglow – but only that this method may not be applicable with very low latency$^6$);

- starting from the patch with highest sky position probability, I check if the detectability map $P(F(t) > F_{\text{lim}} | \alpha, S)$ within that patch exceeds $\lambda$ at some time within the available time slots: if it does not, I discard the patch (I choose not to observe it); if it does, I schedule the observations that cover that patch at the time when the detectability is highest, and I mark the corresponding time slots as not available anymore;

- I proceed to the next patch in descending order of skymap probability until the available time is over, or until all patches have been processed.

To keep the implementation of the above steps as simple as possible, I use a HEALPix tessellation of the sky (Gorski et al., 2005) to define the observable fields: it is a way to divide the sky (i.e. a sphere) into equal area patches, called “pixels”. The “order” $N_{\text{side}}$ of the HEALPix tessellation defines the number of pixels the sky is divided into, namely $N_{\text{pixels}} = 12 N_{\text{side}}^2$. The pixel area is then $A_{\text{pixel}} = 3438 N_{\text{side}}^{-2} \text{deg}^2$. Thus I replace the actual observations of duration $T_{\text{int}}$ and field of view $A_{\text{FoV}}$ with “pseudo observations” of area $A_{\text{pixel}}$ and effective duration $T = T_{\text{int}} \times A_{\text{pixel}} / A_{\text{FoV}}$, choosing $N_{\text{side}}$ in order to minimize the difference between $A_{\text{pixel}}$ and $A_{\text{FoV}}$. The algorithm is thus implemented as follows:

1. I consider the posterior samples produced by a parameter estimation sampler (multiple sampling algorithms are implemented in LALINFERENCE, i.e. LIGO’s parameter estimation tool$^7$) applied to a simulated signal $S$. These are points in the compact binary inspiral parameter space distributed according to the posterior probability density;

2. I find the 90% sky position confidence region of the source based on the posterior samples;

3. I divide the sky into pixels according to a HEALPix tessellation of order $N_{\text{side}}$, and I consider only those pixels which fall inside the 90% sky position confidence region of the simulated signal;

---

$^6$ In the example of §4.5, actually, the only parameters needed to predict the possible SGRB afterglow light curves are $d_L$ and $\iota$. Posterior distributions of these parameters (so called “extrinsic”) can be obtained with a low latency analysis tool such as BAYESTAR (Singer and Price, 2016).

$^7$ http://software.ligo.org/docs/lalsuite/lalinference/
4. I associate to each pixel $p$ the integral of the skymap probability density $P_p = \int_{A_{\text{pixel}}(p)} P(\alpha | S) \, d\Omega$ over the pixel area;

5. I associate to each pixel $p$ the earliest and latest detection times $t_{E,\lambda}(p)$ and $t_{L,\lambda}(p)$ averaged over that pixel; pixels for which the detectability never reaches $\lambda$ are excluded from the list of possible observations;

6. I divide the time axis into contiguous intervals (slots) of duration $T$ (the effective time needed to cover one pixel) and I mark some of the slots as “available” for the follow-up (based on the characteristics of the instrument);

7. I sort the pixels in order of decreasing $P_p$ and, starting from the first ($p = 1$) pixel, I do the following:
   a) I check that at least one available time slot is comprised between $t_{E,\lambda}(p)$ and $t_{L,\lambda}(p)$: if not, no observation of the pixel is scheduled; otherwise, the available time slot where the detectability is maximum is assigned to the observation of the pixel;
   b) I proceed to the next pixel, until all available time slots are assigned, or until all pixels have been processed.

Figure 19 shows a schematic representation of the algorithm described above.

The output of the above algorithm is thus a list of observation times $t_{\text{obs}}(p)$ that cover (part of) the skymap, giving priority to pixels with high skymap probability and trying to use the available time in a way that maximizes the probability to detect the transient. The inputs of the algorithm are:

1. the posterior sample list based on $S$;
2. the detectability threshold $\lambda$;
3. the available time windows;
4. the instrument observing frequency $\nu_{\text{obs}}$, the field of view area $A_{\text{FoV}}$, the limiting flux $F_{\text{lim}}$ and the corresponding integration time $T_{\text{int}}$ (which should also include the slew time);
5. the HEALPix tessellation order $N_{\text{side}}$ which should give a pixel area $A_{\text{pixel}}$ close to $A_{\text{FoV}}$.

The observations are given in order of decreasing “importance” (skymap probability). Once the list of observations is produced, one can decide to perform only the first $N$ observations that fit into the available telescope time. If all observations suggested by the algorithm can be performed within the available time, the excess time can be used e.g. to take comparison images of some fields at different times (for the identification of transients or uncatalogued variable sources, in absence of previously available images) or to perform deeper observations for the characterization of the candidates.

We are now ready to construct an example of how the use of sky-position-conditional posterior probabilities can help in the definition of an EM follow-up strategy.
As a test example, I consider injection number 28840 from the F2Y study. The injection simulates the inspiral of a neutron star binary with $M_1 = 1.59 M_\odot$ and $M_2 = 1.53 M_\odot$ at a luminosity distance $d_L = 75$ Mpc, with orbital plane inclination $i = 14^\circ$, at sky position $(\text{RA}, \text{Dec}) = (23^h 27^m 12^s, -10^\circ 30^\prime 0^\prime\prime)$, detected by the two-detector Advanced LIGO network on MJD 55483.27839 (i.e. at 06:40:53 of the 14th of October 2010 – this is just a simulated event) adopting an early sensitivity curve corresponding to a binary NS range of 55 Mpc (Barsotti and Fritschel, 2012). Assuming that a relativistic jet with isotropic equivalent kinetic energy $E_K = 10^{50}$ erg and half-opening angle $\theta_{\text{jet}} = 0.2$ rad (which is less than the viewing an-

<table>
<thead>
<tr>
<th>Instrument</th>
<th>$v_{\text{obs}}$</th>
<th>$F_{\text{lim}}$</th>
<th>$A_{\text{FoV}}$</th>
<th>$T_{\text{int}}$</th>
<th>$\lambda$</th>
<th>Avail. time$^a$</th>
<th>Det. time$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VST-like (shallow)</td>
<td>$4.8 \times 10^{14}$ (r filter)</td>
<td>$4$ (22.4 m$_{\text{AB}}$)</td>
<td>1</td>
<td>100</td>
<td>0.01</td>
<td>3 h/night</td>
<td>-</td>
</tr>
<tr>
<td>VST-like (deep)</td>
<td>$4.8 \times 10^{14}$ (r filter)</td>
<td>0.58 (24.5 m$_{\text{AB}}$)</td>
<td>1</td>
<td>1000</td>
<td>0.05</td>
<td>3 h/night</td>
<td>7</td>
</tr>
<tr>
<td>MeerKAT-like</td>
<td>$1.4 \times 10^{12}$</td>
<td>$500$</td>
<td>1.7</td>
<td>1000</td>
<td>0.05</td>
<td>20%</td>
<td>47</td>
</tr>
<tr>
<td>VISTA-like</td>
<td>$2.4 \times 10^{13}$ (J filter)</td>
<td>$4$ (22.4 m$_{\text{AB}}$)</td>
<td>1.5</td>
<td>1200</td>
<td>0.5</td>
<td>3 h/night</td>
<td>14.5</td>
</tr>
</tbody>
</table>

$^a$available observing time starts 24h after the event.

$^b$minimum observation time needed for detection.
Figure 21: Mercator projection of the best detection time map \( t_B(\alpha) \) at \( \nu_{\text{obs}} = 1.4 \) GHz for injection 28840 of F2Y. Red lines represent the contours of the 50 and 90 percent sky position confidence areas; the black solid line is the locus of sky positions that yield a GW signal arrival time delay of \( \Delta t = 0.359 \) ms (the “true” arrival time difference associated to the injection position) between the two LIGO detectors, and the black dashed lines correspond to \( \Delta t \pm 1 \) ms. The white contours are isolines of the average antenna pattern of the aLIGO network (the four smaller closed circular contours enclose the local and absolute minima, while the two large closed contours near the lower left and the upper right corners enclose the absolute maxima). The injection true position is marked with a star symbol. Regions marked by blue contours are those where the posterior detectability of the EM counterpart of our test example reaches the required 5 percent threshold.

gle, thus the afterglow is orphan) is launched perpendicular to the orbital plane right after the merger, I computed its afterglow lightcurve assuming an ISM number density \( n_{\text{ISM}} = 0.01 \text{ cm}^{-3} \), adopting the same microphysical parameters as in §4.2.1.1, using BOXFIT v. 1.0 (Eerten, Horst, and MacFadyen, 2012). The lightcurves at \( \nu_{\text{obs}} = 1.4 \) GHz and in the \( r \) filter are shown in Figure 20.

To produce our sky-position-conditional posterior distributions, I use 7962 posterior samples produced by one of the LALINFERENCE parameter estimation samplers. Since I fixed the jet isotropic equivalent kinetic energy \( E_K = 10^{50} \) erg, in this example the only binary parameters which are relevant to the posterior detectability of the jet afterglow are the luminosity distance and the binary plane inclination. Figure 21 shows the best detection time map at \( \nu_{\text{obs}} = 1.4 \) GHz produced using these samples (see the caption for additional information). The 90 percent sky position confidence area (represented by the larger red contours) covers approximately 1500 deg\(^2\).
For our virtual optical EM follow-up, I adopt parameters inspired by the VST follow-up of GW150914 (Abbott et al., 2016a). I consider observations in the \( r \) band (\( v_{\text{obs}} = 4.8 \times 10^{14} \) Hz) with a detection limit \( F_{\text{lim}} = 22.4 \) AB magnitude, reached after \( T_{\text{int}} = 100 \) seconds (slew + integration). The field of view is \( A_{\text{FoV}} = 1 \) deg\(^2\). With the adopted flux limit, the optical lightcurve (Fig. 20) becomes too faint for detection after a few hours. Consistently, the detectability \( P(F(t) > F_{\text{lim}} | \alpha, \delta) \) is below 1 percent at all times \( t > 1 \) d over the whole skymap except for a subregion of the skymap with a total area of 185 deg\(^2\). I run the follow-up strategy construction algorithm allowing 3 hours of available time per night from day 1 to day 15. I set \( N_{\text{side}} = 64 \) which gives \( A_{\text{pixel}} = 0.84 \) deg\(^2\) and \( T = 84 \) s. Even adopting the very low detectability limit \( \lambda = 0.01 \) (i.e. allowing for observations with a detectability as low as 1 percent), only 24 “pseudo observations” (corresponding to 22 observations) are scheduled (all during the first available night), totalling 36 minutes of telescope time. The position of the EM counterpart is not contained in any of the observed fields.

The short integration time and the relatively shallow detection limit of the VST follow-up of GW150914 are good in order to cover the largest possible area during the next few nights after the event; if we wish to have some chance to detect a relatively dim optical afterglow like that in Fig. 20, though, we need to go deeper, i.e. we need longer integration times. I thus repeat the optical search with the same parameters as above, but with \( T_{\text{int}} = 1000 \) s and \( F_{\text{lim}} = 24.5 \) m\(_{\text{AB}}\). I also raise the minimum detectability to \( \lambda = 0.05 \), to avoid pointing fields with a very low detectability. The algorithm outputs 54 “pseudo observations”, corresponding to 46 observations, totalling 15 hours of telescope time. The \( p = 25 \) observation (i.e. the 25th observation in descending order of \( P_p \)), scheduled 2d 2h 34m 45s after the event, contains the EM counterpart. The flux at that time is 24.58 m\(_{\text{AB}}\), slightly dimmer than the required detection threshold, but detectable with a threshold S/N ratio of 5 under optimal observing conditions.

As explained in §4.4.5, the follow-up observations suggested by the algorithm are given in descending order of sky position probability, thus astronomers can choose to perform only the first \( N \) observations if not enough telescope time is available to complete them all. The field containing the counterpart is at \( p = 25 \) in this case. To complete the first 25 observations, 7 hours of telescope time are needed: this represents the minimum amount of telescope time for the EM counterpart to be detected by this facility in this case.

Figure 22 shows the positions and times of the observations scheduled by the algorithm in this case. During the first night, essentially all points of the detectability map are above the limit \( \lambda \), thus observations are concentrated around the centres of the two large uncertainty regions (see Fig. 21 for an all-sky view), which are the points of largest skymap probability. During the second night, the detectability has fallen below the limit in most of the central parts of the two uncertainty regions, thus the algorithm moves towards points of lower skymap probability, but higher detectability. The evolution of the detectability map proceeds in a similar fashion until the fifth night, when the detectability at all points of the skymap eventually falls below \( \lambda \).
Figure 22: Time and position of the Optical follow-up observations of our test example (§4.5). The star marks the injection position.
Radio search

The parameters of our virtual radio follow-up are inspired by MeerKAT, the South African SKA precursor. Sixteen (of the eventual 64) 13.5 m dishes have already been integrated into a working radio telescope and produced their “first light” image\(^8\) in July 2016. I assume a field of view \(A_{\text{FoV}} = 1.7 \ \text{deg}^2\) at \(\nu_{\text{obs}} = 1.4 \ \text{GHz}\). I conservatively estimate 50 \(\mu\text{Jy}\) rms noise for a \(T_{\text{int}} = 1000 \ \text{s}\) observation (slew + integration), assuming 16 working dishes. In a large area survey, usually a 10 sigma detection is required to avoid a large number of false alarms, thus I set \(F_{\text{lim}} = 0.5 \ \text{mJy}\), i.e. I require the flux to be ten times the rms noise for the detection to be considered confident. I allow a maximum of 20 percent of the available time from day 1 to day 100 to be dedicated to the follow-up (in practice, I allow an available time window of 4.8 hours each day), and I adopt a \(\lambda = 0.05\) detectability limit. The best detection time map with the chosen parameters is shown in Figure 21, where the blue contours represent regions where the detectability reaches the required detectability limit at some time \(t > 1 \ \text{d}\) (382 deg\(^2\) in total). Setting \(N_{\text{side}} = 64\), the follow-up construction algorithm outputs 337 “pseudo observations” (corresponding to 169 pointings, totalling 47 hours of telescope time), which are represented in Figure 23. The \(p = 35\) observation contains the counterpart. It is scheduled 5d 0h 9min 27s after the event, when the flux of the EM counterpart is 0.52 mJy (see Fig. 20), which means that the afterglow is detected at better than 10 sigma. The first 35 (pseudo) observations make up 4.7 hours of telescope time: only this amount needs to be actually allocated for the follow-up to successfully detect the radio counterpart.

Infrared search of the associated macronova

For the same event, I computed the Infrared (\(J\) band) lightcurve (see Figure 20) of the associated macronova with the same assumptions as in §4.2.1.1. Due to the rather large masses of the binary components and to their similar mass, the dynamically ejected mass is small \((M_{\text{ej}} \approx 5.6 \times 10^{-3} M_{\odot}\) according to the Dietrich and Ujevic, (2017) fitting formula assuming the H\(_4\) equation of state). The lightcurve is thus quite dim. It peaks between the second and the third day, slightly brighter than 22.4 m\(\text{AB}\) in the \(J\) band. The effective temperature of the photosphere at peak is \(T_{\text{peak}} \sim 2900\) K. To detect such a transient with a telescope like VISTA, an integration time of the order of 1000 s is needed. I thus perform our virtual follow up strategy with the following parameters inspired to VISTA: I choose a limiting flux \(F_{\text{lim}} = 22.4 \ \text{mAB}\) in the \(J\) band with \(T_{\text{int}} = 1200\) s, and I set \(A_{\text{FoV}} = 1.5 \ \text{deg}^2\). Again, I assume that 3 hours per night are dedicated to the follow-up. The a priori detectability (see Figure 15) of the macronova for this limiting flux is high, meaning that most of the possible lightcurves exceed 22.4 m\(\text{AB}\), thus I set \(\lambda = 0.5\) to limit the search to points of the skymap which reach a detectability at least as good as the a priori one.

Adopting the above parameters, the algorithm outputs 123 “pseudo observations” (corresponding to 69 pointings, totalling 23 hours of telescope time), which are represented in Figure 24. The 79th observation in descending order of \(P_p\) con-

\(^8\) Media release at \(\text{http://www.ska.ac.za/media-releases/}\)
Figure 23: Time and position of the Radio follow-up observations of our test example (§4.5). The star marks the injection position.
Figure 24: Time and position of the Infrared follow-up observations of our test example (§4.5). The star marks the injection position.
tains the counterpart. It is scheduled 6 d o h 18 min 27 s after the event, when the flux of the macronova in the J band is just below the 22.9 m_{AB} and the effective temperature of the spectrum is around 2600 K. The flux is lower than \( F_{\text{lim}} \) (which was chosen to represent an indicative limit for achieving a S/N ratio \( \sim 10 \)), but according to ESO’s exposure time calculator\(^9\), such emission would be detected through the VISTA telescope at ESO in the J band with a S/N ratio of 6, assuming a 1000 s integration, in optimal observing conditions. Our virtual Infrared follow-up would thus again result in a detection in a search with a threshold at S/N \( \sim 5 \). To achieve it, 14.5 hours of telescope time are needed (the amount of time for the first 79 most important observations to be performed).

Comparison with follow-up strategies based on the a priori detectability only

For comparison, I performed additional Optical, Radio and Infrared searches using the same parameters as before (listed in Table 4), but replacing the a posteriori detectability \( P(F(t) > F_{\text{lim}} | \alpha, \delta) \) with the a priori detectability \( P(F(t) > F_{\text{lim}}) \) computed in §4.2.1.1. This should simulate a search based on a priori information only. The results are the following:

- **Optical search**: the counterpart is in the field of view 5 d 22 h 53 m 32 s after the event, when the flux of the jet afterglow in the r band is as low as 26 m_{AB}, which is definitely too faint for a detection;

- **Radio search**: the counterpart is in the field of view 20 d 2 h 18 m 0 s after the event, when the flux of the jet afterglow at 1.4 GHz is 120 \( \mu \)Jy, which is significantly below our required limiting flux. Even assuming that the sensitivity was good enough for a detection, the facility should have allocated at least 77.6 hours of telescope time for this single follow-up in order to include the observation that contains the counterpart;

- **Infrared search**: none of the 122 fields of view whose observation is scheduled by the algorithm contains the counterpart.

I conclude that the use of posterior information from the GW signal has a decisive impact on the EM follow-up in our example test case.

**Discussion**

In Salafia et al., (2017b), my co-authors and I proposed the new idea that information on compact binary inspiral parameters extracted from a GW signal can be used to predict (to some extent) the best timing for observation of the possible EM counterpart. In practice, the probability distributions of the binary parameters inferred from the GW signal are fed to a model of the candidate EM counterpart in order to define a family of possible lightcurves. The possible lightcurves are then used to construct the “detectability maps”, which represent an estimate of how likely is the detection of the EM counterpart with a given instrument, if the observation is performed at time \( t \) looking at sky position \( \alpha = (\text{RA}, \text{Dec}) \). In order

---

\(^9\) I queried the VIRCAM ETC at [http://www.eso.org](http://www.eso.org) assuming a 1.2 airmass and a seeing of 0.8 arcsec.
to apply the idea to a practical example, we introduced an explicit method to construct the detectability maps (§4.3.1) and an algorithm which uses these maps to define an EM follow-up observing strategy (§4.4.5). We then applied the method to a synthetic example, showing that it improves significantly the effectiveness of the EM follow-up (§4.5).

In order to keep the treatment as simple as possible, we adopted many simplifications at various stages of the discussion, and we intentionally avoided to mention some secondary details or to include too much complexity. Let me briefly address some of the points that were not discussed in the preceding sections.

**Model dependence and inclusion of priors on unknown parameters**

The approach clearly relies on the availability of models of the EM counterparts, and on our confidence in the predictions of these models. On the other hand, since the lightcurves are treated in a statistical sense, the models only need to represent correctly the peak flux and lightcurve general evolution. Fine details are lost in the processing of the lightcurves, and are thus unnecessary. Moreover, it is very straightforward to include our uncertainty on model parameters unrelated to the GW signal. In all examples discussed in Salafia et al. (2017), we fixed the values of such parameters (e.g. the kinetic energy $E_K$ and ISM density $n_{ISM}$ of the SGRB afterglow, or the NS equation of state). A better approach (at the cost of a higher computational cost) would be to assume priors for these parameters, i.e. to assign a probability distribution to the values of these parameters based on some prior information (e.g. available astrophysical data, if any) or on theoretical arguments. In this case, multiple lightcurves of the counterpart must be computed for each posterior sample, using different values of the unknown parameters sampled from the assumed priors. This should be the most effective way of incorporating the uncertainty on these parameters in the computation of the detectability maps (it applies as well to the a priori detectability and to the a posteriori detectability). We will explore the effect of the inclusion of such priors in the construction of a priori detectabilities, a posteriori detectabilities and detectability maps for a range of potential EM counterpart models in a future work.

**Sky position dependence of the parameters**

In §4.2.2.4 I stated that, in general, the sky-position-conditional posterior distribution of the inspiral parameters depends on the assumed sky position. The main driver of this dependence is the sky projection of the antenna patterns (i.e. the sensitivity to different polarizations) of the interferometers of the network: the distribution of distances, inclinations and mass ratios compatible with a given signal, assuming that the source is at a particular sky position, is especially constrained by what the interferometers can or cannot detect if the source is at that sky position. As an intuitive example, say that a particular sky position corresponds to the maximum sensitivity of one of the detectors with respect to a particular polarization, and say that the incident GW that yields the signal picked up by that detector contains no such polarization: only combinations of parameters for which the corresponding component of the strain is smaller than the limit set by the sen-
sitivity are admissible if that sky location is assumed. If another sky location is assumed, the constraints change accordingly. As a general trend, points of the sky where the network has a higher sensitivity will correspond to a larger average distance of the source, thus implying a lower detectability for both the macronova and the SGRB afterglow. If the sky position uncertainty region is smaller than the typical angular scale over which the antenna pattern varies, the dependence of the posterior distributions of the parameters on sky position becomes less important: the a posteriori detectability contains most of the relevant information in that case. With Advanced Virgo joining the network, in many cases the sky position uncertainty region will still extend over a few hundreds of square degrees (Abbott et al., 2016); on the other hand, better information on the two polarization states of the GW signal will be available (the two interferometers of the aLIGO network are almost anti-aligned, thus the ability of the network to distinguish between the two polarization states is rather poor as of now). In the next decades, third generation interferometers will face again the same issues about sky localization. The use of detectability maps instead of the a posteriori detectability alone is thus likely to remain useful with more advanced networks as well.

The choice of injection 28840

The injection event used to construct the example presented in the last section was selected among those of the F2Y study. We considered a two-detector case (LIGO only), as it leads in general to larger localization uncertainties. We looked through the list for an event which was quite distant and whose orbit inclination was sufficiently inclined for the jet to be off-axis. The 28840 injection event luminosity distance is indeed rather large ($d_L = 75$ Mpc), the jet is slightly off-axis ($\iota = 14^\circ$), the sky position uncertainty is large (more than 1500 deg$^2$) and the injection position is rather far away from the maximum of the skymap probability. The latter condition makes a search based only on the a priori information particularly ineffective, because the exploration of the skymap proceeds slowly (using relatively small field instruments) from the centre of the skymap (where the skymap probability is high) to the periphery (where the source is actually located). In cases when the sky location is better reconstructed (i.e. the source is closer to the point of maximum skymap probability), the improvement in the EM follow-up effectiveness thanks to the detectability maps (with respect to a strategy based only on the skymap probability and on a priori information on the EM counterpart characteristics) could be less striking. We plan to study systematically the relative improvement in a future work.

A better strategy construction algorithm

The algorithm (§4.4.5) used to construct the virtual EM follow-up strategy of our example is admittedly oversimplified. An algorithm suited for real application should be able to:

1. take into account the actual observability constraints on all points of the skymap at a given time, e.g. the setting of tiles below the horizon;
2. use a better (instrument specific) way of dividing the sky into potentially observable fields;

3. consider the impact of airmass, expected seeing, dust extinction, stellar density in the field and other variables on the detectability;

4. potentially use a different integration time for each tile, in order to maximize the detection probability;

5. avoid sequences of widely separated pointings, that would result in a waste of time in slewing.

Some recent works already addressed, at least in part, some of the above points. Ghosh et al., (2015) discussed an algorithm for the optimization of the tiling, which is totally compatible with our approach, since the detectability maps do not set a preferred tiling. Rana et al., (2017) developed and compared some ingenious algorithms that aim at maximizing the sky position probability in the search, taking into account per-tile setting and rising times. Their approach does not account for the time evolution of the EM counterpart luminosity, though, and it can result in the paradoxical situation in which the highest probability tile is observed before the time at which the flux is high enough for a detection. Incorporating the information from the detectability maps in their method could be the starting point for a realistic automated strategy construction algorithm based on the ideas presented in this work. For what concerns point (4) above, both Coughlin and Stubbs, (2016) and Chan et al., (2017) found that an equal integration time in all observations could be sub-optimal with respect to the EM counterpart detection probability. Their assumptions, though, are significantly different from ours: both assume a constant luminosity of the EM counterpart (i.e. they ignore the time variation of the flux), and they do not use astrophysically motivated priors on luminosity (the former use a flat prior, while the latter use Jeffrey’s prior $L^{-1/2}$ limited to the range span by the peak luminosities of the macronova model in Barnes and Kasen 2013). It is not straightforward to figure out if their results are applicable to our more general case as well.

As an additional caveat, we only considered the case where one epoch observation per field is enough to identify transient sources. Realistically, this is only feasible when previous images can be used as reference, or where source catalogs are complete up to the survey limit. The identification of interesting transients and the removal of those unrelated to the GW source require at least two epochs of observations, which are not taken into account in our example algorithm.

Use in conjunction with the “galaxy targeting” approach

The use of detectability maps is entirely compatible with a search based on targeting candidate host galaxies. The observation of each target galaxy would simply need to be performed as close to the corresponding best detection time (as defined in §4.4.4) as possible. Since choosing a target galaxy corresponds to assuming a known distance to the source, an intriguing further refinement of the present method could be to consider the posterior distribution of binary parameters conditioned on both sky position and distance. This would have a great impact
especially on the binary orbit inclination. In other words, it would be possible to associate a fairly well defined binary orbit inclination to each galaxy. The consequence would be that some galaxies (typically the most distant ones) would be better candidate hosts for a SGRB afterglow with respect to others, depending on the associated binary orbit inclination.

**Computational feasibility of the approach**

The Monte Carlo approach adopted in this work, in which “all possible” light-curves of an event must be computed (at least one per posterior sample, which means around $10^4$ light-curves per observing band – which must be increased by one or two orders of magnitude if priors on unknown parameters are included) requires a computationally effective way to produce the lightcurves. Since the aim of the approach is to assess the detectability, rather than to fit the model to observational data, simple analytic models which capture the main features of the expected lightcurves are better than complex numerical models suited for parameter estimation. The macronova model by Grossman et al., (2013) is a good example: few minutes are sufficient to compute $10^4$ lightcurves on a laptop using this model. Parallelization is instead unavoidable when such a large number of off-axis SGRB afterglow lightcurves are to be produced using BOXFIT (Eerten, Horst, and MacFadyen, 2012).

**Reverse-engineering: tuning the models using information from the EM counterparts**

A fascinating possible future application could be to use the detectability maps to test the underlying EM counterpart models and the assumptions on the priors: when a relatively large number of inspirals involving at least one neutron star will have been detected, it will be possible to use the detectability maps to estimate how likely the detection (or non-detection) of the corresponding EM counterparts would have been with a particular choice of priors and adopting a particular model. This could help to tune the models in order to better predict the detectability of subsequent events and may be an alternative way to get insights into the population properties of the EM counterparts.

**Conclusions**

The electromagnetic follow-up of a gravitational wave events is one of the major challenges that transient astronomy will face in the next years. The large localization uncertainty regions and the relatively low expected luminosity of the candidate counterparts call for highly optimized observation strategies. The results

---

10 In the specific case adopted in this paper, actually, the only two free parameters were the luminosity distance $d_L$ and the viewing angle $\theta_v$. Since the involved redshifts are very low, it was sufficient to compute a small number of base lightcurves for each observing frequency, at a fixed distance, with varying viewing angle. The actual lightcurves were then computed by interpolation of the base lightcurves over the viewing angle, after rescaling to the correct distance, ignoring the negligible change in the rest frame frequency corresponding to the observer frame frequency considered. Indeed, interpolation over a table of pre-computed lightcurves can be a general way to reduce the computational cost of the evaluation of the EM counterpart models.
of this work showed that information from the GW signal can be used to make event-specific adjustments to the EM follow-up strategy, and that such adjustments can improve significantly the effectiveness of the search at least in some cases, as shown in the example in §4.5. Advances in the theoretical understanding of the EM counterparts (e.g. in our ability to predict the amount of mass ejected during the merger) will increase the effectiveness of this approach, and are thus of great importance for the astronomy community as well.
Part II

LIGHT AND GRAVITATIONAL WAVES FROM A SINGLE SOURCE: GW170817
A REVOLUTION STARTED ON THE 17TH OF AUGUST 2017

Not many people have had the chance to witness a revolution. Only few have had the luck, and the responsibility, to be part of it. I feel right in between the two cases: I was just entering the field, I had just started to understand some of the aspects of this exciting world, when I suddenly found myself witnessing the first ever detection of gravitational waves from a binary neutron star merger, event GW170817 (Abbott et al., 2017b), and the subsequent discovery of associated electromagnetic emission in several bands (Abbott et al., 2017d). The infrared, optical and ultraviolet emission (Coulter et al., 2017; Valenti et al., 2017; Pian et al., 2017; Andreoni et al., 2017; Arcavi et al., 2017; Chornock et al., 2017; Covino et al., 2017; Diaz et al., 2017; Drout et al., 2017; Evans et al., 2017; Hallinan et al., 2017; Pozanenko et al., 2017; Utsumi et al., 2017; Cowperthwaite et al., 2017; Gall et al., 2017; Kilpatrick et al., 2017; McCully et al., 2017; Nicholl et al., 2017; Smartt et al., 2017) leaves little doubt about its origin: a kilonova, i.e. a supernova-like emission powered by the decay of heavy nuclei produced by r-process nucleosynthesis in the expanding material ejected during the NS-NS merger and post-merger phases (Metzger, 2017). The gamma-ray burst detected by Fermi/GBM (Goldstein et al., 2017) and INTEGRAL/SPI-ACS (Savchenko et al., 2017) and the late-time X-ray (Haggard et al., 2017; Margutti et al., 2017; Troja et al., 2017) and Radio (Hallinan et al., 2017; Alexander et al., 2017) data, instead, have been interpreted in different ways by different groups. It is important to note here that the local rate of NS-NS mergers that can be estimated from this single observation is \( R_0 = 1540^{+3200}_{-1220} \, \text{Gpc}^{-3} \, \text{yr}^{-1} \) (Abbott et al., 2017b), which is three to four orders of magnitude larger than the local rate of SGRBs as estimated in chapter 3. This clearly has implications that must be discussed. Hereafter, I will try to examine some of the possible interpretations of these observations, both from the point of view of the physics behind the emission, and from that of the implications concerning rate associated to this event.

THE SHORT GAMMA-RAY BURST GRB170817A

Before the 17th of August 2017, events like GRB170817A have probably been detected several times by instruments such as Fermi and possibly Swift without attracting too much attention. Due to the low flux and fluence, this kind of burst hardly becomes a candidate for a follow-up aimed at searching for the afterglow and identifying the host galaxy. These events simply end up populating the highly incomplete part of the “\( \log N – \log S \)" (the fluence distribution of the sample), and they are usually cut off the flux limited samples used for population studies.

The association with the gravitational wave event GW170817 reveals though that at least some of these low flux events might represent extremely precious pieces of information about one of the most interesting class of phenomena astrophysicists

---

1 In chapter 4 I referred to the same phenomenon using the term macronova.
have been dreaming of: binary neutron star (NS-NS) mergers (Abbott et al., 2017b; Abbott et al., 2017d; Abbott et al., 2017c).

The fact that such an association entered the scene so early is astonishing under many points of view: even though NS-NS mergers have been among the best candidate short gamma-ray burst progenitors since a long time (Eichler et al., 1989), the most recent predictions of the rate of SGRBs within the Advanced LIGO/Virgo network range during the first and second run were very low (Ghirlanda et al., 2016; Wanderman and Piran, 2014). Moreover, all measurements of SGRB half-opening angles to date (Soderberg et al., 2006; Nicuesa Guelbenzu et al., 2011; Fong et al., 2012; Fong et al., 2013; Troja et al., 2016) point to narrow jets ($\theta_{\text{jet}} \lesssim 10^\circ$ – even though serious selection effects could be at play), implying that the probability to have an on-axis or slightly off-axis jet associated with the very first GW from a NS-NS merger should be very small. Last but not least, several studies (e.g. Ruiz and Shapiro, 2017; Murguia-Berthier et al., 2017; Margalit, Metzger, and Beloborodov, 2015) seem to indicate that not all NS-NS mergers are capable of producing a jet, making the association even more unlikely.

It can certainly be the case that all or some of the above expectations and prejudices about SGRB jets are simply wrong. On the other hand, many features of GRB170817A suggest quite naturally that it does not belong to the SGRB population we are used to. Its isotropic equivalent energy is several orders of magnitude below the least energetic SGRB known so far, despite the spectral peak energy being only moderately low with respect to the known population (Nava et al., 2011a; Zhang et al., 2017). Indeed, these facts have been taken by many as hints that GRB170817A is an ordinary or structured SGRB jet seen off-axis (e.g. Pian et al., 2017; Ioka and Nakamura, 2017; Burgess et al., 2017; Zhang et al., 2017; Lamb and Kobayashi, 2017; He, Tam, and Shen, 2017; Kathirgamaraju, Duran, and Giannios, 2017), while others interpret it as emission from the jet cocoon (e.g. Kasliwal et al., 2017; Bromberg et al., 2017; Piro and Kollmeier, 2017; Gottlieb et al., 2017; Lazzati et al., 2017a).

**Observational Properties of GRB170817A**

The GRB triggered Fermi/GBM on 2017-08-17 12:41:06, just 1.74 s after the estimated merger time of GW170817 (Abbott et al., 2017c). The duration of the burst has been estimated as $T_{90} = 2.0 \pm 0.5$ s. The 64 ms peak flux in the 10–1000 keV band was $3.7 \pm 0.9$ ph cm$^{-2}$ s$^{-1}$, and the fluence was $(2.8 \pm 0.2) \times 10^{-7}$ erg cm$^{-2}$.

A detailed analysis (Goldstein et al., 2017; Zhang et al., 2017) indicates the presence of two components: a non-thermal component, which dominates the early part of the lightcurve, whose spectrum is fit by a power-law with an exponential cut-off with $E_{\text{peak}} = 185 \pm 62$ keV, and has a fluence $(1.8 \pm 0.4) \times 10^{-7}$ erg cm$^{-2}$; a thermal component, visible in the tail of the lightcurve, which is fit by a blackbody with $k_B T = 10.3 \pm 1.5$ keV (where $k_B$ is Boltzmann’s constant) and has a fluence $(0.61 \pm 0.12) \times 10^{-7}$ erg cm$^{-2}$, which corresponds to an isotropic equivalent energy $(1.20 \pm 0.23) \times 10^{46}$ erg at $d_L \approx 40$ Mpc (i.e. the distance to the host galaxy NGC4993, Hjorth et al. 2017; Im et al. 2017). According to Goldstein et al. (2017), the thermal component could be present since the beginning, being initially masked by the non-thermal emission.
5.3 Some Possible Interpretations of GRB170817A

As mentioned in §5.1, the majority of the works published after the announcement of the discovery of GW170817 recognize GRB170817A as a “special” SGRB, i.e. they see the need to add some ingredient to the standard GRB picture in order to explain its properties. Almost all authors seem to agree on the fact that, if a jet was present, it was not seen on-axis, i.e. it was not pointing towards the Earth. The most widespread views are the following:

- the emission was due to a standard SGRB jet seen off-axis (e.g. Burgess et al., 2017; Fraija et al., 2017; Granot et al., 2017; Granot, Guetta, and Gill, 2017; Hallinan et al., 2017; Ioka and Nakamura, 2017; Lamb and Kobayashi, 2017);

- the emission was due to a structured SGRB jet seen off-axis (e.g. Kathirgamaraju, Duran, and Giannios, 2017; He, Tam, and Shen, 2017; Troja et al., 2017; Zhang et al., 2017);

- the emission was from an off-axis jet that was slower than average (e.g. Pian et al., 2017);

- the emission came from a mildly relativistic jet cocoon (e.g. Lazzati et al., 2017a; Bromberg et al., 2017; Gottlieb et al., 2017; Kasliwal et al., 2017; Piro and Kollmeier, 2017);

- aside from the mainstream, I proposed an alternative solution: emission from an isotropic fireball, powered by the strong magnetic field amplified by magnetohydrodynamic turbulence at the beginning of the merger phase (Salafia, Ghisellini, and Ghirlanda, 2018; Salafia et al., 2017a).

First of all, let me comment briefly on how these scenarios relate to the following question: is the rate of SGRBs, as derived in chapter 3, in tension with this event?

**Impact on the Rate of SGRBs**

All the interpretations listed above seek a way out of the same two problems: (1) it is improbable for the first NS-NS merger to be associated to an SGRB jet pointing towards the Earth (if jets are narrow), and (2) the luminosity of GRB170817A is extremely low compared to the known SGRB population. As a byproduct, all these interpretations also imply that the rate of this kind of event does not match that of “normal” SGRBs, i.e. of relatively energetic jets seen on-axis. Let me focus now on the interpretations that require the presence of a jet. Let $R_{SGRB}$ be the local rate of SGRBs pointing towards the Earth, and $R_{NS-NS}$ the local rate of double neutron-star mergers. The two are linked by

$$R_{SGRB} = f_b f_j R_{NS-NS}$$  \hspace{1cm} (39)

where $f_b = \left\langle 1 - \cos \theta \right\rangle$ is the collimation factor (the mean is over the half-opening angles $\theta$ of all SGRBs), which gives the fraction of jets that point towards the Earth over the total, and $f_j$ is the fraction of NS-NS mergers that produce a successful jet (I am neglecting BH-NS mergers for simplicity). Since the orbital plane inclination of the GW170817 binary with respect to the line of sight was $\lesssim 30^\circ$ (Abbott et al.,
we can neglect to first order the possible anisotropy of the cocoon or the weakness at very large angles of the emission from a structured jet, and conclude that the probability to associate such emission to GW170817 was essentially \( f_j \) (because a jet is needed in both scenarios). If we require this probability to be at least, say, \( f_j = 0.3 \), and we use \( R_{\text{NS-NS}} = 1540^{+3200}_{-1220} \text{ Gpc}^{-3} \text{ yr}^{-1} \) as estimated from this single event (Abbott et al., 2017b), from Eq. 39 we have that \( \langle \theta_j \rangle \lesssim 2.3^{+5.5}_{-1.0} \text{ deg} \) if we take \( R_{\text{SGRB}} = 0.36^{+0.53}_{-0.17} \text{ Gpc}^{-3} \text{ yr}^{-1} \) as in Ghirlanda et al., (2016), or \( \langle \theta_j \rangle \lesssim 7.6^{+13.4}_{-4.4} \text{ deg} \) if we take \( R_{\text{SGRB}} = 4.1^{+2.3}_{-1.9} \text{ Gpc}^{-3} \text{ yr}^{-1} \) as in Wanderman and Piran, (2014).

The above argument can be summarized by saying that, adopting an off-axis jet or cocoon interpretation of GRB170817A, the relatively low rate of SGRBs found in recent population studies can be reconciled with the association of GRB170817A with GW170817 provided that a large fraction \( f_j \gtrsim 0.3 \) of NS-NS mergers produce a jet (which must be a successful jet in the structured jet scenario, but may be a choked jet in the cocoon scenario), and that jets feature narrow half-opening angles \( \langle \theta_j \rangle \lesssim \text{ few degrees} \).

In the isotropic fireball scenario proposed in Salafia, Ghisellini, and Ghirlanda, (2018) and Salafia et al., (2017a), on the other hand, no jet is needed to explain the emission: it thus implies no requirements on \( f_b \) and \( f_j \); it simply requires the isotropic fireball to be produced in most NS-NS mergers. In chapter 7, I describe such scenario and show that it could be present in most or even all NS-NS mergers. Before, in chapter 6, I discuss some expected properties of emission from an off-axis jet, and use them to show that an off-axis jet interpretation of GRB170817A presents some difficulties.
PULSES: BUILDING BLOCKS OF GRB LIGHT CURVES

The diversity and complexity of GRB prompt emission light curves is often used to illustrate the difficulty in the classification of these sources and in the unification of their properties. A natural approach to get insight into such complexity is to look for global and average properties, like flux time integral (i.e. fluence), total duration, average spectrum, peak flux. Alternatively, one can try to break down the light curve into simpler parts following some pattern. If a fundamental building block was identified, the analysis of single blocks could be the key to the unification and disentanglement of properties of the underlying processes. Many authors (e.g. Imhof et al., 1974; Golenetskii et al., 1983; Norris et al., 1986; Link, Epstein, and Priedhorsky, 1993; Ford et al., 1995; Kargatis and Liang, 1995; Liang and Kargatis, 1996; Preece et al., 1998; Ramirez-Ruiz and Fenimore, 1999; Lee, Bloom, and Petrosian, 2000; Ghirlanda, Celotti, and Ghisellini, 2002; Hakkila and Preece, 2011; Lu et al., 2012; Basak and Rao, 2014) performed careful analyses of light curves and time resolved spectra looking for patterns and for hints about such fundamental building blocks. As early as 1983, Golenetskii et al. found evidence of a correlation between spectral peak energy and photon flux during the decay of pulses. Such correlation was later confirmed by Kargatis et al., (1994), Kargatis and Liang, (1995) and Borgonovo and Ryde, (2001) and became known as the hardness-intensity correlation (Ryde and Svensson, 1998). Norris et al., (1986) was presumably the first to systematically decompose the light curves into pulses and to look for patterns in the properties of these putative building blocks. Some years later, Woods and Loeb, (1999) developed tools to calculate the emission from a relativistically expanding jet, including the case of an off-axis viewing angle. Ioka and Nakamura, (2001) took advantage of this formulation to model the single pulse, finding that the spectral lag-luminosity and variability-luminosity correlations found by Norris, Marani, and Bonnell, (2000) and Reichart et al., (2001) can be explained as viewing angle effects. The pulse model at that stage assumed emission from a unique radius and from an infinitesimally short time interval (i.e. a delta function in radius and time). In the following years, other authors proposed increasingly refined models of the pulse (e.g. Dermer, 2004; Genet and Granot, 2009), but neglected the possibility for the jet to be observed off-axis.

Off-axis pulses and the interpretation of low-luminosity GRBs

The viewing angle, i.e. the angle between the jet axis and the line of sight, is usually assumed to be smaller than the jet semi-aperture, in which case the jet is said to be on-axis. For larger viewing angles, i.e. for off-axis jets, the flux is severely suppressed because of relativistic beaming. Nevertheless, it can be still above detection threshold if the viewing angle is not much larger than the jet
semi-aperture, especially if the burst is at low redshift. Pescalli et al., (2015) have shown that off-axis jets might indeed dominate the low luminosity end of the observed population.

The idea that nearby low luminosity GRBs could be off-axis events has been a subject of debate since the observation of GRB980425. Soderberg et al., (2004) rejected such possibility, based on radio observations of GRB980425 and GRB031203, but soon later Ramirez-Ruiz et al., (2005) presented an off-axis model for the afterglow of GRB031203 which seems to fit better the observations (including radio) with respect to the usual on-axis modelling. Using the same off-axis afterglow model, Granot, Ramirez-Ruiz, and Perna, (2005) extended the argument to two X-Ray Flashes, thus including them in the category of off-axis GRBs. Based on prompt emission properties, an off-axis jet interpretation of X-Ray Flashes had been already proposed by Yamazaki, Ioka, and Nakamura, (2002) and Yamazaki, Ioka, and Nakamura, (2003), following the work by Ioka and Nakamura, (2001). Ghisellini et al., (2006a) argued that the off-axis interpretation of GRB031203 and GRB980425 is not practicable, because their true energy would then be on the very high end of the distribution, implying a very low likelihood when combined with the low redshift of these two events. More recently, the idea that such events are members of a separate class (e.g. Liang, Zhang, and Zhang, 2007; Zhang, 2008; He et al., 2009; Bromberg, Nakar, and Piran, 2011; Nakar, 2015) has gained popularity. Recent results about the GRB luminosity function (Pescalli et al., 2015), though, still point towards the unification of these events with ordinary GRBs based on the off-axis viewing angle argument. In this chapter, I will address the issue from another point of view, by focusing on the apparently single pulsed, smooth behaviour of prompt emission light curves of these bursts, trying to figure out if such behaviour is expected in the case of an off-axis viewing angle.

PULSE OVERLAP AND LIGHT CURVE VARIABILITY

For a light curve to be highly variable, pulses must be short and not overlap too much. If pulses are produced at a typical radius by material moving close to the speed of light, then the amount of overlap can depend on the viewing angle. To see this, consider two point sources moving at equal constant speed $\beta c$ along the $\mathbf{z}$ axis, separated by a distance $\beta c \Delta T$, as in Fig. 25. Each source starts emitting at radius $R_{in}$ and stops emitting at $R_{off}$. An observer along the $\mathbf{z}$ axis (viewing angle $\theta_v = 0$) sees two separated pulses of equal duration $\Delta t_0$ and peak flux $F_0$, the second starting a time $\Delta T$ after the start of the first. Because of relativistic Doppler effect, an observer with another $\theta_v \neq 0$ measures a lower (bolometric) peak flux $F = F_0/b^4$ and a longer pulse duration $\Delta t = b \Delta t_0$, where $b = (1 - \beta \cos \theta_v)/(1 - \beta)$ is the ratio of the on-axis relativistic Doppler factor $\delta(0) = \Gamma^{-1} (1 - \beta)^{-1}$ to the off-axis one $\delta(\theta_v) = \Gamma^{-1} (1 - \beta \cos \theta_v)^{-1}$ (Rybicki and Lightman, 1979; Ghisellini, 2013). The difference in pulse start times $\Delta T$, on the other hand, is not affected by the viewing angle, because the emission of both pulses begins at the same radius: it can be thought of as emission from a source at rest (for what concerns arrival times). The pulses overlap if $\Delta t > \Delta T$, which corresponds to $\theta_v > \theta_{ov} \approx \Gamma^{-1} \sqrt{\Delta T/\Delta t_0 - 1}$. Consider the case in which the pulse separation is equal to the pulse duration, i.e. $\Delta T = 2\Delta t_0$ and $\theta_{ov} \approx \Gamma^{-1}$. 
6.2 Pulse Overlap and Light Curve Variability

Figure 25: Upper left panel: two point sources (blue and red dots) move at equal constant speed along the $z$ axis, separated by a distance $\beta c \Delta T$. Each starts emitting at $z = R_{\text{on}}$ and stops emitting at $z = R_{\text{off}}$. The blue and red circles represent wavefronts of the emitted light. The first blue wavefront and the first red wavefront reach any observer with a time difference $\Delta T$. Upper right panel: close up. Depending on the viewing angle $\theta_v$, a distant observer sees the blue and red signal separated ($\theta_v < \theta_{\text{ov}}$) or overlapped ($\theta_v > \theta_{\text{ov}}$). The angle $\theta_{\text{ov}}$ is the angle between the $z$ axis and the normal to a plane tangent to both the first red wavefront and the last blue wavefront. Lower panel: sketch of the bolometric light curve as seen by on-axis ($\theta_v = 0$) and off-axis ($\theta_v > \theta_{\text{ov}}$) observers. Letting $b = (1 - \beta \cos \theta_v) / (1 - \beta)$, the single pulse flux as measured by the off-axis observer is decreased by a factor $b^4$ with respect to the on-axis one, while the duration is increased by a factor $b$. The pulse separation $\Delta T$, though, does not depend on the viewing angle, being the emission time difference at a fixed radius. This causes the pulses to overlap as seen by the off-axis observer.
Figure 26: Example light curves constructed by superposition of pulses. All pulses are equal. The pulse shape is a double-sided Gaussian (Norris et al., 1996), which is a common phenomenological description of GRB pulse shapes. The peak flux is $F_0$, and the rise to decay time ratio is 1:3. The start times of the pulses are the same for the two light curves and have been sampled from a log-normal distribution with mean 20 s and sigma 0.35 dex. Pulses in the lower light curve are broadened by a factor of 4 and their flux is lowered by a factor of 256 with respect to the upper light curve, which corresponds to the effect of an off-axis viewing angle $\theta_v = \sqrt{3} \Gamma^{-1}$ as discussed in the text.

Increasing the viewing angle, the amount of pulse overlap increases, reaching half of the pulse width as soon as $b = 4$, which corresponds to $\theta_v \approx \sqrt{3} \Gamma^{-1}$. With this viewing angle, the flux of the single pulse is reduced by $b^4 = 256$, but the flux in the overlapped region is higher by a factor of two, so that the peak flux effectively decreases by 128.

The purpose of this simple argument is to show that if pulses are produced by material moving at relativistic speed, and if a typical emission radius exists, then the apparent variability of the light curve can be significantly smeared out by pulse overlap as seen by an off-axis observer (see also Fig. 26). The viewing angle needed for this to happen is still small enough for the flux not to be heavily suppressed by relativistic (de-)beaming. One may argue that the probability to have a viewing angle in the right range for this to happen without falling below the limiting flux of the instrument is vanishingly small. To address this point, in §6.7 I will give an estimate of the rate of such events, showing that a significant fraction ($\sim 40\%$) of nearby bursts ($z < 0.1$) are likely observed with $\theta_v > \theta_{\text{jet}} + \Gamma^{-1}$. 
Being based solely on geometry and relativity, the above argument does not rely on a specific scenario, e.g. internal shocks. Any model in which photons are produced at a typical radius, being the photospheric radius (e.g. subphotospheric dissipation models like those described in Rees and Mészáros, 2005; Giannios, 2006; Beloborodov, 2010) or beyond (e.g. magnetic reconnection models, Lazarian et al., 2003; Zhang and Yan, 2010) eventually must take into account the pulse overlap as seen by off-axis observers.

**Pulses in the Internal Shock Scenario**

The pulse width in GRB light curves is roughly constant throughout the burst duration (Ramirez-Ruiz and Fenimore, 1999). The internal shock scenario (Rees and Mészáros, 1994) provides a natural framework for the understanding of this kind of behaviour. In this scenario, discontinuous activity in the central engine produces a sequence of shells with different Lorentz factors. When faster shells catch up with slower ones, shocks develop and particles are heated. If the plasma is optically thin and some magnetic field is present, the energy gained by the electrons is promptly and efficiently radiated away by synchrotron (and inverse Compton) emission. Each pulse is thus the result of the merger of two shells beyond the photospheric radius \( R_{\text{ph}} \). The strength of the shock, and thus the efficiency of the electron heating, depends strongly on the relative Lorentz factor of the merging shells (a radiative efficiency of a few percent is achieved only for \( \Gamma_{\text{rel}} \gtrsim 3 \), Lazzati, Ghisellini, and Celotti 1999). Shell pairs with small relative Lorentz factors merge later (they need more time to catch up with each other), thus the highest efficiency is achieved for shells merging just after the photospheric radius. This explains, within this framework, why the typical pulse width is not seen to grow with time: the bulk of the emission happens at a fixed radius, regardless of the expansion of the jet head.

**Time scales**

Three main time scales arise in the internal shock scenario:

- the electron cooling time \( \tau_{\text{cool}} \), i.e. the time needed by electrons to radiate away most of the energy gained from the shock;
- the angular time scale \( \tau_{\text{ang}} \), i.e. the difference in arrival time between photons emitted at different latitudes;
- the shell crossing time \( \tau_{\text{sc}} \), i.e. the time needed for the two shells to merge.

The electron cooling time scale, as measured in the lab frame, is \( \tau_{\text{cool}} \approx \Gamma^{-1} \gamma / \dot{\gamma} \), where \( \Gamma \) is the bulk Lorentz factor, \( \gamma \) is the typical electron Lorentz factor as measured in the comoving frame, and \( \dot{\gamma} \) is the cooling rate. For synchrotron emission, it is of the order of \( \tau_{\text{cool}} \approx 10^{-7} \) s for typical parameters\(^1\) (Ghisellini, Celotti, and Lazzati, 2000).

---

\(^1\) by typical parameters I mean \( \Gamma = 100, \Gamma_{\text{rel}} = \text{a few}, U_{\text{rad}} = U_{\text{B}}, \) a typical synchrotron frequency of 1 MeV and I assume equipartition. See Ghisellini, Celotti, and Lazzati, (2000) and references therein for a complete treatment.
The angular time scale arises when one takes into account the arrival time difference of photons emitted at the same time by parts of the shell at different latitudes. It is defined as the arrival time difference between a pair of photons, one emitted at zero latitude and the other at $\Gamma^{-1}$ latitude at the same time. Given a typical photospheric radius (Daigle and Mochkovitch, 2002a) $R_{\text{ph}} \sim 10^{12}$ cm, this difference is $\tau_{\text{ang}} \sim R/\Gamma^2c \approx 3 \times 10^{-3}$ s $R_{12}/\Gamma^2_2$ (we adopt the notation $Q_x = Q/10^x$ in cgs units).

The shell crossing time is $\tau_{\text{sc}} \sim w/c$, where $w$ is the typical shell width. Being linked to the central engine activity, one may assume $w$ to be of the order of a few Schwarzschild radii. The Schwarzschild radius of a $5 \, M_{\odot}$ black hole is $R_s \approx 1.5 \times 10^6$ cm, thus an estimate might be $\tau_{\text{sc}} \sim 5 \times 10^{-5}$ s $w_6$. In this case, we have $\tau_{\text{ang}} > \tau_{\text{sc}}$, i.e. the effect of shell curvature dominates over (i.e. smears out) intrinsic luminosity variations due to shock dynamics, which take place over the $\tau_{\text{sc}}$ time scale or less.

Temporal analysis of GRB light curves, though, along with simple modelling of internal shocks (Nakar and Piran, 2002a; Nakar and Piran, 2002b), seem to indicate that the shell width must be comparable to the initial shell separation. Taking the two as equal, the time needed for two shells to collide is the same as the shell crossing time, and thus the shell merger is completed within a doubling of the radius. In this case, the shell crossing time and the angular time scale are the same (Piran, 2005). This means that details of the pulse shape and spectral evolution cannot be explained as just being due to the shell curvature effect. Indeed, discrepancies between predictions based on shell curvature only and observations have been pointed out (e.g. Dermer, 2004).

Nevertheless, the description of the pulse in terms of shell curvature qualitatively reproduces the main features of many long GRB pulses, namely the fast rise and slower decay, the hard-to-soft spectral evolution, and the presence of a hardness-intensity correlation (Ryde and Petrosian, 2002). For this reason, since I focus on the effect of the viewing angle rather than on details of the pulse, in what follows I will set up a simple model of the pulse based on the shell curvature effect only.

PULSE LIGHT CURVES AND TIME DEPENDENT SPECTRA IN THE SHELL-CURVATURE MODEL

Main assumptions

Based on the arguments outlined in §6.3.1, I assume that the variation of the flux seen by the observer during a single pulse is due only to the angular time delay described above. The luminosity $L$ of the shell is assumed constant during an emission time $T$ and zero before and after this time interval. The emitting region is assumed geometrically and optically thin. The emitted spectrum, as measured by a locally comoving observer, is assumed to be the same for any shell fluid element.

Woods and Loeb, 1999 and other authors already provided the necessary formulas for the computation of the pulse shape in this case. Nevertheless, I will go
through an alternative derivation, which will allow us to set up an independent notation and a physical understanding of the process.

Let the radius of the shell be $R$ at the beginning of the emission and $R + \Delta R$ at the end of it. The bolometric flux $F(t)$ (specific flux $F_\nu(t)$) is computed by integration of the intensity $I$ (specific intensity $I_\nu$) over the appropriate equal arrival time surface (EATS hereafter) $S(t)$ (i.e. the locus of points of the source whose emitted photons reach the observer at $t$), namely

$$F(t) = \int_{S(t)} I(s) \, ds/r^2$$

where $r$ is the distance between the element $ds$ of $S(t)$ and the observer.

In our situation, it is convenient to use spherical coordinates centred on the emitting sphere, so that (assuming cylindrical symmetry of the intensity) we have $ds = 2\pi \sin \theta \, d\theta \, R(t_e)^2$, where $t_e = t - r/c$ is the emission time. The distance $r$ of the point $(\theta, \phi, t_e)$ from the detector is $r \approx d - R(t_e) \cos \theta$ where $d$ is the distance of the sphere centre from the detector, thus

$$F(t) = 2\pi \int_{S(t)} I(\theta, t_e) \, \sin \theta \, d\theta \frac{R(t_e)^2}{(d - R(t_e) \cos \theta)^2}$$

Since $d \gg R(t_e)$, the last term is well approximated by $R(t_e)^2/d^2$, thus we can write

$$F(t) = \frac{2\pi}{d^2} \int_{S(t)} I(\theta, t_e) \, R(t_e)^2 \, \sin \theta \, d\theta$$

Let me now assume that the luminosity $L$ of the sphere is constant in the time interval $t_0 < t < t_0 + T$. This is different from assuming that the intensity is constant, in that it prevents the expansion of the surface area from causing a rise in the luminosity (this alternative assumption would be more appropriate in the description of an external shock). In terms of intensity, this assumption implies that $I \propto R^{-2}$, which I write as

$$I(\theta, t_e) = I_0(\theta) \frac{R^2}{R^2(t_e)}$$

Inserting this definition into Eq. 42 allows us to bring the radius outside the integral.

Assuming isotropic emission in the comoving frame, in the approximation of infinitesimal shell thickness, the intensity is related to the comoving one (primed quantities throughout the paper will always refer to the comoving frame) through $I = \delta^4 I^\prime$ (or $I_\nu = \delta^3 I_\nu^\prime (\nu/\delta)$ for the specific intensity). Note that the constant luminosity assumption implies $I^\prime \propto R^{-2}$: this is consistent if the number of emitting particles is constant despite the increase of the surface area with the expansion. It would not be appropriate e.g. for external shocks, where the number of emitting particles instead increases with increasing surface area.

In what follows, I will show a simple way to work out the shape of the EATS in our case.
Figure 27: A sphere starts emitting radiation at $t = t_0$ and stops at $t = t_0 + T$. The line of sight of a distant observer is parallel to the $z$ axis. **Left:** a time $t < T$ after the arrival of the first photon, the observer has received radiation from the portion of the sphere with $z > R - ct = R \cos \theta_{on}$; **Right:** later when $t > T$, the observer has stopped receiving radiation from the portion of the sphere with $z > R - c(t - T) = R \cos \theta_{off}$. Thus the effective emitting surface is the portion of the sphere with $R \cos \theta_{on} < z < R \cos \theta_{off}$.

**Equal arrival time surfaces**

**A sphere**  Consider a sphere of radius $R$. The surface of the sphere starts emitting electromagnetic radiation at $t = t_0$ and stops suddenly at $t = t_0 + T$ (as measured in the inertial frame at rest with respect to the centre of the sphere). Emitted photons reach a distant observer at different arrival times. Let the line of sight be parallel to the $z$ axis (as in Fig. 27). The first photon to reach the observer is the one emitted at $t = t_0$ from the tip of the sphere at $z = R$. Let $t = 0$ be its arrival time as measured by the observer. A photon emitted at the same time $t = t_0$ by a point of the surface at $z = R \cos \theta_{on}$ reaches the observer at a later time $t = R(1 - \cos \theta_{on})/c$. Thus, despite the surface turned on all at the same time $t = t_0$, at a given time $t$ the observer has received radiation only from the portion with $z/R > \cos \theta_{on} = 1 - ct/R$ (left panel of Fig. 27). This can be visualized as each point on the sphere being turned on by the passage of a plane traveling in the $-z$ direction with speed $c$, starting from $z = R$ at $t = 0$. The same reasoning applies to the turning-off of the sphere: each point is turned off by the passage of a plane traveling in the $-z$ direction with speed $c$, starting from $z = R$ at $t = T$. As a result, if $T < R/c$, at some time $t$ the observer will “see” only the portion of sphere comprised between $\cos \theta_{on} = 1 - ct/R$ and $\cos \theta_{off} = 1 - c(t - T)/R$ (right part of Fig. 27). Thus, the EATS at time $t$ is this portion of the sphere.

**An expanding sphere**  If the sphere is expanding, the above argument is still valid, with some modification. The radius now is $R(t) = R + \beta c(t - t_0)$ so that the lighting up takes place at $R(0) = R$ and the turning-off at $R(T) = R + \beta c T = R + \Delta R$. Since the lighting up happens all at the same radius, the angle $\theta_{on}$ up to which the observer sees the surface on is still given by $\cos \theta_{on} = 1 - ct/R$.
Light curve of the pulse from an expanding sphere

In sections §6.4.2 and §6.4.2, we found that the EATS are the portions of the sphere comprised between $\cos \theta_{\text{on}}$ and $\cos \theta_{\text{off}}$, so that we have, for $t > 0$

$$F(t) = \frac{2\pi R^2}{d^2} \int_{\theta_{\text{off}}(t)}^{\theta_{\text{on}}(t)} I_0(\theta) \sin \theta \, d\theta \quad (44)$$

The intensity is related to the comoving one by $I_0(\theta) = \delta^4(\theta) I'_0$, where $\delta(\theta) = [\Gamma(1 - \beta \cos \theta)]^{-1}$ is the Doppler factor, and $I'_0$ is the comoving intensity, assumed isotropic. The flux is then

$$F(t) = \frac{2\pi R^2}{d^2} \frac{I'_0(\theta)}{\Gamma^4(\theta)} \int_{\cos \theta_{\text{on}}(t)}^{\cos \theta_{\text{off}}(t)} \frac{d \cos \theta}{(1 - \beta \cos \theta)^4} \quad (45)$$
which yields, after substitution of the expressions for \( \cos \theta_{\text{on}} \) and \( \cos \theta_{\text{off}} \) derived above, the light curve of the pulse

\[
F(t) = F_{\text{max}} \times \begin{cases} 
1 - \left( \frac{1 + \frac{t}{\tau} - t_{\text{off}}}{\tau + t_{\text{off}}} \right)^{-3} & t \leq t_{\text{off}} \\
\left( 1 + \frac{t}{\tau} \right)^{-3} - \left( 1 + \frac{t}{\tau} - 3 t_{\text{off}} \right)^{-3} & t > t_{\text{off}}
\end{cases}
\]

where

\[
F_{\text{max}} = \frac{2\pi R^2 (1 + \beta)^3 \Gamma^2 I'_0}{3d^2} \frac{1}{\beta}
\]

is the (saturation) peak flux if the pulse lasts \( T \gg R/c \),

\[
\tau = \frac{R}{\beta c (1 + \beta) \Gamma^2}
\]

and

\[
t_{\text{off}} = \frac{T}{(1 + \beta) \Gamma^2} = \frac{\Delta R}{\beta c (1 + \beta) \Gamma^2} \equiv t_{\text{peak}}
\]

The fluence, i.e. integral of the flux over time, from \( t = 0 \) to \( t \to \infty \) is

\[
\mathcal{F} = F_{\text{max}} \times \frac{3}{2} t_{\text{off}} = \frac{\pi R^2 (1 + \beta)^2}{d^2} \frac{1}{\beta} I'_0 T
\]

It is worth noting that the light curve parameters are three, i.e. \( \tau \), \( t_{\text{off}} \) and \( F_{\text{max}} \), while the underlying physical parameters are four, namely \( R \), \( T \), \( \Gamma \) and \( I'_0 \). This degeneracy leads to the impossibility to determine all the physical parameters by fitting the pulse shape to an observed light curve.

Figure 29 shows light curves of pulses from an expanding sphere. If the jet viewing angle \( \theta_v \) is small enough so that the angular distance \( \theta_{\text{jet}} - \theta_v \) of the line of sight from the jet border is much larger than \( 1/\Gamma \), the jet pulse lightcurves are just the same as those from an expanding sphere. Since the typical expected Lorentz factor of GRB jets is \( \Gamma \sim 100 \), this means that a viewing angle a few \( 0.01 \) radians smaller than \( \theta_{\text{jet}} \) allows one to consider the jet practically on-axis. On the other hand, if the jet is very narrow, or if the Lorentz factor is low enough (i.e. if \( \theta_{\text{jet}} \) is comparable with \( 1/\Gamma \)), then the finite half opening angle must come into play. In the next section I will show this latter case.

**Light curve of the pulse from an on-axis jet with \( \theta_{\text{jet}} \ll 1/\Gamma \)**

To compute the pulse light curve of a jet of semiaperture \( \theta_{\text{jet}} \), we can just take the pulse of the sphere and “trim” the unwanted part. If the jet is seen on-axis, this amounts to limit the integral of Eq. 45 to angles \( \theta < \theta_{\text{jet}} \). It is straightforward to work out at what time the EATS borders reach the jet border, i.e.

\[
\theta_{\text{on}}(t) = \theta_{\text{jet}} \implies t = R(1 - \cos \theta_{\text{jet}})/c \equiv t_{\text{jet}}
\]

and similarly

\[
\theta_{\text{off}}(t) = \theta_{\text{jet}} \implies t = t_{\text{off}} + R_{\text{off}}(1 - \cos \theta_{\text{jet}})/c \equiv t_{\text{jet},\text{off}}
\]
Figure 29: Bolometric light curves of three pulses from an expanding sphere. The flux is normalized to $F_{\text{max}}$ and the observer time is in units of $\tau$ (see the text for the definition of these quantities). The ratio of $\Delta R$ to $R$ is given near each curve. The black dashed line represents the saturation flux, which is reached if $T \gg R/c$, or equivalently if $\Delta R \gg R$. 
Figure 30: Bolometric light curves of three pulses from an on-axis jet with $\theta_{jet} = 1/\Gamma$. The ratio of $\Delta R$ to R is reported near each curve. The black dashed line represents the time $t_{jet}$ at which the jet border first comes into sight. In this case $t_{jet}$ equals $\tau$.

where $R_{off} \equiv R + \Delta R$. It is then easy to see that the light curve becomes

$$\frac{F(t)}{F_{max}} = 1 - \left( 1 + \frac{\min(t, t_{jet})}{\tau} \right)^{-3}$$

for $t \leq t_{off}$, then

$$\frac{F(t)}{F_{max}} = \left( 1 + \frac{t - t_{off}}{\tau_{off}} \right)^{-3} - \left( 1 + \frac{\min(t, t_{jet})}{\tau} \right)^{-3}$$

for $t_{off} < t < t_{jet,off}$, and zero for $t \geq t_{jet,off}$. This light curve is the same as that of the expanding sphere up to $t = t_{jet}$. After that, if $t_{jet} < t_{off}$ the flux saturates (the whole jet is visible) until $t = t_{off}$, then it drops and reaches zero at $t = t_{jet,off}$. If $t_{jet} \geq t_{off}$ no saturation is reached. The difference between the expanding sphere and the on-axis jet is relevant only if $t_{jet} \lesssim \tau$, i.e. if $\theta_{jet} \lesssim 1/\Gamma$, as expected.

Figure 30 shows light curves (Eqs. 53 & 54) of pulses from an on-axis jet with $\theta_{jet} = 1/\Gamma$.

Off-axis jet

If the jet is off-axis, it is still possible to compute an expression for the light curve. I propose here an approach to the computation, based on geometrical arguments. Let me call $\theta_v$ the angle between the jet axis and the line of sight, $\theta_{jet}$ the jet half-opening angle, and let me set the coordinate system so that the jet axis lies in the
6.4 PULSE LIGHT CURVES AND TIME DEPENDENT SPECTRA IN THE SHELL-CURVATURE MODEL

Figure 31: The off-axis jet can be thought of as being part of an expanding sphere. The axes in the figures above are chosen so that the jet axis lies on the z-x plane. Jet surface elements in the S shaded part all share the same Doppler factor \( \delta \), and thus they all give the same contribution (per unit emitting area) to the flux. For this reason, the ratio of the flux from the annulus to the flux from S is just equal to the ratio of the corresponding surface areas. Left: the jet is seen from the z-axis. The \( y \)-coordinate \( y_i \) of the intersections between the annulus and the jet border is shown. Right: the jet is seen from the x-axis. Angles \( \theta_v \), \( \theta_{jet} \), and \( \theta \) are reported.

\[ z - x \text{ plane, as in Fig. 31. This is what one would obtain by rotating an on-axis jet counter-clockwise by an angle } \theta_v \text{ around the x axis. Let me now consider the ring-shaped part of the sphere surface ("annulus" hereafter) comprised between } \theta \text{ and } \theta + d\theta. \text{ If } \theta > |\theta_v - \theta_{jet}|, \text{ a portion } S \text{ of the annulus lies on the jet surface (shaded part in Fig. 31). Since the annulus width } d\theta \text{ is infinitesimal, the ratio of the area of } S \text{ to the total annulus area is equal to the ratio between the length } l \text{ of } S \text{ and the total annulus length } 2\pi R \cos \theta. \text{ Moreover, this is also equal to the ratio of the flux } dF_S \text{ from } S \text{ to the flux } dF_a \text{ from the whole annulus, namely}
\]

\[
\frac{dF_S(\theta)}{dF_a(\theta)} = \frac{l(\theta)}{2\pi R \cos \theta}
\]

(55)

The flux due to the annulus is easily obtained by deriving the flux of the sphere, Eq. 45, with respect to \( \theta \), which gives

\[
dF_a(\theta) = \frac{dF}{d\theta} d\theta = \frac{2\pi R^2}{d^2} \frac{I_0'}{\Gamma^4} \sin \theta \, d\theta
\]

(56)

To compute the length \( l(\theta) \), we must first find the intersections between the annulus and the jet border. Both are circles on the sphere surface, i.e. they lie on the surface \( x^2 + y^2 + z^2 = R^2 \). The annulus is the circle given by the intersection between the plane \( z = R \cos \theta \) and the sphere; in a coordinate system \( K' \) where the \( z' \) axis coincides with the jet axis, the jet border is the circle given by the intersection between the plane \( z' = R \cos \theta_{jet} \) and the sphere. Applying a rotation of an angle \( \theta_v \) around the x axis, this plane becomes \( z \cos \theta_v - y \sin \theta_v = R \cos \theta_{jet} \).
The intersections between the two circles are then found by solving the linear system

\[
\begin{align*}
  x^2 + y^2 + z^2 &= R^2 \\
  z &= R \cos \theta \\
  z \cos \theta_v - y \sin \theta_v &= R \cos \theta_{\text{jet}}
\end{align*}
\] (57)

The \(y\) coordinate of the intersections (see Fig. 31) is found to be

\[
y_i = \frac{\cos \theta \cos \theta_v - \cos \theta_{\text{jet}}}{\sin \theta_v} R
\] (58)

Consider now the annulus as a circle whose center lies on the \(z = R \cos \theta\) plane. Its radius is \(R \sin \theta\), and the angle \(\alpha\) that subtends \(S\) is \(\alpha = 2 \arccos (-y_i/R \sin \theta)\). The length \(l(\theta)\) is then

\[
l(\theta) = 2R \sin \theta \arccos \left(\frac{\cos \theta_{\text{jet}} - \cos \theta \cos \theta_v}{\sin \theta \sin \theta_v}\right)
\] (59)

Substituting Eqs. 56 and 59 into Eq. 55, I conclude that

\[
dF_S(\theta) = \frac{dF}{d\theta} d\theta \times \frac{1}{\pi} \arccos \left(\frac{\cos \theta_{\text{jet}} - \cos \theta \cos \theta_v}{\sin \theta \sin \theta_v}\right)
\] (60)

This is valid as long as the intersections between the annulus and \(S\) exist, i.e. for \(|\theta_v - \theta_{\text{jet}}| < \theta < \theta_v + \theta_{\text{jet}}\). Let me work out the remaining cases:

- if \(\theta_v < \theta_{\text{jet}}\), i.e. if line of sight is inside the jet border, then for \(\theta < \theta_{\text{jet}} - \theta_v\) also the annulus is inside the jet border, thus \(dF_S(\theta) = dF_a(\theta)\);
- if \(\theta_v > \theta_{\text{jet}}\), i.e. if line of sight is outside the jet border, then for \(\theta < \theta_v - \theta_{\text{jet}}\) the annulus is too small to intercept the jet border, thus \(dF_S(\theta) = 0\); in either case, if \(\theta > \theta_v + \theta_{\text{jet}}\) the annulus is too large to intercept the jet border, thus again \(dF_S(\theta) = 0\).

Summing up, we can define the function \(a(\theta, \theta_v, \theta_{\text{jet}})\) by

\[
a = \begin{cases} 
  H(\theta_{\text{jet}} - \theta_v) & \theta \leq |\theta_v - \theta_{\text{jet}}| \\
  0 & \theta \geq \theta_v + \theta_{\text{jet}} \\
  \frac{1}{\pi} \arccos \left(\frac{\cos \theta_{\text{jet}} - \cos \theta \cos \theta_v}{\sin \theta \sin \theta_v}\right) & \text{otherwise}
\end{cases}
\] (61)

where \(H(x)\) is the Heaviside function, i.e.

\[
H(x) = \begin{cases} 
  0 & x < 0 \\
  1 & x \geq 0
\end{cases}
\] (62)

and write

\[
dF_S(\theta, \theta_v, \theta_{\text{jet}}) = a(\theta, \theta_v, \theta_{\text{jet}}) \frac{dF}{d\theta} d\theta
\] (63)

The light curve of the pulse from the off-axis jet is then obtained by integration of this expression between \(\theta_{\text{on}}(t)\) and \(\theta_{\text{off}}(t)\), namely

\[
F(t, \theta_v, \theta_{\text{jet}}) = \frac{2\pi R^2}{d^2} \int_0^1 I'_0(\theta_{\text{on}}(t)) a(\theta, \theta_v, \theta_{\text{jet}}) \sin \theta d\theta (1 - \beta \cos \theta)^3
\] (64)
6.4 Pulse Light Curves and Time Dependent Spectra in the Shell-Curvature Model

Figure 32: The off-axis jet can be thought of as being part of an expanding sphere. For
simplicity, the EATS of the expanding sphere (hatched area) is represented as
in Figure 27, but it is actually the same as in Figure 28. The actual EATS of the
jet is the intersection between the jet surface and the sphere EATS.

Note that here $t = 0$ is the arrival time of the first photon from the sphere, thus if
$\theta_v > \theta_{\text{jet}}$ the actual light curve of the off-axis jet starts a little later. The actual start
time of the off-axis light curve is

$$t_{\text{start}}(\theta_v, \theta_{\text{jet}}) = \frac{R}{\beta c} \left( 1 - \beta \cos(\theta_v - \theta_{\text{jet}}) \right)$$  (65)

Equation 64 can be easily integrated with a simple numerical procedure. Some ex-
ample light curves computed using a RK4 integration scheme are given in Fig. 33.

Light Curves Figure 33 shows plots of bolometric light curves of the same
pulse seen at different viewing angles, computed by numerical integration of
Eq. 64 using a Runge-Kutta IV order scheme. Both the peak flux decrease and
the duration increase discussed in §6.5 and §6.5 are apparent. The overall shape is
qualitatively insensitive of the viewing angle, apart from the peak being sharper
in the on-axis case.

Spectra and hardness-intensity correlation

All the above arguments can be also applied to the derivation of the observed
spectrum. All we need to do is to compute the flux density

$$\frac{dF}{d\nu}(\nu, t) \equiv F_{\nu}(\nu, t) = \int_{S(t)} \frac{dI}{d\nu}(\nu, t) \cos \alpha \, ds/r^2$$  (66)

over the same EATS as before, using the transformation

$$\frac{dI}{d\nu}(\nu) = \delta^3 \frac{dI'}{d\nu'}(\nu/\delta)$$  (67)

to express the intensity density in terms of the comoving one. It is convenient to
write $dI'/d\nu'$ as follows

$$\frac{dI'}{d\nu'}(\nu') = \frac{I'}{\nu'_0} f(\nu'/\nu'_0)$$  (68)
Figure 33: Light curves of a pulse from a jet with $\theta_{\text{jet}} = 5^\circ$ and $\Gamma = 100$. Each curve refers to a different viewing angle in the sequence (from green to red) $\theta_v = 5^\circ, 5.2^\circ, 5.4^\circ, 5.6^\circ, 5.8^\circ$ and $6^\circ$. $F_{\text{max}}$ refers to the on-axis jet. The inset shows the same curves plotted with logarithmic axes.
where $I'_0$ is the total intensity, $\nu'_0$ is some frequency, and $f(x)$ is a function which describes the comoving spectral shape, and whose integral is normalized to unity. As an example, we can set a power law spectral shape

$$f(\nu'/\nu'_0) = (1 - \alpha) \left( \frac{\nu'}{\nu'_0} \right)^{-\alpha} \quad (69)$$

for $\nu' > \nu'_0$ and zero otherwise, with $\alpha > 1$. Since the integral of $f(x)$ is normalized to unity, we have

$$\int_0^\infty \frac{dI'}{d\nu'}(\nu') d\nu' = \frac{I'_0}{\nu'_0}$$ \quad (70)

The equation for the observed spectrum of a off-axis jet is then

$$\frac{dF}{d\nu}(\nu,t) = \frac{2\pi R^2}{d^2} I'_0 \left( 1 + \beta \right)^{2+\alpha} \left( 1 - \alpha \right) \nu^{-\alpha} \times \left\{ \left( 1 + \frac{t - t_{\text{off}}}{\tau + t_{\text{off}}} \right)^{-2-\alpha} - \left( 1 + \frac{t}{\tau} \right)^{-2-\alpha} \right\} \quad (71)$$

which reproduces the well-known $2 + \alpha$ decay slope due to high latitude emission (Kumar and Panaitescu, 2000; Dermer, 2004). For more general spectral shapes, a numerical approach is necessary to compute the integral in Eq. 71.

Since we are mainly interested in how the peak of the observed spectrum evolves with time, let me assume a simple form of the comoving spectral shape, namely

$$I'_\nu(\nu') = n(a,b) \frac{I'_0}{\nu'_0} \left[ \left( \frac{\nu'}{\nu'_0} \right)^{-a} + \left( \frac{\nu'}{\nu'_0} \right)^{-b} \right]^{-1} \quad (73)$$

where $n(a,b)$ is a normalization constant which depends upon the high and low spectral indices $a$ and $b$; clearly $I'_\nu \propto \nu'^a$ for $\nu' \ll \nu'_0$ and $I'_\nu \propto \nu'^b$ for $\nu' \gg \nu'_0$. If $a > 0$ and $b < -1$, the normalization $n(a,b)$ can be defined so that

$$I'_0 = \int_0^\infty I'_\nu(\nu') d\nu' \quad (74)$$

The break frequency $\nu'_0$ is related to the comoving $\nu'F'_\nu$ peak energy $E'_{\text{peak}}$ through

$$E'_{\text{peak}} = \left( \frac{a + 1}{b + 1} \right)^{\frac{1}{\alpha}} h \nu'_0 \quad (75)$$

where $h$ is Planck’s constant. All the examples in the figures assume the above comoving spectral shape, with $a = 0.2$ and $b = -1.3$, which represent average high and low spectral indices of Fermi GRB spectra (Nava et al., 2011b).
Figure 34: Spectra at different times of a pulse from a jet seen on-axis, with $\Gamma = 100$ and $\Delta R = R$. The comoving spectral shape is given in Eq. 73. The coloured circles in the inset show at which point in the pulse each spectrum (identified by the colour) was calculated.

Figure 35: Peak of the observed spectrum versus the bolometric flux, for a pulse with $\Gamma = 100$ and $\Delta R = R$. A clear hardness-intensity correlation is present. The slope of the black dashed line is $1/3$. The inset is the same as in Fig. 34.
Figure 34 shows spectra from an on-axis jet at six representative times, computed using Eq. 71. The evolution is clearly hard-to-soft (i.e. the peak energy decreases monotonically with time), and the low and high energy spectral indices are the same as those of the comoving spectrum. Figure 35 shows that after the peak of the light curve the peak energy $E_{\text{peak}}$ of the observed $\nu F_{\nu}$ spectrum varies with the bolometric flux $F$ following roughly $E_{\text{peak}} \propto F^{1/3}$, i.e. the model predicts a hardness intensity correlation with index $1/3$ during the decay of the pulse. Let me interpret these results:

1. **Pulse rise:** the maximum of $E_{\text{peak}}(t)$ is at the very beginning of the pulse, when only a small area pointing directly towards the observer (the “tip” of the jet at zero latitude) is visible. As the visible area increases, less beamed contributions from parts at increasing latitude come into sight, reducing $E_{\text{peak}}$ slightly.

2. **Pulse decay:** after the pulse peak, the tip of the jet turns off, causing $E_{\text{peak}}$ to drop. At this time the visible part of the jet is an annulus (see Fig. 28): the spectral peak is determined mainly by the maximum Doppler factor $\delta_{\text{max}}(t) = \Gamma^{-1} [1 - \beta \cos \theta_{\text{off}}(t)]^{-1}$ of the visible area, which corresponds to the innermost circle of the annulus, so that $E_{\text{peak}} \sim \delta_{\text{max}} E'_{\text{peak}}$. The flux $F$ in turn decreases approximately as $\delta_{\text{max}}^4$ times the angular size of the annulus. The latter is proportional to $\cos \theta_{\text{off}} - \cos \theta_{\text{on}}$, which can be shown to be

$$\cos \theta_{\text{off}} - \cos \theta_{\text{on}} \approx \frac{\Delta R}{R} \frac{1}{\beta \Gamma \delta_{\text{max}}}$$

As a result, we have that $F \propto \delta_{\text{max}}^3$, which explains why $E_{\text{peak}} \propto F^{1/3}$.

**Characteristics of pulses from an off-axis jet**

**A longer pulse duration**  If the jet is off-axis, relativistic beaming of the emitted radiation causes both the flux and $E_{\text{peak}}$ to be much lower than the on-axis counterparts. For the same reason, the duration of the pulse becomes longer. This can be understood intuitively as follows: as in the on-axis case, the jet surface is not seen to turn on all at the same time, but progressively from the nearest-to-the-observer point (point $A$ in Fig. 32) down to the farthest. The same holds for the turning off. Thus point $A$ is the first to turn on, and also the first to turn off. As a consequence, the effective emitting area increases as long as point $A$ is seen emitting, then it decreases. In other words, the peak time equals the emission time of point $A$, which is given by

$$t_{\text{peak}}(\theta_v, \theta_{\text{jet}}) = T \left[ 1 - \beta \cos(\theta_v - \theta_{\text{jet}}) \right]$$

thus its ratio to the on-axis peak time is

$$\frac{t_{\text{peak}}(\theta_v, \theta_{\text{jet}})}{t_{\text{peak}}} = \frac{1 - \beta \cos(\theta_v - \theta_{\text{jet}})}{1 - \beta}$$

Figure 36 shows a plot of this ratio as a function of $\theta_v - \theta_{\text{jet}}$ for different values of $\Gamma$. The off-axis pulse is thus intrinsically broader than its on-axis counterpart. The effective duration as seen by the observer, though, depends on the limiting flux and on the amount of overlap with other pulses.
A lower peak flux

The decrease of the pulse peak flux $F_p$ with increasing viewing angle can be understood as follows:

1. when the jet is observed on axis, the bulk of the flux comes from a ring of angular radius $1/\Gamma$ centred on the line of sight. Let me indicate this peak flux with $F^*$;

2. as long as $\theta_v < \theta_{jet} - 1/\Gamma$, we have that $F_p$ is essentially equal to $F^*$;

3. if $\theta_v = \theta_{jet}$, about half of the ring is still visible, thus $F_p$ is reduced to about $F^*/2$;

4. if $\theta_v$ is only slightly larger than $\theta_{jet}$, the flux is dominated by the contribution of the jet border, whose Doppler factor is $\delta_B = \Gamma^{-1} \left[ 1 - \beta \cos (\theta_v - \theta_{jet}) \right]^{-1}$, thus $F_p \propto \delta_B^4 F^*$;

5. as $\theta_v$ increases towards $\theta_v \gg \theta_{jet}$, the relative difference in Doppler factor between different parts of the jet is reduced, and the flux contributions of parts other than the border become increasingly important. This compensates in part the de-beaming of the jet border, the effect being more pronounced for larger jets, because the effective emitting surface area is larger.

Figure 36: Ratio of the off-axis pulse peak time $t_{peak}(\theta_v, \theta_{jet})$ to the on-axis pulse peak time $t_{peak}$. The jet half opening angle is $\theta_{jet} = 5^\circ$. Each curve refers to a different value of the Lorentz factor, from $\Gamma = 50$ to $\Gamma = 300$ with a step of 50.
Based on these considerations, an empirical analytical formula can be constructed to describe how the peak flux depends on the viewing angle \( \theta_v \) and on the jet half opening angle \( \theta_{\text{jet}} \). An example of such an empirical formula is

\[
F_p/F^* \approx \begin{cases} 
1 & \theta_v \leq \theta^*_{\text{jet}} \\
1 - \Gamma (\theta_v - \theta^*_{\text{jet}})/2 & \theta^*_{\text{jet}} < \theta_v \leq \theta_{\text{jet}} \\
\frac{1}{2} \left( \frac{\delta_B}{(1 + \beta) \Gamma} \right)^{(4 - 2 \theta^*_{\text{jet}}^{1/3})} & \theta_v > \theta_{\text{jet}} 
\end{cases}
\]

where \( \theta^*_{\text{jet}} = \theta_{\text{jet}} - 1/\Gamma \). The definition for \( \theta^*_{\text{jet}} < \theta_v \leq \theta_{\text{jet}} \) is just a linear decrease from \( F^* \) to \( F^*/2 \); the exponent of \( \delta_B \) in the definition for \( \theta_v > \theta_{\text{jet}} \) is 4 reduced by an amount\(^2\) which depends on \( \theta_{\text{jet}} \), in order to take into account the flux loss compensation explained in point (v) above.

The flux at time \( t \) of the pulse from an off-axis jet is given by the integral in Eq. 64, which however has no analytical solution for \( \theta_v > 0 \). The coloured solid lines in Figure 37 represent \( F_p \) as computed by numerical integration of Eq. 64 at \( t = t_{\text{peak}}(\theta_v, \theta_{\text{jet}}) \), for five jets with different half opening angles. The orange dashed lines are plots of Eq. 79 for the corresponding parameter values, showing that the best agreement is for half opening angles \( 5^\circ \leq \theta_{\text{jet}} \leq 10^\circ \).

---

\(^2\) The coefficient and exponent of \( \theta_{\text{jet}} \) in Eq. 79 have been chosen to get a good agreement with the results from the semi-analytical formulation developed in the preceding sections.
Figure 38: $E_{\text{peak}}$ at the pulse peak time for three jets with $R = 10^{13}$ cm, $\theta_{\text{jet}} = 5^\circ$ and three values of $\Gamma$, namely (from red to green) $\Gamma = 30$, 100 and 300. The black dashed lines are plots of $\delta B/2\Gamma$ for the corresponding values of $\Gamma$.

**Spectral Peak Energy, Hardness-Intensity Correlation**  With the same assumptions as in the on-axis case, I computed the spectra from the off-axis pulse at different times. The spectrum at each time is dominated by the part of the EATS with the strongest beaming. At time $t_{\text{peak}}$, such part is the border of the jet nearest to the observer, thus one expects $E_{\text{peak}}(t_{\text{peak}})$ to decrease with $\theta_v$ as the Doppler factor of the jet border, i.e. $E_{\text{peak}}(t_{\text{peak}}) \propto \delta B$. Figure 38 is a plot of $E_{\text{peak}}(t_{\text{peak}})$ for three values of $\Gamma$, obtained by using Eq. 71 to compute the spectra, and it shows that indeed $E_{\text{peak}}$ is approximately proportional to $\delta B$. In general, $E_{\text{peak}}$ is a little lower than $\delta B E'_{\text{peak}}$ because of the “blending in” of softer spectra from less beamed parts of the jet.

Figure 39 shows the evolution of $E_{\text{peak}}$ as a function of the flux $F$ during the pulse, for four different off-axis viewing angles. A “hardness-intensity” correlation during the pulse decay is still apparent, with a slightly steeper slope ($\sim 0.5$) just after the pulse peak, getting shallower as the flux decreases and eventually reaching $\sim 1/3$ as in the on-axis case.

**Multi-pulse Light Curves**

Now that we have a detailed (though simple) model of the single pulse, we can proceed to construct a “synthetic” GRB light curve by superposition of pulses. Some non-trivial features emerge from such superposition. Figure 40 shows four light curves of the same series of $N = 100$ pulses seen at four different viewing angles. All pulses are equal in duration and peak flux. Their starting times have been sampled from a uniform distribution within a 2 seconds time span. The jet parameters are $\Gamma = 100$, $\theta_{\text{jet}} = 5^\circ$, $R = 10^{13}$ cm and $\Delta R = R$. The viewing angles...
Figure 39: Logarithmic plot of $E_{\text{peak}}$ versus Flux of the same pulse seen at different off-axis viewing angles. The jet has $\theta_{\text{jet}} = 5^\circ$, $\Gamma = 100$ and $R = 10^{13}$ cm. The four series of points (from green to red) correspond to $\theta_v = 5.1^\circ$, $5.5^\circ$, $6^\circ$ and $7^\circ$. The inset shows the slope of the relation during the decay of the pulse for each viewing angle.

are $\theta_v = 0$, $\theta_{\text{jet}} + 1/\Gamma$, $\theta_{\text{jet}} + \sqrt{3}/\Gamma$ and $\theta_{\text{jet}} + 2/\Gamma$. The comoving spectral shape is the same as before (Eq. 73). The resulting light curves have been binned at 32 ms resolution for a better comparison with actual GRB light curves. For each light curve, the $E_{\text{peak}}$ of the spectrum in each time bin is also given (thin orange histograms). The following features should be apparent:

1. as the viewing angle increases, variability is smeared out by the pulse broadening;
2. the shape of the overall light curve tends to resemble a (long) single pulse when the viewing angle is large enough;
3. the superposition of pulses masks the hard-to-soft spectral evolution of the single pulses, turning it into an intensity tracking behaviour: this is due to the superposition of spectra with different peak energies;
4. the variation of $E_{\text{peak}}$ leads slightly the variation in flux, because of the hard-to-soft nature of the single pulses;
5. there is a general softening of $E_{\text{peak}}$ in time over the entire light curve.

These features are strikingly similar to those found in time resolved spectral analysis of real gamma-ray bursts (e.g. Ford et al., 1995; Ghirlanda, Celotti, and Ghisellini, 2002). I do not advocate this as a proof of the correctness of our model, which is certainly oversimplified, but rather as a further indication that some features of GRB light curves might be explained admitting that the jet is seen at
Figure 40: Light curves (thick histograms) and spectral peak evolution (thin orange histograms) of a sequence of 100 pulses from a jet with $\theta_{\text{jet}} = 5^\circ$, $\Gamma = 100$, $R = 10^{13}$ cm and $\Delta R = R$. The peak of the $\nu F_{\nu}$ comoving spectrum is $E'_p = 1$ keV. The pulse start times are sampled from a uniform distribution within the first 2 s of the observer time. Each panel refers to a different viewing angle in the sequence $\theta_v = 0$, $\theta_{\text{jet}} + 1/\Gamma$, $\theta_{\text{jet}} + \sqrt{3}/\Gamma$, $\theta_{\text{jet}} + 2/\Gamma$ (from left to right, top to bottom). $L_{\text{max}}$ refers to the peak luminosity of the on-axis light curve.
least slightly off-axis. The off-axis viewing angle favours the broadening and superposition of pulses, which is the necessary ingredient to some of the features enumerated above. It can also contribute in a simple way to explain why the slope of the hardness-intensity correlation changes from burst to burst, being influenced by the viewing angle (§6.5).

Figure 40 shows that the simple arguments outlined in §6.2 are valid not only if pulses are produced by point sources, but also in presence of an extended geometry.

**The number of off-axis short GRBs seen by Fermi/GBM**

We can obtain an estimate of the fraction of off-axis SGRBs in the observed population by the simplifying assumption that all jets share the same intrinsic properties, and that their flux in an observer band is uniquely determined by the viewing angle and the redshift. I assume that the majority of SGRBs are observed on-axis, and I choose the following parameters in an attempt to match the average properties of the on-axis population:

1. $E_{\text{peak},o} = 700 \text{ keV}$ as the typical (on-axis, rest frame) peak spectral energy;
2. $\alpha = -0.5$ and $\beta = -2.3$ as typical low- and high-energy spectral indices (Nava et al., 2011b);
3. the redshift distribution as given in Ghirlanda et al., 2016;
4. $L_0 = 2.5 \times 10^{52} \text{ erg s}^{-1}$ as the typical (on-axis) luminosity, which corresponds to the break of the broken power law luminosity function of Ghirlanda et al., 2016. This choice is motivated by the fact that if GRBs can be observed off-axis, then their luminosity function is indeed well described by a broken power law, with the break around the average on-axis luminosity (Pescalli et al., 2015);
5. since the result is sensitive to the assumed typical Lorentz factor $\Gamma$ and half-opening angle $\theta_{\text{jet}}$, I explore the cases $\Gamma = 50, 100$ and $300$, and $\theta_{\text{jet}} = 5^\circ$ and $10^\circ$.

I then define the effective luminosity $L(\theta_v)$ following Eq. 79, namely

$$L(\theta_v) = L_0 \times \begin{cases} 
1 & \theta_v \leq \theta_{\text{jet}}^* \\
1 - \Gamma (\theta_v - \theta_{\text{jet}}^*)/2 & \theta_{\text{jet}}^* \leq \theta_v < \theta_{\text{jet}} \\
\frac{1}{2} \left( \frac{\delta_B}{(1 + \beta) \Gamma} \right) \left( \frac{(4 - \sqrt{2}) \theta_{\text{jet}}^*}{\theta_v} \right)^{1/3} & \theta_v \geq \theta_{\text{jet}}
\end{cases}$$

with $\theta_{\text{jet}}^* = \theta_{\text{jet}} - \Gamma^{-1}$, and the effective peak energy

$$E_{\text{peak}}(\theta_v) = \frac{E_{\text{peak},o}}{1 + z} \times \begin{cases} 
1 & \theta_v \leq \theta_{\text{jet}} \\
\frac{\delta_B}{(1 + \beta) \Gamma} & \theta_v > \theta_{\text{jet}}
\end{cases}$$
Figure 41: Fraction of off-axis GRBs over the total within a given redshift. The curves represent an estimate of the fraction of SGRBs with redshift lower than \( z \) observable by Fermi/GBM (assuming a 64-ms peak photon flux \( P_{64} > 2 \text{ph s}^{-1}\text{cm}^{-2} \) in the 10–1000 keV band) whose viewing angle is larger than \( \theta_{\text{jet}} + \Gamma^{-1} \).
as in §6.5. With these assumptions and prescriptions, one can compute the observed rate of GRBs with a viewing angle in the range \((\theta_v, \theta_v + d\theta_v)\), in the redshift range \((z, z + dz)\), assuming a limiting photon flux \(p_{\text{lim}}\) in a given band, as

\[
\frac{dN}{d\theta_v \, dz} \, d\theta_v \, dz = \frac{\Psi(z)}{1 + z} \frac{dV}{dz} P(\theta_v, z, p_{\text{lim}}) d\theta_v \, dz
\]  

(82)

where \(P(\theta_v, z, p_{\text{lim}})\) is the viewing angle probability, \(dV/dz\) is the differential comoving volume, and the factor \(1 + z\) accounts for cosmological time dilation. The viewing angle probability is

\[
P(\theta_v, z, p_{\text{lim}}) = \begin{cases} 
\sin \theta_v & \theta_v \leq \theta_{v,\text{lim}}(z, p_{\text{lim}}) \\
0 & \theta_v > \theta_{v,\text{lim}}(z, p_{\text{lim}})
\end{cases}
\]

(83)

The limiting viewing angle \(\theta_{v,\text{lim}}\) corresponds (through Eq. 80) to the limiting luminosity \(L_{\text{lim}}\) computed as

\[
L_{\text{lim}} = 4\pi d_L^2 \, p_{\text{lim}} \int_0^\infty \frac{dN}{dE} \frac{E \, dE}{(1+z)E_{\text{high}} \, dE} - \int_0^{(1+z)E_{\text{low}}} \frac{dN}{dE} \, dE
\]

(84)

where \(E_{\text{low}} (E_{\text{high}})\) is the lower (upper) limit of the observer band, \(d_L\) is the luminosity distance, and \(dN/dE(E_{\text{peak}}, \alpha, \beta)\) is the rest frame spectrum.

Let me define the total rate \(\dot{N}_{\text{tot}}(<z)\) of observable GRBs within redshift \(z\) as the integral of Eq. 82 over redshift from 0 to \(z\) and over \(\theta_v\) from 0 to \(\pi/2\); similarly, the rate \(\dot{N}_{\text{off}}(<z)\) of off-axis GRBs within redshift \(z\) is the integral over redshift from 0 to \(z\) and over the viewing angle from \(\theta_v + \Gamma^{-1}\) to \(\pi/2\). Since we are interested in the ratio of these two quantities, we do not need to bother about the normalization.

In Fig. 41 I show the fraction of bursts with \(\theta_v > \theta_{\text{jet}} + \Gamma^{-1}\) at redshift lower than \(z\) for various choices of \(\Gamma\) and \(\theta_{\text{jet}}\), assuming a limiting flux \(p_{\text{lim}} = 2 \text{ ph s}^{-1} \text{ cm}^{-2}\) in the 10-1000 keV band. Standard flat ΛCDM cosmology was assumed, with Planck parameters \(H_0 = 67.3 \text{ km s}^{-1} \text{ Mpc}^{-1}\) and \(\Omega_{m,0} = 0.315\) (Planck Collaboration et al., 2013). These results clearly indicate that at low redshift a large fraction of SGRBs is likely seen off-axis.

THE TIME DELAY BETWEEN GW170817 AND GRB170817A: TOO SHORT FOR AN OFF-Axis JET

The picture of off-axis GRBs developed in this chapter seems to be compatible with the interpretation of GRB170817A as an off-axis jet: it is much less luminous than average, it shows no significant variability, and it is at low redshift. As shown in §6.4.5, though, the delay with respect to the jet launch time (and thus the merger time in our case) increases with the viewing angle (see Eq. 65). I will show in what follows that it is hard to reconcile a \(\sim 2\) seconds delay time with emission from a jet seen under a large viewing angle.

In order to shine in gamma-rays, the jet must expand enough to become transparent.

Let me estimate the transparency radius \(R_t\) as (e.g. Daigne and Mochkovitch, 2002b)

\[
R_t = \frac{L_{K,\text{iso}} \sigma_T}{8\pi m_p c^3 \Gamma^3}
\]

(85)
If the gamma-ray prompt emission is due to an off-axis jet, the arrival time difference between the last gravitational waves and the first gamma-ray photons is dominated by the time it takes for the jet to become transparent in the observer frame.

where $\sigma_T$ is the Thomson cross-section, $m_p$ is the proton mass, and we are assuming an electron fraction of unity. $L_{K,iso}$ in this expression is the isotropic equivalent kinetic luminosity of the outflow, i.e. $L_{K,iso} = \Gamma M_{iso} c^2$. If we assume that the jet is launched a short time ($\ll 1$ s) after the merger, the arrival time difference between the latest gravitational waves and the first photons is

$$t_\gamma - t_{GW} \approx \frac{R_t}{\beta c} (1 - \beta \cos (\theta_v - \theta_{jet}))$$

This accounts for the fact that fluid elements on the jet border must travel up to $R_t$ at speed $\beta c$ before being able to emit the gamma-ray photons (see Fig. 42).

We can relate the kinetic luminosity of the jet to the on-axis gamma-ray luminosity by assuming that 10 percent of the kinetic luminosity is converted into photons, i.e. $L_{iso}(\theta_v = 0) = 0.1 L_{K,iso}$ (as in e.g. Kathirgamaraju, Duran, and Gianinnios 2017; see also Beniamini et al. 2015 who show that this is a typical conversion efficiency). I compute the corresponding off-axis luminosity $L_{iso}(\theta_v)$ using Eq. 80. If we require that $L_{iso}(\theta_v) = 10^{47}$ erg/s as in GRB170817A (Goldstein et al., 2017), and we assume a jet half-opening angle $\theta_{jet}$, we can compute the corresponding on-axis luminosity $L_{iso}(0)$ and arrival time delay $t_\gamma - t_{GW}$ for various combinations of the bulk Lorentz factor $\Gamma$ and off-axis viewing angle $\theta_v - \theta_{jet}$. Figure 43 shows the contours of these quantities for $\theta_{jet} = 0.2$ rad $= 11.5^\circ$ (the figure would be very similar for $\theta_{jet} = 0.1$ rad or $\theta_{jet} = 0.3$ rad). The red solid contour corresponds to the actual 1.7 s delay time as observed in GRB170817A, while the dotted contours represent the on-axis luminosities. The figure shows that a “standard” jet with a large Lorentz factor $\Gamma \gtrsim 70$ and an on-axis luminosity $L_{iso}(0) \sim 10^{51}$ erg/s is formally compatible with the observed time delay, but it requires fine-tuning of the viewing angle, which implies an extremely small probability. Moreover, the afterglow in this case would be very bright and visible very early. A structured
Figure 43: The solid red and dashed brown lines represent contours of the gamma-ray photon arrival time delay with respect to the last gravitational waves for a given jet bulk Lorentz factor $\Gamma$ and off-axis viewing angle $\theta_v - \theta_{jet}$. The dotted lines represent contours of the on-axis jet luminosity $L_{iso}(0)$ corresponding to an off-axis luminosity of $L_{iso}(\theta_v) = 10^{47} \text{erg/s}$. 
jet with slower (but still narrow) wings moving at $\Gamma \sim 10$ does not solve the issue, since the viewing angle cannot be large. A “slow” fireball with a Lorentz factor around $\Gamma \approx 7$ and a much smaller energy, instead, would give the correct time delay both if seen on-axis or a few degrees off-axis. For these reasons, I believe that GRB170817A is more naturally explained by slow material moving towards the observer (which is the case for the jet cocoon, for a structured jet whose structure extends at large angles, and for the isotropic giant flare described in the next chapter) than by a fast off-axis jet.
THE JET-LESS FIREBALL SCENARIO

The possible interpretations of GRB170817A mentioned in the preceding chapters all require the presence of a jet. In this chapter I investigate the possibility to have a detectable isotropic emission in hard X-rays in NS-NS and BH-NS mergers without a jet. I am guided by the fact that isolated magnetars can produce giant flares with $E_{\text{iso}} \sim 10^{46} \text{ erg}$ (e.g. Hurley et al., 2005; Lazzati, Ghirlanda, and Ghisellini, 2005) in non-catastrophic events, probably due to some re-configuration of their magnetic field (Thompson and Duncan, 1995). In NS-NS and BH-NS mergers, the gravitational energy available during the last phase of the coalescence is more than $E_G \sim 10^{53} \text{ erg}$, and only a very small fraction of this needs to be used.

To explore this possibility, I postulate that a small fraction (e.g. less than 0.1%) of $E_G$ can be used, e.g. through magnetic field amplification and subsequent conversion of magnetic field energy into thermal and/or kinetic energy, during the initial phase of the coalescence. The magnetic field in the NS material during the merger can be amplified to values larger than $B \sim 10^{15} \text{ G}$ (Ruiz and Shapiro, 2017; Giacomazzo et al., 2015; Kiuchi et al., 2014; Zrake and MacFadyen, 2013; Price, 2006). This is enough to form a fireball, that can produce an important isotropic emission in hard X-rays, as first suggested by Zrake and MacFadyen, (2013) and Giacomazzo et al., (2015).

Various recent numerical simulations (e.g. Ruiz and Shapiro, 2017; Murguia-Berthier et al., 2017; Just et al., 2016) and theoretical works (e.g. Margalit, Metzger, and Beloborodov, 2015) suggest that the conditions for launching a relativistic jet after a NS-NS or BH-NS merger are not always satisfied, thus the fraction of mergers without jets might be significant. An alternative mechanism to produce a prompt, isotropic, high-energy component is thus particularly relevant to the interpretation of GRB170817A.

SET UP OF THE ISOTROPIC FIREBALL MODEL

The binding energy difference between an isolated neutron star and a black hole of the same mass is

$$\Delta E_G \sim \frac{3}{10} M_{\text{NS}} c^2 \left( 1 - \frac{R_S}{R_{\text{NS}}} \right) \sim 5 \times 10^{53} \text{ erg}$$

where $R_S$ is the Schwarzschild radius, $M_{\text{NS}}$ and $R_{\text{NS}}$ are the mass and radius of the NS, and the numerical value assumes $M_{\text{NS}} = 1.4 M_\odot$ and $R_{\text{NS}} = 14 \text{ km}$.

This energy is of the same order as that released in a supernova explosion. In that case, most of it goes into neutrinos. In a compact binary merger it powers several other processes, such as emission of gravitational waves, dynamical ejection of matter and magnetic field amplification. The latter process is widely believed to produce a magnetic field of order $B \sim 10^{15} - 10^{16} \text{ G}$, which in turn can contain an energy of order $\sim 10^{51} \text{ erg}$ (Giacomazzo et al., 2015).
As first suggested by Zrake and MacFadyen, 2013, if just 1 percent of this energy, \( E_0 \sim 10^{49} \text{ erg} \), is converted into photons in a relatively baryon–free volume surrounding the merger (e.g. with a mechanism similar to that responsible for giant flares in magnetars, see e.g. Thompson and Duncan, 1995), a fireball initially dominated by electron-positron pairs can form and produce a short, high-energy transient.

From this point on, the evolution resembles that of the classical, standard isotropic fireball of gamma-ray bursts: the fireball accelerates to relativistic speed, up to some saturation radius \( R_a \) where the bulk kinetic energy becomes of the same order as the initial internal energy. Thereafter, the fireball coasts with constant velocity. At some point, the transparency radius \( R_t \) is reached, after which the fireball can release radiation.

The origin of the emission is still a controversial issue in the field of GRBs. On one hand, the release of “relic” thermal photons (i.e. the photons that constituted the initial source of pressure of the fireball, diluted by the expansion) at the transparency (photospheric) radius is expected (Meszaros and Rees, 2000). On the other hand, in the case of GRBs the observed spectra suggest a non-thermal origin of the radiation, which could be the result of internal shocks (Rees and Meszaros, 1994) or reconnection of the carried magnetic field (Thompson, 1994) transforming part of the kinetic energy of the fireball back into radiation.

These processes may be present in our isotropic case as well, but let me first consider the thermal photospheric radiation.

The temperature of the initial blackbody can be estimated by equating the energy density of photons to that of the source magnetic field, i.e. \( aT_0^4 \sim B^2/(8\pi) \), giving \( T_0 \approx 4.8 \times 10^{10} B^{1/2} \text{ K} \). The acceleration ends when the bulk Lorentz factor equals \( E_0/(\gamma M c^2) \), i.e. when \( \Gamma \approx 11 E_0,49/M,27 \) (here and in what follows I employ the usual notation \( Q_x \equiv Q/10^x \) in cgs units). This occurs at \( R_a = \Gamma R_0 \), and there the comoving temperature is \( T'_a = T_0(R_0/R_a) \). Beyond \( R_a \) the temperature decreases as \( T' = (R/R_a)^{-2/3} \) and \( \Gamma \) is constant. During these phases, the main radiative processes (pair creation and annihilation and Compton scatterings) conserve the number of blackbody photons. This implies that the total energy contained in this “fossil” thermal component is

\[
E_{BB}(R_t) \approx E_0 \Gamma \frac{T'(R_t)}{T_0} = E_0 \left[ \frac{R_t}{R_a} \right]^{-2/3}
\]

(88)

The transparency radius depends on the thickness of the fireball (Daigne and Mochkovitch, 2002b). In our case the fireball is thin (\( R_t/2\Gamma^2 > c t_{inj} \)), where \( t_{inj} \approx 10^{-3} \text{ s} \), \( \Gamma \approx 10 \) and \( R_t \approx 10^{12} \text{ cm} \) so that \( R_t \) is given by (see Eq. 15 of Daigne and Mochkovitch 2002b):

\[
R_t = \left[ \frac{Y_e \sigma_T E_0}{4\pi m_p c^2 \Gamma} \right]^{1/2} \approx 1.9 \times 10^{12} \left[ \frac{Y_e-1 E_0,49}{\Gamma_1} \right]^{1/2} \text{ cm}
\]

(89)

where \( Y_e = n_e/(n_p + n_n) \) is the electron fraction of the fireball at the transparency. In this case the final blackbody energy (Eq. 88) can be written in terms of the free parameters as:

\[
E_{BB} = E_0 \frac{T_{BB}}{T_0} \approx 1.2 \times 10^{46} E_0,49^{5/3} R_0,7^{2/3} T_0,11 M_27,1^{-1/3} Y_e^{-1/3} \text{ erg}
\]

(90)
and the observed temperature is
\[ k_B T_{BB} = k_B T_0 \frac{R_0}{R_{1/3} t_{2/3}^{2/3}} \approx 10 E_{0,49}^{2/3} R_{0.7}^{2/3} T_{0,11}^{1} M_{27}^{-1} Y_{e,-1}^{-1/3} \text{keV} \] (91)

With the fiducial values of the parameters, this emission is a very small fraction of the initial energy. As discussed above, there can be another mechanism able to convert a larger fraction of the bulk kinetic energy into radiation, just as in GRBs. Since there is no general consensus about this process, I leave it unspecified. This ignorance is encapsulated in the efficiency parameter \( \eta \), such that the energy released in this additional, non-thermal component is:

\[ E' = \eta E_0 \] (92)

With \( \eta = 10^{-2} \eta_2 \) we have \( E' = 10^{47} \eta_2 E_{0,49} \text{erg} \). Assuming a limiting sensitivity of \( 10^{-7} \text{erg cm}^{-2} \), such a fireball can be detected up to a distance of \( \sim 90 \text{Mpc} \), close to the current horizon of LIGO/Virgo. This holds assuming that most of the electromagnetic radiation falls into the observed band of the current instruments (i.e. between 10 and 1000 keV). This requires a radiation mechanism not only able to transform a fraction \( \eta > 10^{-2} \) of \( E_0 \) into radiation, but also able to do that in the hard X-rays. Quasi thermal Comptonization and/or synchrotron emission are the first candidates, as they are in “standard” GRBs.

Delay and pulse duration

A point not always appreciated, in the GRB standard fireball theory, is that the delay time between the initial formation and the arrival of the fireball at the transparency radius \( R_t \) is equal to the duration of the pulse, if the fireball is thin (namely its width \( \Delta R \ll R_t \)).

Indeed, for a thin fireball the pulse duration is comparable to the angular time-scale

\[ t_{\text{ang}} = \frac{R_t}{2 \Gamma^2 c} \approx 0.38 \ E_{0,49}^{-2} M_{27}^{5/2} Y_{e,-1}^{1/2} \text{s} \] (93)

The delay with respect to the initial formation, on the other hand, is given by:

\[ \Delta t_{\text{delay}} = \frac{R_t}{\beta c} (1 - \beta) \approx \frac{R_t}{2 \Gamma^2 c} \] (94)

The two time scales are thus equal. In general, in GRBs the delay time is not measurable, but in the case of the detection of gravitational waves, it can be taken as the time difference \( \Delta t_{\text{delay}} \) between the merger and the arrival of the first photons of the prompt.

Afterglow

After the prompt emission phase, a cold shell is left, which continues its expansion at a constant rate, sweeping the circum-binary medium (CBM). The interaction with the CBM produces a shock, where electrons are accelerated and give rise to an afterglow by synchrotron radiation. The afterglow rises as \( t^2 \) at all frequencies (Sari and Piran, 1999) until the shell has swept up a CBM rest mass comparable
Figure 44: Indicative afterglow light curves in radio (5 GHz), optical (R band) and X-ray (1 keV) for the three scenarios. Parameters in Table 5. Left panel: isotropic fireball. Mid panel: the spine/layer model (which I consider equivalent also to the jet plus cocoon scenario), where the spine is the assumed to have parameters typical of a standard SGRB jet. The layer contributes to the light curve at early times, since $\theta_{\text{layer}} = \theta_{\text{view}} = 30^\circ$. Right panel: off-beam slow jet model. The dotted lines show representative detection limits for the three bands. The corresponding flux values are shown on the figure. The optical and X-ray fluxes are uncertain, since they depend on the presence of high energy electrons. For all models, the behaviour of the light curve is almost the same regardless of the observer frequency: this is due to all three frequencies being between $\nu_m$ and $\nu_c$ (for the chosen parameters) at all times shown on the plot, except for the early rising part of the off-beam slow jet and of the core.

to its own energy, after which it starts to fade as the shell decelerates (Nava et al., 2013). The afterglow peak time is given by

$$t_{\text{onset}} \approx 1.6 \times 10^{1.3} \times 10^{-1.3} \times 10^{-8/3} \times d \ (95)$$

COMPARISON OF THE AFTERGLOW EMISSION FROM THE THREE SCENARIOS

It is instructive to compare the general features of the afterglow of the isotropic fireball with the other competing models able to explain the prompt emission of sub-energetic short GRBs, i.e. the quasi-isotropic cocoon plus relativistic jet model by Lazzati et al., (2017a) and Lazzati et al., (2017b) which is essentially equivalent to the structured jet model by Kathirgamaraju, Duran, and Giannios, (2017), and the off-beam slow jet of Pian et al., (2017), namely a rather weak and homogeneous jet, with moderate $\Gamma$, seen off-beam. In order to describe the general behaviour of the afterglow in the three scenarios, I construct example light curves with repre-
sentative parameters. For all models I use the same microphysical parameters (see Table 5), and I assume that the number density of interstellar medium (ISM) is constant and equal for all models. For definiteness, I locate the source at a luminosity distance $d_L = 60\,\text{Mpc}$ which is comparable to the LIGO interferometer network range at the end of the O1 observing run (Abbott et al., 2016) and represents a conservative estimate of the range during O2. This is also the limiting distance at which a strong magnetar giant flare as the one described in Lazzati, Ghirlanda, and Ghisellini, (2005) (which I take as a prototype for the isotropic fireball scenario) can be detected by a Fermi-like instrument, assuming a limiting fluence of $10^{-7}\,\text{erg}\,\text{cm}^{-2}$. I compute the light curves of the jetted components using the public code BOXFIT (Eerten, Horst, and MacFadyen, 2012), and those of the isotropic component with SPHEREFIT (Leventis et al., 2012). Since neither code accounts for the dynamics before the deceleration time, I effectively correct the light curves by smoothly joining, at the peak time given by Eq. 95, a power law rising as $t^2$ at all frequencies (Sari and Piran, 1999).

**Isotropic fireball**

Fig. 44 shows (left panel) a representative afterglow (in the radio, optical and X-rays) of the isotropic fireball. The onset time is close to 5 days, after which the flux decays as a power law. If a jet is not present, this is all we see. If there is a jet with standard parameters (middle panel), it becomes visible at later times ($\sim 100$ days), when it has slowed down so that $1/\Gamma \gtrsim (\theta_{\text{view}} - \theta_{\text{jet}})$, i.e. when we start to see its border. Its emission in the optical and X-rays depends on the uncertain presence, at late times, of very high energy electrons. Emission in the radio, instead, is more secure. This late emission can be used as a diagnostic to distinguish between the jet-less, isotropic fireball scenario and the cocoon model by Lazzati et al., (2017b), which has similar features, but requires the presence of a jet.

**Jet plus cocoon**

To my understanding, the mildly relativistic layer/sheath described in the structured jet model of Kathirgamaraju, Duran, and Giannios, (2017) is essentially the same as the cocoon described in Lazzati et al., (2017a). I thus consider the core-
layer and the jet-plus-cocoon scenarios as equivalent. The middle panel of Fig. 44 mimics the expected afterglow from this scenario, obtained by summing the emission from a fast, narrow jet and that from a slower, wider jet, which represents the layer or the cocoon (parameters in Table 5). Since the latter has a wider half-opening angle, it is more likely seen within its beam. The layer afterglow peaks earlier (∼5 days for the chosen parameters, which are the same as for the isotropic fireball), with the spine contributing after ∼100 days. As studied in Rossi et al., (2004), in this case a rather strong linear polarization should be present around the time when the light curve peaks, in contrast with the isotropic fireball scenario. This can be used as a diagnostic.

Off-beam slow jet

The energy distribution of SGRB jets may have a low energy tail, accompanied by a corresponding low $\Gamma$ tail. In this case a jet with $\theta_{\text{jet}} \sim 10^\circ$, $E_{k,\text{iso}} \sim 10^{51}$ erg and $\Gamma \sim 5$–15 could be seen also at a relatively large $\theta_{\text{view}} \sim 30^\circ$. If seen on-axis, such a jet would produce $E_{\gamma,\text{iso}} = \eta E_{k,\text{iso}} = 10^{50} \eta^{-1} E_{k,\text{iso},51}$ erg. These values of $E_{k,\text{iso}}$ and $\Gamma$ are roughly consistent (i.e. they are within the rather large dispersion) with the relation shown in Ghirlanda et al., (2012) and Liang et al., (2013). Using Eq. 2 and Eq. 3 in Ghisellini et al., (2006b) one can calculate the observed $E_{\gamma,\text{iso}}$ for any $\theta_{\text{view}}$. It turns out that for $\theta_{\text{view}} = 30^\circ$, the de-beaming factor is $1/2500$, and then $E_{\gamma,\text{iso}}(30^\circ) \approx 4 \times 10^{46} \eta^{-1} E_{k,\text{iso},51}$ erg, detectable up to ∼60 Mpc if the fluence limit is $10^{-7}$ erg cm$^{-2}$. For $\theta_{\text{view}} > 30^\circ$ the de-beaming makes the source undetectable. The probability to see a burst within 30$^\circ$ is $P = (1 - \cos 30^\circ) \sim 0.13$; small, but not impossible. Taking into account the anisotropy of the GW emission (Schutz, 2011), the probability to see the jet within 30$^\circ$ after the progenitor NS-NS binary has been detected in GW is significantly larger$^1$, being ∼0.38. With the parameters listed in Table 5, I calculated the expected afterglow, shown in the right panel of Fig. 44. The peak flux corresponds approximately to the time when the Lorentz factor $1/\Gamma \sim (\theta_{\text{view}} - \theta_{\text{jet}})$, namely when we start to see the border of the jet. This occurs at $t_{\text{peak}} \sim 60$ days for the parameters shown in Tab. 5, and it goes as $t_{\text{peak}} \propto (E_{k,\text{iso}}/n_0)^{1/3}$ for a given viewing angle, i.e. it is independent from the initial Lorentz factor, due to the self-similar nature of the deceleration (Blandford and McKee, 1976).

After $t_{\text{peak}}$, the flux decreases monotonically. After the peak, the light curve is similar to an isotropic fireball with the same $E_{k,\text{iso}}$ and initial $\Gamma$. However, there is an important difference: the flux of an off-beam jet should be strongly polarized at $t_{\text{peak}}$, because the observer sees only the border of the jet (Rossi et al., 2004).

---

$^1$ This relies upon the assumption that the jet is launched perpendicular to the orbital plane of the binary. This most likely holds in NS-NS mergers, while it is less certain for BH-NS mergers, because the BH spin might cause the accretion disk plane in the post-merger phase to be tilted with respect to the original orbital plane.
Let me now see if the isotropic fireball model is able to account for the observed features of GRB170817A. In order to estimate the model parameters, I performed a Markov Chain Monte Carlo using a simple chi-squared likelihood, employing the emcee Python package (Foreman-Mackey et al., 2013). To reduce the dimensions of the parameter space, I fixed $\epsilon_e = 0.1$, following the evidence (Nava et al., 2014; Beniamini and Horst, 2017) that in GRB afterglows this parameter clusters around this value. The set of free parameters is thus $\{E_0, M, R_0, T_0, Y_e, n, \epsilon_B, p\}$. The observables are:

1. the duration $T_{90} = 2.0 \pm 0.5$ s of the prompt emission and the delay (~1.74 s) between the merger and the prompt emission. Both are predicted by the model to be approximately equal to the angular timescale $t_{\text{ang}}$ given by Eq. 93;

2. the temperature of the blackbody component in the prompt emission, i.e. $k_B T = 10.3 \pm 1.5$ keV, to be compared with the prediction given by Eq. 91;

3. the energy in the blackbody component, namely $(1.3 \pm 0.23) \times 10^{46}$ erg as estimated in §5.2, to be compared with Eq. 90;

4. The observed flux of the afterglow, which I model as described in the preceding section. The radio and X-ray observations, published in the literature, that I use to constrain the model parameters are listed in Table 6.

The log-likelihood I use for the fit is given by

$$\ln L = -\frac{1}{2} \sum_{i=1}^{N} \left[ \left( \frac{P_i - O_i}{\sigma_i} \right)^2 - \ln (2\pi \sigma_i^2) \right]$$  \hspace{1cm} (96)

where the index $i$ runs over the observables listed above, the $O_i$ and $\sigma_i$ represent respectively the measurements and their one-sigma uncertainties (assumed symmetric), and $P_i$ represent the model predictions. I assume uniform priors on $\log(E_0),\log(M),\log(R_0),\log(T_0),\log(n),\log(Y_e),\log(\epsilon_B),p$, and I impose the following constraints:

- $\Gamma > 2$ to avoid breaking the fundamental assumption that the fireball is relativistic;

- $Y_e < 1$ to keep the fireball electric charge neutral, and $Y_e > 0.01$ since this is the lowest value of the electron fraction found in NS-NS merger simulations which account for neutrino interaction (e.g. Radice et al., 2016);

- $10^{-6}$ cm$^{-3} < n < 10^2$ cm$^{-3}$ to allow for the widest possible range of ISM number densities;

- $\epsilon_B < 0.2$;

- $2 < p < 4$ not to fall off the validity range of the synchrotron afterglow model used;

- $T_0 < 4.1 \times 10^4$ keV, which corresponds to $B < 10^{17}$ G.
### VLA radio observations

<table>
<thead>
<tr>
<th>Time</th>
<th>$\nu_{\text{obs}}$</th>
<th>$F_\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[days]</td>
<td>[GHz]</td>
<td>[$\mu$Jy]</td>
</tr>
<tr>
<td>15.37</td>
<td>6.2</td>
<td>&lt; 130</td>
</tr>
<tr>
<td>16.42</td>
<td>3.0</td>
<td>18.7 ± 6.3</td>
</tr>
<tr>
<td>16.48</td>
<td>6.2</td>
<td>28.3 ± 5.4</td>
</tr>
<tr>
<td>18.33</td>
<td>3.0</td>
<td>14.5 ± 3.7</td>
</tr>
<tr>
<td>19.35</td>
<td>6.2</td>
<td>15.9 ± 5.5</td>
</tr>
<tr>
<td>21.36</td>
<td>6.2</td>
<td>13.6 ± 2.9</td>
</tr>
<tr>
<td>22.36</td>
<td>3.0</td>
<td>22.5 ± 3.4</td>
</tr>
<tr>
<td>23.36</td>
<td>6.0</td>
<td>22.6 ± 3.4</td>
</tr>
<tr>
<td>24.26</td>
<td>3.0</td>
<td>25.6 ± 2.9</td>
</tr>
<tr>
<td>31.22</td>
<td>3.0</td>
<td>34.0 ± 3.6</td>
</tr>
</tbody>
</table>

### Chandra X-ray observations

<table>
<thead>
<tr>
<th>Time</th>
<th>Band</th>
<th>Flux</th>
</tr>
</thead>
<tbody>
<tr>
<td>[days]</td>
<td>[keV]</td>
<td>[$10^{-15}$ erg cm$^{-2}$s$^{-1}$]</td>
</tr>
<tr>
<td>9</td>
<td>0.3-10</td>
<td>4.0 ± 1.1</td>
</tr>
<tr>
<td>15</td>
<td>0.3-10</td>
<td>5.0 ± 1.0</td>
</tr>
</tbody>
</table>

Table 6: Radio and X-ray detections and upper-limits on the afterglow of GRB170817A used in this work. The X-ray spectral index is $\Gamma = 1.3 \pm 0.4$ in both observations. Radio measurements are from Hallinan et al., (2017); X-ray measurements are from Troja et al., (2017), whose analysis is compatible with Margutti et al., (2017) and Haggard et al., (2017). $^a$I exclude this single observation from the fit, since it seems to be affected by ionospheric scintillation, but I show it on the light curve plot in Figure 45.
Table 7: I summarize here the results of the parameter estimation procedure described in §7.3.

<table>
<thead>
<tr>
<th>Parameter estimation results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fireball parameters</strong></td>
</tr>
<tr>
<td>Kinetic energy              ( E_0 = (3.1^{+2.1}_{-1.3}) \times 10^{49} \text{ erg} )</td>
</tr>
<tr>
<td>Starting radius             ( R_0 = 54^{+15}_{-25} \text{ km} )</td>
</tr>
<tr>
<td>Initial temperature         ( T_0 = (2.6^{+1.5}_{-1.0}) \times 10^{11} \text{ K} )</td>
</tr>
<tr>
<td>Baryonic mass               ( M = (3.6^{+3.3}<em>{-1.1}) \times 10^{-6} \text{ M}</em>\odot )</td>
</tr>
<tr>
<td>Electron fraction           ( Y_e = 0.06^{+0.04}_{-0.04} )</td>
</tr>
</tbody>
</table>

| **Afterglow parameters**   |
| ISM number density         \( n = (0.9^{+0.8}_{-0.7}) \times 10^{-5} \text{ cm}^{-3} \) |
| Shock energy fraction to electrons \( \epsilon_e = 0.1 \) (fixed) |
| Shock energy fraction to magnetic field \( \epsilon_B = 0.09^{+0.08}_{-0.04} \) |
| Electron power-law index    \( p = 2.11^{+0.02}_{-0.02} \) |

| **Derived quantities**     |
| Fireball Lorentz factor     \( \Gamma = 5.0^{+1.2}_{-1.2} \) |
| Non-thermal emission efficiency \( \eta = (0.85^{+0.60}_{-0.35}) \times 10^{-3} \) |
| Magnetic field driving the fireball \( B = (3.0^{+4.5}_{-1.85}) \times 10^{16} \text{ G} \) |

**Results**

The results of the parameter estimation are summarized in Table 7 and in Figure 45, where a corner plot showing the posterior densities in the parameter space is shown. In the upper right panel of the same figure I show the afterglow light curves corresponding to the best fit values of the parameters (thick lines), which have been estimated as the means of the marginalized posteriors, along with 100 light curves corresponding to random posterior samples (thin lines) to give an idea of how the uncertainty on the parameters propagates to the light curves. The uncertainties reported in Table 7 correspond to the 16th and 84th percentiles of these posteriors.

The fit results are generally consistent with the physical picture proposed: the magnetic field is \( B \sim 3 \times 10^{16} \text{ G} \) (which matches the predictions by Zrake and MacFadyen 2013 and Giacomazzo et al. 2015); the mass of the fireball is very small \( (M \sim 3 \times 10^{-6} \text{ M}_\odot) \), it is located close to the merging binary \( (R_0 \sim 54 \text{ km}) \) and is highly neutron rich \( (Y_e \sim 0.06) \), all of which suggests it constitutes the very first dynamically ejected material at the beginning of the merger phase (see e. g. Radice et
Figure 45: The corner plot in the figure shows the marginalized posterior densities of all model parameters and the joint densities of all couples of parameters. The red lines and squares pinpoint the best fit values, estimated by taking the means of the marginalized posteriors. The upper right panel shows the radio and X-ray afterglow light curves of the model with the best fit parameters. Circles, squares and stars represent respectively the 3 GHz, 6 GHz and 1 keV flux densities corresponding to the observations listed in Table 6. The thick solid lines represent the light curves corresponding to the best fit parameter values, while the thin lines are light curves corresponding to 100 random posterior samples, that I draw to show how the uncertainty on the model parameters propagates to the light curves. The X-ray light curves and data are multiplied by $10^4$ for presentation purposes.
The Lorentz factor of the fireball after the accelerated expansion phase is \( \Gamma = E_0/Mc^2 \approx 5 \). At the photospheric radius \( R_t \approx 4 \times 10^{12} \) cm (Eq. 89), the fireball releases relic photons from the radiation dominated phase (which are seen in the thermal tail of the *Fermi*/GBM signal) plus non-thermal radiation from an additional mechanism (e.g. internal shocks or magnetic reconnection) which dissipates a tiny fraction \( \eta \sim 10^{-3} \) of the fireball energy. The cold shell left after the prompt emission contains a large number of free neutrons, which undergo (mainly \( \beta^- \)) decay after a while (the comoving time is approximately \( \Gamma \) times the observer time). The decay of neutrons takes place after the prompt emission (the comoving time at \( R_t \) is \( R_t/(c\Gamma) \sim 27 \) s) but long before the deceleration radius (the mean decay radius is \( 1.3 \times 10^{14} \) cm for a mean lifetime of 900 s), so that the afterglow should not be affected by the presence of neutrons (Beloborodov, 2003a). The afterglow fit requires the number density of the CBM to be very low (\( n \sim 10^{-5} \) cm\(^{-3} \): this can be caused by the binary being outside the host galaxy in our direction (which is supported by the lack of ISM absorption features in the optical spectra of the associated Kilonova, see Levan et al., 2017), or alternatively could be due to the binary being surrounded by a cavity blown by its own Poynting flux (Medvedev and Loeb, 2013).

### Future Light Curve Evolution

As shown in Figure 45, these results imply that the radio and X-ray light curves should stop rising between 30 and 50 days, after which they should decay approximately as \( t^{-0.9} \) (assuming no CBM density gradient is encountered by the blastwave as it expands, and provided that the microphysical parameters \( \epsilon_B, \epsilon_e \) and \( p \) remain constant throughout the afterglow evolution). The alternative scenarios, namely the off-axis (possibly structured) jet and the jet cocoon, should feature a shallower decay (see e.g. the figures in Granot et al., 2017; Hallinan et al., 2017; Troja et al., 2017) or even a further rise of the light curve (depending on the properties of the jet) and polarization should be present near the peak of the afterglow (Rossi et al., 2004), contrary to our case. Future radio and X-ray flux measurements, possibly complemented by polarimetry, can thus be used to falsify or support the discussed scenario.

### Can the Fireball Retain the Low Electron Fraction Up to the Transparency Radius?

A possible difficulty of the physical picture I described in the preceding sections comes from the very low electron fraction of the fireball. The small number of electrons per baryon is essentially needed to reduce the Thomson opacity of the fireball, allowing it to become transparent early enough to produce a short \( \sim 2 \) seconds prompt emission pulse, while having the right energy and Lorentz factor to produce the observed X-ray and Radio afterglow. Studies of the nuclear composition of relativistic outflows from compact objects (e.g. Pruet, Woosley, and Hoffman, 2003; Beloborodov, 2003b; Metzger, Thompson, and Quataert, 2008), though, show that lepton (i.e. electron and positron) and neutrino (in presence of a strong neutrino source) captures by free neutrons and protons can rapidly raise the outflow
electron fraction to $Y_e \sim 0.5$ short after the onset of the acceleration. Our fireball starts expanding at the very beginning of the merger, thus the neutrino flux is probably unimportant in our case, but the presence of a huge density of electron-positron pairs seems inevitable. Moreover, the reaction rates are very high because of the large starting temperature of the fireball, which would suggest that the de-neutronization proceeds very rapidly. A way out is represented by the possibility that leptons are degenerate, in which case the positron density is exponentially reduced, and the equilibrium electron fraction is shifted towards a neutron-rich composition (Beloborodov, 2003b). Such degeneracy happens when the density $\rho$ and temperature $T$ in the fireball satisfy

$$k_B T \lesssim 7.7 \left( \rho \frac{10^{11} \text{g cm}^{-3}}{10^{11} \text{g cm}^{-3}} \right)^{1/3} \text{MeV}$$

The starting temperature of our fireball (Table 7) is $T \sim 2.6 \times 10^{11} \text{K} = 22.4 \text{MeV}$, so that a density $\rho \gtrsim 2.5 \times 10^{12} \text{g cm}^{-3}$ is needed. This is not an unreasonable density for the shock-compressed neutron-star crust material which is ejected at the beginning of the merger, but I admit that it is unclear to me if these conditions are realistic. I need more time to study and think about the feasibility of this scenario.

An alternative way to avoid the de-neutronization would be a Poynting flux dominated outflow (Metzger, Thompson, and Quataert, 2008). In this case, the acceleration is not driven by internal pressure, but by the magnetic field (Drenkhahn, 2001; Drenkhahn and Spruit, 2002), so that the plasma does not need to be radiation-dominated at the beginning and no copious pair production is necessary.
CONCLUSIONS

The first ever double neutron-star merger witnessed by human kind, GW170817, has been also the first event ever to be observed through both GW and EM radiation. This clearly represents a huge scientific achievement, and marks the beginning of a new era in multi-messenger astronomy. The community is glowing with excitement, and I feel very lucky for being part of it right now. In the near future, all of us are required to work towards constructing the best possible understanding of this observation and its consequences. The quality of the observations is very high, and despite the very large number of articles that have been already published about them, there is still much to learn.

On the SGRB side, I would like to solve the question if GRB170817A can be interpreted as off-axis emission from a GRB jet, or as something else. The interpretation described in the previous chapter seems promising, despite the difficulties related to the low electron fraction, but new radio and X-ray observations might disprove it. As I have shown in this thesis, the presence or absence of a jet in GW170817 impacts directly on our understanding of the SGRB population and of the conditions that are needed for a jet to be launched after the merger. I am working right now on a quantitative assessment of how the luminosity function of SGRBs is modified by adding a slower, low-energy component, such as the jet cocoon. I am also working on a comprehensive model of the SGRB population which enables predictions on the rate of orphan afterglows that will be possibly detected in future large surveys.

On the EM follow-up side, GW170817 clearly represents a huge success of current strategies, especially of the “galaxy targeting” approach, and provided the first direct information on the observational appearance of the EM counterparts associated to double neutron-star mergers. This clearly impacts on future strategies. As the horizon of the GW detector network grows larger, the optimal strategies will certainly need to evolve accordingly. As an example, our galaxy catalogues are highly incomplete beyond 100 Mpc, so the galaxy targeting approach will soon become less effective. Moreover, GRB170817A would not have triggered Fermi if it had been twice as far away, so we must not expect the EM follow-up to be always so lucky. I will certainly devote some time in the near future to think about these issues, with the possibility to update and improve the “where and when” method.

For what concerns longer-term research plans, I would like both to continue working on EM counterpart physics and follow-up strategies, and also to expand my research to include other astrophysical transients, such as fast radio bursts and tidal disruption events. In a few years, I would like to find myself having a much more comprehensive understanding of high-energy processes, especially those linked to black-hole accretion, and be able to contribute with new ideas and interpretations in the upcoming era of extremely large facilities and surveys.


RINGRAZIAMENTI

Questi tre anni di dottorato hanno coinvolt, stravolto e trasformato ogni centimetro cubo del mio essere. Durante questo periodo straordinario mi sono trovato a sperimentare un sacco di cose nuove e sempre molto emozionanti. Alcune di queste sono documentate e visibili “dall’esterno”: ad esempio ho scritto e partecipato a un discreto numero di articoli, ho presentato il mio lavoro in molti luoghi, diverse decine di volte, di fronte a centinaia di persone provenienti da ogni angolo del globo, ho guidato tre studenti nel loro percorso di laurea triennale, ho tenuto lezioni all’Università, ho fatto laboratori e interventi di divulgazione in scuole di ogni grado e in altri luoghi di conoscenza, ho prodotto non so quanti resoconti del mio lavoro di ricerca e delle attività di dottorato, ho partecipato a scuole e workshop, ho passato diversi brevi periodi all’estero. Ci sono poi innumerevoli altre esperienze che non rimangono nella documentazione “ufficiale”, ma che conservo nei miei archivi interiori, come la meraviglia di alcune scoperte, il senso di ammirazione verso le persone straordinarie con cui ho lavorato e che ho conosciuto, la paura di confrontarmi con i pari e con me stesso, la fatica della competizione, molte discussioni animate, le frustrazioni, le risate...

È strano e sembra forse un po’ riduttivo pensare che tutto questo percorso possa essere suggellato da una tesi, come se un libretto del genere potesse contenere e attestare la sostanza di quello che significano questi anni. Mi rendo conto però che questa tesi, la sua preparazione e la sua discussione, sono elementi di un passaggio rituale che ha bisogno di prendere una forma esplicita e visibile. Senza uno spartiacque netto, il percorso di dottorato si confonderebbe con il prima e con il dopo: rischierebbe di non essere contenuto (deborderebbe) e di non essere esorcizzato (ne rimarrebbe il daimon frammentato nel mio cuore, nella mia pancia, nelle mie giornate); rischierebbe anche di non essere celebrato, non esprimendo a pieno il suo potenziale catartico e metamorfico.

Sono contento dunque di poter arrivare alla fine di questo cammino, e di potermi voltare indietro un momento, prendendo fiato. Ripercorrendolo mentalmente, vedo molte persone camminare accanto a me, per tratti più o meno lunghi.

Ilaria è stata la mia compagna di viaggio prima ancora di iniziare: era con me quando ho lasciato i lavori precedenti e ho colto l’occasione di poter tornare all’Università. Quando ho vacillato di fronte alle difficoltà legate alla ripresa degli studi, lei mi ha saputo comprendere e sostenere senza mai stancarsi, accompagnandomi attraverso passaggi anche molto dolorosi, incollabile e coraggiosa come Dûrga. Ha battuto al mio fianco contro i miei demoni e le mie paure, come se fossero stati i suoi, e credo che possiamo dire di aver vinto molte grandi battaglie insieme. Durante il dottorato, ha visto cambiare radicalmente le mie abitudini, ha accectato la lontananza imposta dalle missioni in giro per il mondo, ha fatto da sponda a tutti i dubbi e le questioni esistenziali che un lavoro come questo apre continuamente, ha ascoltato con pazienza interminabili spiegazioni di strani fenomeni lontani, a volte difficili da immaginare e spesso incredibili, cogliendone il lato affascinante e suggestivo. So di non esagerare se dico che senza di lei non sarei mai arrivato fin qui. Grazie Ilaria, anche se so che non potrò mai ringraziarti abbastanza.
Giancarlo ha saputo guidarmi e accompagnare la mia crescita, umana e scientifica, con una sensibilità, un ascolto e una pazienza inestimabili. È stato un punto di riferimento, un mentore, un amico, e un grande esempio. Il suo è stato un regalo raro e prezioso: mi ha fatto sentire libero di essere me stesso, di accettare anche le mie debolezze, facendole diventare spunti per una crescita ulteriore. Grazie Giancarlo.

Gabriele è una forza della Natura, uno tsunami, capace di portare la sua energia, il suo entusiasmo, la sua visione penetrante del mondo, la sua genialità con l’umiltà e l’ironia di una persona illuminata. In questi anni ho avuto il privilegio di confrontarmi con lui su questioni scientifiche, politiche, filosofiche ed esistenziali, imparando ogni volta molto più di quanto potessi mai immaginare. Grazie Gabriele.

Monica mi ha dimostrato come si possa unire una grande conoscenza, un’ampia visione scientifica, una grande capacità di insegnare e trasmettere, e una genuina dolcezza. Ci vorrebbero più persone come lei – competenti e umane, idealiste, ma anche pragmatiche – nel mondo, non solo quello della ricerca. In questi anni ha sempre mostrato di credere in me, dandomi fiducia anche quando ero più in difficoltà. Grazie Monica.

Marica ha portato la sua vitalità radiosa nel mio ultimo anno di dottorato, diventando per me un punto di riferimento e aiutandomi tantissimo durante due passaggi: l’entrata nella collaborazione LIGO/Virgo, e la fine del dottorato. È stato un aiuto pratico, ma anche umano, indispensabile. Il suo esempio mi mostra l’esistenza di una rete di persone che non vivono la scienza come una competizione tra eroi overumanì, ma come un’impresa corale e comunitaria, in cui si può trovare anche un equilibrio con una propria vita più semplice e quotidiana. Grazie Marica.

Éric è un altro esempio di genialità e umanità che mi ha ispirato in questi anni. Mi ha accolto durante tre periodi a Parigi, accompagnando con pazienza la mia entrata nel mondo delle onde gravitazionali, dando valore alle mie idee e aiutandomi a svilupparle. Ma non è stato solo un mentore scientifico: il suo sorriso, accanto a quelli di Thibault e di Maria, sono stati la chiave per scoprire un lato più dolce di Parigi, anche in momenti tristi. Grazie Éric.

Mery è stata la prima studentessa che ho avuto l’onore di seguire come relatore della sua tesi. È una delle persone più esplosivamente positive e luminose che abbia mai conosciuto, una supernova di entusiasmo, curiosità e intraprendenza. In quel percorso siamo cresciuti insieme, ed è bellissimo adesso vederla continuare un percorso di travolgente crescita personale e scientifica, senza mai perdere queste sue preziose qualità. Grazie Mery.

Tra le altre persone con cui ho condiviso fisicamente e simbolicamente gli spazi di questi tre anni, non posso non ringraziare Alessio, compagno di avventure e sventure dottoriali e parigine; Tullia, che assieme a Federico ha fatto da fulcro per creare un’atmosfera allegra di condivisione e collaborazione che ha coinvolto tutti gli studenti del corso di astrofisica, e ha permesso di realizzare degli spettacolari eventi di divulgazione al Meet Me Tonight; Federico, stella nascente della radioastronomia e della scultura del legno, paziente compagno di ufficio; Matteo Bormio, simpaticissimo orso che porta un po’ di rigore nel caos del secondo piano dell’U2; i “bulli” Massimo, Francesco, Beppo, Guido e Tommy, che scardinano l’aplomb del suddetto secondo piano con un po’ di sana goliardia, talvolta alcolica,
più spesso nerd; il “gruppo Swift” all’Osservatorio di Brera-Merate, le cui riunioni sono continua fonte di ispirazione e crescita, e dalle quali sono nate molte idee.

Ci sono poi alcune persone non direttamente coinvolte in questo percorso, ma che voglio ringraziare perché sono stati parte del percorso parallelo nella mia sfera più intima. Innanzitutto, Renata e Paolo, che sono i pilastri – i punti fermi – per me e Ilaria, ci sostengono quotidianamente, sono il nostro rifugio, ci donano serenità e un sacco di cose buone. Ringrazio mia mamma, Linda, che anche se siamo lontani mi ascolta e mi accompagna sempre, con la sua dolcezza, la sua creatività, la sua forza inarrestabile, la sua devozione e la sua saggezza. Ringrazio i miei fratelli Ram e Nara, che hanno avuto il coraggio di vivere una vita di musica e di viaggi, alla scoperta continua di sé stessi: sono sempre felice quando li posso abbracciare e ascoltare affascinato tutte le loro storie magiche. Ringrazio mio papà, Claudio, che va sempre più verso la dolcezza, e che mi ha trasmesso l’amore per la conoscenza. Ringrazio mia zia Marcella e mio cugino José per il loro prezioso sostegno, il loro esempio meraviglioso, il loro sprone a continuare.

Ringrazio i miei compagni musicali, Ivan, la Fra, Teo e Alby, per avermi aiutato, nonostante tutte le difficoltà di questi ultimi anni, a mantenere in vita la mia anima musicale, che voglio nutrire per sempre. Grazie amici.

Ringrazio le colorate persone del Cortile, Laura, Nadia, Manu e Marta, i nostri compagni di danza (Laura, Simone, Giancarlo, Mauro, Monica, Monica, Eleonora, Eleonora, Valerio, e tanti altri), e il nostro guru Urs perché con loro ho potuto prendermi cura del mio corpo e della mia anima, condizione indispensabile per poter portare avanti questo percorso.

Grazie a tutti i compagni di Servizio Civile Internazionale, Stefania, Francesca, Sara, Alberto, e tutti gli altri.

Grazie anche a tutti gli amici preziosi che non ho ancora nominato, come Laura e Diego, Diana e Ale, Manu, Laura, Lale e Roby, Pilo e Chiara, Marzia, Ishtiaq, Rici e Claudia, Chiara e Marcello, Chiara e Alberto, Fabrizia e Leonardo, Cristina e Paolo, Lucia e Andrea, e mi sto dimenticando sicuramente qualcuno, ma mancano dieci minuti alla scadenza per la consegna della tesi ed è meglio andare. Vi abbraccio tutti, siete la mia vita. Grazie.
This document was typeset using the typographical look-and-feel classicthesis developed by André Miede. The style was inspired by Robert Bringhurst’s seminal book on typography “The Elements of Typographic Style”. classicthesis is available for both \LaTeX{} and \LaX:

https://bitbucket.org/amiede/classicthesis/

Happy users of classicthesis usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

Final Version as of February 13, 2018 (classicthesis v 1.0).